

Terry D. Oswalt
Editor-in-Chief

Martin A. Barstow
Volume Editor

Planets, Stars and Stellar Systems

VOLUME 4

Stellar Structure
and Evolution



Springer Reference

Planets, Stars and Stellar Systems

Stellar Structure and Evolution

Terry D. Oswalt (Editor-in-Chief)

Martin A. Barstow (Volume Editor)

Planets, Stars and Stellar Systems

Volume 4: Stellar Structure and Evolution

With 315 Figures and 10 Tables

Editor-in-Chief

Terry D. Oswalt
Department of Physics & Space Sciences
Florida Institute of Technology
University Boulevard
Melbourne, FL, USA

Volume Editor

Martin A. Barstow
Department of Physics and Astronomy
University of Leicester
University Road
Leicester, UK

ISBN 978-94-007-5614-4 ISBN 978-94-007-5615-1 (eBook)
ISBN 978-94-007-5616-8 (print and electronic bundle)
DOI 10.1007/978-94-007-5615-1

This title is part of a set with
Set ISBN 978-90-481-8817-8
Set ISBN 978-90-481-8818-5 (eBook)
Set ISBN 978-90-481-8852-9 (print and electronic bundle)

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012953926

© Springer Science+Business Media Dordrecht 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Series Preface

It is my great pleasure to introduce “Planets, Stars, and Stellar Systems” (PSSS). As a “Springer Reference”, PSSS is intended for graduate students to professionals in astronomy, astrophysics and planetary science, but it will also be useful to scientists in other fields whose research interests overlap with astronomy. Our aim is to capture the spirit of 21st century astronomy – an empirical physical science whose almost explosive progress is enabled by new instrumentation, observational discoveries, guided by theory and simulation.

Each volume, edited by internationally recognized expert(s), introduces the reader to a well-defined area within astronomy and can be used as a text or recommended reading for an advanced undergraduate or postgraduate course. Volume 1, edited by Ian McLean, is an essential primer on the tools of an astronomer, i.e., the telescopes, instrumentation and detectors used to query the entire electromagnetic spectrum. Volume 2, edited by Howard Bond, is a compendium of the techniques and analysis methods that enable the interpretation of data collected with these tools. Volume 3, co-edited by Linda French and Paul Kalas, provides a crash course in the rapidly converging fields of stellar, solar system and extrasolar planetary science. Volume 4, edited by Martin Barstow, is one of the most complete references on stellar structure and evolution available today. Volume 5, edited by Gerard Gilmore, bridges the gap between our understanding of stellar systems and populations seen in great detail within the Galaxy and those seen in distant galaxies. Volume 6, edited by Bill Keel, nicely captures our current understanding of the origin and evolution of local galaxies to the large scale structure of the universe.

The chapters have been written by practicing professionals within the appropriate sub-disciplines. Available in both traditional paper and electronic form, they include extensive bibliographic and hyperlink references to the current literature that will help readers to acquire a solid historical and technical foundation in that area. Each can also serve as a valuable reference for a course or refresher for practicing professional astronomers. Those familiar with the “Stars and Stellar Systems” series from several decades ago will recognize some of the inspiration for the approach we have taken.

Very many people have contributed to this project. I would like to thank Harry Blom and Sonja Guerts (Sonja Japenga at the time) of Springer, who originally encouraged me to pursue this project several years ago. Special thanks to our outstanding Springer editors Ramon Khanna (Astronomy) and Lydia Mueller (Major Reference Works) and their hard-working editorial team Jennifer Carlson, Elizabeth Ferrell, Jutta Jaeger-Hamers, Julia Koerting, and Tamara Schineller. Their continuous enthusiasm, friendly prodding and unwavering support made this series possible. Needless to say (but I’m saying it anyway), it was not an easy task shepherding a project this big through to completion!

Most of all, it has been a privilege to work with each of the volume Editors listed above and over 100 contributing authors on this project. I’ve learned a lot of astronomy from them, and I hope you will, too!



January 2013

Terry D. Oswalt
General Editor

Preface to Volume 4

Advances in the technology available to modern astronomers have led to a knowledge explosion. The study of the Sun and other Stars has been transformed by astronomers' access to both modern, electronically instrumented, telescopes and high-performance computers for theoretical modeling and calculations. Many of the topics we write about now were unknown before these tools were available. For example, white dwarfs had barely been studied, neutron stars were just a theoretical idea before the discovery of radio pulsars, and seismology was a topic confined to the study of the Earth's interior. Gamma ray bursts were certainly going off, but the Vela satellites that led to their discovery were a decade away and the release of that "top-secret" information later yet.

The fundamental aim of this volume is to give a comprehensive overview of the current state of our knowledge of stars and to provide a reference of broad use to researchers who may not be experts in a particular subfield. It is hoped that the individual chapters will provide valuable stepping off points for further reading and pursuit of each topic. There is much more to include than before and I am grateful to all the authors for their concise and effective writing.

This volume has been divided into three main sections. Part I broadly covers the theory of stellar structure, while Part II deals with the evolution of stars from their formation through to their various endpoints: white dwarfs, neutron stars, and black holes. These chapters also deal with the mechanisms through which these various forms of stellar death are reached. Finally, Part III does not discuss stars themselves but includes two chapters on their environment and their influence on it, looking at the local interstellar medium and stellar winds. The subjects impact on galactic dynamics and galactic evolution, which are dealt with elsewhere in the series.

The chapters in this volume were mostly commissioned in 2009, with all the material being received between 2010 and 2012. Like almost all of astronomy, in the modern era, this is a fast-moving field. Therefore, the chapters inevitably represent a state of knowledge that will move on over the next few years.

Finally, I would like to acknowledge the friendly cooperation and patience of the authors of the chapters in this volume. I have also valued the support of the staff at Springer, both past and current, in the editorial process and for helping bring this project to fruition. I would also like to thank Terry Oswalt, for inviting me to become an editor on this exciting enterprise. It has been a pleasure to work with everyone.

Martin A. Barstow

Editor-in-Chief



Dr. Terry D. Oswalt

Department Physics & Space Sciences
Florida Institute of Technology
150 W. University Boulevard
Melbourne, Florida 32901
USA
E-mail: toswalt@fit.edu

Dr. Oswalt has been a member of the Florida Tech faculty since 1982 and was the first professional astronomer in the Department of Physics and Space Sciences. He serves on a number of professional society and advisory committees each year. From 1998 to 2000, Dr. Oswalt served as Program Director for Stellar Astronomy and Astrophysics at the National Science Foundation. After returning to Florida Tech in 2000, he served as Associate Dean for Research for the College of Science (2000–2005) and interim Vice Provost for Research (2005–2006). He is now Head of the Department of Physics & Space Sciences. Dr. Oswalt has written over 200 scientific articles and has edited three astronomy books, in addition to serving as Editor-in-Chief for the six-volume Planets, Stars, and Stellar Systems series.

Dr. Oswalt is the founding chairman of the Southeast Association for Research in Astronomy (SARA), a consortium of ten southeastern universities that operates automated 1-meter class telescopes at Kitt Peak National Observatory in Arizona and Cerro Tololo Interamerican Observatory in Chile (see the website www.saraobservatory.org for details). These facilities, which are remotely accessible on the Internet, are used for a variety of research projects by faculty and students. They also support the SARA Research Experiences for Undergraduates (REU) program, which brings students from all over the U.S. each summer to participate one-on-one with SARA faculty mentors in astronomical research projects. In addition, Dr. Oswalt secured funding for the 0.8-meter Ortega telescope on the Florida Tech campus. It is the largest research telescope in the State of Florida.

Dr. Oswalt's primary research focuses on spectroscopic and photometric investigations of very wide binaries that contain known or suspected white dwarf stars. These pairs of stars, whose separations are so large that orbital motion is undetectable, provide a unique opportunity to explore the low luminosity ends of both the white dwarf cooling track and the main sequence to test competing models of white dwarf spectral evolution to determine the space motions, masses, and luminosities for the largest single sample of white dwarfs known; and to set a lower limit to the age and dark matter content of the Galactic disk.

Volume Editor



Martin Barstow

Department of Physics and Astronomy
University of Leicester
University Road
Leicester LE1 7RH
United Kingdom

Dr. Martin A. Barstow received his undergraduate degree in physics from the University of York (UK) in 1979. From there, he became a member of the X-ray Astronomy group in the Department of Physics and Astronomy at the University of Leicester, receiving his Ph.D. in 1983. His Ph.D. work involved the development of a sounding rocket-borne extreme ultraviolet imaging telescope (jointly with MIT), which evolved into the United Kingdom's EUV wide-field camera on board the ROSAT mission. Becoming ROSAT WFC detector scientist in 1984, he received a NASA Group Achievement award on its launch in 1990 and obtained a prestigious SERC (later PPARC) Advanced Fellowship for exploitation of the data in 1990.

Dr. Barstow was appointed to lectureship in the Department of Physics and Astronomy in 1994, was promoted to reader in 1998 and became professor of astrophysics and space science in 2003. He was head of the Physics and Astronomy Department from 2005 to 2009 and is currently pro-vice chancellor and founding head of the College of Science and Engineering. His principal research interests are the study of hot white dwarf stars and the surrounding interstellar medium. In addition to ROSAT, he has been involved in many space missions during his career, including NASA's Voyager probes, during their interplanetary cruises, the ESA EXOSAT mission, IUE, HST, and FUSE. Recently he was a member of the Space Telescope Users' Committee helping to develop the plans for the final servicing mission, which took place in May 2009. He is continues to fly sounding rockets, be involved in research with Hubble, as well as leading the Leicester contribution to the ESA Gaia mission.

During the past few years, Dr. Barstow has played an increasingly important role in scientific funding and advisory structures and has become a member of Science and Technology Facilities Council in 2009. He has also recently been appointed to membership of the ESA Astronomy Working Group and Space Programme Advisory Committee of the United Kingdom Space Agency.

Table of Contents

Series Preface	v
Preface to Volume 4.....	vii
Editor-in-Chief.....	ix
Volume Editor	xi
List of Contributors	xv

Volume 4

1 Stellar Structure	1
<i>Enrique García-Berro · Leandro G. Althaus</i>	
2 Stellar Atmospheres.....	51
<i>Ivan Hubeny</i>	
3 The Sun as a Star.....	87
<i>Giulio Del Zanna · Helen Mason</i>	
4 Asteroseismology.....	207
<i>Gerald Handler</i>	
5 Star Formation	243
<i>Simon Goodwin</i>	
6 Young Stellar Objects and Protostellar Disks	279
<i>Ana Inés Gómez de Castro</i>	
7 Brown Dwarfs	337
<i>I. Neill Reid</i>	
8 Evolution of Solar and Intermediate-Mass Stars.....	397
<i>Falk Herwig</i>	
9 The Evolution of High-Mass Stars	447
<i>Geraldine J. Peters · Raphael Hirschi</i>	

10 Stellar Activity 485
Isabella Pagano

11 White Dwarf Stars..... 559
Detlev Koester

12 Black Holes and Neutron Stars 613
Thomas J. Maccarone

13 Binaries and Multiple Stellar Systems..... 653
Elliott Horch

14 Supernovae and Gamma-Ray Bursts 693
Philipp Podsiadlowski

15 Stellar Winds..... 735
Stan Owocki

Index 789

List of Contributors

Leandro G. Althaus

Facultad de Ciencias Astronómicas y
Geofísicas
Universidad Nacional de La Plata
La Plata
Argentina
and
CONICET-UNLP
La Plata
Argentina

Giulio Del Zanna

DAMTP, Centre for Mathematical Sciences
University of Cambridge
Cambridge
UK

Enrique García-Berro

Departament de Física Aplicada
Universitat Politècnica de Catalunya
Castelldefels
Spain
and
Institut d'Estudis Espacials de Catalunya
Barcelona
Spain

Simon Goodwin

Department of Physics & Astronomy
University of Sheffield
Sheffield, South Yorkshire
UK

Ana Inés Gómez de Castro

Fac. de CC Matemáticas
Universidad Complutense de Madrid
Madrid
Spain

Gerald Handler

Institute for Astronomy
University of Vienna
Vienna, Austria
Warszawa
Poland

Falk Herwig

Department of Physics and Astronomy
University of Victoria
Victoria, BC
Canada

Raphael Hirschi

Astrophysics Group
EPSAM Institute
Keele University
Keele
UK
and
Kavli IPMU
University of Tokyo
Kashiwa, Chiba
Japan

Elliott Horch

Department of Physics
Southern Connecticut State University
New Haven, CT
USA

Ivan Hubeny

Department of Astronomy and Steward
Observatory
The University of Arizona
Tucson, AZ
USA

Detlev Koester

Institut für Theoretische Physik und
Astrophysik
Universität Kiel
Kiel
Germany

Thomas J. Maccarone

School of Physics and Astronomy
University of Southampton
Southampton, Hampshire
United Kingdom

Helen Mason

DAMTP, Centre for Mathematical Sciences
University of Cambridge
Cambridge
UK

Stan Owocki

Bartol Research Institute
Department of Physics and Astronomy
University of Colorado
Newark, DE
USA

Isabella Pagano

INAF, Catania Astrophysical Observatory
Catania
Italy

Geraldine J. Peters

Space Sciences Center and Department
of Physics and Astronomy
University of Southern California
Los Angeles, CA
USA

Philipp Podsiadlowski

Sub-department of Astrophysics
University of Oxford
Oxford
UK

I. Neill Reid

Space Telescope Science Institute
Office of Public Outreach
Baltimore, MD
USA

1 Stellar Structure

Enrique García-Berro^{1,2} · Leandro G. Althaus^{3,4}

¹Departament de Física Aplicada, Universitat Politècnica de Catalunya, Castelldefels, Spain

²Institut d'Estudis Espacials de Catalunya, Barcelona, Spain

³Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina

⁴CONICET-UNLP, La Plata, Argentina

1	<i>Introduction</i>	3
2	<i>Hydrostatic Equilibrium</i>	7
3	<i>Energy Conservation</i>	8
4	<i>Energy Transport</i>	9
4.1	Radiative Transport	9
4.2	Conductive Transport	10
4.3	Convective Transport	11
5	<i>The Virial Theorem</i>	16
6	<i>Physical Inputs</i>	17
6.1	Equation of State	18
6.1.1	Ions	19
6.1.2	Electrons	20
6.1.3	Nonideal Effects	24
6.2	Nuclear Reactions	25
6.2.1	General Concepts	25
6.2.2	Thermonuclear Reaction Rates	26
6.2.3	Electron Screening	28
6.2.4	Nuclear Networks	29
6.2.5	Hydrogen Burning Reactions	30
6.2.6	Helium Burning Reactions	31
6.2.7	Carbon Burning Reactions	32
6.2.8	Other Nuclear Reactions of Interest	32
6.3	Opacities and Conductivities	34
6.3.1	Radiative Opacities	34
6.3.2	Conduction	37
6.4	Neutrino Losses	37
7	<i>Other Physical Processes</i>	38

8	<i>Boundary Conditions and Stellar Atmospheres</i>	40
9	<i>Numerical Techniques: Modern Implementations</i>	42
10	<i>State-of-the-Art Modeling</i>	44
11	<i>Summary</i>	47
	<i>Acknowledgments</i>	47
	<i>References</i>	48

Abstract: In this chapter the equations of stellar structure are introduced. Up-to-date descriptions of the most relevant microphysics needed to build reliable stellar configurations are also provided. These include the equation of state, nuclear reactions, opacities, and other less frequently discussed physical processes, such as diffusion and radiative levitation. The necessary numerical techniques currently employed to build evolutionary stellar models as well as the otherwise necessary boundary conditions needed to integrate the equations of stellar structure are also addressed. Finally, an overview of the current state-of-the-art modeling is given.

Keywords: Stars: evolution, Stars: interiors, Stars: structure

1 Introduction

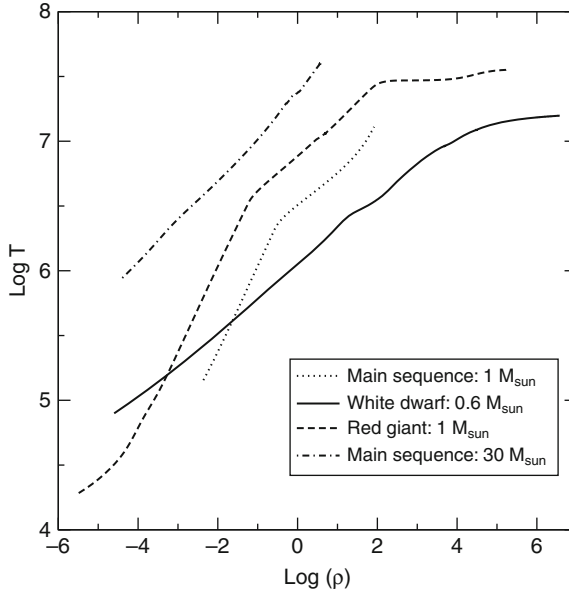
Stars are the fundamental blocks upon which galaxies, and by extension the universe, are built. As such, their study has received continuous interest during several decades. In fact, reliable and up-to-date stellar evolutionary models are routinely required in several fields of modern astrophysics. Hence, understanding the structure of stars is a must, and most branches of astrophysics rely on detailed stellar models. In particular, understanding the structure and evolution of our galaxy, a prototypical spiral galaxy, requires obtaining significant information from statistical studies of the several galactic populations (disk, halo, and bulge). Future space missions, of which Gaia is the archetypical example (Perryman et al. 2001), will obtain an accurate census of the stars in our galaxy, will determine three-dimensional velocities for most of them, and will also quantitatively map its structure. This wealth of information can be accurately and efficiently analyzed only if realistic stellar evolutionary models, based on an accurate description of the properties of stars, are employed. Thus, understanding in detail the structure of stars is of crucial importance. Additionally, Galactic satellite galaxies are proving to be the most natural and successful places where to test and quantify the nature and distribution of dark matter, a key issue in modern cosmology. A precise understanding of these galaxies is a challenge for which again accurate stellar models are needed. Also, the origin of the different chemical elements cannot be assessed without fully understanding the structure and evolution of stars. As a matter of fact, it turns out that all the elements – but hydrogen, helium, lithium, beryllium, and boron – have been synthesized in the deep interiors of stars. Thus, our understanding of the chemical evolution of the universe relies on our ability to describe the details of the internal structure of stars. Moreover, due to the extreme pressures and temperatures that are reached in some of their evolutionary stages, stars can also be used as astroparticle physics laboratories – see the review of Raffelt (1999) – and provide environments that cannot be attained in terrestrial laboratories. Also, stars can be used to test alternative theories of gravitation – see the recent review of García-Berro et al. (2007) – a very hot topic. Finally, a detailed description of stellar structure is also of primordial importance to understand one of the most energetic events in our universe: supernovae. Although important, these are only a few examples for which a detailed knowledge of stellar structure is required and there are several other important applications that are omitted here for the sake of conciseness.

As mentioned, our understanding of stellar structure has been built over several decades. To the best of our knowledge the pioneering studies were done, among others, by Lane (1869), Schwarzschild (1906), and Eddington (1916) between the end of the nineteenth century and the very beginning of the twentieth century. Of course, these initial studies were very crude

and several approximations done in these studies have turned out to be inaccurate. Nevertheless, they helped in setting the stage and they contributed to the foundation of a new branch of astronomy. Since then, the field has experimented notable advances and has reached a degree of maturity and a predictive power that do not have other fields of astronomy. Big steps forward were given by renowned astronomers, of which Chandrasekhar, Fowler, Bethe, Schatzman, Sugimoto, Iben, and several others are conspicuous examples. Their work settled the basements of the modern theory of stellar structure. More recently, the advent of computers and of advanced numerical techniques helped to develop a consistent, coherent, and very detailed theory corpus that nowadays has made possible to compare theoretical models with the extremely large amount of very precise data that is routinely acquired by modern large telescopes.

The study of stellar structure involves several fields of contemporary physics. In order to understand the structure of stars several physical inputs are required. Among them nuclear reactions, which are of fundamental importance to understand the energetics of stars, need to be mentioned first. In turn, an accurate knowledge of nuclear reactions involves not only a proper theoretical description of the structure of nuclei, but also large experimental efforts to measure the cross sections of the most important nuclear reactions. Needless to say that although there exist accurate determinations for several nuclear reactions there are still several others for which reliable determinations are sorely needed. Examples of the former are the nuclear reaction rates relevant for hydrogen burning, and the most paradigmatic example of the latter is, perhaps, the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ nuclear cross section. A detailed knowledge of the equation of state of astrophysical plasmas is also required to accurately model stellar structures. This includes not only the description of the thermodynamical properties of the ideal plasma for a wide range of densities and temperatures – see [Fig. 1-1](#) – but also, if realistic models are to be computed, several nonideal effects must also be taken into account. Also partial ionization must be carefully taken into account. Accurate modeling of energy losses via neutrino emission requires also a good description of weak interactions. Matter at very large densities, like those found in the interiors of neutron stars, require modeling strong interactions. On the other hand, opacities and conductivities of matter are also required. Actually, the list of physical inputs is so large that no attempt is made to be exhaustive at this point, and a detailed discussion of all the most important physical inputs is deferred to subsequent sections.

Before going into the details of stellar structure modeling it is rather convenient to have an approximate and simplistic, but rather effective, idea of how typical stars, of which our own Sun is an example, work. In fact, a main sequence star can be defined as a self-contained, self-controlled, auto-gravitating thermonuclear reactor – something that nuclear engineers have not been able to reproduce on Earth yet. Assume that all stars have spherical symmetry. This is not an unrealistic assumption, because gravity only depends on distance, but it has been only proved for the case of the closest star: our Sun. The most important property (besides isotropy) of gravity is that it is always attractive, contrary to what happens with electromagnetic interactions. In fact, and as it will be shown in subsequent chapters, the life of a star is a battle against gravity. Actually, normal stars succeed in winning successive battles but they lose the war, exploding as supernovae, the most massive ones, or ejecting their external layers and forming compact objects (either white dwarfs or neutron stars), the less massive ones. Since stars seem to be in equilibrium, a force must balance gravity. This force, obviously, is pressure or, to be more precise, the pressure gradient. That is, Archimede's principle comes at work. Given the densities and temperatures of main sequence stars – see [Fig. 1-1](#) – the pressure can be well described by that of a totally ionized ideal gas. This is the same to say that the pressure is proportional to the density (ρ) and the temperature (T). Now, to balance the gravitational force there are



■ Fig. 1-1

Structure of several stellar configurations in the density-temperature plane. Note that the logarithms of both quantities are plotted. The most external layers of these stellar configurations are not displayed. The *dotted line* illustrates the structure of a main sequence star of $1 M_{\odot}$. For comparative purposes the structure of a main sequence star of $30 M_{\odot}$ is also displayed. Both structures correspond to the moment at which hydrogen is ignited at the center. That is, at the zero age main sequence. Note that the $30 M_{\odot}$ star is considerably hotter than our Sun. Also shown is the structure of a red giant star of $1 M_{\odot}$. As can be seen these types of stars have much wider density and temperature ranges. Actually, the central density of this stellar configuration is rather high. The *solid line* shows the stratification of densities and temperatures for an otherwise typical white dwarf of $0.6 M_{\odot}$. It has been chosen to show the structure at an intermediate evolutionary phase, when the central temperature is $\log T \simeq 7.2$. It is important to realize that the central density in this case is much larger than in the previous examples

two possibilities: either to increase the density or to increase the temperature. As increasing the density increases as well the gravitational force per unit volume, it is rather evident that the only possibility that is left to play with is the temperature. This has the consequence that since gravity is stronger at the central regions of the star, the core of a typical main sequence star must be hotter. The second consequence is that more massive stars must be hotter as well, because they need to balance an overall stronger gravitational pull. These two facts can be clearly seen in [Fig. 1-1](#). This issue will be addressed again when the virial theorem, in [Sect. 5](#), and the equation of state, in [Sect. 6.1](#), will be discussed.

Now going one step forward, there is no question that stars shine, that is, stars lose energy. Thus, as they radiate away energy, stars should cool, the pressure should decrease, and, consequently, the radius should decrease as well. But it turns out that this is not the case. Indeed, it can be easily shown that the gravitational potential cannot supply the required amount of

energy during long periods of time. Actually, the release of gravitational energy is governed by the Kelvin–Helmholtz timescale:

$$\tau_{\text{KH}} \sim \frac{GM^2}{RL} \quad (1.1)$$

where G is the gravitational constant, $G = 6.674 \times 10^{-8} \text{ dyn cm}^2 \text{ g}^{-2}$, and M , R , and L are, respectively, the total mass, radius, and luminosity of the corresponding star. Adopting the solar values $M = 1 M_{\odot} \simeq 1.989 \times 10^{33} \text{ g}$, $R = 1 R_{\odot} \simeq 6.955 \times 10^{10} \text{ cm}$, and $L = 1 L_{\odot} \simeq 3.856 \times 10^{33} \text{ erg s}^{-1}$, $\tau_{\text{KH}} \sim 3.0 \times 10^6 \text{ years}$ is obtained, much shorter than the age of the Solar System, which is $\sim 4.5 \times 10^9 \text{ years}$. Thus, gravitational energy cannot be the source of the luminosity of stars. Therefore, another source of energy must be at work. It took several decades to realize that this source of energy were nuclear reactions occurring in the deep interior of stars. In fact, an indication of this is obtained computing the nuclear timescale:

$$\tau_{\text{nuc}} \sim \frac{Mc^2}{L} \quad (1.2)$$

where c stands for the speed of light, and the rest of the symbols have been already defined. Adopting again typical values, it turns out that $\tau_{\text{nuc}} > 10^{10} \text{ years}$, which comfortably fits within the age of the Sun. The sketch previously outlined allows to get a preliminary insight of typical stars that will serve as a guide for more quantitative studies. However, it should be emphasized that this sketch is not valid for compact objects, either white dwarfs or neutron stars. For these stars the key control parameter is not temperature, but density, and nuclear reactions become (in most cases and only as far as it is concerned about isolated stars) irrelevant. This issue will be revisited when studying the equation of state.

With all these considerations in mind the study of the subject of this chapter in a consistent manner can be started. The reader should take into account that the purpose of this chapter is not providing a summary of the several stellar evolutionary phases. This will be found elsewhere in this book. Instead, the chapter will focus on detailing all the equations and physical inputs necessary to compute realistic and up-to-date stellar configurations. Also, the reader should be aware that the selection of papers for explicit citation is necessarily somewhat arbitrary, and is the product of the own special research trajectory and interests of the authors.

The chapter is organized as follows. The equations of stellar hydrostatic equilibrium and energy conservation are first introduced. This will be done in [Sects. 2](#) and [3](#), respectively. The main energy transport mechanisms in stars will be described in [Sect. 4](#). With these tools at hand, an overview of the gross properties of a star, the virial theorem, will be given in [Sect. 5](#). All the necessary physical inputs (equation of state, nuclear reactions, opacities, and neutrino emission rates) will be provided in [Sect. 6](#). A brief introduction to other physical processes relevant for stellar evolutionary calculations, like diffusion and radiative levitation, will be given in [Sect. 7](#). A discussion of how the boundary conditions are usually dealt with will be provided in [Sect. 8](#), and numerical techniques to compute stellar evolution will be detailed in [Sect. 9](#). [Section 10](#) provides insight on modern numerical techniques and available stellar evolutionary codes. Finally, [Sect. 11](#) will close the chapter providing a brief summary.

2 Hydrostatic Equilibrium

In the absence of rotation and magnetic fields, the only forces acting on a given mass element of an isolated star made of matter plus radiation result from pressure and gravity. For most stars, a spherically symmetric configuration can thus be assumed, where functions are constant on concentric spheres at a distance r from the stellar center. The imbalance between gravitational and differential pressure forces yields the equation of motion at r :

$$\varrho \ddot{r} = -\frac{dP}{dr} - \frac{Gm\varrho}{r^2}, \quad (1.3)$$

where \ddot{r} is the local acceleration d^2r/dt^2 , ϱ and P are the local matter density and pressure, and m is the mass in the sphere interior to r . Generally speaking, and for most phases of stellar evolution, stars evolve so slowly that the temporal evolution of the stellar structure can be described by a sequence of models in hydrostatic equilibrium. This being the case, the structure can be assumed to be static. Hence, all time derivatives can be neglected. In this case, the internal pressure gradient balances gravity everywhere in the star, and the equation of motion, (► 1.3), reduces to the equation of hydrostatic equilibrium:

$$\frac{dP}{dr} = -\frac{Gm\varrho}{r^2} \quad (1.4)$$

and mass conservation reads:

$$\frac{dm}{dr} = 4\pi r^2 \varrho. \quad (1.5)$$

These two expressions allow to derive an estimate of the central pressure, P_c , of our Sun using dimensional analysis:

$$\frac{P_c}{R} \sim \frac{GM\langle\varrho\rangle}{R^2} \propto \frac{GM^2}{R^5} \quad (1.6)$$

where $\langle\varrho\rangle$ is the mean density and it has been assumed that the pressure at the surface is much smaller than at the center, a very good approximation. Adopting typical values, $P_c \sim 3.0 \times 10^{15} \text{ dyn cm}^{-2}$ is obtained.

Clearly, the assumption of hydrostatic equilibrium means that the pressure decreases outward. The departures from hydrostatic equilibrium can be characterized using the dynamical timescale τ_{dyn} , which can be computed neglecting the internal pressure gradient in a gravitationally bound configuration. From (► 1.3) it follows that

$$\frac{R}{\tau_{\text{dyn}}^2} \sim \frac{GM}{R^2}, \quad (1.7)$$

thus, we obtain

$$\tau_{\text{dyn}} \sim \frac{1}{\sqrt{G\langle\varrho\rangle}}, \quad (1.8)$$

which is essentially the free-fall timescale. For the Sun τ_{dyn} is about 30 min, for a red giant with $\langle\varrho\rangle = 10^{-6} \text{ g cm}^{-3}$ about 40 days, and for a white dwarf with $\langle\varrho\rangle = 10^6 \text{ g cm}^{-3}$, it is on the order of a few seconds. This implies that hydrostatic equilibrium is always very quickly attained.

3 Energy Conservation

Let l be the net energy per second outflowing from a sphere of radius r , that is, the luminosity, and ϵ_{nuc} the nuclear energy released per unit mass per second. In a stationary situation in which nuclear reactions are the only energy source of the star, the exceeding energy per second, dl , leaving a spherical mass shell of radius r , mass dm , and thickness dr is

$$dl = 4\pi r^2 \rho \epsilon_{\text{nuc}} dr. \quad (1.9)$$

However, heating (or cooling) of the mass element, and the work of expansion (or compression) of the mass shell also contribute to the energy balance. This means that dl can be nonzero even in the absence of nuclear reactions. Using the first law of thermodynamics, the energy equation can then be written

$$\frac{dl}{dm} = \epsilon_{\text{nuc}} + \epsilon_g, \quad (1.10)$$

where ϵ_g is the gravothermal term per unit mass per second

$$\epsilon_g = -\frac{du}{dt} + \frac{P}{\rho^2} \frac{d\rho}{dt}, \quad (1.11)$$

and u is the internal energy per gram. Differentiating the internal energy

$$\frac{du}{dt} = \left(\frac{\partial u}{\partial \rho} \right)_T \frac{d\rho}{dt} + \left(\frac{\partial u}{\partial T} \right)_\rho \frac{dT}{dt}, \quad (1.12)$$

and with the help of the thermodynamic relation

$$\left(\frac{\partial u}{\partial \rho} \right)_T = \frac{P}{\rho^2} - \frac{T}{\rho^2} \left(\frac{\partial P}{\partial T} \right)_\rho, \quad (1.13)$$

Equation 1.11 can be rewritten as

$$\epsilon_g = \frac{T}{\rho^2} \left(\frac{\partial P}{\partial T} \right)_\rho \frac{d\rho}{dt} - C_V \frac{dT}{dt}, \quad (1.14)$$

where C_V is the specific heat at constant density, $C_V = (\partial u / \partial T)_\rho$. Making use again of basic thermodynamics, considering that $\rho = \rho(P, T)$, and the relation

$$\left(\frac{\partial P}{\partial T} \right)_\rho = \frac{P\delta}{T\alpha}, \quad (1.15)$$

with α the isothermal compressibility and δ the volume coefficient of expansion, given by

$$\begin{aligned} \delta &= - \left(\frac{\partial \ln \rho}{\partial \ln T} \right)_P \\ \alpha &= \left(\frac{\partial \ln \rho}{\partial \ln P} \right)_T. \end{aligned} \quad (1.16)$$

Equation 1.14 can be cast, after some algebra, in the form

$$\epsilon_g = -C_P \frac{dT}{dt} + \frac{\delta}{\rho} \frac{d\rho}{dt}, \quad (1.17)$$

where C_P is the specific heat at constant pressure. In this analysis, the variation of the internal energy resulting from the change of local chemical composition has been neglected.

This contribution is usually small for most stages of evolution, as compared to the release of nuclear energy – see Kippenhahn et al. (1965) – but it is relevant in the case of white dwarf stars where nuclear reactions are effectively extinguished (Isern et al. 1997). The energy equation, (1.10), then becomes

$$\frac{dl}{dm} = \epsilon_{\text{nuc}} - C_P \frac{dT}{dt} + \frac{\delta}{Q} \frac{dP}{dt}. \quad (1.18)$$

Integrating (1.11) over mass gives the overall gravothermal contributions to the total energy budget. It is apparent that the integration of the first term in ϵ_g yields the time variation of the total internal energy of the star. As it will be shown later, integration over m of the second term yields the time derivative of the total gravitational energy (Ω) of the star (Kippenhahn and Weigert 1990).

4 Energy Transport

4.1 Radiative Transport

One of the mechanisms by which energy is transferred in stellar interiors is radiation, that is, by photons. Generally speaking, radiation is the usual process by which energy is carried away in stars. In stellar interiors, the photon mean free path ℓ_{ph} is very short compared to the typical length scale over which the structure changes. The mean free path of a photon can be easily estimated:

$$\ell_{\text{ph}} = \frac{1}{\kappa_{\text{rad}} Q}, \quad (1.19)$$

where κ_{rad} is the mean radiative opacity coefficient due to interactions of photons with particles, that is, the radiative cross section per unit mass averaged over frequency. Typical values of κ_{rad} for stellar interiors are $\kappa_{\text{rad}} \approx 1 \text{ cm}^2 \text{ g}^{-1}$. Taking into account the average density of matter in the Sun, $\ell_{\text{ph}} \approx 1 \text{ cm}$. This is much smaller than the stellar radius, thus implying that matter in stellar interiors is very close to local thermodynamic equilibrium, and that the power spectrum of radiation corresponds to that of a blackbody. The mean free path of photons is also so small that the energy transport by radiation can be treated essentially as a diffusive process, introducing an important simplification in the treatment. In the diffusion approximation, the radiative flux is given by

$$\mathbf{F}_{\text{rad}} = -\frac{4\pi}{3\kappa_{\text{rad}} Q} \nabla B = -\frac{4acT^3}{3\kappa_{\text{rad}} Q} \nabla T, \quad (1.20)$$

here $B = (ac/4\pi)T^4$ is the frequency-integrated Planck function, c the speed of light, and a the radiation density constant ($7.564 \times 10^{-15} \text{ erg cm}^{-3} \text{ K}^{-4}$). In the spherical symmetric case, \mathbf{F}_{rad} has only a radial component, $|\mathbf{F}_{\text{rad}}| = F_{\text{rad}}$. Thus, $l = 4\pi r^2 F_{\text{rad}}$, and the diffusion equation becomes

$$\frac{dT}{dr} = -\frac{3}{4ac} \frac{\kappa_{\text{rad}} Q}{T^3} \frac{l}{4\pi r^2}. \quad (1.21)$$

The total energy flux depends on an integral over all radiation frequencies. In the diffusion approximation, the generalization to frequency-dependence leads to the concept of the Rosseland mean opacity, obtained as a harmonic mean of the frequency-dependent opacity. This can

be seen by including in the equation for the radiative flux the frequency dependence

$$F_\nu = -\frac{4\pi}{3\kappa_\nu Q} \frac{dB_\nu}{dr} = -\frac{4\pi}{3Q} \frac{dT}{dr} \frac{1}{\kappa_\nu} \frac{dB_\nu}{dT}, \quad (1.22)$$

where F_ν is the monochromatic flux of frequency ν , and κ_ν the monochromatic opacity resulting from bound–bound, bound–free, and free–free interactions between radiation and electrons. The total flux is obtained by integrating (1.22) over frequency

$$F_{\text{rad}} = -\frac{4\pi}{3} \frac{1}{\kappa_{\text{rad}} Q} \frac{dT}{dr} \int_0^\infty \frac{dB_\nu}{dT} d\nu, \quad (1.23)$$

where κ_{rad} is the Rosseland mean opacity given by

$$\frac{1}{\kappa_{\text{rad}}} = \frac{\int_0^\infty \frac{1}{\kappa_\nu} \frac{dB_\nu}{dT} d\nu}{\int_0^\infty \frac{dB_\nu}{dT} d\nu}. \quad (1.24)$$

With this definition and observing that

$$\int_0^\infty \frac{dB_\nu}{dT} d\nu = \frac{ac}{\pi} T^3, \quad (1.25)$$

(1.21) is recovered. It is worth noting as well that the radiative flux can be cast in the form

$$F_{\text{rad}} = -D \frac{d(aT^4)}{dr}, \quad (1.26)$$

where aT^4 is the radiation energy density and D is a diffusion coefficient given by

$$D = \frac{c}{3\kappa_{\text{rad}} Q} = \frac{1}{3} c \ell_{\text{ph}}. \quad (1.27)$$

In stellar atmospheres, the mean free path of photons becomes much larger, and the diffusion approximation is not valid. There, a more complete and detailed treatment of the full radiative transfer problem is required (Mihalas and Mihalas 1984).

4.2 Conductive Transport

Energy can be transferred not only by photons but also by particles via collisions during the random thermal motions of the particles. This becomes particularly relevant at the high densities characteristic of evolved stars, where electron degeneracy increases both the electron velocity and the mean free path substantially, thus making the diffusion coefficient large. Hence, in the case of stellar matter where electrons are degenerate, electron conduction results in a very efficient energy transfer mechanism, superseding in some cases radiative transfer.

The energy flux due to electron thermal conduction can be written in terms of a coefficient of thermal diffusion, D_e , and the temperature gradient as

$$F_{\text{cd}} = -D_e \frac{dT}{dr}. \quad (1.28)$$

It is convenient to define a “conductive opacity” as

$$\kappa_{\text{cd}} = \frac{4acT^3}{3D_e Q}, \quad (1.29)$$

so that the total energy flux carried by both radiation and thermal conduction can be written as

$$F_{\text{tot}} = F_{\text{rad}} + F_{\text{cd}} = -\frac{4acT^3}{3\kappa_{\text{tot}}\varrho} \frac{dT}{dr}, \quad (1.30)$$

where the total opacity is expressed as

$$\frac{1}{\kappa_{\text{tot}}} = \frac{1}{\kappa_{\text{rad}}} + \frac{1}{\kappa_{\text{cd}}}. \quad (1.31)$$

Note that when $\kappa_{\text{cd}} \gg \kappa_{\text{rad}}$, then $\kappa_{\text{tot}} \simeq \kappa_{\text{rad}}$, and when electron conduction is very efficient, $\kappa_{\text{cd}} \ll \kappa_{\text{rad}}$, $\kappa_{\text{tot}} \simeq \kappa_{\text{cd}}$ is verified.

It is useful to write the diffusion equation, (1.21), for the radiative plus conductive transport in terms of the radiative temperature gradient for a star in hydrostatic equilibrium, ∇_{rad} ,

$$\nabla_{\text{rad}} = \left(\frac{d \ln T}{d \ln P} \right)_{\text{rad}}, \quad (1.32)$$

in the form

$$\nabla_{\text{rad}} = \frac{3}{16\pi acG} \frac{\kappa_{\text{tot}} l P}{m T^4}. \quad (1.33)$$

It should be noted that ∇_{rad} , which is a spatial derivative that relates the variables P and T in two close mass shells, describes the temperature variation with depth for a star in hydrostatic equilibrium where energy is transferred by radiation (and conduction).

4.3 Convective Transport

In a stellar interior, energy is transferred not only by radiation and/or conduction, but also by convection, which is also responsible for chemical element transport. Under certain circumstances, small, local perturbations of elements around their equilibrium positions may grow, thus leading to macroscopic motions. These large-scale motions, or convection, lead to energy transport from the hottest to the coolest regions due to the rising and falling of mass elements under the combined action of buoyancy and gravity. The condition for stability of a mass element that is displaced around its equilibrium position in a certain region of a star (due to, for instance, temperature fluctuations) is

$$\left(\frac{d\varrho}{dr} \right)_{\text{int}} > \left(\frac{d\varrho}{dr} \right)_{\text{ext}}, \quad (1.34)$$

where the subscript “int” denotes the change of internal density of the mass element while it rises a distance dr , and the subscript “ext” indicates the spatial gradient in the star. This condition assumes that the element remains always in pressure equilibrium with the surroundings. That is, the element moves with a speed lower than the local sound speed. The stability condition given by (1.34) simply states that after moving a distance dr , the element will be denser than the fluid in its new environment, so the gravitational force will make the element sink back to its original position.

In order to translate this stability condition into a more tractable form, the equation of state $\varrho = \varrho(P, T, \mu)$ can be used, and can be written as

$$\begin{aligned} \frac{d\varrho}{\varrho} &= \frac{P}{\varrho} \left(\frac{\partial \varrho}{\partial P} \right) \frac{dP}{P} + \frac{T}{\varrho} \left(\frac{\partial \varrho}{\partial T} \right) \frac{dT}{T} + \frac{\mu}{\varrho} \left(\frac{\partial \varrho}{\partial \mu} \right) \frac{d\mu}{\mu} \\ &= \alpha \frac{dP}{P} - \delta \frac{dT}{T} + \varphi \frac{d\mu}{\mu}, \end{aligned} \quad (1.35)$$

where α and δ are given by (1.16), μ is the mean molecular weight, and φ is the chemical potential. Hence, the stability criterion becomes

$$\left(\frac{\alpha}{P} \frac{dP}{dr} - \frac{\delta}{T} \frac{dT}{dr} + \frac{\varphi}{\mu} \frac{d\mu}{dr} \right)_{\text{int}} > \left(\frac{\alpha}{P} \frac{dP}{dr} - \frac{\delta}{T} \frac{dT}{dr} + \frac{\varphi}{\mu} \frac{d\mu}{dr} \right)_{\text{ext}}. \quad (1.36)$$

Since the element is in equilibrium with its surroundings (the pressure is the same), and the values of α and δ for the fluid element are almost equal to the external ones, the first term on either side of the inequality cancel each other. Also, $(du/dr)_{\text{int}} = 0$ since the element does not change its chemical composition while moving, which is the same to say that there is no matter exchange with the surroundings. Then, (1.36) reduces to

$$\left(\frac{\delta}{T} \frac{dT}{dr} \right)_{\text{int}} < \left(\frac{\delta}{T} \frac{dT}{dr} - \frac{\varphi}{\mu} \frac{d\mu}{dr} \right)_{\text{ext}}. \quad (1.37)$$

After multiplying both sides of this inequality by the pressure scale height

$$\lambda_P = -\frac{dr}{d \ln P}, \quad (1.38)$$

which is a measure of the distance over which the pressure is changing by a factor e , the stability condition becomes

$$\left(\frac{d \ln T}{d \ln P} \right)_{\text{ext}} < \left(\frac{d \ln T}{d \ln P} \right)_{\text{int}} + \frac{\varphi}{\delta} \left(\frac{d \ln \mu}{d \ln P} \right)_{\text{ext}}. \quad (1.39)$$

This condition can be used to test the stability of a layer where all energy is transported by radiation and/or conduction. If the star is stable the term on the left-hand side of (1.39) is the radiative temperature gradient defined by (1.32). It will be assumed that the element moves adiabatically. This is same as assuming that the rising fluid element has no time to exchange its energy content with the surrounding environment. Thus, the Ledoux criterion for dynamical stability is obtained

$$\nabla_{\text{rad}} < \nabla_{\text{ad}} + \frac{\varphi}{\delta} \left(\frac{d \ln \mu}{d \ln P} \right), \quad (1.40)$$

where ∇_{ad} is the adiabatic temperature gradient and corresponds to the temperature gradient when the moving element does not exchange heat with the surrounding medium. For a general equation of state, the adiabatic gradient reads

$$\nabla_{\text{ad}} = \frac{P\delta}{C_P \varrho T}. \quad (1.41)$$

It is worth mentioning that for the particular case of a chemically homogeneous region of a star, $d \ln \mu / d \ln P = 0$. Therefore, the condition for stability in this case simply reads

$$\nabla_{\text{rad}} < \nabla_{\text{ad}}, \quad (1.42)$$

which is the Schwarzschild criterion for dynamical stability. In those stellar regions where nuclear reactions produce heavier elements below the lighter ones, the chemical gradient term in (1.39) favors convective stability since $d \ln \mu / d \ln P > 0$. Indeed, in regions of varying μ , a fluid element moving upward is made of matter with a higher molecular weight than that of its surrounding medium, forcing the element to sink down as a result of gravity.

The derivation of the condition for stability can be seen from a slightly different point of view. Assume that a fluid element in a star is displaced vertically and adiabatically from its equilibrium position with the surroundings at r_0 . It will experience a buoyancy force per unit volume equal to $-g (\varrho_{\text{int}} - \varrho_{\text{ext}})$, where g is the absolute value of the gravitational acceleration and ϱ_{int} and ϱ_{ext} are, respectively, the interior and exterior densities of the fluid element. In the absence of viscous effects, the equation of motion of the element is

$$\varrho_{\text{int}} \frac{d^2 r}{dt^2} = -g (\varrho_{\text{int}} - \varrho_{\text{ext}}). \quad (1.43)$$

For a small displacement $(r - r_0)$, we obtain

$$\varrho_{\text{int}}(r) = \varrho_{\text{int}}(r_0) + \left(\frac{d\varrho_{\text{int}}}{dr} \right)_{r_0} (r - r_0), \quad (1.44)$$

and the same for ϱ_{ext} . Since at r_0 , $\varrho_{\text{int}}(r_0) = \varrho_{\text{ext}}(r_0)$, the equation of motion becomes

$$\varrho_{\text{int}} \frac{d^2 r}{dt^2} + g \left(\frac{d\varrho_{\text{int}}}{dr} - \frac{d\varrho_{\text{ext}}}{dr} \right) (r - r_0) = 0, \quad (1.45)$$

the solution of which is of the form $(r - r_0) = A \exp(i N t)$, with N the oscillation frequency of the element around its equilibrium position, also called the Brunt–Väisälä frequency, which is given by

$$N^2 = \frac{g}{\varrho} \left(\frac{d\varrho_{\text{int}}}{dr} - \frac{d\varrho_{\text{ext}}}{dr} \right). \quad (1.46)$$

Note that if $(d\varrho_{\text{int}}/dr) > (d\varrho_{\text{ext}}/dr)$, then $N^2 > 0$, N is real, and the movement is oscillatory. These oscillations are also known as gravity waves (not to be confused with gravitational waves in General Relativity) since gravity is the restoring force. Thus, the layer will be stable against convection. On the other hand, if $(d\varrho_{\text{int}}/dr) < (d\varrho_{\text{ext}}/dr)$, then $N^2 < 0$, N is imaginary, and then the element will move exponentially from the equilibrium position. Clearly, the layer will be unstable against convection. Now compare these results with the stability condition given by (1.34).

The actual temperature gradient in a convective region, the convective gradient ∇_{conv} , will be different from the radiative temperature gradient. It is clear that if a fraction of the total flux is carried by convection, then $\nabla_{\text{conv}} < \nabla_{\text{rad}}$, where ∇_{rad} represents the temperature gradient that would be needed to transport the entire flux by radiation and conduction. The total flux consists of the radiative plus convective fluxes:

$$F_{\text{tot}} = \frac{l}{4\pi r^2} = F_{\text{conv}} + F_{\text{rad}} = \frac{4acGmT^4}{3 \kappa_{\text{tot}} P r^2} \nabla_{\text{rad}}. \quad (1.47)$$

If ∇_{conv} is the actual temperature gradient, it is clear then that the flux carried by radiation is only

$$F_{\text{rad}} = \frac{4acGmT^4}{3\kappa_{\text{tot}}Pr^2}\nabla_{\text{conv}}. \quad (1.48)$$

In a convective zone, the following relations are valid (the second inequality is the criterion for convection)

$$\nabla_{\text{rad}} > \nabla_{\text{conv}} > \nabla_{\text{int}} > \nabla_{\text{ad}}. \quad (1.49)$$

The calculation of ∇_{conv} remains a serious issue. In fact, a model for convection must be specified. Convection is essentially a nonlocal and complex phenomenon that involves the solution of the hydrodynamic equations, and it remains a weak point in the theory of stellar evolution. In most stellar applications, a simple local formulation (Böhm-Vitense 1958) called the mixing-length theory (MLT) is used. This crude model assumes that the convective flux is transported by single size, large fluid elements, which after traveling, on the average, a distance ℓ_{MLT} , the *mixing length*, break up releasing their energy excess into the surrounding medium. The distance ℓ_{MLT} , which is also the characteristic size of the elements, is parameterized in terms of the pressure scale height, $\ell_{\text{MLT}} = \alpha_{\text{MLT}} H_P$, where α_{MLT} is a free parameter not predicted by the theory that must be calibrated using observations, and H_P is the pressure scale height. In particular, in the Böhm-Vitense formulation, the MLT involves three length scales which, in most stellar applications, are reduced to ℓ_{MLT} . Usually, α_{MLT} is found fitting the solar radius ($\alpha_{\text{MLT}} \approx 1.6$), and this value is usually used to model other stars and evolutionary phases.

In the deep interior of stars, convection results from the large values of ∇_{rad} caused by the strong concentration of nuclear burning near the stellar center, for instance during the core hydrogen burning phase via the CNO cycle in stars somewhat more massive than the Sun. The high densities of these regions make the temperature stratification almost adiabatic, and thus $\nabla_{\text{conv}} = \nabla_{\text{ad}}$. This means that a very small excess of ∇_{conv} over the adiabatic value is enough to transport all the flux. Consequently, the uncertainties in the MLT theory become irrelevant and a detailed treatment of convection is not required to specify ∇_{conv} . For the Sun, typical convective velocities are of the order of 400 cm s^{-1} , whereas for more massive stars they are much larger. Hence, the turnover time, or the travel time of the elements over the distance ℓ_{MLT} , ranges from about 1 to 100 days. This time is far much shorter than the main sequence lifetime. In fact, during most evolutionary phases convective mixing is essentially an instantaneous process, thus leading to chemically homogeneous convective zones. However, this may not be true during fast evolutionary stages, where the convective timescale becomes comparable to the evolutionary timescale.

A complete solution of the MLT, with all its associated uncertainties, is required in the low-density, outermost part of convective envelopes where the temperature gradient markedly differs from the adiabatic value (Cox and Giuli 1968). In these layers, large values of ∇_{rad} result from the large opacity values in the ionization zones of hydrogen and helium close to the surface, causing convection in the outer parts of relatively cool stars. Here, the density and the heat content of matter are so low that a temperature gradient largely exceeding ∇_{ad} is required to transport energy. Depending on the efficiency of convection, ∇_{conv} will be somewhere between ∇_{ad} and ∇_{rad} . Typical values for the average convective velocity in the solar envelope are about 1 km s^{-1} , close the local sound speed, and the turnover timescale is of the order of 5 min.

However, processes such as overshooting, that is, the extension of convective zones beyond the formally convective boundaries given by (☛ 1.40) – at the convective boundaries, fluid

elements have zero acceleration but nonzero velocity – cannot be satisfactorily treated using a local theory. More mixing than is expected from the MLT treatment is supported by different pieces of astrophysical evidence, which suggest that real stars have larger convective cores. Overshooting – which is critical in determining the total amount of nuclear fuel available for the star – is a nonlocal process, and its extent depends on the properties of the adjacent layers. In most studies of stellar structure and evolution, overshooting is simulated extending the boundaries of the convective layer and mixing material beyond the formal convective boundary. This is known as instantaneous overshooting. A better approach is to treat overshooting as a diffusion process. This approach enables a self-consistent treatment of this process in the presence of nuclear burning (Herwig 2000). Here, overshooting parameterization is based on hydrodynamical simulations (Freytag et al. 1996), which show that turbulent velocities decay exponentially outside the convective boundaries. This diffusive overshooting gives rise to mixing in the overshoot regions whose efficiency is quantified in terms of the diffusion coefficient

$$D_{\text{os}} = D_0 \exp\left(\frac{-2z}{H_v}\right), \quad (1.50)$$

where D_0 is the diffusion coefficient at the boundary of the convection zone, z is the radial distance from the edge of the convection zone, and H_v is the velocity scale height of the overshoot convective elements at the convective boundary. H_v is parameterized as a fraction f of the pressure scale height, $H_v = f H_p$. The parameter f is a measure of the extent of the overshoot region. Clearly, the larger the f , the extra mixing beyond the convective boundary extends further. Usually $f \approx 0.016$ is adopted. This choice of f accounts for the observed width of the main sequence as well as for the intershell abundances of hydrogen-deficient post-AGB remnants.

Another complication is the occurrence of “semiconvection,” a slow mixing process that is expected to occur in those regions with an inward increasing value of μ that are unstable according to the Schwarzschild criterion but stable according to the Ledoux criterion, namely, those layers where

$$\nabla_{\text{ad}} + \frac{\varphi}{\delta} \left(\frac{d \ln \mu}{d \ln P} \right) > \nabla_{\text{rad}} > \nabla_{\text{ad}}. \quad (1.51)$$

Here, energy losses from the fluid elements (elements are hotter than the surroundings) will cause them to oscillate around their equilibrium positions (vibrational instability) with progressively growing amplitudes (Kippenhahn and Weigert 1990). Because of heat losses, the elements return to the equilibrium position with a temperature lower than that with which they started, thus reaching deeper and hotter regions in their downward excursion. The growth of the oscillation amplitudes is determined by the timescale of thermal adjustment of the fluid elements. The overstability resulting from these growing oscillations and nonlinear effects is believed to result in partial mixing of the corresponding layers. Realistic physical models of all these nonlocal processes require two- and three-dimensional numerical simulations of nonlinear hydrodynamic instabilities and turbulent processes (Young et al. 2003). Finally, heat leakage of the elements is also responsible for another type of instability, thermohaline convection. This process leads to significant turbulent transport in stable regions with negative chemical gradients – see Traxler et al. (2011) for recent three-dimensional simulations of this process in stars.

In closing, it is worth mentioning that various attempts to improve the MLT have been made. In particular, Canuto and Mazzitelli (1991) have considered the full spectrum of turbulence in velocities and sizes of the convective eddies. An extended version of the Mixing Length Theory of convection, for fluids with composition gradients, has been derived by Grossman and

Taam (1996) in the local approximation. These authors – see also Grossman et al. (1993) – have developed the nonlinear Mixing Length Theory of double diffusive convection (GNA), where both the effects of thermal and composition gradients compete to determine the stability of the fluid. The GNA theory is based on the MLT picture and considers the fluid as an ensemble of individual elements or blobs. This ensemble is described by a distribution function that evolves in time according to a Boltzmann-type equation. In its local version, all third-order terms in the second-moment of the Boltzmann equation are neglected. The GNA theory applies in convective, semiconvective, and thermohaline regimes. According to this treatment, the diffusion coefficient D characterizing mixing in the various regimes is given by

$$D = \frac{1}{3} \ell \sigma \quad (1.52)$$

where $\ell = \alpha H_P$ is the mixing length and σ the turbulent velocity. The value of σ is determined by simultaneously solving the equations for the turbulent velocity and flux conservation – see Grossman and Taam (1996) for further details. In this theory of convection, the standard MLT for a fluid of homogeneous composition is a limiting case.

5 The Virial Theorem

Stars are one of the best examples of a virialized system. In fact, it is rather simple to prove this. Assume that they are in hydrostatic equilibrium – and, hence, that (1.4) holds – and that they are made up of an ideal gas. Accordingly, multiplying (1.4) by $4\pi r^3$ and integrating over the entire star,

$$\int_0^R 4\pi r^3 \frac{dP}{dr} dr = - \int_0^R 4\pi r^3 \frac{G \varrho m}{r^2} dr \quad (1.53)$$

is obtained, where R is the stellar radius. The left-hand side of this equation can be easily worked out. Integrating by parts it is found that

$$\int_0^R 4\pi r^3 \frac{dP}{dr} dr = [4\pi r^3 P]_0^R - 3 \int_0^R P 4\pi r^2 dr = -3 \int_0^M \frac{P}{\varrho} dm \quad (1.54)$$

where it has been assumed that the pressure at the surface of the star vanishes, and (1.5) has been used. The right-hand side of the previous equation can also be further worked out, to yield

$$- \int_0^R 4\pi r^3 \frac{G \varrho m}{r^2} dr = - \int_0^R 4\pi r^2 \varrho \frac{Gm}{r} dr = - \int_0^M \frac{Gm}{r} dm = \Omega \quad (1.55)$$

which is the total gravitational energy. Thus,

$$3 \int_0^M \frac{P}{\varrho} dm + \Omega = 0. \quad (1.56)$$

Quite generally, for an ideal gas, $P = (\gamma - 1)\varrho u$, where u is the internal energy and γ is the adiabatic index. Thus, substituting this relationship in the last equation, a rather general result is obtained:

$$3(\gamma - 1)U + \Omega = 0 \quad (1.57)$$

where the total thermal energy, U , has been introduced. On the other hand, the total energy of a star (the binding energy) is $B = U + \Omega$. Thus,

$$B = \frac{3\gamma - 4}{3\gamma - 3} \Omega. \quad (1.58)$$

For a perfect gas $\gamma = 5/3$. Consequently, $2U + \Omega = 0$ and $B = \Omega/2$. This means that if a star contracts – or, equivalently, it releases gravitational energy – half of the gravitational energy is transformed in thermal energy, whereas the other half must be radiated away. Note as well that since $\Omega < 0$, the total energy of the star is also negative, an otherwise expected result that means that the star is bound. Only in the case in which $\gamma = 4/3$, that is, for a completely degenerate relativistic gas, $B = 0$. This last result is of special significance as it is closely related to the concept of Chandrasekhar's mass and to the explosion mechanism of thermonuclear supernovae.

The virial theorem can be used to obtain a relation between the stellar mass and the mean temperature, T . Assume that stars are essentially composed of hydrogen. This assumption is valid for most stars, for which the hydrogen mass fraction is typically ~ 0.75 . The total thermal energy is thus

$$U \sim \frac{3}{2} N k_B T \propto M T \quad (1.59)$$

where N is the total number of particles and $k_B = 1.380 \text{ erg K}^{-1}$ is the Boltzmann constant. Since $\Omega \propto -GM^2/R$ and $U = -\Omega$, it follows that $T \propto M/R$, and thus the more massive a star, the hotter, in qualitative agreement with arguments put forth in [Sect. 1](#). This argument can be pushed forward to obtain a mass-luminosity relationship. The total luminosity of a star can be expressed as $L = 4\pi R^2 F$, where F is the flux. Using ([1.20](#)) we obtain:

$$F \propto \frac{1}{\kappa \varrho} T^3 \frac{dT}{dr} \sim \frac{1}{\kappa \varrho} \frac{T^4}{R}. \quad (1.60)$$

Consequently, after some elementary algebra, and taking into account that the virial theorem states that $T \propto M/R$, and also considering that $\varrho \propto M/R^3$ one arrives at the conclusion that $L \propto M^3/\kappa$. This is same as saying that more massive stars are not only hotter, but also more luminous.

6 Physical Inputs

As has been shown in the previous sections, the basic equations describing the structure and evolution of stars are relatively simple. To this description the microphysics, that is, the properties of stellar matter, has to be added. These properties include the equation of state, opacity, and energy generation rates, among others. The relevant processes occur in an interacting plasma of ions, electrons, and atoms, and the detailed physics of these processes is still an active field of research. Thus, some uncertainties still remain. In this section, all the main physical inputs necessary to model the structure of a star are detailed. These include, of course, the equation of state, described in [Sect. 6.1](#), which provides the pressure as a function of the temperature, density, and chemical composition. Obviously this is needed to solve the equation of hydrostatic equilibrium. The second important input is the nuclear energy generation rate – [Sect. 6.2](#) – which, as mentioned, is needed to maintain stable temperatures during long periods of time. Attention will be paid to only the most important thermonuclear reaction rates, namely, to those relevant for the hydrogen, helium, and carbon burning phases, and other interesting nuclear reactions will be briefly mentioned, but the aim of this section is not to be exhaustive. The interested reader is referred to the several recent works on this particular topic – see, for instance, the detailed and interesting paper of Longland et al. ([2010](#)) and subsequent publications – for detailed and exhaustive information. As already shown, opacities and conductivities are crucial in evaluating the rate at which energy is transported, and they are described

in [Sect. 6.3](#), while [Sect. 6.4](#) is devoted to describe the most recent neutrino emission rates. Finally, in [Sect. 7](#) an overview will be given of the several other physical inputs which are only required under special circumstances.

6.1 Equation of State

The equation of state describes the thermodynamical properties of stellar matter and relates density to pressure, temperature, and chemical composition. It is particularly simple in the case of an ideal gas, but physical processes relevant for stars – that include, among others, ionization, electron degeneracy, molecular dissociation, radiation pressure, or Coulomb interactions – turn the treatment of the equation of state into a rather complex issue. In dealing with these effects, care must be taken to ensure the treatment to be thermodynamically consistent in the sense that it satisfies the thermodynamical identities between the different quantities. One way in which this can be achieved consists in deriving the equation of state from the free energy of the gas. Using this approach, the thermodynamical state of the gas for a given temperature, density, and composition is derived minimizing the free energy, which yields the occupation numbers and ionization states. The relevant thermodynamical quantities can be then obtained as derivatives of the free energy. This is the basis for the so-called chemical picture in determining the equation of state. The second way to ensure thermodynamical consistency is based on the physical picture. Within this less used approach, instead of dealing with the chemical equilibrium of a set of predefined ions, atoms, and molecules, only elementary particles of the problem (nuclei and electrons) are assumed at the beginning, and composite particles appear as a result of the interactions in the system.

It is customary to characterize the chemical composition of stellar matter by means of the mean molecular weight, μ , defined as $\mu = \varrho / (n m_H)$, where n is the total number of particles per unit volume and m_H the atomic mass unit (1.6605×10^{-24} g). For a mixture of fully ionized gases, the molecular weight of the gas is given by the harmonic mean of the molecular weights of the ions (μ_o) and electrons (μ_e):

$$\frac{1}{\mu} = \frac{1}{\mu_o} + \frac{1}{\mu_e}, \quad (1.61)$$

where

$$\frac{1}{\mu_o} = \sum_i \frac{X_i}{A_i}, \quad (1.62)$$

$$\frac{1}{\mu_e} = \sum_i \frac{X_i Z_i}{A_i}, \quad (1.63)$$

with X_i , A_i , and Z_i being the mass fraction (normalized to unity), $X_i = \varrho_i / \varrho$, atomic weight, and charge, respectively, of element i . These definitions are apparent by noting that

$$n = \frac{\varrho}{\mu m_H} = n_e + n_o = \sum_i Z_i n_i + \sum_i n_i = \sum_i Z_i \frac{X_i}{A_i m_H} \varrho + \sum_i \frac{X_i}{A_i m_H} \varrho, \quad (1.64)$$

where n_i is the number of ions per unit volume of element i . In the case of a neutral gas, the molecular weight reduces to μ_o . For complete ionization, a simple expression for μ_e is obtained

by assuming that for all elements heavier than helium ($Z > 2$), $A_i/Z_i \approx 2$. In this case the molecular weight per electron becomes

$$\mu_e = \frac{2}{1 + X_H}, \quad (1.65)$$

where X_H is the abundance by mass of hydrogen. In this approximation and when there is no hydrogen, $\mu_e = 2$.

6.1.1 Ions

In the absence of quantum effects, and assuming that the potential energy of particle interactions is much smaller than the kinetic energy of the particles, the equation of state of ions adopts the simplest form. The particles obey the Maxwell–Boltzmann distribution, and the equation of state corresponding to an ideal (perfect) gas results

$$P = \frac{k_B}{m_H \mu} \varrho T. \quad (1.66)$$

This equation describes the equation of state of fully ionized matter, as found in the deep interior of stars (note that the electron contribution is through μ), as well as matter where all electrons are in the atom (no ionization at all).

If photons contribute appreciably to gas pressure, $P_{\text{rad}} = (1/3)aT^4$, then the equation of state becomes (assuming that radiation is in thermodynamic equilibrium with matter)

$$P = \frac{k_B}{m_H \mu} \varrho T + \frac{1}{3} a T^4. \quad (1.67)$$

From the internal energy per unit mass of the monoatomic gas (it is assumed that, in the case of neutral matter, there are no internal degrees of freedom):

$$E = \frac{3}{2} \frac{k_B}{m_H \mu} T + \frac{a T^4}{\varrho}, \quad (1.68)$$

the specific heat at constant pressure, C_P , and the adiabatic temperature gradient, ∇_{ad} , are derived:

$$C_P = \frac{k_B}{m_H \mu} \left[\frac{3}{2} + \frac{16 - 12\beta - 3\beta^2}{\beta^2} \right], \quad (1.69)$$

and

$$\nabla_{\text{ad}} = \frac{1 + \frac{(1-\beta)(4+\beta)}{\beta^2}}{\frac{5}{2} + \frac{4(1-\beta)(4+\beta)}{\beta^2}}. \quad (1.70)$$

Here, β quantifies the importance of radiation pressure, and is defined as $\beta \equiv P_{\text{gas}}/P$, where P is the total pressure due to gas plus radiation. If radiation pressure is negligible, then $\beta \rightarrow 1$ and $P_{\text{gas}} \rightarrow P$, $C_P \rightarrow 5k_B/(2\mu m_H)$, and $\nabla_{\text{ad}} \rightarrow 0.4$, the well-known values for the ideal gas. For $\beta \rightarrow 0$, $P \rightarrow P_{\text{rad}}$, $C_P \rightarrow \infty$, and $\nabla_{\text{ad}} \rightarrow 0.25$.

This simple treatment ignores a number of important effects that are relevant in the astrophysical context. For instance, at low temperatures, partial ionization of matter must be taken into account. This changes the mean molecular weight and the energetics of the gas. The treatment of partial ionization is usually done assuming chemical equilibrium between the gas constituents. In this case, the ratio n_{j+1}/n_j of atoms $j + 1$ times ionized to those which are j times ionized is given by the Saha equation

$$\frac{n_{j+1}}{n_j} = 2 \frac{u_{j+1}}{u_j} \left(\frac{2\pi m_e}{h^2} \right)^{3/2} \frac{(k_B T)^{5/2}}{P_e} e^{-\chi_j/k_B T}, \quad (1.71)$$

where u_j is the partition function for the ion in the energy state j (which is a function of T), h is the Planck constant, P_e is the electronic pressure, and χ_j is the ionization energy. Note that ionization is favored at high temperatures and low electronic pressures. Quantities such as C_P , ∇_{ad} , and μ are notably affected by ionization. In particular, ∇_{ad} is decreased below 0.4 and C_P markedly increases when compared with the values resulting from fully ionized perfect gas. It is important to note, however, that the treatment of ionization is a difficult task since it involves considering the various ionization degrees of all chemical species. The problem is specified by a full set of coupled Saha equations, the solution of which can only be done numerically.

6.1.2 Electrons

The inner regions of the vast majority of highly evolved stars, including white dwarf stars, are dominated by degenerate electrons. At very high densities and low temperatures, the de Broglie wavelength of electrons, $h/(2m_e k_B T)^{1/2}$, becomes larger than the mean separation of electrons ($d \sim \rho^{-1/3}$) and, as a result, quantum effects become relevant. Hence, electrons become degenerate and quantum mechanics – the Pauli exclusion principle – strongly affects the equation of state. In this case, the distribution of electron momenta obeys the Fermi–Dirac statistics, and the average occupation number at equilibrium of a cell in the phase space – actually, the distribution function of energies – is given by

$$f(\varepsilon) = \frac{1}{1 + \exp[(\varepsilon - \mu_q)/k_B T]}, \quad (1.72)$$

where μ_q is the chemical potential of the gas and ε is the kinetic energy corresponding to the momentum, p , which is given by:

$$\varepsilon(p) = m_e c^2 \left[\sqrt{1 + (p/m_e c)^2} - 1 \right]. \quad (1.73)$$

Note that this expression is valid for both relativistic and nonrelativistic electrons – in the nonrelativistic limit ($pc \ll m_e c^2$), and $\varepsilon \simeq p^2/2m_e$. For noninteracting, degenerate electrons at

a given temperature T , the number density and pressure for the electron gas are, respectively,

$$n_e = \frac{8\pi}{h^3} \int_0^\infty \frac{p^2}{1 + \exp[(\varepsilon - \mu_q)/k_B T]} dp \quad (1.74)$$

$$P_e = \frac{8\pi}{3h^3} \int_0^\infty \frac{p^3}{1 + \exp[(\varepsilon - \mu_q)/k_B T]} \frac{d\varepsilon_p}{dp} dp. \quad (1.75)$$

In the limit of high temperature and low density, $\mu_q/k_B T \ll -1$, and the distribution function reduces to the Maxwell–Boltzmann distribution, $f(\varepsilon) \rightarrow \exp[-(\varepsilon - \mu_q)/k_B T]$. In this case, $P_e \rightarrow n_e k_B T$. At very high densities (or very low temperatures), that is, when the de Broglie wavelength is much larger than the mean separation, the electron gas behaves as a zero-temperature gas. In this zero-temperature approximation, $\mu_q \gg k_B T$ and μ_q is identified with the Fermi energy (ε_F). Hence, the Fermi–Dirac distribution that characterizes electrons reduces to the step-like function:

$$f(\varepsilon) = \begin{cases} 1 & \text{if } \varepsilon \leq \varepsilon_F \\ 0 & \text{if } \varepsilon > \varepsilon_F. \end{cases}$$

In this situation, electrons occupy only the energy states up to ε_F , and not the higher energy states where the distribution function is zero. In particular, those electrons with energies close to ε_F will make the largest contribution to the pressure. In this zero-temperature approximation, the so-called complete degeneracy approximation, the electron pressure can be easily derived by considering only the energy states up to ε_F . Then, (1.74) and (1.75) need only to be integrated up to p_F , the momentum corresponding to the Fermi energy. Using the relation $\varrho = \mu_e m_H n_e$, the electron pressure and mass density become then

$$\begin{aligned} P_e &= \frac{8\pi}{3h^3} \int_0^{p_F} \frac{(p^4/m_e) dp}{\sqrt{1 + (p/m_e c)^2}} \\ &= A [x(x^2 + 1)^{1/2} (2x^2 - 3) + 3 \sinh^{-1}(x)] \\ \varrho &= B \mu_e x^3, \end{aligned} \quad (1.76)$$

where the dimensionless Fermi momentum is given by $x = p_F/m_e c$, $A = \pi m_e^4 c^5/3h^3 = 6.0 \times 10^{22} \text{ dyn cm}^{-2}$, and $B = 8\pi m_e^3 c^3 m_H/3h^3 = 9.74 \times 10^5 \text{ g cm}^{-3}$, and the rest of the symbols have their usual meaning. Note that this expression is valid for any relativistic degree (Chandrasekhar 1939). Relativistic effects have to be taken into account at very high densities – note that as the density is increased so does p_F . In particular, relativistic effects start becoming prominent when $p_F \approx m_e c$ or $x \approx 1$, which corresponds to a density of $\varrho/\mu_e \approx 10^6 \text{ g cm}^{-3}$. This density corresponds to typical values of the central densities in white dwarf stars. Thus, it is expected that electrons in the core of white dwarfs are partially relativistic. Expansions of the equation for the electronic pressure P_e are possible in the limiting cases $x \rightarrow 0$, namely, nonrelativistic, and $x \rightarrow \infty$, extremely relativistic. In both cases it is possible to eliminate the variable x , thus resulting in simple equations for the pressure of a completely degenerate electron gas:

$$\begin{aligned} P_e &\rightarrow A \frac{8}{5} x^5 = \left(\frac{3}{8\pi}\right)^{2/3} \frac{h^2}{5m_e m_H^{5/3}} \left(\frac{\varrho}{\mu_e}\right)^{5/3} \\ &= 1.004 \times 10^{13} \left(\frac{\varrho}{\mu_e}\right)^{5/3} \quad x \ll 1 \end{aligned} \quad (1.77)$$

$$\begin{aligned}
 P_e &\rightarrow A 2 x^4 = \left(\frac{3}{8\pi} \right)^{1/3} \frac{hc}{4m_H^{4/3}} \left(\frac{\varrho}{\mu_e} \right)^{4/3} \\
 &= 1.244 \times 10^{15} \left(\frac{\varrho}{\mu_e} \right)^{4/3} \quad x \gg 1
 \end{aligned} \tag{1.78}$$

where P_e is in dyn cm^{-2} . Although the zero-temperature approximation is an idealized one, the electron gas in some astrophysical environments behaves as if it were indeed at zero temperature. This is the case of the core of cool white dwarfs, where their structure is supported entirely by the pressure of an almost completely degenerate electron gas. This becomes clear by noting that the condition for complete degeneracy, $T_F \equiv \varepsilon_F/k_B \gg T$, where T_F is the Fermi temperature, can be written in the form (for nonrelativistic electrons):

$$T_F = \frac{1}{k_B} \frac{p_F^2}{2m_e} = \frac{m_e c^2}{2k_B} \left(\frac{\varrho}{C} \right)^{2/3} = 3 \times 10^5 \left(\frac{\varrho}{\mu_e} \right)^{2/3} \gg T. \tag{1.79}$$

For a typical white dwarf with a central density of 10^7 g cm^{-3} , $T_F = 9 \times 10^9 \text{ K}$, which is much larger than the core temperature (typical core temperatures range from 10^7 to 10^6 K). Hence, the electron gas behaves as a zero-temperature Fermi gas. In sharp contrast, at the center of the Sun $\varrho/\mu_e \approx 100$, thus $T_F \approx 6 \times 10^6 \text{ K}$, smaller than its central temperature ($\sim 10^7 \text{ K}$). Hence, electron degeneracy is not relevant at the center of the Sun. Finally, note the smaller dependence of pressure on density for relativistic electrons. This relativistic “softening” of the equation of state is responsible for the existence of a limiting mass for white dwarf stars, the so-called Chandrasekhar limiting mass (Chandrasekhar 1939).

As noted, the equation of state of white dwarf interiors can be well approximated by that of an ideal Fermi gas at zero temperature. This has important consequences. In particular, the pressure only depends on the density and not on the temperature. Consequently, the control parameter to balance the gravitational pull in this case is ϱ , at odds with what was discussed in [Sect. 1](#) for normal stars. It should be reminded that for main sequence stars the control parameter is the temperature, and hence more massive stars have larger temperatures. This is not the case for white dwarfs. Massive white dwarfs have larger central densities. It is easy to also show that more massive white dwarfs have smaller radii and, in the limit, Chandrasekhar’s mass appears as the limiting mass for which the radius of the equilibrium configuration is zero.

In the case of finite temperature ($T \neq 0$), the Fermi–Dirac distribution for electrons cannot be approximated by a step-like function as in the case in which complete degeneracy is assumed. Here, $T_F \sim T$ and those electrons with energies near ε_F can occupy levels with energies larger than ε_F . This situation is referred to as partial degeneracy, and the number density and electron pressure are given by [\(1.74\)](#) and [\(1.75\)](#), respectively. Only the case of nonrelativistic electrons, for which $\varepsilon = p^2/(2m_e)$, will be considered here. Introducing the degeneracy parameter $\alpha = -\mu_q/k_B T$, these equations become

$$n_e = \frac{8\pi}{h^3} \int_0^\infty \frac{p^2}{1 + \exp(\alpha + p^2/2m_e k_B T)} dp = \frac{4\pi}{h^3} (2m_e k_B T)^{3/2} F_{1/2}(\alpha), \tag{1.80}$$

$$P_e = \frac{8\pi}{3h^3 m_e} \int_0^\infty \frac{p^4}{1 + \exp(\alpha + p^2/2m_e k_B T)} dp = \frac{8\pi}{3h^3} (2m_e k_B T)^{3/2} k_B T F_{3/2}(\alpha), \tag{1.81}$$

where $F_n(\alpha)$ are the Fermi–Dirac integrals, given by

$$F_n(\alpha) = \int_0^\infty \frac{u^n}{1 + \exp(\alpha + u)} du. \quad (1.82)$$

It should be kept in mind that in the case of partial degeneracy, relativistic effects have to be considered for temperatures in excess, 10^9 K. Indeed, it has been shown that relativistic effects for a degenerate gas become relevant when density exceeds 10^6 g cm^{-3} . At such densities, degeneracy will be essentially complete unless $T > 10^9$ K. These rather extreme conditions are only found in some very late stages of stellar evolution. Hence, the equation of state in the case of relativistic partial degeneracy will not be discussed, instead the reader is referred to the specialized literature for this specific topic. In particular, it is worth mentioning here that the most commonly used equation of state is the so-called Helmholtz equation of state. Full details of the Helmholtz equation of state are provided in Timmes and Swesty (2000). This implementation of the equation of state includes contributions from radiation, completely ionized nuclei, and degenerate and relativistic electrons; has perfect thermodynamic consistency; and covers the density and temperature ranges $10^{-12} \leq \rho \text{ (g cm}^{-3}\text{)} \leq 10^{15}$ and $10^3 \leq T \text{ (K)} \leq 10^{13}$, respectively. It is freely available on the Web page of F. X. Timmes,¹ where several additional physical inputs can also be found. Continuing with our treatment, it is immediately clear from (► 1.80) to (► 1.81) that

$$P_e = n_e k_B T \left[\frac{2}{3} \frac{F_{3/2}(\alpha)}{F_{1/2}(\alpha)} \right]. \quad (1.83)$$

This equation gives the pressure of a nonrelativistic, partially degenerate electron gas. Given ρ (or n_e) and T , the value of α can be found from (► 1.80), and $F_{3/2}(\alpha)$ and P_e can be assessed. Note that the term in brackets in (► 1.83) measures the degree to which the pressure of a degenerate electron gas differs from that of a nondegenerate gas. In the nondegenerate limit, $\alpha \rightarrow \infty$, the term in brackets tends to 1, and $P_e \rightarrow n_e k_B T$. Even for $\alpha > 2$, the electron pressure is basically the pressure of a nondegenerate gas. ► Equation 1.83 also shows that the transition from the nondegenerate regime to that of complete degeneracy as the density increases is not discontinuous, but instead there is a gradual transition between those extreme conditions.

These and other concepts become clear when an expansion of the Fermi–Dirac integrals is considered. This is possible in the limits of weak and strong degeneracy. For weak degeneracy ($\alpha > 0$) it can be shown that

$$n_e = \frac{2(2\pi m_e k_B T)^{3/2}}{h^3} \sum_{n=1}^{\infty} (-1)^{n+1} \frac{e^{-n\alpha}}{n^{3/2}} \quad \alpha > 0, \quad (1.84)$$

$$P_e = \frac{2k_B T (2\pi m_e k_B T)^{3/2}}{h^3} \sum_{n=1}^{\infty} (-1)^{n+1} \frac{e^{-n\alpha}}{n^{5/2}} \quad \alpha > 0. \quad (1.85)$$

Note that for very large values of α , only the first term of each series is relevant, and the classical result $P_e = n_e k_B T$ is recovered. Taking the first two terms of the series, the correction to the pressure of an ideal gas due to very weak degeneracy is

$$P_e = n_e k_B T \left[1 + \frac{n_e h^3}{2^{7/2} (2\pi m_e k_B T)^{3/2}} + \dots \right], \quad (1.86)$$

¹http://cococubed.asu.edu/code_pages/codes.shtml

which is valid for large enough α . On the other hand, for strong, but not complete, degeneracy ($\alpha \ll -1$), expansion of the Fermi–Dirac integrals leads to

$$n_e = \frac{8\pi}{3h^3} (2m_e)^{3/2} \mu_q^{3/2} \left[1 + \frac{\pi^2}{8} \left(\frac{k_B T}{\mu_q} \right)^2 + \frac{7\pi^4}{640} \left(\frac{k_B T}{\mu_q} \right)^4 + \dots \right] \quad (1.87)$$

$$P_e = \frac{8\pi}{3h^3} (2m_e)^{3/2} \frac{2}{5} \mu_q^{5/2} \left[1 + \frac{5\pi^2}{8} \left(\frac{k_B T}{\mu_q} \right)^2 + \frac{7\pi^4}{384} \left(\frac{k_B T}{\mu_q} \right)^4 + \dots \right]. \quad (1.88)$$

These expressions allow to compute some basic thermodynamic quantities. For instance, the specific heat at constant volume

$$C_V = \left(\frac{dE}{dT} \right)_{V, N_e} = \left(\frac{\partial E}{\partial T} \right)_{\mu_q} + \left(\frac{\partial E}{\partial \mu_q} \right)_T \frac{d\mu_q}{dT}, \quad (1.89)$$

where $E = (3/2)P_e V$ is the internal energy of electrons. Taking into account that $dn_e = 0$ – the volume of the gas is fixed – and considering only the first order in T , it can be shown that the specific heat at constant volume for a almost completely degenerate electron gas is

$$C_V \approx \frac{\pi^2}{2} \frac{k_B T}{\varepsilon_F} \ll 1. \quad (1.90)$$

Thus, the specific heat for degenerate electrons is proportional to temperature. In the case of strong degeneracy, the specific heat of the electron gas is much smaller than that of an ion gas. This is a somewhat expected result because, as a result of the exclusion principle, only those electrons with energies near the Fermi energy will be able to change its energy as a result of a temperature variation.

Taking into account the contribution of ions and radiation, the equation of state for a mixture of ions, electrons, and radiation can be written

$$P = P_{\text{ion}} + P_e + P_{\text{rad}} = \frac{k_B}{m_H \mu_0} \varrho T + \frac{8\pi}{3h^3} \int_0^\infty \frac{p^3}{1 + \exp[(\varepsilon - \mu_q)/k_B T]} \frac{d\varepsilon_p}{dp} dp + \frac{a}{3} T^4, \quad (1.91)$$

$$\varrho = \mu_e m_H \frac{8\pi}{h^3} \int_0^\infty \frac{p^2}{1 + \exp[(\varepsilon - \mu_q)/k_B T]} dp. \quad (1.92)$$

Note that ions are treated as a Maxwell–Boltzmann gas. This is so because at any given temperature ions require much higher densities than electrons to become degenerate – because of their larger mass, the de Broglie wavelength corresponding to ions is much shorter than that of electrons. For strong degeneracy, ions barely contribute to the total pressure. As a matter of fact, the stronger the degeneracy, the smaller the contribution of the ion gas to the total pressure. However, ions mostly contribute to the mass density (note the presence of m_H in the equation for density) and strongly influence the thermodynamic properties of the mixture as well.

6.1.3 Nonideal Effects

The ion and electron equations of state just described have to be modified in the presence of interactions between particles. In particular, at high densities and low temperatures, Coulomb

interactions modify the thermodynamical properties of the ion gas, and in particular the specific heat. The strength of Coulomb interactions – relative to the thermal kinetic energy – is determined by the Coulomb coupling parameter, which is defined as:

$$\Gamma = \frac{(Ze)^2}{ak_B T} = 0.2275 \frac{Z^2}{T_8} \left(\frac{\rho_6}{A} \right)^{1/3} \quad (1.93)$$

where T_8 is the temperature in units of 10^8 K, ρ_6 is the density in units of 10^6 g cm $^{-3}$, and a is the radius of the Wigner–Seitz sphere:

$$a = \left(\frac{4\pi}{3} n_i \right)^{-1/3}, \quad (1.94)$$

n_i being the number density of ions. For small values of Γ , Coulomb forces are of minor importance (relative to thermal motions) and the ions behave like an ideal noninteracting gas. But, once Γ becomes of order unity, ions begin to undergo short-range correlations, eventually behaving like a liquid. For large enough values of Γ (~ 180), as usually found in the core of cool white dwarfs, they form a lattice structure. Coulomb interactions between the charged particles of the gas make a negative contribution to the pressure and at low densities affect the energetics and ionization states. Various types of interactions between the constituents have to be taken into account in a realistic treatment that eventually leads to “pressure ionization.” Neglecting these interactions and considering only a simple treatment of ionization given by the Saha equation – see (1.71) – leads to the incorrect prediction that matter at the center of the Sun is far from being completely ionized. Sophisticated equations of state for solar-type stellar interior conditions are those of Mihalas et al. (1988) and OPAL (Rogers and Iglesias 1992; Rogers et al. 1996), based on the chemical and physical picture, respectively, whereas for cool white dwarfs the most up-to-date treatment can be found in Segretain et al. (1994) and references therein.

6.2 Nuclear Reactions

6.2.1 General Concepts

Consider a generic reaction between a particle a and a nucleus X to produce a nucleus, Y and a particle b . Usually, both a and b can be either another nucleus, a photon, a proton, a neutron, or an α particle. These are the most common reactions in stellar interiors, although not the only ones since at very large densities electron captures may also play a role. A symbolic way of expressing this is

$$a + X \rightarrow Y + b \quad (1.95)$$

although in a short and more convenient notation this is frequently expressed as $X(a, b)Y$. In all nuclear reactions charge, nucleon number, momentum, and energy must be conserved. However, the sum of the masses of the reactants differs from the sum of the masses of the products. The difference is called the mass excess. The mass excess is linked to the binding energy E_B of the nuclei involved in the nuclear reaction. Consider a nucleus of mass number A , charge Z , and mass $M(A, Z)$. The binding energy – that is, the energy required to separate the nucleons against their mutual attraction by the strong nuclear forces – is defined as

$$E_B = [(A - Z) m_n + Z m_p - M(A, Z)] c^2, \quad (1.96)$$

and the binding energy per nucleon is E_B/A . The mass excess is then

$$\Delta M = M_a + M_X - M_b - M_Y = \Delta E_B/c^2, \quad (1.97)$$

where ΔE_B is the difference of binding energy of the reaction products and reactants. This energy difference is released in exothermic nuclear reactions and is responsible for keeping stable the temperature in stellar interiors. For instance, when hydrogen is burned to form helium, four hydrogen nuclei are converted in one helium nucleus. Given the masses of the proton and of a helium nucleus, 26.5 MeV are released in the process, corresponding to just 0.7% of the rest mass of the system. The binding energy increases as a function of mass number, peaks for ^{56}Fe , and then decreases. Thus, exothermic reactions are either the consequence of thermonuclear fusion of light nuclei or of fission of nuclei heavier than ^{56}Fe into lighter ones. The former constitutes the main energy source in stars, provided that the temperatures are sufficiently high to overcome the Coulomb barrier of the interacting particles. On the other hand, endothermic nuclear reactions are only produced in those environments where an excess of thermal energy is available.

6.2.2 Thermonuclear Reaction Rates

The nuclear reaction rate between two nuclei can be expressed as:

$$r_{aX} = \frac{n_a n_X}{1 + \delta_{aX}} \langle \sigma v \rangle_{aX} = \frac{1}{1 + \delta_{aX}} \frac{\varrho^2 N_A^2 X_a X_X}{A_a A_X} \langle \sigma v \rangle_{aX}, \quad (1.98)$$

where the Kronecker δ prevents from counting twice identical particles, n_a and n_X are the number densities of the reacting particles, X_a and X_X are the corresponding mass fractions, N_A is Avogadro's number, and $\langle \sigma v \rangle_{aX}$ is the product of the nuclear cross section, σ , and the relative velocity, v , of the intervening nuclei, averaged over the distribution of relative velocities. At the densities and temperatures typical of stellar interiors the distribution of velocities of the interacting nuclei can be well approximated by a Maxwellian thermal distribution, because at the relevant conditions found in most stars ions are nondegenerate:

$$\langle \sigma v \rangle_{aX} = \sqrt{\frac{8}{\pi \mu (k_B T)^3}} \int_0^\infty E \sigma(E) \exp\left(-\frac{E}{k_B T}\right) dE \quad (1.99)$$

where E is the relative kinetic energy, and μ is the reduced mass of the colliding nuclei. For nonresonant reactions induced by charged particles the astrophysical S-factor, $S(E)$, is usually introduced instead of the nuclear cross section:

$$\sigma(E) = \frac{S(E)}{E} \exp[-2\pi\eta(E)], \quad (1.100)$$

where $\eta(E) = Z_a Z_X \alpha / v$ is the Sommerfeld parameter, and α is the fine structure constant. The exponential factor measures the probability of penetration of the Coulomb barrier. This is done because the $S(E)$ is a smoothly varying function and can be better extrapolated to low energies, the regions where most of the relevant nuclear reactions occur. Accordingly, after a

Taylor expansion, one finally obtains

$$\begin{aligned}\langle \sigma v \rangle_{aX} &= \sqrt{\frac{2}{\mu k_B T}} \frac{\Delta E_0}{k_B T} f_0 S_{\text{eff}} \exp\left(-\frac{3E_0}{k_B T}\right) \\ &= 1.301 \times 10^{-14} \text{ cm}^3 \text{ s}^{-1} \left(\frac{Z_a Z_X}{A}\right)^{1/3} f_0 \frac{S_{\text{eff}}}{\text{MeV b}} T_9^{-2/3} \exp\left(-\frac{3E_0}{k_B T}\right),\end{aligned}\quad (1.101)$$

where the energy of the Gamow peak E_0 – that is, the energy where the integrand of (1.99) adopts its maximum value, and thus the reaction is most probable – is given by

$$\frac{E_0}{k_B T} = \left(\frac{\pi Z_a Z_X \alpha}{\sqrt{2}}\right)^{2/3} \left(\frac{\mu}{k_B T}\right)^{1/3}, \quad (1.102)$$

and the full width at half maximum of the integrand is

$$\frac{\Delta E_0}{k_B T} = 4 \sqrt{\frac{E_0}{3k_B T}} \quad (1.103)$$

and

$$A = \frac{A_a A_X}{A_a + A_X}. \quad (1.104)$$

Finally,

$$S_{\text{eff}} = S(0) \left(1 + \frac{5k_B T}{36E_0}\right) + S'(0) E_0 \left(1 + \frac{35k_B T}{36E_0}\right) + \frac{1}{2} S''(0) E_0^2 \left(1 + \frac{89k_B T}{36E_0}\right) + \dots \quad (1.105)$$

and T_9 is the temperature in units of 10^9 K. The factor f_0 accounts for the electron screening and will be discussed later. Hence, the values that need to be measured experimentally (or, in the cases where this is impossible, theoretically computed) are $S(0)$, $S'(0)$, and $S''(0)$.

This formulation is only valid for nonresonant reactions. However, there are several nuclear reactions in which there is a resonance. This happens when the energy of the interacting nuclei is coincident with the energy level of an intermediate nucleus that eventually decays to produce the daughter nuclei:



When this occurs, the nuclear reaction rate is enhanced by several orders of magnitude. For the sake of simplicity, and without any loss of generality, it will be assumed that particle Z has only one single excited state and that, moreover, can only decay to particles a and X and b and Y . The probability of these transitions is given by the corresponding lifetimes, τ_a and τ_b , which can be easily expressed by the inverse of the partial widths of the resonance in the incident and exit channels, Γ_a and Γ_b . Thus, the probability of the direct reaction is Γ_b/Γ , where $\Gamma = \Gamma_a + \Gamma_b$, while that of the backward reaction is Γ_a/Γ . In this case it can be shown that the cross section of the resonant nuclear reaction is given by the well-known Breit–Wigner formula:

$$\sigma_{aX}(E) = \pi \lambda^2 \omega \frac{\Gamma_a \Gamma_b}{(E - E_0)^2 + (\Gamma/2)^2}, \quad (1.107)$$

where $\lambda = h/p$, $p = \sqrt{2\mu E}$ is the momentum, E is the energy of the colliding particles, E_0 is the mean energy of the resonance, and

$$\omega = \frac{2J + 1}{(2J_a + 1)(2J_X + 1)} \quad (1.108)$$

is a statistical factor that takes into account the angular momenta of the intervening particles, J is the total angular momentum of the resonant state, and J_a and J_X are the corresponding angular momenta of particles a and X . Obviously, (1.107) shows that the closer the energy to the mean energy of the resonance, the more likely the reaction. Thus, averaging over the Maxwellian distribution, we obtain

$$\langle \sigma_{aX} \rangle = \sqrt{\frac{8}{\mu\pi}} \left(\frac{1}{k_B T} \right)^{3/2} \int_0^\infty E \sigma_{aX}(E) \exp\left(-\frac{E}{k_B T}\right) dE, \quad (1.109)$$

which, taking into account that the resonances are usually very narrow, and thus the Maxwellian distribution barely changes over the width of the resonance, can be very well approximated by:

$$\langle \sigma_{aX} \rangle = \sqrt{2\pi} \left(\frac{1}{\mu k_B T} \right)^{3/2} \omega \hbar^2 \Gamma_a \Gamma_b \exp\left(-\frac{E_0}{k_B T}\right) \int_0^\infty \frac{dE}{(E - E_0)^2 + (\Gamma/2)^2}, \quad (1.110)$$

which, after some elementary algebra, reduces to:

$$\langle \sigma_{aX} \rangle = \hbar^2 \left(\frac{2\pi}{\mu k_B T} \right)^{3/2} \left(\omega \frac{\Gamma_a \Gamma_X}{\Gamma} \right) \exp\left(-\frac{E_0}{k_B T}\right), \quad (1.111)$$

which, in combination with (1.98), provides the thermonuclear rate.

6.2.3 Electron Screening

The thermonuclear reaction rates previously explained assume that the reacting particles are isolated. In fact, in stellar interiors this is not true, as the plasma is totally ionized. Thus, around each nuclei there is a cloud of electrons that shields the positive charges, hence lowering the Coulomb barrier and enhancing the reaction rates. Consequently, the nuclear reaction rates derived in Sect. 6.2.2 must be conveniently modified to account for this. This is done using a multiplicative factor, f . Here, it will be outlined how this screening factor is computed.

In the weak-screening approximation (valid in the case of low densities) the modified electrostatic potential is given by (Salpeter 1954):

$$V(r) = \frac{\alpha Z_a Z_X}{r} e^{-r/r_{\text{DH}}}, \quad (1.112)$$

where all the symbols but the Debye–Hückel radius have been already defined:

$$r_{\text{DH}} = \sqrt{\frac{k_B T}{4\pi\zeta e^2 \varrho N_A}}, \quad (1.113)$$

being

$$\zeta = \sum_i \frac{Z_i(Z_i + 1)}{A_i} x_i, \quad (1.114)$$

where x_i is the number abundance of chemical species i , with charge Z_i and atomic mass number A_i .

The weak-screening approximation holds when the average interaction energy between particles is smaller than the mean particle kinetic energy. In this case the thermonuclear

reaction rate is enhanced by a factor:

$$\ln f = \frac{Z_a Z_X e^2}{r_{\text{DH}} k_B T} \simeq 0.188 Z_a Z_X \sqrt{\frac{\zeta_Q}{T_6^3}}, \quad (1.115)$$

where T_6 is the temperature in units of 10^6 K.

In the case of strong screening (large densities), however, this expression is no longer valid, and another approximation must be used. For the sake of conciseness the formulation of Itoh et al. (1990) will be closely followed, and for simplicity a one-component plasma will also be assumed. For extremely dense stellar plasmas the enhancement factor is computed from two important parameters that characterize dense plasmas: the Coulomb coupling parameter Γ , given by (1.93), and

$$\tau = \left[\left(\frac{27\pi^2}{4} \right) \frac{A m_p (Ze)^4}{\hbar^2 k_B T} \right]^{1/3}, \quad (1.116)$$

where m_p is the proton mass. This parameter gives the classical turning point at the Gamow peak $r \simeq 3\Gamma/\tau$. In this analysis the electrons are assumed to be strongly degenerate. That is, it is assumed that $T \ll T_F$, where T_F is the Fermi temperature, and furthermore the ions are treated as classical particles. Accordingly, their thermal de Broglie wavelength

$$\Lambda = \frac{2\pi\hbar^2}{A m_p k_B T} \quad (1.117)$$

verifies the condition $n_i \Lambda^3 \leq 1$. Under these conditions the enhancement factor is given by:

$$\ln f = 1.25\Gamma - \tau h(\beta), \quad (1.118)$$

where

$$h(\beta) = \frac{0.0455\beta + 0.348\beta^3 + 9.49\beta^6 - 0.123\beta^{12} + 0.101\beta^{13}}{1 + 100\beta^4 + 0.267\beta^{12}}, \quad (1.119)$$

with $\beta = 3\Gamma/\tau$.

6.2.4 Nuclear Networks

Nuclear reactions not only release energy but also change the chemical composition of the stellar interior. This is an important issue as chemical composition changes are an important driver of stellar evolution. Accordingly, the necessary equations to follow the change of chemical composition must be added to the set of equations presented in Sects. 2–4. The change of molar abundances $Y_i = n_i/\rho N_A$ with time due to nuclear processes is followed using a set of coupled ordinary differential equations. If N is the number of chemical species, the nuclear network has a size N^2 . A typical reaction network reads:

$$\frac{dY_i}{dt} = \frac{1}{\rho N_A} \left(- \sum_j Y_j \lambda_{ij} + \sum_{kl} Y_k Y_l \lambda_{kl} - \frac{Y_i}{\tau_i} + \dots \right). \quad (1.120)$$

The first term in (1.120) corresponds to the net destruction of nucleus i due to its interaction with nucleus j , while the second one corresponds to the creation of nucleus i from nuclear reactions involving nuclei k and l . The third term in this equation is the spontaneous radioactive

decay of nuclei i with timescale τ_i . As can be seen the nuclear reaction equations constitute a set of first-order, nonlinear differential equations. In most stellar evolutionary codes, (1.120) is discretized in a convenient way and the chemical changes are obtained inverting a matrix. For the sake of conciseness, it is assumed that nuclear decays can be neglected. If the beginning and the end of the time interval are denoted by superscripts n and $n + 1$, then (1.120) can be linearized and expressed as (Arnett and Truran 1969)

$$Y_i^{n+1} \left(\frac{1}{\Delta t} + Y_j^n \lambda_{ij} \right) + Y_j^{n+1} Y_i^n \lambda_{ij} - Y_k^{n+1} Y_l^n \lambda_{kl} - Y_l^{n+1} Y_k^n \lambda_{kl} = Y_k^n Y_l^n \lambda_{kl} - Y_i^n Y_j^n \lambda_{ij} \quad (1.121)$$

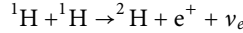
where Δt is the adopted time step. This set of equation can be solved by the Gauss method. Nevertheless, adequate time steps must be chosen. Typically,

$$\Delta t \leq \eta \frac{Y_i^n}{(dY_i/dt)^n} \quad (1.122)$$

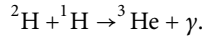
where η is a constant less than unity. Typical values for this constant range from 0.01 to 0.05.

6.2.5 Hydrogen Burning Reactions

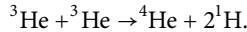
The energy source of main sequence stars is hydrogen burning, and the reaction product is helium. The total energy yield of this nuclear reaction is 26.73 MeV. Depending on whether a star occupies the lower or the upper main sequence, hydrogen burning occurs through the pp-chains or the CNO cycle. For the reader unfamiliar with this terminology, it should be noted that massive stars are located in the upper main sequence, whereas low-mass stars populate the lower main sequence. There are three pp-chains, and in all of them the first step involves a weak interaction to form deuterium and an e^-e^+ annihilation:



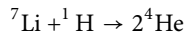
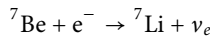
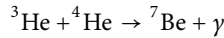
In a second step an isotope of helium is formed:



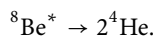
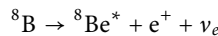
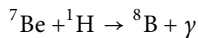
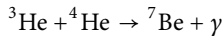
Once enough ^3He is accumulated it follows the so-called pp-I chain:



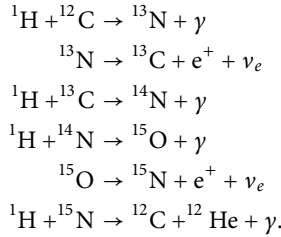
However, for the temperatures typical of main sequence stars ^3He is also burned through the pp-II chain:



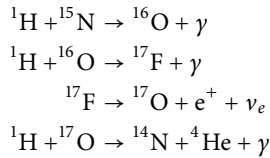
or the pp-III chain, in which an excited state of beryllium decays to form two helium nuclei:



The pp-II chain starts to be relevant for temperatures above $\sim 1.4 \times 10^7$ K, while the pp-III chain operates for temperatures larger than $\sim 2.3 \times 10^7$ K. Finally, it is important to realize that all the nuclei involved in these chains but helium and hydrogen are very unstable, so the interactions are rather inefficient. Actually, the CNO cycle is a more efficient way to produce helium. In fact, CNO cycle is a bi-cycle: the CN cycle, which is dominant for relatively low temperatures, and the NO cycle, which operates mostly for larger temperatures. Both cycles need a seed of carbon, nitrogen, and oxygen to be operative, and these nuclei catalyze the conversion of hydrogen into helium. To begin with, the CN cycle is described next:



In this cycle the reaction with the smallest cross section, which acts as a bottleneck, is the fourth one. Thus, in equilibrium the overall rate is governed by the reaction rate of ${}^{14}\text{N}$. Also, the last reaction of the CN cycle can be substituted by this suite of nuclear reactions:

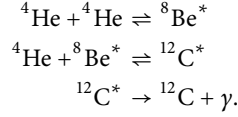


which is known as the ON cycle. According to the previous discussion, the equilibrium CNO abundances are determined by the nuclear cross section of ${}^{14}\text{N}$, and thus the CNO cycle also converts carbon and oxygen into nitrogen. Perhaps, the most commonly used reaction rates for all this suite of reactions are those of Caughlan and Fowler (1988), although there are more recent prescriptions – see, for instance, Adelberger et al. (2011).

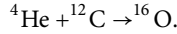
6.2.6 Helium Burning Reactions

Once hydrogen is exhausted in the core of a main sequence star, the pressure gradient cannot be balanced by the energy release of nuclear reactions. Hence, the hydrogen-exhausted core contracts, and becomes hotter. At the same time a hydrogen burning shell is established, the envelope expands, and the star becomes a red giant. Moreover, the envelope becomes convectively unstable and the products of nuclear reactions produced deep in the star are dredged-up to the surface. This is perhaps the most convincing proof of our ability to correctly understand the fine details of stellar structure and evolution, because the theoretically computed abundances of most of the nuclear species match those observationally found. When the temperature reaches $T \sim 10^8$ K helium is ignited through the so-called triple-alpha process, and contraction stops. This nuclear reaction can actually be regarded as a three-body interaction, thereby its name,

and not a normal nuclear reaction. The suite of reactions is




Note that the nuclear isotope ${}^8\text{Be}^*$ is very short-lived at these very high temperatures. This is the reason why this reaction is usually considered a three-body interaction, and also why as soon as some carbon is produced α captures on it rapidly set in:



Thus, the reaction products of helium burning are carbon and oxygen, the ratio between their abundances being extremely dependent on the temperature at which burning proceeds. As mentioned previously, the ${}^{12}\text{C}(\alpha, \gamma){}^{16}\text{O}$ reaction rate is the subject of a vivid debate, as it is one of the most poorly determined experimentally. The nuclear cross section relies on an extrapolation to low energies of the experimental data, which involves several resonances. Since this cross section is crucial in several stellar evolutionary stages, this constitutes a serious drawback of the theory. The most commonly employed nuclear reaction rates of this set of nuclear reactions are those of the NACRE compilation (Angulo et al. 1999).

6.2.7 Carbon Burning Reactions

Once helium is exhausted in the inner regions of the star, its core can no longer support its own gravitational force, and consequently starts to contract again until the temperatures necessary to burn carbon are reached. This occurs at $T \sim 7 \times 10^8$ K. At the same time the star climbs the so-called asymptotic giant branch in the Hertzsprung–Russell diagram. Carbon burning only occurs for stars with masses larger than $\sim 8 M_{\odot}$ (García-Berro et al. 1997; Ritossa et al. 1996). The nuclear reaction network involved in the carbon burning phase is shown in  Fig. 1-2. As can be seen, the number of nuclear isotopes largely increases, and several isotopes of carbon, nitrogen, oxygen, neon, sodium, and magnesium are involved. The most relevant reactions are, nevertheless, the carbon–carbon reaction, yielding either neon or magnesium, and α and proton and neutron captures. The most recent and up-to-date nuclear reaction rates for these last reactions are those of Rauscher and Thielemann (2000, 2001) and Iliadis et al. (2010), whereas in most stellar evolutionary codes the prescription of Caughlan and Fowler (1988) is usually adopted for the fusion of two carbon nuclei.

6.2.8 Other Nuclear Reactions of Interest

After carbon exhaustion there are several other nuclear burning stages. These burning stages only occur for stars with mass larger than $\sim 11 M_{\odot}$ – see, for instance, Ritossa et al. (1999). These burning stages include neon photo-disintegration, oxygen burning, and silicon burning. Neon is burned before oxygen because at the very high temperatures required to overcome the Coulomb barrier of oxygen atoms ($\sim 2.0 \times 10^9$ K), ${}^{20}\text{Ne}$ is photo-disintegrated:



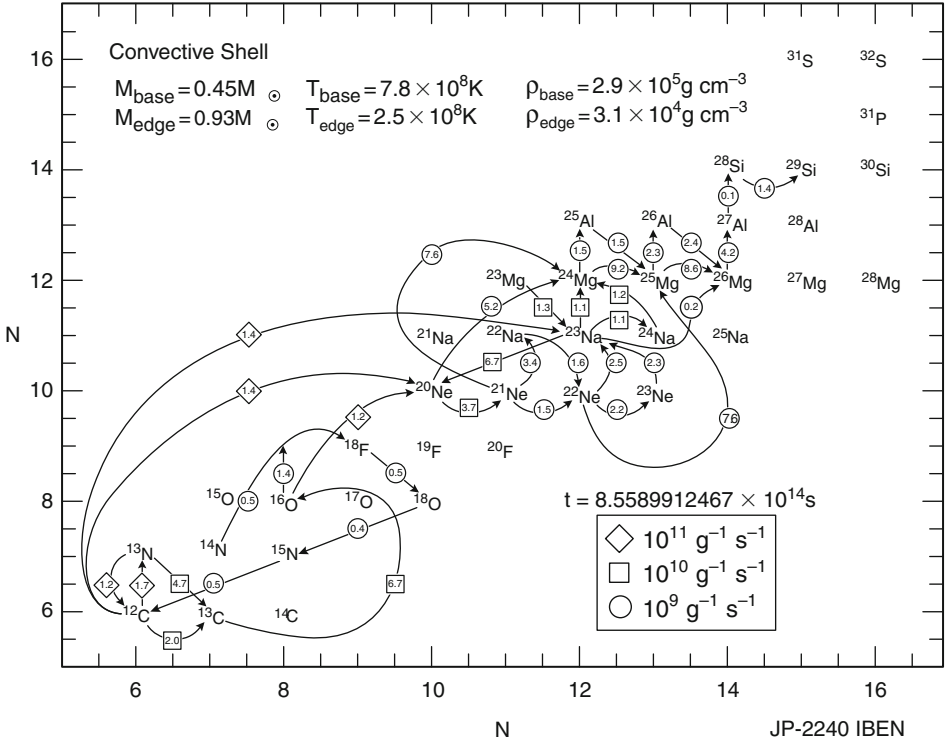


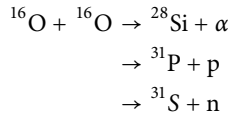
Fig. 1-2

Nuclear reaction network for carbon burning. From García-Berro et al. (1997)

This reaction occurs at $T \sim 1.9 \times 10^9 \text{K}$ and is usually followed by an α capture (using the α particles liberated in the photo-disintegration) on ^{20}Ne to produce ^{24}Mg :



Consequently, the products of neon photo-disintegration are oxygen and magnesium, plus some other products (sodium, zinc, aluminum, silicon, etc.) with much smaller abundances. After neon is exhausted the next nuclear burning stage is oxygen burning:



plus the corresponding recaptures of p, n, and α particles. The final products are mainly ^{28}Si and ^{32}S , and traces of $^{33,34}\text{S}$, $^{35,37}\text{Cl}$, $^{36,38}\text{Ar}$, $^{39,41}\text{K}$, and $^{40,42}\text{Ca}$.

Finally, the last major burning stage is silicon burning. This burning stage involves a very complex network of fast photo-disintegrations, with all sorts of reaction products (protons, neutrons, and α particles), and the corresponding back-reactions. The main products of silicon burning are iron-peak nuclei. However, it is worth mentioning that, at odds with all the previous burning stages, silicon burning occurs at such high temperatures ($T \sim 3.5 \times 10^9 \text{K}$) that reactions

are in quasi-equilibrium. This means that the reactions between the intervening nuclei are so fast that the abundances of such nuclei are close to Nuclear Statistical Equilibrium. Under these conditions the typical reactions are photo-disintegrations to produce neutrons, protons, and α particles, plus the corresponding (γ, α) , (n, α) , (p, α) , and (n, p) reactions – that is, a chain of reactions which are all in equilibrium. Using basic thermodynamical relations the equilibrium abundances can be obtained:

$$n(Z, A) = \frac{g(Z, A)A^{3/2}}{2^A} n_p^Z n_n^{A-Z} \left(\frac{2\pi\hbar^2}{m_p k_B T} \right)^{\frac{3}{2(A-1)}} \exp\left(\frac{E_B(Z, A)}{k_B T} \right) \quad (1.125)$$

where g is the nuclear partition function and the rest of the symbols have been previously defined. To solve this equation the conditions of baryon and charge conservation must also be added. It turns out that in nuclear statistical equilibrium – see, for instance Bravo and García-Senz (1999), and references therein – the composition of matter is essentially determined by the density, the temperature, and the electron fraction, $Y_e = n_e/n$, which is basically the proton-to-nucleon ratio.

6.3 Opacities and Conductivities

6.3.1 Radiative Opacities

As previously discussed in [Sect. 4](#), in stellar interiors the diffusion approximation is valid and the temperature gradient is proportional to the Rosseland mean opacity, obtained as a harmonic mean of the frequency-dependent opacity – see [1.31](#)). Accordingly, opacity is a crucial physical property of stellar matter, since it determines the rate at which energy is transferred from the interior to the outer space. Opacity calculations rely on very detailed descriptions of the interaction between the radiation field and the stellar matter. In fact, bound-free, bound-bound, and free-free interactions between photons and electrons are significant sources of opacity. In turn, these processes depend on the detailed thermodynamical state of the gas, in particular, on the detailed distribution of electrons on the energy levels of the atoms. Consequently, the equation of state plays a major role in the calculation of the opacity.

At high temperatures, when stellar material is almost completely ionized, the main source of opacity is electron scattering with photons. Under these conditions, the cross section for the interaction of photons with electrons is that of Thomson scattering:

$$\sigma_e = \frac{8\pi}{3} \left(\frac{e^2}{m_e c^2} \right)^2 = 0.66 \times 10^{-24} \text{ cm}^2. \quad (1.126)$$

From this equation, the opacity due to electron scattering is computed as $\kappa_e = \sigma_e n_e / \varrho$. Given that $n_e = \varrho / (\mu_e m_H) = 0.5\varrho(1 + X_H)/m_H$, it follows that

$$\kappa_e = 0.2 (1 + X_H) \text{ cm}^2 \text{ g}^{-1}. \quad (1.127)$$

Note that in the case of complete ionization, this opacity source is independent of the temperature and density. Since κ_e is independent of frequency, [1.127](#) provides the Rosseland mean for electron scattering. However, at high enough temperatures, when photons are sufficiently energetic ($k_B T \sim m_e c^2$), the exchange of momentum between electrons and photons has to be considered, thus reducing somewhat the opacity.

The second important source of opacity are free–free transitions. In this case absorption of radiation occurs when a free electron approaches an ion of charge Z . The physical principle operating in this case is quite simple: As an electron passes close to an ionized atom it can be accelerated or decelerated, respectively, absorbing or emitting a photon in the process. Since the photon can be of any wavelength, radiation emitted in this way has a continuous spectrum. These processes can be studied from a classical point of view. It can be shown that the absorption coefficient is proportional to $Z^2 \nu^{-3}$ per ion. At a temperature T , the electron thermal velocity is $\nu^2 \sim T$. Consequently the time interval during which the electron is close enough to the ion scales as $\nu^{-1} \sim T^{-1/2}$. Moreover, at a density ϱ , the number of electron–ion pairs able to take part in the process scales as ϱ , and therefore the contribution to the opacity scales as $\sim \varrho T^{-1/2}$. Since the absorption coefficient per system is proportional to $Z^2 \nu^{-3}$, it follows that

$$\kappa_{\text{ff}}^{\nu} \sim Z^2 \varrho T^{-1/2} \nu^{-3}. \quad (1.128)$$

Thus, the resulting approximate Rosseland opacity due to this process only turns out to be

$$\kappa_{\text{ff}} = 3.68 \times 10^{22} (X_{\text{H}} + X_{\text{He}}) (1 + X_{\text{H}}) \varrho T^{-7/2} \text{ cm}^2 \text{ g}^{-1}, \quad (1.129)$$

which is often referred to as the Kramers opacity. In this expression X_{He} is the abundance by mass of helium.

Absorptions of photons by bound electrons with the result that electrons are removed from the atoms or ions (the so-called bound–free transitions) constitutes another important source of continuum opacity. In the case of photo-ionization of a hydrogen-like system consisting of one electron in the field of charge Ze of an ion, the total bound–free absorption cross section σ_{ν} for a photon of frequency ν is

$$\sigma_{\nu} \sim \frac{Z^4}{\nu^3 n^5} \text{ cm}^2, \quad (1.130)$$

where n is the principal quantum number of the initial state. To compute the total opacity, the number density of atoms in the n -th excited state, N_n , has to be considered. The result is

$$\kappa_{\text{bf}}^{\nu} \sim \sum \frac{N_n \sigma_{\nu}}{\varrho} \text{ cm}^2 \text{ g}^{-1}. \quad (1.131)$$

From this, the Rosseland mean opacity due to bound–free transitions only can be derived. An usual approximation is

$$\kappa_{\text{bf}} = 4.32 \times 10^{25} Z (1 + X_{\text{H}}) \varrho T^{-7/2} \text{ cm}^2 \text{ g}^{-1}, \quad (1.132)$$

where now Z is the mass fraction of elements heavier than helium. Note that this opacity is of the form of Kramers opacity.

Line absorption resulting from photon-induced transitions between bound levels in atoms or ions, the so-called bound–bound transitions, importantly contributes to the total opacity at temperatures below $\sim 10^6$ K. The bound–bound cross section is given by

$$\sigma_{\nu}^{\text{bb}} = \frac{\pi^2 e^2}{m_e c} f_{ij} \phi_{\nu}, \quad (1.133)$$

where f_{ij} is the oscillator strength of the transition from state i to j . This coefficient represents how strongly the line absorbs. Finally, ϕ_{ν} is the line profile function. Because of the large number of transitions (typically a few million lines) that must be taken into account,

bound–bound transitions represent a major complication in opacity calculations. Thermal and pressure broadening effects are important and the total absorption coefficient critically depends on temperature.

Negative hydrogen ions, H^- , provide another source of opacity. The hydrogen atom can harbor a second electron in a bound state with a binding energy of only 0.7544 eV. This energy being so small means that the H^- ion will be easily ionized if temperatures exceed a few thousand degrees. In fact, all photons with wavelengths $\lambda < 1.65 \mu\text{m}$ have enough energy to ionize the H^- ion back to a neutral hydrogen atom plus a free electron. Additionally, it is clear that the total opacity must be proportional to the electron density. However, it also has to be taken into account that the electrons needed to form H^- come from ionized metals. According to these considerations it turns out that in hot stars the H^- ion is so rapidly ionized that this source of opacity is not relevant. On the contrary, in cool stars there are too few free electrons (since no ionized metals are present) to form H^- ions, so the contribution to the opacity is negligible as well, whereas for solar-like stars the H^- ion is the dominant source of continuum opacity at optical wavelengths. In particular, the opacity due to the negative hydrogen ion is the most important source of opacity in the solar atmosphere.

For temperatures lower than $\approx 5,000 \text{ K}$, the contribution from molecules becomes a relevant opacity source. The contributions to molecular opacity arise from transitions between rotational and vibrational molecular states, which are very frequent at these very low temperatures, typical, for instance, of late-type main sequence stars and cool red giants. Also the formation of small dust grains at even lower temperatures ($\approx 1,700 \text{ K}$) has to be considered in the opacities needed to compute reliable atmosphere models of cool stars.

As stems from the preceding paragraphs, which is the dominant source of opacity depends very much on the thermodynamical state of stellar matter. In general, at low temperatures, bound–bound and bound–free processes totally dominate the opacity. At larger temperatures, when ionization is almost complete, free–free transitions take over and become the most important source of opacity. These processes become less relevant at even larger temperatures, where scattering by free electrons is the major source of opacity. Needless to say, all opacity processes contribute simultaneously to the total opacity, and to compute the resulting value of the radiative opacity, κ_{rad} , the sum of the contributions of the monochromatic opacities resulting from all the processes described so far has to be done. After doing that, the integral in (☛ 1.24) must be computed. It is important to realize at this point of our discussion that the sum of the Rosseland mean opacities of each opacity component is not equal to the Rosseland mean of the sum. Hence, the Rosseland mean of a given process lacks meaning, except in the particular case in which the corresponding process dominates by far the rest of the contributions to the opacity. For typical stellar compositions, the largest values of the radiative opacity occur for temperatures between $\sim 10^4$ and 10^5 K , corresponding to the temperatures appropriate to ionize hydrogen and helium, the principal constituents of stellar matter.

Several groups have computed detailed frequency-averaged opacities. Among them the most reliable and frequently used opacities are those of the OPAL (Rogers and Iglesias 1992) and of the OP (Seaton et al. 1994) projects. Both projects have presented their calculations in the form of extensive tabulations for different chemical compositions for a wide range of temperatures and densities. At very low temperatures the most up-to-date and comprehensive set of opacity data has been assembled by Ferguson et al. (2005). Their opacities cover temperatures between 30,000 and 500 K, and, in addition to the usual continuous absorbers, include atomic and molecular line absorption, and grain absorption and scattering (by silicates, iron, carbon, and SiC). More recently, Lederer and Aringer (2009) have presented calculations of Rosseland

opacities for low temperatures that consider the effects of varied abundances of carbon and nitrogen for different metallicities.

6.3.2 Conduction

At the high densities found in the core of evolved stars, the main mechanism of energy transfer is not radiation but conduction by degenerate electrons. For the conditions found in the deep cores of these stars, the mean-free path of electrons is usually very large. This is so because after an interaction in which an electron is involved its energy is in most cases below the Fermi level. Consequently, since below the Fermi energy all the levels are occupied, the electron has to travel long distances to release the excess energy absorbed in the interaction. Of course, this only occurs when $k_B T \ll \varepsilon_F$, that is, in conditions of strong degeneracy. Hence, electrons can transport energy very efficiently and the related opacity is small ($\kappa \sim 1/\ell$). It can be shown that, for strong degeneracy, an approximate expression for the conductive opacity is

$$\kappa_{cd} \sim 5 \times 10^{-3} \frac{\sum Z_j^2 X_j \theta_j / A_j}{(1 + X_H)^2} \frac{T_7^2}{(\varrho/10^5)^2} \text{ cm}^2 \text{ g}^{-1}, \quad (1.134)$$

where $\theta_j \sim 0.85/Z_j^{1/3}$, T_7 is the temperature in units of 10^7 K, and X_j the abundance by mass of element j . Note that the larger the density, the smaller the conductive opacity. The reason for this is that, as already explained, for strong degeneracy it becomes very difficult for an electron to scatter, with the consequent result that the mean-free path becomes very long. For helium composition and $T \sim 10^7$ K and $\varrho \sim 10^5 \text{ g cm}^{-3}$, the conductive opacity is two orders of magnitude smaller than the electron scattering opacity (κ_{es}). Accordingly, the total opacity given by (1.31) is very small, of the order of $10^{-2} \kappa_{es}$. Finally, it is rather evident from the previous discussion that at low densities, electrons transport little energy compared to photons.

More elaborated prescriptions than that given by (1.134) are available in the referenced literature. In particular, detailed calculations of conductive opacities for a wide range of stellar conditions have been given by Itoh et al. (1983). This set of conductive opacities is the most commonly adopted in nearly all the state-of-the-art stellar evolutionary codes. However, there are other prescriptions, the most recent one being that of Cassisi et al. (2007). These authors provide determinations of the conductive opacity covering the entire parameter space relevant to stellar evolution models (i.e., both the regime of partial and high electron degeneracy).

6.4 Neutrino Losses

Photons are not the only way in which energy is lost from stars. Neutrinos also play an important role, especially during the last stages of stellar evolution. These weakly interacting particles escape freely from the dense cores of evolved stars – except in the case of core-collapse supernovae, where the densities are extremely high – carrying out copious amounts of energy. This is because under most circumstances (but in those found in the central very dense regions of these supernovae) their interaction cross section with matter is extremely small ($\sim 10^{-44} x^2 \text{ cm}^2$, x being the neutrino momentum in units of MeV/c). Thus, their mean-free paths are very large and, hence, they can be considered as an instantaneous sink of energy for most stellar evolutionary stages. According to this discussion, (1.18) must be correspondingly modified to include an additional term $-\epsilon_\nu$, which acts as an instantaneous sink of energy.

Besides the nuclear processes in which neutrinos are produced as a regular product of the reaction – like electron captures or the several reactions described in ● Sect. 6.2, where the neutrino energy loss is already included in the nuclear energy released – there are four typical processes in which neutrinos are produced as a result of purely leptonic processes: the photoneutrino, plasmaneutrino, pair annihilation, and bremsstrahlung processes, respectively. In the photoneutrino process the outgoing photon in a regular Compton scattering is substituted by the emission of a neutrino–antineutrino pair. In the plasmaneutrino process a photon moving in the electron gas – the so-called plasmon – is transformed in a neutrino–antineutrino pair. Pair annihilation neutrinos are produced at very high temperatures ($T > 10^9$ K) in the reaction $e^- + e^+ \rightleftharpoons \nu_e + \bar{\nu}_e$, instead of a normal photon. Finally, the emission of a neutrino–antineutrino pair due to the interaction of a high-energy electron with a heavy nucleus is known as the neutrino bremsstrahlung process, where again the photon is replaced by the neutrino pair. Additionally, there is a less common neutrino process, which only occurs for nuclei with relatively large values of Z and at low densities and temperatures – a situation not commonly found in stellar interiors – which is the recombination neutrino process. In this process an electron in the continuum makes a transition to a bound state and a neutrino pair is emitted.

Analytic expressions for all these processes as a function of the density, the temperature, and the chemical composition have been compiled by Itoh et al. (1996). For the sake of conciseness and given that these analytical expressions incorporate coefficients that depend on the detailed chemical composition – which are listed in lengthy tables – their results will not be reproduced here. Instead, the interested reader is referred to their paper. These neutrino emission rates are the most commonly used in stellar evolutionary calculations. However, other expressions can be found in the specialized literature – see, for instance, Haft et al. (1994).

7 Other Physical Processes

During some evolutionary phases there are other physical processes which become of the maximum importance. This is especially true for compact stars, and specifically for white dwarfs, for which element diffusion and radiative levitation are crucial. Element diffusion is of the maximum importance to understand the observed purity of white dwarf atmospheres. The very large surface gravities of these stars makes elements heavier than hydrogen to settle down over timescales shorter than the evolutionary ones, leaving behind pure hydrogen or helium atmospheres. In these stars, the interfaces between the hydrogen (if present) and helium layers, and between the helium buffer and the core are the product of element diffusion. The study of element diffusion processes in white dwarfs dates back to the early studies of Schatzman (1958). Since then, numerous studies have been devoted to this subject, and have convincingly demonstrated that gravitationally induced diffusion is an extremely efficient process in the envelopes of white dwarfs.

Under the influence of gravity, partial pressure, thermal gradients, and induced electric fields, the diffusion velocities in a multicomponent plasma satisfy a set of $N - 1$ independent linear diffusion equations (Burgers 1969):

$$\frac{dp_i}{dr} - \frac{Q_i}{Q} \frac{dp}{dr} - n_i Z_i e E = \sum_{j \neq i}^N K_{ij} (w_j - w_i) + \sum_{j \neq i}^N K_{ij} z_{ij} \frac{m_j r_i - m_i r_j}{m_i + m_j}, \quad (1.135)$$

coupled with the equation of heat diffusion:

$$\begin{aligned} \frac{5}{2} n_i k_B \nabla T = & -\frac{5}{2} \sum_{j \neq i}^N K_{ij} z_{ij} \frac{m_j}{m_i + m_j} (w_j - w_i) - \frac{2}{5} K_{ii} z_{ii}'' r_i \\ & - \sum_{j \neq i}^N \frac{K_{ij}}{(m_i + m_j)^2} (3m_i^2 + m_j^2 z_{ij}' + 0.8m_i m_j z_{ij}'') r_i \\ & + \sum_{j \neq i}^N \frac{K_{ij} m_i m_j}{(m_i + m_j)^2} (3 + z_{ij}' - 0.8z_{ij}'') r_j. \end{aligned} \quad (1.136)$$

In these equations p_i , Q_i , n_i , Z_i , and m_i denote, respectively, the partial pressure, mass density, number density, and charge and mass of chemical species i , N is the number of ionic species plus the electron, and the rest of the symbols have their usual meaning. The unknown variables are the diffusion velocities with respect to the center of mass, w_i , and the residual heat flows r_i . In addition, the electrical field E has to be determined. The resistance coefficients K_{ij} , z_{ij} , z_{ij}' , and z_{ij}'' must also be provided (Paquette et al. 1986a, b) to solve these equations. In total there are $2N - 1$ equations and $2N + 1$ unknowns. The equations for no net mass flow and no electrical current provide two additional conditions that must be fulfilled:

$$\sum_i A_i n_i w_i = 0 \quad (1.137)$$

$$\sum_i Z_i n_i w_i = 0. \quad (1.138)$$

► Equation 1.135 can be manipulated to obtain:

$$\frac{1}{m_i} \left[\sum_{j \neq i}^N K_{ij} (w_i - w_j) + \sum_{j \neq i}^N K_{ij} z_{ij} \frac{m_i r_j - m_j r_i}{m_i + m_j} \right] - Z_i e E = \alpha_i - k_B T \frac{d \ln n_i}{dr}, \quad (1.139)$$

where

$$\alpha_i = -A_i m_H g - k_B T \frac{d \ln T}{dr}, \quad (1.140)$$

g being the gravitational acceleration. The unknowns w_i , r_i , and E can be written in terms of the gradient of ion densities (similarly for r_i and E):

$$w_i = w_i^{\text{gt}} - \sum_j \sigma_{ij} \frac{d \ln n_j}{dr}, \quad (1.141)$$

where w_i^{gt} stands for the velocity component resulting from gravitational settling and thermal diffusion, and the summation is done for the ions only. This equation, together with (► 1.136–1.138) can be solved using matrix inversion to find w_i^{gt} and σ_{ij} , which finally are employed to follow the evolution of the ionic species:

$$\frac{\partial n_i}{\partial t} = -\frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \left(w_{ij}^{\text{gt}} n_i - \sum_j \frac{n_i}{n_j} \sigma_{ij} \frac{\partial n_j}{\partial r} \right) \right]. \quad (1.142)$$

In the otherwise pure atmospheres of white dwarf stars there are occasional appearances of other elements. Consequently, a competing mechanism to gravitational diffusion must be at work. One of these is radiative levitation, especially for hot white dwarfs – those with effective temperatures larger than $\sim 20,000$ K. The physical principle operating here is that ions with large

charges have larger cross sections and, thus, absorb more photons. This is done usually through bound-bound interactions, the net effect being that these ions acquire momentum from the radiation field, thus leading to differential accelerations that make them appear on the surface of the star. In virtually all recent calculations it is assumed that a given levitating element is trace. That is, it is assumed that its abundance has no relevant influence on the structure of the star. Additionally, it is assumed that the approximation of equilibrium abundances holds, and thus that the radiative acceleration can be equated to the local effective gravity at each mass shell of the envelope. To solve for the abundances the total radiative acceleration for element i through bound-bound absorption,

$$g_i^{\text{rad}} = \frac{4\pi}{c} \frac{1}{X_i} \sum_j \sum_{k>j} \int_0^\infty \kappa_i^{jk}(\nu) H_\nu d\nu \quad (1.143)$$

must be computed. In this expression X_i is the mass fraction of the corresponding element, κ_i^{jk} is the monochromatic opacity due to the bound-bound absorption from state j to level k , and H_ν is the monochromatic Eddington flux. In local thermodynamic equilibrium,

$$\kappa_i^{jk}(\nu) = \frac{N_i^j}{Q} \frac{\pi e^2}{m_e c} f_i^{jk} \phi_i^{jk}(\nu) (1 - e^{-h\nu/k_B T}), \quad (1.144)$$

where N_i^j is the number density of ions of type i in excitation state j , f_i^{jk} is the oscillator strength of the bound-bound transition from j to k , and ϕ_i^{jk} is the normalized line profile of such transition. This acceleration must be equated for each ion to the local gravity at each mass shell. Clearly, computing radiative accelerations requires improved oscillator strengths and reliable atomic data. Modern calculations (Chayer et al. 1995) rely on the data bank TOPBASE (Cunto and Mendoza 1992), associated with the Opacity Project (Seaton et al. 1994).

There are other physical processes that affect stellar evolution like mass loss, rotation, and magnetic fields. Mass loss is relevant in giant and luminous stars and it is important to explain some observations. On the other hand, stars are observed to rotate. Rotation affects star evolution particularly via the associated meridional circulation and other processes that lead to chemical mixing – see Maeder (2009). Additional physical processes associated with changes in the chemical abundance distribution may affect the evolution of evolved star. This is the case of old and cool white dwarf stars, where the gravitational energy released during carbon-oxygen phase separation upon crystallization constitutes an additional source of energy that markedly impacts the cooling times of these stars (Isern et al. 2000).

8 Boundary Conditions and Stellar Atmospheres

As it will be shown below, to integrate the equations of stellar structure, boundary conditions are needed at both the stellar center and the surface. At the stellar center, that is, at $m = 0$, both the radius and the luminosity vanish. This leads to singularities. It is not difficult to show that near the center a first-order Taylor expansion of the equations of stellar structure previously derived in [Sects. 2–4](#) yields

$$r = \left(\frac{3}{4\pi Q_c} \right)^{1/3} m_c^{1/3} \quad (1.145)$$

$$P = P_c - \frac{3G}{8\pi} \left(\frac{4\pi\varrho_c}{3} \right)^{4/3} m_c^{2/3} \quad (1.146)$$

$$l = (\varepsilon_{\text{nuc}} - \varepsilon_v + \varepsilon_g)_c m_c \quad (1.147)$$

$$T^4 = T_c^4 - \frac{1}{2ac} \left(\frac{3}{4\pi} \right)^{2/3} \kappa_c (\varepsilon_{\text{nuc}} - \varepsilon_v + \varepsilon_g)_c \varrho_c^{4/3} m_c^{2/3} \quad (\text{radiative}) \quad (1.148)$$

$$\ln T = \ln T_c - \left(\frac{\pi}{6} \right)^{1/3} G \frac{\nabla_{\text{ad},c} \varrho_c^{4/3}}{P_c} m_c^{2/3} \quad (\text{convective}), \quad (1.149)$$

where the suffix “c” means the central values of the physical variables and m_c is the mass of the central sphere.

The treatment of the surface conditions is more difficult, partly because it involves the difficult issue of how energy is transported in the very outer layers of the star. At the surface, $m = M$ and $l = L$, where L is the surface luminosity that has to be determined. Deriving the boundary conditions at the surface requires the computation of model atmospheres. Usually, the Eddington gray approximation for radiative transfer in the case of plane-parallel geometry is used:

$$\frac{dP}{d\tau} = \frac{g}{\kappa_{\text{rad}}} \quad (1.150)$$

$$T^4 = \frac{3}{4} T_{\text{eff}}^4 \left(\tau + \frac{2}{3} \right) \quad (1.151)$$

where g is the surface gravity and τ is the optical depth, defined by $d\tau = \kappa_{\text{rad}} \varrho dr$. The first of these equations corresponds to the equation of hydrostatic equilibrium. Integration of these two equations from starting values ($P \approx 0$, $T \approx 0$ at $\tau = 0$) down to $\tau = 2/3$ yields P_s . The layer at $\tau = 2/3$ defines the location of the photosphere from which most of the radiation is emitted into the space. There the temperature is $T = T_{\text{eff}}$, where T_{eff} is the effective temperature of the star, which is defined by the relation $L_s = 4\pi R^2 \sigma T_{\text{eff}}^4$, $\sigma = ac/4$ being the Stefan–Boltzmann constant. Note that T_{eff} corresponds to the temperature of a blackbody that has the same energy flux of the real star. From here, a two-parameter set of boundary conditions between T_{eff} , P_s , L_s , and R can be cast. For instance, assuming arbitrary values of R and L_s all the other values at the surface can be obtained. In fact, from the values of R and L_s , the value of T_{eff} is determined. Also, the value of g is obtained as well, since the total stellar mass is given. This information together with the chemical composition of the atmosphere (which is assumed to be the same as the composition of the outer layers, and thus known) allows to integrate the model atmosphere and to find P_s .

Finally, it is worth mentioning that during certain stages of evolution, a better treatment for the outer layers than that given by the Eddington approximation is needed. This is the case for instance in giant stars, where the effects of sphericity are usually considered. In cool and dense atmospheres, like those of white dwarfs, the results from full non-gray model atmospheres that take into account convection has to be considered. This is necessary for a proper determination of the envelope structure as well as the cooling time of such stars.

9 Numerical Techniques: Modern Implementations

The basic equations that govern the structure and evolution of a spherically symmetric star in hydrostatic equilibrium have been derived in [Sects. 2–4](#). In terms of the Lagrangian variable, m , they can be written as

$$\frac{\partial r}{\partial m} = \frac{1}{4\pi r^2 \varrho} \quad (1.152)$$

$$\frac{\partial P}{\partial m} = -\frac{Gm}{4\pi r^4} \quad (1.153)$$

$$\frac{\partial l}{\partial m} = \varepsilon_{\text{nuc}} - \varepsilon_v - C_p \frac{\partial T}{\partial t} + \frac{\delta}{\varrho} \frac{\partial P}{\partial t} \quad (1.154)$$

$$\frac{\partial T}{\partial m} = -\frac{GmT}{4\pi r^4 P} \nabla. \quad (1.155)$$

To this set of equations, the equation for the variation of chemical composition with time has to be added. This, for radiative regions, must be done using the set of equations presented previously in [Sect. 6.2.4](#) – see ([1.120](#)) – whereas if chemical changes are due to different mixing processes, a diffusion-like set of equations must be used:

$$\left(\frac{\partial Y_i}{\partial t} \right) = \left(\frac{\partial Y_i}{\partial t} \right)_{\text{nuc}} + \frac{\partial}{\partial m} \left[(4\pi r^2 \varrho)^2 D \frac{\partial Y_i}{\partial m} \right], \quad (1.156)$$

where the first term represents the chemical changes due to nuclear reactions and the second one provides the changes of the chemical abundances resulting from mixing processes, which are characterized by a diffusion coefficient D . In general, the equations for the abundance changes are decoupled from the set of equations providing the thermal and mechanical changes, namely, ([1.152–1.155](#)). This is a valid assumption only in the case in which the nuclear timescale is much larger than the other relevant timescales of the star. This allows to treat the problem by solving first ([1.152–1.155](#)) at a given time t_0 and for a given chemical composition. Then the change in chemical composition during the time step Δt – that is, the chemical composition at time $t_0 + \Delta t$ – is obtained using the physical variables determined at t_0 . Afterward the new chemical compositions are used recursively to find the new physical variables at $t_0 + \Delta t$, and so on. However, in some advanced stages of evolution, where the changes of chemical compositions occur in timescales comparable to that of the physical variables, decoupling the equations for the chemical changes from the other four equations is not a realistic assumption. The term ε_v in ([1.154](#)) represents the energy losses by neutrino in $\text{erg g}^{-1} \text{s}^{-1}$, as discussed in [Sect. 6.4](#). As mentioned, for most stellar conditions, neutrinos do not interact with matter, thus they only act as an energy sink of energy at the location where they are produced. In ([1.155](#)), $\nabla = \nabla_{\text{rad}}$ – as given by ([1.33](#)) – if the energy transport is done by radiation (and conduction). In the case where energy transport is due to convection, $\nabla = \nabla_{\text{conv}}$, where ∇_{conv} is specified by the adopted treatment of convection. In addition, the functions that describe the properties of stellar matter – ϱ , δ , α , C_p , ∇_{ad} , ε_{nuc} , ε_v , κ_{tot} , and ∇_{conv} – that depend on P , T , and chemical composition, must be specified.

[Equations 1.152–1.155](#) constitute a set of nonlinear, partial differential equations that have to be integrated numerically. The independent variables are the mass coordinate m , which runs from zero to the value of the total mass M – the index r is dropped in the notation – and t .

To solve these equations, the total mass M of the star and proper initial values, usually only the chemical composition at the initial instant t_0 , together with appropriate boundary conditions at $m = 0$ and $m = M$ must be specified. These boundary conditions have been discussed previously, in [Sect. 8](#). The solution for r , P , l , T , and chemical composition at a given time in the whole mass interval is named a *stellar model*.

A standard technique for solving the equations of stellar structure and evolution is based on an iterative implicit technique, the so-called the Henyey method or relaxation method, in which the differential equations are replaced by a set of difference equations in a discrete number of mass shells (or grid). Here, an initial or approximate solution is gradually improved by applying successive simultaneous corrections to all variables at all points on the grid. The corrections to the physical variables are applied until they become insignificant and the initial model “relaxes” to the correct one. To implement the technique, the star is divided into N concentric mass shells (very few stellar evolutionary codes use the radius as independent variable), each of them labeled with the lagrangian coordinate m_j , where $m_1 = 0$ and $m_N = M$. In addition, the time derivatives appearing in [\(1.154\)](#) are usually replaced at any point j of the grid by

$$\left(\frac{dP}{dt}\right)_{j+1/2} = \frac{P_{j+1/2} - P_{j+1/2}^*}{\Delta t} \quad (1.157)$$

$$\left(\frac{dT}{dt}\right)_{j+1/2} = \frac{T_{j+1/2} - T_{j+1/2}^*}{\Delta t}, \quad (1.158)$$

where P^* and T^* , evaluated at the middle of the mass interval, denote the values of P and T at the earlier time $t_0 - \Delta t$, and are, thus, known quantities. With these considerations, [\(1.152–1.155\)](#) are converted to a set of ordinary differential equations, which can be replaced by the corresponding difference equations. For instance, the equation of hydrostatic equilibrium is replaced by the following difference equation:

$$\frac{P_{j+1} - P_j}{m_{j+1} - m_j} = -\frac{G}{4\pi} \frac{m_{j+1/2}}{r_{j+1/2}^4}, \quad (1.159)$$

where $r_{j+1/2} = (r_{j+1} + r_j)/2$ and $m_{j+1/2} = (m_{j+1} + m_j)/2$. For each mass shell j , the resulting four difference equations ($i = 1, 4$) can be written in the form

$$G_j^i \equiv \frac{y_{j+1}^i - y_j^i}{m_{j+1} - m_j} - f_i(y_{j+1/2}^1, \dots, y_{j+1/2}^4) = 0, \quad (1.160)$$

where $y^1 = r$, $y^2 = P$, $y^3 = l$, and $y^4 = T$. The index j runs from 2 to $N - 1$. At the center ($j = 1$), singularities occur. There, the radius and luminosity vanish, that is, $r = 0$ and $l = 0$. As already shown – see [Sect. 8](#) – the boundary conditions at the center ($j = 1$) can be obtained by performing a Taylor expansion of [\(1.152\)](#) and [\(1.153\)](#) – see [\(1.145–1.149\)](#) – and it can be easily shown that they adopt the form

$$C^i(y_2^1, y_2^2, y_2^3, y_2^4, y_1^2, y_1^4) = 0, \quad (1.161)$$

with $i = 1, 4$. As for the boundary conditions at the surface ($j = N$) – see [Sect. 8](#) – the following equations can be established:

$$S^1 \equiv y_N^2 - h_s(y_N^1, y_N^3) = 0 \quad (1.162)$$

$$S^2 \equiv y_N^4 - q_s(y_N^1, y_N^3) = 0. \quad (1.163)$$

Usually, a linear approximation is considered for the functions h_s and q_s . Thus, only three integrations of the model atmosphere – (• 1.150) and (• 1.151) – are needed to find all the coefficients that define h_s and q_s .

Note that the total number of unknowns is $4N - 2$ and the total number of equations is also $4N - 2$. Consider now a given time t and assume that $(y_j^i)_1$ with $i = 1, \dots, 4$ and $j = 1, \dots, N$ is a first approximation to the solution of the set of equations, which as a first guess may be taken as the solution at the previous time $t - \Delta t$. Since the values of $(y_j^i)_1$ are only an approximation to the solution to the set of equations and not the real one, it is clear that $G_j^i(1) \neq 0$, $C^i(1) \neq 0$, and $S^{1,2}(1) \neq 0$. Here, the label “1” refers to the initial guess. Let now δy_j^i be the corrections to all variables at all mass points. A second approximation can be then obtained:

$$(y_j^i)_2 = (y_j^i)_1 + \delta y_j^i, \quad (1.164)$$

so that now $G_j^i(2) = 0$, $C^i(2) = 0$, and $S^{1,2}(2) = 0$. If the corrections are small, a first-order Taylor expansion yields:

$$G_j^i(1) + \sum_{m=1}^4 \frac{\partial G_j^i}{\partial y_j^m} \delta y_j^m + \sum_{m=1}^4 \frac{\partial G_j^i}{\partial y_{j+1}^m} \delta y_{j+1}^m = 0 \quad i = 1, \dots, 4 \quad j = 2, \dots, N-1, \quad (1.165)$$

and similarly for (• 1.161), (• 1.162), and (• 1.163). The derivatives of G_j^i are evaluated using the first approximation as arguments. From here, the corrections δy_j^i can be found after inverting a band-type matrix (Henye matrix) which has nonvanishing elements only in blocks near the diagonal (Henye et al. 1959, 1964; Larson and Demarque 1964; Nobili and Turolla 1988). Indeed, note that difference equations depend only on variables at adjacent points. This fact enormously simplifies the inversion technique to find the corrections δy_j^i . After consecutive iterations, the approximate solution can be improved until the absolute value of all corrections drops below a given limit or difference equations are satisfied to the desired accuracy. To compute an evolutionary sequence, after convergence of the model at time t , a new time step is determined, the chemical composition for the new time is assessed from the stellar model converged at t , which is also used as an initial solution for the new time. The Henye technique is a relaxation technique, and has proved to be extremely efficient and stable. It should be stressed, however, that the equations of stellar structure and evolution are strongly nonlinear, so it is expected that, since the corrections are obtained from linearized equations, the solution requires several iterations in some cases, particularly if the initial approximation is not a good representation of the solution or if the time step is not small enough. Nevertheless, in most cases, only few iterations are needed to find the solution.

10 State-of-the-Art Modeling

Building a state-of-the-art stellar evolutionary code requires large efforts. As has been made clear in • Sect. 6, detailed physical inputs are required to produce reasonable models. Not only that, it has also been shown in • Sect. 9 that the system of differential equations that governs the structure and evolution of stars must be solved using specific techniques. This, as mentioned earlier, is a tough task as it involves the solution of a boundary problem.

There are several existing evolutionary codes which incorporate reliable solvers of the equations of stellar structure and up-to-date physical inputs. Among these it is worth mentioning, in alphabetical order, the following ones, but the reader should keep in mind that this list is not intended to be exhaustive whatsoever:

- ASTEC – Aarhus STellar Evolution Code (Christensen-Dalsgaard 1982)
- CESAM – Code d'Evolution Stellaire Adaptatif et Modulaire (Morel and Lebreton 2008)
- DSEP – Darmouth Stellar Evolution Program (Chaboyer et al. 2001)
- EVOL (Blöcker 1995; Herwig 2000)
- EVOLVE (Iben 1965), which was subsequently modified to follow carbon burning (García-Berro and Iben 1994)
- GARSTEC – Garching STellar Evolution Code (Weiss and Schlattl 2008)
- FRANEC – Frascati Raphson Newton Evolutionary Code, of which there are several versions, see for instance Pietrinferni et al. (2004)
- GENEC – Geneva Evolution Code (Meynet and Maeder 2000)
- KEPLER (Heger et al. 2000; Weaver et al. 1978)
- LPCODE – La Plata Stellar Evolutionary Code (Althaus et al. 2005)
- MESA – Modules for Experiments in Stellar Astrophysics (Paxton et al. 2011)
- STARS (Pols et al. 1995)
- STERN (Heger et al. 2000; Langer 1998)
- TGEC – Toulouse-Genève Evolution Code (Richard et al. 1996)
- TYCHO (Young and Arnett 2005)
- YREC – Yale Rotating Stellar Evolution Code (Demarque et al. 2008).

Fortunately, some of these codes are freely available on the Internet – although, of course, some work must be done to adapt them to the specific problem under study – whereas others can be used with permission of the original authors. Finally, others are proprietary codes and permission to use them is not granted.

Describing in detail the variety of physical inputs adopted in each one of these stellar evolutionary codes is beyond the scope of this chapter, so the interested reader is referred to the cited literature. However, a few words regarding the general features of these codes are in order. There are some codes that are specialized at following the initial stages of stellar evolution, for instance ASTEC. Many others are able to follow only the hydrogen and helium burning phases and the thermal pulses on the Asymptotic Giant Branch, but are not able to follow in a detailed and realistic way the carbon burning phase – FRANEC and GARSTEC are good examples, but not the only ones – however they perform extremely well at producing realistic sets of isochrones, while others are specialized in following specific evolutionary stages, for instance the white dwarf evolutionary phase – the leading example is LPCODE – and they produce very accurate cooling tracks. Other codes (GENEC and YREC) incorporate the effects of slow rotation. Finally, other codes are more suitable to follow the evolution of massive stars all the way to the supernova stage – KEPLER being an example of these type of codes. Also, most of the freely available stellar evolutionary codes, but a few ones, rely on gray atmospheres and some adaptations must be made to incorporate non-gray detailed atmospheres.

It is also important to mention that although all these codes are specifically adapted to the research needs and interests of the authors and, consequently, the physical inputs are, as previously said, extremely diverse, there are as well a set of basic ingredients which are common to all of them. One of these is the usage of the OPAL opacities previously discussed in [Sect. 6.3](#). The most up-to-date stellar codes use this set of opacities, which are a reference in the field.

However, prescriptions for the convective instability are very different from one another, as also are other physically interesting inputs. Specifically, most stellar evolutionary codes use the classical mixing length treatment of Böhm-Vitense (1958) and the onset of convection is determined according to the Schwarzschild criterion ($\nabla_{\text{ad}} - \nabla_{\text{rad}} < 0$), see 🔍 Sect. 4.3, although some (variable) amount of convective overshooting is allowed to occur, and only very few stellar evolutionary codes incorporate more sophisticated theories of convection. Finally, it is also important to realize that most of these state-of-the-art codes adopt the nuclear reaction rates of the NACRE compilation (Angulo et al. 1999), see 🔍 Sect. 6.2, although depending on the size of the nuclear network employed to follow the abundances other sources for the relevant cross sections are implemented. Thus, the results obtained using these stellar evolutionary codes differ sizably from one another, although the general characteristics of the different stellar evolutionary phases agree qualitatively. Anyone trying to pursue the hard task of building a stellar evolutionary code from scratch or to improve a previous version of one of his/her own should be aware that most of the necessary pieces of computer code needed to produce realistic physical inputs can be found on the Web page of the IAU Commission 35.² Also, the exact way in which the equations of stellar structure are solved is very different in all the numerical codes. Although all of the previously listed stellar codes share a common feature, namely, that they use Henyey-type techniques, the way in which the equations of stellar structure are discretized varies from one to another code. For instance, most commonly the equations of stellar structure are discretized adopting the interior mass as the independent variable, but some codes adopt the radius as the independent variable. Additionally, the pressure, temperature, luminosity, radius, and interior mass can be all defined at the zone edges, but in some codes the pressure, temperature, and, thus, density are defined midway between l , r , and m . To these differences one must also add the way in which the changes in composition are dealt with, not only in the convective regions, but also in the radiative ones. The most extensively used stellar evolutionary codes employ nuclear networks similar to that introduced in 🔍 Sect. 6.2.4, but the variety of implementations is also very large, as it is also the number of isotopes followed during the course of stellar evolution. Finally, the way in which zoning and the choice of adequate time steps is dealt with is also extremely dependent on the particular taste of the programmer. To put just one example of these procedures, most stellar evolutionary codes introduce new mesh points to allow a smooth variation of the relevant physical quantities, say the temperature, density, pressure, luminosity, etc., inside the star. This is done by allowing a maximum fractional change between adjacent mass shells, η , which is typically of the order of a few percent. Most stellar evolutionary codes use this strategy to avoid numerical instabilities, except the STAR evolutionary code, which incorporates additional equations to determine the adequate zoning (Eggleton 1971). The way in which the boundary conditions are introduced also adds differences to the results obtained using each of these evolutionary codes, and the way in which the initial guess for the next time step is computed is very different as well, since in some codes the initial solution is extrapolated (linear, semilogarithmic, etc.) from the previous one, while in some others this initial guess is simply the last computed model (absolutely no extrapolation). Additional fine details include different prescriptions to deal with the boundaries of convective regions, as in some codes the nature of a given mass shell cannot change once it is determined from the initial model (it is always convective or radiative), while in some other codes matter can change from convective to radiative and vice versa in all trials. Finally, it should be mentioned that regions in which an adverse and abrupt temperature gradient forms, implying large

²<http://iau-c35.stsci.edu/Resources/index.html>

negative luminosities, are difficult to deal with, and that sometimes it is highly advisable to freeze the mesh during certain periods of time to avoid propagation of numerical errors. The way in which this is done by the different stellar evolutionary codes is highly dependent on the specific way in which the code has been designed and written.

11 Summary

In this chapter, our current understanding of stellar structure has been reviewed. A comprehensive discussion of the equations of hydrostatic equilibrium was given in [Sect. 2](#), while in [Sects. 3](#) and [4](#) the energy conservation law and the treatment of energy transport – either by radiation, conduction, or convection – were extensively discussed. All these otherwise classical results were treated in detail, and the most commonly used prescriptions were provided. Also, when necessary, the needed references to construct realistic stellar configurations were given as well. With these tools, the virial theorem was introduced in [Sect. 5](#), which helped to understand the physical behavior of stars.

In the second part of this chapter, the reader was provided with a series of essential physical inputs needed to build a stellar model. This includes, of course, the equation of state, the nuclear reaction rates, the opacities and conductivities, and the neutrino emission rates. This was done in [Sect. 6](#). Additional physical processes, of special importance for some evolutionary stages, were discussed in [Sect. 7](#). When possible, the aim was to provide to the interested reader with the most commonly used and/or most reliable or recent physical inputs. As mentioned in [Sect. 1](#), the intention was to be as complete as possible, without unintentionally missing any key reference.

The subject of stellar atmospheres was superficially treated in [Sect. 8](#), and, certainly, additional details on this topic are needed, as it was only treated here in connection with the boundary conditions needed to stellar evolution. The reader can find excellent descriptions elsewhere. An approximation to the numerical techniques was given in [Sect. 9](#), where the most usually adopted technique was described in full detail. To conclude, in [Sect. 10](#), the most widely used stellar evolutionary codes were critically analyzed.

To conclude we feel necessary to explicitly mention at this point of the discussion that the field of stellar evolution is continuously improving, and most probably several advances during the coming years will be witnessed. More reliable physical inputs will most likely be incorporated to improved versions of the available state-of-the-art stellar evolutionary codes. Better descriptions of convective mixing will become available, an example being the very recent – and impressive – work of Canuto (2011), and companion papers. Certainly, the next decade will see an increase of the predictive power of the theory of stellar structure. Clearly, this will enhance the status of this theory as one of the most important tools for understanding the origin and evolution of our Universe.

Acknowledgments

This work was partially supported by AGAUR, by MCINN grant AYA2008–04211–C02–01, by the European Union FEDER funds, by the ESF EUROGENESIS project (grant EUI2009–04167), by AGENCIA: Programa de Modernización Tecnológica BID 1728/OC-AR, and by PIP 2008–00940 from CONICET.

References

- Adelberger, E. G., et al. 2011, *Rev Mod Phys*, 83, 195
- Althaus, L. G., Serenelli, A. M., Panei, J. A., Córscico, A. H., García-Berro, E., & Scóccola, C. G. 2005, *A&A*, 435, 631
- Angulo, C., et al. 1999, *Nucl Phys A*, 656, 3
- Arnett, W. D., & Truran, J. W. 1969, *ApJ*, 157, 339
- Böhm-Vitense, E. 1958, *Zeit Astrophysik*, 46, 108
- Blöcker, T. 1995, *A&A*, 297, 727
- Bravo, E., & García-Senz, D. 1999, *MNRAS*, 307, 984
- Burgers, J. M. 1969, *Flow Equations for Composite Gases* (New York: Academic)
- Canuto, V. M. 2011, *A&A*, 528, A76
- Canuto, V. M., & Mazzitelli, I. 1991, *ApJ*, 370, 295
- Cassisi, S., Potekhin, A. Y., Pietrinferni, A., Catelan, M., & Salaris, M. 2007, *ApJ*, 661, 1094
- Caughlan, G. R., & Fowler, W. A. 1988, *Atom & Nucl Data Tables*, 40, 283
- Chaboyer, B., Fenton, W. H., Nelan, J. E., Patnaude, D. J., & Simon, F. E. 2001, *ApJ*, 562, 521
- Chandrasekhar, S. 1939, *An Introduction to the Study of Stellar Structure* (Chicago, IL: University of Chicago Press)
- Chayer, P., Vennes, S., Pradhan, A. K., Thejll, P., Beauchamp, A., Fontaine, G., & Wesemael, F. 1995, *ApJ*, 454, 429
- Christensen-Dalsgaard, J. 1982, *MNRAS*, 199, 735
- Cox, J. P., & Giuli, R. T. 1968, *Principles of Stellar Structure, Vol. I* (New York: Gordon and Breach)
- Cunto, W., & Mendoza, C. 1992, *Revista Mexicana de Astronomía*, 23, 107
- Demarque, P., Guenther, D. B., Li, L. H., Mazumdar, A., & Straka, C. W. 2008, *Ap&SS*, 316, 31
- Eddington, A. S. 1916, *MNRAS*, 77, 16
- Eggleton, P. P. 1971, *MNRAS*, 151, 351
- Ferguson, J. W. et al., 2005, *ApJ*, 623, 585
- Freytag, B., Ludwig, H.-G., & Steffen, M. 1996, *A&A*, 313, 497
- García-Berro, E., & Iben, I., Jr. 1994, *ApJ*, 434, 306
- García-Berro, E., Ritossa, C., & Iben, I., Jr. 1997, *ApJ*, 485, 765
- García-Berro, E., Isern, J., & Kubyshin, Y. A. 2007, *A&AR*, 14, 113
- Grossman, S. A., & Taam, R. E. 1996, *MNRAS*, 283, 1165
- Grossman, S. A., Narayan, R., & Arnett, D. 1993, *ApJ*, 407, 284
- Haft, M., Raffelt, G., & Weiss, A. 1994, *ApJ*, 425, 222
- Heger, A., Langer, N., & Woosley, S. E. 2000, *ApJ*, 528, 368
- Heney, L. G., Wilets, L., Böhm, K. H., Lelevier, R., & Levee, R. D. 1959, *ApJ*, 129, 628
- Heney, L. G., Forbes, J. E., & Gould, N. L. 1964, *ApJ*, 139, 306
- Herwig, F. 2000, *A&A*, 360, 952
- Iben, I., Jr. 1965, *ApJ*, 141, 993
- Iliadis, C., Longland, R., Champagne, A. E., Coc, A., & Fitzgerald, R. 2010, *Nucl Phys A*, 841, 31
- Isern, J., Mochkovitch, R., García-Berro, E., & Hernanz, M. 1997, *ApJ*, 485, 308
- Isern, J., García-Berro, E., Hernanz, M., & Chabrier, G. 2000, *ApJ*, 528, 397
- Itoh, N., Mitake, S., Iyetomi, H., & Ichimaru, S. 1983, *ApJ*, 273, 774
- Itoh, N., Kuwashima, F., & Munakata, H. 1990, *ApJ*, 362, 620
- Itoh, N., Nishikawa, A., & Kohyama, Y. 1996, *ApJ*, 470, 1015
- Kippenhahn, R., & Weigert, A. 1990, *Stellar Structure and Evolution* (Heidelberg: Springer)
- Kippenhahn, R., Thomas, H. C., & Weigert, A. 1965, *Z Astrophys*, 61, 241
- Lane, J. H. 1869, *Am J Sci*, 50, 57
- Langer, N. 1998, *A&A*, 329, 551
- Larson, R. B., & Demarque, P. R. 1964, *ApJ*, 140, 524
- Lederer, M. T., & Aringer, B. 2009, *A&A*, 494, 403
- Longland, R., Iliadis, C., Champagne, A. E., Newton, J. R., Ugalde, C., Coc, A., & Fitzgerald, R. 2010, *Nucl Phys A*, 841, 1
- Maeder, A. 2009, *Physics, Formation and Evolution of Rotating Stars* (Berlin, Heidelberg: Springer, A&A library)
- Meynet, G., & Maeder, A. 2000, *A&A*, 361, 10
- Mihalas, D., & Mihalas, B. W. 1984, *Foundations of Radiation Hydrodynamics* (New York: Oxford University Press)
- Mihalas, D., Hummer, D. G., & Däppen, W. 1988, *ApJ*, 331, 815
- Morel, P., & Lebreton, Y. 2008, *Ap&SS*, 316, 61
- Nobili, L., & Turolla, R. 1988, *ApJ*, 333, 248
- Paquette, C., Pelletier, C., Fontaine, G., & Michaud, G. 1986a, *ApJS*, 61, 177
- Paquette, C., Pelletier, C., Fontaine, G., & Michaud, G. 1986b, *ApJS*, 61, 197
- Paxton, B., Bildsten, L., Dotter, A., Herwig, F., Lesaffre, P., & Timmes, F. 2011, *ApJS*, 192, 3
- Perryman, M. A. C., et al. 2001, *A&A*, 369, 339
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, 612, 168
- Pols, O., Tout, C. A., Eggleton, P. P., & Han, Z. 1995, *MNRAS*, 274, 964
- Raffelt, G. G. 1999, *Annu Rev Nucl Sci*, 49, 163
- Rauscher, T., & Thielemann, F.-K. 2000, *Atom & Nucl Data Tables*, 75, 1

- Rauscher, T., & Thielemann, F.-K. 2001, *Atom & Nucl Data Tables*, 79, 47
- Richard, O., Vauclair, S., Charbonnel, C., & Dziembowski, W. A. 1996, *A&A*, 312, 1000
- Ritossa, C., García-Berro, E., & Iben, I., Jr. 1996, *ApJ*, 460, 489
- Ritossa, C., García-Berro, E., & Iben, I., Jr. 1999, *ApJ*, 515, 381
- Rogers, F. J., & Iglesias, C. A. 1992, *ApJS*, 79, 507; *ApJ*, 456, 902
- Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, *ApJ*, 456, 902
- Salpeter, E. E. 1954, *Aust J Phys*, 7, 373
- Schatzman, E. 1958, *White Dwarfs* (North-Holland: Amsterdam)
- Schwarzschild, K. 1906, *Göttingen Nachr.* 195, 41
- Seaton, M. J., Yan, Y., Mihalas, D., & Pradhan, A. K. 1994, *MNRAS*, 266, 805
- Segretain, L., Chabrier, G., Hernanz, M., García-Berro, E., Isern, J., & Mochkovitch, R. 1994, *ApJ*, 434, 641
- Timmes, F. X., & Swesty, F. D. 2000, *ApJS*, 126, 501
- Traxler, A., Garaud, P., & Stellmach, S. 2011, *ApJL*, 728, L29
- Weaver, T. A., Zimmerman, G. B., & Woosley, S. E. 1978, *ApJ*, 225, 1021
- Weiss, A., & Schlattl, H. 2008, *Ap&SS*, 316, 99
- Young, P. A., & Arnett, D. 2005, *ApJ*, 618, 908
- Young, P. A., Knierman, K. A., Rigby, J. R., & Arnett, D. 2003, *ApJ*, 595, 1114

2 Stellar Atmospheres

Ivan Hubeny

Department of Astronomy and Steward Observatory,
The University of Arizona, Tucson, AZ, USA

1	<i>Introduction</i>	52
2	<i>Basic Physics of Stellar Atmospheres</i>	53
2.1	Overview of Basic Equilibrium Conditions: Atmospheric Layers	55
2.2	Interaction of Radiation with Matter	56
2.3	LTE Versus Non-LTE	59
3	<i>Model Atmospheres</i>	61
3.1	Hierarchy of Approximations	61
3.1.1	Approximations of the Geometry	61
3.1.2	Approximations of the Presence of External Forces	62
3.1.3	Approximations of the Dynamical State of the Atmosphere	62
3.1.4	Approximations of the Opacity Sources	63
3.1.5	Approximations Concerning the Thermodynamic Equilibria	65
3.2	Basic Equations of Classical Stellar Atmospheres	65
3.2.1	Radiative Transfer Equation	65
3.2.2	Hydrostatic Equilibrium Equation	66
3.2.3	Radiative Equilibrium Equation	67
3.2.4	Statistical Equilibrium Equations	67
3.2.5	Charge Conservation Equation	69
3.3	Numerical Methods	69
3.3.1	Complete Linearization	69
3.3.2	Accelerated Lambda Iteration	71
3.3.3	Hybrid CL/ALI Method	72
3.4	Available Modeling Codes	73
3.5	Existing Model Atmosphere Grids	73
3.5.1	LTE Models	73
3.5.2	NLTE Models	74
4	<i>Using Model Stellar Atmospheres to Determine the Fundamental Stellar Parameters</i>	79
5	<i>Summary and Outlook</i>	83
	<i>References</i>	83

Abstract: Basic concepts of the stellar atmospheres theory are briefly outlined. After discussing essential assumptions, approximations, and basic structural equations describing a stellar atmospheres, emphasis is given to describing efficient numerical methods developed to deal with the stellar atmosphere problem, namely the method of complete linearization and its recent variants, and the whole class of methods known by name Accelerated Lambda Iteration. The existing computer codes, and some of the most useful grids of model atmospheres that are publicly available, are briefly summarized. Some interesting properties of newly computed NLTE models atmospheres and their comparison to LTE models are shown. Finally, it is briefly shown how the model atmospheres are used to determine basic stellar parameters.

1 Introduction

The term stellar atmosphere refers to any medium connected physically to a star from which the photons escape to the surrounding space. In other words, it is a region where the radiation, observable by a distant observer, originates. Since in the vast majority of cases the radiation is the only information about a distant astronomical object that can be obtained (exceptions being a direct detection of solar wind particles, neutrinos from the Sun and SN1987a, or gravitational waves), all the information that is gathered about stars is derived from analysis of their radiation.

It is therefore of considerable importance to develop reliable methods which are able to decode the information about a star contained in its spectrum with confidence. Having understood the physics of the problem and being able to carry out detailed numerical simulations makes it possible to construct theoretical models of a stellar atmosphere and predict a stellar spectrum. Model stellar atmospheres are one of the building blocks of our understanding of the radiating objects in the Universe. Modeling stellar atmospheres is one of the most mature fields that deals with analysis and spectroscopic diagnostics of astronomical bodies. Being a mature field, one might not expect any significant new developments. But such a view is completely wrong. The last decade brought a renewed interest in modeling stellar atmospheres. There are several reasons for that. The first, more or less obvious, reason is a significantly increased quality and quantity of observational data that bring new challenges for modelers. For instance, thanks to *Hipparcos*, the accurate distances are now known for many stars, which means that the normalization factor to convert the predicted fluxes at the stellar surface to observed fluxes at the Earth is no longer a free parameter.

The last decade also brought an introduction, after almost a century, of completely new stellar spectral types – L and T dwarfs (sometimes referred to as brown dwarfs; Burgasser et al. 1999; Delfosse et al. 1997). At present, there is a vigorous debate about the name for a spectroscopic class of objects cooler than the coolest T-dwarfs; likely candidate being “Y-dwarfs”, essentially because there are not many more remaining letters of the alphabet available.

The second reason, or motivation, for a progress in modeling stellar atmospheres is an ever-increasing computer power, both the memory available, as well as the computer speed. Consequently, much more sophisticated model atmospheres of unprecedented degree of realism may now be constructed in a reasonable amount of computer time.

However, even with most powerful computers and largest memory chips available, one would still not be able to compute sophisticated models without employing clever and powerful numerical methods. Therefore, the third basic reason for a recent progress in modeling stellar atmospheres is a development of new, very efficient numerical methods for solving the highly nonlinear and nonlocal problem, which the solution of the structural equations for stellar atmosphere requires.

Besides being a crucial diagnostic tool for analyzing stellar spectra, the stellar atmospheres theory provides a guide to the methodology of astrophysical quantitative spectroscopy. Stellar atmospheres are the best-studied example of a medium where radiation is not only a *probe* of the physical state, but is in fact an *important constituent*. In other words, radiation in fact *determines* the structure of the medium, yet the medium is probed *only* by this radiation. Unlike laboratory physics, where one can change a setup of the experiment in order to examine various aspects of the studied structures separately, one does not have this luxury in astrophysics: all what is available is an observed spectrum, so the information encoded in it should be extracted as completely as possible. This is exactly what the stellar atmosphere theory is doing for almost a century now. Consequently, it is developed to such an extent that it provides an excellent methodological guide for other situations where the radiation has the dual role of a probe and a constituent. Examples of such astronomical objects are planetary atmospheres, interstellar medium, H II regions, and accretion disks, to name just few.

The main goal of this chapter is to provide a brief introduction to the basic concepts needed to understand the fundamental physics of stellar atmospheres, as well as the leading principles behind recent developments. Particular emphasis will be devoted to the classical plane-parallel atmospheres in hydrostatic and radiative equilibrium. Much more material on stellar atmospheres is presented in several monographs and conference proceedings. The fundamental textbook of the field, Mihalas (1978), is still a highly recommended text, although it does not cover important recent developments, like, for instance, modern numerical methods. The third edition of the book is now in preparation, but it will take some time before it is available. There is a textbook by Rutten (1995), distributed electronically, which covers both the basic concepts as well as some of the modern developments and is recommended to the beginner in the field. There are two books edited by Kalkofen which present a collection of reviews on various mathematical and numerical aspects of radiative transfer (Kalkofen 1984, 1987). A good textbook that covers both the theoretical and observational aspects of the stellar atmospheres is that by Gray (1992). Other related textbooks include Rybicki and Lightman (1979), Shu (1991), and an elementary-level textbook by Böhm-Vitense (1989). An old but excellent textbook on radiative transfer is Jefferies (1968), an even older but classical one by Chandrasekhar (1960), and a newer book by Cannon (1985). There are two excellent books on the topic of radiation hydrodynamics, which contain several chapters related to the present topic, namely Mihalas and Mihalas (1984), and Castor (2004). Besides these monographs, two recent Proceedings (Hubeny et al. 2003, 2009) contain a number of excellent review papers covering various topics in theory and modeling of stellar atmospheres.

2 Basic Physics of Stellar Atmospheres

From the physical point of view, a stellar atmosphere is generally a plasma composed of many kinds of particles, namely atoms, ions, free electrons, molecules, or even dust grains. In an early-type stellar atmospheres, because of the high temperature and strong radiation field, there are typically no molecules nor dust grains present, at least in the layers that are traditionally considered as an atmosphere. Molecules are present in the atmospheres of solar-type stars and cooler, and dust grains (cloud particles) are present in very cool stars and subsolar-mass objects (brown dwarfs and giant planets).

Typical values of temperature range from 10^3 K (or even much less in atmospheres of planets and brown dwarfs; say down to 50 K) to a few times 10^5 K in the hottest stars. Temperature is even higher, 10^6 – 10^7 K, in stellar coronae. The total particle density ranges from, say, 10^6 to 10^{16} cm $^{-3}$. Under such conditions, the natural starting point for the physical description is the kinetic theory. The physical state of the system is then fully described by a set of kinetic (Boltzmann) equation for the distribution functions, f_i , of all kinds of particles that exist in the atmosphere (generally, not only for the individual atoms, ions, free electrons, and possibly molecules, but also for individual excitation states of the atoms/ions),

$$\frac{\partial f_i}{\partial t} + (\mathbf{u} \cdot \nabla) f_i + (\mathbf{F} \cdot \nabla_p) f_i = \left(\frac{Df_i}{Dt} \right)_{\text{coll}} , \quad (2.1)$$

where ∇ and ∇_p are the usual nabla differential operators with respect to position and momentum components, respectively; \mathbf{u} is the particle velocity, and \mathbf{F} is the external force. The term $(Df/Dt)_{\text{coll}}$ is the so-called collisional term, which describes creations and destructions of particles of type i with the position $(\mathbf{r}, \mathbf{r} + d\mathbf{r})$ and momentum $(\mathbf{p}, \mathbf{p} + d\mathbf{p})$.

However, the full distribution functions are usually not needed, and in any case the number of unknowns to be determined would be enormous. The standard procedure to reduce the problem is to form moments of the kinetic equation that lead to the usual hydrodynamic equations. The procedure is described in any textbook of kinetic theory or hydrodynamics; in the astrophysical context, the excellent textbooks are Mihalas and Mihalas (1984) and Castor (2004).

The standard set of hydrodynamic equations are the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 , \quad (2.2)$$

the momentum equation,

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v}) = -\nabla P + \mathbf{f} , \quad (2.3)$$

and the energy balance equation,

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \rho v^2 + \rho \epsilon \right) + \nabla \cdot \left[\left(\frac{1}{2} \rho v^2 + \rho \epsilon + P \right) \mathbf{v} \right] = \mathbf{f} \cdot \mathbf{v} - \nabla \cdot (\mathbf{F}_{\text{rad}} + \mathbf{F}_{\text{con}}) . \quad (2.4)$$

Here, \mathbf{v} is the macroscopic velocity, ρ the total mass density, P the pressure, \mathbf{f} the external force, ϵ the internal energy, \mathbf{F}_{rad} the radiation flux, and \mathbf{F}_{con} the conductive flux. [Equations 2.2–2.4](#) represent moment equations of the kinetic equation ([2.1](#)), summed over all kinds of particles.

In addition to classical hydrodynamics, one also considers a zeroth-order moment equation for the individual kinds of particles in the individual degrees of freedom (excitation states), i.e., the conservation equation for particles of type i ,

$$\frac{\partial n_i}{\partial t} + \nabla \cdot (n_i \mathbf{v}) = \left(\frac{Dn_i}{Dt} \right)_{\text{coll}} , \quad (2.5)$$

which are used for the so-called non-LTE (or NLTE) description – see below. Here, n_i is the number density (population) of the state i .

These equations provide the most general description of the dynamics of a stellar atmosphere. They are still extremely complex to solve in full generality; therefore, one has to resort to various simplifying approximations, which are briefly discussed next.

2.1 Overview of Basic Equilibrium Conditions: Atmospheric Layers

Traditionally, a stellar atmosphere is divided into several basic layers, depending on what kind of equilibrium conditions can be reasonably assumed to hold there. In the case of early-type stars (early A-type and hotter), the two basic layers are the photosphere and the stellar wind. In the case of solar-type stars and cooler, the layers are somewhat different. Typically, there is again a photosphere, then chromosphere and corona, and finally a (typically weak) wind.

The *photosphere* is the innermost part of the atmosphere. The mass outflow velocities are very small there, smaller than the local sound speed, so that they can be neglected. The photosphere is assumed to be an essentially *static* region. The radial extent of the photosphere is typically very small compared to the stellar radius, even for hot massive stars. This region is very important because most of the observed spectral features originate there. Therefore, it is fair to say that the photosphere is crucial for determining basic stellar parameters, such as effective temperature, gravity acceleration, and chemical composition. Stellar photospheres are characterized by the condition of hydrostatic and radiative equilibrium.

The **hydrostatic equilibrium** stipulates that the gradient of the total pressure is balanced by the local gravity acceleration. Because the radial extent of the photosphere is small compared to the stellar radius, the gravity acceleration is essentially constant. The explicit form of this equation is (for a spherically-symmetric star):

$$\frac{dP}{dR} = -\rho GM_*/R^2 \approx -\rho GM_*/R_*^2 \equiv -\rho g, \quad (2.6)$$

where P is the total pressure, R is the radial coordinate, M_* and R_* are the stellar mass and radius, respectively, g is the gravity acceleration at the stellar surface, and ρ the mass density. Notice that (2.6) follows from the general momentum equation (2.3) by setting $\partial/\partial t = 0$ (a stationary situation), $\mathbf{v} = 0$ (static case), and $f = -\rho g$ (external force being given by the gravity force; the minus sign reflects the fact that the gravity acts inward, while the radial coordinate increases outward).

The **radiative equilibrium** simply states that the total radiation flux is conserved. As follows from the energy equation (2.4), its reduction to the static case and neglecting conduction leads to

$$\nabla F_{\text{rad}} = 0 \implies F_{\text{rad}} = \text{const} \equiv \sigma T_{\text{eff}}^4, \quad (2.7)$$

where σ is the Stefan–Boltzmann constant and T_{eff} is the so-called effective temperature.

As follows from (2.6) and (2.7), the gravity acceleration g and the effective temperature T_{eff} are two basic parameters of the stellar photosphere problem.

The atmospheres of cool stars (late A-type and cooler) contain regions that are unstable against convection; the unstable regions are those that satisfy the *Schwarzschild criterion*, (e.g., Mihalas 1978),

$$\nabla_R \equiv \left(\frac{d \ln T}{d \ln P} \right)_R > \left(\frac{d \ln T}{d \ln P} \right)_A \equiv \nabla_A, \quad (2.8)$$

where ∇_R is the logarithmic gradient of temperature with respect to the total pressure assuming radiative equilibrium, while ∇_A is the adiabatic gradient. The latter can be easily evaluated as a function of temperature and density of the stellar material.

Convection is a transport of energy by rising and falling bubbles of material with properties (e.g., temperature) different from the ambient medium. It is therefore, by its very nature, a non-stationary and nonhomogeneous phenomenon. Putting $\mathbf{v} = 0$ and assuming a 1D medium

means, strictly speaking, that convection is a priori neglected in the hydrodynamic equations. However, most modeling approaches use simplified descriptions, such as the mixing-length theory (see any standard textbook, e.g., Mihalas 1978), which recasts the problem in the form of a 1D stationary equation, viz.

$$F_{\text{rad}} + F_{\text{conv}} = \sigma T_{\text{eff}}^4, \quad (2.9)$$

where the convective flux F_{conv} is a specified function of basic state parameters (temperature, density, etc.)

The *stellar wind* is the region where the outflow velocities are comparable or larger than the local sound speed. The radial extent of this region may be comparable to or, in some cases, significantly larger than, the radius of the stellar photosphere.

Stellar winds, as well as stellar chromospheres and coronae, are covered in another chapters. This chapter concentrates mainly on the stellar photospheres.

2.2 Interaction of Radiation with Matter

From the very nature of stellar atmospheres, it is clear that a detailed description of the processes of interaction between radiation and matter is a crucial ingredient of the stellar atmospheres theory. These processes determine (1) how the radiation is transported in the atmosphere and (2) what is the distribution of the microscopic degrees of freedom of the massive particles (e.g., the excitation and ionization states of the individual atomic species, etc.).

The interaction between radiation and matter is described through the radiative transfer equation, which is generally written as

$$\left(\frac{1}{c} \frac{\partial}{\partial t} + \mathbf{n} \cdot \nabla \right) I(\nu, \mathbf{r}, \mathbf{n}, t) = \eta(\nu, \mathbf{r}, \mathbf{n}, t) - \chi(\nu, \mathbf{r}, \mathbf{n}, t) I(\nu, \mathbf{r}, \mathbf{n}, t). \quad (2.10)$$

Here, I is the *specific intensity* of radiation, defined such that it is the energy transported by radiation in a unit frequency range at the frequency ν , across a unit area perpendicular to the direction of propagation, \mathbf{n} , into a unit solid angle, and in a unit time interval. The specific intensity provides a complete description of the unpolarized radiation field from the macroscopic point of view. This description can be generalized to an arbitrarily polarized light by introducing the *Stokes vector* instead of the scalar intensity (e.g., Chandrasekhar 1960). This case will not be considered here.

Quantities χ and η are the *absorption* and *emission* coefficients, respectively. They are defined analogously to the specific intensity, namely as the energy removed or added to a beam of radiation at unit frequency range, solid angle, area, and time.

It is known from the quantum theory of radiation that there are three types of elementary processes that give rise to an absorption or emission of a photon: (1) induced absorption – an absorption of a photon accompanied by a transition of an atom/ion to a higher energy state; (2) spontaneous emission – an emission of a photon accompanied by a spontaneous transition of an atom/ion to a lower energy state; and (3) stimulated emission – an interaction of an atom/ion with a photon accompanied by an emission of another photon with identical properties. In the astrophysical formalism, the stimulated emission is usually treated as negative absorption.

In thermodynamic equilibrium, the microscopic detailed balance holds, and therefore the radiation energy absorbed in an elementary volume in an elementary frequency interval is

exactly balanced by the energy emitted in the same volume and in the same frequency range, i.e., $\chi I = \eta$. Moreover, in thermodynamic equilibrium, the radiation intensity is equal to the Planck function, $I = B$, where

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT) - 1}. \quad (2.11)$$

In thermodynamic equilibrium, one has, therefore, $\eta/\chi = B$, which is called the *Kirchhoff's law*.

The absorption coefficient (or opacity) is given by

$$\chi_\nu = \kappa_\nu + \sigma_\nu, \quad (2.12)$$

where κ_ν is the coefficient of true absorption (or extinction coefficient) and σ_ν is the scattering coefficient. The extinction coefficient is given by

$$\begin{aligned} \kappa_\nu = & \sum_i \sum_{j>i} [n_i - (g_i/g_j)n_j] \sigma_{ij}(\nu) + \sum_i \left(n_i - n_i^* e^{-h\nu/kT} \right) \sigma_{ik}(\nu) \\ & + \sum_\kappa n_e n_\kappa \sigma_{\kappa\kappa}(\nu, T) \left(1 - e^{-h\nu/kT} \right), \end{aligned} \quad (2.13)$$

where the three terms represent, respectively, the contributions of bound-bound transitions (i.e., spectral lines), bound-free transitions (continua), and free-free absorption (also called *brehmstrahlung*).

Here, n_i is the occupation number (population) of an atom in the energy level labeled i , g_i the corresponding statistical weight, and n_i^* denotes an equilibrium population of level i corresponding to temperature T , and electron density n_e . $\sigma(\nu)$ are the corresponding cross sections; subscript κ denotes the “continuum,” and n_κ the ion number density. The negative contributions represent the stimulated emission. The relation between the bound-bound cross section $\sigma_{ij}(\nu)$ and the well-known Einstein coefficients for the photoexcitation is $\sigma_{ij}(\nu) = (h\nu_0/4\pi)B_{ij}\phi(\nu)$, where ν_0 is the frequency of the line center, and $\phi(\nu)$ the *absorption profile coefficient*, normalized to unity, $\int \phi(\nu) d\nu = 1$. It represents the conditional probability density that if a photon is absorbed in the transition $i \rightarrow j$, it is absorbed in the frequency range $(\nu, \nu + d\nu)$.

In hot stars, a photon scattering on free electrons – *electron scattering*, is the dominant scattering process. Moreover, it is to a good approximation coherent, i.e., without a change of frequency – Thomson scattering. In this case,

$$\sigma_\nu = n_e \sigma_e, \quad (2.14)$$

where σ_e is the Thomson cross section. A more general, incoherent case is called Compton scattering. The effects of Compton scattering are negligible in the atmospheres of most stars, with a possible exception of extremely hot subdwarfs, white dwarfs, and pre-white dwarfs. For cooler stars, the Rayleigh scattering becomes important, and for very cool objects (M dwarfs and substellar mass-objects) where clouds are formed, the Mie scattering on clouds particles is important. There is no stimulated emission correction for the coherent scattering, because this contribution exactly cancels with ordinary absorption.

The absorption coefficient defined by (2.13) has the dimension cm^{-1} , and its inverse has a meaning of the photon mean free path. This coefficient is also called *opacity per length*, in contrast to *opacity per mass* which is defined as χ_ν/ρ . The advantage of the latter is that it is typically much less dependent on mass density, because the atomic level populations and the electron density roughly scale with mass density, so that these dependencies almost cancel.

Analogously, the emission coefficient is given by

$$\eta_\nu = \eta_\nu^{\text{th}} + \eta_\nu^{\text{scat}}, \quad (2.15)$$

where the thermal emission coefficient is given by

$$\begin{aligned} \eta_\nu^{\text{th}} = (2h\nu^3/c^2) & \left[\sum_i \sum_{j>i} n_j (g_i/g_j) \sigma_{ij}(\nu) + \sum_i n_i^* \sigma_{i\kappa}(\nu) e^{-h\nu/kT} \right. \\ & \left. + \sum_\kappa n_e n_\kappa \sigma_{\kappa\kappa}(\nu, T) e^{-h\nu/kT} \right]. \end{aligned} \quad (2.16)$$

The three terms again describe the bound-bound, bound-free, and free-free emission processes, respectively.

The scattering part of the emission coefficient is more complex. Generally, it is given by

$$\eta_\nu^{\text{scat}}(\mathbf{n}) = \int d\nu' \oint d\Omega' / (4\pi) R(\nu', \mathbf{n}', \nu, \mathbf{n}) I(\nu', \mathbf{n}'), \quad (2.17)$$

where $R(\nu', \mathbf{n}', \nu, \mathbf{n})$ is the *redistribution function* that represents a probability density that a photon in the frequency range $(\nu', \nu' + d\nu')$ and in an element of solid angle $d\Omega'$ around direction \mathbf{n}' is absorbed, and a photon in the frequency range $(\nu, \nu + d\nu)$ and in an element of solid angle $d\Omega$ around direction \mathbf{n} is emitted in the scattering process. The redistribution function is usually assumed to be separable into the frequency and the angular part,

$$R(\nu', \mathbf{n}', \nu, \mathbf{n}) = R(\nu', \nu) g(\mathbf{n}', \mathbf{n}), \quad (2.18)$$

where $R(\nu', \nu)$ is called frequency redistribution function, and g the phase function. In the case of electron (Thomson) scattering, $R(\nu', \nu) = n_e \sigma_e \delta(\nu' - \nu)$, and assuming isotropic scattering, $g(\mathbf{n}', \mathbf{n}) = 1$, the scattering emission coefficient is simply given by

$$\eta_\nu^{\text{scat}} = n_e \sigma_e J_\nu, \quad (2.19)$$

where J is the mean intensity of radiation – see (► 2.29).

The absorption and emission coefficients are thus described through the corresponding cross sections, given by the atomic physics, the local thermodynamic parameters, T and n_e , and the atomic level populations for all the levels involved in the microscopic processes that give rise to an absorption and emission at frequency ν ; such a number may be enormous. The scattering emission coefficient moreover depends explicitly the radiation field. Besides that, the main difficulty of the stellar atmospheres theory is that the level populations generally depend (in a NLTE situation) on other state parameters and the radiation field.

It should also be realized that the number of cross sections and other parameters to be supplied by atomic physics is enormous. In fact, once efficient numerical methods for solving the coupled problem of radiation transport and hydrodynamics are developed, the main limiting factor in producing realistic model stellar atmospheres is the availability of necessary atomic data. The most popular large-scale source of data for bound-bound transitions, lines, are the Kurucz line lists (Kurucz 1993; the latest additions described in Kurucz 2009). In the last two decades, two major collaborative projects – Opacity Project (OP – OP Team 1995, 1997; Seaton 1987) and its continuation the Iron Project (Hummer et al. 1993; Pradhan et al. 1996), and the OPAL Project (Iglesias and Rogers 1991, 1996) – have produced accurate atomic data on a large scale. There have also been a significant progress on the molecular data; for a review, see Jørgensen (2003) and Sharp and Burrows (2007).

2.3 LTE Versus Non-LTE

It is well known from statistical physics that a description of material properties is greatly simplified if the thermodynamic equilibrium (TE) holds. In this state, the particle velocity distributions as well as the distributions of atoms over excitation and ionization states are specified uniquely by two thermodynamic variables. In the stellar atmospheres context, these variables are usually chosen to be the (kinetic) temperature (T), and the total particle number density (N), or the electron number density (n_e). From the very nature of a stellar atmosphere, it is clear that it cannot be in thermodynamic equilibrium – a star is detectable by a distant observer, so photons must be escaping from it. Because photons carry significant momentum and energy, the elementary fact of photon escape has to give rise to significant *gradients* of the state parameters in the stellar outer layers.

However, even if the assumption of TE cannot be applied to a stellar atmosphere, one may still use the concept of *local thermodynamic equilibrium* – LTE. This assumption stipulates that one may employ the standard thermodynamic relations not globally for the whole atmosphere, but *locally*, for local values of $T(\mathbf{r})$ and $N(\mathbf{r})$ or $n_e(\mathbf{r})$, despite the gradients that exist in the atmosphere. This assumption simplifies the problem enormously, for it implies that all the particle distribution functions may be evaluated locally without reference to the physical ensemble in which the given material is found. Notice that the equilibrium values of distribution functions are assigned to *massive particles*; the radiation field is allowed to depart from its equilibrium, Planckian, distribution function (i.e., $I = B$ is valid only in strict TE).

Specifically, LTE is characterized by the following three distributions:

- Maxwellian velocity distribution of particles

$$f(\mathbf{v})d\mathbf{v} = (m/2\pi kT)^{3/2} \exp(-mv^2/2kT) d\mathbf{v} , \quad (2.20)$$

where m is the particle mass and k the Boltzmann constant.

- Boltzmann excitation equation,

$$(n_j/n_i) = (g_j/g_i) \exp[-(E_j - E_i)/kT] , \quad (2.21)$$

where g_i is the statistical weight of level i , and E_i the level energy, measured from the ground state.

- Saha ionization equation.

$$\frac{N_I}{N_{I+1}} = n_e \frac{U_I}{U_{I+1}} C T^{-3/2} \exp(\chi_I/kT) , \quad (2.22)$$

where N_I is the total number density of ionization stage I , U is the partition function, defined by $U = \sum_i g_i \exp(-E_i/kT)$; χ_I is the ionization potential of ion I , and $C = (h^2/2\pi mk)^{3/2}$ is a constant. It should be stressed that in the astrophysical LTE description, the *same* temperature T applies to all kinds of particles, and to all kinds of distributions, (2.20–2.22).

Equations 2.20–2.22 define the state of LTE from the macroscopic point of view. Microscopically, LTE holds if all atomic processes are in *detailed balance*, i.e., if the number of processes $A \rightarrow B$ is exactly balanced by the number of inverse processes $B \rightarrow A$. A and B refer to any particle states between which there exists a physically possible transition. For instance, A is an atom in an excited state, and B the same atom in another state (either of the same ion, in which

case the process is an excitation/de-excitation; or of the higher or lower ion, in which case the term is an ionization/recombination).

In contrast, the term non-LTE (or NLTE) characterizes any state that departs from LTE. In practice, this usually means that populations of some selected energy levels of some selected atoms/ions are allowed to depart from their LTE values, while the velocity distributions of all particles are assumed to be Maxwellian, all at the same local kinetic temperature, T .

In this case, the number densities (populations) of all energy levels of all atoms and ions that are important in the atmosphere have to be determined. Equations that describe how the level populations behave are the *statistical equilibrium equations* (or rate equations), which are (2.5) reduced to a static case, $(Dn_i/Dt)_{\text{coll}} = 0$. More explicitly, they express the fact that under stationary situation, the total number of transitions *into* a given level i is equal to the total number of transitions *out* of level i ,

$$n_i \sum_{j \neq i} (R_{ij} + C_{ij}) = \sum_{j \neq i} n_j (R_{ji} + C_{ji}) \quad , \quad (2.23)$$

where R_{ij} and C_{ij} are the radiative and collisional rates, respectively, for the transition from level i to level j . The left-hand side of (2.23) represents the total number of transitions *out* of level i , while the right-hand side represents the total number of transitions *into* level i from all other levels.

One of the important issues of the stellar atmospheres theory for the last three decades was whether, and if so to what extent, departures from LTE should be included in numerical modeling. Generally, to understand why and where one may expect departures from LTE, it is best to turn to the microscopic definition of LTE. It is clear that LTE breaks down if the detailed balance in at least one transition $A \rightarrow B$ breaks down. There are two types of transitions: the *collisional transitions* (arising due to interactions between two or more massive particles) and *radiative transitions* (interactions involving particles and photons). Under stellar atmospheric conditions, collisions between massive particles tend to maintain the local equilibrium (because velocities are Maxwellian). Therefore, the validity of LTE hinges on whether the radiative transitions are in detailed balance or not.

The fact that the radiation escapes from a star implies that LTE should eventually break down at a certain point in the atmosphere. Essentially, this is because detailed balance in radiative transitions generally breaks down at a certain point near the surface. Because photons escape (and more so from the uppermost layers), there must be a lack of them there. Consequently, the number of photoexcitations (or any atomic transition induced by absorbing a photon) is less than a number of inverse processes, spontaneous de-excitations (neglecting here, for simplicity, the stimulated emission).

These considerations explain that departures from LTE are expected if the following two conditions are met: (1) radiative rates in some important atomic transition dominate over the collisional rates; and (2) radiation is not in equilibrium, i.e., the intensity does not have the Planckian distribution. Because the collisional rates are proportional to the particle density, the departures from LTE tend to be small for high densities. Likewise, deep in the atmosphere, photons do not escape, and so the intensity is close to the equilibrium value. Departures from LTE are therefore small, even if the radiative rates dominate over the collisional rates. On the other hand, departures from LTE are important for low-density media immersed in a strong radiation field, which are precisely the conditions met in the atmospheres of hot stars.

3 Model Atmospheres

Model stellar atmospheres are basic tools to analyze observed stellar spectra. By fitting the observed spectrum by a grid of theoretically predicted spectra, one can derive the basic parameters used for constructing the models, i.e., the effective temperature, surface gravity, chemical composition, and, in the general case, the mass loss rate. From those parameters, one can derive the fundamental stellar parameters, like the mass, radius, and luminosity. Besides these, there are a number of secondary parameters, such as the rotational velocity, or auxiliary parameters describing the nature of atmospheric velocity fields, etc.

By the term *model stellar atmosphere*, it is understood a specification of all the atmospheric state parameters as functions of position. These parameters are obtained by solving appropriate structural equations, which, in the case of stellar photospheric models, are the equations of hydrostatic equilibrium, radiative equilibrium, radiative transfer equation, and the set of statistical equations (rate equations) for the atomic level populations. In the case of LTE models, the rate equations are not needed because the level populations are given by the Saha–Boltzmann distribution. Because the problem is very complex, it is impossible to find analytical solutions. Therefore, one has to resort to numerical simulations. In order to make the overall problem tractable, it is necessary to make a number of simplifications by invoking various approximations. The quality of an appropriate model, and consequently its applicability to the individual stellar types, is closely related to the degree of approximation used in the construction of the model. Needless to say, the degree of approximation critically influences the amount of computational effort to compute it. It is fair to say that the very art of computing model stellar atmospheres is to find such physical approximations that allow the model to be computed with a reasonable amount of numerical work, yet the model is sufficiently realistic to allow its use for a reliable interpretation of observed stellar data. The adopted approximations are therefore critical. There are several types of approximations that are typically made in the model construction, which are summarized below.

3.1 Hierarchy of Approximations

3.1.1 Approximations of the Geometry

What is meant by the geometrical simplification is that either some prescribed geometrical configuration is assumed or some special kind of overall symmetry is invoked. The goal of those simplifications is to reduce the dimensionality of the problem from a spatially 3D problem to 1- or 2D problem. The most popular approximations are (from simplest to more complex):

- Plane-parallel geometry, with an assumption of horizontally homogeneous layers. This decreases the number of dimensions to one: the depth in the atmosphere. This approximation is typically quite reasonable for stellar photospheres, which indeed are by several orders of magnitude thinner than the stellar radius, so the curvature effects are negligible. The assumption of horizontal homogeneity is made for the sake of simplicity – there is no plausible verification of this approximation, and, moreover, observational evidence mostly shows that stellar surfaces are far from being homogeneous (a notorious example being detailed pictures of the solar surface). Nevertheless, even in the presence of inhomogeneities,

1D models still have their value since, in many cases, one may construct different 1D models for the individual “patches” on the surface.

- Spherical symmetry. Again, the problem is one-dimensional. The approach is used for extended atmospheres, for which the atmospheric thickness is no longer negligible with respect to the stellar radius. Typically, one has to consider such models for giants and supergiants, as well as for earliest types of main-sequence O stars.
- Multidimensional geometry. This field is at its infancy. A numerical solution is extremely demanding on computer time and memory, and only very recently has the computer power reached a stage that calculating such models is becoming feasible. Model atmospheres with a 2D and 3D geometry have been constructed for solar-type atmospheres; for a recent review, see Carlsson and Stein (2003), Stein and Nordlund (2003), and Nordlund and Stein (2009).

3.1.2 Approximations of the Presence of External Forces

In a typical stellar atmospheres, the only external force that is taken into account is the force of gravity of the star, and its effect (in the case of a photosphere) is described through the hydrostatic equilibrium (● 2.6). Other forces are sometimes taken into account, depending above all on the nature of the studied object.

- Centrifugal force. If the star rotates very fast, close to breakup rotation rate, the whole star becomes distorted. A star is still assumed to be cylindrically symmetric, but the overall atmosphere is no longer described by a single model, but instead it is assumed to be composed of a set of latitudinal belts, each described by a unique model atmosphere. This approach implicitly assumes that the physical properties of the atmosphere vary much faster in the vertical direction (toward the stellar center) than in the horizontal direction. This is usually an excellent approximation. The centrifugal force thus does not enter the model construction directly, but only indirectly through the need of constructing several models for a stellar atmosphere instead of one.
- Magnetic force. Some objects (magnetic Ap stars, magnetic white dwarfs, neutron stars) possess a strong magnetic field, which needs to be taken into account in model construction. This topic will not be discussed in this chapter.

3.1.3 Approximations of the Dynamical State of the Atmosphere

This is basically a specification of the realism of the treatment of the macroscopic velocity fields. From the simplest to the most complex, the approaches are the following:

- Static models, in which the macroscopic velocity field is set to zero. As discussed above, these models describe a *stellar photosphere*.
- Models with an a priori-given velocity field. In these models, the velocities are taken into account explicitly, and their influence upon other state parameters, in particular the emergent radiation, is studied in detail. In these models, one can either consider only a dynamical region (i.e., the wind) and take an incoming radiation from the photosphere as given a priori – the so-called core-halo model; or a model which treats the photosphere and the wind on the same footing. Such models are called *unified models*. The most successful computer

- programs for computing such models are CMFGEN (Hillier 2003; Hillier and Miller 1998), Hoefflich's suite of programs (Hoefflich 1995, 2003), PHOENIX (Hauschildt et al. 1997), Munich codes (FASTWIND – Santolaya-Rey et al. 1997; WM-BASICS – Pauldrach et al. 2001), and the Kiel-Potsdam code (Hamann 1985; Koesterke et al. 2002).
- Models where the velocity field is determined self-consistently by solving the appropriate hydrodynamical equations. This problem is very complicated because the wind driving force is given by the absorption of photons in thousands to millions of metal lines, so the hydrodynamical equations should be solved together with at least an approximate treatment of radiative transfer in spectral lines.

3.1.4 Approximations of the Opacity Sources

In real stellar atmospheres, there is an enormous number of opacity sources. It is essentially impossible (or, at least, impractical) to take all of them into account in full detail. The light elements (H, He, C, N, O) have comparatively a small number of lines per ion (say 10^2 – 10^4) because of a relatively simple atomic level structure. The number of lines generally increases with increasing atomic number, and for the iron-peak elements (Fe and Ni being the most important ones), there are of the order of 10^6 – 10^7 spectral lines per ion! Therefore, the opacity (and emissivity) may be an enormously complicated function of frequency. For cool stars, the problem is even worse because the number of molecular lines is even larger (for instance, water has of the order of 10^9 lines that should be taken into account).

There are several approximations that are meant to reduce this complexity considerably:

- Models constructed using certain frequency-averaged opacities; these models are called *grey* models. The approach is based on the implicit assumption that the behavior of the frequency-averaged intensity of radiation is well described by means of some frequency-averaged opacities. There are several possible mean opacities, depending on how exactly the averaging is done. The most used averaged opacity is the *Rosseland mean* opacity, defined by

$$\frac{1}{\chi_R} \equiv \frac{\int_0^\infty (1/\chi_\nu)(dB_\nu/dT) d\nu}{\int_0^\infty (dB_\nu/dT) d\nu}, \quad (2.24)$$

where χ_ν is the opacity (per gram of stellar material). Because averaging is done for $1/\chi$, the largest weight is given to regions of lowest opacity, which are the most efficient regions for the energy flux transport. This explains why the Rosseland mean opacity is well suited for describing the total radiation flux, and why it is the most appropriate mean opacity to be used for modeling stellar interiors.

The grey model atmospheres are no longer used for spectroscopic analysis, but they are useful for providing an initial estimate in any iterative method for constructing more realistic model atmospheres, and they are very useful for pedagogical purposes because they allow one to understand a rough behavior of temperature and radiation field as a function of depth in the atmosphere.

- A possibility is to use stepwise frequency averages for a number of subintervals (frequency bins), called sometimes multifrequency/multi-grey method. This approach was used in constructing model stellar atmospheres only rarely (Anderson 1985, 1987, 1989) but is more used in other branches of astrophysical and laboratory radiative transfer.

- A completely different approach is to construct a model atmosphere neglecting the line opacity completely. Although this may seem very crude, such models may actually provide reasonable results for very metal-poor stars because, as was pointed out above, H and He possess only a small number of lines, which occupy only a very small frequency range and therefore have a small influence on the model structure. (Strictly speaking, this is not completely true because just three of the most important hydrogen lines – $L\alpha$, $L\beta$, and $H\alpha$ – may already have an important indirect effect upon the temperature structure in the outer layers of early-type photospheres (Auer and Mihalas 1969). In any event, this approximation was introduced at the early stages of development of the non-LTE model atmospheres and was motivated by limitations of then available computers and numerical techniques.
- An obvious next level of approximation is to consider a small number of lines (typically tens to hundreds) explicitly while neglecting the bulk of metal lines; the selected lines are those which presumably have the largest effect upon the atmospheric structure. This approach was used in early days of NLTE model atmospheres (Auer and Mihalas 1969).
- Finally, one can take into account, by one way or another, “all” metal lines. Such models are traditionally called *metal line-blanketed model atmospheres*. The problem of constructing such models is computationally very demanding. Under the assumption of LTE, it is, however, considerably simplified because the opacity and emissivity is a function of only local temperature and electron density; the only problem is the complicated frequency dependence of the opacity. Without the approximation of LTE, the problem is significantly more difficult because one has to determine all the atomic level populations and temperature self-consistently with the radiation field.

There are essentially two possibilities:

- *Opacity Distribution Functions* (ODF). In LTE, the use of ODFs is straightforward (Kurucz 1970). In NLTE, this method is used in conjunction with the concept of superlevels. A *superlevel* is a set of individual energy levels with close energies, and with identical or similar quantum numbers (e.g., the same spin quantum number S , the same parity) that share the common NLTE departure coefficient, that is, they are in Boltzmann equilibrium within each other (Anderson 1985). Transitions between superlevels are called superlines. The idea is to resample a complicated frequency dependence of the superline cross section to form a monotonic function of frequency; this function is then represented by a smaller number of frequency quadrature points (Anderson 1989; Dreizler and Werner 1993; Hubeny and Lanz 1995).
- *Opacity Sampling* (OS). The idea is a simple Monte Carlo-like sampling of frequency points of the superline cross sections (Anderson 1989; Dreizler and Werner 1993). The advantage of this approach is that it can easily treat line blends and overlaps; the disadvantage is that considering too few frequency points may easily lead to missing many important line cores. On the other hand, the “exact” method is in fact a variant of the OS with a sufficiently high frequency resolution.

An explicit comparison between results using the ODF and the OS approaches, and with various frequency resolutions in the latter, is presented, e.g., in Lanz and Hubeny (2003). With increasing computational power, both memory and speed, the Opacity Sampling scheme with high-frequency resolution becomes the method of choice.

3.1.5 Approximations Concerning the Thermodynamic Equilibria

As was discussed above, the issue here is whether the approximation of LTE is adopted or not. If one assumes LTE, the resulting model atmospheres are called LTE models. Two state parameters, the temperature, T , and density, ρ (or electron density, n_e), suffice to describe the physical state of the atmosphere at any given position. In practice, LTE models may be useful only for stellar photospheres because for extended atmospheres and/or stellar winds, this approximation breaks down completely and its application would yield erroneous and misleading results.

The models that take some kind of departure from LTE into account are called non-LTE (or NLTE) models. This term is rather ambiguous because it is not a priori specified what is actually allowed to depart from LTE in a given model. In early models, the populations of only few low-lying energy levels of the most abundant species, like H and He, were allowed to depart from LTE; the rest were treated in LTE. During the development of the field, progressively more and more levels were allowed to depart from LTE. The situation is similar for stellar photospheres (static models), as well as for stellar winds and for unified models.

3.2 Basic Equations of Classical Stellar Atmospheres

The basic equations of stellar atmospheres for the case of horizontally-homogeneous, plane-parallel, static atmospheres, that is, stellar photospheres) are summarized below.

3.2.1 Radiative Transfer Equation

For a 1D, static, planar atmosphere, the general transfer (2.10) can be considerably simplified to read

$$\mu \frac{dI(\nu, \mu, z)}{dz} = \eta(\nu, \mu, z) - I(\nu, \mu, z) \chi(\nu, \mu, z), \quad (2.25)$$

where the intensity of radiation is now only a function of the geometrical coordinate z , frequency ν , and the directional cosine μ . Here, $\mu = \cos \theta$, where θ is the angle between the direction of photon propagation and the normal to the surface. Defining the *optical depth* through

$$d\tau_\nu = -\chi_\nu dz, \quad (2.26)$$

and the *source function* as

$$S_\nu = \eta_\nu / \chi_\nu, \quad (2.27)$$

(the functional dependence of position and direction was omitted for notational simplicity, and a dependence on frequency is expressed through subscript ν which is customary in astrophysical literature), the transfer is rewritten to a usual form

$$\mu \frac{dI_\nu}{d\tau_\nu} = I_\nu - S_\nu. \quad (2.28)$$

It is customary to introduce moments of the specific intensity as

$$\begin{pmatrix} J_\nu \\ \mathbf{H}_\nu \\ K_\nu \end{pmatrix} = \frac{1}{4\pi} \oint \begin{pmatrix} 1 \\ \mathbf{n} \\ \mathbf{nn} \end{pmatrix} I_\nu d\omega \longrightarrow \frac{1}{2} \int_{-1}^1 \begin{pmatrix} 1 \\ \mu \\ \mu^2 \end{pmatrix} I_\nu(\mu) d\mu, \quad (2.29)$$

where the last term corresponds to the 1D plane-parallel approximation. In this case, the only nonvanishing component of vector \mathbf{H} is the H_z -component, which is written simply as H ; analogously the only nonvanishing component of tensor \mathbf{K} is the K_{zz} -component, denoted as K . The moment equations of the radiative transfer equation, in a plane-parallel approximation, are written as

$$\frac{dH_v}{d\tau_v} = J_v - S_v , \quad (2.30)$$

and

$$\frac{dK_v}{d\tau_v} = H_v . \quad (2.31)$$

The system of moment equations is not closed, i.e., the equation for n -th moment contains the $(n + 1)$ -th moment, etc. It is therefore necessary to come up with some kind of closure relation. In the stellar atmospheres theory, one defines the Eddington factor, f^K , by (Auer and Mihalas 1970)

$$f_v^K \equiv K_v/J_v . \quad (2.32)$$

It is clear from the definition of moments that in the case of isotropic radiation, $I_v(\mu) = I_v$ being independent of angle, the Eddington factor $f^K = 1/3$. Assuming the Eddington factor to be specified, one may combine the two moment (2.30) and (2.31) together,

$$\frac{d^2(f_v^K J_v)}{d\tau_v^2} = J_v - S_v . \quad (2.33)$$

This equation is very useful. It effectively eliminates one independent variable, the angle μ , from the problem. Numerically, it replaces the original transfer equation, which is a first-order linear differential equation for the specific intensity, $I_{v\mu}$, by a second-order but still linear differential equation for the mean intensity, J_v . However, it cannot be used alone because the Eddington factor is unknown unless the full solution of the transfer equation is known. However, it can be used to advantage in iterative methods in which the current values of J and K are used to determine the current Eddington factor f^K , which is then kept fixed during the subsequent iteration step. This form is usually used in constructing model stellar atmospheres.

3.2.2 Hydrostatic Equilibrium Equation

Recalling (2.6), this equation reads

$$\frac{dP}{dz} = -\rho g , \quad (2.34)$$

where P is the total pressure. Introducing the Lagrangian mass m , defined as the mass in the column of a cross section of 1 cm^2 above a given point in the atmosphere,

$$dm = -\rho dz , \quad (2.35)$$

the hydrostatic equilibrium equation is rewritten as

$$\frac{dP}{dm} = g , \quad (2.36)$$

which, since g is constant in a plane-parallel atmosphere, has a trivial solution, $P(m) = mg + P(0)$. In fact, this is the reason why one usually chooses m as the basic depth variable of the

1D plane-parallel atmospheres problem. Nevertheless, it should be kept in mind that the total pressure is generally composed of three parts, the gas pressure, P_{gas} , the radiation pressure, P_{rad} , and the turbulent pressure, P_{turb} , i.e.,

$$P = P_{\text{gas}} + P_{\text{rad}} + P_{\text{turb}} = NkT + \frac{4\pi}{c} \int_0^\infty K_\nu d\nu + \frac{1}{2} \rho v_{\text{turb}}^2, \quad (2.37)$$

where v_{turb} is the microturbulent velocity. The hydrostatic equilibrium equation may then be written as (neglecting the turbulent pressure)

$$\frac{dP_{\text{gas}}}{dm} = g - \frac{4\pi}{c} \int_0^\infty \frac{dK_\nu}{dm} = g - \frac{4\pi}{c} \int_0^\infty \frac{\chi_\nu}{\rho} H_\nu d\nu. \quad (2.38)$$

The right-hand side of this equation can be interpreted as the *effective gravity acceleration* since it expresses the action of the true gravity acceleration (acting downward, i.e., toward the center of the star), reduced by the radiative acceleration (acting outward).

3.2.3 Radiative Equilibrium Equation

This expresses the fact that the total radiation flux is conserved – (☛ 2.7),

$$\int_0^\infty H_\nu d\nu = \text{const} = \frac{\sigma}{4\pi} T_{\text{eff}}^4. \quad (2.39)$$

This equation may be rewritten, using the radiative transfer equation, as

$$\int_0^\infty (\kappa_\nu J_\nu - \eta_\nu) d\nu = \int_0^\infty \kappa_\nu (J_\nu - S_\nu) d\nu = 0, \quad (2.40)$$

Notice that (☛ 2.40) contains the *thermal* absorption coefficient κ_ν , not the total absorption coefficient χ_ν . This is because the scattering contributions cancel out. To illustrate this mathematically, let us take an example of electron scattering. The absorption coefficient for the process (see ☛ 2.13) is given by $n_e \sigma_e$; σ_e being the electron scattering (Thomson) cross section. The emission coefficient is then given by $n_e \sigma_e J_\nu$. As it is seen from (☛ 2.40), these two contributions cancel. This is also clear physically because an absorption followed immediately by a reemission of a photon does not change the energy balance of the medium, and therefore cannot contribute to the radiative equilibrium equation. The situation is different for noncoherent, Compton, scattering. This case will not be discussed here; the interested reader is referred to Rybicki and Lightman (1979).

3.2.4 Statistical Equilibrium Equations

They are also sometimes called the *rate equations*. These are given by (☛ 2.23),

$$n_i \sum_{j \neq i} (R_{ij} + C_{ij}) = \sum_{j \neq i} n_j (R_{ji} + C_{ji}). \quad (2.41)$$

The collisional rates are given functions of temperature and electron density (since collisions with electrons are usually most efficient), assumed to be known from atomic physics (e.g., Mihalas 1978). The radiative rates are given by:

1. For bound-bound (line) transitions between levels i and j , assuming that $E_i < E_j$ (that is, i is the lower level and j the upper level of the transition)

$$R_{ij} = B_{ij} \int_0^\infty J_\nu \phi_{ij}(\nu) d\nu \equiv B_{ij} \bar{J}_{ij} , \quad (2.42)$$

$$R_{ji} = A_{ji} + B_{ji} \int_0^\infty J_\nu \psi_{ji}(\nu) d\nu \rightarrow A_{ji} + B_{ji} \bar{J}_{ij} , \quad (2.43)$$

where A and B are the Einstein coefficients. Quantity \bar{J} is called the *frequency-averaged mean intensity* of radiation. $\phi_{ij}(\nu)$ is a normalized *line profile coefficient for absorption*, often called shortly “absorption profile”. It represents a probability density that if a photon is absorbed in the transition $i \rightarrow j$, it is absorbed at an elementary frequency range $(\nu, \nu + d\nu)$. Analogously, ψ_{ji} is the profile coefficient for emission, which is generally different from ϕ . However, if one assumes *complete frequency redistribution*, which is a common approximation in the stellar atmospheres theory, both profiles are equal, $\phi_{ij}(\nu) = \psi_{ji}(\nu)$. In other words, absorbed and reemitted photons are not correlated. A more general situation where the correlation between an absorbed and an emitted photon is taken into account is usually referred to in astrophysical literature as *partial frequency redistribution*. This case will not be considered here; a review can be found, e.g., in Mihalas (1978), Hubeny (1985), and Uitenbroek (2003). The last part of (2.43) corresponds to the complete redistribution.

In the case of a pure Doppler profile, i.e., no intrinsic broadening of the spectral line, which is broadened only due to the thermal motion of radiators, the absorption profile is given by

$$\phi(x) = \exp(-x^2)/\sqrt{\pi} . \quad (2.44)$$

where x is a dimensionless frequency displacement from the line center, ν_0 , measured in units of Doppler widths,

$$x \equiv \frac{\nu - \nu_0}{\Delta\nu_D} , \quad (2.45)$$

where $\Delta\nu_D$ is the Doppler width, given by $\Delta\nu_D = (\nu_0/c)v_{\text{th}}$, with the thermal velocity $v_{\text{th}} = (2kT/m)^{1/2}$, m being the mass of the radiating atom.

In a more general case where there is an intrinsic broadening of lines described by a Lorentz profile in the atomic rest frame (the most common types of intrinsic broadening being the natural, Stark, and Van der Waals broadening – see Mihalas 1978, or monograph by Griem 1974), the profile function is given by

$$\phi(x) = H(a, x)/\sqrt{\pi} , \quad (2.46)$$

where $H(a, x)$ is the Voigt function,

$$H(a, x) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{e^{-y^2}}{(x-y)^2 + a^2} dy . \quad (2.47)$$

The Voigt function is a convolution of the Doppler profile (i.e., the thermal motions) and the Lorentz profile (intrinsic broadening). The parameter a is a damping parameter expressed in units of Doppler width, $a = \Gamma/(4\pi\Delta\nu_D)$, where Γ is the atomic damping parameter. For instance, for the natural broadening of a line originating in a two-level atom, $\Gamma = A_{21}$.

2. The radiative rates for bound-free transitions (continua) are given by

$$R_{ij} = 4\pi \int_{\nu_0}^{\infty} \sigma_{ij}(\nu) / (h\nu) J_{\nu} d\nu, \quad (2.48)$$

where $\sigma(\nu)$ is the corresponding cross section, supplied again by the atomic physics.

The set of rate equations for all levels of an atom would form a linearly dependent system. Therefore, one equation of the set has to be replaced by another equation. Usually, this is the *total number conservation* equation (or abundance definition equation), $\sum_i n_i = N_{\text{atom}}$, where the summation extends over all levels of all ions of a given species.

3.2.5 Charge Conservation Equation

This equation expresses the global electric neutrality of the medium,

$$\sum_i n_i Z_i - n_e = 0, \quad (2.49)$$

where Z_i is the charge associated with level i (i.e., equal to 0 for levels of neutral atoms, 1 for levels for once ionized ions, etc.). The summation now extends over all levels of all ions of all species.

3.3 Numerical Methods

The resulting set forms a highly coupled, highly nonlinear system of equations. Even in a classical case of plane-parallel, horizontally homogeneous, static atmospheres, this presents a considerable numerical challenge. The basic numerical strategies are described below.

3.3.1 Complete Linearization

The physical state of a 1D plane-parallel, horizontally-homogeneous atmosphere in hydrostatic and radiative (or radiative + convective) equilibria is fully described by the set of vectors ψ_d for every depth point, $d, d = 1, \dots, ND$, ND being the number of discretized depth points. The state vector ψ_d is given by

$$\psi_d \equiv \{J_1, \dots, J_{NF}, N, T, n_e, n_1, \dots, n_{NL}\}, \quad (2.50)$$

where J_i is the mean intensity of radiation in the i -th frequency point, N the total particle number density, n_e the electron density, and n_i the atomic level populations. The structural equations are discretized, and the resulting set of nonlinear algebraic equations is solved by linearization, i.e., by an application of the Newton–Raphson method. This approach was first used in the context of stellar atmosphere models in a seminal paper by Auer and Mihalas (1969), who coined the term *Complete Linearization* (CL).

Writing the complete set of equations schematically as $P(\mathbf{x}) = 0$, where \mathbf{x} is a vector composed of all state vectors ψ_d at all depths, the iteration scheme is written as

$$\delta \mathbf{x}^{(n)} \equiv \mathbf{x}^{(n+1)} - \mathbf{x}^{(n)} = -J(\mathbf{x}^{(n)})^{-1} P(\mathbf{x}^{(n)}), \quad (2.51)$$

where J is the Jacobi matrix (Jacobian), $J_{ij} = \partial P_i / \partial x_j$, i.e., the ij -element of the Jacobian is the derivative of the i -th equation with respect to the j -th unknown. Since the system (2.51) represents a *finite difference* solution of at most second-order differential equations, the Jacobian J has a particularly simple structure, namely a block-tridiagonal form, and (2.51) reduces to,

$$-A_d \delta \psi_{d-1} + B_d \delta \psi_d - C_d \delta \psi_{d+1} = L_d. \quad (2.52)$$

Here A , B , C are $NN \times NN$ matrices, and $L_d = P_d(\mathbf{x}^{(n)})$ is the residuum vector (of dimension NN) at depth d ; the total number of unknown per depth is given by $NN = NF + NL + NC$; NC is the number of constraint equations, in this case $NC = 3$.

Equation 2.52 is solved as a block-tridiagonal system, which means that one is left with inverting one $NN \times NN$ matrix per depth point. Therefore, the total computer time for ordinary complete linearization scales roughly as

$$t \propto (NF + NL + NC)^3 \times ND \times N_{\text{iter}}, \quad (2.53)$$

where N_{iter} is the number of iterations needed to obtain a converged solution, defined through a suitably selected convergence criterion. It is immediately clear that the original complete linearization, despite its inherent power and robustness, cannot be used as a general numerical scheme because in realistic calculations, one needs a very large number of frequency points NF to describe the radiation field with a sufficient accuracy – of the order of 10^5 – 10^6 points. Likewise, a number of energy levels NL may also be quite large (10^4 – 10^5 levels). Therefore, one has to seek less global, but faster schemes. The basic options are (1) reducing the size of the problem keeping the general framework of CL intact, (2) avoiding repeated inversions of a Jacobian (examples are the Broyden scheme implemented in this context by Koesterke et al. 1992; or Kantorovich scheme – Hubeny and Lanz 1992), or (3) using an idea of the so-called approximate Newton–Raphson method (e.g., Hillier 1990; Hillier and Miller 1998).

The first category of approaches achieve a reduction of computer time by reducing the size and/or the number of matrices to be inverted, and/or the number of iterations needed. Taking (2.53) as a guide, in order to reduce the computer time in the most efficient way, one has to reduce NF or NL . One can trivially reduce these numbers by considering less frequencies (levels), and thus constructing less accurate models, but this ruled out since the goal is to construct as accurate and reliable models as possible. So, the way to go is keep the necessary (large) numbers of these quantities, but only to eliminate them from matrices to be inverted. Since NF is typically the largest contributor to the size of the linearization matrices, reducing NF is most important. This can be achieved by using an idea of Accelerated Lambda Iteration (ALI) method, which will be briefly described below. For a detailed discussion of other possible improvements and modifications of the Complete Linearization scheme, see Hubeny and Lanz (2003).

For LTE models, an interesting possibility is to linearize radiation intensities in all frequency points, but to reorganize matrices as first suggested by Rybicki (1971) in the context of two-level atom. Mihalas (1978) has reformulated the scheme to be applied for constructing LTE model atmospheres. The scheme was implemented in a stellar atmosphere code TLUSTY (Hubeny 1988; Hubeny and Lanz 1995) by Burrows et al. (2006) and Hubeny and Burrows (2007b) and was used to construct comprehensive grids of model atmospheres for L and T dwarfs. It should be noted that in order to make this scheme practical, all the structural parameters have to be expressed as functions of a single parameter, the temperature.

The idea consists in the following: The original complete linearization organizes the master Jacobi matrix in a block-tridiagonal structure, that is, as an $ND \times ND$ grand block matrix, with inner $(NF + 1) \times (NF + 1)$ matrices. The inner matrices are composed of a diagonal $NF \times NF$ upper left corner (because the individual frequency points are not explicitly coupled), plus a full last row and column (arising from the linearization of the energy equation). The Rybicki scheme reverses the order of the inner and outer structure: The grand matrix is an $(NF + 1) \times (NF + 1)$ blocks matrix with inner blocks being $ND \times ND$ matrices. The outer block structure corresponds to structural parameters (radiation intensities and temperature), and the inner structure to depth points. The upper part of the $NF \times NF$ blocks is a block-diagonal matrix, with each block being an $ND \times ND$ tridiagonal matrix, and the last row and column of blocks are full matrices. The bulk of work consists in inverting NF tridiagonal $ND \times ND$ matrices, which scales only as $ND^2 \times NF$, that is linearly with the number of frequency points. The very unfavorable cubic scaling of computer time with the number of frequencies is now completely avoided and is replaced by a linear scaling. The only cubic scaling is now with the number of depth points, which is always kept quite low (up to about 100). The method is most efficient with just one constraint equation; with more constraints the scaling would involve a term with $(NC \times ND)^3$.

3.3.2 Accelerated Lambda Iteration

Soon after the advent of Complete Linearization, another crucial ingredient of the modern numerical stellar atmospheres and radiation transfer was introduced, namely the first astrophysical application of an iteration scheme that later became known as Accelerated Lambda Iteration, or ALI for short. It was introduced by Cannon (1973a, b) by the name “Operator Perturbation” technique. Although recognized for its potential, the method was not widely used until its reformulation by Scharmer (1981). The first application of the method to NLTE model stellar atmospheres was worked out by Werner (1986), and the term ALI itself was coined by Hamann (1985). A historical review is given by Hubeny (1992) and a more recent review by Hubeny (2003).

The idea of the method is usually explained on a simple case of two-level atom with complete frequency redistribution. Very briefly, the source function is written as

$$S = (1 - \epsilon)\bar{J} + \epsilon B, \quad (2.54)$$

where \bar{J} is a frequency-averaged mean intensity of radiation, ϵ is the photon destruction parameter, and B the Planck function. In this problem, ϵ and B are viewed as specified parameters. The unknown quantity \bar{J} is given by

$$\bar{J} \equiv \int J_\nu \phi_\nu d\nu = \Lambda[S], \quad (2.55)$$

where ϕ_ν is the (normalized) absorption profile coefficient. The second equality in (2.55) represents the formal solution of the transfer equation, which is essentially a process of obtaining the radiation intensity from the source function. In the case of two-level atom, the Lambda operator is linear in the mean intensity. The two-level atom problem may be written as a single integral equation

$$S = (1 - \epsilon)\Lambda[S] + \epsilon B, \quad (2.56)$$

or, upon discretization, as a linear algebraic equation. Although it may be solved in a single step without a need to iterate, a matrix representing Λ may be large and thus expensive to invert.

From the mathematical point of view, the ALI scheme is nothing else but an application of the idea of iterative solution of large linear system by *preconditioning*, long known in numerical analysis. The matrix is split into two parts, $\Lambda = \Lambda^* + (\Lambda - \Lambda^*)$, and the iterative procedure is set up as

$$S^{(n+1)} = (1 - \epsilon)\Lambda^*[S^{(n+1)}] + (1 - \epsilon)(\Lambda - \Lambda^*)[S^{(n)}] + \epsilon B. \quad (2.57)$$

The action of the exact Λ operator is thus split into two contributions: an approximate Λ^* operator that acts on the new iterate of the source function, and the difference $\Lambda - \Lambda^*$ that acts on the old, known, iterate of the source function. The latter contribution may be easily obtained by a formal solution of the transfer equation. With the choice $\Lambda^* = \Lambda$, one recovers the exact, non-iterative scheme, while with $\Lambda^* = 0$, one obtains the classical Lambda iteration (e.g., Mihalas 1978) which is known to converge extremely slowly. Thus, in order that Λ^* operator brings an essential improvement over both schemes, it has to incorporate all the essential properties of the exact Λ but, at the same time, be easy and cheap to invert. During the development of the field, several approximate operators were suggested (for a historical review see Hubeny 1992). In a seminal paper, Olson et al. (1986) performed a rigorous numerical analysis of the problem and demonstrated that a nearly optimum operator is the diagonal (local) part of the exact Λ operator. This makes the corresponding matrix inversions particularly easy because they are reduced to simple algebraic divisions. Such a choice of approximate operator is equivalent to the Jacobi method.

Finally, it should be noted that the convergence properties of the ALI scheme can be significantly enhanced by applying various acceleration techniques, such as the Ng acceleration (Auer 1987; Hubeny and Lanz 1992; Ng 1974), or methods based on Krylov subspace scheme such as the Generalized Minimum Residual (GMRES) scheme (Auer 1991; Hubeny and Burrows 2007a; Klein et al. 1989).

In the context of model stellar atmospheres, one does not use the ALI scheme as a means of evaluating the source function; instead, it is used as a means of expressing current mean intensities, namely

$$J_\nu = \Lambda_\nu^* S_\nu + (\Lambda_\nu - \Lambda_\nu^*) S_\nu^{\text{old}}, \quad (2.58)$$

where the second term represents a “correction” which is known from the previous iteration. The source function is a function of other structural parameters (temperature, electron density, atomic level populations). In the context of a linearization scheme, the first term of (2.58) is easy to linearize in terms of these structural parameters, and the second term is known and need not be linearized at all. The mean intensities of radiation are thus effectively eliminated from the state vector. This is essentially the procedure used by Werner and collaborators (Dreizler and Werner 1993; Werner 1986; Werner et al. 2003).

3.3.3 Hybrid CL/ALI Method

Hubeny and Lanz (1995) developed a variant of this approach, in which one uses (2.58) to eliminate the mean intensities in most, but not all, frequency points, while the intensities in a few, most important, frequencies are treated as in original Complete Linearization (cores of the strongest lines; frequencies just above the ionization thresholds of most important continua).

The method is called *hybrid CL/ALI method* and offers a wide spectrum of options ranging from full Complete Linearization to a full ALI scheme. It was demonstrated that the method combines essential advantages of its both constituents: the computer time per iteration is essentially the same as in the case of full ALI scheme, while the number of iterations needed to get a converged model is essentially the same as in the original Complete Linearization, i.e., quite small.

3.4 Available Modeling Codes

There are several publicly available codes for computing model stellar atmospheres. The following list is by no means exhaustive.

For LTE models, the most popular code is Kurucz's *ATLAS* (Kurucz 1970, 1993) and *MARCS* (Gustafsson et al. 1975). It should be noted that most of the NLTE codes listed below can be used to calculate LTE models as well. Appropriately modified versions of *PHOENIX* and *TLUSTY* were actually used for generating grids of LTE models – see ► Sect. 3.5.1.

In the context of NLTE static models, the first publicly available NLTE model atmosphere code was the “NCAR code” (Mihalas et al. 1975). More recently, popular and widely used codes are *TMAP* – Tuebingen Model Atmosphere Package (Dreizler and Werner 1993; Werner 1986, 1989; Werner et al. 2003), and *TLUSTY* (Hubeny 1988; Hubeny et al. 1994; Hubeny and Lanz 1992, 1995). Static models are also being constructed by codes originally designed for expanding atmospheres (by setting the expansion velocity to a very low value), such as *CMFGEN* (Hillier and Miller 1998) or *PHOENIX* (Hauschildt et al. 1997, 1999a).

It should be noted that besides the codes that solve for the complete NLTE model atmosphere structure, there are several codes that take an atmospheric structure (temperature, density) given and fixed, and solve for statistical equilibrium + radiative transfer for a selected chemical element – the so-called restricted NLTE problem. Although recently these codes lost much of their former appeal because the modern codes are capable of solving the structure plus detailed NLTE rate equations for many species, they are still being used for analyzing stellar spectra. The most popular codes of this sort are *PANDORA* (Avrett and Loeser 2003); *DETAIL/SURFACE* (Butler and Giddings 1985), and, perhaps the most widely used, *MULTI* (Carlsson 1986).

3.5 Existing Model Atmosphere Grids

3.5.1 LTE Models

The most extensive grid of LTE plane-parallel line-blanketed models is that of Kurucz (1979, 1993), widely used by the astronomical community. The grid covers effective temperatures between 3,500 and 50,000 K, $\log g$ between -1 and 5 , and for several metallicities. The term “metallicity” traditionally means that all the chemical species heavier than helium share a common ratio of their abundance to the solar abundance; this ratio is called metallicity. Numerically, the metallicity is often taken as a logarithm of the metal abundance ratio.

Using the *MARCS* code, Gustafsson et al. (1975) generated their original grid of models for cool stars, with T_{eff} between 3,750 and 6,000 K; $\log g$ between 0.75 and 3.0 and metallicities

$-3.0 \leq [M/H] \leq 0$. Recently, Gustafsson et al. (2008) made public a new, very extensive grid of MARCS model atmospheres, with T_{eff} between 2,500 and 8,000 K, $\log g$ between -1 and 5 , and metallicities $-5 \leq [M/H] \leq 1$. They also include “CN-cycled” models with $C/N = 4.07$ (solar), 1.5 and 0.5 , and C/O from 0.09 to 5 , which represents stars of spectral types R, S, and N.

Hauschildt et al. (1999a, b) used their code PHOENIX to generate a grid LTE spherical models for cool stars, called NextGen, with T_{eff} between 3,000 and 10,000 K, with step 200 K; $\log g$ between 3.5 and 5.5 , with step 0.5 , and metallicities $-4.0 \leq [M/H] \leq 0$. Another grid (Allard et al. 2000) is for pre-main-sequence cool stars with T_{eff} between 2,000 and 6,800 K, $\log g$ between 2 and 3.5 with step 0.5 , stellar mass $M = 0.1 M_{\odot}$, and metallicities $-4.0 \leq [M/H] \leq 0$. The models are available online at www.hs.uni-hamburg.de/EN/For/ThA/phoenix/index.html. A detailed comparison between the ATLAS and NextGen models was performed by Bertone et al. (2004).

3.5.2 NLTE Models


During the last three decades, it was amply demonstrated that departures from LTE are crucial for spectroscopic studies of early-type stars, even the photospheric layers. Early NLTE models were constructed already in the late 1960s and in the first half of the 1970s by Mihalas and coworkers (Mihalas 1972; Mihalas and Auer 1972). Nevertheless, the numerical problems and sheer amount of computer time and memory needed for computing non-LTE metal line-blanketed model atmospheres have precluded computing such models until the late 1980s. Thanks to the development of a very efficient numerical methods such as ALI or the hybrid CL/ALI scheme, this last barrier of the classical stellar atmosphere problem was broken and non-LTE metal line-blanketed model, including literally millions of spectral lines in NLTE, are now being constructed more or less routinely.

There are several partial grids of NLTE models for various stellar types mostly of hot stars. The models constructed by the TMAP code for very hot white dwarfs, subdwarfs, and pre-white dwarfs (also known as the PG 1,159 stars) are available online at <http://astro-uni-tuebingen.de/~rauch/TMAP/TMAP.html>. Rauch and Werner (2009) describe the so-called *Virtual Observatory*, which is a web-based interface that enables a user either to extract already computed models, or generate specific model using TMAP, for very hot objects (hottest white dwarfs; super-soft X-ray sources).



The effort of the developers of TLUSTY culminated in the construction of a grid of NLTE fully blanketed models atmospheres for O stars (OSTAR2002; Lanz and Hubeny 2003) and early B stars (BSTAR2006; Lanz and Hubeny 2007). It is believed that these grids, which each took several years of computer time of several top-level workstations, represent a more or less definitive grids of models in the context of 1D plane-parallel geometry, with hydrostatic and radiative equilibrium, and without any unnecessary numerical approximations.




The basic characteristics are as follows: The OSTAR2002 grid contains 680 individual model atmospheres for 12 values of T_{eff} between 27,500 and 55,000 K, with a step of 2,500 K, and 8 values of $\log g$, and for 10 metallicities: 2, 1, $1/2$, $1/5$, $1/10$, $1/30$, $1/50$, $1/100$, $1/1000$, and 0 times the solar metal composition. The following species are treated in NLTE: H, He, C, N, O, Ne, Si, P, S, Fe, Ni, in all important stages of ionization; which means that there are altogether over 1,000 (super)levels to be treated in NLTE, and about 10^7 lines, and about 250,000 frequency points to describe the spectrum.


The BSTAR2006 grid is similar. It contains 1,540 individual models for 16 values of T_{eff} between 15,000 and 30,000 K, with a step of 1,000 K, and for 6 metallicities: 2, 1, 1/2, 1/5, 1/10, and 0 times solar. The species treated in NLTE are the same as in OSTAR2002, adding Mg and Al, but removing Ni, which is less important for B stars. There are altogether about 1,450 (super)levels treated in NLTE, about 10^7 lines, and about 400,000 frequency points. The models for both grids are available online at <http://nova.astro.umd.edu>.

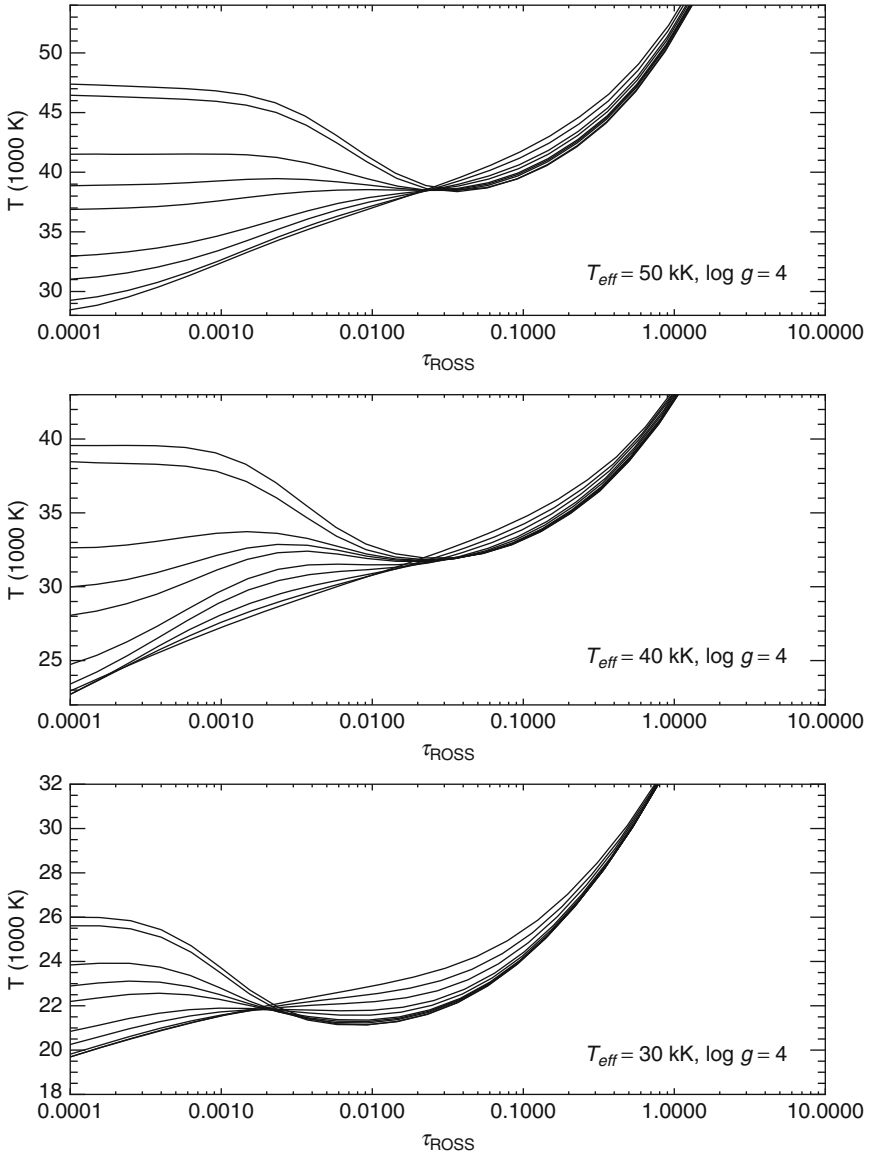
Several representative results from the grids are shown below.  [Figure 2-1](#) displays the temperature structure for three representative effective temperatures of the OSTAR2002 grid. The temperature distribution nicely illustrates the basic features of line blanketing, namely the so-called back-warming (line blanketing leads to a heating of moderately deep atmospheric layers between Rosseland optical depths 0.01 and 1), and a surface cooling. The zero-metallicity and low metallicities models exhibit a temperature rise at the surface, a typical NLTE effect discovered by Auer and Mihalas (1969) and explained as an indirect heating effect of the hydrogen Lyman and Balmer lines. This effect competes with surface cooling caused by metal lines, and these effects nearly cancel at metallicities 1/50 (for hotter models) to 1/10 (cooler models). Interestingly, the temperature curves for all metallicities cross in a very narrow range of optical depths.

From the practical point of view, the most important result of model atmospheres is the prediction of emergent radiation, which is then compared to the observed spectrum in order to determine the basic stellar parameters. Also, theoretical predictions are indispensable for estimating the radiation in unobservable spectrum regions, in particular in the hydrogen Lyman continuum (wavelength less than 912 Å), which produces ionizing photons, but which cannot be directly detected for early-type stars because of the absorption by interstellar hydrogen. (Only two early-type stars, ϵ and β CMa, which are relatively close, and which lie in the direction of a “tunnel” of low density in the local interstellar medium, have detectable Lyman continuum flux as observed by the *EUVE* satellite.)

A sensitivity of the predicted spectra to the effective temperature is depicted in  [Fig. 2-2](#), which shows (from top to bottom) emergent spectra for 50, 45, 40, 35, and 30 kK. Notice a diminishing Lyman jump at 912 Å when going to higher temperatures; this is a consequence of increased ionization of hydrogen. A similar sensitivity to metallicity is shown in  [Fig. 2-3](#). The models with higher metallicities exhibit deeper lines but higher continuum flux.

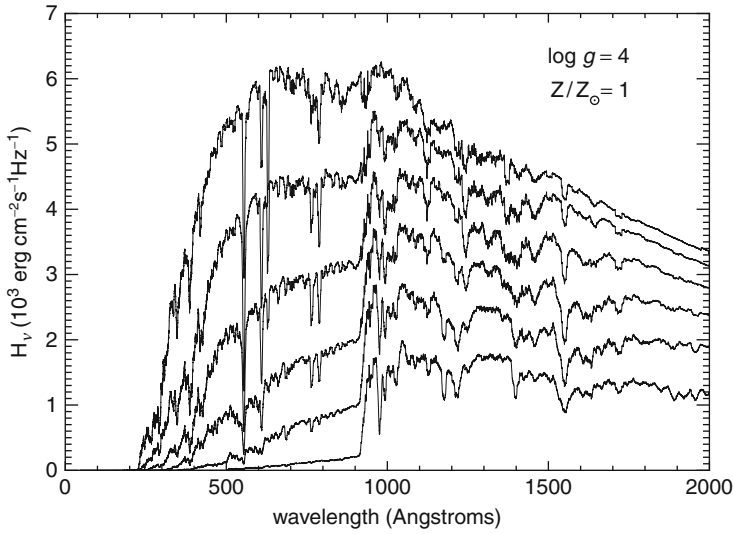
 [Figure 2-4](#) shows the predicted flux from the OSTAR2002 grid for three solar composition model atmospheres with $T_{\text{eff}} = 30,000$ K, 40,000 K, and 50,000 K, for $\log g = 3$, degraded to a lower resolution (about 5 Å), compared to Kurucz model fluxes. Differences both in the continuum level, as well as in the individual line features, are clearly seen. An analogous comparison of the predicted flux from the BSTAR2006 grid for three solar composition model atmospheres with $T_{\text{eff}} = 25,000$ K, 20,000 K, and 15,000 K, for $\log g = 3$, degraded to a lower resolution (about 5 Å), compared to Kurucz model fluxes, is presented in  [Fig. 2-5](#). A similar comparison of high-resolution spectra in several wavelength intervals is presented in  [Fig. 2-6](#).

As an actual example of a fit of OSTAR2002 models to observations,  [Fig. 2-7](#) displays a sample of the predicted flux for a model for $T_{\text{eff}} = 33,500$ K, $\log g = 3.85$, and a high-resolution, high signal-to-noise observation of a late-O main-sequence star 10 Lac secured by the *Goddard High Resolution Spectrograph* (GHRS) aboard the *Hubble Space Telescope* (Hubeny et al. 1998; Lanz et al., in preparation). The agreement between observations and predictions is excellent and demonstrates a power of the present-day model atmospheres of early-type stars.



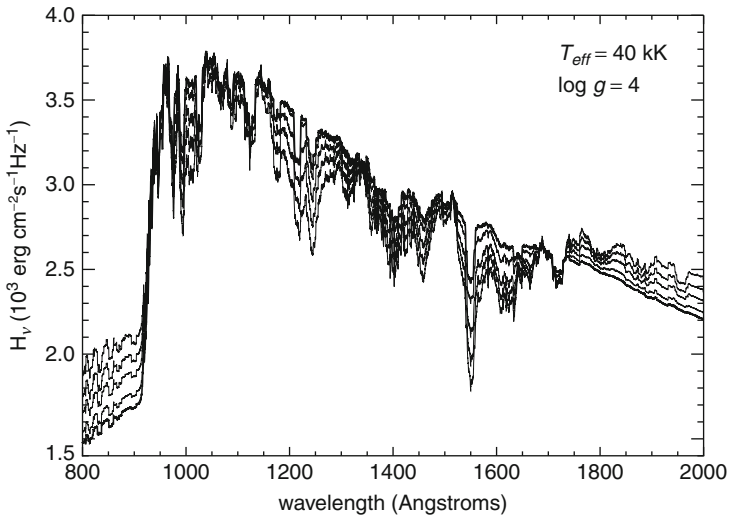
■ Fig. 2-1

Temperature as a function of Rosseland optical depth for OSTAR2002 model atmospheres with $T_{\text{eff}} = 50,000 \text{ K}$ (top), $40,000 \text{ K}$ (middle), and $30,000 \text{ K}$ (bottom); $\log g = 4.0$, and various metallicities. At low optical depths ($\tau_{\text{Ross}} < 10^{-3}$), the top curves are for a pure H-He model, and temperature is progressively lower when increasing the metallicity, while the reverse applies in deep layers ($\tau_{\text{Ross}} > 10^{-2}$) (From Lanz and Hubeny 2003)



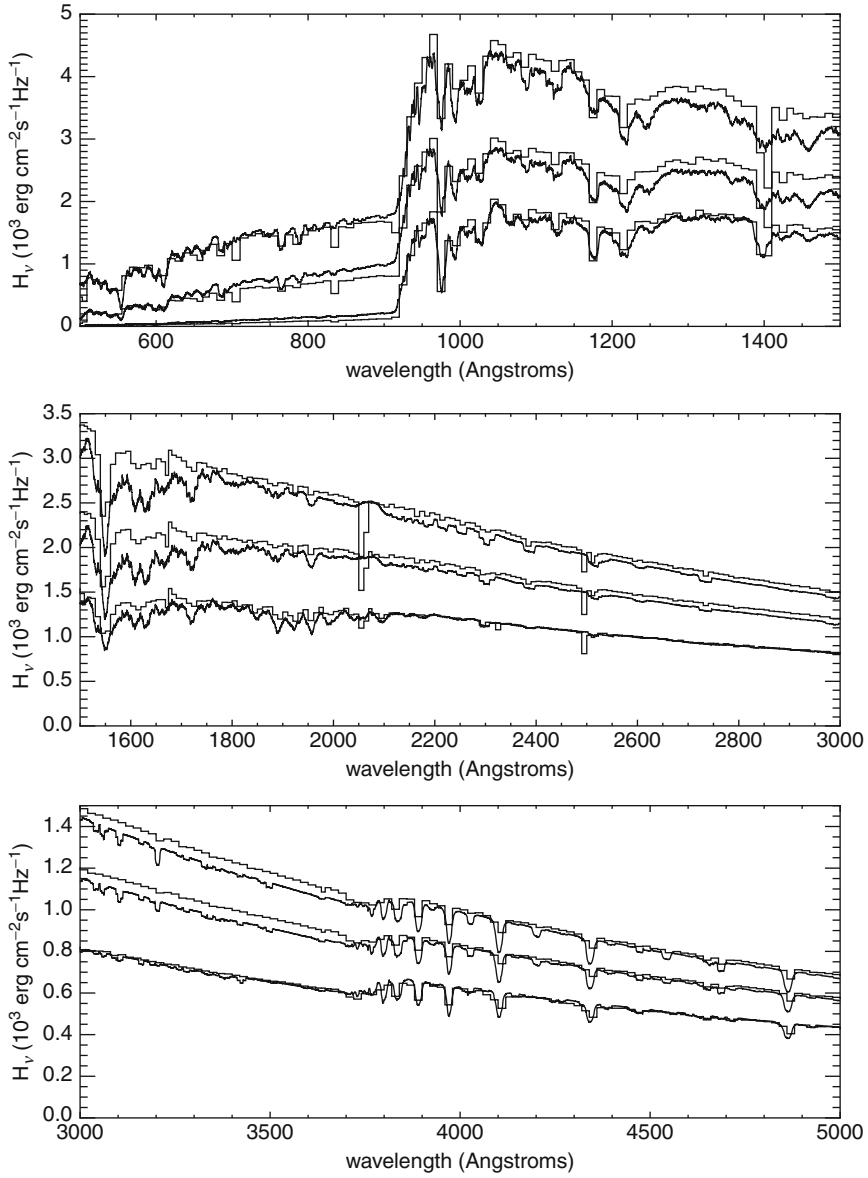
■ Fig. 2-2

Predicted flux for six solar composition OSTAR2002 model atmospheres with T_{eff} between 55 and 30 kK, with a step of 5 kK, and for $\log g = 4$ (From Lanz and Hubeny 2003)



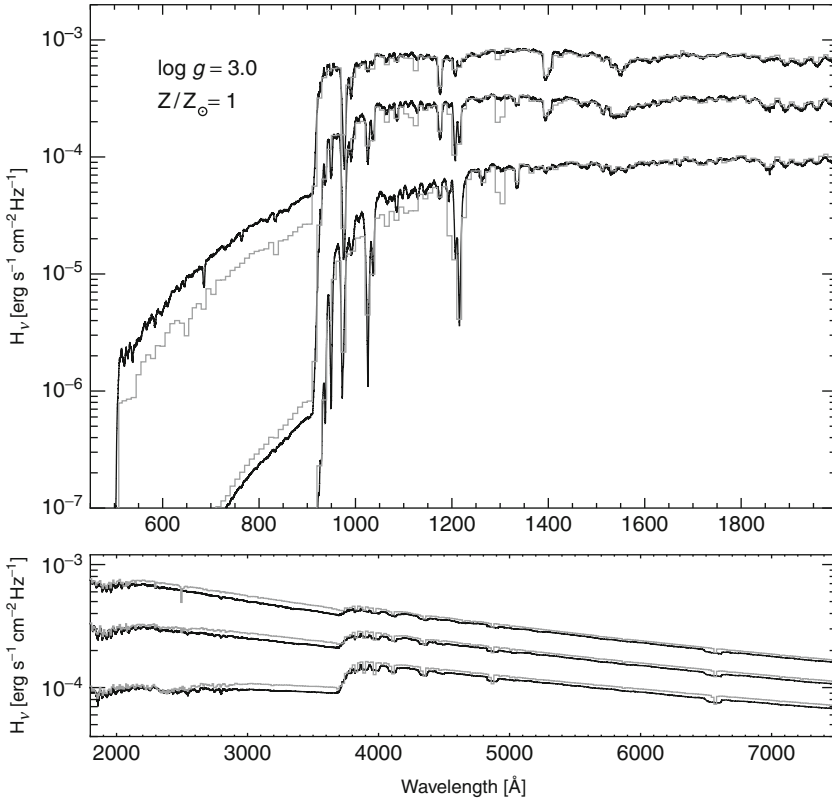
■ Fig. 2-3

Predicted flux for five model atmospheres with $T_{\text{eff}} = 40$ kK and $\log g = 4$, for five different metallicities (From Lanz and Hubeny 2003)



■ Fig. 2-4

Predicted flux for three solar composition mode atmospheres with $T_{\text{eff}} = 50,000 \text{ K}$, $40,000 \text{ K}$, and $30,000 \text{ K}$, for $\log g = 4$ (black lines), compared to Kurucz models with the same parameters (gray histograms). The OSTAR2002 model fluxes were degraded to lower resolution to roughly match the resolution of the Kurucz grid (From Lanz and Hubeny 2003)



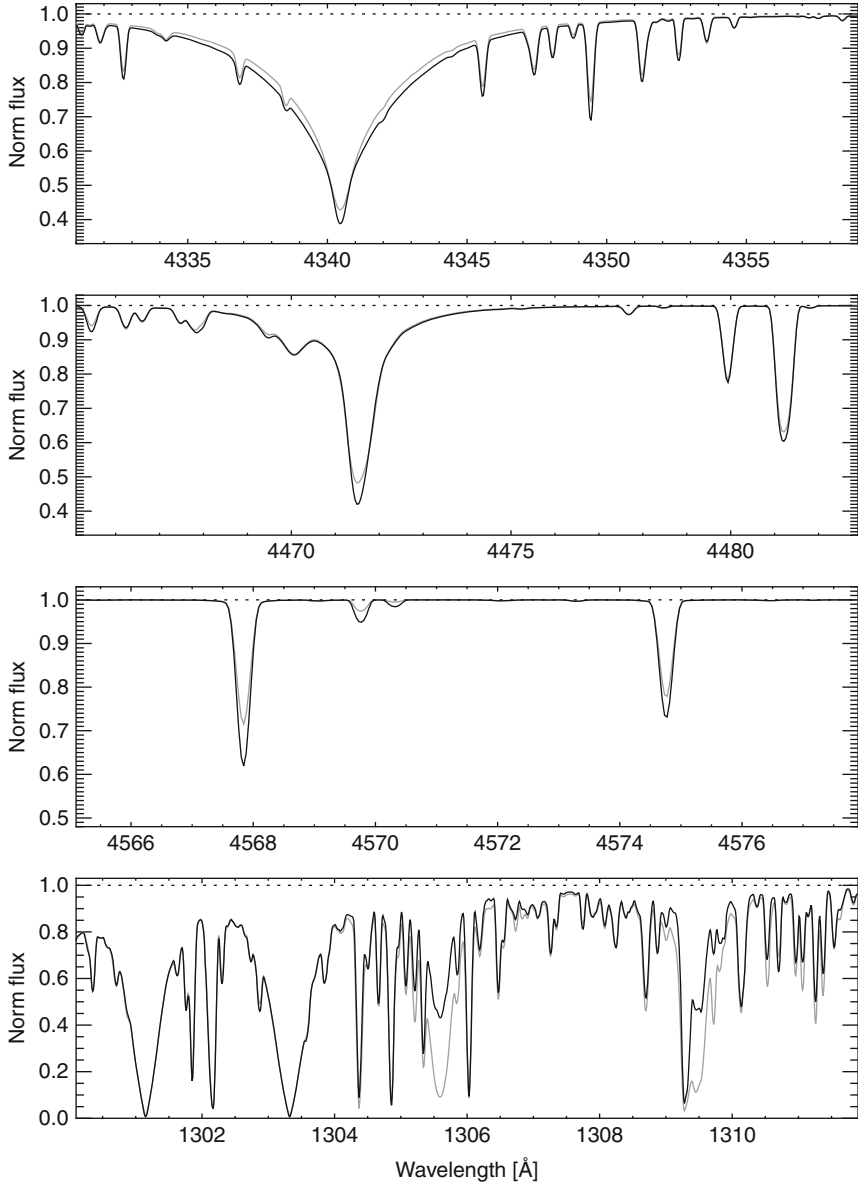
■ Fig. 2-5

Predicted flux for three solar composition model atmospheres with $T_{\text{eff}} = 25,000$ K, $20,000$ K, and $15,000$ K, for $\log g = 3$ (black lines), compared to Kurucz models with the same parameters (gray histograms). The BSTAR2006 model fluxes were degraded to lower resolution to roughly match the resolution of the Kurucz grid (From Lanz and Hubeny 2007)

4 Using Model Stellar Atmospheres to Determine the Fundamental Stellar Parameters

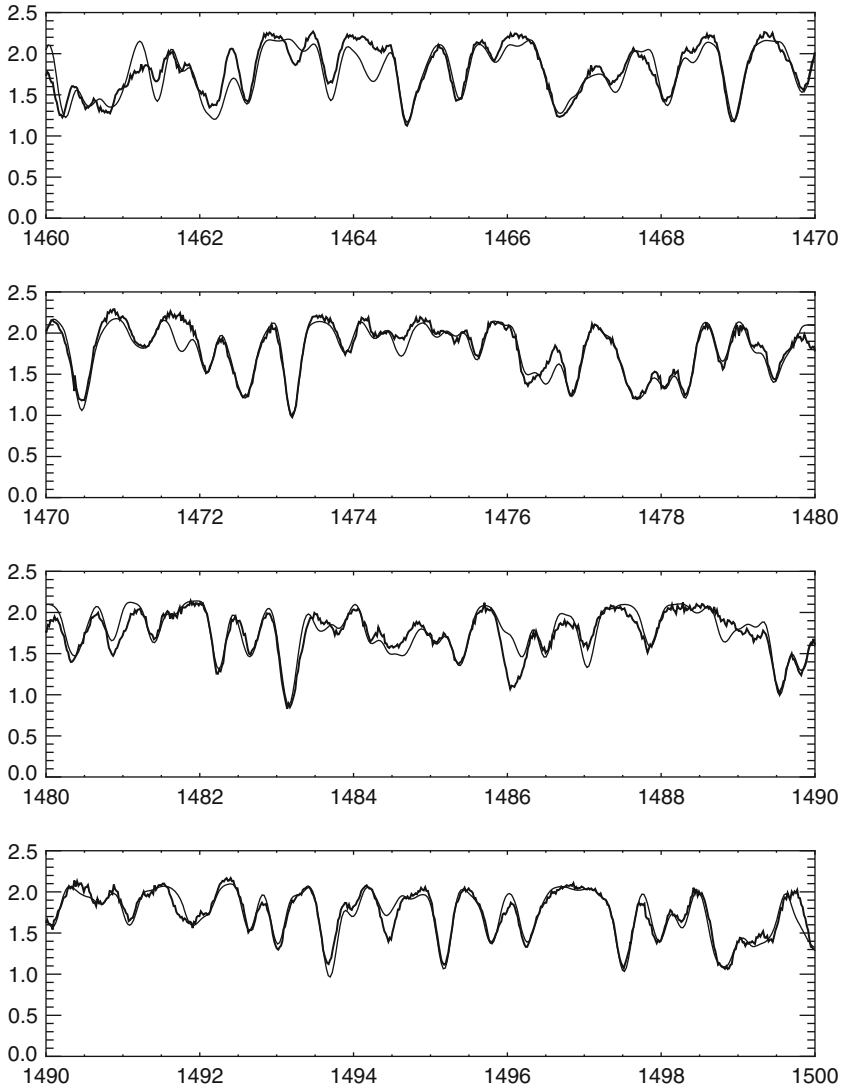
The fundamental stellar parameters to be determined are the stellar mass, M_* , radius, R_* , luminosity, L_* , and the chemical composition, i.e., the set of abundances (ratios of the total number of particles of species A with respect to hydrogen). If the distance to the star, d , is not known, it is added to the list of fundamental parameters to be determined, even if it does not represent an intrinsic stellar property.

A determination of chemical composition is a subject of a vast volume of literature so that there is no need to repeat it here. Essentially, they are determined by comparing the observed line profiles (or integrated quantities, like equivalent widths) with those predicted from model atmospheres.



■ Fig. 2-6

Predicted line profiles for solar-composition model atmospheres with $T_{\text{eff}} = 20,000 \text{ K}$, $\log g = 3.0$, and $v_{\text{turb}} = 2 \text{ km s}^{-1}$ (*black lines*); compared to the Kurucz model with the same parameters (*gray lines*) (From Lanz and Hubeny 2007)



■ Fig. 2-7

A comparison of the observed HST/GHRS flux for 10 Lac (*heavy line*) and the predicted flux from the fully blanketed NLTE model atmosphere with $T_{\text{eff}} = 33,500 \text{ K}$, $\log g = 3.85$, and for the solar abundances of all species (*thin line*). The abscissa is the wavelength in \AA , and the ordinate is the flux in $10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ \AA}^{-1}$. Most spectral features are lines of Fe IV, Fe V, Ni IV, and Ni V. A difference between theory and predictions is hardly seen on the plots

There are, obviously, other stellar parameters, like the rotational velocity and others, but for the purposes of this chapter, it is assumed that these additional parameters and the chemical abundances are determined independently of the fundamental parameters listed above.

The parameters which can be determined directly from observations are the effective temperature, T_{eff} , and surface gravity, g . In addition, one has the measured magnitude, m_{obs} , that

reflects the whole observationally accessible wavelength range. If the flux in the unobservable region is negligible, then this magnitude represents the total, *bolometric*, magnitude, m_{bol} . If not, one has to apply the bolometric correction, which follows from the model atmosphere.

In any case, one ends up with three “measured” quantities, T_{eff} , g , and m_{bol} , but there are four unknown fundamental parameters, M_* , R_* , L_* , and d . The governing relations between them are

$$\sigma T_{\text{eff}}^4 = L_*/(4\pi R_*^2) , \quad (2.59)$$

$$g = GM_*/R_*^2 , \quad (2.60)$$

$$L_* = L_* [m_{\text{bol}} (m_{\text{obs}}, T_{\text{eff}}), d] , \quad (2.61)$$

The last relation expresses the conversion of the observed magnitude to the stellar luminosity.

There are thus three relations for four unknowns. In fact, in some cases, the stellar evolution theory may supply an independent additional relation between the fundamental parameters, for instance, the mass–radius relation for white dwarfs (Hamada and Salpeter 1961) or the mass–luminosity relation for central stars of planetary nebulae (Paczynski 1971). However, in the general situation one does not have such a relation, and, even if so, one may want to check the theoretically predicted relations observationally.

Therefore, from the photospheric analysis only, one cannot derive all four parameters simultaneously. This is easily understood from the physical point of view. A plane-parallel hydrostatic atmosphere is just a thin layer located on the top of a spherical star. The only information about a dimension of the underlying star is contained in the surface gravity g which depends also on the stellar mass. Since the atmosphere is thin, the emergent spectrum does not carry any independent information about the atmospheric extent.

To remove the radius–mass degeneracy, either an independent geometrical information (knowing the radius or the distance) or an independent knowledge of the mass is needed. A typical situation is that the distance d is known (e.g., from Hipparchos parallaxes); then the other parameters are determined as follows:

1. From known m_{obs} and d (and, possibly, T_{eff}), the absolute bolometric magnitude, M_{bol} and, therefore, luminosity, L_* , is determined.
2. From L_* and T_{eff} , the stellar radius, R_* , is determined.
3. From R_* and g , the stellar mass, M_* , is determined.

As it turns out, if the mass of an early-type O stars is determined in this way, the so-called spectroscopic mass, and if the mass is also determined by comparing the evolutionary tracks and the position of the star in the H-R diagram, the so-called evolutionary mass, one finds a significant discrepancy (e.g., Herrero et al. 1992). The sense of discrepancy is that the spectroscopic masses are systematically lower than the evolutionary masses. The discrepancy arises either by inaccuracies of the stellar atmospheres theory, or the stellar evolution theory, or, most likely, both. From the stellar atmospheres side, there has been a progress in understanding the reasons for the discrepancy (e.g., Lanz et al. 1996), namely as an effect of a previously neglected effects of metal line blanketing on the atmospheric structure. However, the problem is not yet fully solved.

5 Summary and Outlook

Thanks to the concentrated effort of several groups of researchers, starting with pioneering work of Mihalas and Auer in the late 1960s and early 1970s, and a continuing effort during the last three decades, the problem of constructing 1D stationary model atmospheres, even with full metal line blanketing and in NLTE, is now conceptually solved, although a significant amount of new models still need to be computed.

Even in the domain of LTE models, new discoveries led to new challenges. In particular, the rapidly evolving field of modeling atmospheres of substellar-mass objects, such as brown dwarfs and extrasolar giant planets, provides a natural extension of the traditional stellar atmosphere theory to new and exciting domains, and contributes to a revival of stellar atmospheres theory in general.

The next big step will be to go to 3D radiation hydrodynamic models, which will undoubtedly be one of the main themes of theoretical astrophysics in the next decade. The effort in this area is under way, but much remains to be done. The future of the stellar atmospheres theory thus seems quite bright and exciting.

References

- Allard, F., Hauschildt, P. H., & Schweitzer, A. 2000, *ApJ*, 539, 366
- Anderson, L. S. 1985, *ApJ*, 298, 848
- Anderson, L. S. 1987, in *Numerical Radiative Transfer*, ed. W. Kalkofen (Cambridge: Univ. Cambridge Press), 163
- Anderson, L. S. 1989, *ApJ*, 339, 588
- Auer, L. H. 1987, in *Numerical Radiative Transfer*, ed. W. Kalkofen (Cambridge: Univ. Cambridge Press), 101
- Auer, L. H. 1991, in *NATO ASI Ser. C 341, Stellar Atmospheres: Beyond Classical Models*, ed. L. Crivellari, I. Hubeny, & D. G. Hummer (Dordrecht: Kluwer), 9
- Auer, L. H., & Mihalas D. 1969, *ApJ*, 158, 641
- Auer, L. H., & Mihalas, D. 1970, *MNRAS*, 149, 60
- Avrett, E. H., & Loeser, J. 2003, in *IAU Symp. 210, Modeling of Stellar Atmospheres*, eds. W. Weiss & N. Piskunov (Dordrecht: Kluwer), 14
- Bertone, E., Buzzoni, A., Chavez, M., & Rodriguez-Merino, L. H. 2004, *AJ*, 128, 829
- Böhm-Vitense, E. 1989, *Introduction to Stellar Astrophysics II. Stellar Atmospheres* (Cambridge: Univ. Cambridge Press)
- Burgasser, A. J., Kirkpatrick, J. D., Brown, M. E., Reid, I. N., Gizis, J. E., Dahn, C. C., Monet, D. G., Beichman, C. A., Liebert, J., Cutri, R. M., & Skrutskie, M. F. 1999, *ApJ*, 522, L65
- Burrows, A., Sudarsky, D., & Hubeny, I. 2006, *ApJ*, 640, 1063
- Butler, K., & Giddings, J. 1985, *Newsletter on Analysis of Astronomical Spectra. No. 9* (Daresbury, England: Daresbury Laboratory)
- Cannon, C. J. 1973a, *J Quant Spectrosc Radiat Transfer*, 13, 627
- Cannon, C. J. 1973b, *ApJ*, 185, 621
- Cannon, C. J. 1985, *The Transfer of Spectral Line Radiation* (Cambridge: Univ. Cambridge Press)
- Carlsson, M. 1986, *A Computer Program for Solving Multi-Level Non-LTE Radiative Transfer Problems in Moving or Static Atmospheres* (Melville, NY: AIP)
- Carlsson, M., & Stein, R. F. 2003, in *ASP Conf. Ser. 288, Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 505
- Castor, J. I. 2004, *Radiation Hydrodynamics*, (Cambridge: Univ. Cambridge Press)
- Chandrasekhar, S. 1960, *Radiative Transfer* (New York: Dover)
- Delfosse, X., et al. 1997, *A&A*, 327, L25
- Dreizler, S., & Werner, K. 1993, *A&A*, 278, 199
- Gray, D. F. 1992, *Observations and Analysis of Stellar Photospheres* (2nd eds.; Cambridge: Univ. Cambridge Press)
- Griem, H. R. 1974, *Spectral Line Broadening by Plasmas* (New York: Acad. Press)
- Gustafsson, B., Bell, R. A., Eriksson, K., & Nordlund, A. 1975, *A&A*, 42, 407

- Gustafsson, B., Edvardsson, B., Eriksson, K., Jørgensen, U. G., Nordlund, A., & Plez, B. 2008, *A&A*, 486, 951
- Hamada, T., & Salpeter, E. E. 1961, *ApJ*, 134, 683
- Hamann, W.-R. 1985, *A&A*, 148, 364
- Hauschildt, P. H., Baron, E., & Allard, F. 1997, *ApJ*, 483, 390
- Hauschildt, P. H., Allard, F., & Baron, E. 1999a, *ApJ*, 512, 377
- Hauschildt, P. H., Allard, F., Ferguson, J., & Baron, E. 1999b, *ApJ*, 525, 871
- Herrero, A., Kudritzki, R. P., Vilchez, J. M., Kunze, D., Butler, K., & Haser, S. 1992, *A&A*, 261, 209
- Hillier, D. J. 1990, *A&A*, 231, 116
- Hillier, J. D. 2003, in ASP Conf. Ser. 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 199
- Hillier, D. J., & Miller, D. L. 1998, *ApJ*, 496, 407
- Hoeflich, P. 1995, *ApJ*, 443, 89
- Hoeflich, P. 2003, in ASP Conf. Ser. 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 185
- Hubeny, I. 1985, in *Progress in Stellar Spectral Line Formation Theory*, NATO ASI Series 152, eds. J. E. Beckman & L. Crivellari (Dordrecht: Reidel), 27
- Hubeny, I. 1988, *Comp Phys Commun*, 52, 103
- Hubeny, I. 1992, in *Lecture Notes in Phys.* 401, *The Atmospheres of Early-Type Stars*, eds. U. Heber & C. J. Jeffery (Berlin: Springer), 377
- Hubeny, I. 2003, in ASP Conf. Ser. 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 17
- Hubeny, I., & Burrows, A. 2007a, *ApJ*, 659, 1458
- Hubeny, I., & Burrows, A. 2007b, *ApJ*, 669, 1248
- Hubeny, I., Heap, S. R., & Lanz, T. 1998, in ASP Conf. Ser. 131, *Boulder- Munich II: Properties of Hot, Luminous Stars*, eds. I. Howarth (San Francisco: ASP), 108
- Hubeny, I., Hummer, D. G., & Lanz, T. 1994, *A&A*, 282, 151
- Hubeny, I., & Lanz, T. 1992, *A&A*, 262, 501
- Hubeny, I., & Lanz, T. 1995, *ApJ*, 439, 875
- Hubeny, I., & Lanz, T. 2003, in ASP Conf. Ser. 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 51
- Hubeny, I., Mihalas, D., & Werner, K. (eds.) 2003, *ASP Conf. Ser. 288 Stellar Atmosphere Modeling* (Melville, NY: AIP)
- Hubeny, I., Stone, J. M., MacGregor, K., & Werner, K. (eds.) 2009, *AIP Conf. Proc.* 1171, *Recent Directions in Astrophysical Quantitative Spectroscopy and Radiation Hydrodynamics* (Melville, NY: AIP)
- Hummer, D. G., Berrington, K. A., Eissner, W., Pradhan, A. K., Saraph, H. E., & Tully, J. A. 1993, *A&A*, 279, 298
- Iglesias, C. A., & Rogers, F. J. 1991, *ApJ*, 371, 408
- Iglesias, C. A., & Rogers, F. J. 1996, *ApJ*, 464, 943
- Jefferies, J. 1968, *Spectral Line Formation* (Waltham: Blaisdel)
- Jørgensen, U. G. 2003, in ASP Conf. Ser. 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 303
- Kalkofen, W. (ed.) 1984, *Methods in Radiative Transfer* (Cambridge: Univ. Cambridge Press)
- Kalkofen, W. (ed.) 1987, *Numerical Radiative Transfer* (Cambridge: Univ. Cambridge Press)
- Klein, R. I., Castor, J. I., Greenbaum, A., Taylor, D., & Dykema, P. 1989, *J Quant Spectrosc Radiat Transfer*, 41, 199
- Koesterke, L., Hamann, W.-R. & Kosmol, P. 1992, *A&A*, 255, 490
- Koesterke, L., Hamann, W.-R. & Graefener, G. 2002, *A&A*, 384, 562
- Kurucz, R. L. 1970, *Smithsonian Astrophys Obs Spec Rep No.* 309
- Kurucz, R. L. 1979, *ApJS*, 40, 1
- Kurucz, R. L. 1993, *Kurucz CD-ROM No. 13* (Cambridge, MA: Smithsonian Astrophys. Obs.)
- Kurucz, R. L. 2009, in *AIP Conf. Proc.* 1171, *Recent Directions in Astrophysical Quantitative Spectroscopy and Radiation Hydrodynamics*, eds. I. Hubeny et al. (Melville, NY: AIP), 45
- Lanz, T., & Hubeny, I. 2003, *ApJS*, 146, 417
- Lanz, T., & Hubeny, I. 2007, *ApJS*, 169, 83
- Lanz, T., de Koter, A., Hubeny, I., & Heap, S. R. 1996, *ApJ*, 465, 359
- Mihalas, D. 1972, *Non-LTE model atmospheres for B and O stars*, NCAR-TN/STR-76 (Boulder: NCAR)
- Mihalas, D. 1978, *Stellar Atmospheres* (2nd eds.; San Francisco: Freeman)
- Mihalas, D., & Auer, L. H. 1972, *ApJS*, 24, 193
- Mihalas, D., Heasley, J. N., & Auer, L. H. 1975, *A Non-LTE model stellar atmospheres computer program*, NCAR-TN/STR-104 (Boulder: NCAR)
- Mihalas, D., & Mihalas, B. 1984, *Foundations of Radiation Hydrodynamics* (New York: Univ. Oxford Press)
- Ng, K. C. 1974, *J Chem Phys*, 61, 2680
- Nordlund, Å., & Stein, R. F. 2009, in *AIP Conf. Proc.* 1171, *Recent Directions in Astrophysical Quantitative Spectroscopy and Radiation Hydrodynamics*, eds. I. Hubeny et al. (Melville, NY: AIP), 242
- Olson, G. L., Auer, L. H., & Buchler, J. R. 1986, *J Quant Spectrosc Radiat Transfer*, 35, 431

- The Opacity Project Team 1995, *The Opacity Project*, Vol. 1 (Bristol: Inst. of Physics Publications)
- The Opacity Project Team 1997, *The Opacity Project*, Vol. 2 (Bristol: Inst. of Physics Publications)
- Paczynski, B. E. 1971, *Acta Astr.*, 21, 417
- Pauldrach, A. W. A., Hoffmann, T. L., & Lennon, M. 2001, *A&A*, 375, 161
- Pradhan, A. K., Zhang, H. L., Nahar, S. N., Romano, P., & Bautista, M. A. 1996, *BAAS*, 189, 72.11
- Rauch, T., & Werner, K. 2009, in *AIP Conf. Proc.* 1171, *Recent Directions in Astrophysical Quantitative Spectroscopy and Radiation Hydrodynamics*, eds. I. Hubeny et al. (Melville, NY: AIP), 85
- Rutten, R. J. 1995, *Radiative Transfer in Stellar Atmospheres* (2nd WWW edition, 1995, Robert J. Rutten, Utrecht, Netherlands)
- Rybicki, G. B. 1971, *J Quant Spectrosc Radiat Transfer*, 11, 589
- Rybicki, G. B., & Lightman, A. P. 1979, *Radiative Processes in Astrophysics* (New York: Wiley-Interscience)
- Santolaya-Rey, A. E., Puls, J., & Herrero, A. 1997, *A&A*, 323, 488
- Scharmer, G. B. 1981, *ApJ*, 249, 720
- Seaton, M. J. 1987, *J Phys B* 20, 6363
- Sharp, C. S., & Burrows, A. 2007, *ApJS*, 168, 140
- Shu, F. H. 1991, *The Physics of Astrophysics I. Radiation* (Mill Valley: University Science Books)
- Stein, R. F., & Nordlund, Å. 2003, in *ASP Conf. Ser.* 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, & K. Werner (San Francisco: ASP), 519
- Uitenbroek, H. 2003, in *ASP Conf. Ser.* 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, K. Werner (San Francisco: ASP), 597
- Werner, K. 1986, *A&A*, 161, 177
- Werner, K. 1989, *A&A*, 226, 265
- Werner, K., Deetjen, J. L., Dreizler, S., Nagel, T., Rauch, T., & Schuh, S. L. 2003, in *ASP Conf. Ser.* 288, *Stellar Atmosphere Modeling*, eds. I. Hubeny, D. Mihalas, K. Werner (San Francisco: ASP), 31

3 **The Sun as a Star**

Giulio Del Zanna · Helen Mason

DAMTP, Centre for Mathematical Sciences, University of
Cambridge, Cambridge, UK

1	<i>Introduction</i>	90
2	<i>Historic Perspective</i>	93
3	<i>Observing the Solar Atmosphere from Space</i>	98
3.1	Early Space Missions	98
3.2	Yohkoh	101
3.3	Ulysses	102
3.4	The Solar and Heliospheric Observatory – SoHO	103
3.5	Transition Region and Coronal Explorer (TRACE)	108
3.6	Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)	108
3.7	CORONAS Missions	108
3.8	Hinode	110
3.9	STEREO	113
3.10	Solar Dynamics Observatory (SDO)	114
3.11	Imaging Versus Spectroscopic Instruments	115
4	<i>The Lower Solar Atmosphere</i>	115
4.1	The Photosphere and the Sun's Magnetic Field	115
4.2	The Chromosphere	118
4.3	The Transition Region	122
5	<i>The Interior of the Sun</i>	125
5.1	The Standard Model and the Solar Neutrinos	125
5.2	Solar Oscillations	126
5.3	Some Results from Helioseismology Inversions	130
5.4	Abundances and Helioseismic Models	133
5.5	Local Helioseismology	134
5.6	The Solar Dynamo	135
6	<i>Radiation from the Solar Corona: Atomic Processes and Plasma Diagnostics</i> ...	135
6.1	Line Flux and Intensity	137
6.2	Differential Emission Measure and Emission Measure	138
6.3	Atomic Processes Affecting the Ion Excitation	138
6.4	Atomic Processes Affecting the Ion State	140
6.5	Coronal Model Approximation: Two-Level Ion	141
6.6	Emission Measure Approximations	142

6.7	A Word of Warning: Problems with Emission Measures	145
6.8	Electron Density Determination	145
6.9	Spectroscopic Filling Factors	147
6.10	Electron Temperature Determination	148
6.11	Continuum	149
6.12	Line Widths, Ion Temperatures, and Doppler Motions	150
6.13	The CHIANTI Atomic Package	150
6.14	Benchmarking the Atomic Data	152
7	<i>Chemical Abundances</i>	152
7.1	Photospheric Abundances	153
7.2	Coronal Abundances	153
7.3	Abundances from In Situ Measurements	155
7.4	Depletion, Enhancement or Both?	155
7.5	Helium, Neon, and Argon	156
7.6	Modeling	158
7.7	Stellar and Galactic Abundances	158
8	<i>The Solar Corona</i>	159
8.1	Solar Features as Seen in X-rays and EUV	159
8.2	Solar Active Regions	160
8.2.1	Active Regions Observed with SoHO CDS	161
8.2.2	Active Regions Observed with Hinode EIS	162
8.3	Coronal Heating	162
8.3.1	Coronal Heating in Loops	165
8.3.2	Hydrostatic Models and the RTV Scaling Laws	165
8.3.3	Hydrodynamic Models of Active Region Loops	166
8.3.4	Microflares and Nanoflares	169
9	<i>Solar Flares and Coronal Mass Ejections</i>	170
9.1	Solar Flares	170
9.2	Coronal Mass Ejections (CMEs)	174
9.3	Theoretical Concepts of CME Initiation	176
9.4	Flares on Other Stars	178
10	<i>Solar Wind</i>	182
10.1	The Heliosphere	185
10.2	Physical Characteristics of the Solar Wind and Models	185
10.3	The Sources of the Solar Wind	187
10.4	Ulysses	188
11	<i>Solar Irradiance</i>	189
11.1	Introduction	189
11.2	Total Solar Irradiance, TSI, and Surface Temperature	190
11.3	Irradiance in the UV	195
11.3.1	Irradiance in the EUV and X-rays	196

12 *Future Prospects* 199

12.1 Solar Orbiter 200

12.2 Solar Probe+ 200

References 201

Abstract: In this chapter we provide an overview of various observational and theoretical aspects about the Sun, paying more attention to global ones, i.e., to those where the Sun is considered as a star, so they can be more or less directly related to those studied for other stars. The emphasis is on the processes by which energy is transferred to the solar atmosphere and then flows from the Sun in the form of radiation and particles. The chapter starts with an historical introduction, and then discusses various subjects, starting with a brief description of the photosphere, the chromosphere, and the transition region of the Sun. The interior of the Sun is then introduced, with the standard model and solar neutrinos, and with the observations and theory of solar oscillations. We present some important results from helioseismology, and briefly discuss the problematics involved with the production and regeneration of the solar magnetic fields during a solar cycle.


We then discuss the atomic processes which lead to the formation of the optically thin radiation that is emitted from the X-rays to the UV from the solar corona. In particular, the spectroscopic diagnostic techniques which are used to determine the physical parameters of a stellar corona are discussed in some depth.

We describe some of the in situ and remote-sensing measurements and modeling that have been used to obtain the solar chemical abundances, and how they compare with stellar and galactic abundances. We then introduce the main features of the solar corona, together with observations and modeling of flares and coronal mass ejections. We conclude by describing two subjects of relevance for the entire solar system and in particular for our planet, the solar wind and the solar irradiance.

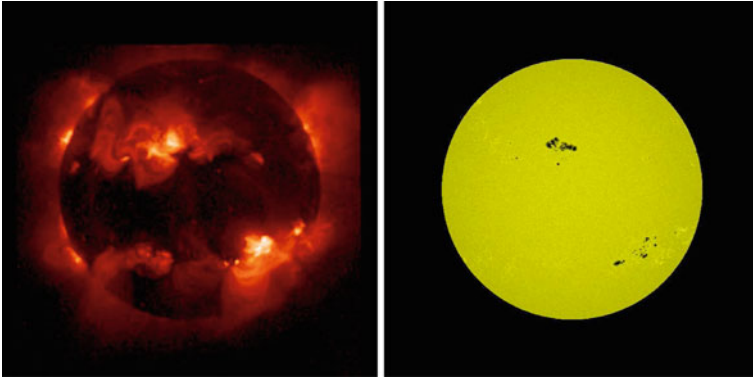
For each subject, we provide a brief historical overview, and then refer the reader to standard textbooks and recent literature. Our knowledge of the Sun as a star has progressed significantly over the last century, however a clear understanding of the physical processes at work in the interior of the Sun, in its atmosphere, and in the solar wind is still incomplete, and further challenges lie ahead. We should be able to understand our Sun, before we can attempt a complete description of other stars and astrophysical objects.

Keywords: Atomic processes; Helioseismology; Solar corona; Sun

1 Introduction

The Sun is a typical middle-aged, low-mass star, but it is the only star which can be studied in great detail, due to its proximity. Viewed in white light, and at low resolution, the Sun might appear as a featureless disk, except for the occasional presence of a few sunspots. However, when viewed at ultraviolet (UV) or X-ray wavelengths, the hot corona provides a spectacular display, as shown in  Fig. 3-1.

With the advent of space-borne instrumentation over the past few decades, some fantastic images and movies have been obtained, together with new insights into the nature of the solar atmosphere. Extreme ultraviolet (EUV) and X-ray observations of the solar atmosphere from the most recent solar missions, such as the Yohkoh, SoHO, TRACE, Stereo, Hinode and SDO satellites, have revealed complex and dynamic structures, which are clearly governed by the Sun's magnetic field. The energy channeled by and stored in solar magnetic fields is released in the solar atmosphere in the form of thermal and kinetic energy, both on small and large scales.



■ Fig. 3-1

The Sun in X-rays (*lhs*) and white light (*rhs*) (Courtesy: Yohkoh)

This energy release can be observed in a number of interesting and dramatic phenomena, ranging from small brightenings and jets to huge solar flares and coronal mass ejections (CMEs), seen as intense emission in a broad wavelength range: from X-rays down to radio frequencies.

Stimulated by these recent observations, research in solar physics, both observational and theoretical, has progressed at a rapid pace. Major progress has been made in recent years on some fundamental issues: the main dynamo responsible for the magnetic field and sunspots is believed to be based in the tachocline (the thin layer just below the convective zone); magnetic fields emerge through the photosphere over a wide range of scales and create many small regions of opposite polarity magnetic flux in the quiet Sun (forming what is termed *the magnetic carpet*); the corona has a highly complex topology and may well be heated by the coronal tectonics mechanism; magnetic flux emergence into the solar atmosphere can trigger large-scale explosive events (solar flares and coronal mass ejections); the solar wind can be highly variable, with an as yet unknown acceleration mechanism. Significant advances have been made in theory and modeling but further work is needed. For example, it is only in the last few years that the process of magnetic reconnection (by which magnetic field lines change connectivity and energy can be released) has been studied in 3D. It turns out that 3D reconnection is completely different from 2D reconnection. New insights have been provided concerning the impact of the Sun on the heliosphere and near-Earth environment, using in situ measurements from SoHO, Cluster and Ulysses.

The Sun is a yellow dwarf star with a mass of 2×10^{30} kg, a diameter of 1,392,000 km (just over 109 times the diameter of the Earth) and is 149.6 million km from the Earth (termed an Astronomical Unit, 1 AU). The Sun is 4.5 billion years old, almost half way through its life cycle. It will eventually grow into a red giant. The color of a star is determined by the surface temperature, most stars being between 3,000 K (red) and 40,000 K (blue). The Sun appears yellow because it has a surface temperature of around 6,000 K, but the temperature of the core is 15 million K. The nearest star to the Sun, Proxima Centauri, is 4 light-years away (63,240 AU). Sirius, often called *the Dog star*, is the brightest star in the sky. It is a blue star with a surface temperature of 10,000 K and is 9 light-years away. It has a very faint and dense companion, a white dwarf known as Sirius B, which has the mass of the Sun, but is only the size of the Earth. The first spatially resolved image which astronomers were able to obtain of another star was

of Betelgeuse, which is a red supergiant in the far-off constellation of Orion. Its mass is about ten times that of the Sun, but it is 50,000 times brighter than the Sun. If it were in the same position as the Sun, it would extend almost out to the orbit of Jupiter. When the Sun is about 7 billion years old it will slowly start to change and will become bigger and cooler. By the time it is 10 billion years old it will have changed into a red giant and its atmosphere will stretch out to near where the Earth is today. The Sun is just one star in a spiral galaxy, the Milky Way, which is 75,000 light years across and contains 100 billion stars.

The Sun provides a unique physical laboratory with extreme conditions. The level of detail which we can now observe is almost overwhelming, from the visible to the X-rays. It is indeed a challenge to understand and model the underlying physical processes. Studies of the Sun have helped to deepen our understanding of processes in other stars. Interdisciplinary research in the *solar-stellar connection* reveals insights into properties as diverse as astrophysical dynamo action, starspots, flares, prominences, and variations in elemental abundances. Stellar coronae seem to be a general phenomena rather than an exceptional occurrence. It has become clear that all late-type stars share the same basic coronal characteristics: the presence of thermal plasma at several million degrees, magnetic confinement, the presence of flares, etc. The basic plasma physics governing coronae is also probably very similar in all of these stars. Otherwise similar stars may have very different levels of stellar activity, UV and X-ray luminosity, and starspots. This is believed to be due to very different rotation speeds which probably determine how effectively the stellar magnetic dynamo generates magnetic fields. One might then expect that differences between the coronae of late-type stars are just due to rather different regimes of the dynamo at work inside the stars, while the basic plasma phenomena occurring in the outer magnetized coronal plasma are probably very similar. EUV and X-ray spectroscopic observations, such as those from EUVE, FUSE, XMM, and Chandra have progressed our knowledge of stellar atmospheres enormously.

Listed below are some of the key questions confronting us in solar research in the modern era, which also relate to other stars and their stellar systems.

- How are magnetic fields generated and transported in the interior of the Sun?
- What causes the solar activity cycle?
- How is the solar corona heated?
- What is the source of and acceleration mechanism for the coronal mass ejections and the solar wind?
- How does the Sun interact with the Earth's environment?
- What is the relationship between solar radiation and the Earth's climate?

In this chapter it is impossible to cover or do justice to the whole spectrum of solar research. We have therefore chosen to focus primarily on what has been learnt from recent observations of the Sun and measurements of the solar wind. We relate these studies to theoretical models where possible, but have left out much of the detailed theoretical work, for example MHD, dynamo, convection and helioseismology models, which could in themselves comprise a whole book, or indeed several volumes! Instead, we refer the reader to standard texts in these subjects. In this chapter we present a historical perspective and the latest solar observations, techniques, interpretation of these observations linked to theoretical models, concepts and possibilities for future directions in solar research, in the context of the Sun as a star. The emphasis is on the processes by which energy is transferred to the solar atmosphere and then flows from the Sun in the form of radiation and particles. In particular, the spectroscopic diagnostic techniques which

are used to determine the physical parameters of the solar atmosphere are discussed in depth. They provide a powerful tool for probing the nature of the solar atmosphere. Similar techniques are now also being applied to the study of other stars and astrophysical objects.

Several excellent books on the Sun have been written in recent years which feature the latest solar observations and theories (Aschwanden 2006; Foukal 2004; Golub and Pasachoff 2010; Phillips 1992; Phillips et al. 2008; Priest and Forbes 2000; Thomas and Weiss 2008). Earlier books, such as Athay (1976), Jordan (1981), Priest (1982), Zirin (1988), Mariska (1992), and Stix (2002) are timeless. These are highly recommended for further study.

2 Historic Perspective

Chinese astronomers reported seeing sunspots several thousand years ago. They did not have telescopes, but when looking through fog or the smoke from forest fires, they noticed that the Sun had very small, dark dots on it. However, it was not until the invention of the telescope that it was possible to study sunspots in detail. The first telescopic observations of sunspots were made by Galileo and Harriot in 1610. However, Harriot never published his observations of sunspots so his work was not known about until after his death. Galileo noted that sunspots appeared, disappeared, and changed their shapes so he deduced that they could not possibly be planets just passing in front of the Sun. He also noted that the sunspots moved (the Sun's rotation is around 25 days in the equatorial region, but slower toward the polar regions). Galileo was not the first to use a telescope, but he was one of the first to carry out extensive observations of the skies (day and nighttime) with his telescope.

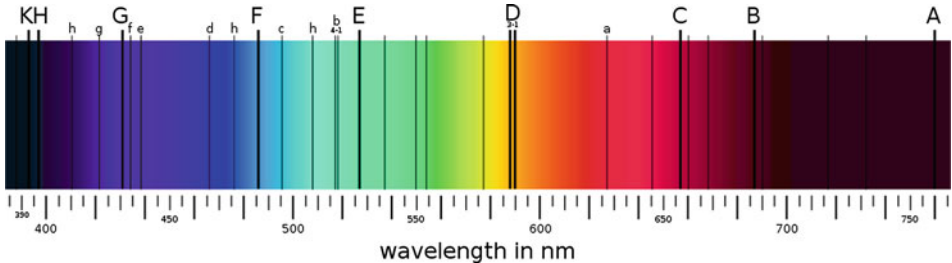
In 1843, Samuel Schwabe discovered a periodic variation in the average number of sunspots seen on the disk of the Sun from year to year. By studying these and other observations, Rudolf Wolf was able to reconstruct the solar cycle back to 1745, eventually pushing these reconstructions back to the earliest observations of sunspots by Galileo and his contemporaries. Cycles as short as 9 years and as long as 14 years have been observed, but the average is 11 years. During the solar minimum, the Sun can go for long periods without any sunspots. During solar maximum, the Sun is very active, with many active regions (and corresponding sunspots) producing solar flares and coronal mass ejections (CMEs). The solar activity cycle can have consequences for the Earth's environment, space weather, and the climate. For example, the period between 1645 and 1715, a time during which very few sunspots were observed coincides with the Little Ice Age. This epoch is now known as the Maunder minimum. However, it should be stressed that recent (during the last three decades) global warming cannot be attributed to changes in the solar cycle.

Sir Isaac Newton was fascinated by the nature of light. It was known that when white light shone through a prism, the colors of the rainbow were produced. However, it was thought that the nature of light was changed within the prism. Newton took another prism and isolated one color. This remained unchanged when passing back through the second prism. His work was published by the Royal Society *Opticks: or a Treatise of the Reflexions, Refractions, Inflexions and Colors of Light*. His interests lay in many fields of science and mathematics, including astronomy, and he built the first reflecting telescope, which used mirrors instead of lenses. Hence he removed the aberrations of colors which were problematic with refracting telescopes. Wollaston in 1802 commented on the *gaps* in the visible spectrum. Thereafter spectroscopic techniques

developed rapidly. In 1811, Fraunhofer produced a spectrum of the Sun, using a spectrograph with a slit, showing an array of dark absorption lines (see ► [Fig. 3-2](#)). The Fraunhofer spectrum is formed by the continuous spectrum of the photosphere, with a characteristic temperature of 6,000 K, superimposed by absorption lines.

The second half of the nineteenth century saw rapid advances in observations of the solar corona during eclipses. Pietro Angelo Secchi SJ (1818–1878) was an Italian astronomer who made fundamental contributions to astrophysics and solar physics. His photographs of the solar corona taken during the eclipse of 1860 in Spain, together with those of Warren de la Rue, provided the first proof that prominences are real effects of the solar atmosphere. ► [Figure 3-3](#) shows the corona in visible light during a total solar eclipse. Because the corona is a million times fainter than the solar disk, we can observe it from the Earth only during a total solar eclipse, or by using a coronagraph which simulates an eclipse by occulting the bright solar disk. The most striking physical property of the corona is its extent, with streamers stretching out for great distances.

We now know that the emission from the solar corona is diverse. The dominant contributions are the so-called K- (Kontinuierlich), F- (Fraunhofer), and E- (Emission) corona. The F-corona is photospheric light reflected by dust particles which are present close to the Sun. The reflection largely keeps the photospheric emission unmodified. The K-corona is photospheric light scattered by coronal electrons. The scattering polarizes the light and broadens the Fraunhofer lines so much that they become undetectable.



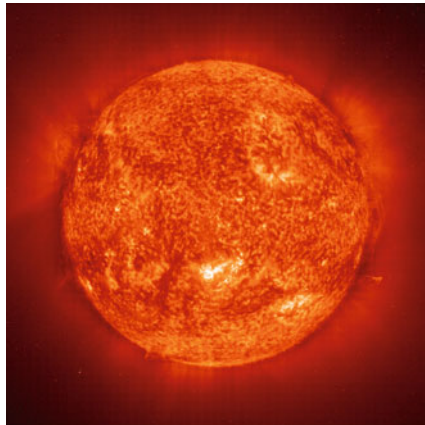
■ Fig. 3-2
The Fraunhofer spectrum



■ Fig. 3-3
White light image of the solar corona during a total solar eclipse (Courtesy of A. Sterling)

Spectroscopy techniques were used to analyze the solar atmosphere during eclipses. Janssen was the first to apply slit spectroscopy during the 1868 eclipse in India, and to observe the chromosphere and solar prominences in emission in the red $H\alpha$ (line C in the Fraunhofer spectrum) line. At the 1868 eclipse, as the photosphere was covered, Young attempted to observe the *flash spectrum* of the chromosphere. He eventually succeeded in observing it at the 1870 eclipse in Spain. Young (1870) realized that these emission lines corresponded to the dark absorption lines seen in the spectrum from the solar disk. The dark Fraunhofer lines in the photospheric spectrum are in fact due to absorption by chromospheric material. Lockyer was the person who actually coined the name chromosphere (color-sphere). Both Janssen and Lockyer discovered a bright yellow line due to an unknown element. This line was close to the sodium doublet (named D_1, D_2), so the new line was named D_3 . Because it was found only on the Sun, the element was named *helium* by Lockyer after the Greek name *Helios* for the Sun. Only later, in 1895, was helium identified on Earth. We now know that helium is the most abundant element after hydrogen, and that helium also produces the strongest emission lines in the extreme-ultraviolet. An image of the Sun in helium emission is shown in ► Fig. 3-4.

The 1868 eclipse was also particularly important because the first of the coronal lines, the green one (5303 \AA), was observed by Rayet observing in Siam (now Thailand). The following year, in 1869, at the US eclipse, Harkness and Young also observed the green coronal emission line, although they incorrectly identified it with a nearby chromospheric line (5317 \AA). It was only during the 1898 eclipse that these two lines were separated by Lockyer. The observation of the green coronal emission line provided a weight of evidence that the corona (as seen during an eclipse) was a characteristic of the Sun rather than the Moon. However, the coronal *green line* and other coronal visible emission lines provided astronomers with a major puzzle. This presented the single most outstanding problem in astrophysics for 50 years. The wavelength of the green line, 1474 (on Kirchhoff's scale, 5303 \AA) did not match the emission spectrum from any known element. This presented a serious challenge to the new science of spectroscopy. In addition, the corona extended much further (by a factor of 100) from the solar disk than seemed



■ Fig. 3-4

The Sun in the helium $\text{Ly}\alpha$ emission (304 \AA) as observed on August 11, 1999, from SoHO EIT (Courtesy of the SoHO consortium)

possible. The name *coronium* was assigned to the unknown element responsible for the coronal green line. Lyot (1939) used his coronagraph to study the coronal spectrum and found that the green line was very wide (0.9 Å). A major advance came in 1939 through the meticulous laboratory work of Edlén, who following a suggestion by Grotrian, assigned the strange coronal lines to highly ionized stages of iron, calcium and nickel (Edlén 1942). The green line is due to the forbidden transition in the ground configuration of Fe XIV (Fe^{+13}), that is iron with 13 electrons stripped from its outer shells. The *red line* is due to Fe X and the *yellow line* to Ca XV. The identifications of these coronal lines became one of the great triumphs in astrophysics. Temperatures in excess of 1 MK are needed to produce these ion stages, which also explained the large extent of coronal emission.

How energy was produced in stars posed another major puzzle for astronomers. In the nineteenth century, Lord Kelvin thought that the Sun might be powered by gravity. He suggested that if the Sun were collapsing and gradually getting smaller, the gravitational potential energy might be transferred into the heat and light the Sun emits. From the size and mass of the Sun, Kelvin calculated that the Sun could not be more than 25 million years old. The problem with Kelvin's theory was that scientists had already shown that the Earth was at least 400 million years old (we now know it to be a lot older than that). It hardly seemed likely that the Earth could be a lot older than the Sun. Bethe finally solved the mystery of the Sun's energy in the 1930s, but he based his ideas on the work Einstein had done many years before, when he proposed his famous formula $E = mc^2$, where E is energy, m is mass, and c is the speed of light. Bethe suggested that if the Sun could change just a small amount of its mass into energy, there would be enough energy to power the Sun for billions of years. He suggested that a way of doing this was to convert hydrogen to helium. Bethe won a Nobel Prize in 1967 for his *discovery concerning the energy production in stars*. Actually, Bethe's work was also based on pioneering work by a famous English astronomer, Eddington who said in 1926 "It is not too much to hope that in the not too distant future we shall be able to understand so simple a thing as a star".

At its core, the Sun is dense and hot enough (15 MK) to allow thermonuclear reactions to occur and to convert hydrogen into helium. The energy generated at the core is transferred by radiation out to about 0.7 of the solar radius. From this point, the energy is transported to the photosphere through convection. Through the combined actions of convection and differential rotation, the Sun operates as an alpha-omega dynamo where the magnetic field is constantly generated and sustained. The global solar magnetic field, driven by the regenerative dynamo, oscillates between the poloidal and toroidal components every 11 years, creating what is known as the solar cycle. The solar tachocline, at the bottom of the convection zone, is now known to play a key role in the dynamo process (Hughes et al. 2007).

The structure of the solar atmosphere can be crudely divided into different layers of gas and plasma threaded by magnetic fields. When viewed in white light the Sun appears to have a sharp edge, suggesting the existence of a solar surface. However, the Sun is a gas (plasma) and this apparent surface arises because the opacity of the solar plasma rises rapidly with temperature. The thin layer over which this rapid change from transparent to opaque takes place is termed the photosphere and has a characteristic temperature of 6,000 K. In fact, the base of the photosphere depends on the wavelength of the radiation. A common but somewhat arbitrary definition is where the optical depth is unity for radiation of wavelength 5,000 Å (Phillips et al. 2008).

Lying above the photosphere, the chromosphere has a typical thickness of 2,000 km and temperatures up to around 20,000 K (see ► Fig. 3-5). The photosphere and lower chromosphere are largely made up of neutral H atoms. In the upper chromosphere, the H atoms begin to be ionized, and large numbers of free protons and electrons may be found together with the

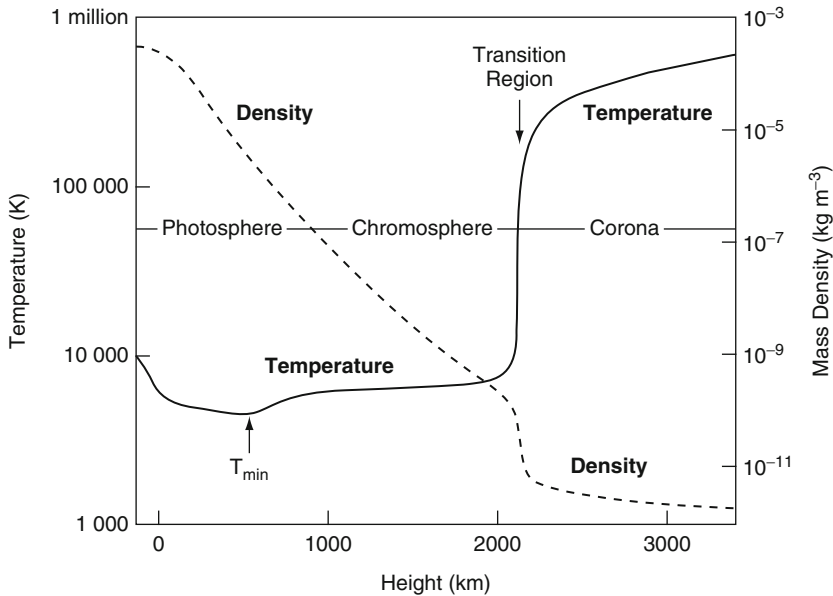
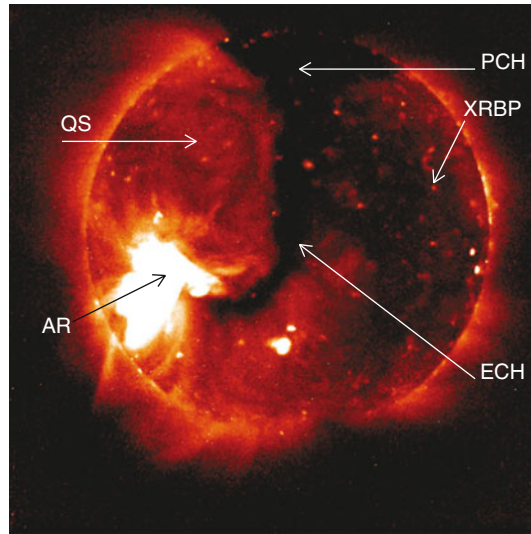


Fig. 3-5
Layers of the solar atmosphere

neutral H and He atoms. The transition region (TR) is a very thin region above the chromosphere, in which the temperature drastically increases to about 1 MK. The corona is the very hot, extended outer layer of the solar atmosphere. At these high temperatures, the coronal plasma emits radiation strongly in the EUV and X-rays. The unexpected and extraordinarily high temperature of the corona has been one of the main focuses for solar research over the past seven decades. The heating mechanism for the solar corona remains a hotly debated topic. A great deal of progress has been made in this regard, but a full explanation still eludes us.

The resolution of the coronal heating problem is likely to involve a mechanism by which magnetic energy can be dissipated as heat, such as reconnection. The appearance of the solar corona reflects the geometry and strength of the magnetic fields that are produced and emerge continuously from below the photosphere. A magnetic field of strength B exerts a pressure $B^2/8\pi$ on the plasma, which adds to the gas pressure P . The ratio of the gas pressure P to the magnetic pressure is known as the plasma β . While in the photosphere the pressure of the thermal gas dictates the behavior of the magnetic field lines, the opposite is true in the solar corona. Due to the low density, the gas pressure is significantly smaller, so that $\beta \ll 1$, i.e., the magnetic field dominates the dynamics of the tenuous coronal plasma.

As seen in X-ray emission (► [Fig. 3-6](#)) the corona is made up of three main structures: active regions (ARs), displaying intense X-ray emission; quiet regions, or the quiet Sun (QS), displaying faint, diffuse emission; and coronal holes (CHs) from which practically no coronal emission is detected. CHs are regions where the magnetic field is principally of one polarity. They are often found at the solar poles, but can also appear in other regions. The QS has a mixed-polarity magnetic field and is peppered with small bright points. A characteristic of ARs is the presence of hot loop arches (typically 1–4 MK) mirroring the magnetic field configuration and



■ Fig. 3-6

Image of the solar corona recorded by Yohkoh's Soft X-ray Telescope (SXT) on August 27, 1996, when the "Elephant's Trunk" equatorial coronal hole (Del Zanna and Bromage 1999) passed the meridian. QS quiet Sun regions, AR active region, ECH equatorial coronal hole, PCH polar coronal hole, XRBP X-ray bright point

extending up to 10^5 km from the photosphere, joining regions with opposite magnetic field polarity. One example is shown in ● Fig. 3-7. Sunspots, dark regions of intense magnetic field, are frequently observed (in white light) in mature ARs (see ● Fig. 3-8). The most tantalizing aspects of the atmospheric activity such as solar flares, erupting filaments, and coronal mass ejections (CMEs) are often associated with ARs which have a strong magnetic field concentration. These highly energetic manifestations of the solar activity play an important role in understanding, and ultimately predicting, space weather.

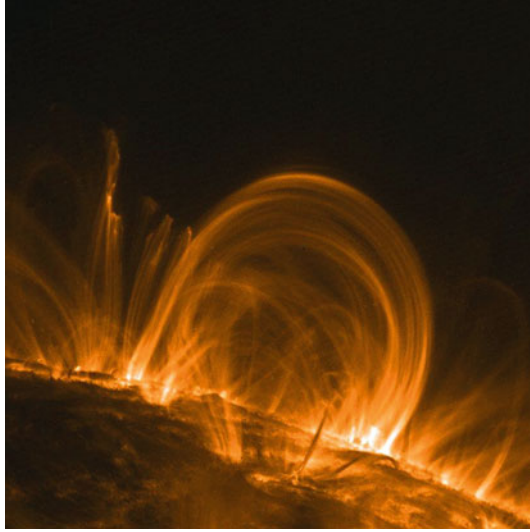
3 Observing the Solar Atmosphere from Space

The solar corona emits radiation over a very broad range of wavelengths, but because it is so hot, the UV and X-ray emission is very strong. Observing the Sun in spectral regions outside the visible, infrared, and radio wavelength ranges is only possible above the Earth's atmosphere.

3.1 Early Space Missions

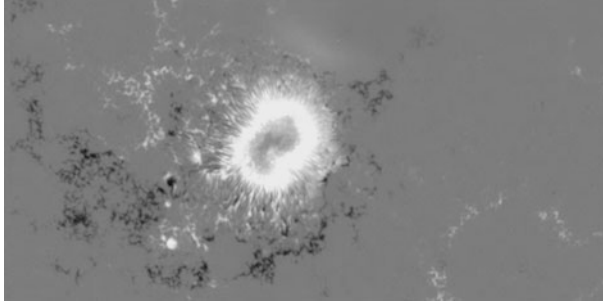
The 1970s saw a rapid development in techniques for high spatial resolution imaging of the X-ray corona. A number of rockets observing the corona with spectrometers in the X-rays and EUV were launched.

In May 1973, the first NASA space station, Skylab (see ● Fig. 3-9), was launched. The Apollo Telescope Mount (ATM) on Skylab carried several state-of-the-art solar instruments, both



■ Fig. 3-7

Image of coronal loops observed at the solar limb in the EUV by the Transition Region and Coronal Explorer (Courtesy of the TRACE consortium)



■ Fig. 3-8

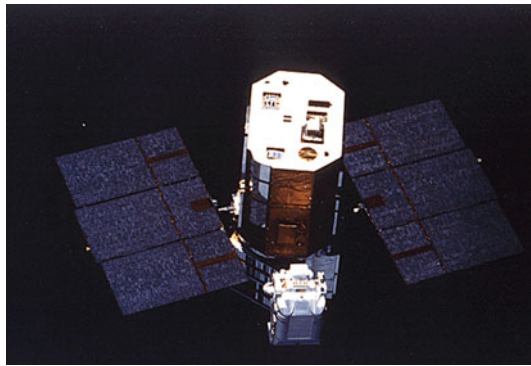
Hinode/SOT image of the magnetic field of a sunspot. The signal is proportional to the line-of-sight magnetic field density in the photosphere

imaging and spectroscopic, observing from the UV to the X-rays. Naval Research Laboratory's S082A instrument (often affectionately called the “overlap-ogram”) on the ATM had an exceptionally good spatial resolution (better than $2''$) and covered a wavelength range 170–630 Å. Skylab/ATM was a highly successful mission, followed by several data analysis workshops, which brought together observers and theoreticians. The quality and quantity of data (mainly on photographic plates) provided a major impetus for space-based solar physics. Three topics were chosen for study at the Skylab workshops: Coronal Hole and High Speed Wind Streams (Zirker 1977), Solar Flares (Sturrock 1980) and Solar Active Regions (Orrall 1981). These volumes are an excellent starting point as a foundation for research in these fields.



■ Fig. 3-9

The Skylab satellite and Apollo Telescope Mount (courtesy of NASA)

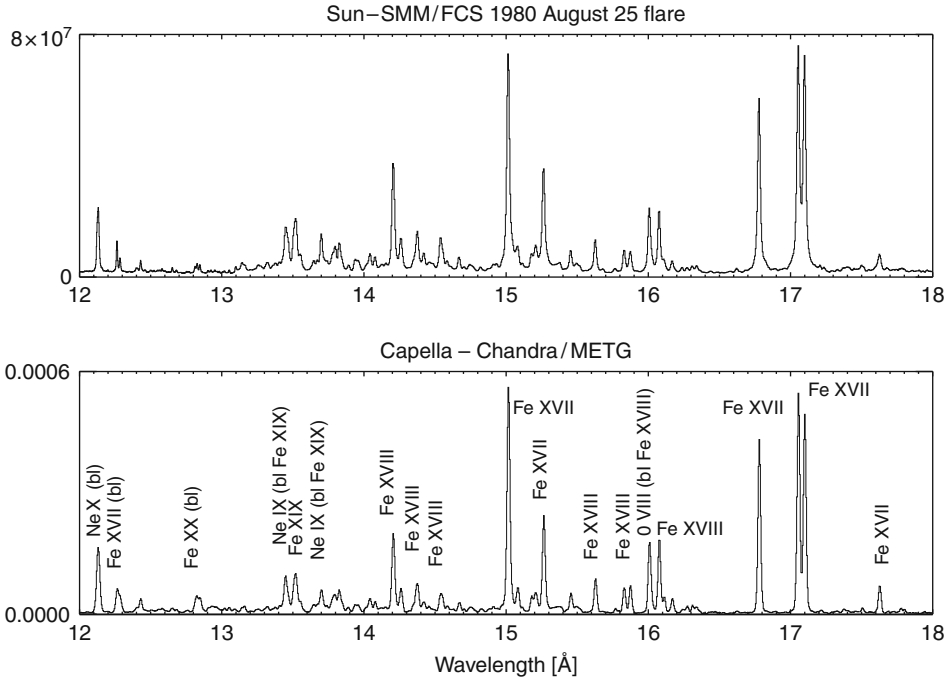


■ Fig. 3-10

The SMM satellite being repaired in space (courtesy of NASA)

February 1980 saw the launch of a new solar space satellite, the Solar Maximum Mission, SMM (cf. ► [Fig. 3-10](#)), dedicated to the study of the active Sun, active regions, and solar flares. SMM remained operational until December 1989, with an in-orbit repair by the NASA Space Shuttle in April 1984. SMM carried instruments to record gamma-rays, X-ray emission (both hard and soft), together with a UV spectrometer, a coronagraph, and also an irradiance monitor (ranging from UV to infrared). Preliminary results from SMM were published in a special volume of *ApL*, 1981, (244) and further results were published in a special NASA publication NASA, CP-2439, *Energetic Phenomena on the Sun* (1986).

The SMM X-ray polychromator (XRP) Flat and Bent Crystal Spectrometers (FCS and BCS) instruments produced excellent X-ray spectra of the solar corona, which remain unsurpassed even nowadays, three decades later. ► [Figure 3-11](#) shows an XRP/FCS spectrum of a flare, which is remarkably similar to the X-ray spectrum of active stars such as Capella, also shown in the Figure. These X-ray spectra are dominated by Fe xvii and Fe xviii lines in the 12–18 Å range. The XRP/FCS had limited spatial resolution. These data allowed measurements of the



■ Fig. 3-11

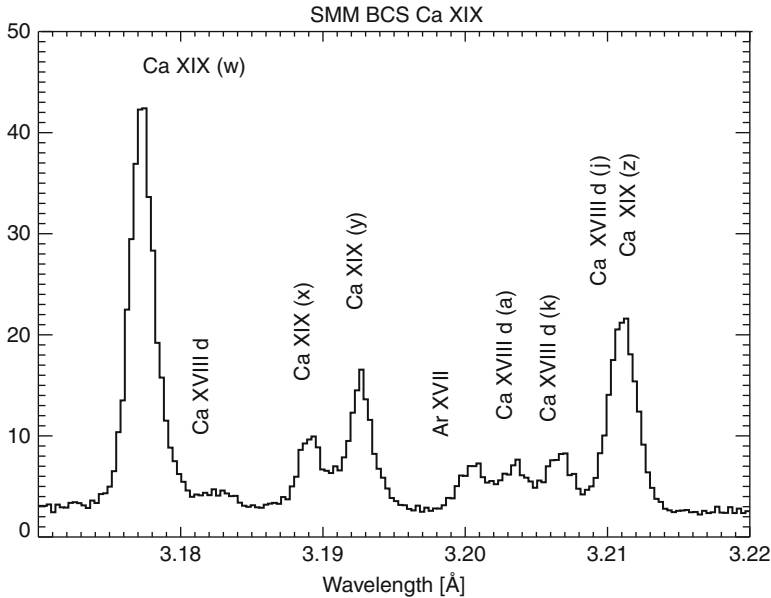
Solar spectrum of a flare from the SMM XRP/FCS instrument (*above*), compared to a Chandra METG spectrum of Capella (Adapted from Del Zanna 2002). Notice the similarities in the spectra

temperature distribution and electron density of solar active regions and flares together with measurements of relative coronal abundances of various elements. The XRP/BCS, which was full disk, observed X-ray line complexes of high temperature (in excess of 10 MK) coronal lines: Fe xxvi, Fe xxv, Ca xix, and S xv. The observed spectral range allowed a wide range of plasma diagnostics to be applied. The He-like ions (Fe xxv, Ca xix and S xv) allowed for example measurements of electron densities and temperatures (Gabriel and Jordan 1969). A sample XRP/BCS spectrum of a flare is given in ● Fig. 3-12. The resonance w line is the strongest transition, followed by the intercombination z and forbidden x, y. The satellite lines (i.e., those formed by dielectronic recombination) are indicated with a 'd'. The presence of satellite lines allowed some important diagnostic measurements (see, e.g., Phillips et al. 2008).

Other solar X-ray missions included the P78-1 satellite launched in 1979 and the Japanese satellite, Hinotori (meaning *firebird*) launched in 1981. A good overview of results from SMM and Hinotori can be found in Culhane and Jordan (1991).

3.2 Yohkoh

Yohkoh (Japanese for *sunbeam*) was a very successful joint Japanese, UK, NASA satellite launched in August 1991, which operated successfully for a decade until 2001. The instruments on Yohkoh were able to observe the Sun in X-ray and Gamma Ray emission. The spacecraft



■ Fig. 3-12

Solar spectrum of a flare from the SMM BCS instrument, with the resonance (w), forbidden (x,y) and intercombination (z) lines from the He-like Ca. Lines denoted with "d" are formed by dielectronic recombination

carried four instruments: the Soft X-Ray Telescope (SXT), the Bragg Crystal Spectrometer (BCS), the Wide Band Spectrometer (WBS), and the Hard X-Ray Telescope (HXT). The SXT instrument was widely used to study the plasma properties of active regions and solar flares. The SXT was a grazing-incidence X-ray telescope capable of imaging the Sun and resolving features down to $2.5''$ in size (see ● Fig. 3-6). Yohkoh/SXT could be used to derive some information about the temperature and density of the plasma by comparing images recorded using different filters. The BCS observed the X-ray H- and He-like line complexes observed by SMM/BCS. Early results from YOHKOH were presented in a special issue of the Publications of the Astronomical Society of Japan, PASJ, 1992, Vol. 4. There have been many conferences and workshops dedicated to results from YOHKOH.

3.3 Ulysses

Most satellites have orbits within the ecliptic. The Ulysses mission was the first and only space probe dedicated to the exploration of the heliosphere out of the ecliptic plane. Indeed, much of our knowledge about the solar wind comes from this unique mission. Ulysses was launched in October 1990 and, after an encounter with Jupiter, started on its journey out of the ecliptic.

3.4 The Solar and Heliospheric Observatory – SoHO

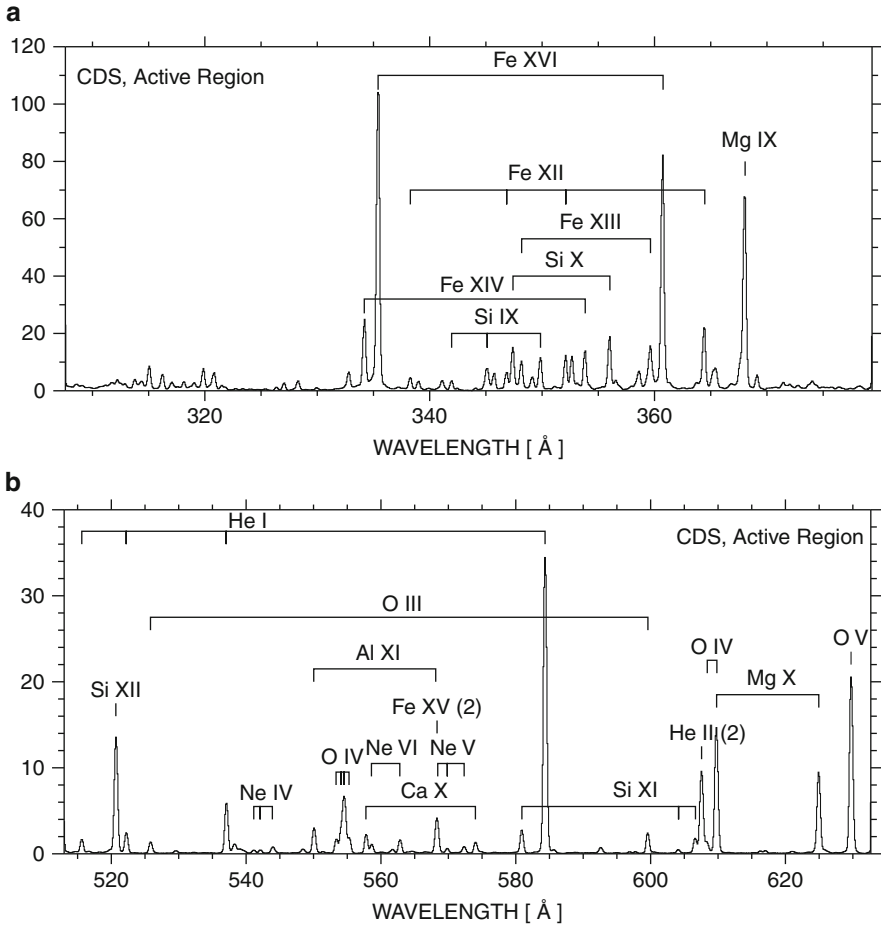
The Solar and Heliospheric Observatory is the most comprehensive and successful solar observatory ever built. SoHO is a joint NASA and ESA mission, which was launched in December 1995 and is still operational. The scientific objectives of SoHO are to study the solar interior, using helioseismology; to investigate the heating mechanism for the solar corona, together with the origin and acceleration mechanism of the solar wind. The SoHO satellite has an unusual orbit, it orbits the Sun at the first Lagrangian (L1) point, allowing the instruments an uninterrupted 24 h view of the Sun.

SoHO carries 12 instruments, including 4 EUV/UV instruments for studying the solar atmosphere. The Coronal Diagnostic Spectrometer (CDS) is a UK-led instrument designed at the Rutherford Appleton Laboratory (RAL). CDS has been used to study the relationship of the transition region and coronal emission – it is designed to determine the physical parameters of the solar atmosphere. The instrument measures EUV emission lines from the Sun's atmosphere in the wavelength range 150–800 Å. The instrument comprises two spectrometers: the Normal Incidence Spectrometer (NIS) and the Grazing Incidence Spectrometer (GIS). The wavelength range covered by the two detectors contains many emission lines. The NIS detector can receive radiation in the two wavelength bands, NIS-1, 308–381 Å and NIS-2, 513–633 Å (see ● Fig. 3-13). To construct monochromatic images (rasters) using the emission lines measured by NIS, a scan mirror is moved across the solar region with a chosen slit (2'' or 4''). One example is shown in ● Fig. 3-14. The CDS instrument has a maximum Field of View (FOV) of $4 \times 4 \text{ arcmin}^2$ with a spatial resolution of about 5''.

The Solar Ultraviolet Measurement of Emitted Radiation (SUMER) is a high-resolution (spatial and spectral) instrument covering the wavelength range 500–1,600 Å in first and second order (see ● Fig. 3-15). SUMER can resolve fine structures, with a spatial resolution of about 2''. As in the case of the CDS, SUMER has been able to scan solar regions to obtain monochromatic images in selected spectral lines (see, e.g., ● Fig. 3-16). SUMER can also measure Doppler motions (flows) with an accuracy better than 2 km/s and nonthermal line broadening in the chromosphere and transition region.

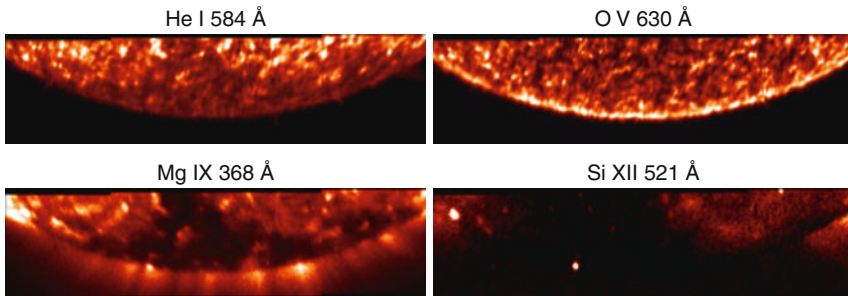
The Extreme-ultraviolet Imaging Telescope (EIT) has provided superb images and movies of the solar atmosphere with a spatial resolution of 2.6'' per pixel. The instrument uses multilayer technology, whereby the reflecting surfaces are coated with layers of various materials. High reflectivities in the EUV are achieved, but images are formed over passbands which are typically 10–50 Å wide. The EIT has four passbands centered on the following wavelengths: 171, 195, 284, 304 Å (see ● Fig. 3-17). In most conditions, the first band is dominated by Fe IX, Fe X formed at 1×10^6 K. The second band is dominated in active regions by Fe XII, formed around 1.6×10^6 K, but has significant contributions from much lower and much hotter lines. The third band is dominated in active regions by Fe XV at 2×10^6 K, but otherwise by transition-region lines. The last band is dominated by the He II chromospheric line 304 Å, but off-limb it is dominated by the coronal line Si XII.

The Ultraviolet Coronagraph Spectrometer (UVCS) can observe the solar corona from its base out to $10 R_{\odot}$. The UVCS spectroscopic diagnostics are based on the measurement of the intensities and spectral line profiles of resonantly and Thompson scattered H I Lyman- α , the collisionally excited and resonantly scattered Li-like resonance lines of O VI (1032 and 1037 Å); it can also detect other coronal lines such as Si XII (499 and 521 Å) and Fe XII (1242 Å) and is rich



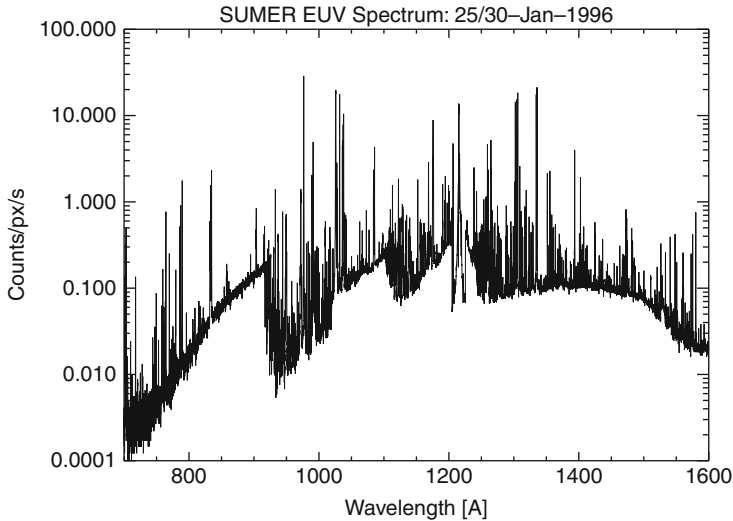
■ Fig. 3-13

SoHO CDS NIS-1 and NIS-2 spectra of an active region (Courtesy of P.R.Young)

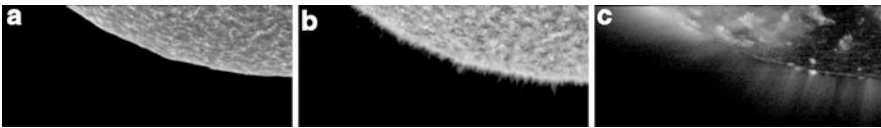


■ Fig. 3-14

Monochromatic images of the south pole on August 11, 1996, from SoHO CDS, as part of the Whole Sun Month campaign (Del Zanna 1999). He I is formed in the chromosphere, while O V in the middle transition region (0.5 MK, notice the limb-brightening), Mg IX in the upper transition region (1 MK), and Si XII in the hot (2 MK) corona



■ Fig. 3-15
A SoHO/SUMER spectral atlas (Courtesy of SoHO/SUMER)



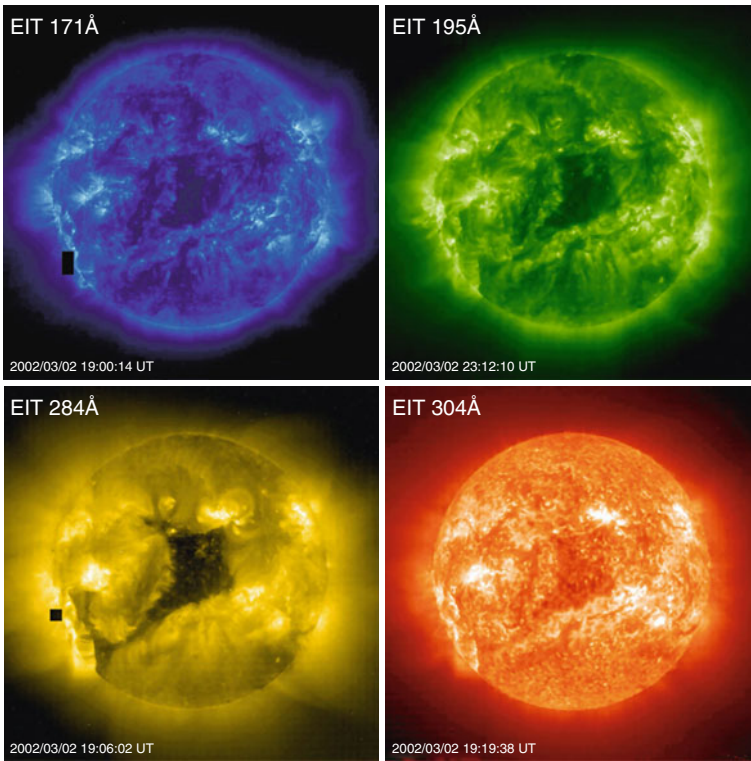
■ Fig. 3-16
A portion of the SE limb of the Sun observed by SoHO/SUMER in 1996 during the Whole Sun Month campaign. The monochromatic images are in C I, O V, and Mg X (a, b, and c respectively). C I is a photospheric line, the O V clearly shows the spicules at the limb, while the hotter (1 MK) Mg X shows the presence of a south polar coronal hole, dark in the image (Courtesy of SoHO/SUMER)

in diagnostic possibilities. ▶ [Figure 3-18](#) shows an O VI monochromatic image of the extended corona obtained by multiple UVCS slit positions around the Sun.

The Large Angle Spectrometric Coronagraph (LASCO) is a wide-field, white-light coronagraph which has provided images of the solar corona from 1.1 to 30 R_{\odot} with different fields of view (C1, C2, C3) and in several wavelength bands in the visible 5,300–6,000 Å. LASCO has provided fascinating observations of streamers, coronal mass ejections (CMEs), and many other large-scale dynamic phenomena. One image is shown in ▶ [Fig. 3-19](#).

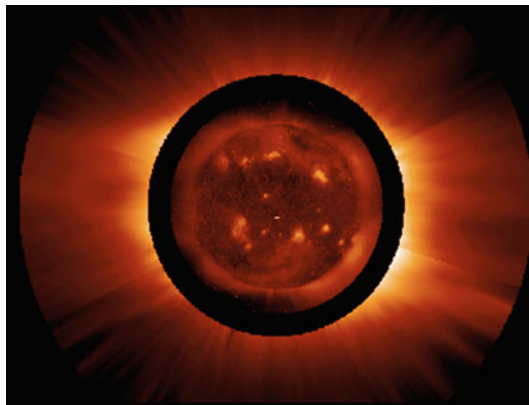
In addition, SoHO carries an instrument for studying the photosphere: measuring the photospheric magnetic field and carrying out helioseismology studies. The Michelson Doppler Imager (MDI) uses a pair of tunable Michelson interferometers to record filtergrams. Polarizers are inserted to measure the line of sight, i.e., the longitudinal component of the magnetic field from the Zeeman splitting of the Ni I 6768 Å line with higher resolution (in a central region) or lower resolution (on the full Sun, cf. ▶ [Fig. 3-20](#) left).

First results from SoHO were published in a special issue of *Solar Physics* in 1997. There have been many SoHO workshops, focusing on particular topics, during which results from different



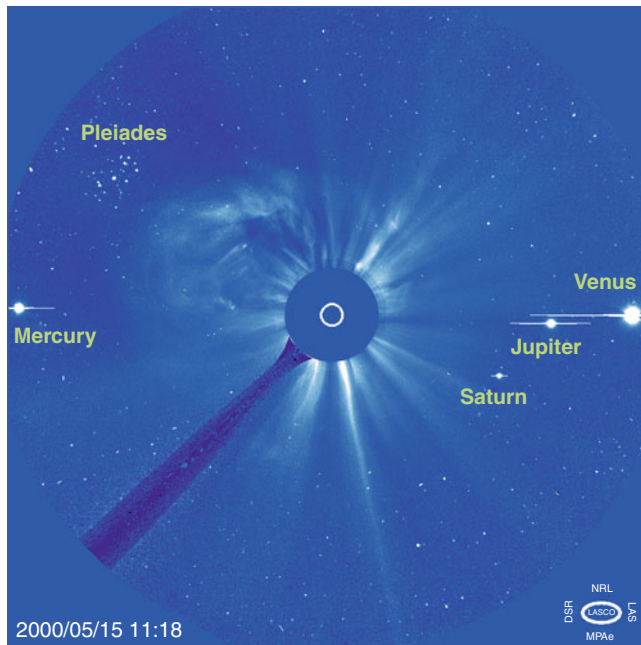
■ Fig. 3-17

SoHO EIT images in different filters (Courtesy of SoHO/EIT consortium)



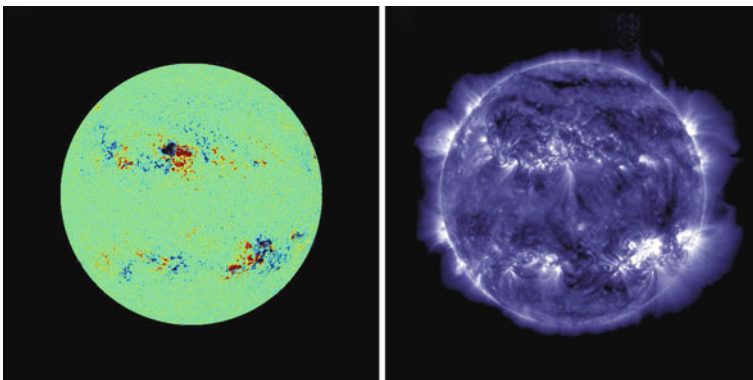
■ Fig. 3-18

Composite of O VI monochromatic image observed from SoHO UVCS. The inset shows a SoHO EIT image (Courtesy of SoHO/UVCS consortium)



■ Fig. 3-19

A SoHO LASCO C3 image, with various planets in the field of view. The white circle in the *middle* of the image indicates the size of the Sun, while the dark feature pointing toward the lower *left corner* is the shadow of the pylon holding the occulter of the coronagraph (Image courtesy of the SoHO/LASCO consortium and Steele Hill (NASA GSFC))



■ Fig. 3-20

Left: line-of-sight magnetogram from SoHO MDI; *right:* mosaic of coronal TRACE 171 Å images from the same day. Notice the obvious correlation between regions of strong magnetic fields and strong intensities in the corona (Courtesy of the SoHO and TRACE consortia)

instruments have been presented and published, starting with the First SoHO workshop held in Annapolis, Maryland, in 1992. One of the particularly successful SoHO workshops was held at St Andrew's University, Scotland in 2004.

3.5 Transition Region and Coronal Explorer (TRACE)

The Transition Region and Coronal Explorer (TRACE) is a very successful NASA small-explorer mission launched in April 2001. The instrument comprises a Cassegrain telescope to observe the solar corona in three EUV wavelength bands (similar to the SoHO/EIT ones), centered around: 171 Å; 195 Å; 284 Å. It also images the H I Lyman-alpha (1,216 Å), C IV (1,550 Å) and UV continuum (1,600 Å), and white light (5,000 Å). TRACE's unprecedented high spatial resolution (of approximately 1'', see [Fig. 3-20](#), right) and cadence (down to 10 s) render this instrument suitable for observing the detailed dynamics of the solar corona.

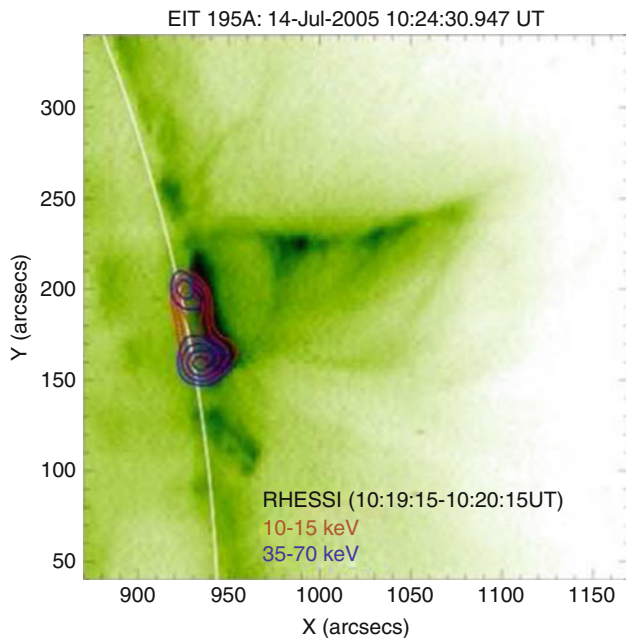
3.6 Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)

The Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) is a NASA mission launched in February 2002 designed to study the particle acceleration and energy release in solar flares. Since its launch, the mission has been providing ground-breaking imaging and spectroscopy measurements from soft X-rays to gamma rays. RHESSI's uniquely high sensitivity at energies 3–15 keV enables studies of large flares, but also small-scale thermal and nonthermal emissions to much lower levels than ever before, from features such as micro-flares.

Obtaining images of high-energy emission from solar flares is extremely difficult. Sophisticated software has been developed to enable X-ray images to be reconstructed from the RHESSI data (see, e.g., [Fig. 3-21](#)). RHESSI also provides high spectral resolution. Spectral fitting techniques enable measurements of the physical parameters (temperature and emission measure) describing the hottest part of the thermal plasma. RHESSI observes thermal free-free and free-bound continuum together with emission lines emitted in the X-ray range 5–20 keV by flare plasmas with temperatures from approximately 7 MK to tens of MK. Furthermore, nonthermal plasma emission can be characterized by fitting models such as *bremsstrahlung* from a power-law mean electron distribution. An example of a RHESSI spectrum acquired during a small pre-flare brightening is shown in [Fig. 3-22](#) (Chifor 2008). The ability to measure and image the Fe line complexes provides an additional powerful plasma diagnostic, needed to help separate the thermal and nonthermal continuum and obtain accurate energy estimates.

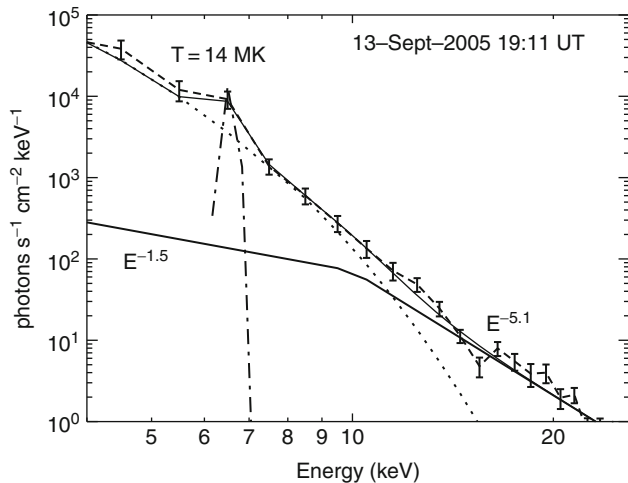
3.7 CORONAS Missions

CORONAS-F, launched in 2001, carried a suite of 15 instruments. It provided XUV monochromatic full-Sun imaging and XUV spectroscopy with the SPIRIT and RESIK suite of instruments. One of the SPIRIT instruments provided images in the Mg XII doublet, formed at a temperature of about 8 MK. A new class of highly dynamic plasma structures were discovered with various characteristic sizes and lifetimes, which still await an explanation (Zhitnik et al. 2003).



■ Fig. 3-21

RHESSI image (red and blue contours), superimposed to a SoHO EIT image during a solar flare at the limb (Courtesy of S. Krucker and H. Hudson)



■ Fig. 3-22

RHESSI spectrum (dashed curve, with error bars) of a pre-flare brightening. The thermal continuum was modeled with one isothermal component ($T = 14$ MK, dotted curve). One Gaussian function was added to fit the Fe line feature observed ~ 6.7 keV (dashed-dot curve). The high-energy tail of the spectrum was fitted with a power-law component $E^{-5.1}$ (dashed dot-dot-dot) breaking to $E^{-1.5}$ at energies below 10.5 keV (Courtesy of C. Chifor)

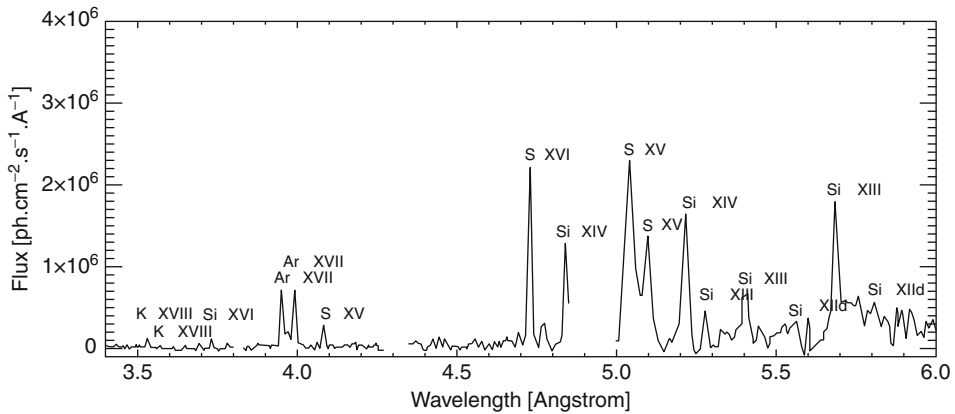


Fig. 3-23
A CORONAS-F RESIK X-ray spectrum of a solar flare (Courtesy of C. Chifor)

The RESIK wavelength range (see ► [Fig. 3-23](#)) is of particular interest for various reasons, one being that line-to-continuum measurements can be used for absolute abundance determinations (see ► [Sect. 7](#)).

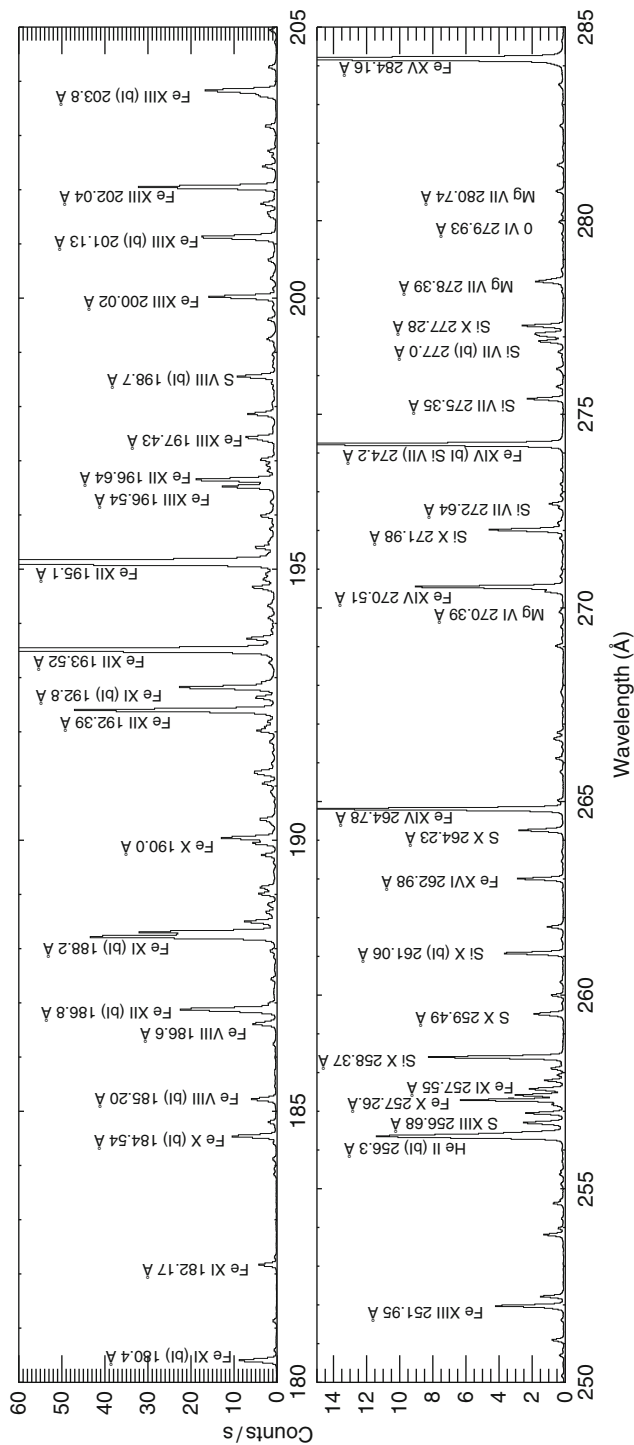
Full-Sun imaging and XUV spectroscopy have also recently been provided by CORONAS-Photon (launched in 2009) with the THESIS and SPHINX suite of instruments.

3.8 Hinode

On September 23, 2006, a satellite called Hinode was successfully launched. Hinode (Japanese for *sunrise*) is a highly sophisticated satellite equipped with three advanced solar telescopes: the EUV Imaging Spectrometer (EIS), the X-Ray Telescope (XRT), and the Solar Optical Telescope (SOT). Positioned in a Sun-synchronous orbit, the mission provides continuous monitoring of the solar activity without interruption for 8 months a year. The instruments mounted on Hinode are used as an integrated observatory aimed at understanding the magnetic connectivity between the photosphere and the corona, with particular emphasis on the phenomena of solar activity, including structure, dynamics, plasma heating, and transient events.

The Extreme-ultraviolet Imaging Spectrometer (EIS) is a UK (Mullard Space Science Laboratory) led instrument, and the first EUV solar spectrometer capable of obtaining high spectral resolution data with both high spatial and temporal resolution. The EIS wavelength bands, 170–211 Å and 246–292 Å, cover a wide range of temperatures, from chromospheric to flare temperatures ($\log T \text{ (MK)} = 4.7 - 7.3$) with a spatial resolution of about $3''$. The high spectral resolution allows velocity measurements of a few km/s. High spectral resolution images (maximum FOV: $6 \times 8.5 \text{ arcmin}^2$) can be obtained by rastering with one of the narrow slits ($1''$ or $2''$). In addition, monochromatic images at a high cadence can be obtained using a slot ($40''$ or $266''$).

Figure 3-24 shows a sample average spectrum obtained with EIS for an active region (Young et al. 2007). EIS contributes to achieving the goals of the Hinode mission by providing the capability of measuring the physical properties of plasma, such as its temperature and

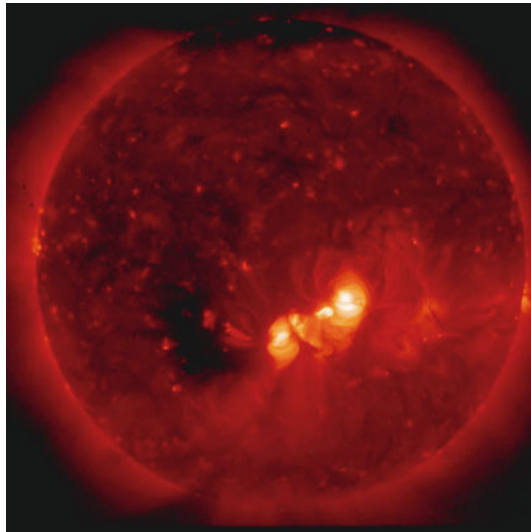


■ Fig. 3-24
Hinode EIS spectrum of an active region

density, together with bulk flow velocities (Doppler shifts), and nonthermal velocities (from line widths). The ten times greater effective areas than previous spectrometers, coupled with a higher data rate than was available from previous SoHO instruments (such as CDS) allow for high-cadence studies of transient phenomena to be undertaken.

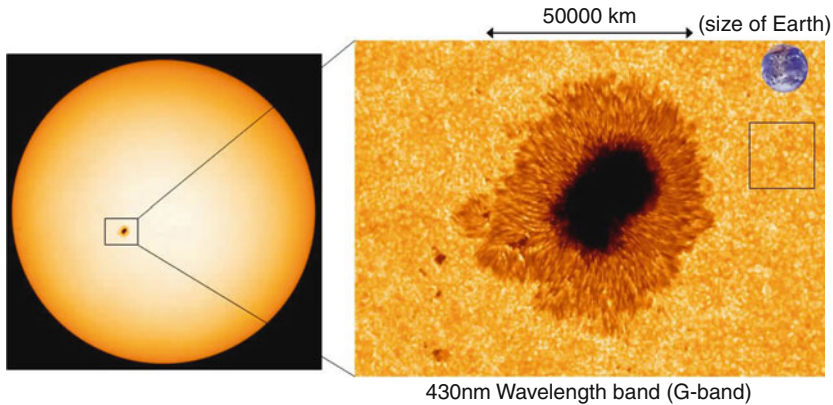
The Hinode/X-ray Telescope was designed to provide an unprecedented combination of spatial and temporal resolution in solar coronal studies. A primary purpose of the Hinode XRT is to observe the generation, transport, and emergence of solar magnetic fields, as well as the ultimate dissipation of magnetic energy in forms such as flares and micro-flares, coronal heating, and coronal mass ejections. High-resolution soft X-ray images reveal magnetic field configuration and its evolution, allowing one to observe the energy build up, storage and release process in the corona for any transient event. One of the unique features of XRT is its wide temperature coverage (1–20 MK) through nine X-ray filters, with $2''$ spatial resolution (see, e.g., [Fig. 3-25](#)). XRT images complement the spectroscopic observations from EIS, by enabling higher cadence and a wider (up to full-Sun) field of view.

The Solar Optical Telescope was designed to provide photometric and vector magnetic images of the photosphere and chromosphere with unprecedented high resolution (see, e.g., [Fig. 3-26](#)) for a long time (ground-based observations can achieve similar resolutions but are limited by *seeing*). The instrument comprises the Optical Telescope Assembly (OTA) consisting of a Gregorian telescope, and the Focal Plane Package (FPP) including a narrow-band (NFI) and a wide-band (BFI) filtergram, plus a Stokes spectro-polarimeter (SP). The time cadence ranges from tens of seconds for both photometric images and vector magnetograms in selected NFI lines to a few hours for a wide-field scan with SP. The maximum field of view is $328 \times 164 \text{ arcsec}^2$. SOT/SP provides the full line profiles in polarized light. The measurements of the Stokes I, Q, U, V signals allows a reconstruction of the vector magnetic field. Initial results from the Hinode satellite have been published in special issues of the PASJ, Science, and A&A journals in 2007 and 2008.



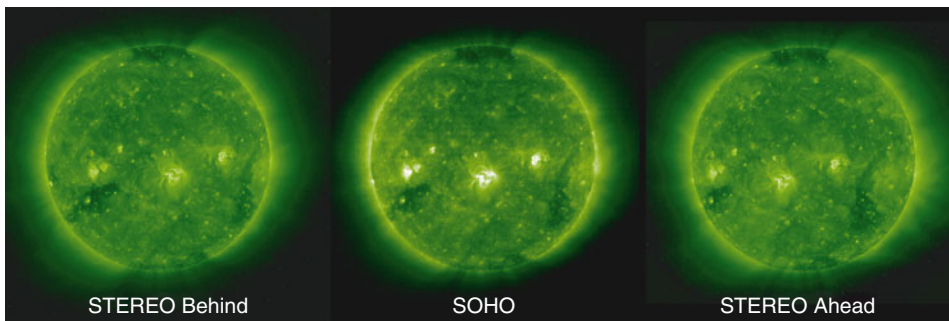
■ Fig. 3-25

Full-Sun X-ray image from Hinode/XRT (Courtesy of the Hinode/XRT team)



■ Fig. 3-26

Image from Hinode/SOT of a sunspot (Courtesy of Hinode/SOT team)



■ Fig. 3-27

EUV images from the STEREO SECCHI behind and ahead spacecraft, together with one from SoHO, showing three different views of the solar corona (Courtesy of STEREO SECCHI consortium)

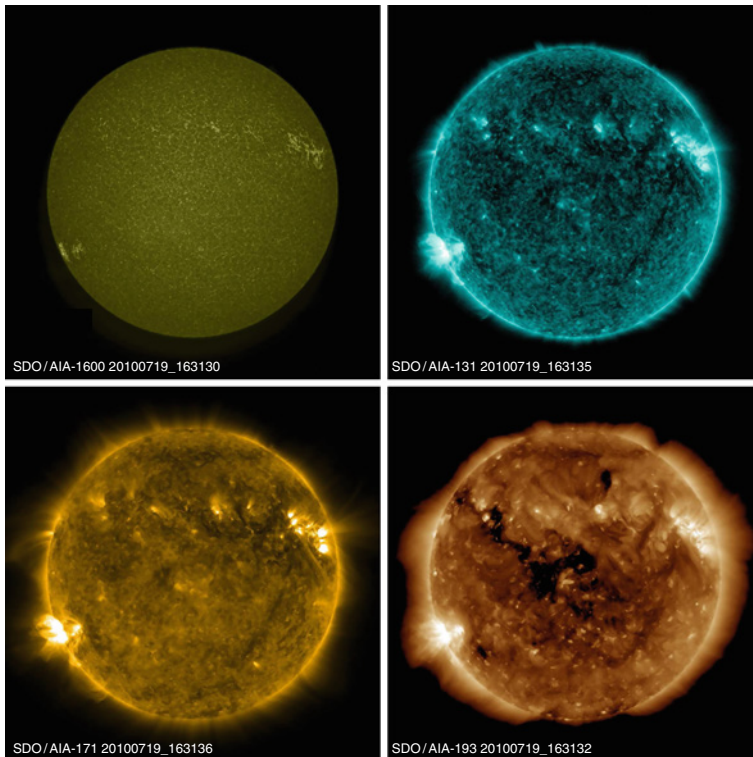
3.9 STEREO

The Solar TERrestrial Relations Observatory (STEREO) is a NASA mission launched in October 2006. The mission comprises the two nearly identical observatories – one Ahead (A) of Earth in its orbit, the other trailing Behind (B), designed to reveal the 3D structure of solar active regions and CMEs. The two spacecraft were launched together and used gravity assistance from the moon to slingshot the spacecraft into a heliocentric orbit. They drift away from Earth at an average rate of about 22.5° per year. Mounted on each of the STEREO spacecraft are four instrument packages: SECCHI, SWAVES, IMPACT, and PLASTIC. Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) comprises four instruments: an Extreme Ultraviolet Imager (EUVI), two white-light coronagraphs (COR1 and COR2), and a wide angle imaging system for viewing the inner heliosphere (HI1 and HI2). EUVI has a field of view out to 1.7 solar radii, and has four channels (similar to the SoHO/EIT ones) (see 🔗 Fig. 3-27). The other instruments on STEREO sample the plasma, solar wind, and CMEs.

3.10 Solar Dynamics Observatory (SDO)

The Solar Dynamics Observatory (SDO) was launched in February 2010 and is providing full-Sun, multiwavelength observations with excellent spatial and temporal resolution. Full-disk magnetic field information is provided from the Helioseismic and Magnetic Imager (HMI). The Atmospheric Imaging Assembly (AIA) provides an unprecedented view of the solar corona, taking images that span at least 1.3 solar diameters in 10 wavelength ranges almost simultaneously, at a resolution of approx $1''$ and a cadence of about 10 s. One example set of images in four passbands is given in [Fig. 3-28](#). Combined with spectroscopic data from existing instruments, the SDO observations promise to advance our knowledge of both small-scale and large-scale solar activity.

The SDO EVE instrument includes two spectrographs and multiple photometers providing the solar EUV irradiance with an unprecedented wavelength range ($1\text{--}1,220\text{ \AA}$) and temporal resolution (10 s). The EVE spectra are from the Multiple EUV Grating Spectrographs (MEGS) and have about 1 \AA spectral resolution. The MEGS A channel is a grazing incidence spectrograph for the $50\text{--}380\text{ \AA}$ range, and the MEGS B channel is a double-pass normal incidence



■ Fig. 3-28

SDO/AIA images in the 1,600, 131, 171 and 193 \AA passbands. In quiet Sun areas, these bands are dominated by continuum (with C IV), by Fe VIII (0.4 MK), Fe IX (0.7 MK), and Fe XII (1.5 MK). Courtesy of the SDO

spectrograph for the 350–1,050 Å range. A prototype of the SDO EVE instrument was launched on board a rocket on April 14, 2008.

3.11 Imaging Versus Spectroscopic Instruments

In order to understand the physical conditions which lead to small and large-scale features of the solar atmosphere, it is essential to determine the local plasma parameters (temperature, density, abundance, flows, magnetic field, etc.) as a function of space and time. This can only be achieved by combining both multiwavelength imaging and spectroscopic observations of the solar atmosphere.

Imaging observations (from instruments such as SoHO/EIT, TRACE, Hinode/XRT, STEREO/EUVI, SDO/AIA etc.) generally provide a higher spatial resolution and cadence, with a wider FOV than spectroscopic observations (e.g., rasters which use a narrow slit when building up images). On the other hand, spectroscopic observations from instruments such as SoHO/CDS and Hinode/EIS are capable of observing individual spectral lines which provide both line profiles and line intensities.

Phillips et al. (2008) provide a recent review of solar UV and X-ray spectroscopy. Spectral line profiles provide information about the line shift (Doppler motion) and line width (from which thermal and nonthermal broadenings can be determined). The intensities of the spectral lines from some ions are sensitive to electron temperature or electron density. These may be used as diagnostics to probe the physical properties of the solar plasma, including temperature distributions, densities, filling factors (indicating to what extent the observed features are spatially resolved), and elemental abundances.

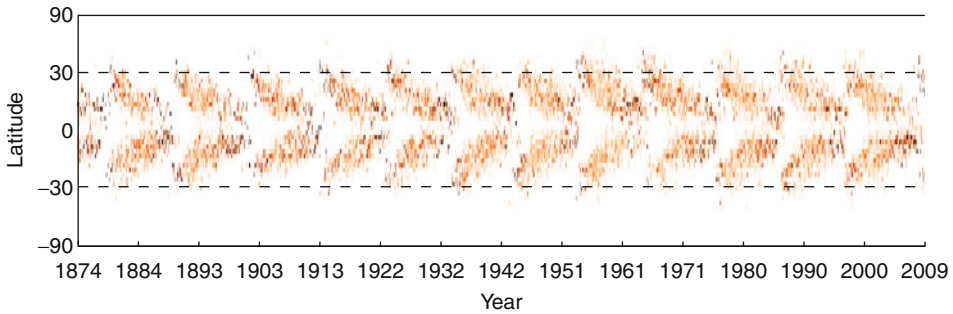
The vast array of solar instrumentation over the past three decades has enabled us to examine the solar atmosphere in microscopic detail, which sometimes seems overwhelming. It is a great challenge to be able to explain the dynamic and transient nature of the solar atmosphere.

4 The Lower Solar Atmosphere

4.1 The Photosphere and the Sun's Magnetic Field

For a detailed description of the photosphere see Thomas and Weiss (2008). The most prominent feature of the photosphere are the sunspots. Detailed observations of sunspots have been obtained by Zürich Observatory since 1849 and the Royal Greenwich Observatory since 1874. These records include information on the size and position of sunspots as well as the number. Various aspects of sunspots have been defined and used as a proxy of solar variability. The first was introduced by Rudolf Wolf of the Zürich Observatory, the Wolf Sunspot Number $WSN = c(10G + N)$, where G is the number of sunspot groups, N the number of individual ones, and c a constant factor. Many different observations existed before 1849, and efforts have been ongoing to build a more complete sample. One noteworthy example is that of Hoyt and Schatten (1998), who defined a Group Sunspot Number (GSN) and collected all archival data available at the time.

Heinrich Schwabe in 1844 was the first to indicate the presence of a solar cycle in the observed variations of the number of sunspot groups. In the second half of the nineteenth



■ Fig. 3-29

The solar butterfly diagram

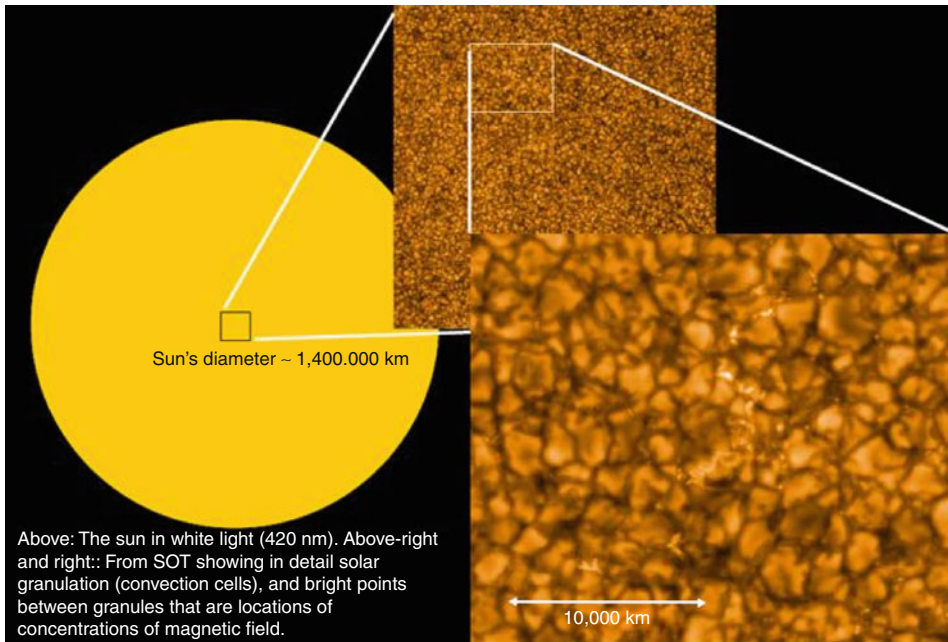
century it was noted by both Carrington and Spörer that as the cycle progresses, sunspots appear first at mid-latitudes, and then closer and closer to the equator until solar minimum is reached. This pattern is best visualized in the form of the so-called butterfly diagram, the first version of which was elegantly produced by Maunder in 1904, using data he collected together with his wife. A butterfly diagram obtained from a collection of data (rescaled and adjusted by David Hathaway.¹) is shown in ► Fig. 3-29. Sunspots for a new solar cycle tend to appear at midlatitudes while larger, more isolated sunspots from the previous cycle still persist.

Hale was the first, in 1908, to measure magnetic fields on the Sun (in sunspots where it is strongest, up to several 1,000 G). He also showed in 1919 that the orientation of the magnetic field is opposite in the northern and southern hemisphere and that the orientation of the polarities changes every 11 years during a solar cycle. During solar minimum, the solar magnetic field is approximately dipolar. Sunspots tend to be bipolar and slightly tilted compared to the direction of solar rotation (approximately east-west). The polar fields were also found to invert their polarity (Joy's law).

Magnetic flux emergence occurs continuously everywhere on the Sun. Magnetic fragments of opposite polarity tend to emerge and separate in the east-west direction. Sunspots normally form after the coalescence of magnetic fragments, and predominantly in the leading polarity (the western one). Sunspots appear dark because their temperature is much lower than the surroundings, an effect caused by the strong magnetic field (convection beneath the sunspot is inhibited). Most sunspots have an umbra and a penumbra, together with an associated (bright) plage area. Small bipolar regions continuously emerge at all latitudes on the Sun. This was seen clearly with the SoHO/MDI observations. The whole *magnetic carpet* for the quiet Sun was found to be replaced over approximately 40 h. The better the spatial and temporal resolution which is achieved in the observations, the more dynamic the Sun appears. For example, the Hinode/SOT observations show flux emergence and cancellation occurring on spatial scales smaller than 1'', and with temporal cadence of minutes or less.

On closer inspection the apparently smooth surface of the Sun, the photosphere, is seen to actually be made up of many small granules each of around 15–2,500 km across. ► Figure 3-30 shows as an example a recent image from the Hinode SOT instrument. Time sequences of images show that these granules are in continuous motion. Granules last on average 18 min,

¹<http://solarscience.msfc.nasa.gov/greenwch.shtml>



■ Fig. 3-30

First-light image from Hinode SOT showing the granular pattern in the solar photosphere (Courtesy of Hinode SOT)

and are formed by convection – upward motions of hotter material which then cools and falls back down in the intergranular lanes (observed as dark channels). Relative velocities of the up- and down-flows are only about 2 km/s.

There are also large-scale motions that form the *supergranular pattern*. Slow (0.4 km/s), almost horizontal flow from the center of cells, which are about 30,000 km in size, brings structures toward a boundary, where downflows of about 0.1 km/s are observed (Leighton et al. 1962).

The typical strength of magnetic flux elements in the supergranular cell centers is of the order of 10–20 G. About 90% of the emergent magnetic flux is located in the network, with discrete elements that have strengths up to two orders of magnitude higher (Stenflo 1989).

The region of the upper photosphere and chromosphere extends up to several thousand kilometer above the visible surface. Here, the density and pressure decrease with a scale height of only a few 100 km. The magnetic flux elements naturally expand their horizontal cross section to keep the pressure balance with the surrounding plasma. Within this layer, it is predicted that most flux tubes would either merge with flux of the same polarity or connect to nearby flux elements of opposite polarity to form the so-called *magnetic canopy*.

A significant amount of information is obtained from imaging and spectroscopy of the photosphere in the visible light. Some examples (oscillations, chemical abundances) are provided later in this chapter. However, to achieve a good understanding of what is observed, detailed modeling which includes radiative transfer effects is needed. This is beyond the scope of this chapter.

4.2 The Chromosphere

The chromosphere is the thin region above the photosphere, discovered during eclipse observations through a *flash spectrum*. Chromospheric jets, spicules, and prominences are often readily visible to the naked eye at the solar limb during an eclipse, having a distinctive pink-red color, mostly arising from the strong neutral hydrogen $H\alpha$ and $H\beta$ lines in emission (see [Fig. 3-31](#)).

In the visible, the chromosphere is observable in the strong neutral hydrogen $H\alpha$ and in the singly ionized Ca II H and K lines, which are formed mostly in the middle/higher chromosphere. These lines, in on-disk observations, show wide absorption profiles, which are occasionally self-reversed (e.g., in emission) in the core. A detailed understanding of the Ca II H and K line profiles requires in-depth modeling which takes into account radiative transfer, non-local thermodynamic equilibrium (NLTE), and nonequilibrium ionization effects, not discussed in this chapter. Further recommended reading is Athay (1976), Athay and Thomas (1961), Judge (2006), Rutten (2007).

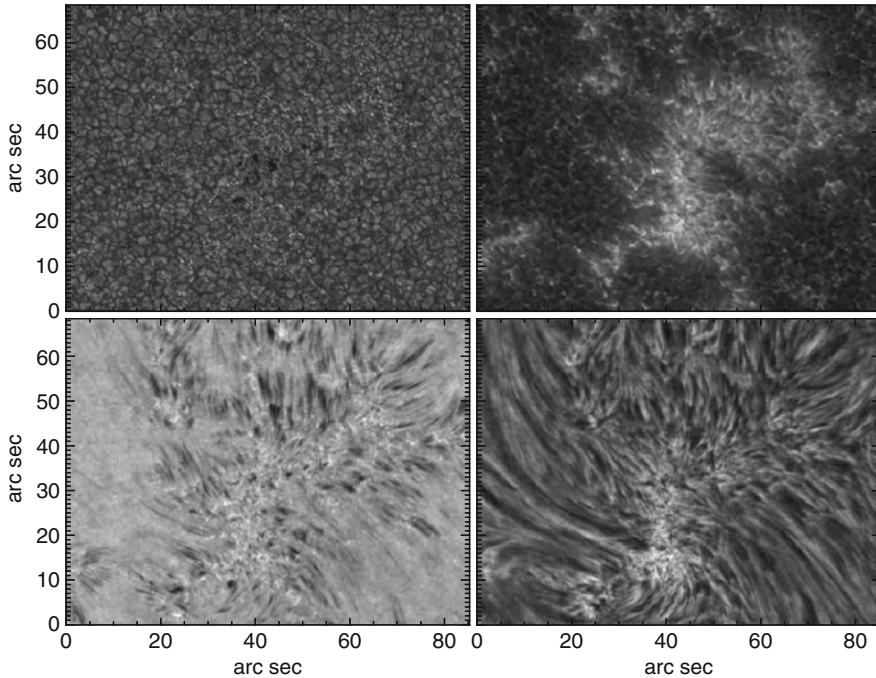
In on-disk observations, these lines show an extremely fine structure, organized in a supergranular cell-network pattern (see [Fig. 3-32](#)), and also very dynamic small-scale activity. At the limb, spicules are observed (see, e.g., [Figs. 3-33](#), and [3-34](#)).

Father Secchi SJ was the first one to make regular spectroscopic observations of the Sun, in particular of the chromosphere. He discovered the solar spicules, dynamic jets about 10,000 km long and 500 km wide, with short lifetimes (10–20 min), and with upward velocities of about 20 km/s (although even higher velocities, up to 100 km/s, have been recently reported using new high-resolution HINODE observations). These spicules tend to either fall back or fade from view. The entire chromosphere is dominated by them, indeed it is estimated that at any one time there are 70,000–400,000 spicules on the Sun. This is the reason why Secchi termed them *prateria ardente* (burning fields). Since Secchi, the dynamical character of the chromosphere has



■ Fig. 3-31

Image of the chromosphere during a total solar eclipse (Courtesy of S.Koutchmy)



■ Fig. 3-32

On-disc images from Dutch Open Telescope (DOT) ground observations in G-band (*top left panel*), Ca II H (*top right panel*), $H\alpha-0.07$ nm (*bottom left panel*), and $H\alpha$ line center (*bottom right panel*) (Figures courtesy of K. Tziotziou)

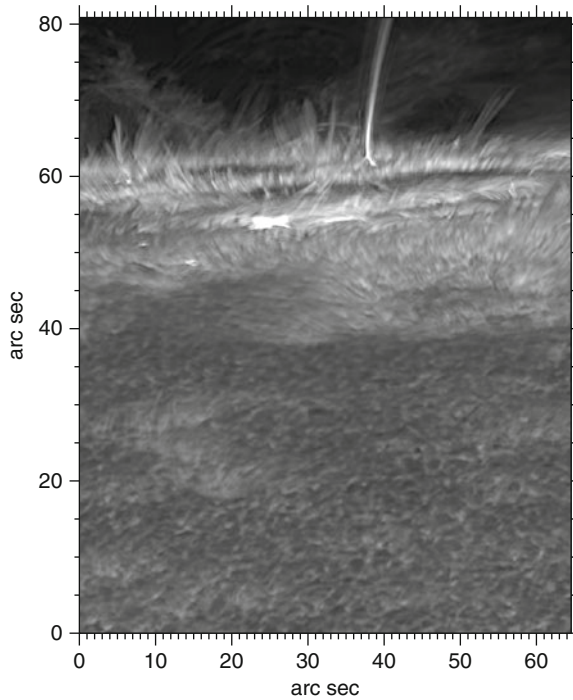
been explored extensively, for example, by Hinode/SOT (see ● Fig. 3-34). The chromosphere in coronal holes is similar, although there is the tendency to have a few macro-spicules, i.e., spicules much more extended in length. Interestingly, the thickness of the chromosphere increases at the poles during solar minimum conditions.

Cool “chromospheric” material is often found also at coronal heights, the most prominent example being quiet Sun or active region *filaments*, seen in absorption in on-disc observations, and in emission when observed above the limb called *prominences* (see ● Fig. 3-35).

Nowadays, the chromosphere is believed to be the fundamental supplier of mass and energy to the corona, and is also the place where large part of the coronal thermal energy, transferred via thermal conduction and enthalpy, is radiated.

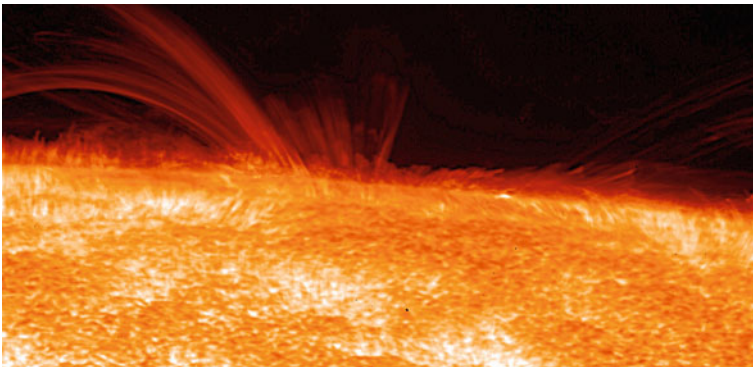
In the UV, observations of the continuum and emission lines provide detailed information about the chromosphere. The $H\text{ I Ly } \alpha$ alone radiates most of the energy emitted at chromospheric temperatures, which is an enormous amount, about 100 times the total energy required to power the extended solar corona. Other EUV-UV emission lines from neutral and singly ionized elements, in particular He, O, C, and Si, also contribute significantly.

In broad terms, chromospheric emission from the Sun arises from plasma at temperatures between 7,000 K and 25,000 K. The chromosphere is a complex region where hydrogen and the other elements become ionized (hydrogen becomes fully ionized around 20,000 K), and where chemical fractionation processes occurs (see the chemical abundance ● Sect. 7). The lower chromosphere has a similar behaviour to the photosphere, being mostly dominated by



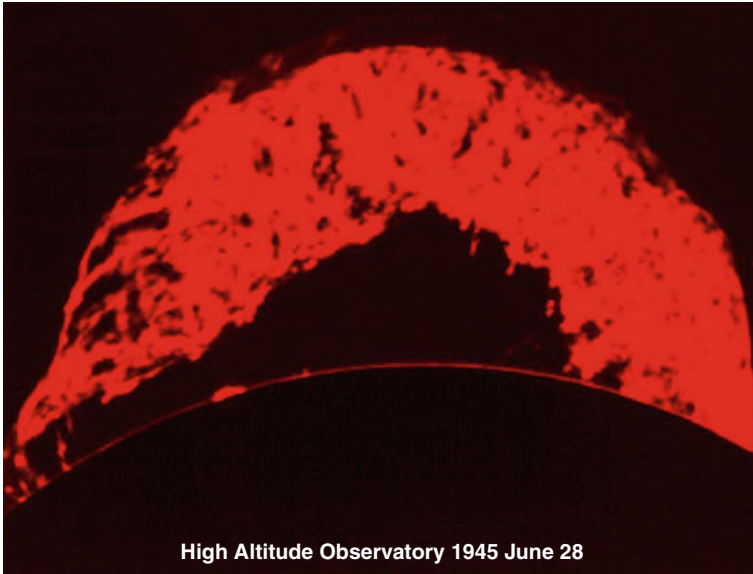
■ Fig. 3-33

A sample DOT Ca II H filtergram showing an active region on the limb on November 4, 2003. The dark pancake structures near the limb are sunspots. The bright large-scale cellular pattern at the bottom of the image is the chromospheric network, while reverse granulation is seen in the inter-network. Numerous jet-like structures are clearly visible on the limb in addition to a large surge (Courtesy of K. Tziotziou)



■ Fig. 3-34

Image of the chromosphere at the solar limb from Hinode SOT (Courtesy of Hinode SOT)



■ Fig. 3-35

The “grand dad” prominence eruption as seen in 1945 in hydrogen $H\alpha$

hydrodynamics. The upper chromosphere (above 1,300 km), on the other hand, is more like the corona, in that the morphology and dynamics of structures (mostly spicules) appear loop-like, being dominated by the magnetic fields. In fact, chromospheric emission is much brighter at the supergranular cell boundaries, where magnetic elements converge, having been advected by the supergranular flows (see ● Sect. 4.1).

Unfortunately, because of the many observational and theoretical complexities, a unified physical explanation of the chromosphere in general, and of spicules in particular, is still lacking. Many theories have been proposed for the formation of spicules, such as shocks resulting from the granular buffeting of flux tubes (e.g., Roberts 1979) and of course magnetic reconnection (e.g., Pikel’Ner 1969; Tziotziou et al. 2003). Another addition to the formation theories is the suggestion (see, e.g., De Pontieu et al. 2004; Suematsu 1990) that spicules form as a result of shocks which arise from the leakage of 5-min p-mode oscillations along tilted field lines in the Sun’s surface.

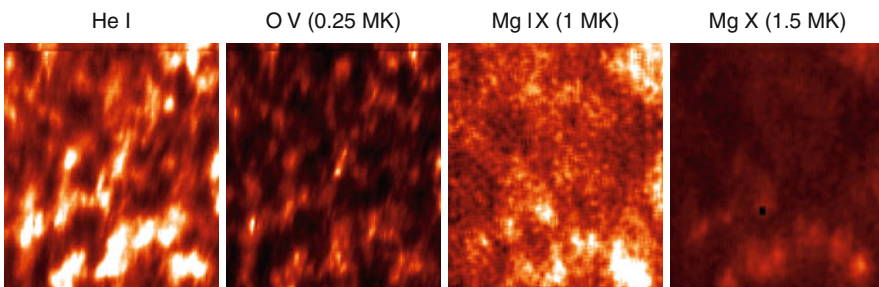
Waves with periods of about 3 min are present in the chromospheric inter-network regions (see ● Sect. 5.2). It was originally proposed that these are standing acoustic modes reflecting within the walls of a cavity formed by the temperature structure of the chromosphere. It was also thought that these p-modes in the chromosphere are not important because the waves could not propagate outside the cavity, however a clear relation between photospheric and lower chromospheric oscillations is observed. Also, some evidence for the presence of traveling waves was found in Skylab data (see the review in Mariska 1992). SoHO SUMER has provided a wealth of observations of photospheric and chromospheric lines and oscillations, which have stimulated modeling. For example, Carlsson and Stein (1997) developed a 1D hydrodynamic model where they studied the response of the lower chromosphere driven by a piston calculated to match the observed variations in a photospheric line. The piston induces the upward propagation of

3-min compressive waves and a shock near 1,000 km above the photosphere. As a result, the inter-network chromosphere below 1,300 km oscillates in response to the p modes. The results of the modeling are in agreement with chromospheric observations, which is a significant progress.

4.3 The Transition Region

In the quiet Sun, the transition region (TR) is clearly observed in emission lines formed between 0.3 and 0.9 MK. The upper TR is the region where lines between 0.6 and 0.9 MK are formed. The emission is optically thin in lines formed above 0.4 MK. Eclipse and spatially resolved observations have shown that the thickness of the TR in the quiet Sun is only about 2,000 km. Observations at the solar limb clearly show the *limb brightening*, an increase in intensity (density), due to the increased line of sight across the transition region. Limb observations show that lines formed at higher temperatures have peak emission at progressively greater heights. Observations on the solar disk instead show that the TR emission is mostly concentrated in the supergranular network, where photospheric magnetic fragments of opposite polarity continuously coalesce. This has been known for a long time (see, e.g., the Skylab ATM observations, Reeves 1976) and has been very clearly shown from SoHO observations (see, e.g. [Fig. 3-36](#)). On-disk observations also show that emission from the quiet Sun at higher temperatures becomes more and more diffuse, until it reaches 1 MK when it becomes totally unresolved at the best spatial resolution ($1''$). Interestingly, there is no significant difference between the transition region in the quiet Sun and in coronal holes.

Some authors (e.g., Feldman 1983 and subsequent papers) have argued that the TR is mostly composed of unresolved fine structures (UFS) which do not physically connect the chromospheric structures with the coronal ones. This view is hotly debated, and future instrumentation with very high spatial and temporal resolution might be needed to clarify this issue. It is however clear that the majority of lower-temperature structures appear to be low-lying, as also shown in Landi et al. (2000), where more details of the quiet Sun TR in the supergranular network can be found. It is also interesting to note that the only coronal structures easily resolvable in



■ Fig. 3-36

Monochromatic images (From a SoHO/CDS synoptic observation near Sun center) of the chromosphere-transition region in the quiet Sun, showing the supergranular network which progressively disappears at higher temperatures

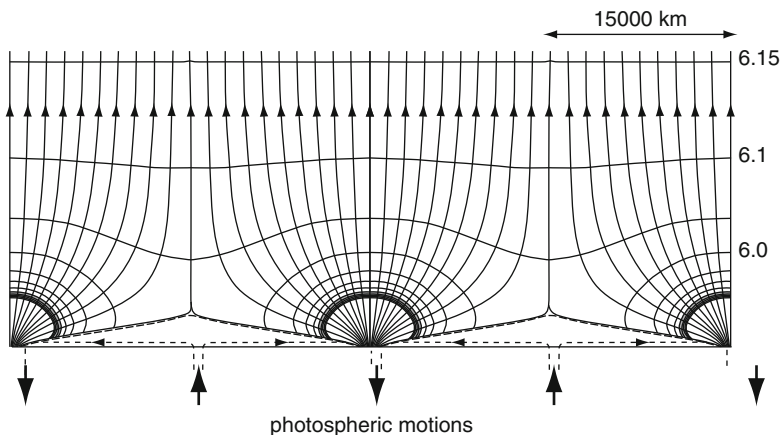
the corona, such as coronal hole plumes and coronal loops, do show a clear connection from chromospheric to coronal temperatures (Del Zanna 2003; Del Zanna et al. 2003).

At high resolution, the TR emission is structured in fine-structure emission similar to the chromospheric plasma (spicules). Dere et al. (1987) used HRTS measurements to show that chromospheric and TR spicular structures *are* the same, however this does not seem to always be the case. Dere et al. (1987) also measured very small filling factors for the TR at 10^5 K of the order of 10^{-4} (see Sect. 6.9), which is interpreted by many as evidence that this region has a filamentary structure.

Early static models of the solar atmosphere (see, e.g., the book by Mariska (1992)) proposed a plane parallel model with the existence of a thin layer, the *transitions region* (TR), with the presence of constant conductive flux from the corona toward the chromosphere, covering temperatures between say 0.3 and 1 MK. The first more physically realistic model of the TR was that of Gabriel (1976), as shown in Fig. 3-37. The strong magnetic fields at the supergranular boundaries exert a strong magnetic pressure, which naturally forces a very large expansion of the strong field, which occupies the majority of space in a small height (1,500 km). The model had various assumptions, one of which was that the supergranular cell had a unipolar magnetic field, something not confirmed by observations. A more refined model was proposed by Dowdy et al. (1986), where only a fraction of the lower-lying magnetic structures would reach the corona.

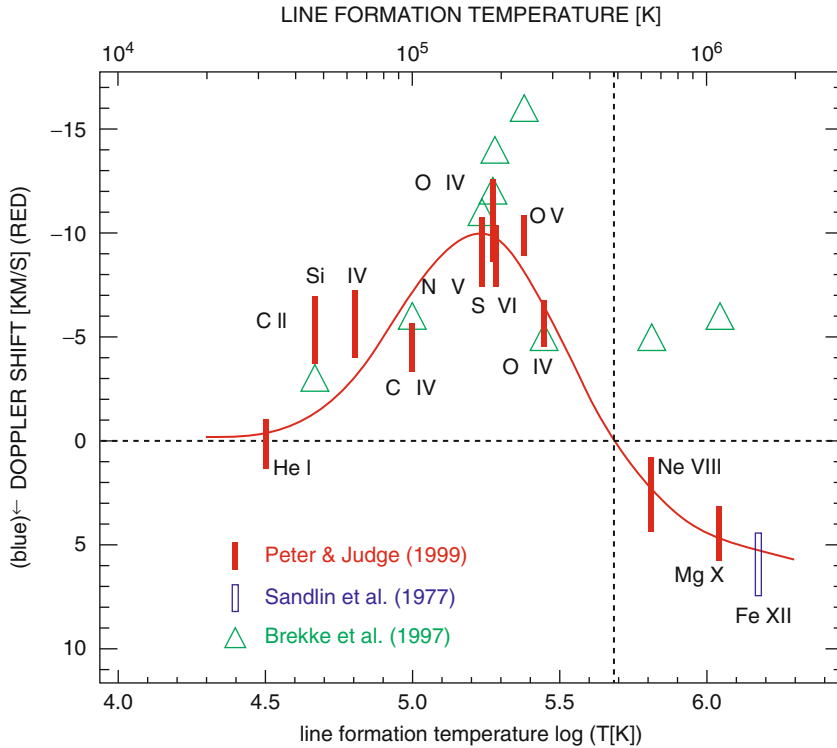
Models of a static plane parallel TR have now become obsolete, mainly because the TR is now known to be very dynamic, in terms of line intensity variations and Doppler motions. However, a clear physical picture of the TR has still not fully emerged.

Transition region spectra are characterized by line profiles much broader than their thermal width. The excess broadening varies with the temperature of formation of the line (see Sect. 6.12). Observations have been obtained by Skylab (Kjeldseth Moe and Nicolas 1977), rocket instruments (e.g., HRTS, see Dere and Mason 1993), and more recently with SoHO SUMER. The nature of this excess broadening puts constraints on possible heating processes.



■ Fig. 3-37

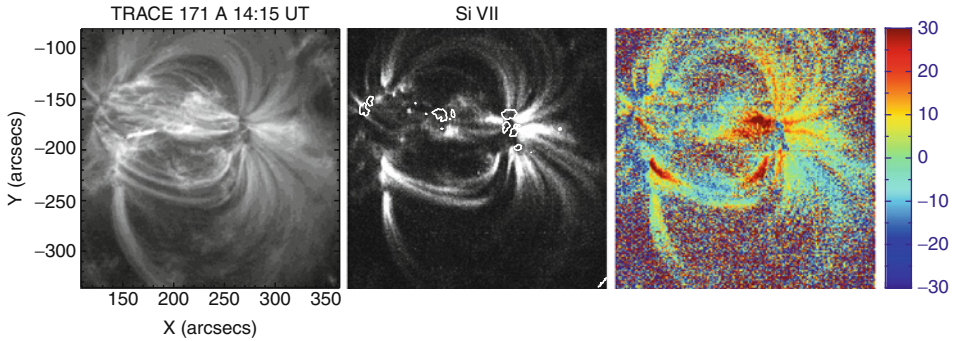
A model of the transition region according to Gabriel (1976) (Revised version courtesy of A. Gabriel), showing the magnetic field and contours of $\log T [K]$



■ Fig. 3-38
Dopplershifts measured in the transition region

At transition-region temperatures, lines appear red-shifted (see, e.g., ► Fig. 3-38), as shown from the spectrometers aboard OSO 8, the HRTS flights, Skylab, and more recently from the CDS and SUMER aboard SoHO and EIS on Hinode. Stellar UV spectra show the same pattern, which has been considered a puzzle for a long time. Many solar measurements were limited because either did not have good spatial resolution, or they were confined to small regions (Sunspots), or only considered the center-to-limb behavior of the quiet Sun (see, e.g., Peter and Judge 1999). It was only with routine measurements from EIS on Hinode that it became clear that almost all active region loops exhibit strong redshifts in both legs (Del Zanna 2007, 2008). ► Figure 3-39 shows one example. The redshifts at below 1 MK are ubiquitous, sustained, and become stronger toward the footpoints of the loop structures.

These redshifts are possibly the signature of plasma that is cooling and flowing back toward the chromosphere. Indeed, from theoretical considerations, the lifetime of a plasma that is cooling is much longer than the lifetime of a low-density plasma that is suddenly heated. Cooling plasma in the corona is dense and therefore readily observable. The cooling is mostly due to energy lost by radiation that freely escapes and by the enthalpy flux. In fact, the enthalpy flux (Bradshaw 2008; Pneuman and Kopp 1978) is a considerable mechanism by which energy can be transferred from the corona to the chromosphere. Indeed the energy transferred by thermal conduction is estimated to be a small fraction of the energy required to sustain the chromospheric radiative losses.



■ Fig. 3-39

Left: TRACE 171 Å image of an active region, showing warm loops connecting the two opposite magnetic polarities. **Middle:** monochromatic image in Si VII from Hinode/EIS. **Right:** Dopplergram in Si VII from Hinode/EIS, showing redshifts in the legs of all loops (Adapted from Del Zanna 2007)

5 The Interior of the Sun

5.1 The Standard Model and the Solar Neutrinos

The standard solar model assumes that the Sun is a spherically symmetric nonrotating star. The energy produced by nuclear fusion in the core is transferred by radiation until it reaches the outer 30% region of the solar interior. In this region, the radiative gradient exceeds the adiabatic one, and the gas becomes unstable to convection. The convection is observed in the solar photosphere in the form of the granular and the supergranular motions.

The standard model equations are integrated in time from a set of initial conditions in order to match the observed mass, luminosity, diameter, and chemical composition of the Sun at this time. The five structural variables are pressure, opacity coefficient, energy production rate, entropy, and adiabatic exponent, all of which are functions of density, temperature, and chemical abundance.

The age of the Sun is estimated to be 4.6 billion years. Knowledge of the chemical composition of the Sun and stars is nontrivial as discussed in Sect. 7. In particular, the helium abundance of the Sun is not well known. Moreover, it is not clear if the chemical abundances measured at the photosphere are representative of those inside the Sun. The convective motions should maintain sufficient mixing, however diffusion and gravitational settling are also present and could alter the internal composition. The current mass loss of the Sun through the solar wind and eruptions such as coronal mass ejections is negligible, however it is unclear how this has changed with time. The Sun is a relatively slow rotator, however it is likely that this is due to loss of angular momentum via the solar wind, magnetically coupled to the outer convection zone (see the book by Mestel 1999). In fact all young stars are fast rotators.

The rate of energy generation depends on the various ways in which hydrogen is converted into helium by fusion:



The positrons, e^+ , immediately become annihilated by encounters with free electrons, while the electron neutrinos, ν_e , largely escape from the Sun, given their extremely low cross section

for interaction with matter. The energy carried away by the neutrinos is small (about 3%) compared to the total energy released. The neutrinos offer a beautiful link between different branches of physics. The standard hydrodynamic equations, together with atomic physics calculations which provide the opacity term, can predict the status of the solar interior, defining its overall characteristics. Nuclear physics can predict the energy released by fusion and the number of neutrinos produced. We cannot easily measure the status of the solar core; however measurements of neutrinos can confirm the predictions of the various theories. Details about the observations of neutrinos and their relevance for the *solar standard model* can be found in the book *Neutrino Astrophysics* by J.N. Bahcall.

The most efficient path is the pp chain, where a continuum of low-energy (<0.4 MeV) neutrinos is produced. In the ${}^7\text{Be}$ chain, mono-energetic neutrinos are produced. In the ${}^8\text{B}$ chain, a continuum of high-energy neutrinos is produced. The number of neutrinos at 1 AU is about $7 \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$. Over the last four decades, various experiments have measured the number of solar neutrinos. Results are normally summarized in terms of the Solar Neutrino Units (SNU), which correspond to 10^{-36} reactions per target atom per second. The first experiment was in the Homestake Gold Mine, South Dakota, and led by R. Davis and J.N. Bahcall, following a suggestion from B. Pontecorvo to use chlorine: $\nu_e + \text{Cl}^{37} \rightarrow \text{Ar}^{37} + e^-$. The first measurements in 1968 gave 3 SNU. The average of results from 1970 to 1995 gave 2.56 ± 0.16 SNU. This experiment is sensitive to high-energy neutrinos, and hence sensitive to the central temperature in the core of the Sun. Standard models predict about 8 SNU for the high-energy neutrinos. This large discrepancy was termed the “solar neutrino problem”.

The Kamiokande and Super-Kamiokande experiments, which measured neutrino scattering on electrons in water, also only detected the high-energy neutrinos. The result was a flux smaller by about a factor of 2 than that predicted by the standard model.

Detection of the lower-energy neutrinos has been made possible by using gallium as a target. The GALLEX/GNO (Gran Sasso laboratory, Italy) and SAGE (Baskan underground laboratory, Caucasus) experiments have yielded values of 77.5/62.9 and 70.8 SNU respectively, again much lower (by almost a factor of 2) than estimated for lower-energy neutrinos (see the review of Dore and Orestano 2008).

A possible solution to the problem came with the idea first introduced by B. Pontecorvo in the 1950s that neutrinos might have mass, and interact with matter by oscillating between the three different types, the electron, muon, and tau neutrino. Details of the oscillation as neutrinos interact with matter were published in the 1970s and 80s.

In 2001–2002, the Sudbury Neutrino Observatory (Ontario, Canada) measured high-energy neutrinos interacting with deuterium and producing Cherenkov radiation. The experiment can measure processes which are sensitive to the different neutrino flavors. Results confirm the predictions based on the neutrino oscillation theory. The same holds for recent measurements from the Borexino experiment (Dore and Orestano 2008). It therefore appears that the solar neutrino problem has been solved. The missing neutrinos were electron neutrinos which had changed to other types on their way to Earth. They therefore escaped detection.

5.2 Solar Oscillations

The first complete observations of solar oscillations were obtained by Leighton et al. (1962). Oscillations with periods of about 5 min in the local Doppler velocity in photospheric absorption lines were observed. The amplitude of the velocity oscillation was about 0.5 km/s,

which is the result of a random combination of a large number (10^7) of individual modes, with a large degree of spatial and temporal coherence. Single modes have tiny velocity amplitudes of less than 20 cm/s. In terms of the relative intensity variation, this corresponds to a few part per million, hence is very difficult to observe. The same observations revealed the supergranular scales in the photosphere.

Ulrich (1970) proposed that these oscillations were due to standing acoustic waves in the solar interior. The theory was further developed by Stein and Leibacher (1974). Deubner (1975) found the predicted ridges in the wavenumber frequency diagram. Each ridge corresponds to a fixed number of nodes in the radial direction (n , see below). Claverie et al. (1979) first reported the 5-min oscillations at low angular, l , degree from full-Sun Dopplergrams.

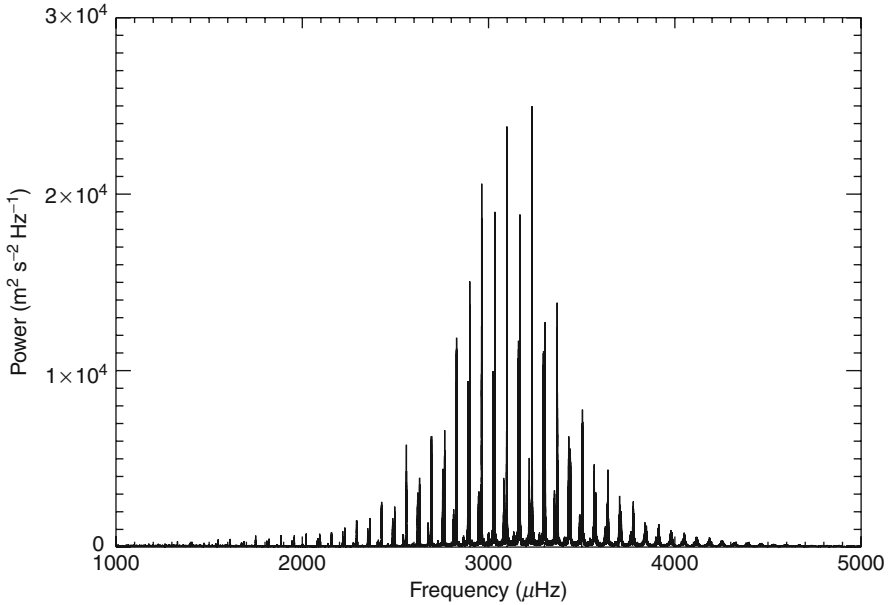
Solar-like oscillations in other stars are very difficult to observe because of their very small amplitudes. However, significant advances in instrumentation have been achieved in the search for extrasolar planets. Accurate radial-velocity measurements are now routinely carried out, to search for signals caused by the motion in the star induced by the gravitational effect of a planet. A number of ground-based observations have found that stars other than the Sun, as expected, also exhibit oscillations. New spacecraft have been launched or are planned in this field of astrophysics. For example, the French-led ESA satellite Convection, Rotation and planetary Transits (CoRoT) has been obtaining data since 2007 from a number of stars and has been measuring stellar oscillations.

Chromospheric lines also show oscillations, clustered around 3 min. These oscillations are characteristic of propagating waves rather than the evanescent waves observed in the photosphere. The modes are spatially coherent and long lasting, hence observations are needed over extended periods of time. Various ground-based networks have been established to study solar oscillations. They often observe the Sun as a star, revealing truly global oscillations. The Birmingham Solar Oscillation Network (BiSON) was established in 1981 and carries out disk-averaged velocity observations by means of potassium-vapor resonant-scattering cells. This provides very accurate measurements of the low degree modes. An example of a power spectrum is shown in ► Fig. 3-40. The Global Oscillation Network Group (GONG) has been operational since 1995. Other important observations of low- l modes have come from the LOWL instrument. These ground-based observations are important because they provide long-term records. From these measurements, it has been found that the frequencies of oscillations vary with the solar cycle, for reasons yet to be understood.

SoHO carries three helioseismic instruments. GOLF (Global Oscillations at Low Frequency) was designed as a resonant scattering Doppler-velocity instrument using sodium vapor, but has suffered some technical problems. VIRGO (Variability of solar Irradiance and Gravity Oscillations) measures disk-integrated and low-resolution intensities in three different wavelength regions. SOI/MDI (Solar Oscillations Investigation Michelson Doppler Imager) provides observations of Doppler velocities over the entire solar disk with a spatial resolution of $2''$. SDO/HMI is now providing measurements with a spatial resolution of $0.5''$.

The theory of solar oscillations is complex, however the basic characteristics can be obtained by adopting some basic approximations. Readers are referred to Leibacher and Stein (1981), Christensen-Dalsgaard (2002), Stix (2002); and references therein for more details.

To a first approximation, the rotation and small asphericity of the Sun can be neglected. We know that the timescales for the oscillations are very short when compared to the timescales for energy transport, even the fast timescales given by convection. So, a common assumption is for the oscillations to be adiabatic.



■ Fig. 3-40

Power spectrum of the solar oscillations, obtained from full-Sun Doppler observations. (Courtesy of A.M. Broomhall and the BiSON network)

The amplitude of the oscillations is really tiny, when compared to the size of the Sun, so we are only interested in the linear variations of density and pressure, i.e., small variations around equilibrium. The basic hydrodynamic equations are linearized and a search is made for a stationary solution. The equations become greatly simplified when the approximation due to Cowling (1941) is adopted. Cowling's idea was that the effect of the oscillations on the gravity can be neglected.

During the analytical derivation, three fundamental frequencies are found. The first is the acoustic cutoff frequency $\omega_a = \gamma g / 2c_s$, where $\gamma = 5/3$ for a perfect fully ionized, g is the local gravity, and c_s the sound speed. The second is the frequency of the gravity waves ω_g (known as *Brunt-Väisälä frequency*), which is the natural frequency of oscillation of a plasma element displaced from equilibrium in the presence of gravity. The third is the Lamb frequency ω_L , the frequency of a horizontally propagating sound wave.

By searching for a local solution in terms of a plane wave for the velocity $\vec{v} \simeq e^{i(\vec{k} \cdot \vec{r} - \omega t)}$, dispersion relations can be found between the frequency of oscillation ω and the wavenumber k . Different dispersion relations are obtained when e.g., radial or oblique oscillations are considered.

● **Figure 3-41** (left) presents a diagnostic diagram obtained for a simple isothermal atmosphere, i.e., frequencies versus the horizontal wavenumber k_x . Shaded areas are regions where waves can propagate. For $\omega > \omega_a > \omega_g$ we have acoustic waves that propagate radially, named *p-modes* by Cowling. They are fundamentally acoustic waves modified by the gravity field. These waves can propagate radially only if their frequency is higher than ω_a . In other words, acoustic waves that oscillate too slowly are not allowed to propagate because of the presence of the density stratification created by the gravity. At large wavenumbers, the

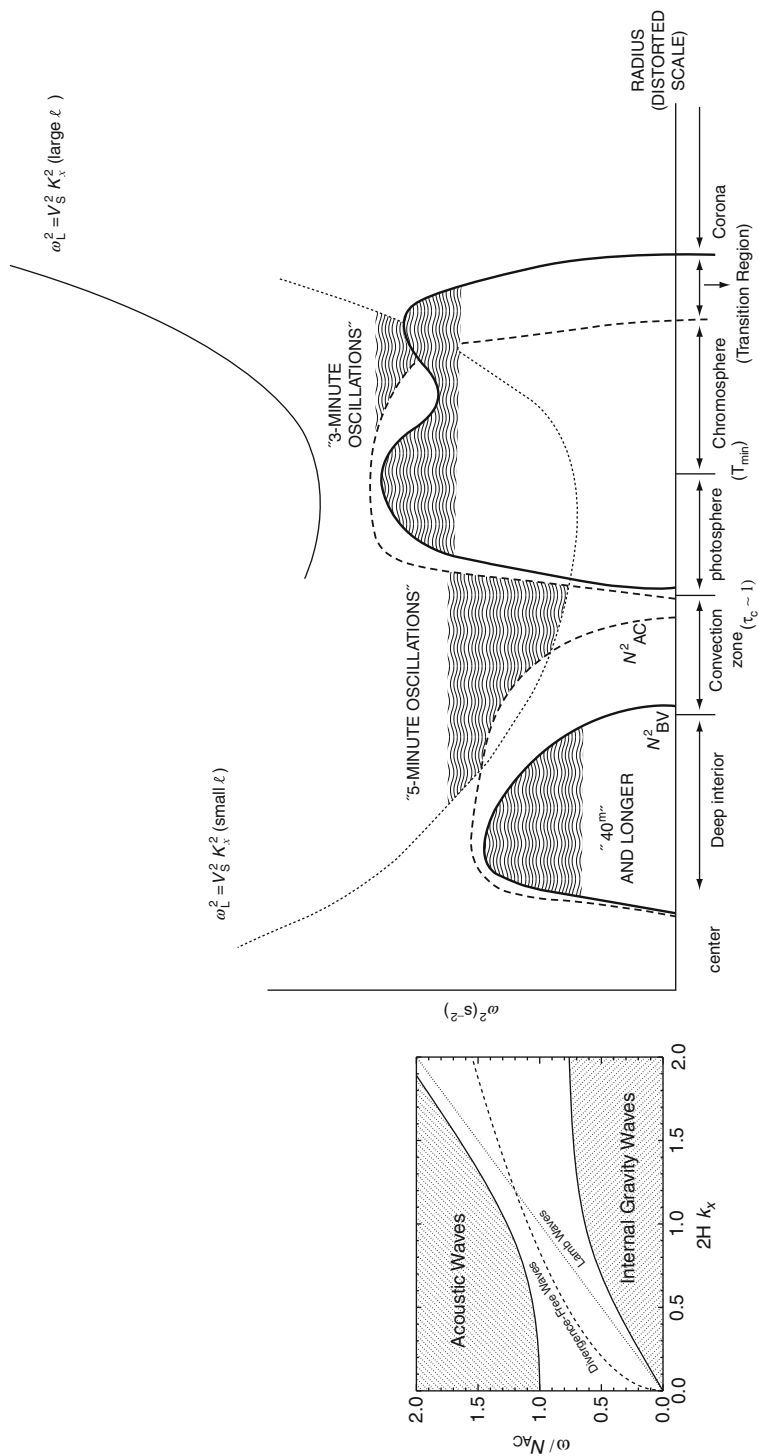




Fig. 3-41


Left: diagnostic diagram for an isothermal atmosphere. The frequency ω , normalized to the cutoff frequency N_{AC} , is shown as a function of the horizontal wavenumber k_x . H is the isothermal scale height for the pressure (and density). Shaded regions indicate where waves can propagate. Right: schematic diagram showing the trapping of 5- and 3-min modes in the solar atmosphere (Courtesy of J. Leibacher). The solid lines indicate the Brunt-Väisälä frequency N_{BV}^2 . The dashed lines indicate the cutoff frequency N_{AC}^2 .

p-modes are bound by the asymptotic Lamb frequency $\omega_L = c_s k_h$. The propagation speed for the high-frequency p-modes is the sound speed c_s . Since their discovery on the Sun, p-modes have received much attention, because it was thought that they could significantly contribute to the heating of the chromosphere.

For $\omega < \omega_g < \omega_a$ we have the *g-modes*, which are *internal gravity waves*, i.e., waves for which the restoring force is the Archimedean buoyancy (still, modified by the fact that the gas is compressible). At large wavenumbers, the g-modes are bounded by the Brunt-Väisälä ω_g frequency, and by the Lamb frequency $\omega_L = c_s k_x$ at low wavenumbers. The g-modes are essentially transverse waves, like the waves on the surface of the sea. There have been contradicting reports about observations of g-modes, and no convincing evidence has yet been found, because of the great difficulty in observing them.

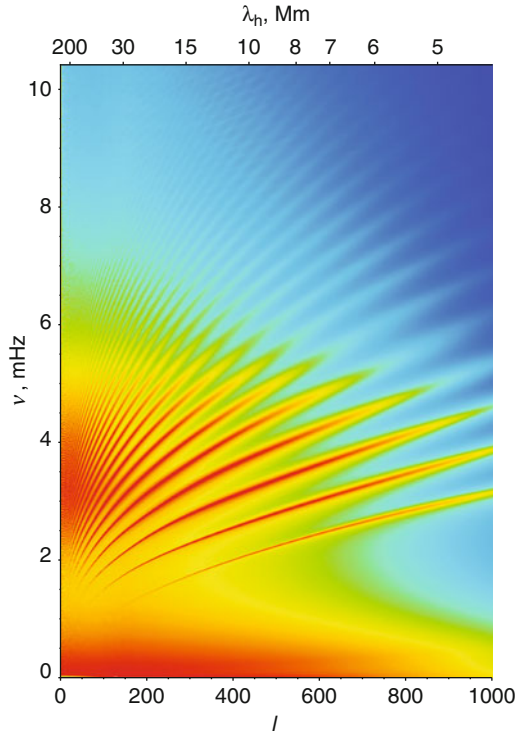
The fundamental modes (*f-modes*) are located in the evanescent-wave region, between the p- and g-modes and are essentially surface gravity waves, i.e., modes without compression. They are observed at large wavenumbers and their dispersion relation is $\omega^2 \simeq g_s k_x$ where g_s is the local surface gravity and k_x the horizontal wavenumber. They are indicated in  Fig. 3-41 (left) as the *divergence-free modes*, because $\nabla \cdot \vec{v} = 0$.

The sound speed is directly related to the temperature ($c_s \sim \sqrt{T}$ for an isothermal atmosphere), so the rapid changes of the temperature with height across the photosphere–chromosphere produce large variations in the sound speed, hence in the acoustic cutoff and Lamb frequencies. As shown in the diagnostic diagram, this in turn can affect the propagation of the acoustic waves. The lower photosphere and upper chromosphere act as two cavities where the acoustic waves become significantly trapped (see  Fig. 3-41, right). The temperature reaches a minimum at around 4500 K, and acoustic modes with periods larger than 3 min and large l become trapped in the photosphere. Acoustic waves also become trapped in the chromosphere when they have periods of 3 min or longer.

The stationary solutions (*modes*) are uniquely characterized by their frequency ω and by their spatial distribution, i.e., their different pattern on the surface of the Sun. This is characterized by two integer numbers, n, l . n is the radial order and represents, to a first approximation, the number of nodes of the radial part of the solution, while the degree l the number of nodes of the angular part on a sphere. When solar rotation is introduced, a third azimuthal number m is required to characterize the solution, and a splitting in the frequencies occurs. Radial modes, with $l = 0$, penetrate to the center, whereas the modes of highest degree observed in the Sun, with $l = 1,000$ (see  Fig. 3-42), are trapped in the outer shell of the Sun, which occupies only a small fraction of a percent of the solar radius. In other words, oscillations of different frequencies probe different layers of the solar interior and carry a great deal of information.

5.3 Some Results from Helioseismology Inversions

The observed frequencies depend on the local values of quantities such as density and temperature in the Sun, and so offer an opportunity to learn more about the solar interior. Densities and temperatures in turn depend on the local opacity and chemical abundances, as well as the details of the energy transport. It is therefore possible, with some assumptions and modeling, to obtain a variety of information from the observed frequencies. The methods and theory are quite involved, see for example Gough and Toomre (1991), Stix (2002), Christensen-Dalsgaard (2002); and references therein. The inversion to obtain the internal solar rotation is the simplest case, since it is linear with the observed splittings in the frequencies.



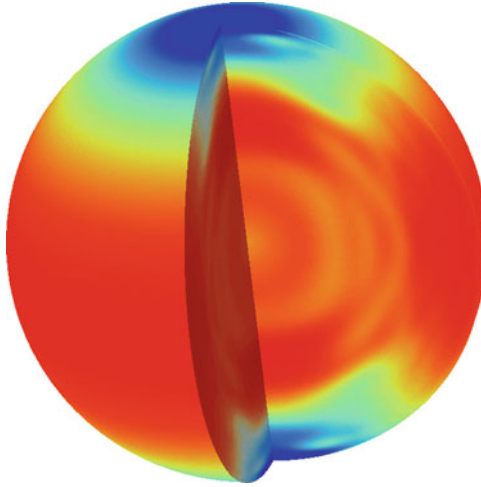
■ Fig. 3-42

Cyclic frequencies $\nu = \omega/2\pi$ as a function of the degree of the mode, l , obtained from SoHO/MDI high-cadence full disk data. Shows mode frequencies up to 10 mHz and $l = 1,000$. Each ridge corresponds to a given value of the radial order n . Courtesy of SoHO/MDI

It has long been known, since the tracking of sunspots (after Galileo), that the solar surface has a differential rotation, with the equatorial regions spinning much faster than the poles. Inversions of ground-based observations (e.g., BBSO) of oscillations did indicate that the differential rotation continued down to the base of the convection zone, but large uncertainties were present in the interior. SoHO/MDI has confirmed that the change in rotation occurs in the zonal belts and extends deep into the convection zone.

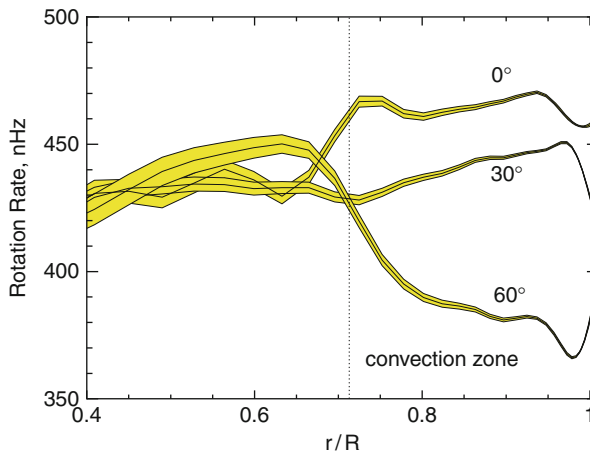
An important discovery from SoHO/MDI has been that the Sun rotates almost rigidly in the radiative zone (cf. ► Figs. 3-43, and ► 3-44). These observations were totally unexpected from previous models, and have recently transformed our understanding on the internal rotation in a star (see the review of Thompson et al. 2003). The region where strong shear is present, known as the *tachocline*, is close to the base of the convection zone (as estimated from inversion methods). It is now thought to be the region where magnetic fields are created via hydromagnetic dynamo processes and then transported to the surface by the convective motions. See Tobias and Weiss (2007), Miesch (2005), Charbonneau (2005), Weiss and Thompson (2009) for details.

SoHO/MDI also confirmed the presence of a meridional flow, from the equator to the poles. It was previously deduced from ground-based observations, but thanks to MDI it is now clear that it affects a significant depth of the outer layer of the Sun, down to 10,000 km. The velocity



■ Fig. 3-43

This false-color image represents surface and internal velocities of the Sun as obtained from 1 year of SoHO MDI observations, beginning in May 1996. *Red* indicates greater angular rotation rates, and *dark blue* the slowest. The inner 70% of the Sun is thought to rotate almost as a solid body, whilst in the outer layer there is a marked differential rotation, with the equator rotating faster than the poles. Internal flows of gas at the equator and near the poles are also present (Courtesy of P. Scherrer and the SoHO/MDI consortium)



■ Fig. 3-44

Solar rotation rate inferred from 2 months of MDI data as a function of radius at three latitudes (Courtesy of SoHO/MDI and A. Kosovichev)

of the meridional flow is very slow (about 1%), when compared to the rotation speed, but is fast enough to recycle material within 1 year, a relatively short timescale, compared to the solar cycle. Note that the velocity of this flow is opposite to the changes in magnetic field emergence along the cycle (from high latitudes toward the equator), and it is still not clear what happens

to the flow when it reaches the poles. However, it is clear that the flow must have an important role for the solar cycle.

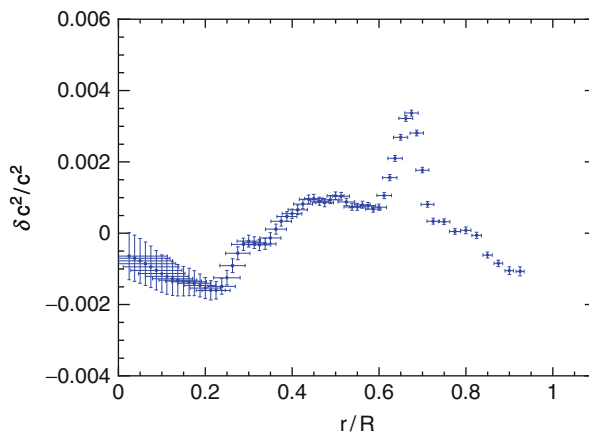
Helioseismological inversions to obtain information on the solar structure are more complex, because the relation between frequencies and structural quantities is highly nonlinear. The inversion methods are normally based on linearizing the equations for the oscillations (often assumed adiabatic) around a known reference model. The differences between the observed and computed frequencies are then related via some kernels to two model parameters, for example density and sound speed. Various inversion techniques are applied to obtain the variation of the model parameters with radius. The helioseismology results are often presented as the relative differences between the observed and predicted ones.

One parameter, the sound speed, can be measured accurately and does not depend greatly on the solar model used. Accurate measurements were obtained from early ground-based observations, but more accurate ones come from SoHO/MDI. [Figure 3-45](#) shows the difference in the square of the sound speed between that obtained from 2 months of MDI data and the reference solar model “S” of Christensen-Dalsgaard of 1996. These types of differences are very small but are nonetheless significant, and models are being continuously refined to reduce them.

Another important parameter for the solar structure that is obtained from helioseismology is the location of the bottom of the convective zone.

5.4 Abundances and Helioseismic Models

Parameters such as the chemical abundances are more difficult to derive from helioseismology. The sound speed is directly related to the chemical composition, hence it should be possible to assess whether or not the adopted solar abundances are correct by comparing the variation of the sound speed as predicted by solar structure models with the observed one. However,



■ Fig. 3-45

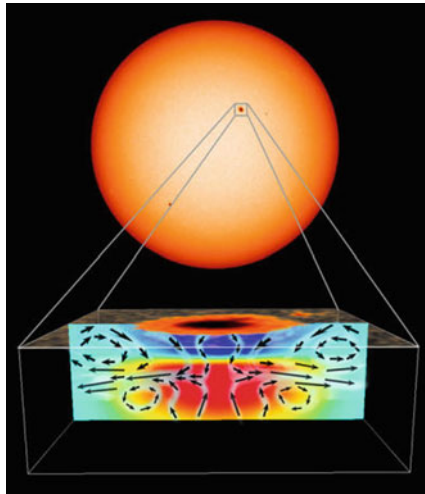
The relative differences between the squared sound speed in the Sun and in a standard solar model, as a function of radial distance (Courtesy of A. Kosovichev)

the methods are complex and depend on several physical assumptions, in particular on the calculation of opacities. For a recent review see Basu and Antia (2008).

Until recently, the predicted sound speed (as a function of depth) from standard solar models showed very good agreement with the helioseismic values. However, adopting the reduced photospheric abundances obtained from the new 3-D simulations causes the opacity to decrease, which in turns reduces the depth of the convection zone, to a level inconsistent with the seismic estimate. See, for example, Serenelli et al. (2009). It appears that the discrepancies cannot be resolved, even with the most recent opacity calculations from the Opacity Project, see Badnell et al. (2005). The He abundance from the latest solar structure models is also at odds with the value inferred from helioseismology. More work is needed in this important area of research.

5.5 Local Helioseismology

Recent helioseismic techniques such as Ring-analysis, Time-distance and Acoustic-holography (see review by Gizon and Birch 2005) measure large-scale flows and sound speed perturbations in and around magnetically active regions (see, e.g., [Fig. 3-46](#)). The interaction of flow fields and observed magnetic fields are important for understanding the stability and long-term evolution of the active regions. It was also established from the innovative work of Braun and coworkers (see Braun and Birch 2008 and references therein) that the sunspots and plages absorb 50–70% of the acoustic power. Unfortunately, it is not clear what changes magnetism causes to the helioseismically determined flows and sound speeds near and within the active regions. Some modeling efforts to assess these have only just begun (e.g., Jain et al. 2009). With the much awaited high-resolution data from the HMI instrument of the Solar Dynamics Observatory, these are exciting times for local magnetic helioseismology.



■ Fig. 3-46

Flow pattern underneath a sunspot, from SoHO MDI helioseismology (Courtesy of the MDI consortium)

5.6 The Solar Dynamo

The issue of the continuous regeneration of the solar magnetic fields is too extensive to be covered in detail in this review chapter. Readers are referred to reviews such as Weiss (1994), Tobias (2004), Miesch (2005), Charbonneau (2005), Solanki et al. (2006). As with the coronal heating problem, many ideas and models have been put forward, but a comprehensive coherent model based on solid physical grounds is still needed. Even the most advanced computations are still not sufficient to model both the fine and large-scale structures in the convection zone. The major problem is how to explain the observed cyclic behavior of magnetic flux emergence on the Sun. During solar minimum conditions, the solar magnetic field is mainly bipolar (poloidal). The differential solar rotation stretches the magnetic field lines and creates a toroidal component. This occurs because the magnetic field is expected to be frozen-in with the plasma motions. In the 1950s, E.N. Parker suggested the idea of a turbulent dynamo, i.e., that within the convection zone, up-flowing volumes of plasma would rotate because of the Coriolis force. This would twist the toroidal field to produce loops in the poloidal plane. Dynamo mean field theory was subsequently developed and became very popular. This theory later failed to explain a variety of features, most importantly the migration patterns and the internal solar rotation as inferred from helioseismology.

In different approaches, the rearrangement of the solar magnetic fields in the solar corona plays a major role. In the Babcock–Leighton (Babcock 1961; Leighton 1964, 1969) mechanism, toroidal field is still produced by the differential rotation of the Sun in the convection zone. This toroidal component creates largely bipolar sunspots that emerge through the photosphere by magnetic buoyancy. It is known (Joy’s law) that bipolar sunspots are normally tilted, so when they decay they tend to create a diffuse poloidal component of the field. The pattern is opposite in the two hemispheres. One current idea is that these weak diffuse poloidal fields are advected toward the poles by the meridional circulation and regenerate the solar poloidal field. There is little doubt now that this occurs on the Sun. What is still not clear is if this is a key component of the dynamo cycle, or just a consequence of it.

In recent years, the importance of considering further effects occurring in the corona has been recognized. For example, Fisk and Schwadron (2001) proposed a theory for the evolution of “open” magnetic flux, by which open field lines continuously reconnect with closed structures, and are subject to the convective diffusive motions on the solar surface. These diffusive motions can explain the accumulation of open flux in the polar regions of the Sun during minimum, and their reversal from cycle to cycle. As a by-product, the reconnection provides mass and acceleration to the solar wind. Another process which is clearly occurring on the Sun is the generation of Coronal Mass Ejections, CMEs, which are global coronal events and can similarly rearrange the large-scale solar fields and become important contributors to the 11-year solar cycle (see, e.g., Low 2001).

6 Radiation from the Solar Corona: Atomic Processes and Plasma Diagnostics

Emission lines result from the spontaneous decay of an excited electron within an ion to a lower energy state, the excess energy being carried off by a photon. The intensity of a particular emission line as measured by an observer clearly depends on the number of emitting

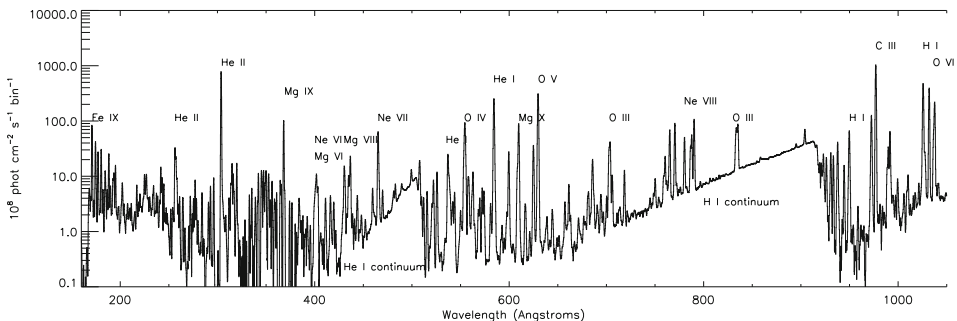
ions in the line of sight and the fraction of these ions that are in the particular excited state that gives rise to the spectral line. We are thus interested in both the *ionization state* of the plasma and the *level balance* within an ion. For each atom there is an ongoing competition between processes changing the ionization state of the ions and processes changing their energy state.

The outer solar atmosphere consists of a hot, low-density plasma. The corona is a fully ionized plasma, being composed mainly of hydrogen and helium fully stripped of their electrons, while heavier elements are at least partially ionized. Hot coronal plasma itself gives rise to line as well as to continuum radiation in the UV to X-ray wavelengths. One can also detect the visible and UV radiation from the lower atmospheric layers, since the low-density corona is *optically thin*.

The EUV spectral range (see ► Fig. 3-47) is rich in chromospheric lines from He II, lines from the transition region, and coronal lines. It also has continuum emission from hydrogen and helium. The far UV has strong chromospheric lines, the brightest ones being those from H I, but also lines from the transition region, and a few coronal lines.

In principle, knowing the relevant atomic parameters, at a given temperature and electron density, one can determine both the *ionization state* and the energy *level balance* within an ion. From this, the intensity of the emission lines and continuum due to *bound-bound* (from one energy to another within an atom), *bound-free*, and *free-free* electron processes, can be predicted. In practice, however, the calculations represent a very complex task. For each atom, there is an ongoing competition between processes changing the ionization (charged) state of ions (ionization/recombination) on one hand, and processes changing their energy state (excitation/decay) on the other hand. All atomic processes are required to balance their corresponding inverse processes, according to the *principle of detailed balance*.

Reviews of atomic processes and spectroscopic diagnostics for the solar transition region and corona are provided by Dere and Mason (1981), Foukal (2004), Gabriel and Mason (1982), Mariska (1992), Mason and Monsignori Fossi (1994), Del Zanna et al. (2002b), and Phillips et al. (2008).



■ Fig. 3-47

The EUV spectrum of the whole Sun, as measured by the prototype SDO/EVE instrument flown aboard a rocket in April 14 2008 (Chamberlin et al. 2009; Woods et al. 2009)

6.1 Line Flux and Intensity

When the Sun is observed as a star, the spectral flux (i.e., total irradiance in a line), $F_T(\lambda_{ji})$, for an optically thin line of wavelength λ_{ij} (frequency $\nu_{ij} = \frac{c}{h\lambda_{ij}}$, where h is Planck's constant) is defined as:

$$F_T(\lambda_{ji}) = \frac{1}{4\pi d^2} \int_V P_{ji} dV \quad (\text{erg cm}^{-2} \text{ s}^{-1}) \quad (3.2)$$

where d is the Sun/Earth distance, $P_{ji} = N_j(X^{+m})A_{ji}h\nu_{ji}$ is the power emitted ($\text{erg cm}^{-3} \text{ s}^{-1}$) in the line per unit volume (also called the line emissivity). Here i, j are the lower and upper atomic levels, A_{ji} is the spontaneous radiative transition probability, and $N_j(X^{+m})$ is the number density of the upper level j of the emitting ion. The integral is over the volume V of the emitting plasma.

Most solar observations are spatially resolved, so what is measured is the radiance, or intensity $I(\lambda_{ij})$, of an optically thin spectral line emitted from a volume V , which is:

$$I(\lambda_{ij}) = \frac{h\nu_{ij}}{4\pi} \int_V N_j(X^{+m}) A_{ji} dV \quad [\text{ergs s}^{-1} \text{ sr}^{-1}] \quad (3.3)$$

In practice, the spatial resolution of the instrument defines the volume of plasma, $V = Sh$, where S is the projected area of the instrument aperture, the solid angle and h is the line of sight through the emitting plasma.

There are many processes that cause changes in the energy state for an emitting ion, due to its electrons transitioning between different energy levels. For instance, an ion in its ground state configuration becomes excited when one of its electrons makes a transition to a higher energy level. Similarly, the same ion may return to the ground state configuration when the excited electron falls back down to the ground energy level.

Generally, the processes that populate/depopulate the excited levels of an ion are much faster than the processes affecting the charged state of the ions (ionization/recombination). Thus, the problem of calculating the excited level populations can be separated from that of calculating the ionization state. The population of the level j is then often expressed as:

$$N_j(X^{+m}) = \frac{N_j(X^{+m})}{N(X^{+m})} \frac{N(X^{+m})}{N(X)} \frac{N(X)}{N(H)} \frac{N(H)}{N_e} N_e \quad (3.4)$$

- $N(X^{+m})/N(X)$ is the ionization ratio of the ion X^{+m} relative to the total number density of element X
- $A(X) = N(X)/N(H)$ is the chemical abundance of element X relative to hydrogen ($N(H)$ or N_H);
- $N(H)/N_e$ is the hydrogen density relative to the free electron density. In the literature, it is often assumed to be equal to 0.83 (assuming fully ionized plasma and assuming $N(H)/N(\text{He}) = 10$), however this value depends on the ionization state and helium abundance.
- The fraction $N_j = \frac{N_j(X^{+m})}{N(X^{+m})}$ of ions X^{+m} lying in the state j .

In this case we can express the line intensity as:

$$I(\lambda_{ji}) = \int_h N_e N_H A(X) G(N_e, T, \lambda_{j,i}) dh \quad (3.5)$$

where $G(N_e, T, \lambda_{j,i})$, sometimes denoted as $G(T)$, the *contribution function* is calculated from atomic parameters for that spectral line. $G(T)$ has a very strong dependence on temperature, through the ionization balance calculations, and a weak dependence on electron number density, N_e .

6.2 Differential Emission Measure and Emission Measure

If a unique relationship exists between N_e and T , a differential emission measure $\text{DEM}(T)$, a function of only the plasma temperature, can be defined (Withbroe 1978):

$$\int_T \text{DEM}(T) dT = \int_h N_e N_H dh \quad (3.6)$$

i.e.,

$$\text{DEM}(T) = N_e N_H \frac{dh}{dT} \quad [\text{cm}^{-5} \text{K}^{-1}] \quad (3.7)$$

For example, if we assume that the plasma pressure, P , along the line of sight is constant, then from the perfect-gas law $N_e^2 \sim (P^2/T^2)$, and the electron density is only a function of temperature, and the $\text{DEM}(T)$ can be defined.

The DEM gives an indication of the amount of plasma along the line of sight that is emitting the radiation observed and has a temperature between T and $T + dT$. For a plane parallel atmosphere, the DEM can be related to the conductive flux and can give information about the structure of the atmosphere. With this definition, the intensity integral becomes

$$I(\lambda_{ij}) = A(X) \int_T G(T) \text{DEM}(T) dT, \quad (3.8)$$

where the abundance of the element $A(X)$ is assumed to be constant along the line of sight. We therefore have a system of Fredholm integrals of the first species to be inverted, in order to deduce the DEM from a set of observed intensities. This inversion procedure is notoriously difficult (see, e.g., Craig and Brown 1976).

Once the differential emission measure DEM is known, the *total column emission measure* EM_h can be calculated by integrating the DEM over the whole temperature range:

$$EM_h \equiv \int_h N_e N_H dh = \int_T \text{DEM}(T) dT \quad [\text{cm}^{-5}] \quad (3.9)$$

When fluxes are observed, instead of intensities, the *volume emission measure* $EM_V = \int_V N_e^2 dV \text{ (cm}^{-3}\text{)}$ and its differential in temperature $\text{DEM}_V(T) = N_e^2 \frac{dV}{dT}$ are defined accordingly.

6.3 Atomic Processes Affecting the Ion Excitation

The number density population of level j is calculated by solving the statistical equilibrium equations for a number of low-lying levels and including all the important collisional and

radiative excitation and de-excitation mechanisms:

$$\frac{dN_j}{dt} = \sum_{i<j} N_i R_{i,j} + \sum_{i>j} N_i A_{i,j} + \sum_{i>j} N_i N_e C_{i,j}^e + \sum_{i<j} N_i N_e C_{i,j}^e - N_j (\sum_{i<j} A_{j,i} + \sum_{i>j} R_{j,i} + N_e \sum_{i<j} C_{j,i}^e + N_e \sum_{i>j} C_{j,i}^e) \quad (3.10)$$

- * C^e ($\text{cm}^{-3} \text{s}^{-1}$) are the electron collisional excitation rate coefficients. $N_e C_{i,j}^e$ gives the number of electron collisional transitions per unit time such that an ion changes its state from i to j .

The proton collisional coefficient can be considered as an additional term. Proton collisions are important when excitation energies are small, $\Delta E_{i,j} \ll kT_e$. This happens, for instance, for transitions between fine structure levels within the ground configuration of an ion at high temperatures, as for example the Fe XIV transition in the ground configuration: $3s^2 3p(^2P_{1/2} - ^2P_{3/2})$ as explained in Seaton (1964a).

- * $A_{j,i}$ (s^{-1}) are the spontaneous radiation transition probabilities.
- * $R_{j,i}$ (s^{-1}) are the stimulated absorption rate coefficients, proportional to the mean intensity of the radiation field. Photoexcitation and stimulated emission are important processes which also need to be included in the level balance equations. It is possible to show that this can be simply done by replacing the $A_{i,j}$ value with a generalized radiative decay rate.

In practice, because of the short timescales of the relevant processes, for the solar corona it is normally assumed that the plasma is in a steady state ($\frac{dN_j}{dt} = 0$). The set of equations 3.10 is then solved for a number of low-lying levels, with the additional requirement that $N(X^{+m}) = \sum_j N_j$.

The electron collisional excitation rate coefficient for a Maxwellian electron velocity distribution with a temperature $T_e(K)$, is given by

$$C_{i,j}^e = \frac{8.63 \times 10^{-6}}{T_e^{1/2}} \frac{\Upsilon_{i,j}(T_e)}{\omega_i} \exp\left(\frac{-\Delta E_{i,j}}{kT_e}\right) \quad (3.11)$$

where ω_i is the statistical weight of level i , $\Delta E_{i,j}$ is the energy difference between levels i and j , k is the Boltzmann constant and $\Upsilon_{i,j}$ is the thermally averaged collision strength:

$$\Upsilon_{i,j}(T_e) = \int_0^\infty \Omega_{i,j} \exp\left(-\frac{E_j}{kT_e}\right) d\left(\frac{E_j}{kT_e}\right) \quad (3.12)$$

Here the collision strength (Ω) is a symmetric, dimensionless quantity, which is related to the electron excitation cross section; E_j is the energy of the scattered electron relative to the final energy state of the ion. The electron de-excitation rates are obtained by the application of the principle of the detailed balance.

$$C_{j,i}^e = \frac{\omega_i}{\omega_j} C_{i,j}^e \exp\left(\frac{\Delta E_{i,j}}{kT_e}\right) \quad (3.13)$$

The solution of the electron-ion scattering problem is complex and takes a great deal of computing resources. The accuracy of a particular calculation depends on two main factors. The first is the representation which is used for the target wavefunctions, the second is the type of scattering approximation chosen. The collision strengths usually have a slowly varying part and spikes that are due to resonances (dielectronic captures). The theory on electron-ion collisions has been developed by Seaton (1964b). Calculations can be very accurate (better than 10%),

but are not trivial for many ions. The target must take account of configuration interaction and allow for intermediate coupling for the higher stages of ionization. The main approximations used for electron-ion scattering are *Distorted Wave* (DW), *Coulomb Bethe* (CBe) and the more elaborate *Close-Coupling* (CC) approximation. In the CC approximation, the scattering electron sees individual target electrons, the channels are coupled and a set of integro-differential equations are solved.

There are an abundance of electron scattering calculations for individual ions in the literature. The complexity of iron ions leads to many spectral lines and useful diagnostics, but also means atomic calculations are difficult. The Iron Project, an international collaboration of atomic physicists, was set up to calculate electron excitation rates for all the iron ions (Hummer et al. 1993). A series of Iron Project papers have been published in *Astronomy and Astrophysics*. A large set of excitation data have also been produced recently by the UK Rmax and APAP networks. The most accurate calculations for coronal iron have been performed by P.J.Storey and collaborators. Other international groups, for example at Queen's University of Belfast, have also produced extensive atomic data. Atomic data has been assessed and gathered in various databases, for example the CHIANTI database discussed below.

6.4 Atomic Processes Affecting the Ion State

Since the coronal plasma is optically thin, most of the radiation escapes, the plasma is not in local thermodynamic equilibrium (LTE), and the main processes that regulate the ionization balance are spontaneous radiative recombination, di-electronic recombination, photoionization, and collisional ionization. By considering the detailed balance equations that relate successive ionization stages for these processes and requiring that the population of ions in a given state is constant with time, it is possible to derive for each state the ratio of two successive ionization stages $N(X^{+m+1})/N(X^{+m})$, which is mainly a function of the temperature, and only slightly depends on the electron density. With the additional condition $N(X) = \sum_m N(X^{+m})$, it is possible to derive the ionization ratio.

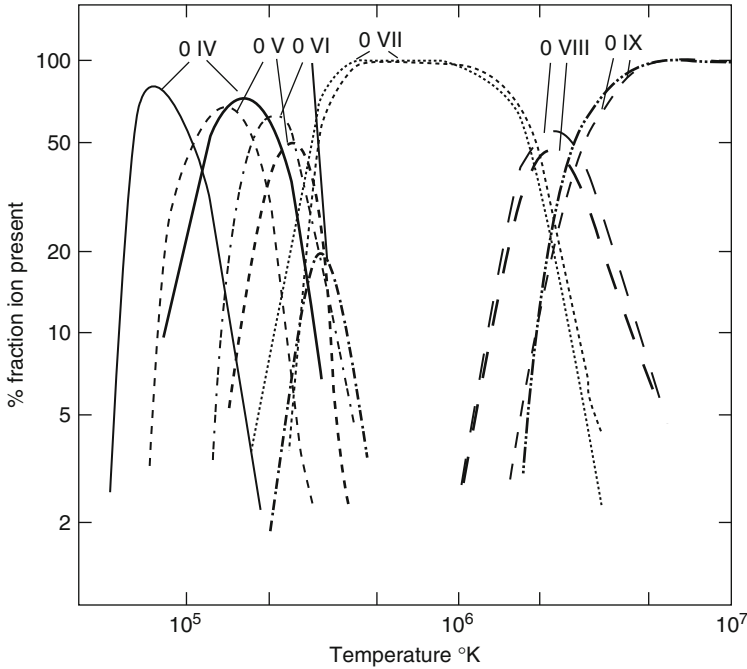
In the case of local thermodynamic equilibrium (LTE), if we write the detailed balance including photoionization and radiative recombination we obtain the Saha equation. At low densities, the plasma becomes optically thin and most of the radiation escapes, therefore processes like photoionization and radiative recombination are attenuated and the plasma is no longer in LTE.

If the timescales of the observed phenomena in the coronal plasma are longer than those for ionization and recombination, ($\frac{dN(X^{+m})}{dt} = 0$), the degree of ionization of an element is obtained by equating the ionization and recombination rates that relate successive ionization stages:

$$N(X^{+m})(q_{\text{col}} + q_{\text{au}} + q_{\text{ct}}) = N(X^{+m+1})(\alpha_r + \alpha_d + \alpha_{\text{ct}}) \quad (3.14)$$

The dominant processes in optically thin plasmas are *collisional ionization – direct ionization from the inner and outer shells* (q_{col}) and excitation followed by *autoionization* (q_{au}); *radiative recombination* (α_r) and *dielectronic recombination* (α_d); and *charge transfer* (α_{ct} , q_{ct}).

The process of dielectronic recombination was shown by Burgess (1964, 1965), and Seaton (1964b) to be a very important effect for the solar corona. Figure 3-48 shows the effect that inclusion of dielectronic recombination has on the ion abundances. Dielectronic recombination occurs when a free electron is captured into an autoionization state of the recombining



■ Fig. 3-48

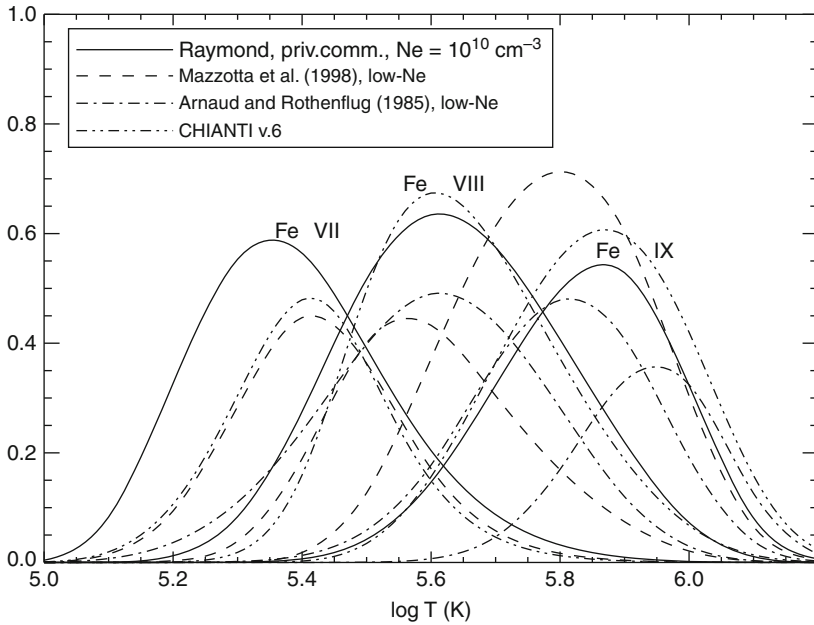
The effect that dielectronic recombination has on ion abundances (*thin lines* are calculated without), from Pottasch (1967)

ion, the ion can then autoionize (releasing again a free electron) or produce a radiative transition into a bound state of the recombined ion. The transition can only occur at specific wavelengths. There are many diagnostic applications using these satellite lines, which we cannot describe here. See, for example, Gabriel and Mason (1982), Phillips et al. (2008); and references therein. A significant number of recombination rates have been calculated by N.R.Badnell and collaborators.

Most ion abundances are calculated in the low-density limit, assuming that all the population in an ion is in the ground state. This is a good approximation for many ions, but not those which have metastable levels. The ion population slightly depends on the electron density through the rate coefficients, as shown by Burgess and Summers (1969). An example of how different ion abundances can be obtained using different rates is shown in ► Fig. 3-49.

6.5 Coronal Model Approximation: Two-Level Ion

In the low-density corona, radiative decay (spontaneous emission with Einstein coefficient A_{ji}) occurs much more often than collisional excitation processes, and, in general, one can neglect the population of higher excited states, j , relative to the ground state, g . Under this so-called *coronal-model approximation*, it is only the electron *collisional excitation* from the ground state



■ Fig. 3-49

The abundances of three iron ions as calculated in equilibrium using different sets of ionization/recombination rates

of an ion and the *spontaneous radiative decay* that are competing. Thus, the radiative emission rate may be inferred from the equilibrium between the excitation/de-excitation processes expressed by the following equation:

$$N_g N_e C_{g,j}^e = N_j A_{j,g} \quad (3.15)$$

If the coronal approximation breaks down, if for example $A_{j,g}$ is small (a forbidden or intersystem transition), then collisional de-excitation can compete with radiative decay as a depopulating mechanism.

6.6 Emission Measure Approximations

Following Pottasch (1963), many authors have approximated the above expression (● 3.5) by removing an averaged value of $G(T)$ from the integral:

$$I_{\text{th}} = A(X) \langle G(T) \rangle \int_h N_e N_H dh \quad (3.16)$$

A *line emission measure* EM_L can therefore be defined, for each observed line of intensity I_{ob} :

$$\text{EM}_L \equiv \frac{I_{\text{ob}}}{A(X) \langle G(T) \rangle} \quad [\text{cm}^{-5}] \quad (3.17)$$

where $\langle G(T) \rangle = 0.7G(T_{\max})$. The method assumes that each line is mainly formed at temperatures close to the peak value T_{\max} of its contribution function, $G(T_{\max})$, and produces plots of the *line emission measures* (multiplied by the corresponding abundance value)

$$A(X) \text{ EM}_L = \frac{I_{\text{ob}}}{0.7G(T_{\max})} \quad (3.18)$$

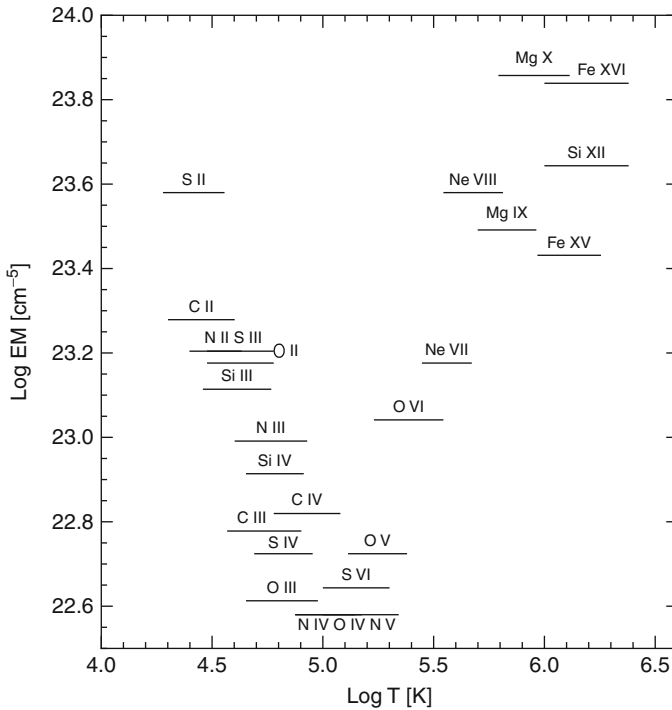
as a function of T_{\max} .

The method was developed to measure the relative abundances of elements contributing to the observed lines. The relative abundances of the elements are derived in order to have all the *line emission measures* of the various ions lie along a common smooth curve. One example is given in [Fig. 3-50](#), obtained from the original data as in Pottasch (1963).

Many other approximations and methods have been proposed. One approximation widely used and originally proposed by Widing and Feldman (1989) was to define a *line differential emission measure* DEM_L value:

$$\text{DEM}_L \equiv \left\langle N_e N_H \frac{dh}{dT} \right\rangle \quad [\text{cm}^{-5} \text{K}^{-1}] \quad (3.19)$$

and adjust the abundances $A(X)$ so that the $A(X) \text{ DEM}_L = I_{\text{ob}} / \int_T G(T) dT$ values form a smooth curve as a function of T_{\max} .

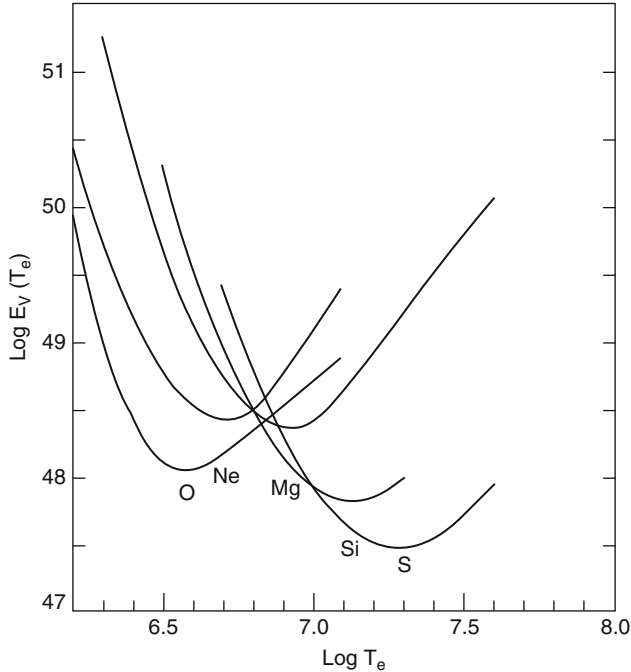


■ Fig. 3-50

The line emission measure values obtained by Pottasch (1963) from irradiance measurements of the quiet Sun [values from the original paper]

The above methods are limited because of the assumption that the bulk of each line is formed at the T_{\max} temperature. This is broadly correct if the plasma is uniformly distributed in temperature, but if not, most of the intensity in each line could come from either lower or higher temperatures. Another problem is that many ions (e.g., He I-like, Li I-like) have broad contribution functions, so are naturally formed over large temperature ranges. The major problem, however, lies in the assumption that the plasma is uniformly distributed in temperature. There are many cases on the Sun (e.g., coronal loops, coronal hole plumes) where the distribution is actually nearly isothermal. In these case, it has been shown by Del Zanna et al. (2001), Del Zanna (2003), and Del Zanna et al. (2003) that the method seriously overestimates the relative chemical abundances.

Another way to estimate the $EM(T)$ is the *Emission Measure Loci* method, by which the loci of the curves $I_{\text{ob}}/(A(X) \times G(T))$ as a function of temperature are considered. In fact, for each line and temperature T_* , the value $I_{\text{ob}}/(A(X) \times G(T_*))$ represents an upper limit to the value of the emission measure $EM(T_*)$ at that temperature, assuming that all the observed emission I_{ob} is produced by an isothermal plasma at the temperature T_* . The EM Loci method was first introduced by Strong (1978) and later applied by Veck et al. (1984) to the analysis of solar X-ray flare spectra (see ► Fig. 3-51) and by Del Zanna and Mason (2003) to solar active regions.



■ Fig. 3-51

The EM Loci method used by Veck et al. (1984)

6.7 A Word of Warning: Problems with Emission Measures

The Li- and Na-like ions give rise to some of the strongest lines in the UV. However, it has long been known (see, e.g., Burton et al. 1971) that the emission measures for lines from ions of the Li- and Na-like isoelectronic sequences are at odds with those of the others. This fact has been largely neglected or underestimated in the literature.

Del Zanna (1999) reanalyzed Skylab data using up-to-date atomic data to show the extent of this anomalous behavior. Li-like N V and C IV are underestimated by factors of 3 and 10, while those of Ne VIII and Mg X are overestimated by factors of 5 and 10, respectively.

Del Zanna et al. (2002b) showed, for the first time, that the same problem occurs when stellar emission measures are considered (see ● Fig. 3-52). This was only possible by combining UV observations from different instruments (FUSE, HST/STIS). All previous literature considered stellar observations at wavelengths where by far the brightest lines are those from Li- and Na-like ions (e.g., IUE and HST/STIS). As a consequence, results in previous literature should be revised.

6.8 Electron Density Determination

The electron (number) density can be defined as the ratio of the emission measure (EM) to the volume (V) of the emitting plasma: $N_e^2 = EM/V$. This approximation is reasonable under the assumption of plasma being distributed homogeneously throughout the emitting volume. To determine the electron density using this method, assumptions about the element abundance and ionization balance are also needed.

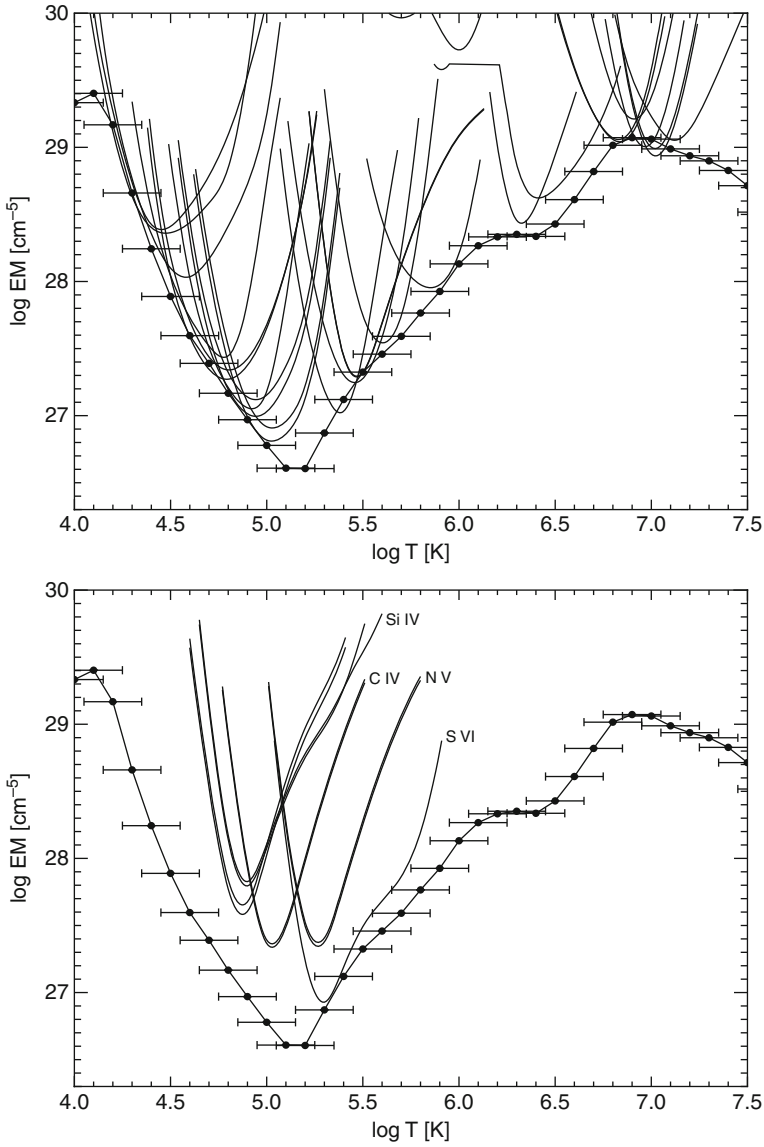
A more reliable diagnostic method for measuring density is the use of density-dependent line ratios from a single ion which has *metastable levels*. A metastable level, m , within an ion is defined as a level that has a small radiative decay rate, so that collisional de-excitation competes with radiative decay as a depopulating process. Metastable levels could for example be excited levels within the ground configuration which can only decay via forbidden transitions. The ion population is shifted from the ground level into the metastable(s) as the electron density of the plasma increases. Therefore, two lines principally excited from the ground and the metastable levels, respectively, have different dependencies on density.

For simplicity we again consider a two-level model. For forbidden and intersystem transitions the radiative decay rate is generally very small ($A_{m,g} \simeq 10^0 - 10^2 \text{ sec}^{-1}$), and collisional de-excitation then becomes an important depopulating mechanism ($A_{m,g} \simeq N_e C_{m,g}^e$) and may even be the dominant mechanism; moreover the population of the metastable level becomes comparable with the population of the ground level and we have:

$$N_m = \frac{N_g N_e C_{g,m}^e}{N_e C_{m,g}^e + A_{m,g}} \quad (3.20)$$

For small electron densities, $N_e \rightarrow 0$, $A_{m,g} \gg N_e C_{m,g}^e$, then the intensity has the same dependence on the density as an allowed line (I^A):

$$I_{m,g} \simeq N_e^2 \quad (3.21)$$



■ Fig. 3-52

The emission measure obtained by Del Zanna et al. (2002b) from irradiance measurements of the dMe star AU Mic. The dots indicate the emission measure calculated over a temperature interval $\Delta \log T = 0.3$. The curves are the $I_{\text{ob}}/(A(X) \times G(T))$ values, obtained from the FUSE, HST/STIS lines used for the emission measure calculation (Top) and from the lines of the “anomalous” ions, from the Li- and Na-like isoelectronic sequences (Bottom)

For very large values of the electron density, $N_e \rightarrow \infty$, the collisional depopulation dominates, $N_e C_{m,g}^e \gg A_{m,g}$; the metastable level is in Boltzmann equilibrium with the ground level:

$$\frac{N_m}{N_g} = \frac{C_{g,m}^e}{C_{m,g}^e} = \frac{\omega_m}{\omega_g} \exp\left(\frac{-\Delta E_{g,m}}{kT}\right) \quad (3.22)$$

And the line intensity has the form:

$$I_{m,g} \simeq N_e \quad (3.23)$$

For intermediate values of the electron density $A_{m,g} \simeq N_e C_{m,g}^e$, the population of the metastable level is significant and the intensity varies as

$$I_{m,g} \simeq N_e^\beta \quad 1 < \beta < 2 \quad (3.24)$$

The intensity ratio of a forbidden to an allowed transition (I^F/I^A) for different spectral lines from the same ion can be used to determine an average electron density for the emitting volume. This value is independent of the elemental abundance, ionization ratio, and any assumptions about the size of that volume.

If the population of a metastable level (m) is comparable with the ground level (g), then other excited levels (k) can be populated from this metastable level as well as from the ground level and the dependence of the intensity on electron density becomes

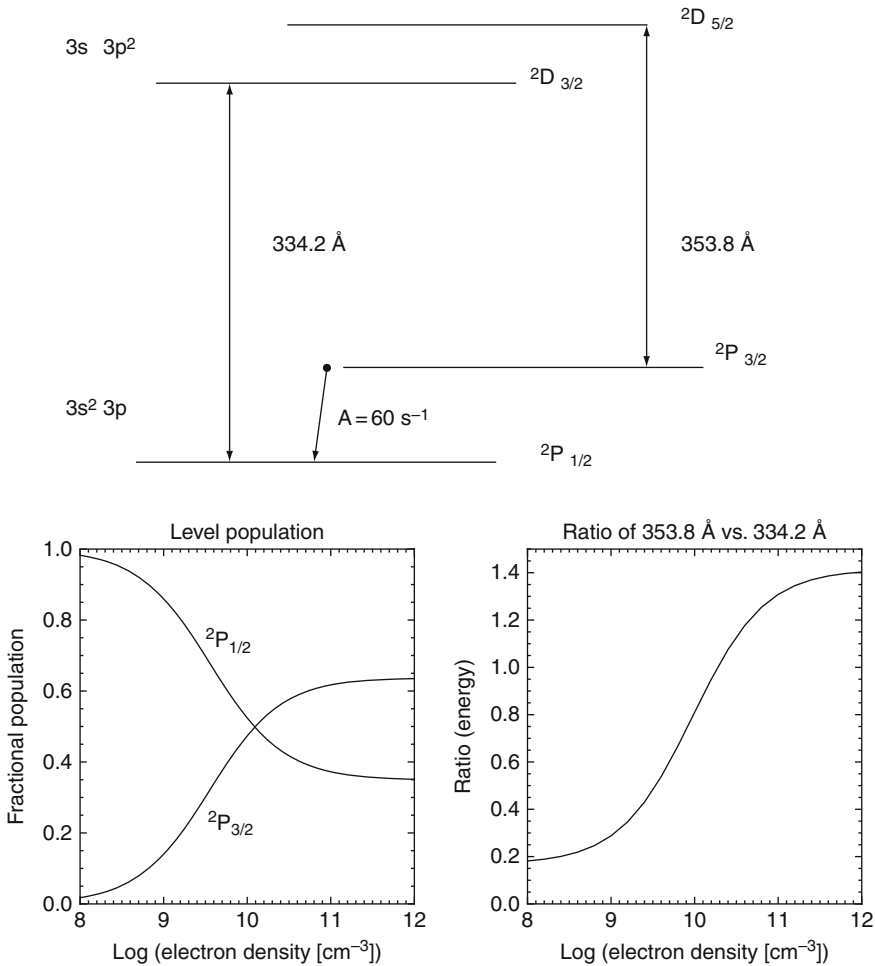
$$I_{k,m} \simeq N_e^\beta \quad 2 < \beta < 3 \quad (3.25)$$

Taking Fe XIV as an example (☉ Fig. 3-53), the ground configuration $3s^2 3p$ has two levels: $^2P_{1/2}$ and $^2P_{3/2}$; the transition between these two levels gives rise to the coronal green line at 5303 \AA . The metastable level is actually the upper level in the ground configuration ($3s^2 3p^2 P_{3/2}$). Its transition probability to the ground level is in fact only 60 s^{-1} . For low electron densities (10^8 cm^{-3}), almost all the population for Fe XIV is in the ground level, but as electron density is increased (10^{10} cm^{-3}), the upper level begins to have a significant population as well. The spectral line at 334.2 \AA is excited from the ground level ($3s^2 3p^2 P_{1/2}$), whereas the line at 353.8 \AA is excited from the upper level in the ground configuration ($3s^2 3p^2 P_{3/2}$). The intensity ratio of these two lines (☉ Fig. 3-53) varies with electron density, reflecting the level population changes in the ground configuration. For active region observations, the $[353.8/334.2]$ ratio gives electron densities varying from around 10^9 cm^{-3} in the quiet part to around $10^{10.5} \text{ cm}^{-3}$ in the core (Mason et al. 1999).

The CDS and SUMER instruments on board SoHO have provided a wide range of density diagnostics, but density measurements from these two instruments are generally only accurate to the 30–40% level at best. The high sensitivity of the EIS instrument on board Hinode now allows electron number density measurements to an unprecedented precision of up to 5% in active regions.

6.9 Spectroscopic Filling Factors

If one assumes a homogeneous slab of plasma of thickness dh , once elemental abundances are known, and a line intensity measured, one can obtain an average $EM_L = \langle N_e^2 \rangle dh$ from



■ Fig. 3-53

Electron density sensitivity for a pair of Fe XIV lines

which the path length dh can be estimated once the electron density N_e is measured from, e.g., a line ratio. Dere et al. (1987) used HRTS transition region C IV intensities and densities form a line ratio of O IV (both formed at around 10^5 K) to obtain path lengths of 0.1–10 km, which was much smaller than the observed sizes of the spicular structures (2,400 km). This is often interpreted as due to the transition region having a filamentary structure, so that most of the plasma occupies only a small fraction (0.01–0.00001 – the filling factor) of the observed volume. However, translating the spectroscopic measurements into a geometrical filling factors is not straightforward (see, e.g., Judge 2000).

6.10 Electron Temperature Determination

One way to measure the electron temperature is to consider the intensity ratio of two allowed lines excited from the ground level, g having significantly different excitation energies.

The intensity ratio can be written as

$$\frac{I_{g,j}}{I_{g,k}} = \frac{\Delta E_{g,j} \Upsilon_{g,j}}{\Delta E_{g,k} \Upsilon_{g,k}} \exp \left[\frac{\Delta E_{g,k} - \Delta E_{g,j}}{k_B T} \right], \quad (3.26)$$

where j and k denote the excited levels; Υ is the thermally averaged collision strength, and k_B is the Boltzmann constant. The ratio becomes sensitive to changes in the temperature if the thermal energy of the electrons is much smaller than the difference between the excitation energies

$$\frac{\Delta E_{g,k} - \Delta E_{g,j}}{k_B T} \gg 1 \quad (3.27)$$

assuming that the lines are emitted by the same isothermal volume with the same electron density.

Such spectral lines are normally far apart in wavelength and it is often necessary to use observations from different instruments, which can lead to inaccuracies (e.g., O VI 1032, 173 Å lines observed by SoHO SUMER and CDS/GIS). Be-like ions offer a good opportunity for diagnostics, as the ratios of the strong resonance and intercombination lines are good temperature indicators. Of all the Be-like ions, Mg IX is that one formed closest to the average coronal T (1 MK). The atomic structure of Be-like ions is such that the resonance line in second order is very close to the intercombination line. For Mg, the resonance 368.07 Å line has been routinely observed in second order by the SoHO/GIS close to the 706.06 Å intercombination line. The 368.07, 706.06, and 749.55 Å lines are basically density-insensitive but have a different temperature dependence. Therefore, any combination of two of these three lines is a good temperature diagnostic. The 706.06, and 749.55 Å lines have been observed with SUMER and used as a temperature diagnostic in coronal holes. Del Zanna et al. (2008) have calculated the atomic data for Mg IX, and revised previous electron temperature measurements.

6.11 Continuum

The continuum emission can be significant in the UV, but diminishes in the EUV. *Free-free* emission is produced when an electron interacts with a charged particle X^{+m+1} and makes a transition from an energy E to an energy E' releasing a photon of energy $h\nu$: $X^{+m+1} + e(E) \Rightarrow X^{+m+1} + e(E') + h\nu$. For a Maxwellian velocity distribution of electrons the process is called *thermal bremsstrahlung*. The formulation for the emitted energy per unit of time and volume is based on the mean Gaunt factors computed by Karzas and Latter (1961). Free-free emission is the main radiative loss mechanism for low-density plasmas at $T > 10^7$ K.

Free-bound emission is produced when a free electron of energy E is captured by an ion (X^{+m+1}) into a bound state of X^{+m} :

$$X^{+m+1} + e(E) \Rightarrow X_n^{+m} + h\nu \quad (3.28)$$

a photon of energy $h\nu = E + I_n$ is emitted and I_n is the ionization energy of the bound state n . For a Maxwellian electron velocity distribution, the continuum emission is characterized by discontinuities at the ionization thresholds. The process is called *radiative recombination*. The free-bound emission depends on the chemical abundances.

For hydrogen and helium-like ions, another continuum (*two-photon*) is important, in particular at $T \leq 3 \cdot 10^4$ K. The metastable 2s state of hydrogenic and helium-like ions depopulates via collisional excitation to the 2p level and by the two-photon decay process.

6.12 Line Widths, Ion Temperatures, and Doppler Motions

Spectral line profiles give information about the dynamic nature of the solar and stellar atmospheres. The Dopplershift $\Delta\lambda$ in a spectral line is directly proportional to the line-of-sight velocity v of the plasma element emitting the line: $\Delta\lambda/\lambda = v/c$, where c is the speed of light. A blueshift indicates plasma moving toward the observer. Naturally, different plasma velocities along the line of sight can be present. These would modify the line profile from the simple thermal broadening of the line.

An emission line has a natural width (called the thermal width) that is characteristic of the temperature of formation of the emitting ion and is due to the Doppler motions of the emitting ions. An additional broadening is also introduced by the observing instrument and depends on the quality and alignment of the optical components. If a line width is found to be broader than the combined thermal and instrumental widths, then we refer to the additional component as the non-thermal width. The line width of an optically thin spectral line has a Gaussian profile and is expressed mathematically in terms of three components.

The intensity per unit wavelength I_λ is defined as

$$I_\lambda = \frac{I}{\sqrt{2\pi}\sigma} \exp[-(\lambda - \lambda_0)^2/2\sigma^2] \quad (3.29)$$

where $I = \int I_\lambda d\lambda$ is the integrated intensity and σ is the Gaussian width given by

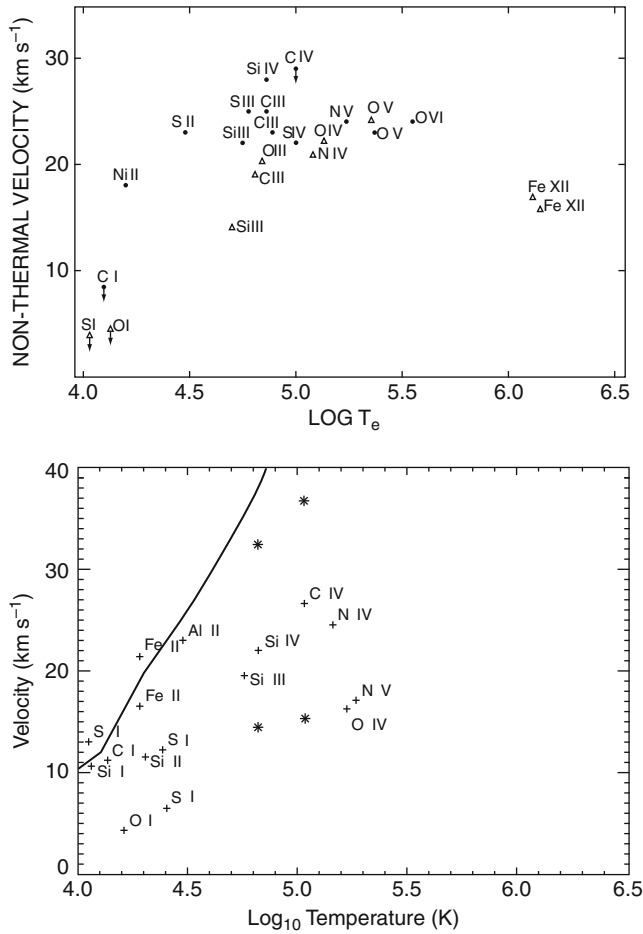
$$\sigma^2 = \frac{\lambda^2}{2c^2} \left(\frac{2kT_i}{M} + \xi^2 \right) + \sigma_I^2 \quad (3.30)$$

for a Maxwellian velocity distribution for the ion emitting the line at a temperature T_i . Here M is the ion mass, σ_I is the Gaussian instrumental width, and ξ is the most probable nonthermal velocity ntv . Nonthermal velocities can be obtained by assuming that T_i is the temperature corresponding to peak abundance of the ion (see [Fig. 3-54](#)). Conversely, if assumptions are made on the nonthermal velocities, the ion temperature T_i can be obtained.

Several coronagraph experiments (e.g., Spartan, SoHO/UVCS) have provided information on hydrogen atom and electron temperatures from H I Ly α profiles (see, e.g., the review of Kohl et al. 2006). The spectral line profiles, at the distances observed by these instruments, have two components. The resonantly scattered, relatively narrow component provides empirical constraints on the electron and proton temperatures. The weaker, electron scattered component is wider and related to the electron density and temperature.

6.13 The CHIANTI Atomic Package

Quantitative modeling of the EUV and soft X-ray plasma emission provides a powerful temperature and density diagnostic of the coronal plasma. However, to calculate optically-thin synthetic spectra and to apply spectral analysis and plasma diagnostics to solar EUV and soft X-ray emission, data from thousands of atomic energy levels are required. For this purpose, the CHIANTI atomic database (Dere et al. 1997), which was first released in 1996, provides a comprehensive dataset covering the UV and X-ray wavelength ranges. The database contains state-of-the-art energy levels, radiative data, and electron and proton collisional data for the majority of astrophysical ions of interest. Furthermore, the CHIANTI codes represent a diagnostics toolbox for astrophysical and solar plasmas. In CHIANTI, the atomic data (energy levels,



■ Fig. 3-54

Nonthermal velocities in the quiet Sun obtained from Skylab (Kjeldseth Moe and Nicolas 1977) and HRTS (Dere and Mason 1993)

wavelengths, radiative transition probabilities and excitation data) are stored in ASCII files. The wavelengths are based on experimental data. Lines for which only theoretical energy levels are available are also provided. The radiative data are taken from published literature and, where necessary, supplemented by new calculations. CHIANTI also includes ancillary data such as standard differential emission measures, elemental abundance, and ionization fraction files.

In the CHIANTI database, the assumption of Maxwellian electron distributions is implicit through the storage of Maxwellian-averaged electron collision strengths. However, recent observations of the small-scale transient solar activity have shown that the solar atmosphere is dynamic and nonthermal (e.g., Dzifčáková et al. 2008). Therefore, to progress our knowledge of the solar atmospheric activity, nonequilibrium and nonthermal spectral plasma emission models are required in the future.

6.14 Benchmarking the Atomic Data

Ongoing work has been carried out to benchmark the atomic data against observations. A large part of the work is done as part of the assessment for each CHIANTI release.

The atomic data, identifications, and diagnostics from the X-rays to the visible for the most important lines on an ion by ion basis have been assessed against laboratory and astrophysical observations in a series of papers published in *Astronomy and Astrophysics*. The assessments include atomic calculations and provide the most accurate identifications to date. The first paper in the series is on Fe x (Del Zanna et al. 2004).

7 Chemical Abundances

An accurate knowledge of the chemical composition of the Sun is of paramount importance for a variety of reasons, the most important one being to obtain an accurate model of the Sun as a Star (the standard model). Chemical abundances have been measured in a variety of ways, from samples of meteoritic materials, to remote-sensing of the photospheric (visible) and coronal (XUV) emission of the Sun, to helioseismology and to in situ measurements of the solar wind and of the solar energetic particles (SEP). Gamma-ray spectroscopy is another powerful tool for measuring the abundances of various elements, as described in the review from Ramaty and Murphy (1987). The drawback is that very few observations of strong flares which produce gamma-rays are available. One was observed on April 27, 1981, and a few have been observed more recently with RHESSI.

The photospheric and coronal abundances of many elements are still uncertain, because they depend critically on the spectral resolution, the method adopted, and the accuracy of atomic physics calculations. Some useful review articles on elemental abundances are Anders and Grevesse (1989), Meyer (1985), Bochsler (2007), Asplund et al. (2009), Lodders et al. (2009). Various review articles can also be found in Wimmer-Schweingruber (2001), the proceedings of a joint SoHO/ACE workshop on solar and galactic composition in 2001.

Elemental abundances, A , are normally provided as number densities relative to hydrogen (often as $\log_{10}(\text{dex})$, i.e., assuming $\log_{10}(A(H)) = 12$). In terms of number densities, the Sun and its neighborhood are composed primarily (90%) of hydrogen, with significant (10%) helium, but with only about 0.1% of the remaining elements. Abundances of Li, Be, and B are very low. The Li abundance is an interesting proxy for estimating the age of a star.

J. von Fraunhofer (1787–1826) was the first to observe the dark lines in the visible spectrum of the Sun. G. Kirchhoff (1824–1887) correctly proposed that these were due to absorption of light by cooler, overlying layers of the solar atmosphere. Russell (1929) produced a pioneering piece of work, establishing the photospheric abundances of a large number of elements from the visible spectrum of the Sun.

Goldschmidt (1938) compiled chemical data for terrestrial rocks and meteorites in a pioneering piece of work. Along with others, he suggested that the abundances of non-volatile elements in the CI carbonaceous chondrites meteorites should be representative of the average values for our solar system. Among the many types of meteorites, the CI carbonaceous chondrites (CC) are thought to originate from asteroids which were formed at the same time as the rest of the planets in the solar system, but that were not subject to the many changes which occurred during the evolution. Indeed, there is a remarkable agreement between CC

and the most up-to-date photospheric abundances of non-volatile elements. A recent review on abundances in chondritic meteorites is given by Lodders et al. (2009).

The solar (photospheric) abundances of non-volatile elements such as Mg, Si, Ca, Fe, Ni, can be established from the absorption lines in the visible spectrum, or obtained from meteoritic samples. The abundances of volatile elements such as H, C, N, O can in principle also be obtained from photospheric spectra, but not from meteoritic samples because they are not fully retained. The abundances of the important noble gases He, Ne, Ar cannot be obtained either way, in the first instance because these elements do not produce usable absorption lines, and in the second case because they are volatile elements.

Detailed examination of the Sun's upper atmosphere reveals that elemental abundances vary from the photospheric values between different solar regions. In addition, measurements of the solar wind composition differ from photospheric values. There is a correlation between the abundance and the first ionization potential (FIP) of the different elements. In particular, the low-FIP (≤ 10 eV) elements are more abundant than high-FIP in the slow solar wind (SSW) and in coronal loop plasma.

7.1 Photospheric Abundances

Many semiempirical 1D models adopting LTE have been developed to study the photosphere. Sophisticated model atmospheres include departures from LTE and, in general, comparisons between observed high-resolution spectra and theoretical ones are very satisfactory.

In recent years, it has become possible to perform 3D time-dependent hydrodynamic simulations of solar convection, where the hydrodynamic equations are solved together with 3D radiative transfer equations (for a review see Nordlund et al. 2009). The calculations are carried out adopting a limited set of opacities when computing the radiative heating/cooling rate in the energy equation. These models successfully reproduce the main features of the solar granulation, in terms of spatial, temporal, and velocity scales. These new models produce significantly lower abundances of carbon, nitrogen, and oxygen, compared with the widely used earlier values.

Oxygen is the most abundant element after H and He. The derived “photospheric” abundance of oxygen has changed from $10^{8.93}$ (Anders and Grevesse 1989) to $10^{8.69}$ (Asplund et al. 2009). The latest result is based on the analysis of forbidden and permitted O I transitions as well as vibration–rotation and pure rotation lines of OH. Various discrepancies that have been present for a long time appear to have been resolved, once severe blending and the new simulations are taken into account.

An up-to-date review on photospheric abundances is given by Asplund et al. (2009), where references to recent work can be found. There is still a debate in the literature about the accuracy of these new abundances. If confirmed, they have far-reaching implications in a variety of fields. As most previous measurements have been carried out *relative* to oxygen, for example, a large part of literature may need to be revised. The revised abundances have important implications for the *standard solar model*, as discussed for example by Basu and Antia (2008).

7.2 Coronal Abundances

For a review of some of the uncertainties related to abundance determinations derived from UV and X-ray spectra, see Mason (1995). A vast literature exists on abundance measurements

using various techniques from the Skylab, Yohkoh, and SMM missions. Only some examples are given here. Pottasch (1963) applied an approximate method (see ► Sect. 6.6) to show that the abundances in the solar corona are different from those of the photosphere. Walker et al. (1974) used the intensities of resonance lines of hydrogen-like and He-like ions in the X-rays to obtain relative chemical abundances for a number of elements. Interestingly, for the active regions and flares that they studied, they found near-photospheric abundances.

An extensive literature based on Skylab EUV observations exists. For example, Widing and Feldman (2001) used Skylab spectroheliographs and measured the rate of change of the radiance ratios of Mg/Ne lines of active regions as they evolved with time, as a proxy for the FIP effect. Interestingly, at birth of an active region, they measured ‘photospheric’ composition, while after 2–3 days after the emergence of the active region, the relative abundance changes by factors of 4–5. However, most results from Skylab need to be revisited, since the Widing and Feldman (1989) approximation (see ► Sect. 6.6) which was used extensively may have some limitations. This approximation tends to overestimate the FIP effect when the plasma distribution is not multi-thermal, as is the case of active region loops (as discussed in Del Zanna 2003). For example, the EM Loci method, applied to a Skylab observation, was shown by Del Zanna (2003) to be consistent with an FIP bias of 3.7, rather than a large FIP bias of 14 obtained previously.

Some results based on EUV SoHO CDS spectra exist, and tend to indicate that the Mg/Ne abundances in quiescent 1 MK loops have near-photospheric abundances (cf. ► Fig. 3-59 and Del Zanna 2003; Del Zanna and Mason 2003). On the other hand, Young and Mason (1997) studied the Mg/Ne abundance ratio in an active region and found that emerging flux regions had photospheric abundances in contrast to longer-lived active region loops which showed a large FIP bias.

UV measurements in the outer corona have been obtained by SoHO UVCS. Raymond et al. (1997) analyzed SoHO UVCS spectra where lines from hydrogen and other elements were observed at a radial distance $R = 1.5 R_{\odot}$. In the core of a quiescent equatorial streamer during solar minimum, high-FIP elements were found to be depleted by an order of magnitude (compared with photospheric abundances), while low-FIP were also depleted by about a factor of 3. Along the edges of a quiescent streamer, and in a streamer associated with an active region, high-FIP elements were found to be depleted by a factor of 3. The abundances along the edges of the streamer legs were similar to those measured in the slow solar wind, which would be an important indication that these are the source regions for the slow wind. Different streamers observed when the Sun was more active, however, have shown different abundances and much reduced depletion.

Feldman et al. (1998) report an analysis of SoHO SUMER spectra very close to the solar limb, at a radial distance $R \leq 1.03 R_{\odot}$ in a quiet Sun region and above the north polar coronal hole. Low-FIP elements were found to be enriched by about a factor of 4 in the corona above the quiet equatorial region, but not in the coronal hole. One significant problem is that *the Ne VIII, Na IX, Mg X lines, all of which belong to the Li-like isoelectronic sequence, appear to indicate a systematic lower effective FIP bias value than the rest of the lines.* The authors used the O VI 1032 Å to normalize all the other lines, and since O VI is also in the same isoelectronic sequence, the results are quite uncertain. The same issue applies to the interpretation of SoHO UVCS spectra, and is related to the problem of anomalous ions in emission measure analyses (see ► Sect. 6.7).

Various spectral measurements in the EUV and UV from the SoHO CDS UVCS and SUMER instruments as a function of height from the solar limb have now indicated the presence of gravitational settling in the sense that emission lines from elements with larger atomic weights such as Fe have radiances which decrease much rapidly than lines which should be formed at similar temperatures but are from lighter elements (e.g., Mg, Si). Feldman et al. (1998)

found evidence of gravitational settling and suggested that the Raymond et al. (1997) results may just be caused by gravitational settling. This is reasonable, if the streamer core is formed of quiescent closed structures so gravitational settling acts in reducing the abundances of the heavier elements, a fact that is not yet well established.

7.3 Abundances from In Situ Measurements

In situ chemical composition measurements are now available for over 25 years, from foil collections during the Apollo missions (and Genesis), to mass spectrometers such as the Solar Wind Ion Composition Spectrometer (SWICS) on Ulysses and CELIAS on SoHO, and more recently from the ACE and the STEREO missions.

In situ observations of the SW and of solar energetic particle (SEP) have shown that elemental abundances are highly variable. SEP events are transient and are now known to be associated with acceleration processes within CME-driven shocks and flares. SEP events are extremely variable and are therefore not suitable for determining the bulk of the solar composition. See Reames (1999) for a review of the characteristics of SEP events.

For many years, in situ measurements have indicated a FIP effect of about 4–5 in the slow SW, while the fast SW does not seem to have a significant FIP effect (as shown e.g., by Ulysses observations of a proxy, the Mg/O ratio). In the slow and fast SW, the He abundance is very variable but on average it is 2–3 times lower than the photospheric value.

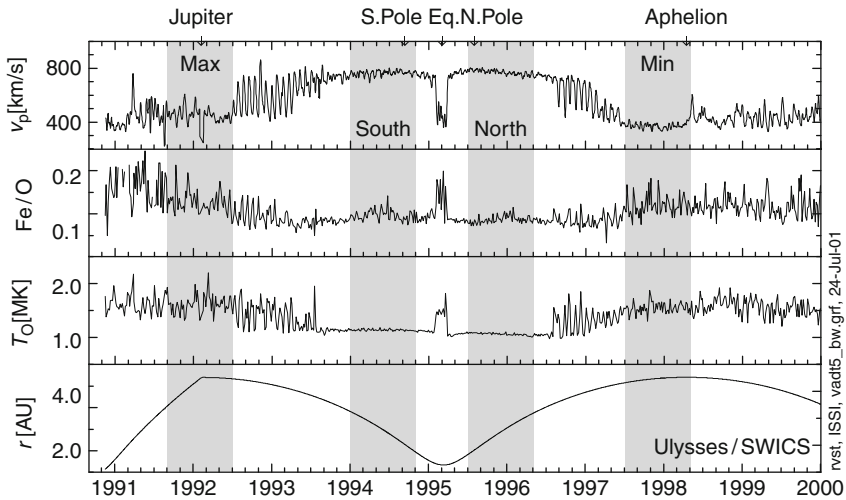
von Steiger et al. (2000) adopted an improved analysis technique to measure the average solar wind charge state and chemical composition of about 40 ion species of He, C, N, O, Ne, Mg, Si, S, and Fe observed with the SWICS on Ulysses. The FIP effect was confirmed, however it was found that on average the differences between slow and fast SW were only about a factor of 3, much lower than what was previously found. [Figure 3-55](#) shows as an example the Fe/O relative abundance as it varies with the solar wind speed. It is evident that during the polar passes at solar minimum (1994, 1996, fast solar wind), the variability is much reduced and the relative abundance of Fe/O is lower.

Genesis was a NASA mission devoted to the collection of solar wind data at L1 from December 2001 to April 2004. Initial results show results consistent with other measurements. Accurate values, in particular for isotopic composition, should soon be available.

7.4 Depletion, Enhancement or Both?

It is still unclear whether the low-FIP elements such as Fe are *enhanced* in the corona, or high-FIP ones such as O are *depleted*, relative to their photospheric values, or if a “hybrid” solution such as that one proposed by Fludra and Schmelz (1999) applies.

Veck and Parkinson (1981) were the first to measure elemental abundances from line-to-continuum X-ray measurements from OSO-8. They found that the low-FIP elements seem to have photospheric values, and that high-FIP elements seem to be *depleted*. Since then, a large body of conflicting evidence on the subject has been presented. However, most direct remote-sensing measurements seem to confirm the early results from Veck and Parkinson (1981). For example, Fludra and Schmelz (1995) applied a DEM and line-to-continuum analysis on flare observations with the SMM Bent Crystal Spectrometer, and found high-FIP elements *depleted*, compared with their photospheric values. Direct measurements of inner-shell Fe lines formed during large solar flares by fluorescence of X-rays exciting photospheric iron, observed



■ Fig. 3-55

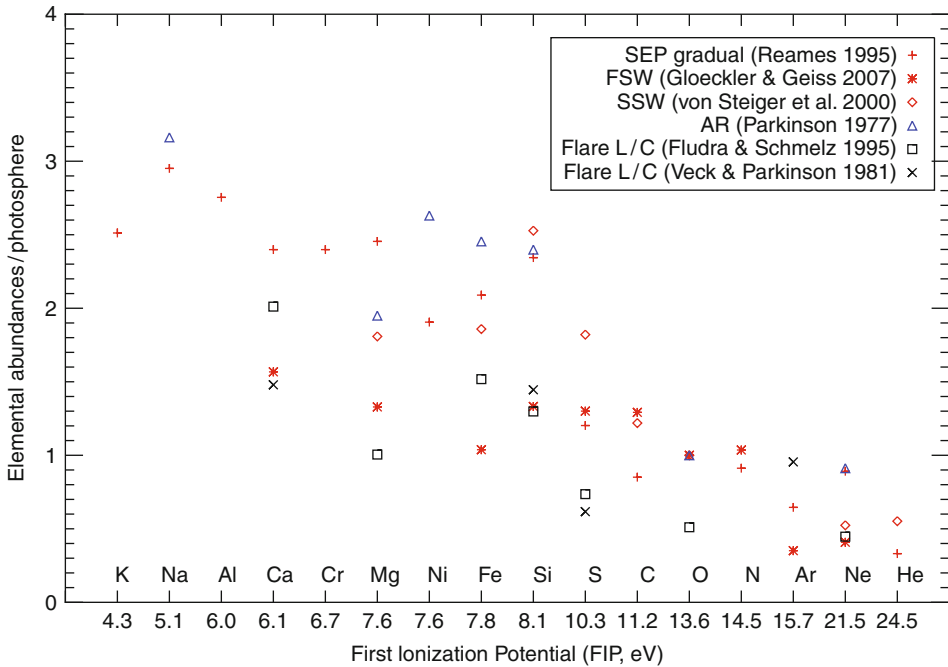
Results from the Solar Wind Ion Composition Spectrometer on Ulysses. From top to bottom: proton velocity, Fe/O relative abundance, freezing-in temperature, and distance of the spacecraft from the Sun. (Figure courtesy of R. von Steiger)

with various crystal spectrometers on Yohkoh, SMM, and P78-1 are consistent with the coronal Fe abundance being equal to the photospheric value, within a factor of 2 (Phillips et al. 1995). The RESIK crystal spectrometer on board the CORONAS-F spacecraft has also allowed line-to-continuum analyses of solar flares in the X-rays. For example, Chifor et al. (2007) found near-photospheric abundances for K, Si, and Ar and S abundances consistent with older measurements from SMM.

► **Figure 3-56** shows a sample of a few abundance measurements, some in situ (in red), and some from remote sensing. For the in situ, we show measurements of Solar Energetic Particles (SEP) gradual events, which are thought to be associated with ejecta from the solar corona. We also show two average Ulysses solar wind measurements, one from the fast solar wind (FSW) from Gloeckler and Geiss (2007), and one from the slow solar wind (SSW) during maximum (von Steiger et al. 2000). All these in situ measurements were relative ones, and are plotted here taking the Oxygen abundance as a reference. The Figure also shows two measurements from X-ray observations of active regions (AR), again rescaled to the Oxygen abundance. Finally, we have added two sets of “absolute” abundances obtained from X-ray spectra of flares, and a line-to-continuum (L/C) analysis. All the measurements are plotted as a ratio relative to the “photospheric” abundances of Asplund et al. (2009). There is a clear trend for low-FIP elements to be overabundant and for high-FIP ones to be underabundant; however the scatter, mostly due to real solar variability, is large.

7.5 Helium, Neon, and Argon

Helium chromospheric lines were observed by Sir Norman Lockyer during a solar eclipse in the nineteen century, at a time when the element was still unknown. After H, neutral and ionized



■ Fig. 3-56

Plot of coronal/photospheric elemental abundances as a function of element, ordered by the first ionization potential. In situ measurements (in red), remote sensing measurements (in blue and black)

He lines are the strongest lines in the XUV. The Coronal Helium Abundance Spacelab Experiment (CHASE) on the Spacelab 2 Mission was specifically designed to determine the helium abundance from the ratio of the scattered He II 304 Å to Lyman- α 1218 Å, on the disk and off the limb. A value for the He/H abundance of $7.0 \pm 1\%$ for the quiet corona was deduced by Gabriel et al. (1995).

Helium abundance measurements have also been indirectly obtained from helioseismology. A variety of measurements exist in the literature. For example, Basu and Antia (2004) obtained a value of 10.93 dex, equivalent to 8.5% by number, using OPAL opacities. This is close to the value of 8.3% predicted from standard Big Bang nucleosynthesis theory.

The modeling for all He lines needs detailed radiative transfer calculations, taking into account many effects, such as photon scattering, or the so-called photoionization–recombination mechanism, whereby coronal radiation can be a significant source of photoionization for neutral helium. Another complexity relates to the fact that all He lines have high excitation energies, which makes them very sensitive to non-thermal electron velocity distributions and other non-equilibrium effects. For a discussion of some of these effects, see e.g., Andretta et al. (2003), Jordan et al. (2005). In general, the EUV radiances in quiet Sun for all He lines are much stronger than realistic models predict.

Andretta et al. (2008) devised a new technique which involved modeling and observations from SoHO instruments of a C-class flare to obtain the first accurate measurement of the He abundance in the chromosphere, $7.5 \pm 1\%$, a value in marginal agreement with the helioseismic value.

In situ measurements of the relative abundance of helium to hydrogen show a very low abundance, less than 5%, with a large variability which is partly correlated with the solar wind speed and partly with the solar activity and cycle.

Neon and argon photospheric abundances are not directly measured, but have been inferred relative to oxygen or other elements from XUV spectra. They are also measured in situ. Gloeckler and Geiss (2007) suggest that Ulysses SW observations of Ne and Ar can be used to provide reliable “photospheric” abundances. They suggest a Ne/O abundance of 0.078 and an Ar/O one of 0.0018.

7.6 Modeling

The presence of an FIP effect implies an ion-neutral fractionation, which probably takes place low down in the solar atmosphere, possibly in the chromosphere at $T \leq 10^4$ K, where the ions start to become ionized. The ionization is mainly by the coronal EUV photons, but the low-FIP elements are more easily ionized by electron collisions. It was proposed some time ago (see, e.g., Geiss 1982) that once the fractionation has taken place in the chromosphere, the ions then drift up into the corona, guided by the magnetic field, to become ions of higher stages. The amount of the FIP effect would depend on how ions drift up into the corona, hence on the topology of the magnetic field. The key factor is the Coulomb drag of the ions with the protons. Low-FIP elements are more easily ionized and have a stronger coupling with the proton flow.

The issue is complex, and there are various physical processes at work that need to be studied in detail. On the one hand there are the ionizing/recombining processes. In the chromosphere, the H I L_α is by far the strongest contributor to ionizing UV photons, and is able to ionize all neutrals from elements with FIP ≤ 13.6 eV. Then, there are the collisions between all the particles, which produce charge exchange and diffusion, for example. In addition, gravitational settling is also present. In hydrostatic equilibrium, each ion is distributed in the corona with a pressure scale height which is proportional to its mass. Heavier ions tend to accumulate at lower heights, unless mixing due to e.g., flows is present.

Thermal diffusion is also strong for ions (see, e.g., Geiss and Buergi 1986), and occurs in the presence of a steep temperature gradient. This drags ions of heavier elements toward higher temperatures. In the transition region, temperature gradients are large and positive, and thermal diffusion acts in the opposite direction to gravitational settling.

Many models, based on first principles, have been proposed. Various authors followed von Steiger and Geiss (1989) by assuming that diffusion processes in the upper chromosphere are the main factors. Recently, Laming (2004) have presented an unified picture of the FIP and inverse FIP effects based on the action of the ponderomotive force of Alfvén waves in the chromosphere. The model simulations reproduce some of the main observational features (i.e., overabundance of low-FIP and underabundance of high-FIP), which is very encouraging.

7.7 Stellar and Galactic Abundances

The solar metallicity is quite typical when compared to nearby stars of similar spectral type (G). Accurate measurements have recently been obtained as a by-product of the search of extrasolar planets. It turns out that the majority of planet-bearing stars are metal-rich by 0.1–0.3 dex compared to the Sun, which in turn is metal-rich by about 0.2 dex compared to stars with no

planets (see, e.g., Butler et al. 2000). A recent compilation from Fuhrmann (2008) of nearby stars (within 25 pc) confirms that the Sun has an average metallicity.

When chemical abundances within our Galaxy are studied, a large scatter is normally found, but also an overall gradient, as a function of galactocentric distance. The latest measurements indicate that the Sun has quite typical photospheric abundances. This can be seen e.g., in Hou et al. (2000), with recently compiled chemical abundances obtained from H II regions, planetary nebulae, and B stars.

In contrast, the situation regarding stellar coronae is not clear. Some early measurements from the EUVE satellites provided chemical abundances for a few stars which showed a solar-like FIP effect (see, e.g. Drake et al. 1997), but it has only been since the launch of the *XMM-Newton*, *Chandra* satellites in 1999 that systematic measurements in the X-rays have been obtained. Some line-to-continuum measurements have been obtained, however in the majority of cases simple emission measure analyses have provided relative abundances between low-FIP elements (mainly iron) and high-FIP ones (mainly oxygen, neon). A complex picture has emerged, with many active stars and binaries showing an *inverse FIP effect*, with low-FIP elements being a lot more depleted than the high-FIP ones (both compared to hydrogen). This effect becomes more prominent with increasing activity of the star (see, e.g., Güdel and Nazé 2009 and references therein). These results have been puzzling, however some caveats apply. First, measurements of photospheric abundances have often been lacking, due to the strong stellar rotation which broadens the spectral lines. It is often assumed that stars have the same photospheric composition as the Sun. In only a few cases the inverse FIP effect seems to be confirmed. Second, the emission measure methods provide meaningful results only if the emission in lines emitted by different ions/elements is co-spatial, something very difficult to ascertain. Third, density effects and departures from ionization equilibrium may play a role.

8 The Solar Corona

8.1 Solar Features as Seen in X-rays and EUV

The most dramatic change in our perception of the solar atmosphere in the past few decades is due to X-ray and EUV observations from space. It is now evident how dynamic and active the solar atmosphere is, both on small and large temporal and spatial scales. Early concepts of relatively uniform, plane parallel models for the solar atmosphere are no longer sustainable. The structured and dynamic nature of the solar corona was first evident from the Skylab X-ray telescope observations. Spectacular observations were then made over an extended period with YOHKOH/SXT, showing dramatic changes from solar minimum to solar maximum, later shown in more detail with SoHO EIT and CDS observations (see ► Sect. 11.3.1).

The SoHO/EIT observations showed the dynamic nature of the solar corona in the EUV, while the SoHO/MDI observations showed changes in the photospheric magnetic field over short timescales, even in the quiet Sun regions. TRACE has provided some stunning EUV images and movies of solar features, especially active regions.


When viewed in X-ray emission (► Fig. 3-6) the structures on the Sun can clearly be divided into different characteristics. Coronal hole areas are dark with quite well-defined boundaries. This occurs because the averaged density is about half the density of the averaged quiet regions, and because there is hardly any plasma at temperatures above 1 MK (see, e.g.,

results from SoHO CDS: Del Zanna and Bromage 1999). During solar minimum coronal holes are apparent at the north and south poles, but during solar maximum, coronal holes frequently extend down to lower latitudes. Coronal holes represent areas in the solar atmosphere where the magnetic field is predominantly unipolar and is “open”, that is goes out into the heliosphere. The fast coronal wind flows from these regions. Coronal holes contain sub-features, long plumes of bright emission, bright points (small bipolar regions), and dynamic jets.

The quiet sun is the more diffuse area of weak X-ray emission. On closer inspection, even the quiet sun regions are covered by a multitude of small transient brightening (and small bipolar regions), due to the constant motion of the photospheric footpoints of the magnetic field. Other solar features include prominences, seen in the quiet Sun and active regions. These can sometimes erupt and can be associated with Coronal Mass Ejections (CMEs) which shoot out from the Sun.

In this section we discuss the characteristics and properties of the solar corona, in particular with regard to active regions. We pay particular regard to the measurements of plasma properties: temperature, electron density, flows etc, which are an essential requirement to constrain heating models.

8.2 Solar Active Regions

Solar active regions (ARs) are areas of enhanced UV and X-ray emission that lie directly above areas of enhanced magnetic field, sometimes with associated sunspots. The heliographic position of active regions is typically confined within latitudes of $\pm 40^\circ$ from the solar equator, as is the case of the sunspots. ARs typically exhibit a strongly concentrated leading magnetic polarity, followed by a more fragmented trailing group of opposite polarity. Because of this bipolar nature, active regions are mainly made up of closed magnetic field lines. Due to continuous magnetic flux emergence, and reconfiguration, a number of dynamic processes such as plasma heating, flares, and coronal mass ejections occur in active regions. Active regions have the familiar appearance of numerous filled loops, which are hotter and denser than the background corona, producing bright emission in soft X-rays and EUV wavelengths (see  Fig. 3-6). It is normally accepted that loops trace out the magnetic field structure in the corona.

The Monograph from the Skylab Workshop III on Solar Active Regions (Orrall 1981) is a good starting point for the study of solar active regions. In particular, the S082A instrument provided some good observations of active regions and flares. The S082A observations indicated that the *hot core loops* (seen e.g., in Fe xvi and X-rays) were mostly low lying, compact and closely packed (unresolved). They seemed to connect across the magnetic neutral line. The cooler *warm loops* (seen in Ne vii) seemed better defined, were larger and slender. There were also many bright patches of emission or *kernels* in strong magnetic regions on either side of the neutral line (corresponding to the bright chromospheric plage areas). These kernels could be identified as the footpoints of the Fe xvi and X-ray loops. Dere and Mason (1981) provided a summary of electron density values derived for active regions, which ranged from $0.7\text{--}6.0 \times 10^9 \text{ cm}^{-3}$ for a temperature range of 1.2–2 MK.

After Skylab, the YOHKOH/SXT instrument was widely used to study the plasma properties of active regions and solar flares, as well as SoHO/EIT, TRACE, and other broadband imagers in the X-rays and EUV.

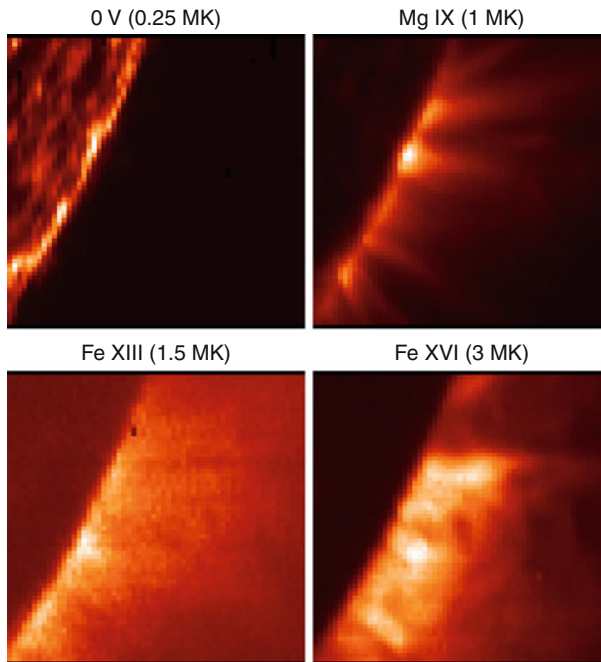
A specific type of emission known as *active region moss* was clearly visible in the 173 Å passband of TRACE. Although this type of emission was mentioned in early observations taken

by Skylab, TRACE provided the first detailed observations of moss. It is currently thought that the moss emission corresponds to the footpoints of the hot loops seen e.g., in the Yohkoh/SXT images (see, e.g., Martens et al. 2000).

8.2.1 Active Regions Observed with SoHO CDS


Monochromatic imaging in spectral lines emitted in a wide range of temperatures became routine with the SoHO/CDS instrument. Active regions appear very different at different temperatures. Mason et al. (1999) studied an active region at the limb using CDS observations. The core of the active region was found to be very hot in comparison to the surroundings. The density map of the active region derived from Si x line ratio shows that the density is highest (greater than $2.3 \times 10^9 \text{ cm}^{-3}$) in the hot core of the active region. Further studies with CDS, and later Hinode/EIS gave similar results.


Del Zanna and Mason (2003) used both imaging (TRACE) and spectroscopic (CDS) observations to show that the dominant EUV emission in active regions is multi-thermal and unresolved even at $1''$ resolution (TRACE). Its peak emission is around 1.5 MK where Fe XIII is formed (cf. ► Fig. 3-57). This diffuse emission forms a background and foreground which needs to be taken into account when analyzing loop structures.




■ Fig. 3-57


Monochromatic images of an active region observed by SoHO/CDS at the limb (Adapted from Del Zanna and Mason 2003). Notice the limb brightening in O V, the warm loops seen around 1 MK, the unresolved emission at 1.5 MK, and the bright hot (3 MK) loops in the core

The SoHO/CDS instrument also allowed measurements of EM and DEM from the observation of several spectral lines. These can be compared with simulated emission measures derived from theoretical models.  [Figure 3-58](#) shows a sample DEM for an active region. The peak emission is at $\log T[K] \simeq 6.3$, (2 MK) compared to the quiet Sun which has peak emission at about 1.5 MK, and coronal holes where there is hardly any emission above 1 MK.


The large, quiescent 1 MK loops are best seen in the TRACE 173 Å band, sensitive to Fe IX/Fe X emission, and a considerable number of studies has been published. Early results from TRACE indicated that these warm loops were almost isothermal along their length and “over dense”, compared to what expected in hydrostatic equilibrium. However, TRACE measurements of temperature and density along and across coronal loops have severe limitations, which only spectroscopic measurements can resolve. SoHO/CDS showed that these quiescent warm (1 MK) coronal loops are nearly isothermal at each location along their length (e.g., Del Zanna 2003, see  [Fig. 3-59](#)), but are not isothermal along their length, something later confirmed with Hinode/EIS observations (Tripathi et al. 2009). With CDS, it was also possible to follow the coronal loops down to the chromosphere (Del Zanna 2003).

SoHO/CDS also showed (Del Zanna et al. 2006b) that hot and cool loops are intermingled and tend to keep their characteristics, in particular the temperature, for a very long time ( [Fig. 3-60](#)). In essence, most structures must continuously be heated in a steady way to maintain their temperatures. These results radically changed our knowledge of loops, given that it was widely thought that cooler loops naturally form via cooling of the hotter ones.

8.2.2 Active Regions Observed with Hinode EIS

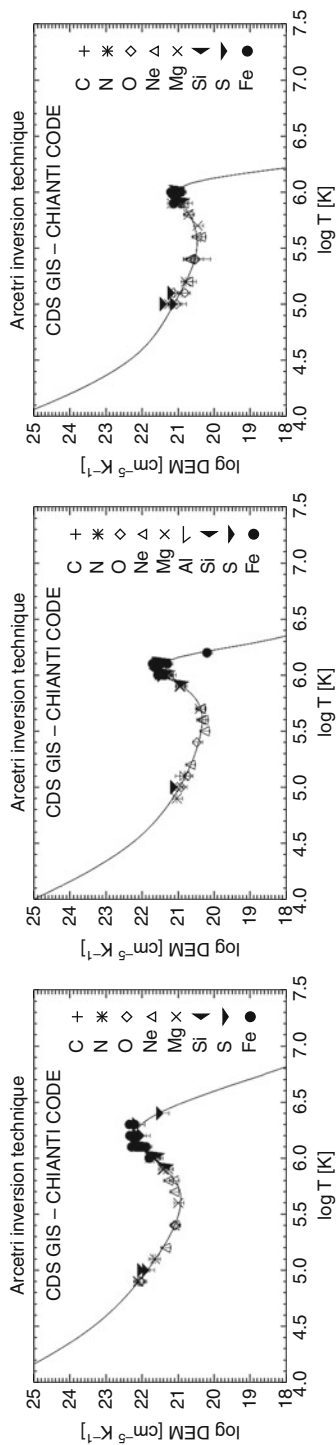
A major advantage of the Hinode/EIS instrument over previous ones is the ability to measure Dopplershifts and line widths in coronal lines emitted above 1 MK. New features have emerged. EIS observations of active regions have shown a clear pattern of blueshifts (Del Zanna 2007, 2008; Doschek et al. 2007), increasing at higher temperatures. The blueshifted areas are always in low-density, low-emission regions, and have associated large line widths, indicating that they are places where strong heating is taking place.  [Figure 3-61](#) shows one example, from Del Zanna (2008).

Some signatures of blueshifted emission in “coronal” ($T > 1$ MK) lines were previously obtained from SMM/UVSP, SoHO/CDS, and SUMER, but most observations had a poor spatial coverage and/or resolution. Perhaps surprisingly, loops tend to have smaller line widths.

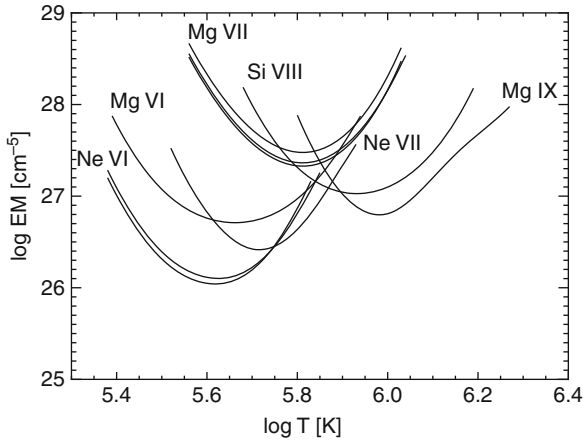
Hinode/EIS has also shown that almost all active region loops exhibit strong redshifts in both legs, with increasing values toward the footpoints, in progressively cooler emission lines (Del Zanna 2007, 2008), as described in  [Sect. 4.3](#).

8.3 Coronal Heating

The amount of energy required to sustain and heat the corona is very small, compared to e.g., the total energy radiated from the Sun. However, a well-accepted explanation of how the corona is heated is still lacking. Different coronal features might also be heated by different processes. A great variety of mechanisms for coronal heating have been proposed over the last 70 years. Most theories can be grouped into two categories, one hydrodynamic and the other magnetic. In the first one, purely hydrodynamic effects are important, for example, heating by

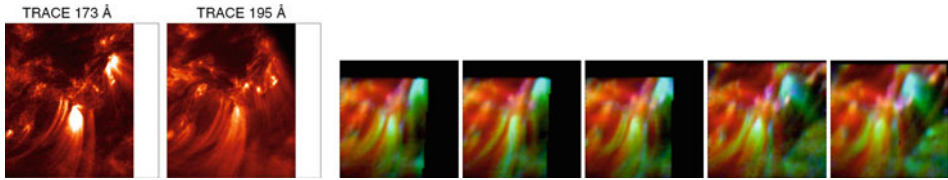


■ Fig. 3-58 DEMs for an active region (*left*), quiet Sun (*centre*) and coronal hole (*right*), obtained from SoHO/ GIS observations (Del Zanna 1999)



■ Fig. 3-59

EM loci curves from SoHO/CDS observations of a warm active region loop, showing nearly isothermal plasma (Del Zanna 2003) and photospheric Mg/Ne abundances

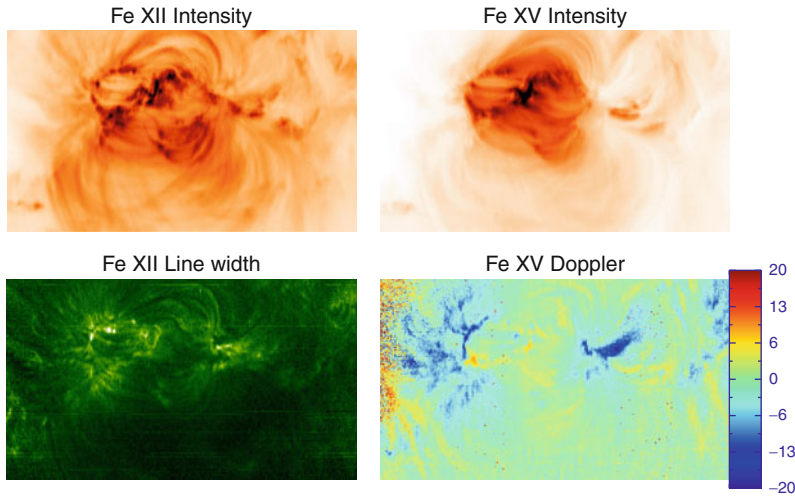


■ Fig. 3-60

From the *left*: TRACE 173 Å and 195 Å EUV images of the legs of warm (1 MK) loops; false-color images obtained from monochromatic CDS images (Ne VII, 0.7 MK, in blue; Ca X, 1 MK, in green; and Si XII, 3 MK, in red over a long time period, 12 h (Figures adapted from Del Zanna et al. 2006b)). The warm loops (*green*) are clearly seen in TRACE; however there are cooler (*blue*) and hotter (*red*) loops as well. Cooler and hotter loops are intermingled and fill the entire AR volume. They are also nearly isothermal at each location

the acoustic waves which are observed in the photosphere and chromosphere. In the second one, energy is channeled via the magnetic fields, for example via microflares, current dissipation, or magnetic wave dissipation. Within the above categories, theories can be further grouped into DC and AC processes. With the AC processes, waves are generated and then dissipated to heat the plasma. With the DC mechanism, stresses in the magnetic structures are slowly built up, until a point is reached when instability occurs and a relaxation process releases the energy. For recent reviews on the subject, see Narain and Ulmschneider (1996), Klimchuk (2006), Erdélyi and Ballai (2007).

Current efforts are both on the observational side, to measure plasma properties, and on the theoretical side. Modeling entire active regions is computationally very intensive and complex. Three-dimensional MHD codes have begun to produce interesting results, however they still



■ Fig. 3-61

Hinode EIS observations of an active region. *Left*: monochromatic (*negative*) image in Fe XII (*top*) and (*below*) line width. *Right*: monochromatic (*negative*) image in Fe XV (*top*) and (*below*) a Dopplergram (± 20 km/s), showing strong blueshifts in regions of low emission outside the hot loops in the core of the active region (Adapted from Del Zanna 2008). The field of view is $400'' \times 230''$

have inadequate spatial resolution, in particular to resolve the transition region, where steep gradients in density and temperature occur.

8.3.1 Coronal Heating in Loops

Active region loops are the brightest and geometrically simplest structures observed in the solar corona. They have been studied extensively, in an attempt to better define the coronal heating mechanism. It is well established that loops are arch-like structures of nearly constant cross section which are connected to the photosphere where strong magnetic fragments of opposite polarity are located. The pressure exerted by the magnetic field present in the corona is much stronger than the dynamic pressure of the plasma, so loops can be considered as magnetic flux tubes. In these conditions, the magnetic field is known to be *frozen in* with the plasma, i.e., magnetic fields always move with the plasma. Vice versa, plasma elements can freely move along field lines but not across. It is also known that thermal conduction across magnetic field lines is many orders of magnitude lower than conduction along field lines, so effectively coronal loops can be modeled to a first approximation as one-dimensional flux tubes.

8.3.2 Hydrostatic Models and the RTV Scaling Laws

Historically, the first attempts to model coronal loops was with hydrostatic models. Rosner et al. (1978) [RTV] studied observations recorded by the X-Ray Telescope aboard Skylab and developed a model of the solar corona, based on coronal loops. Rosner et al. (1978) extended the

coronal loop model of Landini and Monsignori Fossi (1975) to study the relation of the physical quantities in the case of static equilibrium of a symmetric loop having constant cross section. The equation for energy conservation then becomes

$$\frac{d}{ds} \left(\kappa_s T^{5/2} \frac{dT}{ds} \right) = E_R - E_H(s) \quad (3.31)$$

where s is the spatial coordinate along the loop; $E_R = N_e^2 \Lambda(T)$ are the radiative losses; E_H is the volumetric heating rate; T is the temperature; and κ_s is the coefficient for thermal conductivity.

The assumption of symmetry means that the heat flux should vanish at the loop top ($s = L$): $\left. \frac{dT}{ds} \right|_{s=L} = 0$, and the equation can be integrated by prescribing boundary conditions at the foot-points and the form of the heating function $E_H(s)$. With the assumption that $E_H(s)$ and the pressure are constant along the loop, and using an approximation for the radiative losses, RTV showed that the maximum temperature T_{\max} (at the apex) is related to the pressure and the semi-length L by the scaling law:

$$T_{\max} \simeq 1.4 \times 10^3 (PL)^{1/3} \quad (3.32)$$

and showed that, within a factor of 2, it appeared to fit the Skylab observations of the X-ray loops well. RTV also derived a second scaling law:

$$E_H \simeq 9.8 \times 10^3 P^{7/6} L^{-5/6}. \quad (3.33)$$

Loops having the same pressures but longer lengths require less energy input. These scaling laws have been applied for a long time in solar and stellar physics, however solar observations have clearly shown that loops are very dynamic structures and detailed hydrodynamic modeling is needed.

8.3.3 Hydrodynamic Models of Active Region Loops

It is common to neglect the details of the effect of the magnetic field on the heating by defining a general heating function E_H (the volumetric heating rate) and to solve the 1D hydrodynamic equations of mass, momentum, and energy conservation for a single-fluid loop, which in their Eulerian form are:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial s}(\rho v) = 0, \quad (3.34)$$

$$\frac{\partial}{\partial t}(\rho v) + \frac{\partial}{\partial s}(\rho v^2) = \rho g_{\parallel} - \frac{\partial P}{\partial s}, \quad (3.35)$$

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial s}((E + P)v) = \rho v g_{\parallel} + \frac{\partial}{\partial s}(\kappa_s T^{5/2} \frac{\partial T}{\partial s}) - N_e^2 \Lambda(T) + E_H, \quad (3.36)$$

where $E = 1/2 \rho v^2 + 3/2 N_e k T$ is the total energy, kinetic and internal of the fluid element; s is the spatial coordinate along the loop; ρ is the mass density; g_{\parallel} is the component of gravity along the loop axis; T is the temperature.

The use of a single-fluid approach is only valid if all particles are thermalised. For the low solar corona in quiet conditions, the thermal assumption (i.e., that particles follow a Maxwellian

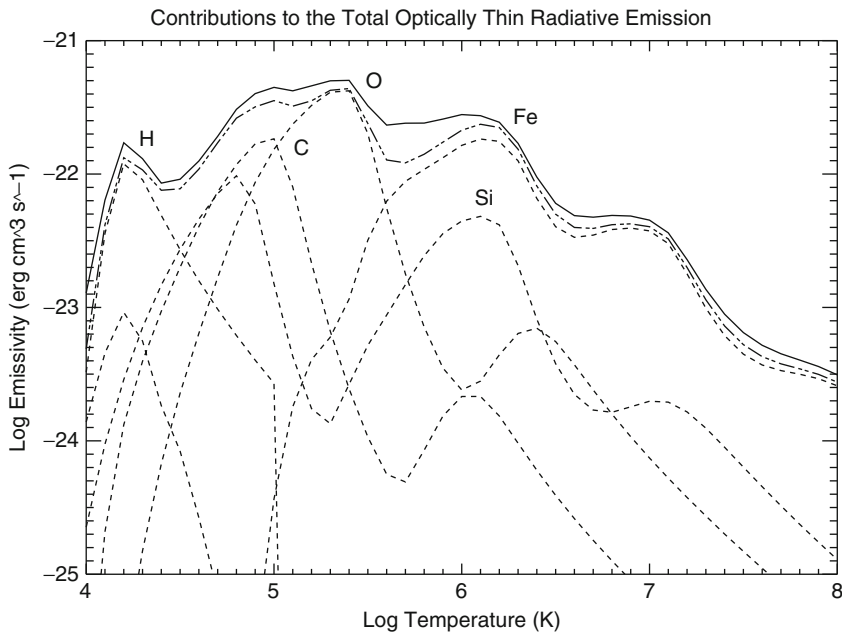
distribution function) is not very stringent, since the relaxation time for e-e collisions is $\tau_{ee} \simeq 0.01 T^{3/2} N_e^{-1} s = 0.01s$ for $T = 10^6$ and $N_e = 10^9$. After a time longer $(m_i/m_e)^{1/2} \simeq 43$, the ions also become thermal, and after the same time factor, both species are thermalized, and only one temperature can be used: $T \equiv T_e = T_i$.

Given its high temperature and low density, the coronal plasma follows the *perfect gas law*, which can be written, once a single temperature is defined, as $P = (R/\mu)\rho T = k(N_e T_e + N_i T_i) \simeq 2NkT$.

Coronal plasma is mostly composed of protons, helium nuclei and electrons, hence the temperature can be thought of an average value. $P = 2N_e kT$ is the total pressure, where N_e is the electron number density. k is Boltzmann's constant. v is the bulk velocity along the loop; $\kappa_s = 9.7 \times 10^{-7}$ (cgs units) is the coefficient of thermal conduction, assuming classical conductivity, see Spitzer (1965).

$\Lambda(T)$ is the optically thin radiative loss function, i.e., is the sum of all the emissivities of the lines and continuum (total power per unit emission measure). It is a function strongly dependent on the temperature and the chemical abundances. [Figure 3-62](#) shows that the main contributions are from oxygen, silicon, iron. The radiative losses also depend on the ion abundances within each element.

The heating function E_H is an often prescribed function of space and time, and the hydrodynamic response of the plasma is studied in detail and compared to observations. Priest et al. (1998) attempted to match models with a specific set of X-ray observations. They showed that

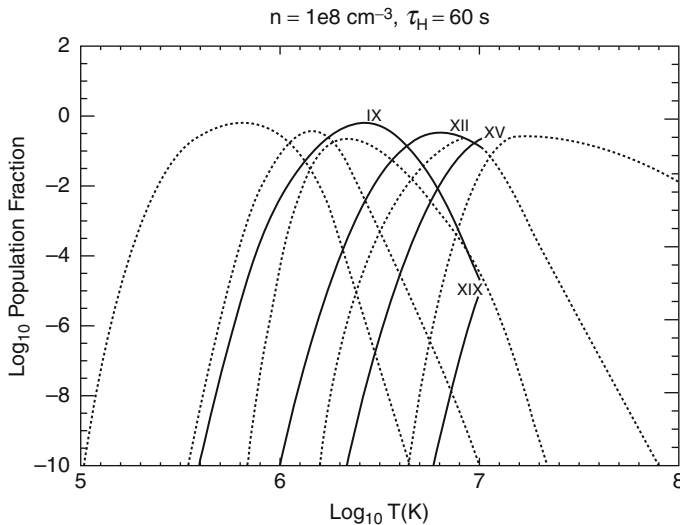


■ Fig. 3-62

The total optically thin radiative loss function. The plot was calculated using coronal abundances. The solid line is the total radiative loss function for the plasma; the dashed lines show the contributions from the individual elements

the footpoint to apex temperature profile, calculated under the assumption of static equilibrium within the loop, is highly sensitive to the spatial distribution of heat deposition along the loop. They compared predicted intensities from various theoretical models with specific Yohkoh observations and found that a uniform heating model was most appropriate. Even though their analysis was later shown to have some severe limitations, it did start a long and fruitful debate on the heating mechanism for coronal loops, with further attempts to match theory and observations.

To this purpose, several hydrodynamic codes have been developed to model the response of coronal loops to different heating functions. For example, the HYD(rodynamic) and RAD(iation) code (HYDRAD) (Bradshaw and Mason 2003) is a hydrodynamics code which incorporates time-dependent ionization and forward modeling of emission from the solar atmosphere. The predicted intensities for the spectral lines are critically dependent upon the population of emitting ions and in traditional hydrodynamic models, these have been assumed to remain in equilibrium throughout the evolution of the plasma. However, the timescale for ionization in the solar atmosphere can be on the order of minutes. Thus, given activity timescales on the order of seconds in solar loops, it is easy to see that the assumption of equilibrium ion populations may not be valid (see, e.g., [Fig. 3-63](#)). This has extremely important consequences for the correct interpretation of observations, the accurate treatment of optically-thin radiation in numerical models and forward-modeling studies derived from model results. HYDRAD handles non-equilibrium ion populations in the solar atmosphere and the consequences for optically-thin radiation, in a self-consistent way. The equations of hydrodynamics and ionization balance are coupled together at each time step during the evolution of the plasma.



■ Fig. 3-63

The *solid curves* show the relative abundances of a selection of Fe ions out of equilibrium, while the *dotted curves* show the equilibrium ion population, as a function of the electron temperature (Adapted from Bradshaw 2009)

8.3.4 Microflares and Nanoflares


Small flares with an energy content of 10^{-6} to 10^{-9} of a typical large flare are classified into *microflares* and *nanoflares*. Microflares are observed not only in ARs but also in the QS regions. Small-scale coronal activity is believed to be of importance for solving the coronal heating problem. Parker (1988) was the first one to suggest that frequent small-scale magnetic reconnections producing nanoflares could be responsible for heating the solar corona. The idea was further developed by Cargill (1994) and others. Basically these small-scale magnetic reconnections prevent indefinite tangling of coronal field lines and help dissipate the magnetic energy stored in the magnetic field lines due to photospheric motions. Magnetic reconnection would lead to particle acceleration. When the particles hit the loop footpoints, this would lead to the evaporation of material. Searching for evidence of upflows (blueshifts in spectral lines) could confirm such models. We do have some observational evidence for magnetic reconnections in the corona and dissipation of magnetic energy into thermal energy in form of micro-flares, flares, and CMEs. However, detection of nanoflares has so far been elusive. Although TRACE, with its $1''$ spatial resolution, seems to resolve loops, it is possible that these loop structures could be made up of many bundles of magnetic flux tubes (strands), each of which could be heated separately.

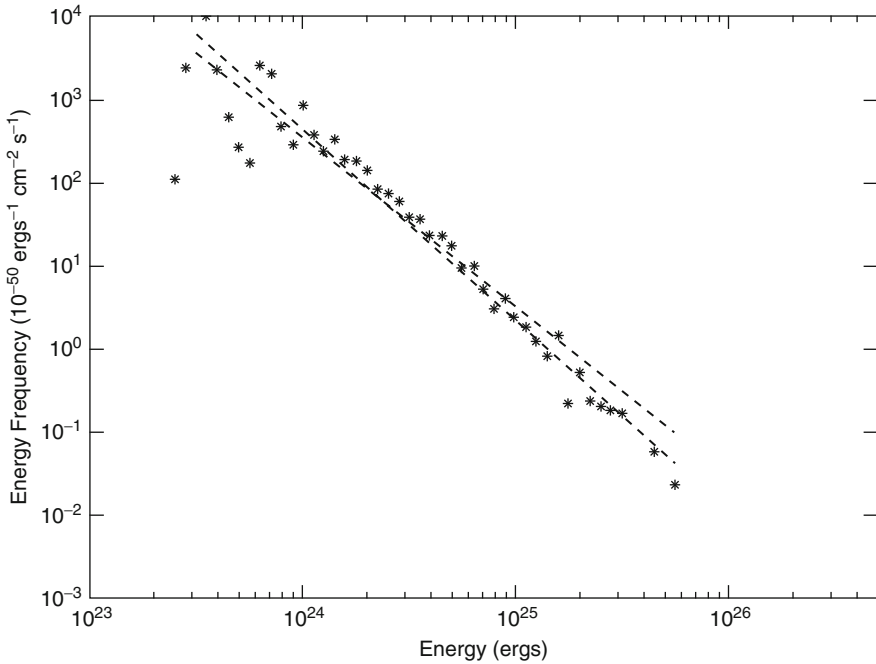
One prediction from nanoflare heating models is the presence of plasma in the corona with temperatures in excess of 5 MK. Various studies based on broadband imaging have suggested the possibility that this hot plasma is present. However, spectroscopic measurements have so far not shown conclusive evidence, and further studies are needed. With Hinode/EIS one can study high-temperature coronal emission from Fe xvii and Ca xvii ions formed at peak temperatures of $\log T = 6.6$ and 6.7 , respectively. Maps made in Fe xvii and Ca xvii emission indicate that the hottest portions of the active region are localized at low heights in the active region core, while further out the emission from hot material is much weaker (O' Dwyer et al. 2010).

The solar corona exhibits a plethora of transient energy releases over many scales, from large solar flares down to microflares. Active region microflares (or active region transient brightenings) were first identified in X-rays with YOHKOH/SXT, possibly associated with the emergence of new magnetic flux. About 1–40 of these events per active region per hour have been observed, lasting a few minutes and reaching high temperatures (up to 10 MK) and electron densities (10^{10} cm^{-3}). Microflares have about a million times less energy than a large flare. Microflares have also been observed in EUV, however, five times as frequent as X-ray microflares. Microflares originating from active regions seem from RHESSI observations to have a non-thermal electron spectrum with a low energy cutoff.

Drake (1971) first noted that the distribution of flare energies W can be fit by a power law function:

$$dN = AW^{-\alpha} dW \quad (3.37)$$

with a slope of $\alpha = 1.7$ – 1.8 . Similar power laws were found in the radio and soft X-ray regions by Lin et al. (1984), with α about 2. These authors pointed out that one may integrate this power law distribution to determine the total power emitted in transient brightening or micro-flare events. If the slope is steep enough ($\alpha > 2$), these transient events would be sufficient to explain the *quiescent heating* of the solar corona. Many authors have found power law indexes lower or higher than two, depending on which instrument was used. One example is given in  Fig. 3-64. One major problem in these types of studies is the fact that what is often observed is a by-product of the primary energy release, not the actual energy, as shown for example with a simulation by Parenti et al. (2006).



■ Fig. 3-64

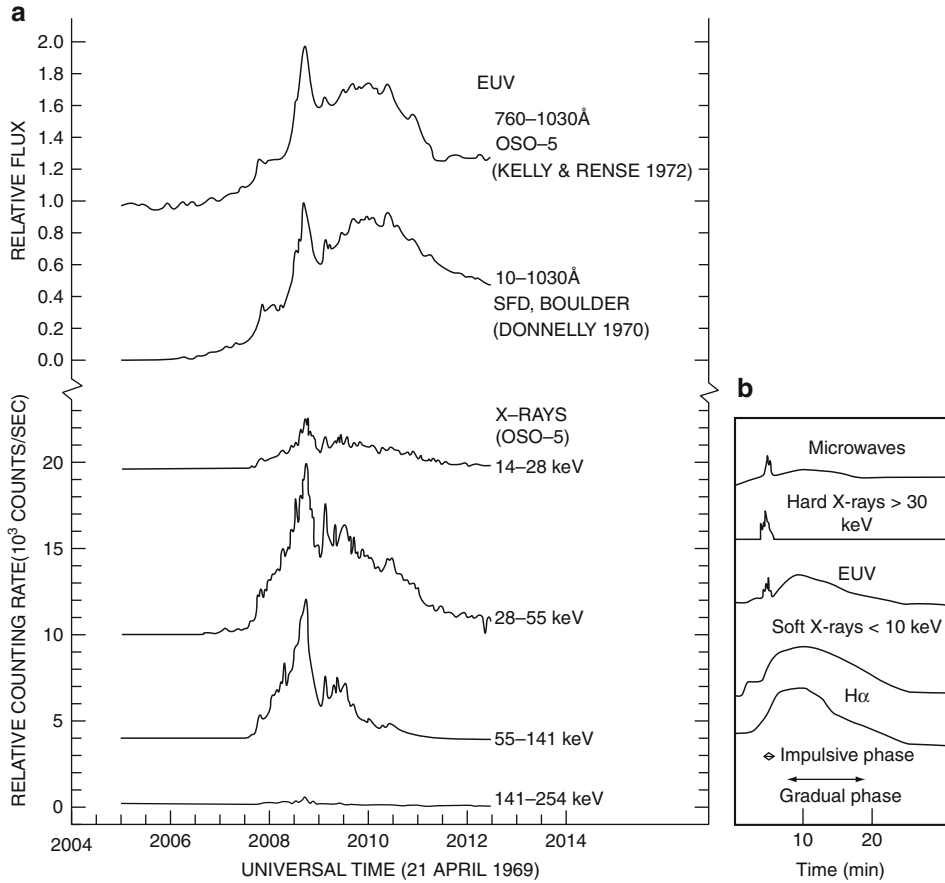
Frequency of occurrence of events vs. event energy obtained from TRACE EUV images (Parnell and Jupp 2000)

9 Solar Flares and Coronal Mass Ejections

9.1 Solar Flares

Solar flares are violent explosions in the solar atmosphere involving the sudden release of energy: bursts of particle acceleration, plasma heating, and bulk mass motion (for reviews, cf. Aschwanden 2006; Culhane and Jordan 1991; Phillips 1992; Sturrock 1980; and the *living review* by Benz 2008). Over periods of less than an hour, a flare can release energy of up to 10^{32} ergs, emitting high levels of radiation in wavelengths ranging from radio to short wavelength X-rays and even gamma rays during large events. The energy range of flares extends over many orders of magnitude. The currently accepted classification of flare size is based on the power irradiated in the 1–8 Å channel of the *Geostationary Orbital Environmental Satellite (GOES)* family. The events are classified as belonging to class A, B, C, M, or X, if their peak emission increases over 10^{-8} , 10^{-7} , 10^{-6} , 10^{-5} , 10^{-4} W m⁻², respectively.

The first observations of solar flares were taken independently by R. C. Carrington and R. Hodgson in 1859 (Carrington 1859). Carrington said “While engaged in the forenoon of Thursday, Sept 1, in taking my customary observation of the forms and positions of the solar spots, an appearance was witnessed which I believe to be exceedingly rare.” He goes on to describe the appearance of “two patches of intensely bright and white light”, which we now associate with the flare ribbons (footpoints of post-flare loops).

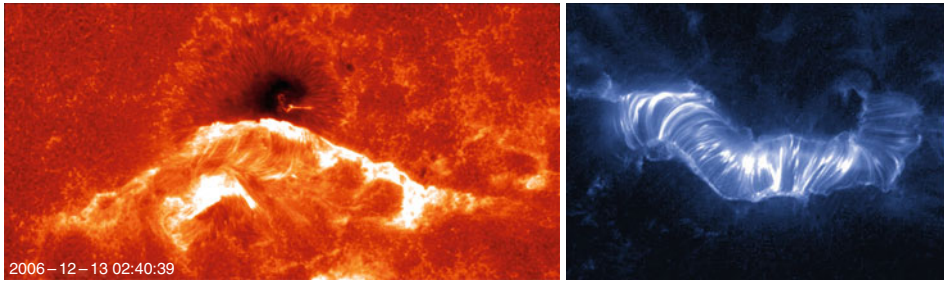


■ Fig. 3-65

Observed *lightcurves* in the EUV, X-rays, and hard X-rays of a solar flare (a (left), after Kane et al. 1979, and b (right) sketch of the typical behavior at different wavelengths as a function of time)

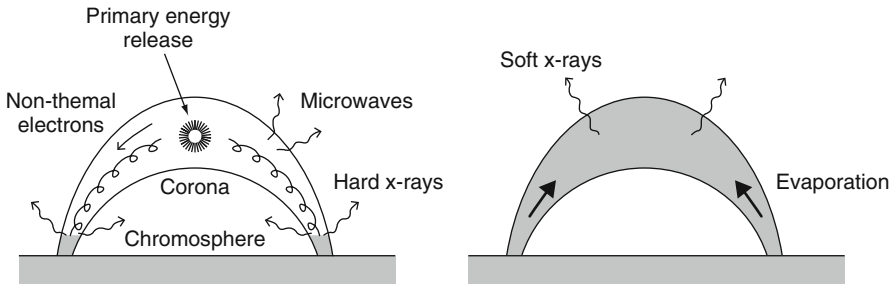
When observed at different wavelengths, flares have a very different temporal evolution. ▶ [Figure 3-65](#) (left) shows a typical example. Hard X-rays normally show an *impulsive* behavior, the same observed in microwave radio emission, and thought to be related to nonthermal processes. The EUV emission also shows some impulsive behavior, due to low-temperature line emission, however the bulk of the emission is thermal and occurs during the *gradual* decay phase (see ▶ [Fig. 3-65](#), right). Soft X-ray emission normally shows peak emission after the hard X-ray peak, and is followed by the gradual phase.

During a solar flare, the temperature can rise in excess of 10 MK, so that the plasma emits strongly in the EUV and X-rays. A typical *two-ribbon* flare is characterized by morphological features which include separating ribbons of chromospheric H α emission (▶ [Fig. 3-66, lhs](#)) joined by a rising arcade of hot but rapidly cooling EUV loops (▶ [Fig. 3-66, rhs](#)), with hard X-ray emission at their summits and at their feet (as seen by RHESSI). Many events are normally associated with large flares, for example, filament eruption, acceleration of large number of particles, triggering of vast ejecta (Coronal Mass Ejections, see below).



■ Fig. 3-66

Left: Image in the Ca II H spectral line of a large flare on December 13, 2006 observed with the Hinode/Solar Optical Telescope (SOT). The image shows the separating flare ribbons in the chromosphere. **Right:** Post-flare loop arcs during the 'Bastille day flare' on July 14, 2000 observed in EUV with the *TRACE* satellite



■ Fig. 3-67

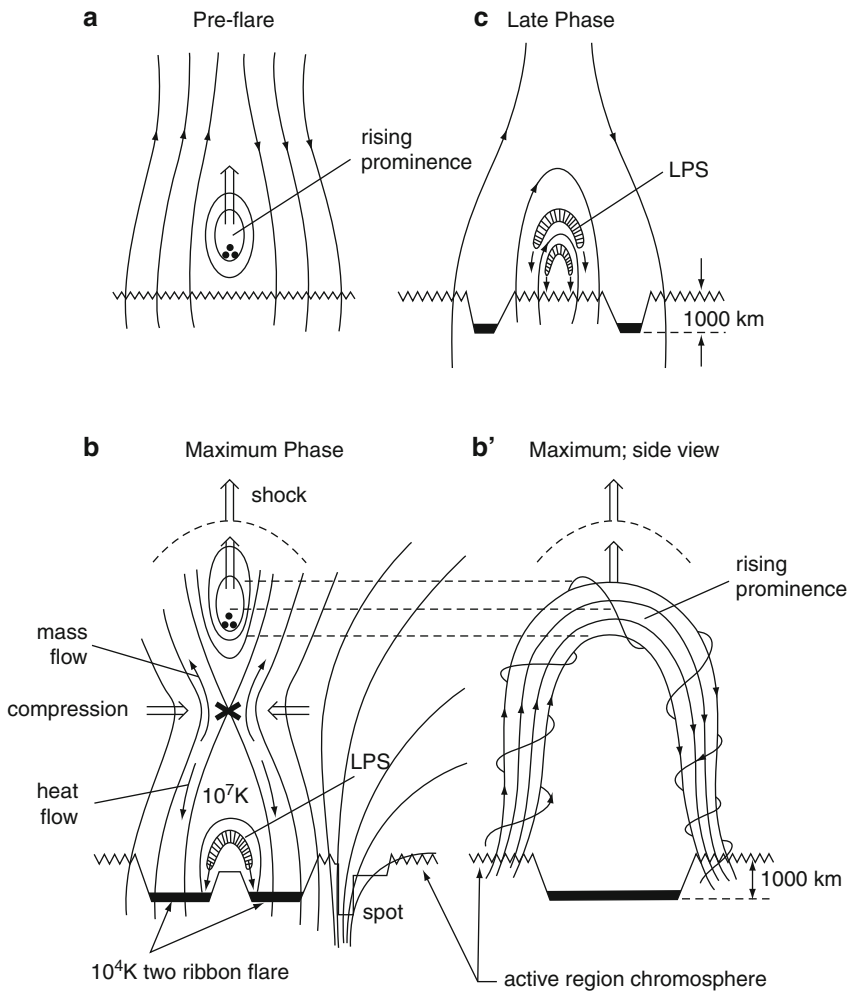
Illustration of the production of microwaves, soft X-rays, and hard X-rays in solar flares after Priest and Forbes (2002)

Flares are believed to be driven by stored non-potential magnetic energy and triggered by an instability in the magnetic configuration. This sudden release of energy and the transport of this energy are still not well understood. There is evidence (for example from the hard and soft X-rays) that both thermal and nonthermal processes occur in flares.

As described in ● Fig. 3-67, a general interpretation of the observed flare features is that particles are somehow accelerated by the magnetic energy released higher up in the corona and are transported down to the footpoints of the loop structures. The energetic electrons then lose all their energy in Coulomb collisions with the dense ambient cold plasma (thick-target nonthermal bremsstrahlung, see, e.g., Brown 1971), giving rise to strong radio, hard X-ray, UV, and even white light. The thick-target nonthermal bremsstrahlung is however very inefficient, and far too many electrons (compared to those available in the corona) need to be accelerated to high energies. Other theories still accept the presence of coronal reconnection in the corona, but advocate the presence of Alfvén waves to transport the energy down (see, e.g., Fletcher and Hudson 2008 and references therein).

Although not all flares can be explained by a single model, it is generally accepted that a standard model for the gradual phase that fits most observations and that is theoretically well understood exists. The most widely accepted flare model is known as the *CSHKP standard model*

(based on the original work of Carmichael, Sturrock, Hirayama, Kopp and Pneuman). In this scenario, an instability causes a filament to rise, which, in turn, causes the magnetic field to be opened up (🔗 Fig. 3-68). The rapid temperature increase during the impulsive phase produces an enhanced pressure in the heated region. This overpressure drives downward-moving cool plasma and evaporation from the lower solar atmosphere (chromosphere) during the flare *gradual* (or *decay*) phase. This results in a hot (over 10 MK) plasma filling the magnetic loop structures in the corona which have several thousand kilometers altitude. This plasma then cools by conduction and radiation forming post-flare loop structures.



■ Fig. 3-68


Temporal evolution of a flare after Hirayama (1974), starting from (a) a rising prominence, (b) which triggers X-point reconnection beneath an erupting prominence, shown in side-view (b'), and (c) ending with the draining of the chromospheric, evaporated, hot plasma from the flare loops

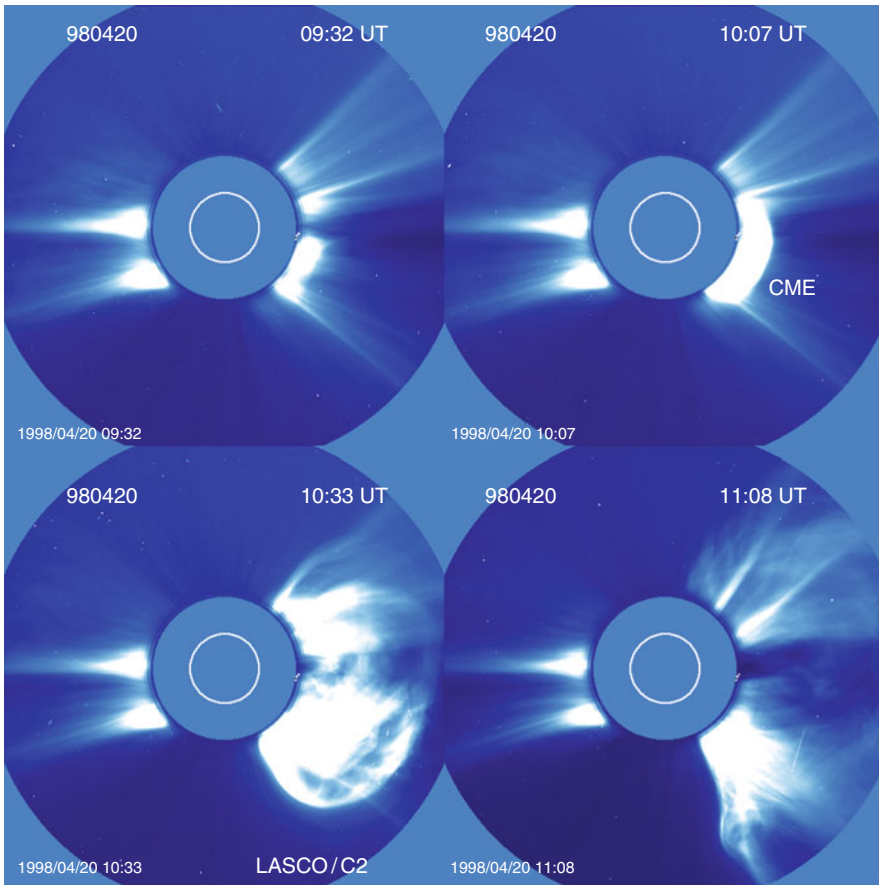
One of the limitations in testing theoretical predictions of chromospheric evaporation and cooling has been a lack of spectroscopic measurements that cover a broad range of temperatures during the lifetime of a flare with sufficient spatial resolution. Earlier observations have been performed with various instruments, but there were limitations mainly in their spatial capability. Blue-shifted components of a few hundred km s^{-1} have been reported in H- and He-like lines observed in the X-rays, for example, with the SMM X-ray polychromator (XRP) Bent Crystal Spectrometer (BCS) (see, e.g., Antonucci et al. 1982) which did not have spatial resolution. The blueshifted components were thought to be located at the footpoints of the X-ray loops. Some spatial resolution was obtained with the SMM Ultraviolet Spectrometer and Polarimeter (UVSP, see Mason et al. 1986) and XRP Flat Crystal Spectrometer (FCS) observations, however it was only with SoHO CDS observations that more information was obtained. Czakowska et al. (1999) confirmed that strong blueshifts in hot Fe XVI (3×10^6 K) and Fe XIX (9×10^6 K) lines were located at the outer edges of the flare ribbons, as expected with the CSHKP model. Del Zanna et al. (2002a) presented the first spatially resolved spectral observations of the entire evolution of a small flare, where long-lasting (over an hour) blueshifts of 30 km s^{-1} at the footpoints of a small flare ($T = 9 \text{ MK}$) loop were present. Del Zanna et al. (2006a) confirmed expectation from theory that at the flare kernels, during the impulsive phase of an M1 class flare, strong upflows, progressively larger at higher temperatures (reaching 140 km/s for $T = 9 \text{ MK}$) are seen together with small downflows in the cooler lines. Important contributions in this field are expected from Hinode EIS observations.

9.2 Coronal Mass Ejections (CMEs)

Coronal Mass Ejections (CMEs), often associated with solar flares, are also spectacular and violent phenomena in the solar atmosphere which propagate through the heliosphere and can impinge on the near-Earth environment. The flare which was observed by Lord Carrington in 1859 led to a gigantic CME, large enough to have serious consequences had it happened in the modern-day technology-based era. Remote sensing and in situ measurements from various spacecraft over the past decade or more have revealed that CMEs are the prime agents for interplanetary shock waves, solar energetic particle events, and geomagnetic storms.

A CME is defined as an observable change in the coronal structure which occurs on a timescale between a few minutes and several hours, and also involves the appearance and outward motion of a new, discrete, bright, white-light feature in the coronagraph field of view (Hundhausen et al. 1984). CMEs consist of clouds of hot plasma expelled from the Sun with propagation speeds of up to $3,000 \text{ km s}^{-1}$, and with a mass of 10^{12} kg or more. Both flares and CMEs can accelerate electrons and ions which travel through interplanetary space as far as the Earth's orbit and beyond. The CME is often preceded as a shock front which, when it reaches the Earth, may result into a magnetic storm. Therefore, flares and CMEs can influence the space weather and human activities both in space and on the ground.

CMEs were systematically observed with coronagraphs aboard *Skylab* (1973–1974), *SOL-WIND* on board *P78-1* and the Corona/Polarimeter on board *Solar Maximum Mission* (SMM, 1984–1989). Since 1996, the *Large Angle Spectrometric Coronagraph* (LASCO) aboard SoHO has provided unique observations of CMEs. An example of a spectacular CME is shown in  Fig. 3-69 as a sequence of four images obtained by LASCO on Apr-20-1998. In the figure the white solid circles indicate the solar limb in the shadow of the coronagraphic occulting disk. In the images, north points upward and west toward the right. The bright feature which appeared at the west limb in the south-west quadrant (top right panel at 10:07 UT) expanded



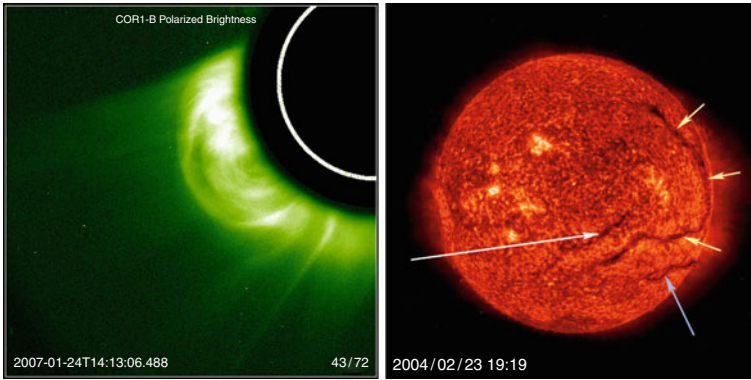
■ Fig. 3-69

Series of LASCO images showing a coronal mass ejection at the west limb. The appearance of the CME was first detected by LASCO at 10:07 UT (*Top right panel*). The white circles in the images mark the solar limb. In the images north points upward and west toward the right. Courtesy: SoHO/LASCO consortium

outward in the following images. This feature moved outward through the corona with a speed of about $1,863 \text{ km s}^{-1}$. A recent collection of reviews and papers can be found in ‘Coronal mass ejections’ (Kunow et al. 2006). The recent launch of the STEREO spacecraft has provided more CME observations.

The identification of the source regions, the triggering and acceleration mechanisms of CMEs have been the subject of intensive research in solar physics since their detection in the 1970s. Many features have a close but not unique association with CMEs and are still not well understood. For example, solar flares are often observed in close association with CMEs. It is generally believed that both flares and CMEs are triggered by a common magnetic instability, the details of which remain unknown.

Filaments (also referred to as *prominences* if observed as bright features over the solar limb) consist of cool material at chromospheric temperatures of about 60,000 K (🔗 Fig. 3-70 - right).



■ Fig. 3-70

Left: A CME blasting out from the Sun on January 24, 2007. The observation was taken by the COR1 coronagraph on the *STEREO* mission. The Sun is represented by the white circle superimposed on the occulting disk (*the larger black circle*). The field of view of this coronagraph shows details of events closer to the Sun than either of the two coronagraphs currently operating on *SoHO*. **Right:** EIT image showing He II plasma (304 Å). The darker, cooler filaments of around 60,000 K (indicated by arrows) are observed in absorption (Courtesy: NASA/ESA)

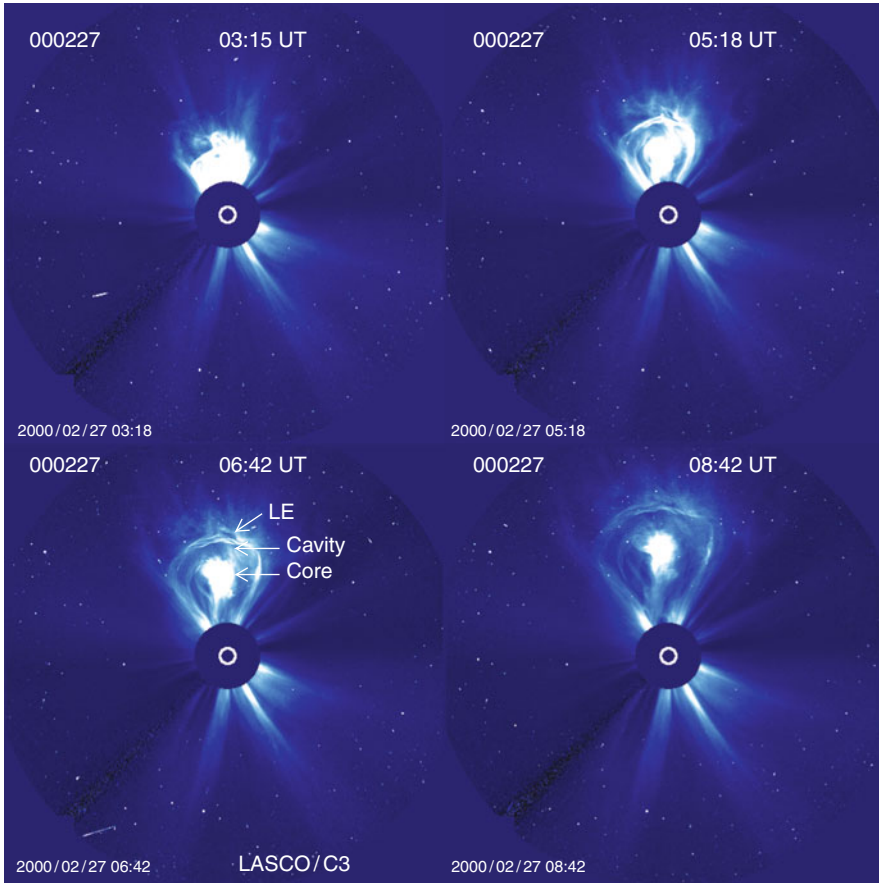
Filaments are often observed to erupt during a flare or CME process. Very often, filaments during the pre-eruption phase slowly rise into the corona. During the eruption, the height normally grows exponentially, and both the filament and the filament-carrying field move together (e.g., Rust 1976). When erupting filaments are associated with a CME, they are sometimes recognized as a bright core in white light images, following a leading edge and a dark cavity, as shown in ● Fig. 3-71. However, many CMEs do not show this structure.

Other features associated with CMEs are coronal dimmings (temporary darkenings in the corona) and apparent “waves” (EUV brightenings that are seen propagating outward). The brightening in X-rays of S-shaped ($\sim 2 \times 10^6$ K) coronal loop structures, called sigmoids (see ● Fig. 3-72), has been found to frequently precede CMEs. The sigmoid structures are probably the result of twisted magnetic fields. Shearing and stressing of magnetic field lines above the magnetic neutral line may lead to the helical structure which are S-shaped in projection. When the helical twist exceeds a critical value, the structure might become unstable, producing a disruption of the magnetic field leading to the expulsion of a filament or CME. The S-shaped structure may flare and transform itself into a set of bright loops during the launch of a CME. The active region itself evolves from a sheared sigmoid seen in X-rays to a potential-like post-eruption arcade, as shown in ● Fig. 3-72.

Several studies have found preflare activity such as small X-ray and EUV brightenings at the onset of the filament’s slow rise, which could trigger the eruption. These brightenings are often associated with emerging magnetic flux.

9.3 Theoretical Concepts of CME Initiation

The driver in the CSHKP model is a rising filament, but the magnetic pre-evolution that leads up to the flare and CME instability is not quantified in the various concepts of the CSHKP models.

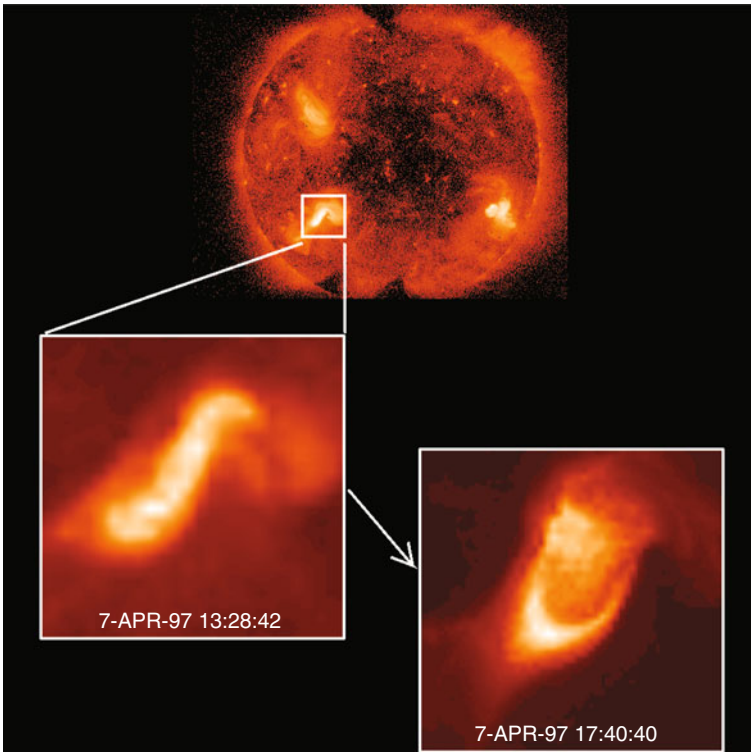


■ Fig. 3-71

A CME observed with the LASCO/C3 coronagraph, showing the typical three-part structure (Figure courtesy of SoHO/LASCO and D.Tripathi)

Several eruption models proposed in the literature involve emerging flux. A review of the ideas to date for the initiation of solar eruptions may be found in Moore and Sterling (2006). They described each of the three basic triggers models (internal tether-cutting (● Fig. 3-73), external tether-cutting, and an ideal MHD [e.g., kink] instability) for the case when the erupting filament sustained by a sheared core field is situated in the central lobe of a quadrupolar magnetic field configuration.

Breakout models (e.g., Antiochos 1998) predict that reconnection occurs high above the prominence before the eruption, while tether-cutting mechanisms imply that reconnection *beneath* the prominence unleashes the explosion. Prior to the explosion, the core magnetic field suspending the filament is in force-free equilibrium (Moore and Sterling 2006), the magnetic pressure being balanced by the field's own magnetic tension as well as tension and pressure from surrounding fields. An evolutionary model that starts with a stable (force-free) magnetic field configuration, then applies converging flows as a continuous driver, and demonstrates how



■ Fig. 3-72

Collage showing the sequence from pre-eruption sigmoid to post-eruption cusp/arcade imaged by Yohkoh/SXT on Apr-07-1997 (Courtesy: Yohkoh/SXT team)

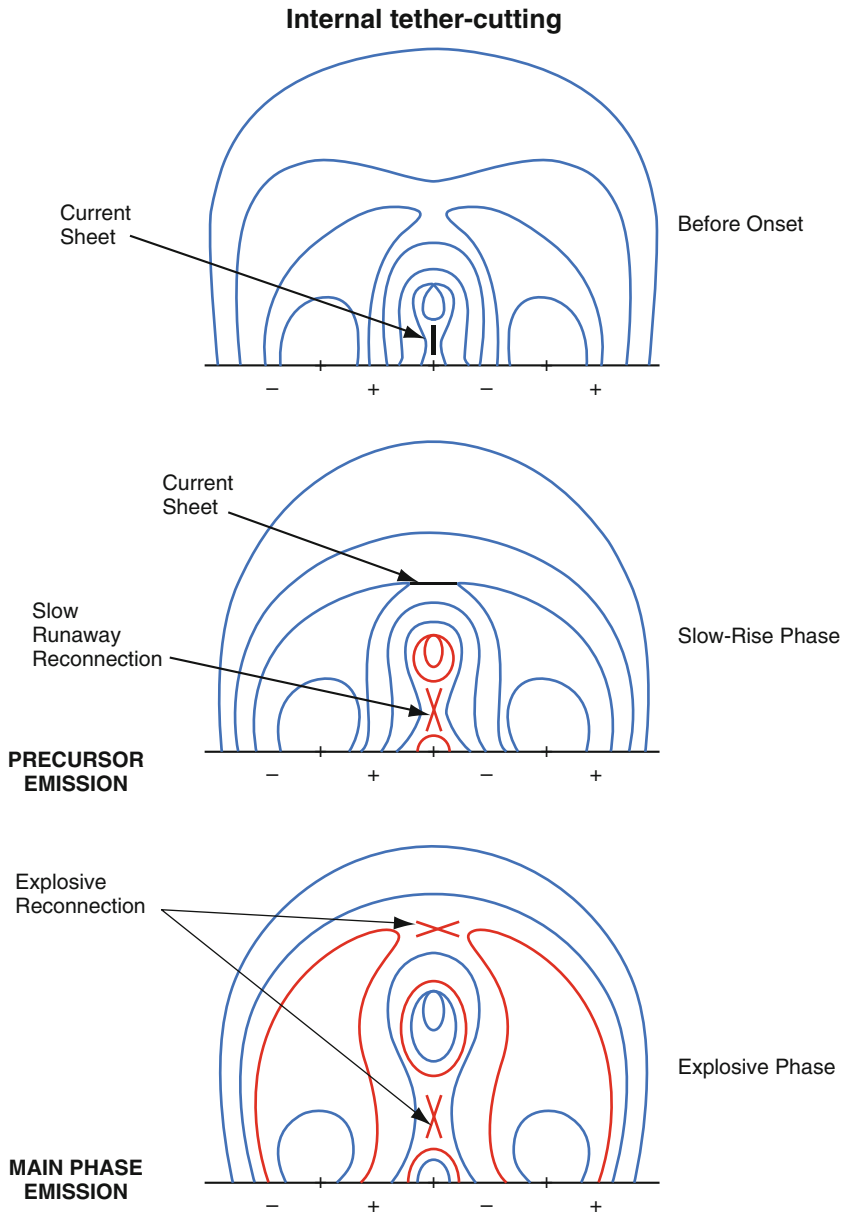
(force-free) evolution passes a critical point where the system becomes unstable and triggers the rise of a filament has been developed by Forbes and Priest (1995) in 2D.

The exponential growth in the early phases of filament eruptions associated with CMEs naturally invokes the presence of some MHD instability. Various successful models of filament eruptions include the helical kink instability of the filament/flux rope (Kliem et al. 2004) and the torus instability (Kliem and Török 2006).

9.4 Flares on Other Stars

Stellar flares have been observed in the visible since the late 1930s in active M-type stars. There is a strong nonthermal continuum in the UV which contributes significantly to the ultraviolet broadbands used in ground-based observations. The dMe stars are the most active and flare-productive stars in our neighborhood. Stellar flares were later observed at all wavelengths, in the radio and with satellites from the UV to the X-rays. As we have seen, one key feature of flares is the sudden heating to temperatures greater than 10 MK, giving rise to strong emission in the EUV and X-rays.

The EXOSAT carried Transmission Grating Spectrometers (TGS) which recorded XUV (50–400 Å) spectra for a few stars. EINSTEIN carried three spectrometers with different



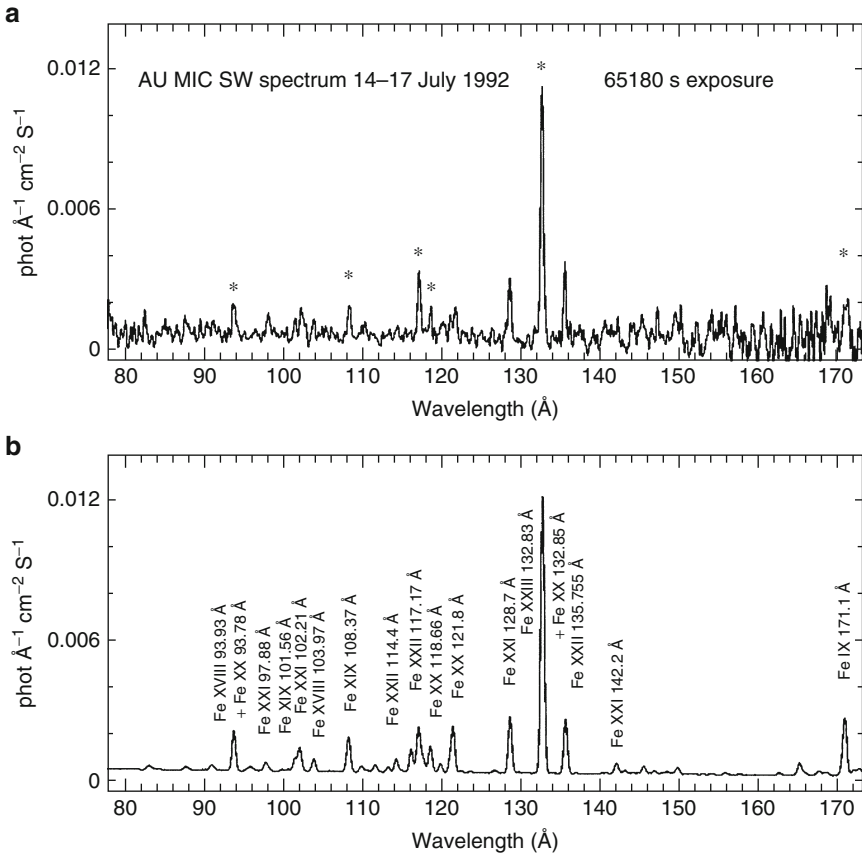
■ Fig. 3-73

Schematic representation of one trigger mechanism for CME eruption, the internal tether-cutting. The magnetic field is shown prior to eruption (*top*), just after eruption onset (*middle*), and when the eruption is well under-way (*bottom*). The sites of the precursor and main phase emissions (observed respectively as preflare brightenings and flare brightenings during the impulsive phase of the eruption) are indicated. Modified figures courtesy of Chifor, Moore and Sterling, after Moore and Sterling (2006)

resolutions which also obtained X-ray spectra of a few sources. ROSAT carried broadband X-ray instruments, which was of limited use for diagnostics. All these early X-ray satellites showed the presence of hot plasma up to a few times 10^7 K in many stellar sources.

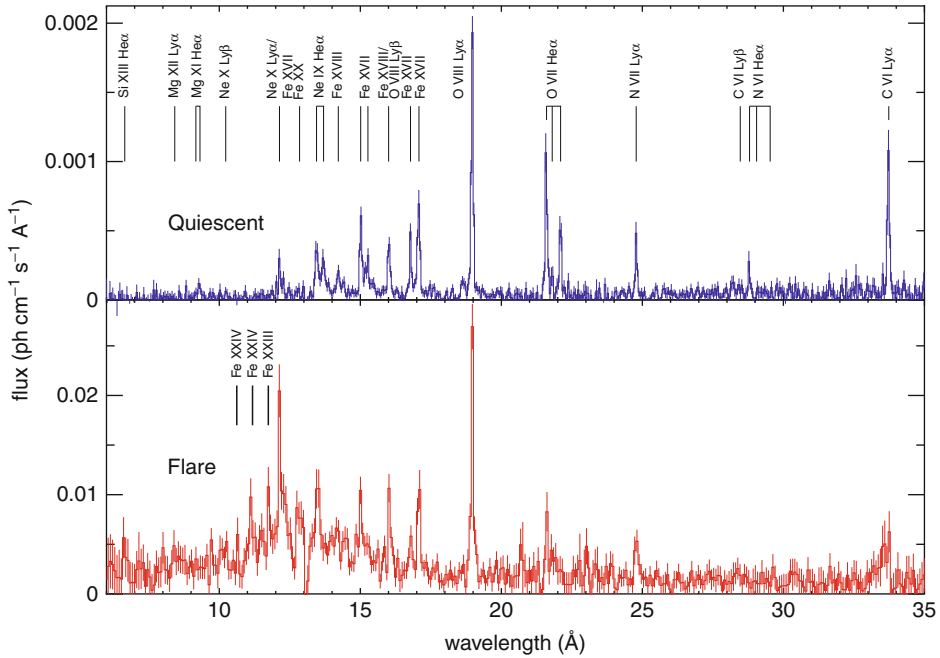
The International Ultraviolet Explorer (IUE), the Hubble Space Telescope (HST), and later FUSE have provided the capability of observing UV emission with spectroscopic instrumentation. With the exception of a few weak forbidden lines formed at flare temperatures, the UV is dominated by emission lines formed in the stellar chromospheres and transition regions.

The Extreme Ultraviolet Explorer (EUVE) satellite carried three grazing incidence spectrometers which observed (in three bands) the 70–760 Å wavelength region. These observations are very rich in diagnostics for flares, in particular with regard to the measurement of electron densities. Temperature sensitivity is also provided, but only by observations of lines from different ions, e.g., Fe XVIII, Fe XX, Fe XXI, Fe XXIII. The sensitivity of EUVE allowed the first time-resolved spectroscopy of a flare in the EUV (Del Zanna 1995, later published in ApJ, see ► Fig. 3-74), indicating, as in the solar case, that high temperatures and densities were reached during the peak phase.



■ Fig. 3-74

EUVE spectrum of AU Mic, observed (a) and simulated (b) (Del Zanna 1995)



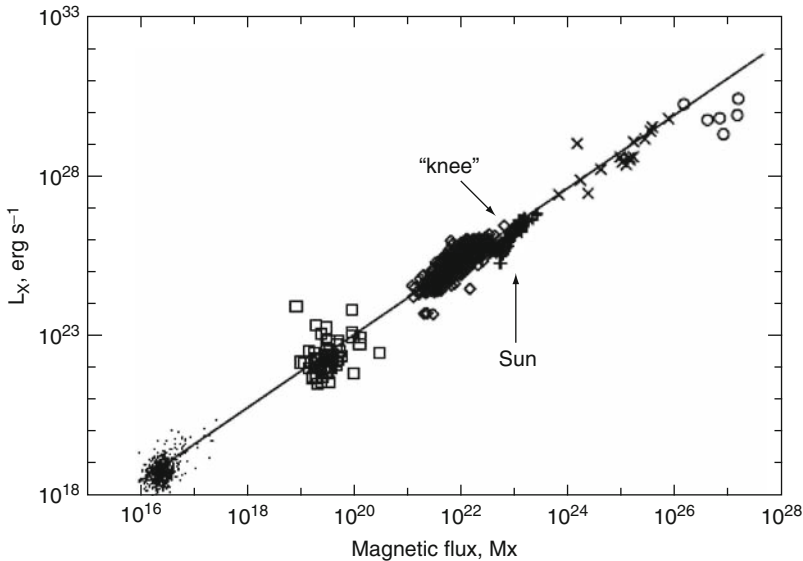
■ Fig. 3-75

XMM-Newton RGS spectra during quiescence and a flare in Proxima Centauri (Adapted from Güdel et al. 2004, courtesy of M. Güdel)

The *XMM-Newton* and *Chandra* satellites, launched in 1999, carried X-ray spectroscopic instruments and have been used to observe many flares in a few nearby stars. The X-rays are rich in strong lines emitted at high temperatures (cf. ● Fig. 3-75), but are comparatively poor in diagnostic capabilities, except for the He-like ions. The spectral resolution of the *XMM-Newton* and *Chandra* instruments is much lower than that attained with solar instruments, so diagnostic capabilities are more limited. The next generation of X-ray instrumentation (IXO) should allow more detailed studies of stellar flares, having much higher spectral resolution and sensitivity.

Time-resolved spectroscopy of flares is still largely beyond the current instrumental capabilities, although some diagnostics based on a few of the strongest lines have been obtained. For example, a large flare in Proxima Centauri allowed the measurement of flare lines (see ● Fig. 3-75, Fe XXIV, Fe XXIII) and of changes in He-like lines (Güdel et al. 2004).

In the large majority of cases, stellar flares have similar characteristics to the solar ones, with the main difference being that luminosities and temperatures are often much larger. For some large stellar flares, temperatures of 100 MK or more have been reported (see the review of Güdel and Nazé 2009), however direct measurements were not obtained and did not take into account the large nonthermal effects which occur in large flares. In the Sun, larger flares have the tendency to have larger emission measures and temperatures, however direct measurements (from line ratios) of temperatures from high-resolution X-rays rarely indicated temperatures above 15 MK (see the review of Doschek 1990).



■ Fig. 3-76

X-ray luminosity L_X as a function of total unsigned magnetic flux for solar features and stars (From Pevtsov et al. 2003)

Young stars have very hot and dense coronae (typical of large solar flares) even in quiescence. The strong magnetic fields measured in these stars suggests that the quiescent state is actually due to continuous flaring, an issue which has not yet been confirmed unambiguously. It has been known for a long time that there is a correlation between estimates of magnetic flux and X-ray luminosity, and that this extends from solar active region core loops to the largest stellar flares, as ► Fig. 3-76 shows.

A large number of open questions about the X-ray quiescent and flaring emission are still present. It is not clear yet what dynamo mechanism is active in different stars. It is not known if stars have active regions as is the case of the Sun, and if flares are really similar and can be explained with the same models as the solar ones.

10 Solar Wind

By the late nineteenth century, there was substantial evidence of terrestrial effects (magnetic storms and aurorae) being correlated with solar activity (e.g., flares), so it was clear that the Sun had a direct influence on a variety of terrestrial phenomena. An obvious candidate was the direct irradiance from the Sun, but other suggestions soon emerged. As early as 1900, Birke-land proposed that aurorae were due to particles (electrons) emanating from the Sun. The idea was superseded in the 1930s with Chapman and Ferraro's explanation in terms of solar plasma clouds, which is close to our present understanding.

Observations of the tails of comets (Biermann 1951) also suggested the presence of a *solar wind* (SW) emanating from the Sun in all directions, with velocities of 500–1,000 km/s. From 1939, it was also known that the solar corona had a temperature of 1 MK. S. Chapman in 1957

suggested that the solar corona was made up of ionized gas in a stationary condition. At that time the heat flux for thermal conduction by a cloud of electrons and protons was known, the main conductors being the electrons. By assuming spherical symmetry and a static corona, as well as a constant conductive flux, it was straightforward to show that Chapman's model was unrealistic because it predicted finite pressures and divergent densities toward infinity. E.N. Parker knew that the energy in a gas naturally goes into an ordered flow (kinetic energy) and a disordered one (pressure), so it was sufficient to introduce an ordered flow (a coronal expansion) to fix the problem. Parker (1958) brilliantly explained that Biermann's suggestion was perfectly consistent with the natural thermal expansion of the solar corona, a plasma at 1 MK.

To understand Parker's solution it is necessary to simplify the hydrodynamic equations, by considering spherical symmetry, with a stationary ($\partial/\partial t = 0$) and adiabatic ($dS/dt = 0$, where S is the entropy of the gas) flow for a perfect gas (single fluid). The equations for conservation of mass, momentum, and energy can then be written in radial coordinates:

$$\frac{1}{r^2} \frac{d}{dr} (r^2 \rho v) = 0 \quad (3.38)$$

$$\rho v \frac{dv}{dr} = -\frac{dP}{dr} - \frac{GM\rho}{r^2} \quad (3.39)$$


$$v \frac{dS}{dr} = 0 \quad (3.40)$$

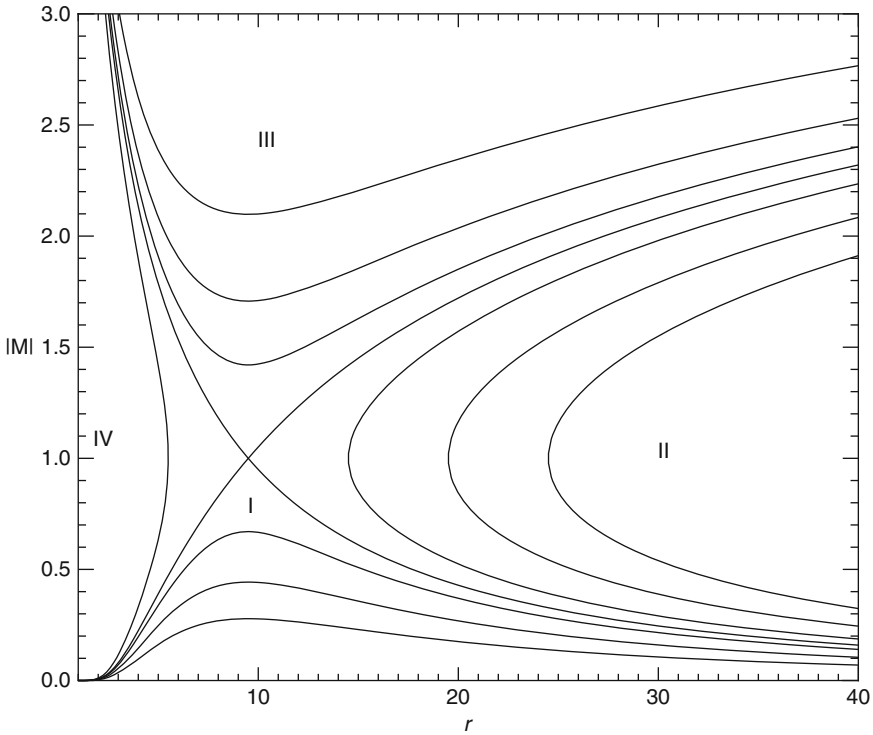
where r is the radial distance from the Sun, P the pressure, and M the solar mass. A relation between the velocity v and r is found by eliminating dP/dr from the conservation of mass and momentum equations. It is straightforward to obtain

$$\left(1 - \frac{c_s^2}{v^2}\right) \frac{dv^2}{dr} = -\frac{2GM}{r^2} \left(1 - \frac{r}{r_c}\right) \quad (3.41)$$

where

$$r_c = \frac{GM}{2c_s^2} = \frac{GM\rho}{2\gamma P} = \frac{GM}{2\gamma RT/\mu} \quad (3.42)$$

is the critical radius, i.e., the distance where the velocity is equal to the sound speed $c_s^2 = \gamma P/\rho$. γ is the ratio of specific heats, and $P = R/\mu T/\rho$ is the perfect gas law, with R the Avogadro number and μ the mean molecular weight. The solutions of the equation (mass loss) can be grouped into four types, and are shown in the  Fig. 3-77. The profile of the velocity as a function of radial distance is shown in terms of the Mach number, the ratio of the velocity v to the sound speed c_s . There are subsonic solutions (the *breezes*, region I in the figure), where the velocity is always below the sound speed. These lie below the two transonic solutions, which are the curves that intersect at Mach number = 1. The transonic solution that has a very large velocity close to the Sun is not physical, and is not supported by observations. The accelerating transonic solution is of particular interest not only because the velocity starts from zero at the Sun, but also because the density and pressure tend to zero at infinity. This behavior close to and very far from the Sun represents not only a physical solution for the solar case, but also the only one which is stable. This solution also implies that the velocity of the solar wind has to be supersonic near Earth.



■ Fig. 3-77

The solution for mass loss from a star in spherical symmetry and stationary conditions (Courtesy of M.Velli). The plot shows the plasma Mach number (The ratio of the velocity v to the sound speed c_s) as a function of the distance from the Sun, in solar radii

Parker's initial simple model assumed an isothermal gas. This means that c_s^2 is constant and (3.41) can be integrated to find

$$v^2 - c_s^2 \ln v^2 = \frac{2GM}{r} + 4c_s^2 \ln r + \text{const.} \quad (3.43)$$

for large values of r , the velocity $v \sim 2c_s(\ln r)^{1/2}$. To obtain an estimate of the density, one considers that $4\pi r^2 \rho v = \text{const.}$ from which $\rho \sim c_s^{-1} r^{-2} (\ln r)^{-1/2}$, i.e., the density tends to zero for large values of r . The same is true for the pressure, since $P \sim \rho T$. For an isothermal temperature of 10^6 K, the simple model predicts a velocity of 100 km/s at 1 AU, a value which is of the same order as the measured one.

The idea that a supersonic solar wind could exist was not initially accepted by the majority of researchers; however very soon Parker's predictions were confirmed. The Soviet *Lunik* 1 and 2 missions were launched in 1959, and made the first measurements of the solar wind. The definitive proof of a supersonic solar wind was made public after the measurements of the NASA *Mariner* 2 satellite which flew to Venus in 1962. Velocities of 300–700 km/s were measured. It has been clear from the first solar wind measurements that Parker's idea was correct. Indeed, Parker's model explains quite well the mainstream slow component of the solar wind, which has a velocity of about 400 km/s and is predominant within the ecliptic.

10.1 The Heliosphere

Parker also suggested that a radial flow of the SW, combined with the solar rotation, would produce a spiral heliospheric magnetic field in the solar equatorial plane, which is what is observed. A good review of the heliosphere near solar minimum (1996) with significant Ulysses contributions is the book by Balogh et al. (2001). A more recent review of the heliosphere through the last activity cycle is the book edited by Balogh et al. (2008). A recent review on the structure of the global heliospheric field and its connection, both observationally (Ulysses, SoHO) and theoretically, is given by Zurbuchen (2007).

The heliospheric magnetic field is weak (about 5 nT) but omnipresent and has the important effect of shielding the solar system from cosmogenic cosmic rays (GCR). The charged energetic particles are deflected when strong magnetic fields are present. Indeed, there is a clear relation between solar activity (hence heliospheric magnetic field density) and the flux of GCRs observed within the solar system. This was already suggested by Davis (1955) even before the presence of the solar wind was proposed.

It is well known now that a vast number of stellar classes have a hot corona similar to the Sun, so it is natural to expect that stellar winds are a common feature. P Cygni line profiles have been observed in hot stars, and are thought to be caused by radiation-pressure-driven winds. Also red giants and supergiants produce massive winds. Winds in solar-like stars are difficult to observe, and it is only since HST observations of H I absorption in the Ly α line that indirect evidence was provided for stellar winds similar to those of the Sun (see the review by Wood 2004).

One subtle issue is that the external pressure of the interstellar medium is not zero. It would not require a strong pressure to actually inhibit the solar wind and to have a static corona. For example, a pressure of 1.24×10^{-12} dyne cm $^{-2}$ would be enough to confine a 4×10^5 K static corona with a density of 10^9 cm $^{-3}$. The dependence of the solar wind on the external conditions has been studied in detail by Velli (1994).

Several satellites have orbited nearby planets, but others have gone further afield, to the edge of the solar system, for example *Pioneers* 10 (launched in 1972) and 11 and the *Voyager* 1 and 2, launched in 1977. By 1997, the *Pioneer* missions were terminated, while the *Voyager* missions are still operating.

At the edge of the Heliosphere, the solar wind must encounter and interact with the local interstellar medium (LISM). (see Zank (1999) for a good theoretical review). It was once thought that the boundary was very distant, about 100 AU or more. It was expected that there should be a termination shock followed by a region, called the heliosheath, where the solar wind is slowed down by the pressure of interstellar gas. Indeed *Voyager* 1 and *Voyager* 2 crossed the termination shock in 2004 and 2007 and are currently inside the heliosheath, continuing their amazing journey. The data from these crossings suggest the presence of an unexpectedly strong interstellar magnetic field of about 5 μ G.

10.2 Physical Characteristics of the Solar Wind and Models

A great deal is known about the solar wind from a range of satellites, in particular from the two German Helios satellites, which were launched in 1976 and which observed the inner Heliosphere between 0.29 and 1 AU for many years, providing a wealth of data. For a summary of early results from these two missions, see the two books by Schwenn and Marsch (1990) and Schwenn and Marsch (1991). Peculiar aspects of these satellites were their closeness to the

Sun, and the ability to measure the particle distribution functions. Anisotropic (i.e., far from Maxwellian) distributions have been found, in particular in the fast SW. More recently, a large number of satellites with in situ instruments have been operating, for example, the Advanced Composition Explorer (ACE), Wind, SoHO, and STEREO.

The properties of the solar wind close to the Sun are also obtained by indirect measurements, by the scintillation of the radio signals from distant objects such as quasars when their line of sight is close to the Sun. The signal is modulated by the density and flow velocity fluctuations, and with many line-of-sight measurements it is possible to construct a map of the solar wind speed, which appears to increase very close to the Sun (within 20–30 R_{\odot}). The first array of detectors were built by A. Hewish at Cambridge University, and nowadays many such arrays exist over the world.

The particle velocity is radial, with values in the range 300–800 km s^{-1} . Note that the wind needs about 4–5 days to reach Earth, and a few months to reach the outermost planets. The main plasma constituents (95%) are protons and electrons, with the rest mainly α particles. The Coulomb drag is caused by the interactions between ions and protons, and results in all minor ions (i.e., ions form all elements except H, He) to move at equal speeds with the protons, a fact which is observed in the SW. The velocity distribution is mainly bimodal, with a highly variable, slow component around 400 km s^{-1} , and a fast one, more steady around 750 km s^{-1} .

Close to Earth, at 1 AU, the characteristics of the slow SW are: the mean temperatures: $T_e \simeq 1.5 \times 10^5$; $T_p \simeq 4 \times 10^4$; $T_{\alpha} \simeq 2 \times 10^5$ K; the intensity of the magnetic field is in the range $2 - 10 \times 10^{-5}$ G; the plasma β is typically around 1–2, though becomes larger than 3 in the fast SW; the sound speed is about 40 km/s, of the same order of the Alfvén speed. So the solar wind is not only supersonic, but also super-Alfvénic. The number density is very low, of the order of 15 particles cm^{-3} .

The fast SW has lower densities by about one third, similar electron temperatures but much higher proton and α -particle temperatures, by a factor of 4.

Close to the Sun, the atmosphere can be approximated quite well to first order with a hydrostatic solution, which predicts that the density falls off exponentially with an hydrostatic scale height $H = RT/\mu g$. The approximate exponential drop off in density has been confirmed by observations. The very low density close to the Sun means that at just two solar radii the plasma is not collisionally dominated, i.e., the electrons and protons do not exchange sufficient energy to thermalise. As a consequence, electrons and protons can have very different temperatures. Considering a 2-fluid plasma composed of electrons and protons, it is expected that the proton temperature would decrease much faster than the electron temperature, given that the thermal conduction for the electrons is about $(m_p/m_e)^{1/2} \simeq 50$ times larger. The first two-fluid model of the solar wind, by Sturrock and Hartle (1966), produced T_e quite close to the observed values at 1 AU, but T_p was about an order of magnitude lower than observed. The predicted plasma velocity was about 250 km/s at 1 AU.

Clearly, the simple thermal expansion idea from Parker was able to explain the gross properties of the solar wind, but in reality there must be some acceleration processes that increase the velocity of the particles and heat them (in particular the protons). The situation is even more complex in the case of the high-speed component which cannot be explained by Parker's model.


The presence of a high-speed solar wind creates compression regions corotating with the Sun, as slow and fast wind streams originating from the same solar latitude, but from different longitudes interact dynamically.

The discovery of Alfvén waves appeared to provide a way to solve the problem of the heating and acceleration of the solar wind. Belcher and Davis (1971) used data from Mariner 5 to

show that for a large fraction of the time Alfvén waves are present and dominate the microscale structure of the wind, in particular of the high-speed one. The waves are of large amplitude and propagate outwards from the Sun, with long periods (hours). They have energy densities comparable both to the unperturbed magnetic field energy density and to the thermal energy density. Wave-driven wind models started to be constructed, by including a wave pressure and dissipation/plasma heating. These models were successful in explaining the fast solar wind, but only far from the Sun. Subsequent models assumed that close to the Sun Alfvén waves of high frequencies are created, which in turn would preferentially heat the protons via the cyclotron resonance (when the wave frequency matches the cyclotron frequency). According to these models, the solar wind would mainly be driven by the hot protons. Details and references can be found in a recent review by Hollweg (2008).

The solar wind is a highly turbulent medium. For a review of the observational and theoretical aspects, see the *living review* by Bruno & Carbone (2005). One major new findings from the SoHO UVCS spectrometer was the very large widths in the H I and O VI lines above coronal holes. By assuming that nonthermal broadening due to waves or turbulence is negligible, very large ion temperatures were found. For example, 3×10^8 K at $3.5 R_{\odot}$ from O VI. Many authors now regard these measurements as proof that preferential heating due to ion-cyclotron resonance is occurring, however more measurements of different ions would be needed to confirm this, and a theory of how high-frequency Alfvén waves can be generated close to the Sun needs to be developed. For observational aspects, see the review by Kohl et al. (2006), while for theoretical considerations see the *living review* by Marsch (2006).

10.3 The Sources of the Solar Wind

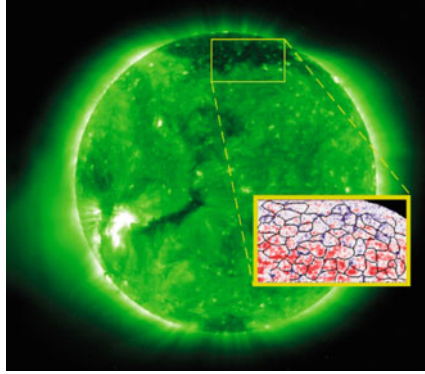
Coronal holes as the source of the fast solar wind was first established from Skylab X-ray observations and in-situ measurements by Krieger et al. (1973). This is now well-accepted. Doppler measurements by SoHO/SUMER clearly show blueshifted emission in coronal holes at the network boundaries (see, e.g., Hassler et al. 1999, and  Fig. 3-78), confirming the source of the fast solar wind.

Whether or not coronal hole plumes and jets contribute to the fast solar wind is still a matter of hot debate. Opposite arguments, mainly based on SoHO and Hinode observations, have been put forward. The difficulty lies in mapping back the in situ measurements, mainly velocities and magnetic fields, onto the Sun's surface.

It is still not entirely clear where the slow SW originates from. Many believe that comes from the boundaries between coronal holes and large streamers that undergo strong super-radial expansion in the corona, while others believe that it comes from narrow plasma sheets that extend out from the tops of streamer cusps.

One major contribution from the SoHO LASCO coronagraph has been the finding of the continuous large number of *bubbles* leaving the lower corona all the time. The plasmoids are observed to accelerate to an average velocity of about 300 km/s between 5 and $25 R_{\odot}$. The average velocity profile is consistent with an isothermal solar wind at 1.1 MK and a sonic point near $5 R_{\odot}$, in good agreement with the simple original Parker's model.

LASCO has also shown that CMEs provide an important contribution to the solar wind (in terms of mass and magnetic field), in particular during high solar activity, when many CMEs occur per day. During minimum conditions, the rate of CME production is about one order of magnitude lower.



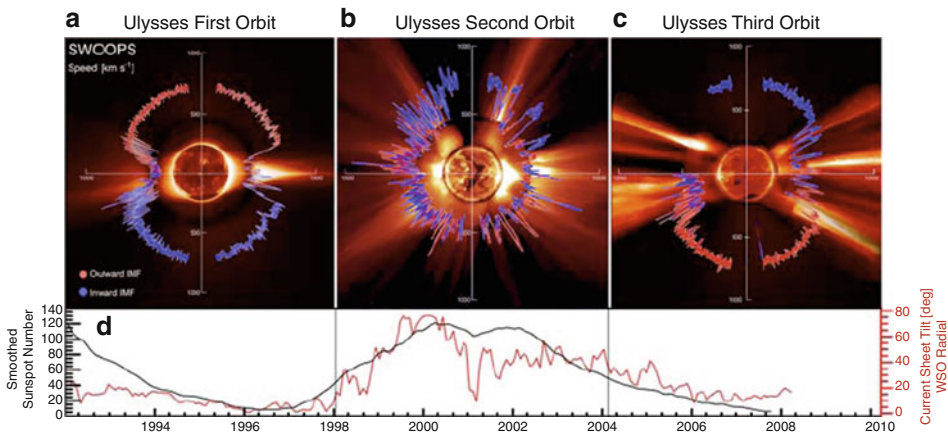
■ Fig. 3-78

The polar coronal hole regions show blueshifted emission in Ne VIII (formed around 0.6 MK) as observed by SoHO SUMER (Hassler et al. 1999). The image is from the SoHO EIT broadband at 195 Å, while the inset shows a Doppler velocity map from SoHO SUMER with superimposed the boundaries of the supergranular network (Image courtesy of D. Hassler)

10.4 Ulysses

Ulysses encountered the fast SW streaming out from the polar holes in 1992–1993. It crossed the south pole in June–November 1994 and the north pole in June–September 1995. During this time, the Sun was in solar minimum conditions for cycle 22. Extended polar coronal holes were present and symmetric, well-defined, streamers were present. Ulysses showed that the solid angle occupied by the fast SW is much larger than the solid angle occupied by the coronal hole at the solar surface, an indication of super-radial expansion. It also showed that the heliospheric magnetic field was mainly symmetric and dipolar. Pressure-balance (a decrease of magnetic pressure balanced by an increase of plasma pressure) structures were observed in the fast SW and were interpreted as the interplanetary manifestations of polar plumes.

Contrary to expectations, and Parker's model, the radial magnetic field strength was found to be nearly constant with latitude. Indeed the magnetic flux density $r^2 B_r$ was found to be constant with latitude by Ulysses. Ulysses clearly showed that the magnetic field embedded in the fast solar wind originates in coronal holes and fills over 60% of the heliosphere near solar minimum. Ulysses performed its second polar pass during the rise and peak phase of the solar maximum for cycle 23. It showed a complex, highly variable structure at all heliolatitudes, composed of flows arising from different sources (streamers, coronal mass ejections, coronal holes, and active regions). Ulysses has also provided important chemical abundance measurements, discussed in the abundance section, and empirical constraints on the temperatures at the freezing-in radius. The freezing-in radius is the distance from the Sun center where the expansion time of the solar wind equals the exchange time between the ions. Various assumptions have to be made, in order to deduce these kinds of temperatures, e.g., the expansion model, the charge state distribution, the elemental abundance, the distribution function of the electrons, the relative flow speeds of the different charge states. Different values are present in the literature. For many years, it seemed that a large discrepancy between results from the in situ and the remote-sensing observations was present, with the freezing-in temperatures being 2–3 times larger. Gloeckler and Geiss (2007) have recently re-evaluated Ulysses SWICS measurements to



■ Fig. 3-79

Polar plot of the solar wind speed as a function of heliographic latitude observed by Ulysses during the two/three polar passes during minimum and maximum (From McComas et al. 2008)

find a freezing-in temperature in fast SW from coronal holes of only 1.1 MK, in relatively good agreement with recent remote-sensing measurements at the base of the corona (cf. Del Zanna et al. 2008). The plasma freezing-in temperature, calculated from the observed O^{7+}/O^{6+} ratio by Ulysses, correlates well with the solar wind speed, confirming that the fast SW originates in low-temperature regions of the corona.

Another important result from Ulysses is that, unexpectedly, the number of GCRs measured did not increase at higher latitudes. A North-South asymmetry in the radial field component, with the field being stronger in the South was another key discovery. These types of asymmetries have recently received a great deal of attention, because they might be related to the overall change in the heliosphere which started around 2004, just after the last peak of activity (2001–2002). The polar fields also appear to be significantly lower than expected.

Ulysses completed a third polar pass during its third orbit in 2007. Preliminary results are presented in McComas et al. (2008) (see ► Fig. 3-79). The last pass showed that the fast solar wind occupied a much reduced solid angle, compared to the previous minimum conditions, where the heliospheric magnetic field was closer to dipolar. The third orbit also showed that the fast wind was slightly slower and significantly less dense and cooler, when compared to the 1996 solar minimum. This resulted in much lower dynamic (by 22%) and thermal (by 25%) pressures. The significant, long-term trend toward lower dynamic pressures means that the heliosphere has been shrinking. This could possibly be the cause of the early crossings of the termination shock by Voyager 1 in 2004 and Voyager 2 in 2007. After more than 18 years of operation, the Ulysses mission terminated on June 30, 2009.

11 Solar Irradiance

11.1 Introduction

No one can deny the crucial influence that the Sun has on the Earth's climate. The total amount of solar radiation reaching unit area of the Earth's atmosphere is called the total solar

irradiance (TSI). The TSI is the main heating input to our climate, and directly affects the surface temperatures, given that the bulk of the photons are emitted at visible wavelengths, hence are not significantly absorbed by the atmosphere.

The solar irradiance is of particular importance for the study of the Sun as a star, since it is the only case where detailed information about the distribution of the sources on its surface is available. Hence, it is the only case where predictions of the irradiance can be tested in detail. The only drawback is the fact that all the measurements to date have been done within the ecliptic.

The variations in the solar influence on our planet are not simply due to inherent variations in solar energy. Actually, significant variations in solar forcing on our planet are due to the Milankovitch cycles, i.e., changes in the eccentricity ($\sim 100,000$ years), the obliquity (angle between the equator and the orbital plane) ($\sim 40,000$ years) and the precession of the Earth's axis ($\sim 20,000$ years).

Aside from the total solar irradiance, there are three other key mechanisms which might impact the Earth's climate: (1) Solar UV and X-ray irradiance, which directly affects the temperature and composition of the upper and middle atmosphere (e.g., the amount of Ozone). Changes in UV, X-ray, and total solar irradiance vary in phase with the 11-year solar cycle. When the Sun is active, there is higher solar irradiance, with large increases in UV and X-ray emission. (2) Solar energetic particles, which ionize constituents of the middle and upper atmosphere, affecting atmospheric chemistry and temperature. (3) Galactic cosmic rays, which ionize constituents of the lower atmosphere and may thereby seed the formation of clouds.

The linkages between total solar irradiance (the dominant mechanism) and climate are included in climate models, but the other three mechanisms are generally not included. Some models do include changes in UV, but not the associated effects on atmospheric chemistry. It is important that sophisticated, integrated models are developed, in order to fully understand and predict the complex interaction between the Sun and the Earth.

11.2 Total Solar Irradiance, TSI, and Surface Temperature

The TSI was once known as the solar constant. It has been measured on the ground for a long time. An extensive set of measurements were obtained by Langley and others (Smithsonian Institution of Washington) from 1902 until the 1960s, with an averaged value of $1,353 \text{ W m}^{-2}$. Large (50% or so) corrections due to atmospheric absorption are required, and therefore some uncertainties are present.

More accurate measurements have been carried out with instruments in space. The Earth Radiation Budget (ERB) cavity instrument on board the NASA Nimbus satellite, launched in 1978, started to obtain irradiance measurements from space. The Active Cavity Radiometer Irradiance Monitor (ACRIM) instrument aboard SMM provided data between 1980 and 1989. These instruments were very precise and detected variations on all timescales, from daily ones of 0.01% or so, to weekly ones of up to 0.25%, and long-term ones of about 0.15% along the solar cycle [see Fröhlich and Lean (2004) for a review of measurements of the solar irradiance]. It soon became clear that the dips in the total irradiance which lasted a few days were related to the meridian passage of large Sunspot groups. On the other hand, general variations are due to a fine balance between the decrease in irradiance due to cooler (darker) emission in Sunspots and the bright UV emission in plage/faculae areas. One interesting fact is that the areas covered by

faculae vs. sunspots grows with activity relative to sunspots. It is still unclear how this relation applies to active stars.

Many other instruments have measured the TSI, including the Earth Radiation Budget Experiment (ERBE), the Upper Atmosphere Research Satellite (UARS), the European Retrievable Carrier (EURECA), the SoHO VIRGO (with two instruments), ACRIMSAT, and most recently the Solar Radiation and Climate Experiment (SORCE), with the Total Irradiance Monitor (TIM).

► **Figure 3-80** (left) shows a composite of some of the measurements. The Nimbus, SMM, and UARS data were obtained from the US National Geophysical Data Center at NOAA. The SoHO VIRGO data are the version 2, which includes various corrections for instrument degradation. They were obtained from PMOD/WRC, Davos, Switzerland. The SORCE/TIM data were obtained from the instrument database. Large uncertainties in the absolute value of the solar constant exist, given that over the last 30 years there has been a large scatter of values. The latest measurements from SORCE are not consistent with previous ones but are thought to be the most accurate, providing a value of about 1361 W m^{-2} during the period of low solar activity in 2007–2009. The long-term trends due to the solar cycle are evident. ► **Figure 3-80** (right) shows a composite of the TSI values. It seems that the latest minimum was significantly lower than previous ones (Fröhlich 2009).

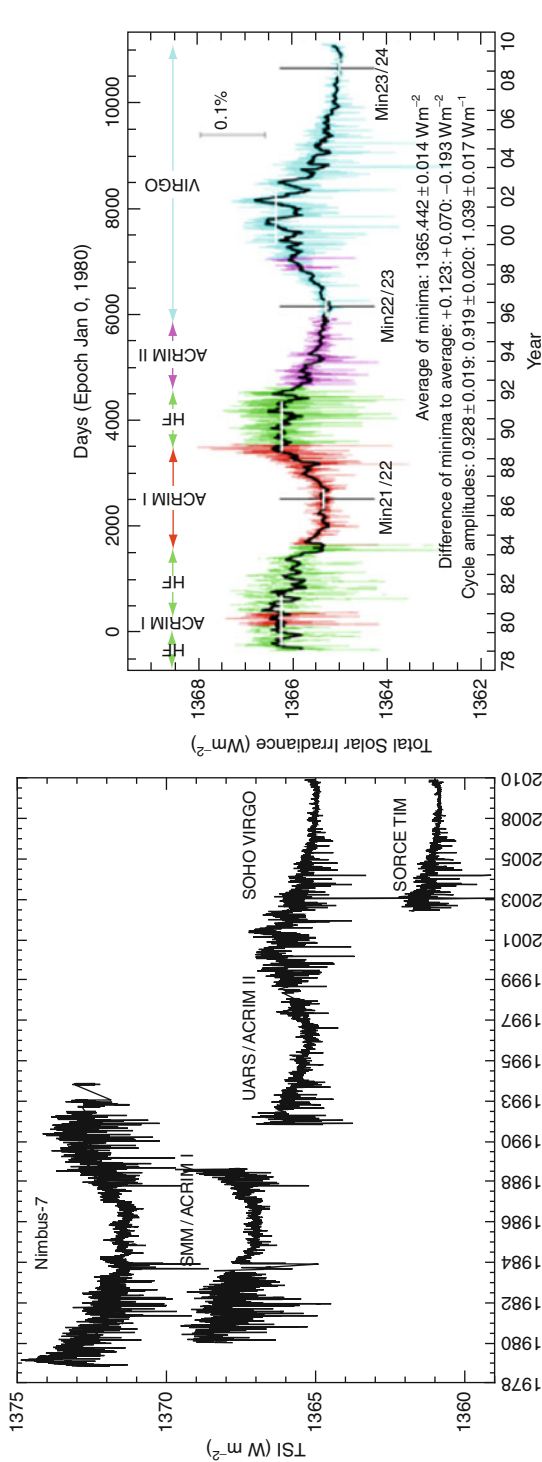
Further measurements will be obtained in the future. For example, PICARD is a new space mission dedicated to simultaneous measurements of the solar diameter, and of the spectral and total solar irradiance.

The issue as to whether the characteristics of the regions of the Sun with small magnetic field density are changing with the cycle is a complex one. Radiance measurements from space in the EUV and UV suffer from instrument calibration issues. Ground-based measurements of equivalent widths (EW) are independent from radiometric calibration issues, hence are very reliable. The EW of photospheric and chromospheric lines (e.g., Ca II) over the “quiet Sun” (often Sun center) have not changed over the last three solar cycles (Livingston et al. 2007), thus providing evidence that the basal photospheric–chromospheric emission has been relatively constant over this period.

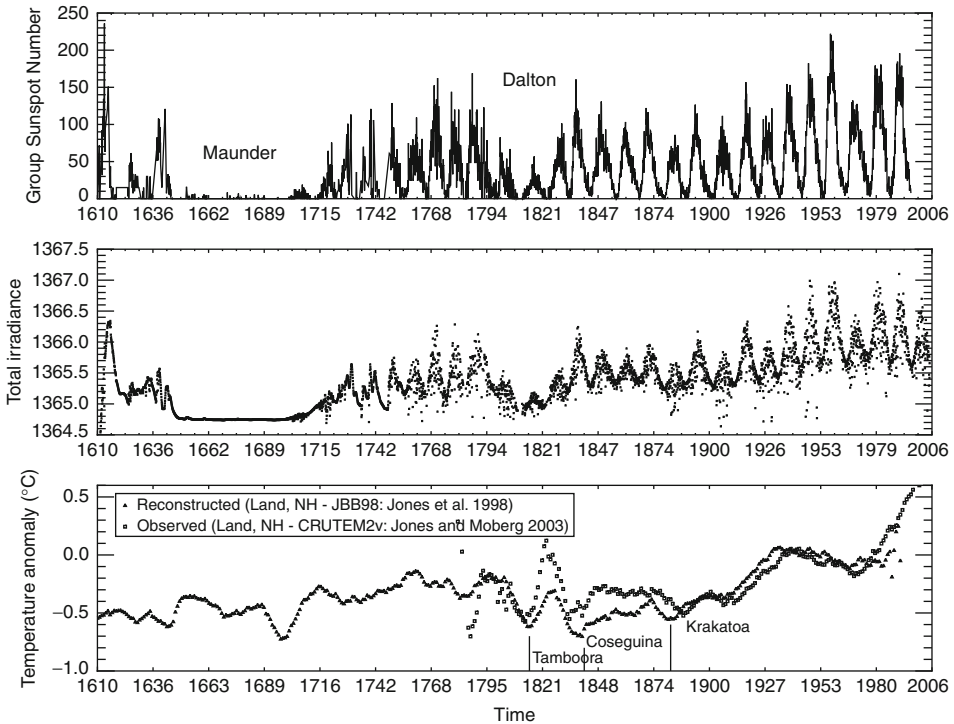
In the last decade, some effort has gone into modeling the solar irradiance. The reasons are twofold. One goal is to provide reliable input to climate models. Currently, most climate models use various proxies for the solar activity and some use averaged spectral irradiances. Another goal is to model the effects of the magnetic activity in young stars (including the young Sun) on planets. Recently, there has been particular interest in predicting the environment in which young planets were formed. The strong UV and EUV irradiance in young stars is an important factor for the development of life. The same is true for the Sun.

Models of solar irradiance are inherently semiempirical and have large uncertainties, given that are based upon a large number of assumptions and indirect measurements (i.e., on proxies of solar activity). One proxy widely used is the *sunspot number*. In fact, records of sunspots exist from the early seventeenth century, for example, from Galileo Galilei. This sunspot proxy is rather crude, given that various visibility effects are present (see, e.g., Dalla et al. 2008; Maunder 1907). ► **Figure 3-81** (top) shows the Group Sunspot Number, obtained by Hoyt and Schatten (1998) from a vast collection of archival data.

For a recent example of irradiance reconstruction see, e.g., Krivova et al. (2007), who estimated, based on the ideas developed by Solanki et al. (2002), the total irradiance since 1611 using the Group Sunspot Number. The data are displayed in ► **Fig. 3-81** (middle). Many such reconstructions exist in the literature, each differing in some respect. Large uncertainties are



■ Fig. 3-80 Absolute measurements of the total solar irradiance (TSI) from various satellites (left). A composite normalized TSI, courtesy of PMOD/WRC, Davos, Switzerland (right)



■ Fig. 3-81


Top: the Group Sunspot Number, indicating the level of solar activity since 1611. The long-term Maunder minimum is evident, as well as the shorter Dalton minimum. **Middle:** reconstructed total solar irradiance, from Krivova et al. (2007) (data have been smoothed). **Bottom:** temperature anomaly (data have been smoothed), observed (land, northern hemisphere), and reconstructed (land, northern hemisphere). The timings of large-scale volcanic eruptions (Tambora, Coseguina and Krakatoa) are also shown

present, in particular for the early years. However, all reconstructions predict a modest variation in TSI in the last 400 years.

It has been suggested (cf. Eddy 1976) that a relation exists between total solar irradiance and surface temperatures in the northern hemisphere. The most famous example is the Maunder Minimum in sunspot numbers, which occurred during the second half of the seventeenth century, and coincided with a period, known as the *Little Ice Age*, during which temperatures in the Northern Hemisphere (Europe in particular) were exceptionally low.

Direct surface temperature measurements are available only from the late nineteenth century. However, various proxies are used to estimate surface temperature variations, all with a significant degree of uncertainty, given that temperature variations are not the sole driving factor. Among the various methods, one uses tree rings. Another method is to look at isotopic concentrations in ice cores, obtained from Greenland and Antarctica. Another method uses sedimentary deposits.

❶ **Figure 3-81** (bottom) shows the variation in a set of observed and reconstructed temperatures. The observed ones are land-only summer temperatures for the northern hemisphere

(latitudes between 20° and 90°) from Jones and Moberg (2003). The reconstructed ones are land temperatures for the northern hemisphere obtained by Jones et al. (1998). The temperatures are displayed as “anomaly”, i.e., as the variation with respect to the average temperatures recorded during the 1961–90 period. The data were obtained from the Fourth Assessment Report (AR4) of the Working Group 1 on paleoclimate, for the Intergovernmental Panel on Climate Change (IPCC). The timings of large-scale volcanic eruptions (Tamboora, Coseguina and Krakatoa), known to cause short-term global decreases in temperature, are also shown in  Fig. 3-81.

As is the case of TSI reconstructions, a number of temperature reconstructions exist, all differing quite significantly. However, some correlation between solar activity and surface temperatures appears to exist. However the recent global warming, the sharp rise in the Earth's surface temperature over the past 50 years, cannot be explained by the Sun's influence. The solar irradiance, although varying with the solar cycle over the past 50 years, has increased but not dramatically. The vast majority of scientists are convinced that global warming is man-made and is due to the increase in the amount of “greenhouse gases” in the Earth's atmosphere.

One major problem in accepting a link between total solar irradiance and surface temperatures has been the fact that global temperatures show little correlation, and that the forcing due to the small irradiance variations is too small to have a direct effect. However, there is ample evidence for solar forcing on global circulation patterns and pressures (e.g., the North Atlantic Oscillation (NAO) which has a strong influence on the climate of Northern Europe), and there is growing consensus that even a small variation of the solar energy output could produce significant effects on some local climate patterns. For a review of some of these issues, see, e.g., Haigh (2007).

One interesting question is whether the Sun's variability on long timescales is peculiar or not. This issue is not yet resolved. Baliunas and Jastrow (1990) used the Mount Wilson Ca II H&K survey on what was believed to be a good sample of solar analogs, to find that a significant number of K,G stars had activity levels significantly below that of the Sun at solar minimum. Some of the findings have been subsequently questioned; however, it is quite evident that cycle variations are present in many stars. One puzzling issue is that a large number of stars have not shown cyclic variations so far. Stars that have cyclic variations are generally more active. One problem with the use of Ca II H&K observations is that the excess flux in the cores of these absorption lines only varies by small (10% or so) amounts during the solar cycle, and it decreases with stellar age, hence it is not a very sensitive proxy. Much better proxies would be fluxes in the EUV and X-rays, however most observations (e.g., ROSAT) have been carried out in broadbands, which, even in the X-rays, can be contaminated by emission formed in very different atmospheric layers.

A significant amount of data are now available since 1999 from the XMM-Newton and Chandra X-ray missions, and a more complete picture of stellar variability in the X-rays will become available. Missions such as Kepler, Corot, GAIA are going to give a substantial contribution with more accurate measurements of stellar characteristics (e.g., age, size, mass, rotation), so more detailed comparative studies with the Sun as a star will be possible.

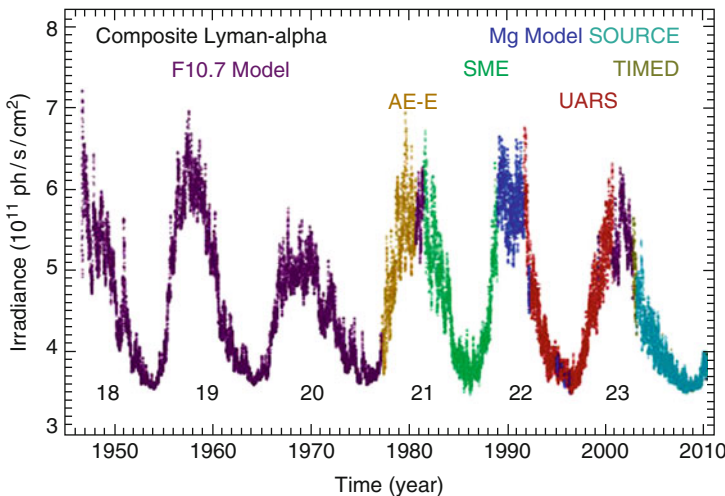
It is also possible to look back in time and see if the past few centuries were typical or not. This is done by using cosmogenic isotopes as proxy indicators of past variations in solar irradiance on long timescales. These isotopes are spallation products of galactic cosmic rays (GCRs) impacting Earth's atmosphere, and then absorbed and deposited into e.g., ice sheets, tree rings etc. The number of GCRs is strongly reduced during periods of solar activity, given that the heliospheric magnetic fields act as a shield. Therefore, GCRs are actually a proxy of heliospheric magnetic fields and not directly of the photospheric ones which modulate the solar irradiance.

The cosmogenic isotopes ^{10}Be (polar ices) and ^{14}C are mostly used. Their variations follow to some degree the long-term variations in solar activity (e.g., Sunspot numbers), and the Maunder minimum is clearly visible. The records show that Maunder-like minima have occurred quite frequently. See, e.g., Beer et al. (2006) for a recent review based on ^{10}Be measurements. The ^{10}Be records also show that the Sun has been unusually active in the last 50 years or so, although this is not uncommon within the last 10,000 years (see, e.g., Abreu et al. 2008). The increased solar activity in the last 50 years has led to suggestions that global warming could also be related to the Sun's activity. However, anthropogenic forcing remains the dominant effect.

11.3 Irradiance in the UV

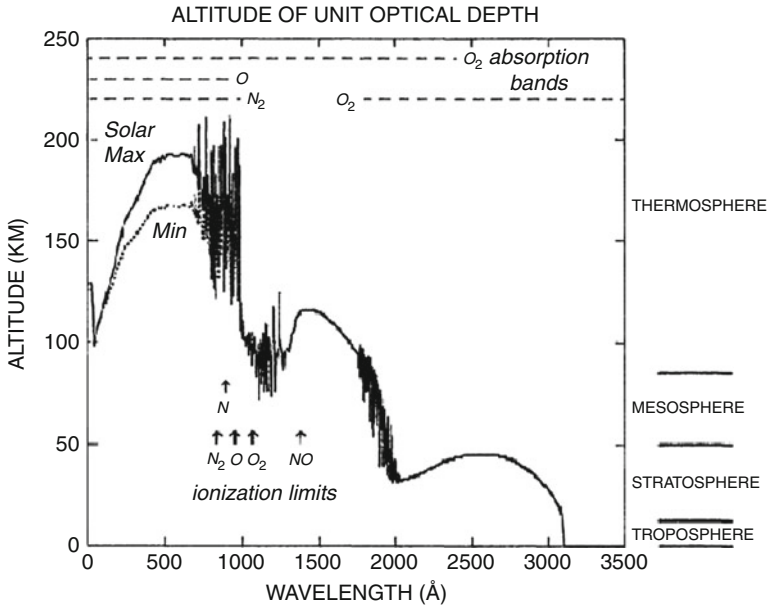
Variations with the solar cycle occur at all wavelengths and are up to 20%, depending on which UV wavelength is considered. The neutral hydrogen Ly- α is by far the strongest transition in the UV. It is mainly absorbed in the upper mesosphere. The neutral hydrogen Ly- α varies significantly (see Woods et al. 2000), as [Fig. 3-82](#) shows.

UV light controls the amount of stratospheric ozone through a series of reactions. The most important one is the photodissociation of O_2 , which occurs at wavelengths less than 2420 Å in the stratosphere (cf. [Fig. 3-83](#)). The oxygen atoms then react with oxygen molecules to produce ozone molecules. Various other processes are at work, at different heights in the atmosphere, the net effect being a peak in ozone around 25 km in equatorial regions. Stratospheric ozone abundance increases during periods of higher solar activity, although other effects such as global circulation and energetic particles change its abundance. Energetic particles from cosmic rays or from solar events such as flares or CME tend to reduce on short timescales the ozone abundance. The ozone concentration, among other parameters, is predicted to have important



■ Fig. 3-82

A composite of the irradiance in the neutral hydrogen Ly- α , courtesy of the Laboratory for Atmospheric and Space Physics, University of Colorado in Boulder



■ Fig. 3-83

Unity optical depth of the atmosphere for absorption of the incoming solar radiation, as a function of altitude (Meier 1991)

effects on global climate, as simulations show (see Haigh 1996, 2007). The processes are highly nonlinear and are complex to model.

The Solar Stellar Irradiance Comparison Experiment (SOLSTICE) and the Solar Ultraviolet Spectral Irradiance Monitor (SUSIM) on UARS have been providing UV measurements of the UV spectral irradiance since 1991. Solar Mesosphere Explorer (SME) also monitored the UV irradiance from 1981 until 1989. The Solar EUV Experiment (SEE) aboard the Thermosphere, Ionosphere, Mesosphere, Dynamics, and Energetic (TIMED) satellite, have made measurements since 2002.

Broad-band measurements have also been done, recently by Lyra aboard Proba2, which measures the UV irradiance in 4 UV passbands. Variations on all timescales have been observed, from short ones due to flares to those lasting a few days and related to the meridian passage of solar features. Plage areas are much brighter than sunspots in the UV, compared to the difference in the visible, which explains why larger variations are observed. How such variations produce some effect – however small – on Earth's climate is still unclear. The same applies to the EUV radiation.

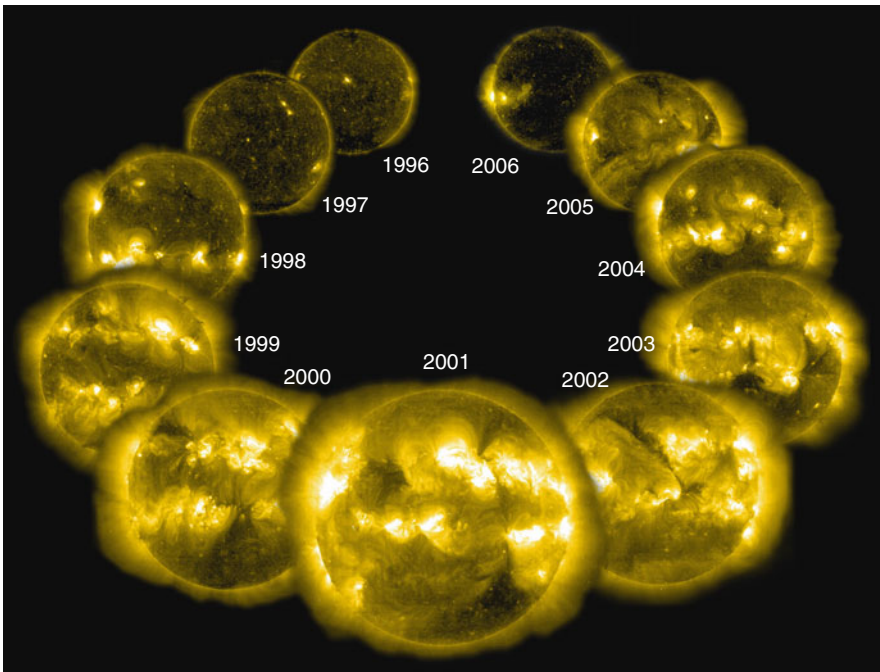
11.3.1 Irradiance in the EUV and X-rays

Some effects of the EUV radiation on Earth's atmosphere – particularly on the Thermosphere – are well known. The solar EUV radiation is the dominant factor in the formation of our ionosphere via photoionization and molecular photo-dissociation. EUV photons with wavelengths shorter than the neutral hydrogen Ly- α are absorbed in the Thermosphere and are

those which ionize neutral oxygen and N_2 , as well as O_2 and other molecules. Variations in the EUV irradiance directly affect the density of the ionosphere and have a significant impact on satellite drag.

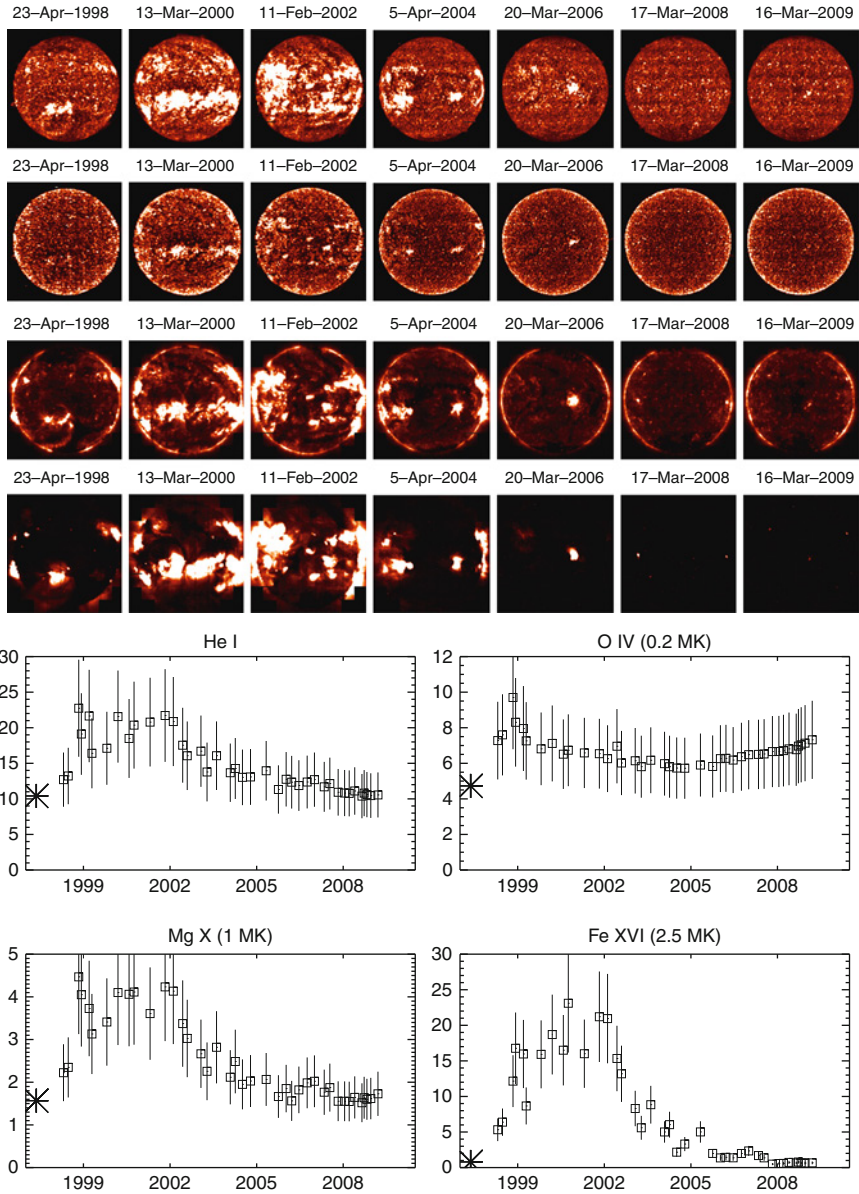
Only a few measurements are available, from occasional sounding rockets, which give only a snapshot of the solar irradiance, or cover less than one solar cycle or partial spectral ranges. Continuous measurements have only been carried out since 2002 with the NASA TIMED SEE experiment, that has provided EUV spectra with a coarse spectral resolution.

SoHO has been observing the Sun since 1996. Apart from the broadband measurements of EIT (see ● Fig. 3-84) and the CELIAS/*Solar Extreme Ultraviolet Monitor* (SEM), the Coronal Diagnostics Spectrometer (CDS) has been routinely observing the full solar disc in the ranges 31–38 and 51.5–63 nm with its *Normal Incidence Spectrograph* (NIS). The CDS instrument has been radiometrically calibrated using sounding rockets and by studying the long-term trends in the quiet Sun radiances. The calibrated radiances have been converted to irradiances (Del Zanna and Andretta 2010). Lines formed in the low transition region show relatively modest changes (by 50% at most). Lines forming around 1 MK change during the cycle by a factor ~ 5 ; for hotter lines (2.5 MK) the variability reaches factors of the order of 40 (see ● Fig. 3-85). The presence of active regions during the solar cycle has a strong influence even in regions of the more “quiet corona” which change considerably with the cycle, in particular for lines formed above 1 MK.



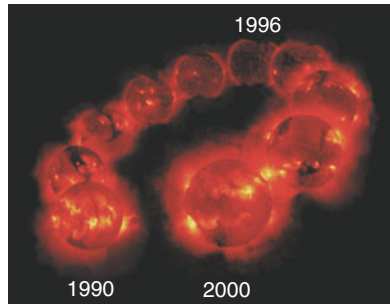
■ Fig. 3-84

SoHO EIT images in the 284 Å band, which is dominated by active region emission at 3 MK during periods of high solar activity (Courtesy of the SoHO/EIT consortium and Steele Hill (NASA GSFC))



■ Fig. 3-85

SoHO CDS radiances and irradiances in a selection of emission lines (Adapted from Del Zanna and Andretta 2010) from 1998 until 2009. The *top* images correspond to He I, O IV, Mg X, Fe XVI emission and each row has the same intensity scale. The *bottom* plots show the SoHO CDS irradiances in the same lines



■ Fig. 3-86

Yohkoh SXT X-ray images during one solar cycle, from maximum (1990) to minimum (1996) and last maximum (2000)

The extreme ultraviolet variability experiment (EVE) on board the NASA Solar Dynamics Observatory (SDO), the first mission of the NASA Living With a Star (LWS) programme, is measuring the solar EUV irradiance from 1 to 1050 Å with medium spectral resolution (1 Å) and high temporal cadence (10 s).

Many observations in the XUV have been done with photometers or broadband instruments; hence no calibrated irradiance measurements are available, in contrast with the stellar case where spectroscopic observations from Chandra and XMM-Newton are available since 1999. The solar observations include X-ray Sensors (XRS) aboard the NOAA GOES spacecrafts, the SoHO solar EUV monitor (SEM), the solar XUV photometers (SXP) on the Student Nitric Oxide Explorer (SNOE). X-ray broad-band images from Yohkoh SXT obtained in the 1990s (see ● Fig. 3-86) showed variability by two orders of magnitude.

12 Future Prospects

The future couple of decades offer an exciting prospect for solar physics. Various new satellites are being proposed and built. A follow-up of the successful Solar-A (Yohkoh) and Solar-B (Hinode) is being proposed. New missions to probe the inner heliosphere are being designed: the Russian InterHelios and the ESA-NASA Solar Orbiter satellites will go close to the Sun and above the ecliptic. The NASA solar Probe+ will go even closer to the Sun.

Besides the new observatories, many smaller rockets and SMEX (small explorer missions) are planned. The level of detail we shall be able to measure and see in the solar atmosphere continues to increase dramatically. The challenge is for the theory and models to keep up with these observations. Increases in the size and capacity of computers helps in this regard, but the modeling codes (MHD, kinetic, hydrodynamic etc.) require faster and larger processors. To date many codes have been developed independently, but the theoretical challenge is to bring all these codes (MHD + kinetic theory + hydrodynamic processes) together. This is likely to need the modeling of nonequilibrium processes (non-Maxwellian velocity distributions, two fluid models, time-dependent ionization and recombination, radiative transfer in the lower atmosphere).

The great challenge as always in solar physics is to keep pace with the complexity of the observations, to develop new theoretical techniques to accurately model the phenomena which

the Sun exhibits. The Sun, our nearest star, does not make this challenge easy. However, unless we can understand the physics of the Sun we have little chance of understanding the complex processes in other stellar atmospheres. Observations of other stars do however also provide a spectrum of stellar conditions in which we can place the Sun. In this regard, studies of the Sun and other stars go hand-in-hand.

12.1 Solar Orbiter

Solar Orbiter is one of ESA's candidates for a medium-class mission part of the Cosmic Vision programme to be launched no earlier than 2017. The orbit of the Solar Orbiter is unique in that it is highly elliptical with perihelion as low as 0.23 AU and with increasing inclination up to more than 30° with respect to the solar equator over the years, obtained with Venus flybys. This is technically very challenging, in particular with the temperature extremes which will be experienced.

The main goal of Solar Orbiter is to address the following questions:

- What are the origins of the solar wind streams and the heliospheric magnetic field?
- What are the sources, acceleration mechanisms, and transport processes of solar energetic particles?
- How do coronal mass ejections evolve in the inner heliosphere?
- Explore, at all latitudes, the energetics, dynamics, and fine-scale structure of the Sun's magnetized atmosphere.

This will be achieved with a suite of instruments, built in collaboration with NASA. The primary ones will be the in situ ones, which will measure the solar wind close to the Sun, a fundamental requirement. The main advantages over previous missions are: close-up view point; out-of-ecliptic view point; near corotating orbital phases; multi-point in situ observations by Solar Orbiter and Solar Probe Plus. The close-up view point will provide unprecedented spatial resolution, although it should be noted that the same could be achieved by Earth-orbiting satellites. The key advantage is the near corotating orbital phases, i.e., it should be possible to relate the in situ measurements with the sources, as observed with remote sensing (EUV imaging and spectroscopy). A coronagraph will provide unprecedented details about the inner corona.

12.2 Solar Probe+

The Solar Probe+, now scheduled for launch in 2015, is a modification of a previous NASA mission. It is designed to approach the Sun as close as 8.5 radii from its surface with a number of flybys over a period extending over both solar minimum and solar maximum. The mission will be highly complementary to Solar Orbiter and will make measurements of the solar wind plasma within its birthplace. Its baseline payload is an integrated package of in situ instruments and a white-light Hemispheric Imager (HI) for imaging coronal structures. Its scientific goals are similar to the Solar Orbiter ones:

- To determine the structure and dynamics of the magnetic fields at the sources of the solar wind.
- To trace the flow of energy that heats the corona and accelerates the solar wind.

- To determine what mechanisms accelerate and transport energetic particles.
- To explore dusty plasma near the sun and its influence on solar wind and energetic particle formation.

A quote from E.N.Parker, the father of the solar wind:

- ▶ *"It is ironic that voyaging spacecraft have roughed out the general picture of the solar wind region-the heliosphere-with its spiral magnetic field and the fast and slow stream interaction regions all the way out to the termination shock at 100 AU, while the heat sources that create the wind are still poorly understood. We expect that the coronal heating is a complex phenomenon, different over the quiet and active regions of the Sun. The ideal space mission to explore this problem would involve detailed in-situ investigation of plasma temperature, composition, and bulk motion; fast particle populations; magnetic field; and wave motions from the base of the corona out to 1 AU at many locations around the Sun at times of both high and low activity. Such a space mission is impossible, of course, because the Sun, being a star, is not readily approachable, but we would like to approach it as closely as money and technology permit."*

References

- Abreu, J. A., Beer, J., Steinhilber, F., Tobias, S. M., & Weiss, N. O. 2008, *Geophys Res Lett*, 35, 20109
- Anders, E., & Grevesse, N. 1989, *Geochim. Cosmochim. Acta*, 53, 197
- Andretta, V., Del Zanna, G., & Jordan, S. D. 2003, *A&A*, 400, 737
- Andretta, V., Mauas, P. J. D., Falchi, A., & Teriaca, L. 2008, *ApJ*, 681, 650
- Antiochos, S. K. 1998, *ApJ*, 502, L181
- Antonucci, E., et al. 1982, *Sol Phys*, 78, 107
- Aschwanden, M. 2006, *Physics of the solar corona*. (Chichester, UK: Praxis Publishing Ltd)
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
- Athay, R. G., eds. 1976, *The Solar Chromosphere and Corona: Quiet Sun, Astrophysics and Space Science Library*, Vol. 53 (Dordrecht: Reidel Publishing Co.)
- Athay, R. G., & Thomas, R. N. 1961, *Physics of the Solar Chromosphere* (New York: Interscience Publishers)
- Babcock, H. W. 1961, *ApJ*, 133, 572
- Badnell, N. R., Bautista, M. A., Butler, K., Delahaye, F., Mendoza, C., Palmeri, P., Zeippen, C. J., & Seaton, M. J. 2005, *MNRAS*, 360, 458
- Baliunas, S., & Jastrow, R. 1990, *Nature*, 348, 520
- Balogh, A., Lanzerotti, L. J., & Suess, S. T. 2008, *The Heliosphere Through the Solar Activity Cycle* (New York, NY: Springer-Praxis Books and Springer Science+Business Media)
- Balogh, A., Marsden, R. G., & Smith, E. J. 2001, *The Heliosphere Near Solar Minimum. The Ulysses Perspective* (Springer-Praxis books in Astrophysics and Astronomy)
- Basu, S., & Antia, H. M. 2004, *ApJ*, 606, L85
- Basu, S., & Antia, H. M. 2008, *Phys. Rep.*, 457, 217
- Beer, J., Vonmoos, M., & Muscheler, R. 2006, *Space Sci Rev*, 125, 67
- Belcher, J. W., & Davis, L., Jr. 1971, *J Geophys Res*, 76, 3534
- Benz, A. O. 2008, *Living Reviews in Solar Physics*, 5, 1
- Biermann, L. 1951, *Zeitschrift fur Astrophysik*, 29, 274
- Bochsler, P. 2007, *A&A Rev.*, 14, 1
- Bradshaw, S. J. 2008, *A&A*, 486, L5
- Bradshaw, S. J. 2009, *A&A*, 502, 409
- Bradshaw, S. J., & Mason, H. E. 2003, *A&A*, 401, 699
- Braun, D. C., & Birch, A. C. 2008, *Sol Phys*, 251, 267
- Brown, J. C. 1971, *Sol Phys*, 18, 489
- Bruno, R., & Carbone, V. 2005, *Living Reviews in Solar Physics*, 2
- Burgess, A. 1964, *ApJ*, 139, 776
- Burgess, A. 1965, *ApJ*, 141, 1588
- Burgess, A., & Summers, H. P. 1969, *ApJ*, 157, 1007
- Burton, W. M., Jordan, C., Ridgeley, A., & Wilson, R. 1971, *Philos Trans R Soc Lond A*, 270, 81
- Butler, R. P., Vogt, S. S., Marcy, G. W., Fischer, D. A., Henry, G. W., & Apps, K. 2000, *ApJ*, 545, 504
- Cargill, P. J. 1994, *ApJ*, 422, 381
- Carlsson, M., & Stein, R. F. 1997, *ApJ*, 481, 500
- Carrington, R. C. 1859, *MNRAS*, 20, 13
- Chamberlin, P. C., Woods, T. N., Crotser, D. A., Eparvier, F. G., Hock, R. A., & Woodraska, D. L. 2009, *Geophys Res Lett*, 36, 5102
- Charbonneau, P. 2005, *Living Reviews in Solar Physics*, 2

- Chifor, C. 2008, PhD thesis. Univ. of Cambridge, UK
- Chifor, C., Del Zanna, G., Mason, H. E., Sylwester, J., Sylwester, B., & Phillips, K. J. H. 2007, *A&A*, 462, 323
- Christensen-Dalsgaard, J. 2002, *Rev Mod Phys*, 74, 1073
- Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B., & Cortes, T. R. 1979, *Nature*, 282, 591
- Cowling, T. G. 1941, *MNRAS*, 101, 367
- Craig, I. J. D., & Brown, J. C. 1976, *A&A*, 49, 239
- Culhane, J. L., & Jordan, C. 1991, *The Physics of Solar Flares*. Proceedings of the Royal Society, UK
- Czaykowska, A., de Pontieu, B., Alexander, D., & Rank, G. 1999, *ApJ*, 521, L75
- Dalla, S., Fletcher, L., & Walton, N. A. 2008, *A&A*, 479, L1
- Davis, L. 1955, *Phys Rev*, 100, 1440
- De Pontieu, B., Erdélyi, R., & James, S. P. 2004, *Nature*, 430, 536
- Del Zanna, G. 1995, Master thesis, Univ. of Florence, Italy
- Del Zanna, G. 1999, PhD thesis, Univ. of Central Lancashire, UK
- Del Zanna, G. 2002, in *High Resolution X-ray Spectroscopy with XMM-Newton and Chandra*, ed. Branduardi-Raymont, G. (London: University College London, UK)
- Del Zanna, G. 2003, *A&A*, 406, L5
- Del Zanna, G. 2007, *ASPC*, Vol. 397, *First Science Results from Hinode*, 87
- Del Zanna, G. 2008, *A&A*, 481, L49
- Del Zanna, G., & Andretta, V. 2010, in *IAU Symposium*, Vol. 264, eds. A. G. Kosovichev, A. H. Andrei, & J.-P. Roelot, 78–80
- Del Zanna, G., Berlicki, A., Schmieder, B., & Mason, H. E. 2006a, in *10 Years of SOHO and Beyond*, *Sol Phys*, Vol. 617
- Del Zanna, G., Berrington, K. A., & Mason, H. E. 2004, *A&A*, 422, 731
- Del Zanna, G., & Bromage, B. J. I. 1999, *J Geophys Res*, 104, 9753
- Del Zanna, G., Bromage, B. J. I., & Mason, H. E. 2001, in *Solar and Galactic Composition*, *AIP Conf Proc* 598, 59
- Del Zanna, G., Bromage, B. J. I., & Mason, H. E. 2003, *A&A*, 398, 743
- Del Zanna, G., Gibson, S. E., Mason, H. E., Pike, C. D., & Mandrini, C. H. 2002a, *Adv Space Res*, 30, 551
- Del Zanna, G., Landini, M., & Mason, H. E. 2002b, *A&A*, 385, 968
- Del Zanna, G., & Mason, H. E. 2003, *A&A*, 406, 1089
- Del Zanna, G., Mason, H. E., & Cirtain, J. 2006b, in *10 Years of SOHO and Beyond*, *ESA Special Publication*, Vol. 617
- Del Zanna, G., Rozum, I., & Badnell, N. 2008, *A&A*, 487, 1203
- Dere, K. P., Bartoe, J., Brueckner, G. E., Cook, J. W., & Socker, D. G. 1987, *Sol Phys*, 114, 223
- Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., & Young, P. R. 1997, *A&AS*, 125, 149
- Dere, K. P., & Mason, H. E. 1981, in *Solar Active Regions: A Monograph from Skylab Solar Workshop III*, eds. F. Q. Orrall, 129–164 (Boulder: Colorado Associated University Press)
- Dere, K. P., & Mason, H. E. 1993, *Sol Phys*, 144, 217
- Deubner, F. 1975, *A&A*, 44, 371
- Dore, U., & Orestano, D. 2008, *Rep Prog Phys*, 71, 106201
- Doschek, G. A. 1990, *ApJS*, 73, 117
- Doschek, G. A., et al. 2007, *ApJ*, 667, L109
- Dowdy, J. F., Jr., Rabin, D., & Moore, R. L. 1986, *Sol Phys*, 105, 35
- Drake, J. F. 1971, *Sol Phys*, 16, 152
- Drake, J. J., Laming, J. M., & Widing, K. G. 1997, *ApJ*, 478, 403
- Dzifčáková, E., Kulinová, A., Chifor, C., Mason, H. E., Del Zanna, G., Sylwester, J., & Sylwester, B. 2008, *A&A*, 488, 311
- Eddy, J. A. 1976, *Science*, 192, 1189
- Edlén, B. 1942, *Zeitschrift für Astrophysics*, 22, 30
- Erdélyi, R., & Ballai, I. 2007, *Astronomische Nachrichten*, 328, 726
- Feldman, U. 1983, *ApJ*, 275, 367
- Feldman, U., Schühle, U., Widing, K. G., & Laming, J. M. 1998, *ApJ*, 505, 999
- Fisk, L. A., & Schwadron, N. A. 2001, *ApJ*, 560, 425
- Fletcher, L., & Hudson, H. S. 2008, *ApJ*, 675, 1645
- Fludra, A., & Schmelz, J. T. 1995, *ApJ*, 447, 936
- Fludra, A., & Schmelz, J. T. 1999, *A&A*, 348, 286
- Forbes, T. G., & Priest, E. R. 1995, *ApJ*, 446, 377
- Foukal, P. V. 2004, *Solar Astrophysics*, 2nd, Revised Edition (Wiley-VCH Verlag GmbH)
- Fröhlich, C. 2009, *A&A*, 501, L27
- Fröhlich, C., & Lean, J. 2004, *A&A Rev.*, 12, 273
- Fuhrmann, K. 2008, *MNRAS*, 384, 173
- Gabriel, A. H. 1976, *Philos Trans R Soc Lond A*, 281, 339
- Gabriel, A. H., Culhane, J. L., Patchett, B. E., Breeveld, E. R., Lang, J., Parkinson, J. H., Payne, J., & Norman, K. 1995, *Adv Space Res*, 15, 63
- Gabriel, A. H., & Jordan, C. 1969, *MNRAS*, 145, 241
- Gabriel, A. H., & Mason, H. E. 1982, in *Applied Atomic Collision Physics, Atmospheric Physics and Chemistry*, Vol. 1, eds. H. S. W. Massey, & D. R. Bates (New York: Academic Press), 45–397
- Geiss, J. 1982, *Space Sci Rev*, 33, 201
- Geiss, J., & Buergi, A. 1986, *A&A*, 159, 1
- Gizon, L., & Birch, A. C. 2005, *Living Reviews in Solar Physics*, 2, 6
- Gloeckler, G., & Geiss, J. 2007, *Space Sci Rev*, 130, 139
- Goldschmidt, V. 1938, *Skrifter Norske Videnskaps-Akad., Oslo I Math.-Naturv. Klasse*, 4, 148

- Golub, L., & Pasachoff, J. M. 2010, *The Solar Corona* (Cambridge: Cambridge University Press)
- Gough, D., & Toomre, J. 1991, *ARA&A*, 29, 627
- Güdel, M., Audard, M., Reale, F., Skinner, S. L., & Linsky, J. L. 2004, *A&A*, 416, 713
- Güdel, M., & Nazé, Y. 2009, *A&A Rev.*, 17, 309
- Haigh, J. D. 1996, *Science*, 272, 981
- Haigh, J. D. 2007, *Living Reviews in Solar Physics*, 4, 2
- Hassler, D. M., Dammasch, I. E., Lemaire, P., Brekke, P., Curdt, W., Mason, H. E., Vial, J., & Wilhelm, K. 1999, *Science*, 283, 810
- Hirayama, T. 1974, *Sol Phys*, 34, 323
- Hollweg, J. V. 2008, *Astron Astrophys*, 29, 217
- Hou, J. L., Prantzos, N., & Boissier, S. 2000, *A&A*, 362, 921
- Hoyt, D. V., & Schatten, K. H. 1998, *Sol Phys*, 179, 189
- Hughes, D. W., Rosner, R., & Weiss, N. O., eds. 2007, *The Solar Tachocline* (Cambridge: Cambridge Univ. Press)
- Hummer, D. G., Berrington, K. A., Eissner, W., Pradhan, A. K., Saraph, H. E., & Tully, J. A. 1993, *A&A*, 279, 298
- Hundhausen, A. J., Sawyer, C. B., House, L., Illing, R. M. E., & Wagner, W. J. 1984, *J Geophys Res*, 89, 2639
- Jain, R., Hindman, B. W., Braun, D. C., & Birch, A. C. 2009, *ApJ*, 695, 325
- Jones, P. D., Briffa, K., Barnett, T., & Tett, S. 1998, *The Holocene*, 8, 455
- Jones, P. D., & Moberg, A. 2003, *J Climate*, 16, 206
- Jordan, C., Smith, G. R., & Houdebine, E. R. 2005, *MNRAS*, 362, 411
- Jordan, S. 1981, in *The Sun as a star*, ed. S. Jordan, Vol. 450 (NASA Special Publication)
- Judge, P. 2006, in *Solar MHD Theory and Observations: A High Spatial Resolution Perspective*, eds. H. Uitenbroek, J. Leibacher, & R. F. Stein, 354, 259 (Astronomical Society of the Pacific Conference Series)
- Judge, P. G. 2000, *ApJ*, 531, 585
- Kane, S. R., Frost, K. J., & Donnelly, R. F. 1979, *ApJ*, 234, 669
- Karzas, W. J., & Latter, R. 1961, *ApJS*, 6, 167
- Kjeldseth Moe, O., & Nicolas, K. R. 1977, *ApJ*, 211, 579
- Kliem, B., Titov, V. S., & Török, T. 2004, *A&A*, 413, L23
- Kliem, B., & Török, T. 2006, *Phys Rev Lett*, 96, 255002
- Klimchuk, J. A. 2006, *Sol Phys*, 234, 41
- Kohl, J. L., Noci, G., Cranmer, S. R., & Raymond, J. C. 2006, *A&A Rev.*, 13, 31
- Krieger, A. S., Timothy, A. F., & Roelof, E. C. 1973, *Sol Phys*, 29, 505
- Krivova, N. A., Balmaceda, L., & Solanki, S. K. 2007, *A&A*, 467, 335
- Kunow, H., Crooker, N. U., Linker, J. A., Schwenn, R., & von Steiger, R. 2006, *Space Sci Rev*, 123, 1
- Laming, J. M. 2004, *ApJ*, 614, 1063
- Landi, E., Mason, H. E., Lemaire, P., & Landini, M. 2000, *A&A*, 357, 743
- Landini, M., & Monsignori Fossi, B. C. 1975, *A&A*, 42, 213
- Leibacher, J. W., & Stein, R. F. 1981, in *The Sun as a star*, ed. Jordan, S., Vol. 450 (NASA Special Publication), 263
- Leighton, R. B. 1964, *ApJ*, 140, 1547
- Leighton, R. B. 1969, *ApJ*, 156, 1
- Leighton, R. B., Noyes, R. W., & Simon, G. W. 1962, *ApJ*, 135, 474
- Lin, R. P., Schwartz, R. A., Kane, S. R., Pelling, R. M., & Hurley, K. C. 1984, *ApJ*, 283, 421
- Livingston, W., Wallace, L., White, O. R., & Giampapa, M. S. 2007, *ApJ*, 657, 1137
- Lodders, K., Palme, H., & Gail, H. 2009, *ArXiv e-prints*
- Low, B. C. 2001, *J Geophys Res*, 106, 25141
- Lyot, B. 1939, *MNRAS*, 99, 580
- Mariska, J. T. 1992, *The Solar Transition Region* (Cambridge: Cambridge University Press)
- Marsch, E. 2006, *Living Reviews in Solar Physics*, 3
- Martens, P. C. H., Kankelborg, C. C., & Berger, T. E. 2000, *ApJ*, 537, 471
- Mason, H. E. 1995, *Adv Space Res*, 15, 53
- Mason, H. E., Landi, E., Pike, C. D., & Young, P. R. 1999, *Sol Phys*, 189, 129
- Mason, H. E., & Monsignori Fossi, B. C. M. 1994, *A&A Rev.*, 6, 123
- Mason, H. E., Shine, R. A., Gurman, J. B., & Harrison, R. A. 1986, *ApJ*, 309, 435
- Maunder, A. S. D. 1907, *MNRAS*, 67, 451
- Maunder, E. W. 1904, *MNRAS*, 64, 747
- McComas, D. J., Ebert, R. W., Elliott, H. A., Goldstein, B. E., Gosling, J. T., Schwadron, N. A., & Skoug, R. M. 2008, *Geophys Res Lett*, 35, 18103
- Meier, R. R. 1991, *Space Sci. Rev.*, 58, 1
- Mestel, L. 1999, *Stellar Magnetism* (Oxford: Oxford University Press)
- Meyer, J. 1985, *ApJS*, 57, 173
- Miesch, M. S. 2005, *Living Reviews in Solar Physics*, 2
- Moore, R. L., & Sterling, A. C. 2006, *Washington DC American Geophysical Union Geophysical Monograph Series*, 165, 43
- Narain, U., & Ulmschneider, P. 1996, *Space Sci Rev*, 75, 453
- Nordlund, Å., Stein, R. F., & Asplund, M. 2009, *Living Reviews in Solar Physics*, 6, 2
- O' Dwyer, B., Del Zanna, G., Mason, H. E., Sterling, A., Tripathi, D., & Young, P. 2010, *A&A*, 525, A137
- Orrall, F. Q., ed. 1981, *Solar Active Regions: A Monograph from SKYLAB Solar Workshop III* (Boulder: Colorado Associated University Press)

- Parenti, S., Buchlin, E., Cargill, P. J., Galtier, S., & Vial, J. 2006, *ApJ*, 651, 1219
- Parker, E. N. 1958, *ApJ*, 128, 664
- Parker, E. N. 1988, *ApJ*, 330, 474
- Parnell, C. E., & Jupp, P. E. 2000, *ApJ*, 529, 554
- Peter, H., & Judge, P. G. 1999, *ApJ*, 522, 1148
- Pevtsov, A. A., Fisher, G. H., Acton, L. W., Longcope, D. W., Johns-Krull, C. M., Kankelborg, C. C., & Metcalf, T. R. 2003, *ApJ*, 598, 1387
- Phillips, K. J. H. 1992, *Guide to the sun* (Cambridge: Cambridge University Press)
- Phillips, K. J. H., Feldman, U., & Landi, E. 2008, *Ultraviolet and X-ray Spectroscopy of the Solar Atmosphere* (Cambridge: Cambridge University Press)
- Phillips, K. J. H., Pike, C. D., Lang, J., Zarro, D. M., Fludra, A., Watanabe, T., & Takahashi, M. 1995, *Adv Space Res*, 15, 33
- Pikel'Ner, S. B. 1969, *Soviet Astronomy*, 13, 259
- Pneuman, G. W., & Kopp, R. A. 1978, *Sol Phys*, 57, 49
- Pottasch, S. R. 1963, *ApJ*, 137, 945
- Pottasch, S. R. 1967, *Bull. Astron. Inst. Netherlands*, 19, 113
- Priest, E., & Forbes, T. 2000, *Magnetic Reconnection: MHD theory and applications* (Cambridge: Cambridge University Press)
- Priest, E. R. 1982, *Solar Magnetohydrodynamics* (Dordrecht: D.Reidel)
- Priest, E. R., Foley, C. R., Heyvaerts, J., Arber, T. D., Culhane, J. L., & Acton, L. W. 1998, *Nature*, 393, 545
- Priest, E. R., & Forbes, T. G. 2002, *A&A Rev.*, 10, 313
- Ramaty, R., & Murphy, R. J. 1987, *Space Sci Rev*, 45, 213
- Raymond, J. C., et al. 1997, *Sol Phys*, 175, 645
- Reames, D. V. 1999, *Space Sci Rev*, 90, 413
- Reeves, E. M. 1976, *Sol Phys*, 46, 53
- Roberts, B. 1979, *Sol Phys*, 61, 23
- Rosner, R., Tucker, W. H., & Vaiana, G. S. 1978, *ApJ*, 220, 643
- Russell, H. N. 1929, *ApJ*, 70, 11
- Rust, D. M. 1976, *Sol Phys*, 47, 21
- Rutten, R. J. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 368, *The Physics of Chromospheric Plasmas*, eds. P. Heinzel, I. Dorotović, & R. J. Rutten, 27
- Schwenn, R., & Marsch, E. 1990, *Physics of the Inner Heliosphere I. Large-Scale Phenomena* (Berlin: Springer)
- Schwenn, R., & Marsch, E. 1991, *Physics of the Inner Heliosphere II. Particles, Waves and Turbulence* (Berlin: Springer)
- Seaton, M. J. 1964a, *MNRAS*, 127, 191
- Seaton, M. J. 1964b, *Planet. Space Sci.*, 12, 55
- Serenelli, A. M., Basu, S., Ferguson, J. W., & Asplund, M. 2009, *ApJ*, 705, L123
- Solanki, S. K., Inhester, B., & Schüssler, M. 2006, *Rep Prog Phys*, 69, 563
- Solanki, S. K., Schüssler, M., & Fligge, M. 2002, *A&A*, 383, 706
- Spitzer, L. 1965, *Physics of Fully Ionized Gases* (New York: Interscience Publishers, Inc.)
- Stein, R. F., & Leibacher, J. 1974, *ARA&A*, 12, 407
- Stenflo, J.O., 1989, *A&AR* 1, 3-48
- Stix, M. 2002, *The Sun: An Introduction* (Berlin: Springer)
- Strong, K. 1978, PhD thesis, University College London, UK
- Sturrock, P. A., eds. 1980, *Solar Flares: A Monograph from SKYLAB Solar Workshop II* (Boulder: Colorado Associated University Press)
- Sturrock, P. A., & Hartle, R. E. 1966, *Phys Rev Lett*, 16, 628
- Suematsu, Y. 1990, in *Lecture Notes in Physics, Progress of Seismology of the Sun and Stars*, Vol. 367, eds. Y. Osaki & H. Shibahashi (Berlin: Springer Verlag), 211
- Thomas, J. H., & Weiss, N. O. 2008, *Sunspots and Starspots* (Cambridge: Cambridge University Press)
- Thompson, M. J., Christensen-Dalsgaard, J., Miesch, M. S., & Toomre, J. 2003, *ARA&A*, 41, 599
- Tobias, S. 2004, in *Fluid Dynamics and Dynamos in Astrophysics and Geophysics*, eds. A. Soward, C. Jones, D. Hughes, & N. Weiss (Oxford, UK: Taylor & Francis Group), 193
- Tobias, S., & Weiss, N. 2007, in *The Solar Tachocline*, eds. D. W. Hughes, R. Rosner, & N. O. Weiss (Cambridge: Cambridge Univ. Press), 319
- Tripathi, D., Mason, H. E., Dwivedi, B. N., del Zanna, G., & Young, P. R. 2009, *ApJ*, 694, 1256
- Tziotziou, K., Tsiropoulou, G., & Mein, P. 2003, *A&A*, 402, 361
- Ulrich, R. K. 1970, *ApJ*, 162, 993
- Veck, N. J., & Parkinson, J. H. 1981, *MNRAS*, 197, 41
- Veck, N. J., Strong, K. T., Jordan, C., Simnett, G. M., Cargill, P. J., & Priest, E. R. 1984, *MNRAS*, 210, 443
- Velli, M. 1994, *ApJ*, 432, L55
- von Steiger, R., & Geiss, J. 1989, *A&A*, 225, 222
- von Steiger, R., et al. 2000, *J Geophys Res*, 105, 27217
- Walker, A. B. C., Jr., Rugge, H. R., & Weiss, K. 1974, *ApJ*, 188, 423
- Weiss, N. 1994, in *Lectures on Solar and Planetary Dynamos*, eds. M. Proctor & A. Gilbert (Cambridge, UK; New York, USA: Cambridge University Press), 59
- Weiss, N. O., & Thompson, M. J. 2009, *Space Sci Rev*, 144, 53
- Widing, K. G., & Feldman, U. 1989, *ApJ*, 344, 1046
- Widing, K. G., & Feldman, U. 2001, *ApJ*, 555, 426

- Wimmer-Schweingruber, R. F., eds. 2001, in American Institute of Physics Conference Series, Vol. 598, Joint SoHO/ACE Workshop "Solar and Galactic Composition" (New York: American Institute of Physics)
- Withbroe, G. L. 1978, *ApJ*, 225, 641
- Wood, B. E. 2004, *Living Reviews in Solar Physics*, 1, 2
- Woods, T. N., Tobiska, W. K., Rottman, G. J., & Worden, J. R. 2000, *J Geophys Res*, 105, 27195
- Woods, T. N., et al. 2009, *Geophys Res Lett*, 36, 1101
- Young, C. 1870, *J. Franklin Inst*, 60, 232
- Young, P. R., & Mason, H. E. 1997, *Sol Phys*, 175, 523
- Young, P. R., et al. 2007, *PASJ*, 59, 857
- Zank, G. P. 1999, *Space Sci Rev*, 89, 413
- Zhitnik, I. A., et al. 2003, *MNRAS*, 338, 67
- Zirin, H. 1988, *Astrophysics of the Sun* (Cambridge: Cambridge University Press)
- Zirker, J. B., ed. 1977, *Coronal holes and high speed wind streams: a monograph from Skylab solar workshop I* (Boulder: Colorado Associated University Press)
- Zurbuchen, T. H. 2007, *ARA&A*, 45, 297

4 **Asteroseismology**

Gerald Handler

Institute for Astronomy, University of Vienna, Vienna, Austria

Warszawa, Poland

1	<i>Introduction: Variable and Pulsating Stars</i>	208
2	<i>Astrophysical Background</i>	211
2.1	Driving Mechanisms	211
2.2	Asteroseismology	212
3	<i>From the Telescope to a Seismic Model</i>	219
3.1	Basic Methods for Analyzing Asteroseismic Data	219
3.2	Methods for Mode Identification	224
3.3	Asteroseismic Modeling	226
4	<i>Applications</i>	229
4.1	Pulsating White and Pre-white Dwarf Stars	229
4.2	Delta Scuti Stars	231
4.3	Slowly Pulsating B and Gamma Doradus Stars	232
4.4	Beta Cephei Stars	233
4.5	Pulsating Subdwarf Stars	234
4.6	Rapidly Oscillating Ap Stars	234
4.7	Solar-Like Oscillators	235
4.8	Hybrid Pulsators	235
4.9	Pulsation in Eclipsing Binaries and Open Clusters	236
4.10	A New Era in Precision	237
4.11	Prospects and Problems	238
	<i>Acknowledgments</i>	239
	<i>References</i>	239

Abstract: Asteroseismology is the determination of the interior structures of stars by using their oscillations as seismic waves. Simple explanations of the astrophysical background and some basic theoretical considerations needed in this rapidly evolving field are followed by introductions to the most important concepts and methods on the basis of example. Previous and potential applications of asteroseismology are reviewed and future trends are attempted to be foreseen.

Keywords: Binaries: eclipsing, Cepheids, Convection, Delta Scuti stars, Hertzsprung-Russell diagram, Hydrodynamics, Line: profiles, Magnetic fields, Methods: data analysis, Nuclear reactions, Space vehicles, Stars: atmospheres, Stars: chemically peculiar, Stars: early-type, Stars: evolution, Stars: fundamental parameters, Stars: horizontal-branch, Stars: interiors, Stars: magnetic fields, Stars: oscillations (including pulsations), Stars: premain-sequence, Stars: rotation, Techniques: photometric, Techniques: radial velocities, Techniques: spectroscopic, Variables: other, White dwarfs

1 Introduction: Variable and Pulsating Stars

Nearly all the physical processes that determine the structure and evolution of stars occur in their (deep) interiors. The production of nuclear energy that powers stars takes place in their cores for most of their lifetime. The effects of the physical processes that modify the simplest models of stellar evolution, such as mixing and diffusion, also predominantly take place in the inside of stars.

The light that we receive from the stars is the main information that astronomers can use to study the universe. However, the light of the stars is radiated away from their surfaces, carrying no memory of its origin in the deep interior. Therefore, it would seem that there is no way that the analysis of starlight tells us about the physics going on in the unobservable stellar interiors.

However, there are stars that reveal more about themselves than others. *Variable stars* are objects for which one can observe time-dependent light output, on a time scale shorter than that of evolutionary changes. There are two major groups of variable star, extrinsic and intrinsic variables.


Extrinsic variables do not change their light output by themselves. For example, the light changes of eclipsing binary stars are caused by two stars passing in front of each other, so that light coming from one of them is periodically blocked. The individual components of eclipsing binary stars are not necessarily variable. By analyzing the temporal light variations and orbital motion of eclipsing binaries, one can determine their fundamental properties, and by assuming that their components are otherwise normal stars, determine fundamental properties of all stars, most importantly their masses. In this way, stars and stellar systems can be understood better.

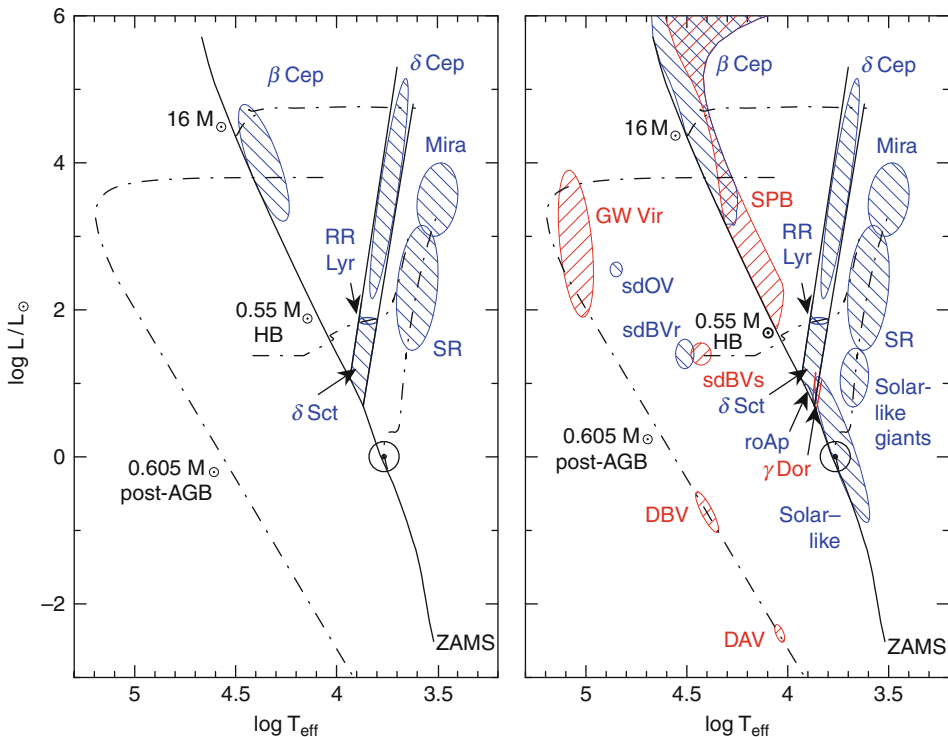
Intrinsic variables, on the other hand, change their light output physically. Supernovae, which are stellar implosions/explosions, can become brighter than their host galaxies because of the ejection of large amounts of material. Even more revealing are stars that vary their sizes and/or shapes: *pulsating variables*.

The first pulsating star was discovered more than 400 years ago. In 1596, David Fabricius remarked that the star *o Ceti* (subsequently named “Mira”, the wonderful) disappeared from the visible sky. About 40 years later, it was realized that it did so every ~ 11 months; the first periodic


variable star was known (although we know today that in this case, the term “periodic” is not correct in a strict sense).

In 1784, John Goodricke discovered the variability of δ Cephei, and in 1914 enough evidence had been collected that Harlow Shapley was able to demonstrate that the variations of δ Cephei and related stars (also simply called “Cepheids”) were due to radial pulsation. Also in the teens of the previous century, Henrietta Leavitt pointed out that the Cepheids in the Small Magellanic Clouds follow a period-luminosity relation, still one of the fundamental methods to determine distances in the visible universe – and one of the major astrophysical applications of pulsating stars.

With the ever increasing precision in photometric and radial velocity measurements, a large number of groups of pulsating stars are nowadays known.  [Figure 4-1](#) shows theoretical (in the sense that the logarithm of the effective temperature is plotted versus the logarithm of the stellar luminosity) HR diagrams containing the regions in which pulsating stars were known some 40



 **Fig. 4-1**

Theoretical HR diagram schematically showing locations of selected confirmed types of pulsating star. Parts of model evolutionary tracks for main sequence, horizontal branch, and post-AGB stars are shown as *dashed-dotted lines* for orientation. *Left*: known classes of pulsating star about 40 years ago. *Right*: classes of pulsator known to date. The names of the different groups are comprehensively listed in  [Table 4-1](#). Note the enormous increase in the number of classes of pulsating star and the improved knowledge about their loci in the recent past

■ Table 4-1

Selected classes of pulsating star

Name	Approx. periods	Discovery/definition
Mira variables	100–1,000 days	Fabrizius (1596)
Semiregular (SR) variables	20–2,000 days	Herschel (1782)
δ Cephei stars	1–100 days	1784, Pigott, Goodricke (1786)
RR Lyrae stars	0.3–3 days	Fleming (1899)
δ Scuti stars	0.3–6 h	Campbell and Wright (1900)
β Cephei stars	2–7 h	Frost (1902)
ZZ Ceti stars (DAV)	2–20 min	1964, Landolt (1968)
GW Virginis stars (DOV)	5–25 min	McGraw et al. (1979)
Rapidly oscillating Ap (roAp) stars	5–25 min	1978, Kurtz (1982)
V777 Herculis stars (DBV)	5–20 min	Winget et al. (1982)
Slowly Pulsating B (SPB) stars	0.5–3 days	Waelkens and Rufener (1985)
Solar-like oscillators	3–15 min	Kjeldsen et al. (1995)
V361 Hydrae stars (sdBvr)	2–10 min	1994, Kilkenney et al. (1997)
γ Doradus stars	0.3–1.5 days	1995, Kaye et al. (1999)
Solar-like giant oscillators	1–18 h	Frandsen et al. (2002)
V1093 Herculis stars (sdBVs)	1–2 h	Green et al. (2003)
Pulsating subdwarf O star (sdOV)	1–2 min	Woudt et al. (2006)

years ago and today. ➤ Table 4-1 gives a rough overview of the classes of pulsator contained in ➤ Fig. 4-1.

The different types of pulsator have historically been classified on a phenomenological basis. The separation between those types has usually later turned out to have a physical reason. The individual classes are different in terms of types of excited pulsation mode (or, less physical, pulsation period), mass and evolutionary state, hence temperature and luminosity. The names of these classes are assigned either after the name of a prototypical star or give some description of the type of variability and star.

It must be pointed out that the present overview does by far not contain all types and subgroups of pulsating star that have been suggested. The Cepheids are subdivided according to population and evolutionary state, into δ Cephei, W Vir, RV Tau, and BL Her stars. Jeffery (2008) proposed a number of types of evolved variable, there are the Luminous Blue Variables, and there may be new classes of white dwarf pulsator, oscillating red and brown dwarfs, etc. Furthermore, some of the instability domains of different pulsators overlap and indeed, some objects called “hybrid” pulsators that show oscillations of two distinct types, have been discovered. Also, the instability boundaries of some of these variables may need to be (considerably) extended and/or revised in the near future. For instance, there may be super-giant SPB stars, and solar-like oscillations are expected in all stars having a significant surface convection zone.

Whereas the classification of and distinction between the different classes of pulsating stars, which are historically grown and modified designations, can in some cases be called arbitrary today, one recognizes an important fact: pulsating stars populate almost the entire HR diagram, and this means that they can be used to learn something about the physics of most stars.

2 Astrophysical Background

2.1 Driving Mechanisms

What can make a star oscillate? After all, stars are in hydrostatic equilibrium: the gravitational pull on the mass elements of normal stars is balanced by gas pressure. If something would hit a star, the inward moving regions will be heated, and the increased heat loss damps the motion. Consequently, self-excited pulsations require a driving mechanism that overcomes this damping and results in a periodic oscillation.

Four major driving mechanisms have been proposed. The ϵ *mechanism* (Rosseland and Randers 1938) assumes a variation in the stellar nuclear reaction rate: if a nuclear burning region is compressed, the temperature rises and more energy is produced. This causes expansion, the pressure drops, and so does the energy generation: the motion is reversed and oscillations develop. The ϵ mechanism (where ϵ is the usual designator for the nuclear reaction rate in formulae), that operates similar to a Diesel engine, has been proposed for several different types of pulsating star, such as our Sun and pulsating white dwarfs, but observational proof for oscillations driven by it is still lacking.


Considerably more successful in explaining stellar oscillations is the κ - γ *mechanism* (Baker and Kippenhahn 1962 and references therein). In layers where the opacity κ increases and/or the third adiabatic exponent Γ_3 decreases with increasing temperature, flux coming from inner layers can be temporally stored. Such layers in the stellar interior are generally associated with regions where (partial) ionization of certain chemical elements occurs.

The energy accumulated in this layer during compression is additionally released when the layer tries to reach its equilibrium state by expanding. Therefore, the star can expand beyond its equilibrium radius. When the material recedes, energy is again stored in the stellar interior, and the whole cycle repeats: a periodic stellar oscillation evolves. This mechanism is also called the *Eddington Valve*, and it explains the variability of most of the known classes of pulsating star.

The classical pulsators in the instability strip, ranging from the δ Cephei stars to the RR Lyrae stars and the δ Scuti stars draw their pulsation power from the HeII ionization zone, whereas the oscillations of the roAp stars are believed to be excited in the H I and He I ionization zones, those of the Mira variables in the H I ionization zone, and those of the β Cephei and SPB stars are triggered in the ionization zone of the iron-group elements.

A very similar mechanism, in the sense that it is also due to a region in the star behaving like a valve, is convective blocking (or *convective driving*). In this scheme (e.g., Brickhill 1991), the base of a convection zone blocks the flux from the interior for some time, releasing the energy stored during compression in the subsequent expansion phase. The pulsations of white dwarf stars of spectral types DA and DB as well as γ Doradus stars are thought to be excited (at least partly) via this mechanism, which may also be of importance in Cepheids and Mira stars.

Finally, the pulsations of the Sun and solar-like stars, that are intrinsically stable and therefore not called self-excited, are *stochastically excited* due to turbulence in their surface convection zones. The vigorous convective motion in the outer surface layers generates acoustic noise in a broad frequency range, which excites solar-like oscillation modes. Due to the large number of convective elements on the surface, the excitation is of random nature, and the amplitudes of the oscillations are temporally highly variable.

Given the physical nature of these driving mechanisms, the existence of the different instability domains in the HR diagram (cf.  Fig. 4-1) easily follows. A star must fulfill certain physical conditions that it can pulsate, as the driving mechanism must be located in a specific

part of the star to give rise to observable oscillations. More physically speaking, in the case of self-excited pulsations the driving region must be located in a region where the thermal, and/or convective, time scale closely corresponds the dynamical (pulsational) time scale.

The consequence of the previous requirement is a constraint on the interior structure of a pulsating variable: if the instability region of some class of pulsating variable is accurately known, models of stars incorporating its excitation mechanism must be able to reproduce it. In this way, details of the input physics describing the interior structures of stars can be modified to reflect the observations.

However, this is not the only method available to study stellar structure and evolution from stellar pulsation. Models also need to explain the oscillation properties of individual stars. We are fortunate to be in the presence of stars having very complex pulsation patterns, multiperiodic radial and nonradial oscillators. The research field that determines the internal constitution of stars from their pulsations is called *asteroseismology*.

It is now due to make clear that the star whose interior structure is best known is the star closest to us, the Sun. Its surface can be resolved in two dimensions and millions of pulsation modes can be used for seismic analyses. The related research field is *helioseismology*, and it has been extensively reviewed elsewhere (e.g., Christensen-Dalsgaard 2002; Gizon et al. 2010). The present chapter will not touch upon helioseismology.

2.2 Asteroseismology

The basic idea of asteroseismology (see Gough 1996 for a discussion of why “astero”) is analogous to the determination of the Earth’s inner structure using earthquakes: these generate seismic waves that travel through rock and other interior structures of our planet, which can then be sounded. Today, the Earth’s interior has been completely mapped down to scales of a few hundred kilometers.

The strongest earthquakes can even cause *normal mode oscillations* (e.g., see Montagner and Roullet 2008), that is, the whole planet vibrates with its natural frequencies (also called *eigenfrequencies* in theoretical analyses). These normal modes are most valuable in determining the deep interior structure of the Earth. Asteroseismology does just the same: it uses the frequencies of the normal modes of pulsating stars (that may be seen as “starquakes”) as seismic waves. The eigenfrequencies of stellar models, that are dependent on their physical parameters and interior structures, are then matched to the observed ones. If a model succeeds in reproducing them, it is called a *seismic model*.

The pulsation modes are waves in the stellar interior, just like the waves that musical instruments resonate in. Popular articles often compare the frequency pattern of stellar oscillations to the sounds of musical instruments. The superposition of the normal mode frequencies in which a star oscillates can therefore be seen as its sound, although it generally does not involve frequencies the human ear is susceptible to.

There is a large variety of normal modes that stars can pulsate in. The simplest are *radial modes*: the star periodically expands and shrinks, and its spherical symmetry is preserved. The mathematical description of the displacement due to the oscillations results in differential equations of the Sturm–Liouville type that yields discrete eigensolutions: the radial mode frequencies of the given model.

Pulsation in *nonradial modes* causes deviations from spherical symmetry: the star changes its shape. Mathematically, this no longer results in an eigenvalue problem of the Sturm–Liouville type, and a large number of possible oscillation modes originate. The eigenfunctions are proportional to *spherical harmonics*:

$$Y_l^m(\theta, \phi) = N_l^m P_l^{|m|}(\cos \theta) e^{im\phi}, \quad (4.1)$$

where θ is the angle from the polar axis (colatitude), ϕ is the longitude, $P_l^{|m|}$ is the associated Legendre polynomial, N_l^m is a normalization constant and l and m are the spherical degree and azimuthal order of the oscillation. With

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x) \quad (4.2)$$


we obtain

$$P_0^0(\cos \theta) = 1 \quad (4.3)$$


$$P_1^0(\cos \theta) = \cos \theta \quad P_1^1(\cos \theta) = -\sin \theta \quad (4.4)$$

$$P_2^0(\cos \theta) = \frac{1}{2}(3 \cos^2 \theta - 1) \quad P_2^1(\cos \theta) = -3 \cos \theta \sin \theta \quad P_2^2(\cos \theta) = 3 \sin^2 \theta$$

etc. etc. (4.5)

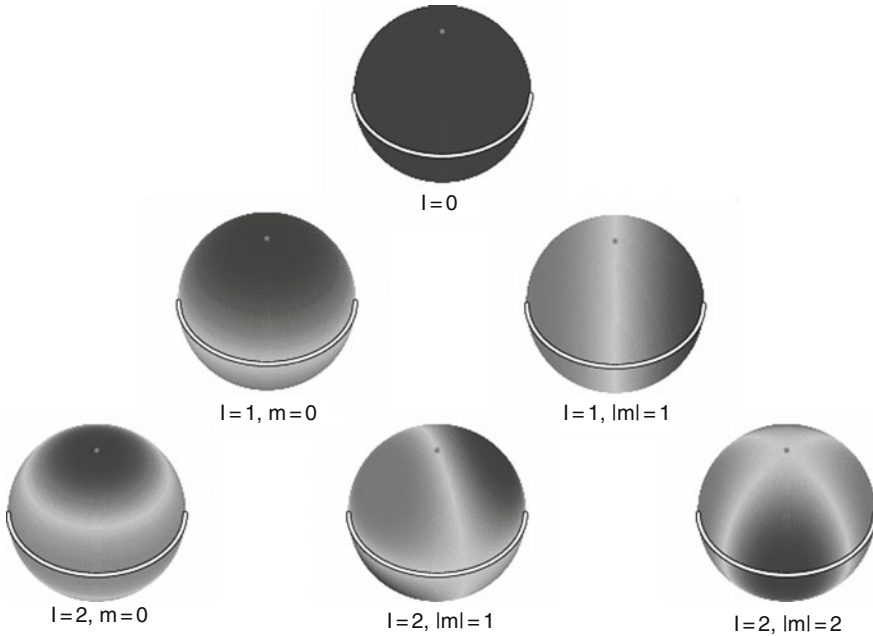
What does this mean in practice? Nonradial pulsation modes generate distortions on the stellar surface described by these spherical harmonics. The oscillations separate the stellar surface into expanding and receding as well as heating or cooling areas. A graphical example of these is shown in  Fig. 4-2.

Between the expanding and receding surface areas, no motion takes place. The lines along which this is the case are called the *node lines*; the nulls of the previously specified associated Legendre polynomials specify the locations of these node lines in polar direction. The total number of node lines on the stellar surface is the *spherical degree* l , which must in all cases be larger than or equal to zero. The number of node lines that are intersected when traveling around the stellar equator is the *azimuthal order* m . l and m are the quantities that appear in the expressions for the eigenfunctions and for the associated Legendre polynomial given above, and they are used for the classification of the pulsation modes.

Pulsation modes with $m \neq 0$ are traveling waves (as can also be seen in the defining  Equation 4.1 for spherical harmonics), and as they can run either with or against the rotation of the star, m can lie in the interval $[-l, l]$. Modes with $m = 0$ are called *axisymmetric*, modes with $|m| = l$ are named *sectoral*, and all other modes are referred to as *tesseral* modes.

The third quantity needed to describe pulsation modes is the *radial overtone* k (sometimes also denoted n in the literature), which is the number of nodes in the stellar interior. A mode that has no node in the interior is called a *fundamental mode*. A mode with one interior node is called the *first overtone*, modes with two interior nodes are the *second overtone*, etc. An accurate account of mode classification from the theoretical point of view is given by Deubner and Gough (1984).

Historically, observationally, and inconsistently with the theoretical definition, radial overtone modes have also been called the first and second harmonics, respectively, and have been



■ Fig. 4-2

Schematic description of the surface distortions produced by pulsation modes with spherical degrees $0 \leq l \leq 2$, arranged in the same way as the previous expressions for the associated Legendre polynomials. Whilst the outward moving areas of the star are colored in *dark grey*, the *light grey* areas move inward, and vice versa. The pole and equator of the star are indicated (Adapted from Telting and Schrijvers (1997))

abbreviated with F for the fundamental, as well as 1H (or 1O), 2H (or 2O), etc. for the overtones. Radial pulsations can be seen as modes with $\ell = 0$ (remember that $P_0^0(\cos \theta) = 1$); all other modes are nonradial oscillations. Modes with $l = 1$ are also called *dipole modes*; $l = 2$ modes are *quadrupole modes*.

There are two major restoring forces for stellar oscillations that attempt to bring the star back in its equilibrium configuration: pressure and buoyancy (gravity). For radial motion, the gravitational force in a star increases during compression, so it would actually accelerate, and not restore, the oscillation. Therefore, pressure must be the restoring force. On the other hand, for a predominantly transverse motion, gravity restores the motion through buoyancy, similar to what can be observed when throwing a stone into a pond. Therefore, aside from their identification with the pulsational quantum numbers k , l , and m , nonradial pulsation modes are also classified into *p pressure (p) modes* and *gravity (g) modes*. These two sets of modes thus differ by the main direction of their motion (radial/transverse), and their frequencies. Pulsation modes with periods longer than that of the radial fundamental mode are usually *g modes*, whereas *p modes* have periods equal or shorter than that; radial pulsations are always *p modes*.

The different modes are often labeled with their radial overtone number, for example, a p_3 mode is a pressure mode with three radial nodes, and a g_8 mode is a gravity mode with eight radial nodes. Modes with no interior nodes are fundamental modes, or *f modes*. Note that the

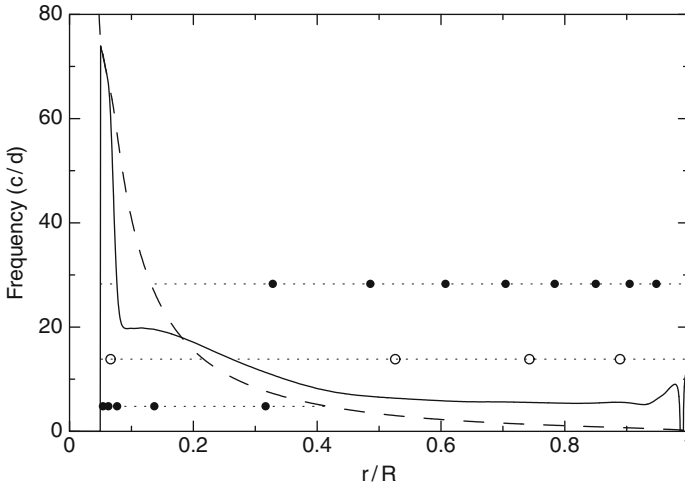
f mode for $l = 1$ does not exist, as a dipole motion of the entire star would require a movement of the stellar center of mass, which is physically impossible.

The propagation of pulsation modes in the stellar interior is governed by two frequencies. One of these is the *Lamb frequency* L_l , which is the inverse of the time needed to travel one horizontal wavelength at local sound speed. The other frequency describes at what rate a bubble of gas oscillates vertically around its equilibrium position at any given position inside a star; it is called the *Brunt-Vaisälä frequency* N . These two quantities are defined as:

$$L_l^2 = \frac{l(l+1)c^2}{r^2} \quad N^2 = g \left(\frac{1}{p_0 \gamma_1} \frac{dp_0}{dr} - \frac{1}{\rho_0} \frac{d\rho_0}{dr} \right), \quad (4.6)$$

where l is the spherical degree, c is the local velocity of sound, r is the radius, g is the local gravitational acceleration, p_0 and ρ_0 are local pressure and density in the unperturbed state, respectively, and $\gamma_1 = (\rho dp/pd\rho)_{\text{ad}}$ is the first adiabatic exponent.


The Lamb and Brunt-Vaisälä frequencies have the following implications: an oscillation with a frequency higher than both experiences pressure as the main restoring force in the corresponding part of the star. On the other hand, a vibration with a frequency lower than both L_l and N is restored mostly by buoyancy. In other words, if we have a stellar oscillation with an angular frequency ω , it is a p mode wherever $\omega > L_l, N$, and it is a g mode wherever $\omega < L_l, N$. In stellar interior regions where ω lies between the Lamb and Brunt-Vaisälä frequencies, the amplitude of the wave decreases exponentially with distance from the p and g mode propagation regions; such parts in the stellar interior are called evanescent regions. A *propagation diagram* aids the visualization of this discussion (► Fig. 4-3).



■ Fig. 4-3


A propagation diagram for an evolved $2 M_{\odot}$ stellar model. The run of the Lamb (*dashed line*) and Brunt-Vaisälä (*full line*) frequencies with respect to fractional stellar radius is shown. Some stellar pulsation modes are indicated with *thin horizontal dashed lines*, the *circles* are interior nodes. The lowest frequency oscillation (*lowest dashed line*) shown is a g_5 mode; that with the highest frequency is a p_8 mode. The oscillation with intermediate frequency is a mixed mode (Data kindly supplied by Patrick Lenz)

Whereas the Lamb frequency decreases monotonically toward the model's surface, the Brunt-Vaisälä frequency shows a sharp peak near the model's center and then rapidly drops to zero. This is because this model possesses a convective core, where $N^2 < 0$, and the spike in N is due to a region of chemical inhomogeneity. Over a range of stellar models, the behavior of L_1 in the interior is usually simple, whereas N may show considerable changes with evolutionary state and mass. However, N is always zero in the stellar center.


In  [Fig. 4-3](#), the g mode is confined to the innermost parts of the star. It is trapped in the interior, and therefore unlikely to be observed on the surface. The p mode is concentrated near the stellar surface and may be observable. The intermediate frequency mode shows remarkable behavior: it has three nodes in the outer regions of the star, but also one node in the g mode propagation region. This particular mode is capable of tunneling through its narrow evanescent region; it is a g mode in the deep interior, but a p mode closer to the surface. Such modes are called *mixed modes*.

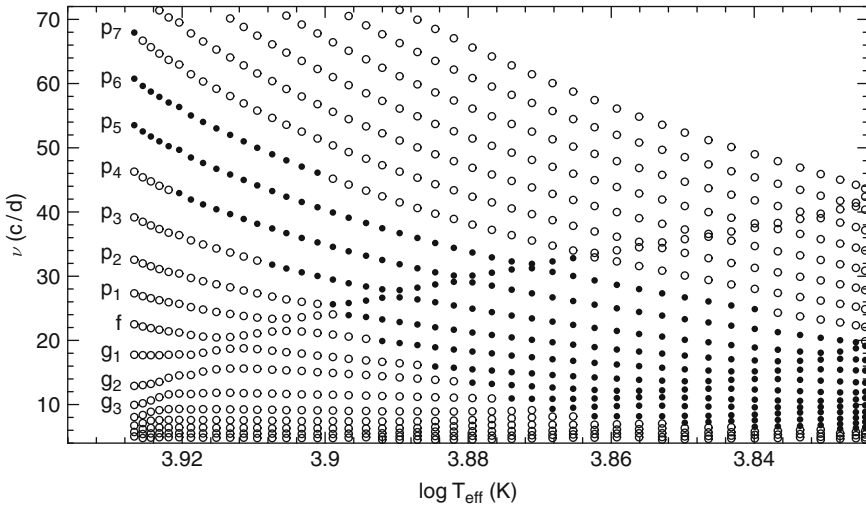
As mentioned before, stellar pulsation modes can be excited in certain parts of the stellar interior and they can propagate in some regions, whereas in other regions they are damped. The *work integral* is the energy gained by the pulsation mode averaged over one oscillation period. An evaluation of the work integral from the stellar center to the surface is used to determine whether or not a given mode is globally excited in a stellar model. For excitation to occur, the exciting forces must overcome those of damping and the work integral W will be positive.

The *growth rate* $\gamma = W/2\pi\omega I$, where ω is the pulsation frequency and I is the mode inertia, parameterizes the increase of oscillation energy during a pulsation cycle, and also indicates how rapidly the amplitude of a given mode increases. The normalized growth rate (or stability parameter) η , which is the ratio of the radius-integrated work available for excitation to the radius-integrated total work, is used to evaluate which pulsation modes are excited in the given model. If $\eta > 0$, a mode is driven and may reach observable amplitude; if $\eta = 1$, a mode is driven in the entire stellar model, and if $\eta = -1$, a mode is damped everywhere in the model. The most widely used application of this stability parameter is the comparison of the excited modal frequency ranges as determined from observations and those predicted by theory.

 [Figure 4-4](#) shows how the frequency domains predicted to be excited change with the evolution of a $1.8 M_{\odot}$ main sequence model. On the Zero-Age Main Sequence (ZAMS), only pure p modes of high radial overtone, and corresponding high frequencies, are excited in the model. Later on, some of these become stable again, but modes with lower overtone become excited. At the end of main sequence evolution, a large range of p modes, mixed modes, and g modes is predicted to be excited, and the frequency spectrum becomes dense (and even denser as the model leaves the main sequence). The range of excited frequencies is an observable that, when compared with models, can give an estimate of the evolutionary state of the star.

The reason for the evolutionary change in the pulsation frequencies is that as the model evolves, its convective core, in which the g modes are trapped at the ZAMS, shrinks and the g mode frequencies increase. At the same time the envelope expands due to the increased energy generation in the contracting nuclear burning core, causing a decrease of the p mode frequencies.

At some point, some p and g modes attain the same frequencies, and the modes begin to interact: they become mixed modes, with g mode character in the core and p mode character in the envelope (see also  [Fig. 4-3](#)). The frequencies of these modes never reach exactly the same value; the modes just exchange physical properties. This effect is called *avoided crossing* or *mode bumping* (Aizenman et al. 1977).



■ Fig. 4-4

Theoretically predicted $l = 2$ oscillation spectra of a $1.8 M_{\odot}$ main sequence model, evolving from hotter to cooler effective temperature. Pulsation modes excited in this model are shown with *filled circles*, stable modes with *open circles*. The types of mode on the Zero-Age Main Sequence (ZAMS) are given. Note the g modes intruding into the p mode domain

Of the individual modes, particular astrophysical potential is carried by the mode that originates as g_1 on the ZAMS, as pointed out by Dziembowski and Pamyatnykh (1991). This mode is mostly trapped in the shrinking convective core (for stars sufficiently massive to possess a convective core), because the rapid change in mean molecular weight at its edge causes a spike in the Brunt-Vaisälä frequency (cf. ► Fig. 4-3). The frequency of this mode is thus dependent on the size of the convective core.

The most important parameter determining the evolution of main sequence stars is the amount of nuclear fuel available. Stellar core convection can mix material from the radiative layer on top of the core into it, providing more nuclear fuel. This mixing (or penetration) is often named *convective core overshooting* and parameterized in theoretical models. Overshooting also decreases the gradient in the mean molecular weight at the edge of the core, which means that the frequency of the mode that is g_1 on the ZAMS measures the convective core size, a most important quantity for astrophysics in general. To emphasize its sensitivity to the extent of stellar core convection, this mode has also been named g_c mode.

In the following, only p and g modes, and mixed modes of these types will be considered. Other types of modes, such as r modes (torsional oscillations that may occur in rotating stars), g^- modes (convectively excited g modes in rotating stars), strange modes (showing up in calculations of highly nonadiabatic environments) or gravitational-wave w modes, will not be discussed as they have been of little practical importance for the seismic sounding of stars at the time of this writing.

The frequencies of the p and g modes of pulsating stars depend strongly on their structure. However, when high radial overtones are considered, some simple relations between mode

frequencies emerge. These are derived from *asymptotic theory*. The classical reference on the subject is Tassoul (1980), the instructive is Gough (1986). In the high-overtone limit one finds

$$\omega_{n,l} \simeq \Delta\omega_0 \left(n + \frac{l}{2} + \epsilon \right) - D_0 l(l+1) \quad (4.7)$$

for p modes, whereas for g modes

$$P_{n,l} \simeq \frac{P_0(n+\delta)}{\sqrt{l(l+1)}}, \quad (4.8)$$

where $\Delta\omega_0$ is the inverse sound travel time through the center of the star, D_0 is a frequency separation dependent on the stellar evolutionary state, P_0 is the asymptotic period that is proportional to the integral of the Brunt-Vaisälä frequency throughout the star, l and n are the spherical degree and radial overtone, respectively, and ϵ and δ are stellar structure parameters, respectively.

These relations have important consequences: low-degree p modes of consecutive high radial overtones of the same spherical degree are equally spaced in frequency, whereas low-degree g modes of consecutive high radial overtones of the same spherical degree are equally spaced in period. Furthermore, if the parameter D_0 were zero, (4.7) indicates that the frequencies of high-overtone p modes of even and odd degrees would be the same, respectively, and odd-degree modes would have frequencies intermediate between those of even degree. In realistic stellar models, these relations hold approximately, but not exactly.

In a nonrotating star, the frequencies of modes with $m \neq 0$ are the same as those of the $m = 0$ modes. However, as the $m \neq 0$ modes are traveling waves, their frequencies separate in the observer's frame when looking at a rotating star: the mode moving with rotation appears at higher frequency, the mode moving against rotation appears at lower frequency and the frequency difference to their nonrotating value is m times the rotation frequency (e.g., see Cox 1984). This effect is called *rotational frequency splitting*, and is one basic tool of asteroseismology: if such splittings are observed, the rotation frequency of the star can be determined.

Unfortunately, reality is not quite as simple as that. The Coriolis force acts on the traveling waves and modifies their frequencies. In addition, the $m \neq 0$ modes cause tidal bulges, on which centrifugal forces act. Therefore, the frequencies of these modes are often expressed as

$$\omega_{k,l,m} = \omega_{k,l,0} + m(1 - C_{k,l})\Omega + m^2 \frac{D_{k,l}\Omega^2}{\omega} \quad (4.9)$$

in case of (moderately) slow stellar rotation (Dziembowski and Goode 1992). $\omega_{k,l,m}$ is the observed frequency of the mode k, l, m , Ω is the stellar rotation frequency, and $C_{k,l}$ and $D_{k,l}$ are constants that describe the effects of the Coriolis and centrifugal forces, respectively; they are also called Ledoux constants. These constants are usually determined from stellar model calculations. The rotational splitting constant $C_{k,l}$ also approaches asymptotic values for high radial overtones. For g modes, $C_{k,l}$ becomes $(l(l+1))^{-1}$, and for p modes $C_{k,l} = 0$ in the asymptotic limit.

It should be made clear that the previous formula is only an approximation for the case that rotation can be treated as a perturbation to the equilibrium state: stellar rotation distorts the spherical shape of a star. As a consequence, the individual modes can no longer be described with single spherical harmonics. For instance, radial modes receive contamination of $l = 2$ and other modes with even l . Vice versa, $l = 2$ modes obtain some $l = 0, l = 4$, etc. contributions.

In the previous formula, this means that for rapid rotation higher order terms may be added, but to arrive at reliable results, two-dimensional numerical calculations are really required. *Rotational mode coupling* can affect the properties of oscillation modes with close frequencies (e.g., see Daszyńska-Daszkiewicz et al. 2002). Furthermore, the rotational distortion causes the stellar temperature to increase at the flattened poles and to decrease at the equatorial bulge. It follows that asteroseismology of rapidly rotating stars has additional degrees of complexity.

Even more complexity to the theoretical treatment of stellar pulsation is added by the presence of a magnetic field. A weak magnetic field would generate a second-order perturbation to the pulsation frequencies, just like the centrifugal force in case of rotation, but with opposite sign, that is, the observed oscillation frequencies would increase with respect to the non-magnetic case (e.g., Jones et al. 1989).

In the presence of a strong magnetic field, the effect would be more severe: if the field was oblique to the rotation axis (just like the Earth's magnetic field!), the pulsation axis would align with the magnetic axis, and no longer with the rotation axis, as implicitly assumed so far. This means that an observer would see each pulsation mode at a varying angle over the stellar rotation period, creating amplitude and phase variations with exactly that period, and therefore separable from the effects of rotational m-mode splitting. A star oscillating in this way is called *oblique pulsator* (Kurtz 1982). We note for completeness that the most general case of rotational splitting is the *nonaligned pulsator* (Pesnell 1985), where $(2l + 1)^2$ components of a given mode appear, with frequency separations proportional to the rotation period and to the angle between the rotation and pulsation axes.

To summarize, different stellar pulsation modes propagate in different interior regions, and their energy within those regions is not equally distributed. Each single pulsation mode has a different cavity, and its oscillation frequency is determined by the physical conditions in its cavity. This means that different modes are sensitive to the physical conditions in different parts of the stellar interior. Some modes teach us more about stellar envelopes, whereas other modes tell us about the deep interior. The more modes of different type are detected in a given star, the more complete our knowledge about its inner structure can become.

Some stars do us the favor to oscillate in many of these radial and/or nonradial modes simultaneously. Interior structure models of the stars can then be refined by measuring the oscillation frequencies of these stars, identifying them with their pulsational quantum numbers, and by reproducing all these with stellar models. This method is very sensitive because stellar oscillation frequencies can be measured to extremely high precision, up to one part in 10^9 (Kepler et al. 2005). In the following it will be described what observables are available to base asteroseismic models on, how the measurements can be interpreted, and how observations and theory are used to develop methods for asteroseismic interpretations.

3 From the Telescope to a Seismic Model

3.1 Basic Methods for Analyzing Asteroseismic Data

Because stellar oscillations generate motions and temperature variations on the surface, they result in observable variability. The interplay of these variations causes light, radial velocity, and line profile changes. Pulsating stars can thus be studied both photometrically and spectroscopically, via time series measurements.

These time series are subjected to *frequency analysis*, meaning that the constituent signals are extracted from the data. In many cases, this is done by harmonic analysis, transforming the time series into frequency/amplitude space, for example, by using the Discrete Fourier Transformation

$$F_N(f) = \sum_{k=1}^N x(t_k) e^{i2\pi f t_k}, \quad (4.10)$$

of the input function $x(t_k)$, corresponding to the time series of the measurements. Periodograms, amplitude, or power spectra are means of visualizing the results; an example is given in **Fig. 4-5**.

The amplitude spectrum in **Fig. 4-5** can be used to estimate the frequencies of the dominant signals in the time series. In many cases, the analysis is carried forward by fitting sinusoids to the data, determining and optimizing their frequencies, amplitudes and phases, often by least squares methods. It has also become common practice to subtract this optimized fit from the data and to compute periodograms of the residuals, a procedure called *prewhitening*. This process is repeated until no more significant signals can be found in the data (a most delicate decision!). Depending on the specific case and requirements of the data set, a large number of alternative frequency analysis methods can also be applied, such as phase-dispersion minimization, autocorrelation methods, wavelet analysis, etc. Care must be taken to keep the limitations of the data sets and of the applied methods in mind. Periods no longer than the data set itself can be reliably determined, and adjacent frequencies spaced no more than the inverse length of the data set can be resolved. Nonsinusoidal signals cause harmonics and combination frequencies in Fourier amplitude spectra that must not be mistaken for independent mode frequencies.

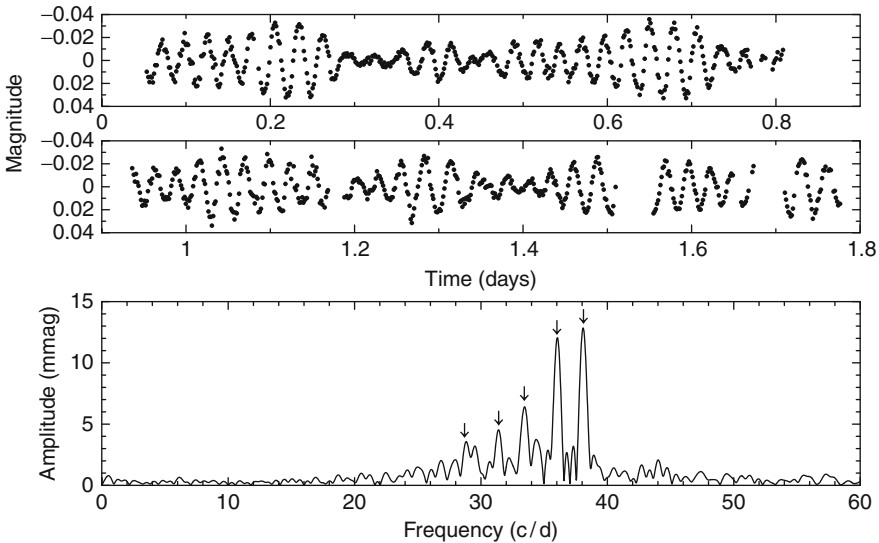


Fig. 4-5

The light variations (*upper two panels*) of a pulsating star and the corresponding Fourier amplitude spectrum. The complicated beating in the light curve is reflected by the presence of several signals in the periodogram; the strongest are labeled with *arrows*

No frequencies shorter than the inverse of twice the median distance between consecutive data points can be unambiguously detected; the highest frequency that can be retrieved in a given set of measurements is also called its *Nyquist frequency*.

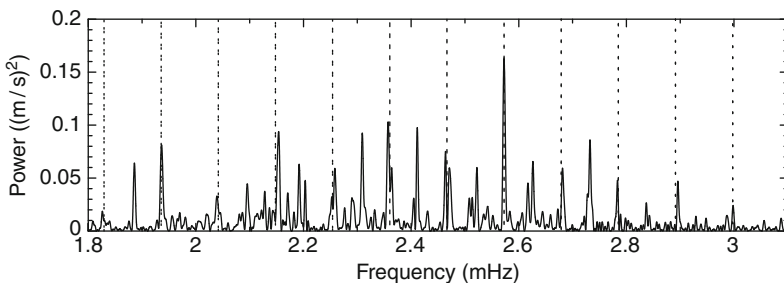
Having determined the frequencies characterizing the stellar variability, the next step is their interpretation. Measurements of distant stars have an important limitation: nonradial oscillations create patterns of brighter and fainter, approaching and receding areas on the stellar surface. However, as a distant observer can usually not resolve the stellar surface, she or he can only measure the joint effect of the pulsations in light and radial velocity. As a consequence, the effects of oscillations with high spherical degree average out in disk-integrated measurements, and their observed amplitudes are reduced with respect to the intrinsic value. This effect is called *geometric cancellation* (Dziembowski 1977). Calculations show that the amplitude drops as $\approx 1/\sqrt{l}$ for high l . In ground based observational studies it is mostly assumed (and often confirmed) that only modes with $l \leq 2$ are observed in light and radial velocity, with a few exceptions of l up to 4. Radial velocity measurements are somewhat more sensitive to higher l than observations in integrated light.

Once the oscillation frequencies of a given star have been determined, how can they be made asteroseismic use of? As an example, high-precision radial velocity measurements of the close-by star α Centauri A showed the presence of solar-like oscillations; the power spectrum (square of amplitude vs. frequency) of these data is shown in [Fig. 4-6](#).

This graph contains a series of maxima equally spaced in frequency: a high-overtone p mode spectrum, as predicted by asymptotic theory. The mean frequency spacing is $106.2 \mu\text{Hz}$. However, it is obvious that only every other of the strongest peaks conforms to this spacing; there are other signals in between. The signals halfway between the vertical lines in [Fig. 4-6](#), denoting modes of spherical degree l , are a consequence of ([4.7](#)): these are modes with $l \pm 1$. Given the effects of geometrical cancellation, it is straightforward to assume that these are the modes of lowest l , viz. $l = 0$ and 1 .

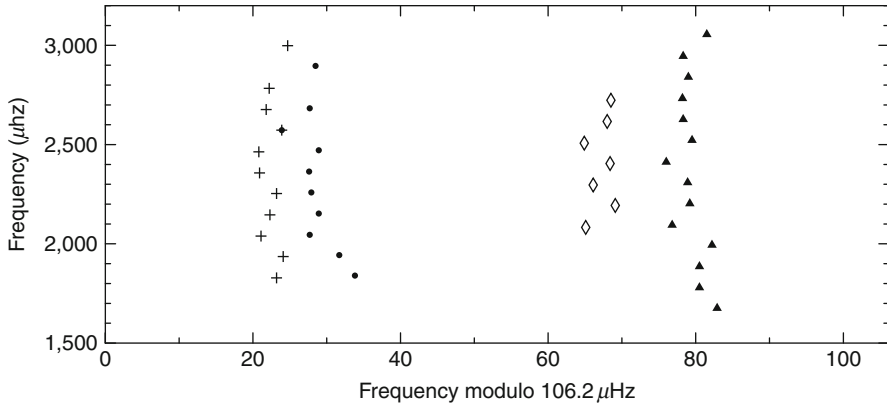
However, there is more information present. A close look at [Fig. 4-6](#) reveals that many of the strongest peaks seem to be split. Again, the explanation for this finding lies within ([4.7](#)): the close neighbors are modes with $l + 2$. These frequencies no longer have the same values as those with degree l because the stellar interior has structure.

As stars evolve on the main sequence, their nuclear burning cores shrink, increase in density, and change in chemical composition as evolution progresses. This alters the acoustic sound



■ Fig. 4-6

High-overtone p mode spectrum of the solar-like oscillator α Centauri A (Bedding et al. 2004). The vertical dotted lines are separated by $106.2 \mu\text{Hz}$



■ Fig. 4-7

Echelle diagram of the frequencies determined from ► Fig. 4-6 (values from Bedding et al. 2004). Full dots represent radial modes, triangles mark dipole modes, plus signs stand for quadrupole modes and diamonds are for $l = 3$ modes

speed in the core and is reflected in the frequency differences between modes of degrees l , $l + 2$, also called the *small frequency separation* $\delta\nu_{02}$. On the other hand, evolution causes expansion of the outer regions of stars that become more tenuous, which increases the sound travel time through the star. This is measurable via the frequency difference between consecutive radial overtones and is called the *large frequency separation* $\Delta\nu_0$.

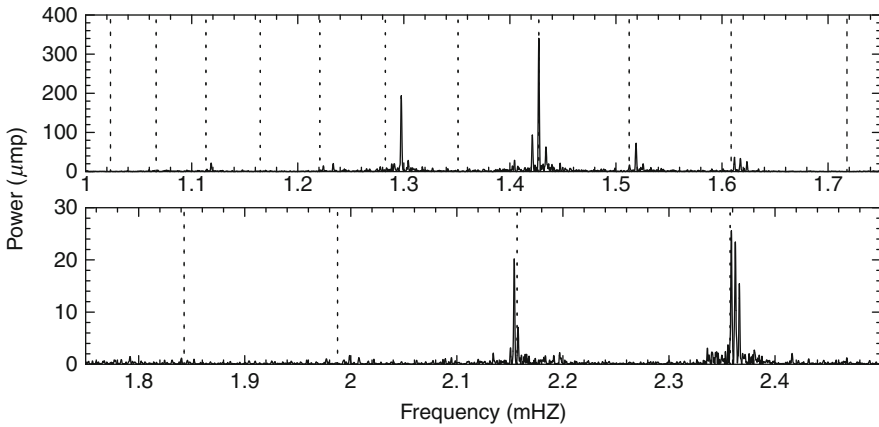
The large and the small separations can be computed for a range of theoretical stellar models. It turns out that a plot of $\delta\nu_0$ vs. $\Delta\nu_0$ allows an unambiguous determination of stellar mass and evolutionary state, and this method works particularly well for models with parameters similar to our Sun (main sequence, $M \lesssim 1.3 M_\odot$). This diagnostic is called an *asteroseismic HR diagram* (Christensen-Dalsgaard 1988).

Another important tool to analyze high-overtone p mode pulsation spectra is the *Echelle diagram*. This diagram plots the oscillation frequencies versus their modulus with respect to the large separation. ► Figure 4-7 shows the Echelle diagram for α Cen A constructed from the frequencies of the signals apparent in ► Fig. 4-6.

The frequencies fall onto four distinct ridges, corresponding to modes of the same spherical degree l . In this way, l can be identified. From left to right, the ridges correspond to $l = 2, 0, 3$ and 1, respectively, and it can be seen that, within the errors, $\delta\nu_{13} = 5/3\delta\nu_{02}$, consistent with (► 4.7). The scatter in this diagram is due to a combination of the temporal resolution of the data and, more importantly, the finite lifetimes of the stochastically excited modes.

The example of α Cen A shows that once a sufficient number of intrinsic pulsation frequencies of a given star has been determined, one may identify their spherical degrees by *pattern recognition*. This method is also applicable to the oscillation spectra of pulsating white dwarf stars, as exemplified in ► Fig. 4-8.

The series of peaks discernible in ► Fig. 4-8 does not form a pattern of equally spaced frequencies, but of equally spaced periods. This is the signature of a high-overtone g mode pulsation spectrum (► 4.8). In this case, all the strong peaks are roughly aligned with the vertical lines; there are no strong modes in between. Some of the expected signals are missing, but in



■ Fig. 4-8

High-overtone g mode spectrum of the pulsating DB white dwarf star GD 358 (Winget et al. 1994). The vertical dotted lines denote periods with an equal spacing of 39.5 s

most cases only apparently: they are just much weaker in amplitude. A mean period spacing of 39.5 s has been determined from this analysis, and the obvious identification of the strongest modes is $l = 1$. Equation 4.8 then results in $P_0 = 55.9$ s. For pulsating white dwarf stars, P_0 is a measure of the stellar mass, which was consequently determined.

Furthermore, several of the strongest peaks in Fig. 4-8 are split into triplets. These triplets are equally spaced in frequency, and they are the signature of rotationally split m modes (Equation 4.9). The rotation period of GD 358 could therefore also be determined, just from an inspection of the power spectrum of its photometric time series, and basic application of theory. The fact that the m -mode splitting only results in triplets strengthens the previous identification of the spherical degree as $l = 1$. The relative amplitudes within those triplets are dependent on the inclination of the stellar pulsation axis (Equation 4.4), which may then be determined, but in reality there are other, presently unknown, effects that modify the relative multiplet amplitudes.

There is even more information present. It is noticeable that the peaks in the amplitude spectrum do not perfectly conform to their asymptotically predicted locations. This is a sign of *mode trapping*. White dwarf stars consist of a degenerate core, with subsequent outer layers of different chemical elements. The transition regions between these layers create spikes in the Brunt-Vaisälä frequency. The pulsation modes prefer to place their radial nodes in these transition regions, and be standing waves to both sides of the nodes. This modifies their frequencies compared to the case of homogeneous interior structure and gives rise to the observed deviations from equal period spacing.

The observant reader will have noticed three apparent inconsistencies in the previous paragraphs. How can two stars with similar oscillation frequencies be high-overtone p and g mode pulsators, respectively? How does one know that 106.2 μHz is the mean frequency spacing for α Cen A, and not half this value? Why can the claimed small frequency separation not be the effect of rotation?

This is because some additional constraints are available that help in the interpretation of pulsational mode spectra. As the p mode pulsation periods depend on the sound travel time

through the star, they must be related to its size, or more precisely, to its mean density. The *pulsation constant*

$$Q = P \sqrt{\frac{\bar{\rho}_{\odot}}{\bar{\rho}}}, \quad (4.11)$$

where P is the pulsation period and $\bar{\rho}$ is the mean stellar density, is a useful indicator of what type of mode one sees in a given star. Over the whole HR diagram, the Q value for the radial fundamental mode is between 0.03 and 0.04 days; for the Sun, it is 0.033 days (0.8 h). As mentioned before, radial modes can only be p modes, (pure) p modes always have frequencies of the same order or larger than that of the radial fundamental, and (pure) g modes always have frequencies lower than that.

Because GD 358 is a white dwarf star and therefore has high mean density, its radial fundamental mode period would be of the order of 4 s. On the other hand, α Cen A, a little more massive and more evolved than our Sun, has a radial fundamental mode period of about 1 h. Therefore, the similar 5–10 min pulsation periods of the two stars correspond to completely different types of mode.

The asymptotic p mode frequency separation also relates to the stellar mean density

$$\Delta\nu_0 = \Delta\nu_{0,\odot} \sqrt{\frac{\rho_{\odot}}{\rho}}, \quad (4.12)$$

where $\Delta\nu_{0,\odot} = 135 \mu\text{Hz}$. Given the knowledge on the mass and evolutionary state of α Cen A, it immediately be inferred that $106.2 \mu\text{Hz}$ must be the large frequency spacing. Finally, the rotation period of α Cen A is much too long to generate m -mode splitting with the observed $\delta\nu_{02}$.

Apart from these examples, there is another clue toward the nature of observed pulsation modes just from observed frequency spectra. Radial modes are global oscillations that may reveal themselves because the period ratios of consecutive overtones are well known. On the main sequence, the period ratio of the radial fundamental mode and the first overtone is around 0.773, and the period ratio of the first to second overtone radial modes is 0.810. For more evolved stars, such as the δ Cephei stars, these ratios change to 0.705 and 0.805, respectively. Such period ratios are therefore suggestive to correspond to radial modes.

3.2 Methods for Mode Identification

It is now clear that besides the pulsational mode spectra of stars themselves, the incorporation of other constraints is useful for their interpretation. This is particularly important for oscillation spectra that do not show obvious imprints of the underlying modes, like those of stars pulsating in high radial overtones. Examples are pulsations of low radial overtone or stars rotating so rapidly that the rotational splitting is of the same order as the frequency spacing of consecutive radial overtones of the same l or of modes of the same k , but different l . In practice, most observed stellar oscillation spectra are incomplete, either because the star chooses so or because the observations do not have sufficient sensitivity, which makes it even more difficult to recognize patterns in the observed frequencies or periods and to type the modes accordingly. Therefore, *mode identification methods* have been developed, of which a variety is available.

The first method uses photometric data only. The flux change for nonradial pulsation in the linear regime can be expressed as (Watson 1988)

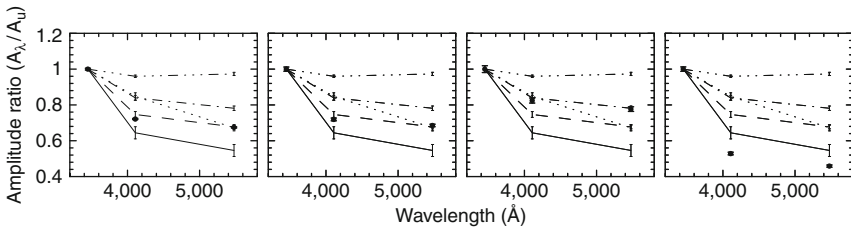
$$\Delta m(\lambda, t) = -1.086\epsilon P_l^{[m]}(\mu_0)((T_1 + T_2) \cos(\omega t + \psi_T) + (T_3 + T_4 + T_5) \cos(\omega t)), \quad (4.13)$$

where $\Delta m(\lambda, t)$ is the time and wavelength dependent magnitude variation of an oscillation, -1.086ϵ is an amplitude parameter transformed from fluxes to magnitudes, $P_l^{[m]}$ is the associated Legendre polynomial, μ_0 is the cosine of the inclination of the stellar pulsation axis with respect to the observer, ω is the angular pulsation frequency, t is time, and ψ_T is the phase lag between the changes in temperature and local geometry (mostly originating in convection zones near the stellar surface). The term T_1 is the local temperature change on the surface, T_2 is the temperature-dependent limb darkening variation, T_3 is the local geometry change on the stellar surface, T_4 is the local surface pressure change, and T_5 is the gravity-dependent limb darkening variation.

The T_i terms can be determined for different types of pulsator from theoretical model atmospheres, and the observables best suited to reveal the types of mode present can also be deduced. These would for instance be the photometric amplitude ratios or phase shifts between different filter passbands, but also the optimal passbands themselves. An example of such a mode identification is shown in [Fig. 4-9](#).

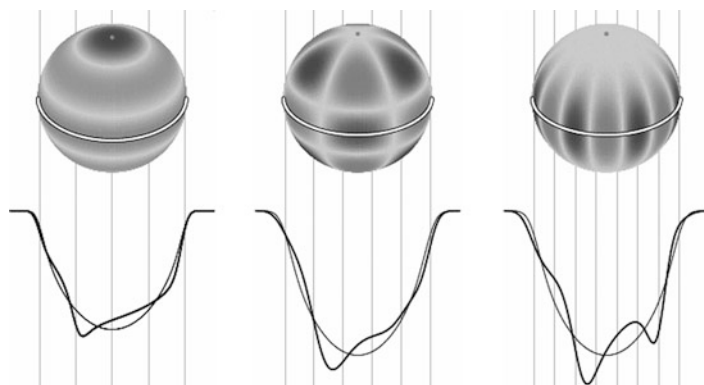
From photometry alone, only the spherical degree of a given pulsation mode can be identified. The method can be supported by adding radial velocity measurements, which increases its sensitivity (Daszyńska-Daszkiewicz et al. 2005), but still does not supply a determination of the azimuthal order. To this end, high-resolution spectroscopy must be invoked.

The lines in a stellar spectrum are broadened by rotation through the Doppler effect: the intrinsic line profile is blueshifted on the parts of the stellar surface approaching the observer, and redshifted on the areas moving away. The effect is strongest on the stellar limb and decreases towards the center. As a consequence, a rotationally broadened line profile contains spatial information of the stellar surface. For instance, in a dark starspot flux is missing, and a spike will move through the line profile as the spot rotates over the visible disk. The method to reconstruct stellar surface structure from line profile variations is called *Doppler Imaging* (Vogt and Penrod 1983).



■ Fig. 4-9

Identification of the four strongest pulsation modes of the β Cephei star 12 Lacertae from multi-color photometry (taken from Handler et al. 2006). The amplitudes are normalized to unity in the ultraviolet and compared with theoretical predictions. The *full lines* are for $l = 0$, the *dashed lines* for $l = 1$, the *dashed-dotted lines* for $l = 2$, the *dotted lines* for $l = 3$, and the *dashed-triple dotted lines* for $l = 4$. The modes investigated in the *two left panels* are $l = 1$, the next is $l = 2$, and the rightmost one is $l = 0$



■ Fig. 4-10

Line profile variations due to stellar pulsation. The *upper* parts of the graph show the shape of the oscillation mode on the surface, whereas the *thin lines* in the lower halves represent the unperturbed rotationally broadened line profile, and the *thick lines* are the superpositions with the pulsation. Each individual mode (from *left to right*: $l = 4$, $m = 0$; $l = 5$, $|m| = 3$; $l = |m| = 7$) generates a different distortion of the line profile (Adapted from Telting and Schrijvers (1997))

Apart from radial velocity changes, stellar oscillations also cause variations in rotationally broadened line profiles. The areas on the stellar surface that have an additional approaching component to their motion with respect to the observer have their contribution to the line profile blueshifted, whereas the receding parts are redshifted by the corresponding amount of pulsational Doppler velocity. The net result of all these motions is bumps traveling through the line profile, and their shapes are governed by the oscillation mode causing them (e.g., see Telting 2003 for a review). Examining stellar line profiles, pulsation modes of up to $l \approx 20$ can be observed and identified, a vast extension in l compared to photometric and radial velocity techniques. Some examples of pulsational line profile variations are shown in ► Fig. 4-10.

The task now is to extract the correct values of l and m from the observed variations. The principle is to fit the theoretically calculated 3-D velocity field to the observed line profiles, and a wide range of spectroscopic mode identification methods is available. Some of the most commonly used are the Moment method (most suitable for low l), the pixel-by-pixel method, the Fourier Parameter Fit method, or Doppler reconstruction of the surface distortions. High-resolution spectroscopy is better suited for the determination of m rather than l which makes it complementary to photometric and radial velocity methods.

When resorting to photometric or spectroscopic methods, it is not required to arrive at unique identifications of all observed pulsation modes in each given star. What is needed is the secure identification of a sufficient number of modes to rule out all possible alternative interpretations. An example will be presented later.

3.3 Asteroseismic Modeling

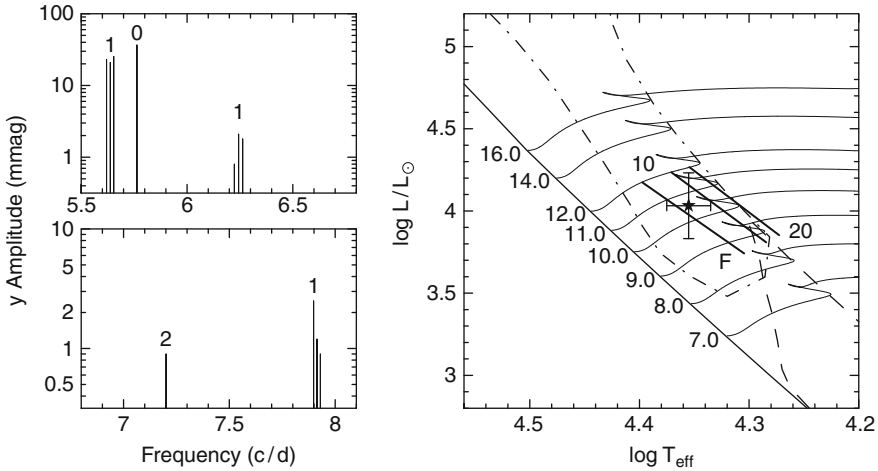
The observed pulsation frequencies and their identifications are then matched to theoretical models (see Kawaler, this volume, for details). These would ideally be full evolutionary models with pulsation codes operating on them, although a few codes based on envelope models are still in use.

Most pulsation codes use the *linear approximation*, which means that the oscillations are treated as linear perturbations around the equilibrium state. This allows the evaluation of the excitation of oscillation modes and thus the computation of theoretical domains of pulsation in the HR diagram. Nonlinear computations, that would allow predictions of oscillation amplitudes, are still rather the exception than the rule because they are, even today, expensive in terms of computing time, as are numerical hydrodynamical simulations.

Many stellar pulsation codes employ the *adiabatic approximation* (sometimes called isentropic approximation) to compute oscillation frequencies, which is the assumption that no energy exchange between the oscillating element and its surroundings takes place. Other codes perform nonadiabatic frequency computations. A wide variety of stellar oscillation codes is available, and most theory groups use their own routines, optimized for the application to their objects of main interest.

There are several strategies to find seismic models from observations. Some compare the observed oscillation frequencies with those of a grid of stellar models and perform automatic matching between them. This is computationally extensive, which means that supercomputers or parallel processing are invoked, or that intelligent optimization methods such as genetic algorithms are employed, or both.

Other strategies start with first imposing observational constraints. Besides the observed oscillation frequencies themselves and the identification of the underlying pulsation modes, these would often be estimates of the objects' positions in the HR diagram. As an example, these are shown for the β Cephei star ν Eridani in [Fig. 4-11](#).

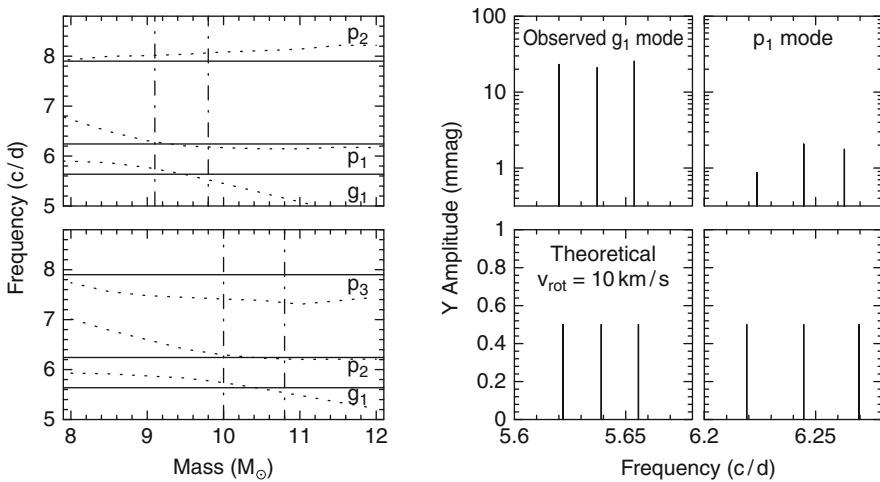


■ Fig. 4-11

Left: schematic oscillation spectrum of ν Eridani. The numbers on top of each mode (group) are their l identifications, consistent in photometry and spectroscopy (De Ridder et al. 2004). **Right:** a plot of the star's position in the theoretical HR Diagram (*star symbol*) with its error bars and lines of equal mean density for the observed radial mode periods assuming them to be the fundamental, first, and second overtones, respectively (*thick lines*). Some model evolutionary tracks labeled with their masses, the Zero-Age Main Sequence, and the borders of the β Cephei (*dashed-dotted line*) and SPB star (*dashed lines*) instability strips are also shown

The detection of a radial mode in the frequency spectrum is an asset for asteroseismic studies of this star: there are only three possibilities for the value of its mean density. These depend on whether its frequency corresponds to the fundamental, the first or the second radial overtone (► 4.11). A comparison of the observed position of the star in the HR diagram (right panel of ► Fig. 4-11) with its error bars leads to the rejection of the second overtone hypothesis, and to the elimination of models with masses below $8.5 M_{\odot}$.

Now the $l = 1$ modes come into play: moving along the lines of constant mean density in the HR diagram, a comparison between their observed and theoretically predicted frequencies can be made. This is done on the left side in ► Fig. 4-12. Given the uncertainties and assumptions in the asteroseismic model construction, all $l = 1$ modes are reproduced by models between 9.1 and $9.8 M_{\odot}$ that have the radial mode as the fundamental. However, the frequency of the highest-overtone dipole mode cannot be explained within the errors for the models that assume the first overtone for the radial mode. This interpretation can therefore also be rejected. The modal assignment to the observed frequencies is now unambiguous, and the range of models to be explored for seismic fitting is severely reduced.



■ Fig. 4-12

Left: a figure showing the match of the $l = 1, m = 0$ modes of ν Eri for models of different mass but same mean density. The *full horizontal lines* are the observed frequencies, the *dotted lines* are theoretical model frequencies. *Upper panel:* assuming that the radial mode is the fundamental. The *vertical dashed-dotted lines* show the mass range in which the first g mode and the first p mode fit the observed frequencies (9.1 – $9.8 M_{\odot}$). *Lower panel:* the same, but assuming the radial mode is the first overtone. Here it would be the first g mode and the second p mode that give an acceptable fit between 10.0 and $10.8 M_{\odot}$. However, the observed $l = 1$ mode with highest frequency is not compatible with the more massive models. **Right:** a comparison of the rotational splittings of a rigidly rotating model, with a rotation rate chosen to fit the observed $l = 1, g_1$ triplet, with the observations. The observed splitting of the $l = 1, p_1$ triplet is not reproduced with this assumption, demonstrating the presence of differential interior rotation

The reason why this way of mode identification was successful is that the g_1 mode and its respective p mode neighbor are in the process of avoided crossing in the parameter space of interest. Small changes in the evolutionary states of the models depending on mass therefore lead to significant changes in their frequencies. The observed $l = 2$ mode has been excluded from the mode identification procedure as all models in the parameter domain under consideration reproduce its frequency correctly: it is nearly a pure p mode and its frequency hardly changes in models of the same mean density.

In this restricted parameter space, an interesting observation can be made: the observed $l = 1$ m -mode splittings do not agree with those predicted by uniformly rotating models (► Fig. 4-12, right-hand side). Fitting the rotational splitting for the g_1 mode, that samples the deep stellar interior, results in predicted splitting about 30% larger than that observed for the p_1 mode, which is concentrated closer to the surface. This means that the star's rotation rate increases toward its interior, as predicted by theory (Talon et al. 1997). In addition, as the $l = 1$, g_1 mode frequency is sensitive to the size of the convective core (it is the g_c mode!), a constraint on the convective core overshooting parameter was also obtained (Pamyatnykh et al. 2004).

This example shows the potential of asteroseismology, even when only few pulsation modes are available. Radial pulsation modes are quite valuable in the identification process as they immediately result in accurate constraints on the stellar mean density, that can become unambiguous when the radial overtone of the mode can be inferred, supported by other estimates such as on effective temperature and luminosity. Theory and observations work hand in hand. Once a unique identification of the normal modes is achieved, some firm results on stellar structure can be obtained, even given uncertainties in the modeling procedure.

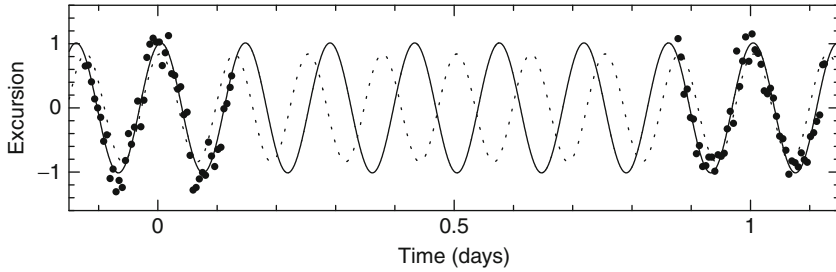
In the case of ν Eridani, the most severe problem is that the $l = 1$, p_2 mode is hard to be excited and accurately matched in frequency given the observed effective temperature of the star and its chemical surface composition. Only changes to the input physics would solve this problem. This is just the goal of asteroseismology: to improve our knowledge about stellar physics. Such an improvement can only be achieved using observational results that present models cannot account for!

4 Applications

4.1 Pulsating White and Pre-white Dwarf Stars

The first pulsators that could be studied asteroseismically were white dwarf and pre-white dwarf stars, pulsating in high-overtone low-degree g modes. A remarkable initial result was the theoretical prediction of a then new class of pulsating white dwarfs of spectral type DB, and their subsequent observational confirmation (Winget et al. 1982). Upon the realization that some pulsating white dwarf stars have very complicated frequency spectra, and having a well-developed theory of white dwarf pulsation available, the main obstacle to extracting correct mode frequencies from observational data was the daily interruptions of the measurements by sunlight. ► Figure 4-13 explains why.

The interruptions cause ambiguities in the determination of the frequencies of the signals present in the data: a frequency different by one cycle per sidereal day from the real oscillation frequency would fit the data to comparable precision. In the presence of complicated variability



■ Fig. 4-13

Simulation of observations from a single astronomical site. The *filled circles* represent the measurements, including noise. The *full line* is a fit with the correct frequency present in the data, and the *dotted line* is a fit with a frequency different by one cycle per day. The two fits represent the data almost equally well. However, they would be completely out of phase, and the correct one immediately obvious, if measurements were available in between the present data, from a site 180° different in geographical longitude. This led to the setup of multisite telescope networks

and most notably for signals of low signal-to-noise this could lead to erroneous frequency determinations. Any seismic model based on incorrect observational input is misleading.

The solution to this problem is to avoid, or at least minimize, daytime gaps in time resolved measurements. This can be accomplished by concerted observational efforts, involving interested colleagues over the whole globe, passing on the asteroseismic target from one observatory to the next. The best known of these collaborations is the Whole Earth Telescope (WET, Nather et al. 1990), invented for the study of pulsating white dwarf stars.

One of the first WET runs was devoted to the prototypical pulsating pre-white dwarf star PG 1159-035 = GW Vir. It resulted in the detection of over 100 g mode frequencies of $l = 1$ and $l = 2$ modes of high radial overtone, leading to precise determinations of the stellar mass and rotation period, an asteroseismic detection of compositional stratification, and an upper limit to the magnetic field strength (Winget et al. 1991). Subsequent WET observations of the prototypical pulsating DB white dwarf star GD 358 = V777 Her showed a mode spectrum dominated by high-overtone $l = 1$ g modes (cf. ► Fig. 4-8), resulting in determinations of its total and surface Helium layer mass, luminosity, and rotation rate (Winget et al. 1994).

The evolution of white dwarf stars is dominated by cooling, at (nearly) constant radius. As they cool, they pass through a number of instability strips. Evidence is that all white dwarf stars located in such instability domains in the HR diagram do pulsate. This has an important consequence: the interior structures of the pulsators must be representative of all white dwarf stars. Thus asteroseismic results for white dwarf pulsators can be extended to the stars that do not oscillate.

Cooling of pulsating white dwarf stars changes their oscillation periods, and the rate of period change is directly related to their energy loss. The hottest DB pulsators have reasonably high neutrino emission rates, and their evolutionary period changes may be able to tell us whether their neutrino emission is compatible with the standard model of particle physics (Winget et al. 2004). Measurements to detect such a period change are ongoing. On the other hand, the period changes of DA pulsators (some of which are the most stable optical clocks in the Universe, Kepler et al. 2005) could reveal the masses of axions, if the latter existed (Kim et al. 2006).

Asteroseismology of white dwarf pulsators does not only allow to detect stratification in their chemical profiles near the surface, it also gives evidence of their core composition. This, in turn, is dependent on their history of evolution on the Asymptotic Giant Branch (AGB) and can be used to obtain constraints on the nuclear reaction rates in AGB stars. Present results (Metcalf 2005) indicate consistency with evolutionary models.

As white dwarf stars cool, their cores become crystallized. Being composed mainly of carbon and oxygen, such cores can be seen as giant diamonds! Massive white dwarf stars begin to crystallize when still in the DAV (ZZ Ceti) instability strip, and asteroseismic investigations of one such massive pulsator have proven substantial crystallization in its interior (Brassard and Fontaine 2005; Metcalfe et al. 2004).


The light curve shapes of pulsating white dwarf stars of spectral types DB and DA are often nonsinusoidal. The nonlinearities originate in their convection zones, that cannot instantly adjust to the pulsational variations (the g mode pulsations of pulsating white dwarf stars are almost exclusively due to temperature changes; Robinson et al. 1982). As the light curve shapes of such pulsators depend on the thermal response time of the convection zone, the latter parameter can be determined from nonlinear light curve fits (Montgomery 2005).

As a final example, there are pulsating white dwarf stars in mass accreting close binary systems. If the mass transfer rate is in a certain range, the surface temperature of the accreting white dwarf places it in an instability strip. A handful of such oscillators is known to date (Mukadam et al. 2007), but attempts at asteroseismology have proven difficult due to low amplitudes and unstable mode spectra.

Among all classes of pulsating star (aside from the Sun itself), asteroseismology of pulsating white dwarf stars is certainly in the most advanced state. A new class of such oscillators has recently been proposed, hot DQ stars with carbon-dominated atmospheres and temperatures similar to that of the DB pulsators (Dufour et al. 2009; Montgomery et al. 2007). There is little doubt remaining that the variability of these stars is due to pulsation. Theory predicts yet another new type of white dwarf oscillator, DA white dwarf stars somewhat hotter than the DB pulsators, but observational searches for them have so far been inconclusive (Kurtz et al. 2008). We refer to Montgomery (2009), and references therein, for more information on asteroseismology of pulsating white dwarf stars.

4.2 Delta Scuti Stars

At about the same time when the necessity for worldwide observing efforts for pulsating white dwarf stars was realized, the same conclusion was reached for δ Scuti stars, some of which also exhibit complex oscillation spectra, but in this case composed of g, p, and mixed modes of low overtone. The Delta Scuti Network, founded over 25 years ago, was the first multisite observing collaboration for these stars, followed by a number of others such as STEPHI or STACC; also a few WET runs were devoted to δ Scuti stars.

The asteroseismic potential of δ Scuti stars is enormous, but could so far not be fully exploited. Part of the reason is visible in  Fig. 4-4: the pure p mode spectrum on the ZAMS mainly allows a determination of the stellar mean density. When the scientifically more interesting mixed modes appear, the frequency spectrum is fairly dense and requires a large amount of data and long time base to be resolved observationally. However, the real stars seldom cooperate in showing many of the potentially excited modes at observable amplitude, inhibiting mode

identification by pattern recognition. In addition, mode amplitudes are often small, counter-acting reliable identifications of many modes by applying the methods discussed before. Even though dozens or even hundreds of pulsation modes have been detected in some δ Scuti stars, little has been learnt on their interior structures from asteroseismology so far.

Pulsational amplitude limitation of δ Scuti stars is a major problem for theory: what makes the star excite only certain modes, which modes would these be and what determines their amplitudes?

For slowly rotating δ Scuti stars gravitational settling and radiative levitation give rise to chemical surface peculiarities and are believed to deplete the pulsational driving regions. Consequently, Am and Ap stars are not expected to pulsate, although a few of them do (Kurtz 2000). This also means that many δ Scuti stars rotate rapidly, which requires special calculations to extract information from their distorted pulsation modes, a field that has made considerable progress in the recent past (Reese et al. 2009).

Some pre-main sequence stars cross the δ Scuti instability strip on their way to the ZAMS, and consequently pulsate. The interior structures of these stars are fairly simple, which may make them more accessible to asteroseismic investigation compared to their main sequence counterparts. The oscillation spectra of pre-main sequence and main sequence δ Scuti stars in the same position in the HR diagram are predicted to be different, which may allow an asteroseismic separation (Suran et al. 2001).

4.3 Slowly Pulsating B and Gamma Doradus Stars

These two classes of high-overtone g mode pulsator, although well separated in effective temperature, share most of their asteroseismic characteristics. They also share the problems with respect to observations, caused by their long periods: resolving their oscillation spectra requires measurements over a long time baseline, possibly many years. It is therefore no surprise that most of the known members of these two groups of pulsator were discovered with the HIPPARCOS (High Precision PARallax COLlecting Satellite) mission from its data set spanning over 3 years (Handler 1999; Waelkens et al. 1998).

As in pulsating white dwarf stars, effects of inhomogeneities in stellar structure would manifest themselves in mode trapping and thus in oscillatory behavior in the g mode periods with respect to their radial overtone (Miglio et al. 2008). The dominant inhomogeneity is the change in mean molecular weight at the edge of the convective core, whose size can therefore be measured. This is a method alternative to measuring the frequency of the g_c mode in p mode pulsators.

Several of the SPB and γ Doradus stars rotate with periods comparable to their oscillation periods. This, again, calls for models that take rotation into account with a more sophisticated approach than perturbation theory; the corresponding work is in progress.

The γ Doradus stars are located in a domain where the influence of convection on the pulsations is significant; convection is also responsible for the red edge of the δ Scuti instability domain. Modeling with a time-dependent convection approach allowed Dupret et al. (2005) to reproduce the observed boundaries of these instability strips, and also to perform predictions of mode excitation in these stars. The situation for the SPB and γ Doradus stars with respect to asteroseismology is therefore the same as for δ Scuti stars: the basic theory is in place, the difficulty remains to find stars permitting the extraction of the required information from the observations.

4.4 Beta Cephei Stars

The β Cephei stars are massive ($\sim 9\text{--}17 M_{\odot}$) early-B main sequence stars that oscillate radially and nonradially in p, g, and mixed modes of low radial overtone. This is roughly the same type of modes as excited in the δ Scuti stars, but asteroseismology has been more successful for β Cephei stars in the recent past due to several reasons.

The observed frequency spectra are simple enough to provide initial clues for mode identification, yet complicated enough to reveal information about the stars' interior structures. Photometric and spectroscopic mode identification methods (and combinations of both) could be applied successfully to some β Cephei stars (e.g., De Ridder et al. 2004), additionally aided by the large radial velocity to light amplitude ratios (of the order of several hundreds km/s/mag) of their pulsation modes. Sufficient information for unique identifications of all modes was obtained; an example was shown earlier.

Asteroseismic modeling was also eased for β Cephei stars, as radial modes have sometimes been identified (☛ Sect. 3.3, ☛ Fig. 4-11). This immediately reduces the parameter space in which a seismic model must be sought by one dimension. Due to the evolutionary state of β Cephei stars (near the center of the main sequence), a few of the observed nonradial modes are of mixed p/g type, which has provided information about the convective core size and/or differential interior rotation in a number of stars (e.g., Aerts et al. 2003; Pamyatnykh et al. 2004).

Apart from learning about stellar interiors, asteroseismology of β Cephei stars has interesting astrophysical implications. Given their high masses, they are progenitors of supernovae of Type II which are largely responsible for the chemical enrichment of galaxies. The evolution of massive stars is strongly affected by rotational mixing and angular momentum transport (Maeder and Meynet 2000); their internal rotation profile is testable by asteroseismology of β Cephei stars.

However, the field has not yet matured to a point where we can claim satisfactory understanding of all aspects of the physics governing the interior structures of massive main sequence stars. Several questions may be answered by seismic sounding of β Cephei stars, such as: How strong is their differential interior rotation? How efficient is internal angular momentum transport (Townsend 2009)? How strong is convective core overshooting? Can all stars between $\sim 9\text{--}17 M_{\odot}$ be modeled with the same convective overshooting parameter?

There are additional questions related to the pulsation physics of β Cephei stars that need to be addressed. Only the center of the theoretically predicted β Cephei instability strip is populated by observed pulsators. Is this an observational shortcoming or a weakness of theory? What is the upper mass limit of the β Cephei stars? Are there post-main sequence β Cep stars, contrary to theoretical predictions?

Five different observables strongly depend on the opacities and element mixtures used for theoretical modeling of β Cephei stars: the radial fundamental to first overtone period ratio, the excited range of pulsation modes, the frequencies of p modes with radial overtones larger than two, the dependence of bolometric flux amplitude on the surface displacement (Daszyńska-Daszkiewicz and Walczak 2009), and in case of a "hybrid" pulsator (see below) the excited range of g modes. Most of these observables are largely independent of each other; modeling of some stars shows that no standard input opacities and element mixtures can explain the pulsation spectra in detail. Therefore, the last question that may be answered from asteroseismology of β Cephei stars is: where must we improve the input physics for stellar modeling?

Some β Cephei and SPB stars have emission-line spectra and are thus Be stars. These objects rotate rapidly and have circumstellar disks, occasional outbursts, etc. They may also be studied asteroseismically, but their oscillations are hard to be detected and identified, and their rapid rotation requires special theoretical treatment – that is underway.

4.5 Pulsating Subdwarf Stars

Three types of pulsating subdwarf stars are known: long-period subdwarf B (sdB) stars that oscillate in high-overtone g modes (V1093 Herculis stars), short-period sdB stars pulsating in low-overtone p and g modes (V361 Hydrae stars), and the only oscillating subdwarf O (sdO) star known to date, a low-overtone p mode pulsator.

Although their g modes would allow the sounding of deep interior regions, their faintness, long periods, and low amplitudes made V1093 Herculis stars escape asteroseismic study so far. Theoretical studies of the sdO pulsator have so far been focused on the problem of mode excitation. Therefore, the only subdwarf pulsators that have been asteroseismically modeled are among the V361 Hydrae stars.

This is no easy undertaking as the problem with mode identification and mostly sparse (but for a few stars rich and highly variable) frequency spectra again occurs. It is possible that these objects have steep interior rotation gradients. Pulsation models must also be built upon evolutionary models including the effects of gravitational settling and radiative levitation. In practice, modeling is carried out by surveying parameter space in effective temperature, gravity, mass and hydrogen mass fraction, and by seeking best agreement between observed and theoretically predicted oscillation frequencies. Results have been obtained on about a dozen of those stars, and a mass distribution consistent with that expected from a double star evolutionary scenario has been obtained. Charpinet et al. (2009) elaborated on many aspects of asteroseismic modeling of pulsating subdwarf stars.

4.6 Rapidly Oscillating Ap Stars

The rapidly oscillating Ap (or short: roAp) stars are special among the pulsating stars because their high-overtone p mode oscillations are predominantly governed by a magnetic field that aligns their pulsation axis with the magnetic, and not with the rotation, axis of the stars. The magnetic field also distorts the pulsations modes, so they can no longer be described with a single spherical harmonic.

Because of their short periods and very low amplitudes (below 1%) these stars have for many years been studied photometrically only; an extensive review was given by Kurtz and Martinez (2000). Spectroscopic observations of roAp stars, however, provided a new level of insight, particularly into atmospheric physics of these pulsators.

The reason is that the vertical wavelength of the pulsation modes is shorter than or about the same order of the layer thicknesses of the chemical elements in their atmospheres that are highly stratified by radiative levitation. Therefore the radial velocity amplitudes of the oscillations change with line depth and from one chemical species to the other. The chemical elements are also inhomogeneously distributed over the surface, allowing three-dimensional studies of the abundances and pulsational geometry (Kurtz 2009).

Because of the unique possibilities offered by the atmospheric structure of roAp stars, spectroscopy is also much more sensitive in detecting oscillations compared to photometry. As an outstanding example, Mkrtichian et al. (2008) detected a complete $l = 0-2$ mode spectrum for Przybylski's star over 3–4 radial orders and performed some initial seismic modeling.

4.7 Solar-Like Oscillators

The low amplitudes of the stochastically excited oscillations of solar-type stars made their observational detection elusive for a long time. In retrospect, the first detection was made by Kjeldsen et al. (1995), but confirmed only several years later. Meanwhile, the observational accuracy has improved to an extent that detections were made in hundreds of stars (mostly giants), and seismic analyses of several were performed.

The potential of solar-like oscillations for seismic sounding is large. Once detected, the pulsation modes are rather easy to be identified because they are high-overtone p modes; see [Sect. 3.1](#) for an example. So are the large and the small frequency separations (if detectable), immediately placing main sequence stars on the (asteroseismic) HR diagram.

Most interesting, as for all pulsators with nearly-asymptotic frequency spectra, are irregularities in the latter, caused by features in the stellar interiors. These would for instance be the base of their envelope convection zones, or the helium ionization region. Houdek (2009) gave an overview of the expected seismic signatures of such features and their astrophysical importance.

The phenomenon of avoided crossing ([Fig. 4-4](#)) also takes place in solar-like oscillators once they have reached the subgiant stage, and it makes itself obvious in Echelle diagrams (Bedding et al. 2007). As for other types of pulsator, this would allow an asteroseismic determination of the convective core size (for stars sufficiently massive to possess a convective core).

The limited lifetimes of the intrinsically damped solar-like oscillations enable inferences concerning pulsation mode physics. The observed power spectra at individual mode frequencies show a multitude of peaks whose overall shape would correspond to a Lorentzian. The half-widths of these Lorentzians yield a determination of the mode damping rates; the *mode lifetimes* are inversely proportional to those. The mode lifetimes are in turn dependent on properties of the surface convection zone.

Theoretical predictions of the amplitudes of solar-like oscillators are important not only for understanding their physics, but also for planning observational efforts. After years of predictions resulting in amplitudes larger than were observed, the incorporation of the mode lifetimes and subsequent computation of *mode heights* appear to result in a scaling law that estimates observed amplitudes well (Chaplin et al. 2009), and seems in agreement with measurements up to oscillating giants (Hekker et al. 2009).

Finally, it is worth to note that as all cool stars possess a convective envelope, it can be expected the solar-like oscillations are excited in all of them, up to red supergiants (Kiss et al. 2006).

4.8 Hybrid Pulsators

Some of the pulsational instability strips shown in [Fig. 4-1](#) partly overlap. It is therefore logical to suspect that stars that belong to two different classes of pulsating stars, having two different

sets of pulsational mode spectra excited simultaneously, may exist. Indeed, a number of those have been discovered. This is good news for asteroseismology as the information carried by both types of oscillation can be exploited. The confirmed cases have so far always been high-overtone g modes and low-overtone p modes, as evaluations of the pulsation constants of the oscillations show. Mixed-mode pulsations by themselves are not “hybrid” pulsations because they occupy the same frequency domain as pure p modes.

There are δ Scuti/ γ Doradus stars, β Cephei/SPB stars, and long/short-period subdwarf B pulsators. The main physical difference between the B-type and A/F-type “hybrid” pulsators is that in the first group the same driving mechanism excites both types of oscillation, whereas in the δ Scuti/ γ Doradus stars two main driving mechanisms are at work. The cooler δ Scuti stars and all γ Doradus stars should have thin surface convection zones that would support the excitation of solar-like oscillations (e.g., Samadi et al. 2002).

The revision in our knowledge of the solar chemical element mixture (Asplund et al. 2004) resulted in the prediction of a much larger overlap region between β Cephei and SPB stars in the HR diagram, and the frequency ranges of the excited long- and short-period modes in “hybrid” B-type pulsators suggest that the heavy-element opacities used for stellar model calculations are still too low (Dziembowski and Pamyatnykh 2008; Handler et al. 2009). Intriguingly, a similar conclusion has been independently obtained from helioseismology (Guzik et al. 2009).

One of the first hybrid pulsators reported in the literature (HD 209295, Handler et al. 2002) turned out to have its g mode oscillations tidally excited by another star in a close eccentric orbit. The changing gravitational influence of the companion gives rise to forced oscillations with frequencies that are integer multiples of the orbital frequency. The tidal deformation of the pulsating star is similar to a sectoral $l = 2$ mode, and such modes are therefore most easily excited. There are also several known cases of pulsation in ellipsoidal variables. Whereas close binarity represents an additional complication for theoretical asteroseismic studies due to the gravitational distortion of the pulsator, some other cases of binarity can be used to obtain additional constraints for seismic modeling.

4.9 Pulsation in Eclipsing Binaries and Open Clusters

The fundamental way to determine stellar physical parameters, in particular masses, is the analysis of detached eclipsing binary systems whose components can be assumed to have evolved as if they were single (Torres et al. 2010). Such constraints are most welcome for asteroseismology and therefore it is self-evident that pulsators be sought for in such binaries. To date, several dozens of such systems are known. Most of these would be δ Scuti pulsators, some are β Cephei stars, but the best studied case is a short-period sdB pulsator.

These objects provide another possibility for mode identification (Nather and Robinson 1974): Throughout the eclipse, different parts of the stellar surface become invisible. In case of nonradial oscillations, only part of the pulsation mode is seen, and the light amplitudes and phases change according to the type of mode. In this way, the oscillation mode can be identified, a method known as *eclipse mapping* or *spatial filtration* (e.g., Reed et al. 2005).

Another way to gain support for asteroseismic modeling is to study pulsators in stellar clusters. Besides the possibility to observe several objects simultaneously (e.g., via CCD photometry or multifiber spectroscopy), cluster stars can be assumed to have originated from the same

interstellar cloud. Therefore, they should be of the same age and chemical composition. These parameters can be well determined from the properties of the cluster itself, and be imposed as a constraint on the seismic modeling procedure of all pulsating cluster members, also known as *ensemble asteroseismology*.

4.10 A New Era in Precision

Most observational results reported in this chapter so far were based on classical ground based observing methods, such as single- and multisite photometry and spectroscopy. However, a new era in observational asteroseismology has begun.

Asteroseismology requires knowledge about as many intrinsic stellar oscillation frequencies as possible, and these often have low amplitude, calling for measurements with the highest accuracy. As this is also a requirement for the search for extrasolar planets, synergies between the two fields have emerged.

Spectroscopically, asteroseismology has benefited from high-precision radial velocity techniques, such as the iodine cell method, originally invented to find extrasolar planets. Measurements of oscillations in distant stars with amplitudes down to 20 cm/s, about one-tenth of human walking speed, have become possible in that way. Only about a dozen spectrographs worldwide are capable of reaching the required precision. Such observations are still therefore expensive in terms of observing time and complexity of data analysis, but new observing networks, such as the Stellar Observations Network Group (SONG, <http://astro.phys.au.dk/SONG/>) aim at achieving similar precision on a regular basis. SIAMOIS (Sismomètre Interférentiel À Mesurer les Oscillations des Intérieurs Stellaires, Mosser et al. 2009) is expected to work at the same precision and duty cycle with its single node placed in Antarctica.

Concerning photometry, the main problem of ground based observations is scintillation, irregular changes in the measured intensity due to anomalous atmospheric refraction caused by turbulent mixing of air with different temperatures. The solution to this problem is to observe stellar variability from space. Here the synergy with extrasolar planet research is that the measurements have sufficient precision to detect the signature of transits of planets when they pass in front of their host star.

Aside from fine-guidance sensor Hubble Space Telescope photometry, the first asteroseismic data from space were due to an accident. The main science mission of the WIRE (Wide-Field Infrared Explorer) satellite failed due to loss of cooling fluid for the main instrument, but the star trackers, of 52 mm aperture, were consequently used for time-resolved photometry. An overview of the results can be found in the paper by Bruntt and Southworth (2008); WIRE has fully ceased operation in October 2006.

The first dedicated asteroseismic space mission that was successfully launched is MOST (Microvariability and Oscillations of STars, <http://www.astro.ubc.ca/MOST/>, Matthews 2006; Walker et al. 2003). The spacecraft is in orbit since June 2003 and still continues to provide asteroseismic data with its 15-cm telescope as main instrument. One of the most interesting results from MOST was the discovery of high-overtone g mode oscillations in a supergiant B star (Saio et al. 2006), suggesting an extension of the SPB star instability domain to the highest luminosities.

The CoRoT (Convection, Rotation and Transits, Baglin 2003) mission was successfully launched in December 2006 and hosts a 27-cm telescope. It observes two fields of $2 \times 3^\circ$ on

the sky each, one devoted to asteroseismology of a few bright stars, and the other searching for planetary transits in many stars, at the same time performing a high-precision stellar variability survey. A special volume of *Astronomy & Astrophysics* (2009) reports some of the early CoRoT science.

Asteroseismically, the detection of solar-like oscillations in almost a thousand giant stars is remarkable (Hekker et al. 2009). In addition, one of the most intriguing CoRoT results was the detection of large numbers of variability frequencies in two δ Scuti stars. Interpreted as independent pulsation modes, these would supply hundreds of oscillations to be asteroseismically modeled, reverting the basic problem for the study of these objects: first, there were too few modes available, now there would be too many! However, doubts have been raised whether all the frequencies extracted from those data would really correspond to normal modes of pulsation, or would rather be a signature of granulation (Kallinger and Matthews 2010).

Originally designed for the search for earth-like planets in the habitable zone, the latest addition to the asteroseismic space fleet became Kepler (<http://www.kepler.nasa.gov>), launched in March 2009. Asteroseismology can measure the radii of solar-like oscillators, among which should be planet-hosting stars, to a relative accuracy of 3%. Thus, the radii of transiting planets would be known to 3%. Therefore, it was decided to devote a small percentage of the observing time of this space telescope with an effective 95-cm aperture to asteroseismology. Kepler is the most powerful photometry tool for asteroseismology to date and observes a $\sim 10 \times 10^\circ$ field for 3 years, practically without interruption, providing a time base considerably longer than all other present missions. Initial results of the Kepler Asteroseismic Investigation, based on the first 43 days of science operations, were summarized by Gilliland et al. (2010).

A possible shortcoming of all these asteroseismic space missions is that they observe in only one passband. Therefore, all information that is available for seismic modeling are the targets' oscillation frequencies, unless mode identifications are provided by ground-based support observations. These are often difficult because the larger space telescopes mostly observe faint targets. The BRITE-Constellation (BRiGht Target Explorer, <http://www.brite-constellation.at>) mission therefore adopts quite a different strategy. BRITE consists of three pairs of nanosatellites hosting a 3-cm telescope each that will observe in at least two different passbands (one per satellite), facilitating mode identification with the photometric method. BRITE-Constellation will preferentially observe stars brighter than fifth magnitude and has a large $24 \times 24^\circ$ field of view. Given the brightness of the science targets, mode identification from high-resolution spectroscopy can also easily be done. The first pair of satellites is to be launched in early 2012.

Finally, PLATO (PLANetary Transits and Oscillations, <http://sci.esa.int/plato>) is a mission designed to provide a full statistical analysis of exoplanetary systems around nearby and bright stars. Currently in the definition phase, it will host 28 telescopes of 10 cm aperture that will observe two 557 sq. deg fields for 2.5 years each. It is intended to observe 100,000 stars to a precision of 1 ppm per month and 500,000 stars to somewhat poorer accuracy to determine stellar and planetary masses to 1%.

4.11 Prospects and Problems

Asteroseismology is a research field evolving so rapidly that some of the results reported here will already be outdated when this book appears in print. Given the ~ 30 -year headstart of helioseismology in comparison, the field now is in its teenage years, but matures rapidly.

The theoretical basis for asteroseismic studies is laid, although far from being perfect. Some of the problems that require solution comprise improved treatment of magnetic fields, convection, internal flows, and fast rotation. It is still poorly known what causes stellar cycles, and what makes certain classes of pulsator select the types of mode they oscillate in. Some asteroseismic results point toward a requirement of still higher heavy-element opacities, which current calculations do not seem capable of providing.

Observationally, pulsating white dwarf stars, β Cephei stars, and solar-like oscillators have been studied asteroseismically, and continue to be. There are high hopes that present asteroseismic space missions will open the δ Scuti, SPB, γ Doradus, and V1093 Herculis stars for interior structure modeling, and further improve the situation for V361 Hydrae stars. Future high-precision radial velocity networks and sites will improve our knowledge mostly for solar-like oscillators and roAp stars, with the latter guiding theory of stellar pulsation under the influence of rotation and magnetic fields. The Kepler mission will provide asteroseismic results for solar-like oscillators en masse, and a large number of massive stars are expected to be studied with BRITE-Constellation. The Gaia mission (<http://www.rssd.esa.int/gaia>) is expected to provide accurate luminosity determinations for a vast number of asteroseismic targets, tightly constraining the modeling.

It is therefore only appropriate to finish with a quote by Eyer and Mowlavi (2008): now is the time to be an asteroseismologist!

Acknowledgments

I am grateful to Victoria Antoci, Tim Bedding, Günter Houdek, and Mike Montgomery for their comments on this manuscript, as well as to Joris De Ridder and Thomas Lebzelter for helpful input. Tim Bedding and Patrick Lenz provided some results reproduced here. I apologize to all colleagues whose work was not properly cited here due to a rigorous restriction on the total number of literature sources to be quoted in this chapter.

References

- Aerts, C. 2007, *Lecture Notes on Asteroseismology* (Leuven: Katholieke Universiteit Leuven)
- Aerts, C., Thoul, A., Daszyńska, J., Scuflaire, R., Waelkens, C., Dupret, M. A., Niemczura, E., & Noels A. 2003, *Science*, 300, 1926
- Aizenman, M., Smeyers, P., & Weigert, A. 1977, *A&A*, 58, 41
- Asplund, M., Grevesse, N., Sauval, A., Allende Prieto, C., & Kiselman, D. 2004, *A&A*, 417, 751
- Astronomy & Astrophysics*. 2009, *A&A*, 506 (Special Issue)
- Baglin, A. 2003, *Adv Space Res*, 31, 345
- Baker, N., & Kippenhahn, R. 1962, *Zeitschrift für Astrophysik*, 54, 114
- Bedding, T. R., et al. 2004, *ApJ*, 614, 380
- Bedding, T. R., et al. 2007, *ApJ*, 663, 1315
- Brassard, P., & Fontaine, G. 2005, *ApJ*, 622, 572
- Brickhill, A. J. 1991, *MNRAS* 251, 673
- Bruntt, H., & Southworth, J. 2008, *JPCS*, 118, 012012
- Campbell, W. W., & Wright, W. H. 1900, *ApJ*, 12, 254
- Chaplin, W. J., Houdek, G., Karoff, C., Elsworth, Y., & New, R. 2009, *A&A*, 500, L21
- Charpinet, S., Brassard, P., Fontaine, G., Green, E. M., Van Grootel, V., Randall, S. K., & Chayer, P. 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 585
- Christensen-Dalsgaard, J. 1988, in *Advances in Helio- and Asteroseismology*, Proc. IAU Symp., Vol. 123, ed. J. Christensen-Dalsgaard & S. Frandsen (Dordrecht: Reidel), 295
- Christensen-Dalsgaard, J. 2002, *Rev Mod Phys* 74, 1073

- Christensen-Dalsgaard, J. 2010, *Lecture Notes on Stellar Oscillations* (Århus: Aarhus University)
- Cox, J. P. 1984, *PASP*, 96, 577
- Daszyńska-Daszkiewicz, J., & Walczak, P. 2009, *MNRAS*, 398, 1961
- Daszyńska-Daszkiewicz, J., Dziembowski, W. A., Pamyatnykh, A. A., & Goupil, M.-J. 2002, *A&A*, 392, 151
- Daszyńska-Daszkiewicz, J., Dziembowski, W. A., & Pamyatnykh, A. A. 2005, *A&A*, 441, 641
- De Ridder, J., et al. 2004, *MNRAS*, 351, 324
- Deubner, F.-L., & Gough, D. O. 1984, *Ann Rev Astron Astroph*, 22, 593
- Dufour, P., Green, E. M., Fontaine, G., Brassard, P., Francoeur, M., & Latour, M. 2009, *ApJ*, 703, 240
- Dupret, M.-A., Grigahcène, A., Garrido, R., Gabriel, M., & Scuflaire, R. 2005, *A&A*, 435, 927
- Dziembowski, W. A. 1977, *Acta Astron*, 27, 203
- Dziembowski, W. A., & Goode, P. R. 1992, *ApJ*, 394, 670
- Dziembowski, W. A., & Pamyatnykh, A. A. 1991, *A&A*, 248, L11
- Dziembowski, W. A., & Pamyatnykh, A. A. 2008, *MNRAS*, 385, 2061
- Eyer, L., & Mowlavi, M. 2008, *J Phys Conf Ser* 118, 012010
- Fabritius, D. 1596, reported by Wolf, R. 1877, *Geschichte der Astronomie* (Munich: Verlag R. Oldenbourg), 116
- Fleming, W. P. 1899, reported by Pickering, E. C., et al. 1901, *ApJ*, 13, 226
- Frandsen, S., et al. 2002, *A&A*, 394, L5
- Frost, E. B. 1902, *ApJ*, 15, 340
- Gautschi, A., & Saio, H. 1995, *Ann Rev Astron Astroph*, 33, 75
- Gilliland, R. L., et al. 2010, *PASP*, 122, 131
- Gizon, L., Birch, A. C., & Spruit, H. C. 2010, *Ann Rev Astron Astroph*, submitted (arXiv:1001.0930), 48, 289
- Goodricke, J. 1786, *Phil Trans R Soc Lond*, 76, 48
- Gough, D. O. 1986, in *Hydrodynamic and Magnetodynamic Problems in the Sun and Stars*, ed. Y. Osaki (Tokyo: University of Tokyo Press), 117
- Gough, D. O. 1996, *Observatory*, 116, 313
- Green, E. M., et al. 2003, *ApJ*, 583, L31
- Guzik, J. A., Keady, J. J., & Kilcrease, D. P. 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 577
- Handler, G. 1999, *MNRAS*, 309, L19
- Handler, G., et al. 2002, *MNRAS*, 333, 262
- Handler, G., et al. 2006, *MNRAS*, 365, 327
- Handler, G., et al. 2009, *ApJ*, 698, L56
- Hekker, S., et al. 2009, *A&A*, 506, 465
- Herschel, W. 1782, reported by Pickering, E. C. 1884, *Proc Am Acad Arts Sci*, 19, 269
- Houdek, G. 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 519
- Jeffery, C. S. 2008, *Inf Bull Var Stars*, 5817
- Jones, P. W., Hansen, C. J., Pesnell, W. D., & Kawaler, S. D. 1989, *ApJ*, 336, 403
- Kallinger, T., & Matthews, J. M. 2010, *ApJ*, 711, L35
- Kaye, A. B., Handler, G., Krisciunas, K., Poretti, E., & Zerbi, F. M. 1999, *PASP*, 111, 840
- Kepler, S. O., et al. 2005, *ApJ*, 634, 1311
- Kilkenny, D., Koen, C., O'Donoghue, D., & Stobie, R. S. 1997, *MNRAS*, 285, 640
- Kim, A., Montgomery, M. H., & Winget, D. E. 2006, in *New Horizons in Astronomy*, ASP Conf. Ser., Vol. 352, ed. S. J. Kannappan et al. (San Francisco, CA: Astronomical Society of the Pacific), 253
- Kiss, L. L., Szabó, G. M., & Bedding, T. R. 2006, *MNRAS*, 372, 1721
- Kjeldsen, H., Bedding, T. R., Viskum, M., & Frandsen, S. 1995, *AJ*, 109, 1313
- Kurtz, D. W. 1982, *MNRAS*, 200, 807
- Kurtz, D. W. 2000, in *Delta Scuti and Related Stars*, ASP Conf. Ser., Vol. 210, ed. M. Breger & M. H. Montgomery (San Francisco, CA: Astronomical Society of the Pacific), 287
- Kurtz, D. W. 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 491
- Kurtz, D. W., & Martinez, P. 2000, *Baltic Astron*, 9, 253
- Kurtz, D. W., Shibahashi, H., Dhillon, V. S., Marsh, T. R., & Littlefair, S. P. 2008, *MNRAS*, 389, 1771
- Landolt, A. U. 1968, *ApJ*, 153, 151
- Maeder, G., & Meynet, G. 2000, *Ann Rev Astron Astroph*, 38, 143
- Matthews, J. M. 2006, *Highlights of Recent Progress in the Seismology of the Sun and Sun-Like Stars*, 26th Meeting of the IAU, Joint Discussion 17, #21
- McGraw, J. T., Liebert, J., Starrfield, S. G., & Green, R. 1979, *White Dwarfs and Variable Degenerate Stars* (Rochester: University of Rochester), 377
- Metcalfe, T. S., 2005, *MNRAS*, 363, L86
- Metcalfe, T. S., Montgomery, M. H., & Kanaan, A. 2004, *ApJ*, 605, L133
- Miglio, A., Montalbán, J., Noels, A., & Eggenberger, P. 2008, *MNRAS*, 386, 1487
- Mkrtichian, D. E., Hatzes, A. P., Saio, H., & Shobbrook, R. R. 2008, *A&A*, 490, 1109
- Montagner, J.-P., & Roullet, G. 2008, *JPCS*, 118, 012004
- Montgomery, M. H. 2005, *ApJ*, 633, 1142

- Montgomery, M. H. 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 605
- Montgomery, M. H., Williams, K. A., Winget, D. E., Dufour, P., De Gennaro, S., & Liebert, J. 2007, *ApJ*, 678, L51
- Mosser, B., et al. 2009, *Comm Asteroseis*, 158, 337
- Mukadam, A., et al. 2007, *ApJ*, 667, 433
- Nather, R. E., & Robinson, E. L. 1974, *ApJ*, 190, 637
- Nather, R. E., Winget, D. E., Clemens, J. C., Hansen, C. J., & Hine, B. P. 1990, *ApJ*, 361, 309
- Pamyatnykh, A. A., Handler, G., & Dziembowski, W. A. 2004, *MNRAS*, 350, 1022
- Percy, J. R. 2007, *Understanding Variable Stars* (Cambridge, UK: Cambridge University Press)
- Pesnell, W. D. 1985, *ApJ*, 292, 238
- Reed, M. D., Brondel, B. J., & Kawaler, S. D. 2005, *ApJ*, 634, 602
- Reese, D. R., Thompson, M. J., MacGregor, K. B., Jackson, S., Skumanich, A., & Metcalfe, T. S. 2009, *A&A*, 506, 183
- Robinson, E. L., Kepler, S. O., & Nather, R. E. 1982, *ApJ*, 259, 219
- Rosseland, S., & Randers, G. 1938, *Astrophisica Norvegica*, 3, 71
- Saio, H., et al. 2006, *ApJ*, 650, 1111
- Samadi, R., Goupil, M.-J., & Houdek, G. 2002, *A&A*, 395, 563
- Suran, M., Goupil, M., Baglin, A., Lebreton, Y., & Catala, C. 2001, *A&A*, 372, 233
- Talon, S., Zahn, J.-P., Maeder, A., & Meynet, G. 1997, *A&A*, 322, 209
- Tassoul, M. 1980, *ApJS*, 43, 469
- Telting, J. H. 2003, *Ap&SS*, 284, 85
- Telting, J. H., & Schrijvers, C. 1997, *A&A*, 317, 723
- Torres, G., Andersen J., & Giménez, A. 2010, *A&ARv*, 18, 67
- Townsend, R., 2009, in *Stellar Pulsation: Challenges for Observation and Theory*, AIP Conf. Proc., Vol. 1170, eds. J. A. Guzik & P. A. Bradley (Melville, NY: American Institute of Physics), 355
- Unno W., Osaki Y., Ando H., Saio H., & Shibahashi H. 1989, *Nonradial Oscillations of Stars* (2nd ed., Tokyo: University of Tokyo Press)
- Vogt, S. S., & Penrod, G. D. 1983, *PASP*, 95, 565
- Waelkens, C., & Rufener, F. 1985, *A&A*, 152, 6
- Waelkens, C., Aerts, C., Kestens, E., Grenon, M., & Eyer, L. 1998, *A&A*, 330, 215
- Walker, G., et al. 2003, *PASP*, 115, 1023
- Watson, R. D. 1988, *Ap&SS*, 140, 255
- Winget, D. E., Robinson, E. L., Nather, R. E., & Fontaine, G. 1982, *ApJ*, 262, L11
- Winget, D. E., et al. 1991, *ApJ*, 378, 326
- Winget, D. E., et al. 1994, *ApJ*, 430, 839
- Winget, D. E., Sullivan, D. J., Metcalfe, T. S., Kawaler, S. D., & Montgomery, M. H. 2004, *ApJ*, 602, L109
- Woudt, P. A., et al. 2006, *MNRAS*, 371, 1497

5 Star Formation

Simon Goodwin

Department of Physics & Astronomy, University of Sheffield,
Sheffield, South Yorkshire, UK

1	<i>Introduction</i>	245
1.1	The Basic Model of Star Formation	245
1.2	What Is a Star?	246
1.3	Open Questions	247
2	<i>From Gas to Stars</i>	247
2.1	Star Forming Regions	248
2.2	Molecular Clouds	249
2.2.1	Observations of MCs	249
2.2.2	Molecular Tracers	249
2.2.3	Dust as a Tracer	250
2.2.4	The Appearance of Star Forming Regions	250
2.2.5	The Structure and Kinematics of MCs	252
2.2.6	The Formation of Cores in Molecular Clouds	253
2.2.7	The Formation of Molecular Clouds	254
2.3	Low-Mass Star Formation	254
2.3.1	The Physics of Core Collapse	255
2.3.2	The Stages of Star Formation	257
2.3.3	Different Types of Young Stars	259
2.4	The Initial Mass Function	259
2.4.1	The Origin of the IMF	261
2.4.2	The IMF from Cores	261
2.4.3	Competitive Accretion	262
2.4.4	The CMF Versus Competitive Accretion	263
2.5	The Formation of Massive Stars	264
2.6	The Formation of Brown Dwarfs	265
2.7	Star Formation Efficiency	266
3	<i>Multiple Stars, Star Clusters, and the End of Star Formation</i>	266
3.1	Binary and Multiple Systems	266
3.1.1	Binary Formation	268
3.1.2	Binary Destruction	269
3.2	Binaries or Singles?	270
3.3	Star Clusters	270
3.4	The End of Star Formation	272

3.4.1	Gas Expulsion	273
3.4.2	Dynamical Evolution	274
3.5	Is Star Formation Universal?	274
4	<i>Conclusions</i>	275
	<i>References</i>	275

Abstract: Stars are one of the most important constituents of the Universe, and understanding their formation is crucial to many areas of astrophysics. Stars form from dense molecular gas, and they tend not to form in isolation. Stars often form in binary and multiple systems, and these systems tend to form in clusters with 10^2 – 10^5 members. Stars also form with a wide range of masses, from substellar brown dwarfs with masses $<0.1 M_{\odot}$ to massive stars $>100 M_{\odot}$, and wherever stars form the distribution of their masses seems always to be the same. This chapter will review our current understanding of star formation from cold gas to young star clusters.

Keywords: Initial mass function, Molecular clouds, Stars, Star formation, Star clusters

1 Introduction

How stars form is one of the big questions in modern astrophysics. Stars are, in many ways, the fundamental (baryonic) components of the Universe. Most of the electromagnetic radiation detected by our telescopes comes directly or indirectly from stars, they are the basic visible components of galaxies, they are the crucibles in which heavy elements are produced, and they are the hosts of planetary systems and even, in at least one case, life.

Understanding star formation requires us to understand gravity, turbulence, magnetic fields, chemistry, thermodynamics, and radiative transfer processes, all acting together with extremely complex interactions and interdependencies. The study of star formation involves understanding physical processes that work on galactic kpc-scales, star cluster formation on pc-scales, and star (and planet) formation on au- and stellar-scales, with mass ranges from whole galaxies of $>10^{12} M_{\odot}$ to Jupiter-sized planets of only $10^{-3} M_{\odot}$.

The enormous range of scales which need to be probed to understand star formation make it extremely challenging. Only in recent years have computers become powerful enough to start dealing with the problem of star formation, informed by observations which are now available (often only from space) across the whole electromagnetic spectrum.

This chapter will not even attempt to cover much of the theory and observations of star formation (entire books are not enough). Rather, it will give a general overview of some of the more important processes and our understanding of them.

1.1 The Basic Model of Star Formation

The basic theory of star formation dates back to 1734 when Emanuel Swedenborg first proposed the Nebula Hypothesis: that the Sun and the Solar System formed from a rotating cloud of gas that collapses under gravity to form the Sun at the center and a disk around it from which the planets formed. This model was expanded later in the eighteenth century by Immanuel Kant and Pierre-Simon Laplace. Perhaps surprisingly, the basic idea of the nebula hypothesis is still the basis of our understanding of star formation. Current thinking is that star formation can be divided into several main stages.

Molecular cloud formation. A large (10^3 – $10^6 M_{\odot}$) cloud of gas forms within the interstellar medium. The high column and volume densities allow the cloud to cool, and remain cool, and molecular hydrogen (and other molecules) to form.

Prestellar core formation. The molecular cloud fragments into self-gravitating condensations known as “clumps” and, on smaller scales, “cores.” “Prestellar cores” of roughly a solar mass are the birth places of stars.

Embedded star formation. Prestellar cores collapse and form a protostar (actually, often a binary or multiple system) surrounded by a disk of gas near their centers. Initially these protostars are large (AU-size) hydrostatically supported objects deeply embedded in the gas of the core.

Pre-main sequence stars. Once most of the mass in the core is accreted onto the star(s), and young stars on the pre-main sequence are observed to be surrounded by massive disks. These disks may well be in the process of planet formation and fairly rapidly disappear. Eventually the star will reach the main sequence and become a “normal” main sequence star.

Star clusters. A molecular cloud usually fragments into many cores, forming a star cluster of between 10^2 (arguably even just ten stars can comprise a “cluster”) and 10^6 stars in only a few pc^3 in a few Myr.¹ The extreme densities in star clusters result in frequent encounters between stars and multiple systems which can destroy young multiple systems and disrupt circumstellar (and planet forming) disks.

The end of star formation. Once one or more massive ($>10 M_\odot$) stars form, the input of energy from their UV radiation fields (and associated HII regions), stellar winds, and eventual supernovae will expel any gas that has not already formed stars. This prevents more star formation and often results in the destruction of the star cluster.

These stages do not necessarily happen one after the other. Obviously stars cannot form before a molecular cloud has formed, but prestellar cores and pre-main sequence stars are often observed in the same cluster, while at the same time massive stars may have started to clear some regions of the cluster of gas so stopping star formation while star formation is still occurring in other regions (indeed, the removal of gas from one part of a cloud may well induce star formation in another).

1.2 What Is a Star?

Before proceeding, it is useful to define what is meant by a “star.” Formally a star is defined as an object that will, is, or has produced energy through hydrogen fusion in its core. This is any object with a mass in excess of the hydrogen-burning limit of $\sim 0.08 M_\odot$. For the purposes of this chapter, the “will” in this definition is crucial as many of the young objects that are discussed are far from reaching the main sequence and beginning core hydrogen burning.

A brown dwarf is an object too small to fuse hydrogen, but large enough to undergo a short-lived deuterium-burning phase (which produces very little energy as deuterium is very rare). Brown dwarfs span the mass range between $10^{-2} M_\odot$ (about ten Jupiter masses) to $\sim 0.08 M_\odot$. Planets are therefore objects with masses $< 10^{-2} M_\odot$, but this chapter will be generally unconcerned with planets and objects of even lower mass.

While this formal distinction between stars and brown dwarfs exists, throughout this chapter we will generally assume that brown dwarfs and stars are fundamentally the same type of object. Star formation should have no reason to draw a distinction between brown dwarfs and

¹The difference between a massive star cluster and a small dwarf galaxy is that a star cluster forms all of its stars at roughly the same time with roughly the same metallicity (a “simple stellar population”), while a dwarf galaxy has many episodes of star formation over many Gyr.

stars. Only the most massive objects in young star forming regions have yet fused hydrogen to generate energy (even if they eventually will), and the physics of star formation should find nothing special about the hydrogen-burning limit. Therefore every statement about the formation of low-mass stars (especially very low-mass stars) should also apply to the formation of brown dwarfs.

The most massive stars that are forming now have masses of 150–300 M_{\odot} . It is difficult to estimate the masses of the most massive stars as young, massive stars are deeply embedded in gas. By the time the most massive stars are easily observable, after 1–2 Myr, they have evolved significantly and may have lost >50% of their initial mass.

1.3 Open Questions

There are a number of critical open questions in star formation:

The nature of molecular clouds. Stars tend to form in massive gas and dust clouds with sizes of 10s pc and masses up to $10^6 M_{\odot}$. These clouds appear to contain high levels of turbulence which drives and controls star formation. What is the source of their turbulence and what is the lifetime of molecular clouds?

The origin of the initial mass function (IMF). Observations suggest that stars form with the same distribution of masses everywhere. There are very few brown dwarfs, and very few very high-mass stars. Most stars (90%) are M-dwarfs with masses in the range 0.1–0.5 M_{\odot} . Interestingly, and strangely, the IMF appears almost universal – observations of star forming regions show that their IMF always seems to be approximately the same. Why is it universal? And why is the typical mass always around 0.2 M_{\odot} ?

The origin of multiple systems. Many (possibly most) stars do not form as single, isolated, objects, but in multiple systems with two or more members. The distribution of separations between members of multiple systems is extremely wide, with some companions almost touching, while others have orbits of millions of years. Why do stars form in multiple systems? And how do they produce such a wide range of separations between companions?

Is star formation universal? Do stars everywhere form in the same way? Star clusters range in mass from 10^2 to $10^6 M_{\odot}$. Do more massive clusters just form more stars, or is there a fundamental difference in how they form stars? Do some regions form more or different binaries to other regions, or is it always the same?


There are many good reviews of, and introductions to, star formation, in particular: the *Protoplasts and Planets V* volume from 2007, and many articles from *Annual Reviews of Astronomy & Astrophysics* (especially recently McKee and Ostriker 2007, and Zinnecker and Yorke 2007), as well as the textbook *The Formation of Stars* by Stahler and Palla from 2005.

2 From Gas to Stars

Stars form from the gravitational fragmentation of cold gas. In the early Universe there were no stars; as the Universe cooled and the first dark matter halos formed, gas was able to collapse and cool in these halos. The formation of the first stars in a zero metallicity environment was probably very different to the formation of stars today, and they were probably very massive, short-lived stars (this is beyond our remit, but see Stacy et al. 2011 and references therein).

The introduction of metals to the interstellar medium (ISM) by the first generation of stars created dust and massively increased the efficiency of cooling, and later generations of stars were able to form in much the same way that stars currently form.

2.1 Star Forming Regions

Stars usually form in large complexes. In  [Fig. 5-1](#) we see the Carina Nebula, a massive star forming complex. The image is about 20 pc across and illustrates almost every feature of star formation. One of the most obvious features is the complex and filamentary nature of the gas throughout the region. This is due to a combination of supersonic turbulence in the gas and the feedback of energy from the most massive stars. Stars are forming throughout the region including many very massive stars. η Carina is a $\sim 100 M_{\odot}$ star found at the far left as a bright blob just to the left of the Keyhole Nebula.² To the middle right is the star cluster Trumpler 14, while to the far left is the cluster Trumpler 16 (less easily visible but containing η Carinae) which are only a few Myr old. Along the bottom is a molecular ridge with very obvious dust pillars carved by the feedback from massive stars in Trumpler 14.

The Carina Nebula is a particularly extreme star forming region forming some $10^5 M_{\odot}$ of stars in several clusters. Many stars also form in midrange clusters such as the Orion Nebula Cluster (the fuzzy blob in Orion's sword) with masses of $10^3 M_{\odot}$, and many in very small clusters with only around 10^2 members such as the Taurus star forming region. As is described later, each of these types of star cluster contribute roughly the same mass of stars to the general field population of galaxies.



 Fig. 5-1

An amazing HST ACS image of the Carina Nebula which is a massive star forming region. The image covers roughly 20 pc (Credit: NASA, ESA, N. Smith (University of California, Berkeley), and The Hubble Heritage Team (STScI/AURA), Credit for CTIO image: N. Smith (University of California, Berkeley) and NOAO/AURA/NSF)

²There is a lovely clickable version of this image available at <http://heritage.stsci.edu/2007/16/supplemental.html> describing many of the features.

2.2 Molecular Clouds

The most abundant element in the Universe is hydrogen. Hydrogen is most often found in one of three forms: ionized (HII), atomic (HI), and molecular (H_2). The interstellar medium (ISM) can be broadly divided into three phases: the hot ISM with temperatures $>10^4$ K which contains mainly HII, the warm ISM with temperatures between about 100 and 10^4 K containing HI, and the cold ISM at <100 K containing mainly H_2 .

Stars form in the cold phase in clouds of gas and dust known as molecular clouds (MCs). Most molecular gas is found in Giant Molecular Clouds which typically have sizes of 10–100 pc, and masses of 10^3 – 10^6 M_\odot , but some molecular gas is found in smaller clouds known as Bok Globules. MCs have high column densities, shielding them from the interstellar radiation field and cosmic rays. This means that their temperatures can become very low (only 10s K) which means that they can form molecular hydrogen and many other molecular species.

2.2.1 Observations of MCs

Clouds of molecular hydrogen are unfortunately remarkably hard to observe. In the optical they are only observable as “dark clouds” as MCs can have tens of magnitudes of optical extinction meaning it is impossible to see their contents or anything behind them (which is how they were first discovered). Observations in the IR are able to see embedded (proto)stars within MCs, but not the MCs themselves.

Obviously it would be extremely useful to observe the structure and kinematics of the gas within MCs. But molecular hydrogen in MCs is most often unobservable: at the low (10s K) temperatures of MCs H_2 has no easily excited states. The only possible way of observing cold H_2 directly is to use UV absorption along the line of sight from a massive star, unfortunately there are very few of these lines of sight available, and almost never where you would like them to be. Therefore “tracers” must be used: other molecules or dust in the MC which are assumed to trace the underlying H_2 distribution.

For a far fuller description of the physics and techniques of MC observations see Evans (1999) and early chapters of Stahler and Palla (2005).

2.2.2 Molecular Tracers

Even though H_2 is not excited at low temperatures, many other molecular species are, and these are observable at radio (and mm) wavelengths. Molecular lines can be extremely useful as line ratios can provide information on the local (kinetic) temperature, and line widths provide information on bulk flows and nonthermal motions.

The emission of a particular line from a molecule depends on both the temperature and *volume* density (i.e., collision rate). However, the strength of the emission also depends on the column density of material.

Probably the most common tracer is CO which has a number of easily excitable lines at low temperature in the radio. CO is destroyed at low column densities by UV radiation and cosmic rays and so it can only exist in high (column) density environments such as MCs, and the CO column density is assumed to trace the H_2 column density, therefore allowing CO to be used to map H_2 . Due to its relatively high abundance CO lines can saturate at high column densities.

In such situations isomers of CO can be used (such as C^{18}O) which are less abundant and so do not saturate. However, CO freezes-out onto dust grains at very low temperatures ($<10\text{ K}$), forming an ice mantle around the dust grain.

Other molecules are used to trace regions of high volume density, for example, CS, HCN, NH_3 , and many others. Different molecules (and lines) can be used to trace regions of different density. NH_3 is a particularly common tracer for densities of $\sim 10^3\text{ cm}^{-3}$, while HCN traces volume densities of $10^6\text{--}10^8\text{ cm}^{-3}$. Line profiles can also be used to detect signatures of outflows and infall; however, the details of using and interpreting molecular tracers are quite complex and involved and the reader is referred to Evans (1999) for more information.

The power of using several molecular tracers in tandem is that the general column density structure of molecular clouds can be mapped (usually in CO), and regions of high volume density can be located using other tracers. In addition, the thermal and kinetic properties of the gas can be determined.

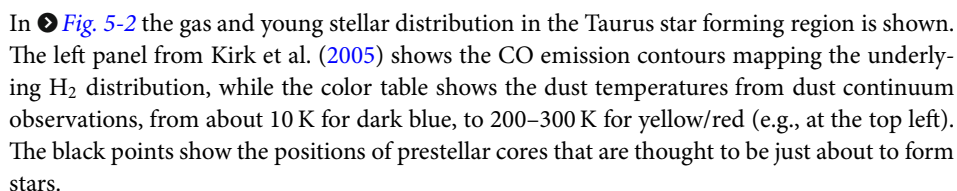
2.2.3 Dust as a Tracer

MCs contain significant quantities of dust (estimated to be roughly 1% of their mass). Unlike molecular tracers, dust does not emit line radiation, rather dust particles are large enough to emit a thermal continuum. The dust has the same kinetic temperature as the surrounding gas as collisions with gas particles are able to keep the dust and gas in thermal balance. At the typical temperatures of MCs of $10\text{--}100\text{ K}$ the peak thermal emission from dust lies in the sub-mm at wavelengths of hundreds to tens of microns.

Thus the total sub-mm flux in a region provides a measure of the column density of dust (which is converted into a gas column density using a gas-to-dust ratio that is usually taken to be 100-to-1), and the spectral energy distribution provides the gas (kinetic) temperature.

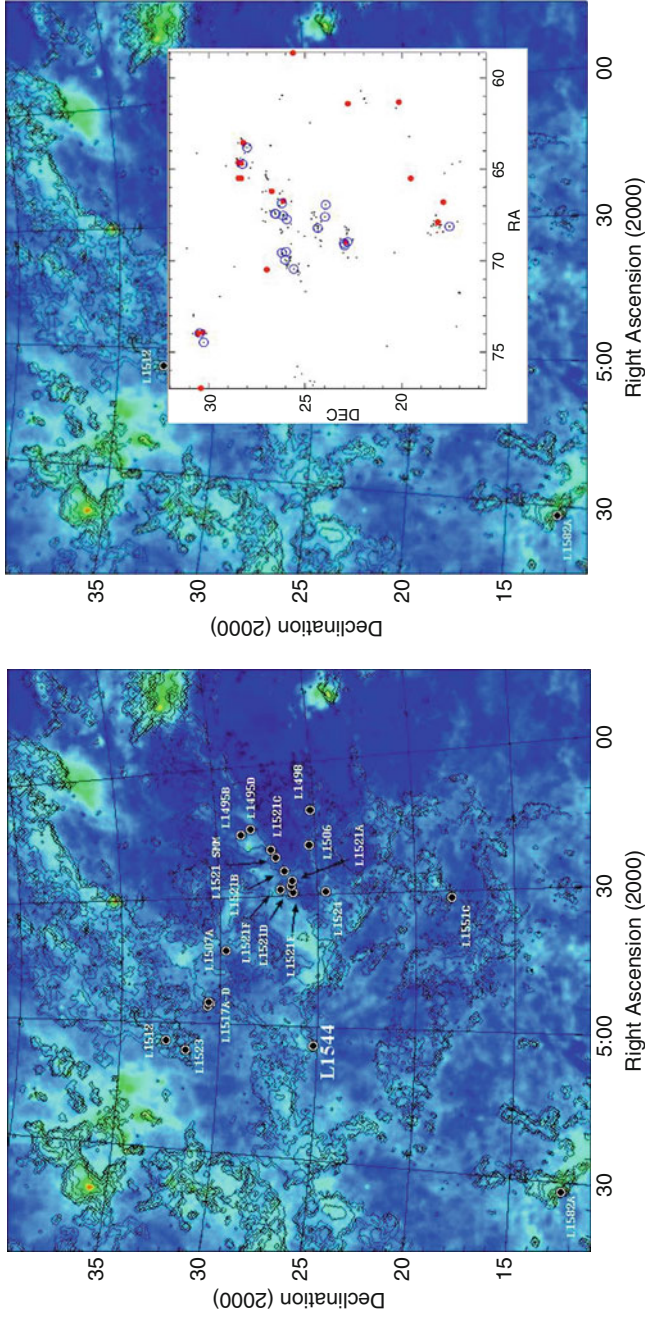
Unfortunately from the ground there are only a few atmospheric windows in the sub-mm, most notably at 450 and $850\text{ }\mu\text{m}$, and so the whole sub-mm is not available (obviously this is not a problem from space and the Herschel mission is able to observe the whole of the sub-mm).

2.2.4 The Appearance of Star Forming Regions

In  Fig. 5-2 the gas and young stellar distribution in the Taurus star forming region is shown. The left panel from Kirk et al. (2005) shows the CO emission contours mapping the underlying H_2 distribution, while the color table shows the dust temperatures from dust continuum observations, from about 10 K for dark blue, to $200\text{--}300\text{ K}$ for yellow/red (e.g., at the top left). The black points show the positions of prestellar cores that are thought to be just about to form stars.

The right panel from Parker et al. (2011) shows the positions of the young stars in Taurus. Blue circles show the positions of brown dwarfs and very low-mass stars, red points the positions of stars $>1\text{ M}_\odot$, and the black points the positions of intermediate-mass stars.

There are several things to note about these figures which will become important in our discussions below. Firstly, the gas is not distributed evenly, it is clumpy and substructured with regions of very low density and filamentary structure in the gas. Secondly, prestellar cores – the sites of new star formation – are distributed in a similar way to the gas, generally forming



■ Fig. 5-2

Left panel: The gas distribution in Taurus from Kirk et al. (2005) showing CO emission contours, dust temperature in the background color table, and black points show the position of prestellar cores. *Right panel:* The stellar distribution in Taurus from Parker et al. (2011), blue circles show the positions of brown dwarfs and very low-mass stars, red points the positions of stars $> 1 M_{\odot}$, and the black points the positions of intermediate-mass stars. Note that the projection of the two figures is slightly different

where the gas is densest. Thirdly, young stars also follow the gas distribution, but less closely as they have had some chance to disperse from their formation sites, or for gas to clear from their formation sites.

That cores, and therefore stars, form where the gas is densest should not be a surprise. If the gas is substructured, then the initial stellar distributions in clusters will also be substructured. However, substructure does not remain for long. Most young clusters appear relatively smooth and circular which suggests that some process occurs very rapidly to remove the memory of the initial distribution of the stars/gas.

2.2.5 The Structure and Kinematics of MCs

Almost every MC is observed to have a large degree of internal density and kinematic structure down to the resolution limits of the observations. All MCs appear similar to Taurus (see [Fig. 5-2](#)) with clumps and filaments on all scales (Williams et al. 2000).

Often the structure in MCs is divided into “clouds,” “clumps,” and “cores”:

Clouds are the largest structures representing the whole molecular cloud with typical sizes of pcs, masses of 10^3 – $10^6 M_\odot$, and densities of $\sim 100 \text{ cm}^{-3}$.

Clumps are denser regions within clouds which are associated with star cluster formation. They have sizes of around a pc, masses of a few hundred M_\odot , and typical densities of 10^3 – 10^4 cm^{-3} .

Cores are the sites of star formation in single or multiple systems with sizes of $< 0.2 \text{ pc}$, masses of around a M_\odot , and densities of $> 10^4 \text{ cm}^{-3}$.

It should be noted that most of the mass in MCs is at relatively low density, clumps and cores only make up a few percent of the total mass of a MC. The distinctions between clouds, clumps, and cores are rather arbitrary, and in many ways the structure in MCs appears scale-free which has led to suggestions that the structure of MCs is fractal (e.g., Elmegreen 2002).

This of course raises the question as to why MCs have such complex (fractal even?) structure? The answer is probably that MC structure is created and maintained by turbulence.

Observations of molecular line widths are a probe of the velocity dispersions of MCs. Line ratios show that the kinetic temperatures of MCs are fairly uniform, and very low, with typical temperatures throughout a MC of $< 100 \text{ K}$, down to only 10 – 20 K . However, line widths show significant supersonic motions in MCs. At 10 – 20 K , the sound speed of the gas is around 0.2 km s^{-1} ; however velocity dispersions of many km s^{-1} are not unusual in MCs.

Larson (1981) found that the velocity dispersions of clouds as measured from the line widths σ are proportional to the size of the region observed L

$$\sigma \propto L^\beta$$

where $\beta \sim 0.5$. Thus the larger the region that is observed, the larger the velocity dispersion. This relationship has been ascribed to turbulence, which has become a major factor in modern theories of star formation.

Turbulence is a rather complex subject with a huge literature on both terrestrial and astrophysical turbulence. Turbulence is a process by which energy on one scale can be transferred to smaller scales through a series of “eddies” where an eddy is a local vorticity.³ Turbulence proceeds through a “cascade” where energy on large scales can be transferred to smaller scales,

³Energy is transferred through vortex stretching in which local velocity gradients can amplify the vorticity.

and then to smaller scales still through eddies. Eddies are not isolated, larger eddies will contain smaller eddies, which can contain still smaller eddies.

Importantly, turbulence is a dissipative process. Energy is transferred from the largest scale downward and eventually the scale will be such that the energy is able to dissipate. At the smallest scales, turbulence will dissipate due to molecular viscosity, although in astrophysical turbulence the scale on which self-gravity dominates adds a somewhat larger scale on which the behavior of turbulence changes.

In supersonic turbulence energy is also dissipated by shocks. Supersonic collisions cause shocks which are discontinuities in the gas properties, usually seen as a strong density jump across the shock. The strong enhancement of density in shocks provides the initial conditions for self-gravitating fragments to form which will eventually become stars.

Turbulence can be described by examining the relationship between the wavenumber (scale⁻¹) κ of the turbulence, and the energy contained on that scale $E(\kappa)$

$$E(\kappa) \propto \kappa^n$$

where n measures the spectrum of the turbulence. For pressureless turbulence, the turbulence is expected to have the Kolmogorov–Burgers spectrum of $n = 1.75$ (Boldyrev 2002).

It is hopefully obvious that if there is more energy on larger scales, the observed velocity dispersion would increase with the size of the region observed as seen by Larson (1981). Indeed, the linewidth-size relation with $\beta \sim 0.5$ that is observed in MCs is what would be expected from a Kolmogorov–Burgers spectrum (McKee and Ostriker 2007).

2.2.6 The Formation of Cores in Molecular Clouds

In shocks and clumps the increased density can form gravitationally bound prestellar cores. Prestellar cores are the basic unit of star formation as they will each produce an individual stellar system (a single star, or often a binary or multiple system) typically with a size of $<1,000$ AU. Prestellar cores may form in relative isolation (such as Bok Globules), or in their hundreds of thousands to make massive clusters.

An object will collapse if its mass exceeds the local Jeans mass. The Jeans mass M_J is given by

$$M_J = \frac{\pi}{6} \frac{c_s^3}{G^{3/2} \rho^{1/2}} \quad (5.1)$$

where c_s is the sound speed of the gas, ρ is its density, and G is the gravitational constant. For a typical molecular cloud with a temperature of around 10 K the sound speed is $\sim 0.2 \text{ km s}^{-1}$, and the typical number density in a clump is $n = 10^3\text{--}10^4 \text{ cm}^{-3}$ and so the Jeans mass can be rewritten in more useful units as

$$M_J \sim (2M_\odot) \left(\frac{c_s}{0.2 \text{ km s}^{-1}} \right)^3 \left(\frac{n}{10^3 \text{ cm}^{-3}} \right)^{-1/2}$$

Therefore the Jeans mass in a typical molecular clump in a cloud is roughly a solar mass. This is remarkably close to the mean mass of a star of $\sim 0.4 M_\odot$.

A prestellar core can be created if around $1 M_\odot$ of gas can be compressed to $n > 10^3 \text{ cm}^{-3}$ which shocks from supersonic turbulence can easily manage. It should be noted for later that to make a core of $0.1 M_\odot$ the density needs to be roughly 100 times greater, and that cores of $100 M_\odot$ will typically contain 100 Jeans masses of material.

2.2.7 The Formation of Molecular Clouds

The mechanism by which MCs form is unclear. Opinion is split on whether MCs are long-lived and virialized (Mouschovias et al. 2006; Tan et al. 2006) or transient, nonequilibrium structures (Ballesteros-Paredes 2006; Elmegreen 2000). This has important consequences as to the timescale of star formation: Is it long-lived and quasi-static, or rapid and dynamic? The fundamental question is whether star formation occurs on only one or two crossing times (rapid) or many crossing times (quasi-static). (See also reviews by Mac Low and Klessen 2004, and Ballesteros-Paredes et al. 2007).

Because turbulence is dissipative (see above), it does not last forever. Turbulence will decay on roughly the crossing time of the system, where the crossing time, t_{cross} , of the system is

$$t_{\text{cross}} \sim L/c_s.$$

where c_s is the (kinetic) sound speed, typically $\sim 0.2 \text{ km s}^{-1}$.

Therefore if supersonic turbulence is observed it means that either (a) the MC is less than a few crossing times old, or (b) the turbulence is being “driven” (i.e., energy is being added to the MC) to maintain it. Which of these is occurring is unclear, and the timescale of star formation is somewhere between one and ten crossing times (roughly between 1 and 10 Myr for a typical pc-scale clump).

An obvious way to determine the timescale of star formation is to measure the age-spread of stars within a cluster. However, this is extremely problematic. For example, Palla et al. (2005) claim a 10 Myr age-spread amongst stars in the Orion Nebula Cluster. Given that the Orion Nebula Cluster is only about 1 pc in size with a crossing time of around 1 Myr this would seem to argue for quasi-static star formation. However, Burningham et al. (2005) and Mayne and Naylor (2008) show that at least some (but possibly not all) of the age-spread may be accounted for by binarity, variability, and photometric errors. Rather worryingly, Naylor (2009) suggests that the ages of young clusters have been under-estimated by a factor of 2 due to uncertainties in PMS tracks.

Another clue might be from the formation mechanism of MCs. If the initial conditions of MCs are known, then how star formation will then proceed in the cloud may become clearer. Unfortunately, the formation mechanism(s) of MCs are unknown. It is thought that dense molecular gas forms when the warm neutral medium that makes up the majority of the volume of a galaxy is compressed or overrun by a shock; however, many possible mechanisms for this exist and there is little agreement on which are most important (see, e.g., Vázquez-Semadeni et al. 2007 and references therein).

2.3 Low-Mass Star Formation

Once a *bound* prestellar core has formed within a MC it will collapse to form a star or stars.

Before proceeding it is useful to define the two classes of young stars: protostars and pre-main sequence (PMS) stars. Usually “protostar” is used for very young stars before they have collapsed to stellar densities when they have radii of order 1 AU. A “PMS star” is a star that has collapsed to stellar densities but has not yet started to produce energy through hydrogen fusion at its center.

2.3.1 The Physics of Core Collapse

As described above, star formation typically begins with a prestellar core of mass $\sim 1 M_{\odot}$ at a temperature of ~ 10 K. Initially, the H_2 in the core loses energy by being in thermal balance with the dust. Impacts with dust grains transfers thermal energy from the gas to the dust which is then able to radiate that energy away as a blackbody. At the typical densities of a prestellar core of $\sim 10^{-20} \text{ g cm}^{-3}$ the core is optically thin to this radiation with a peak wavelength of $\sim 200 \mu\text{m}$ and the core temperature remains constant.

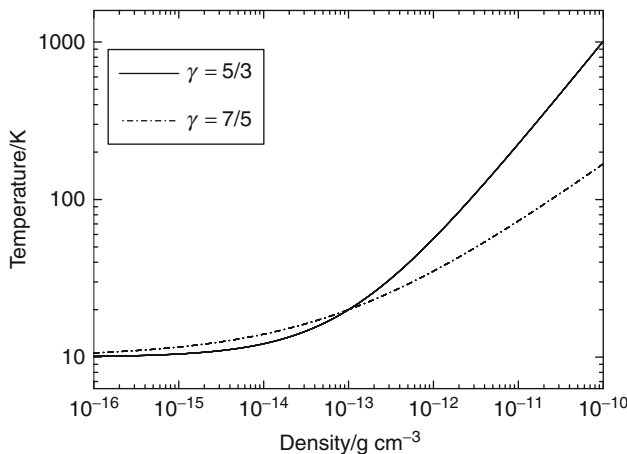
As the core collapses it is initially able to retain this thermal balance and the dust is able to radiate away the kinetic energy released by the collapse and so the temperature remains at ~ 10 K and the collapse is isothermal.

At a critical density of $\rho_{\text{crit}} \sim 10^{-13} \text{ g cm}^{-3}$ the combination of accelerating release of gravitational energy and the increasing column density of the gas mean that it is no longer able to radiate away the heat of the collapse efficiently. At this point the core begins to increase in temperature and the collapse becomes adiabatic, and it is around this point that the first proto-star(s) will form. Protostars initially have a size of order 1 AU and remain at this size for some time as they are only able to radiate on a Kelvin–Helmholtz timescale. This situation continues as protostars accrete more material and increase in temperature.

The thermal behavior of gas in a core can be simply described by a barytropic equation of state of the form


$$\frac{P(\rho)}{\rho} \equiv c_s^2(\rho) = c_0^2 \left[1 + \left(\frac{\rho}{\rho_{\text{crit}}} \right)^{1-\gamma} \right] \quad (5.2)$$

where P is the pressure, ρ the density, c_s is the general isothermal sound speed, and $c_0 \sim 0.2 \text{ km s}^{-1}$ is the isothermal sound speed in low-density gas. γ is the adiabatic index which is $5/3$ for a monatomic gas, and $7/5$ for a diatomic gas. [Figure 5-3](#) shows the temperature–density relationship for a barytropic equation of state with $\gamma = 5/3$ and $7/5$: an isothermal collapse followed by adiabatic heating.



■ Fig. 5-3

The dependence of temperature with density during core collapse for a barytropic equation of state ([Eq. 5.2](#)) with adiabatic index $\gamma = 5/3$ and $7/5$

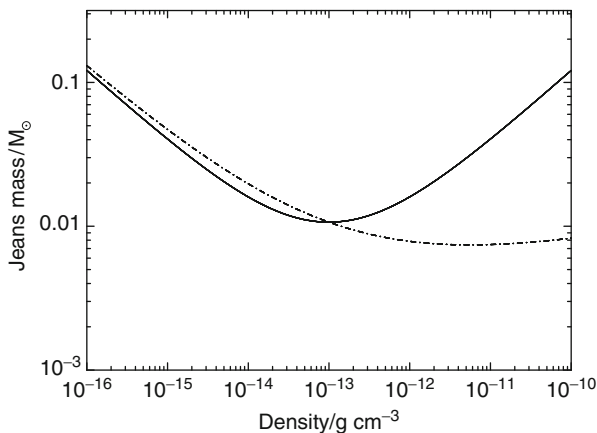
A consequence of this thermal behavior is that the Jeans mass of the core changes in an unusual way. A core is supported against collapse by thermal energy (c_s), and encouraged to collapse by its mass/density (ρ). The effect of the thermal behavior of the core is to lower the Jeans mass during isothermal collapse (ρ increases while c_s remains constant), before increasing it during adiabatic collapse (as the c_s^3 term dominates over $\rho^{-1/2}$). This means that there is a minimum Jeans mass during core collapse that is reached at a density of around $10^{-13} \text{ g cm}^{-3}$ of around $10^{-2} M_\odot$ (roughly ten Jupiter masses).  Figure 5-4 shows the behavior of the Jeans mass with density for barytropic equation of state with $\gamma = 5/3$ and $7/5$. The Jeans mass has a lower minimum when $\gamma = 7/5$ as the temperature rises more slowly with density. The minimum Jeans mass that is reached during core collapse is known as the *opacity limit for fragmentation* and is the minimum mass that any object can have if it forms by gravitational fragmentation.


Note that there is another minimum during the second collapse as molecular hydrogen dissociates which is potentially lower; however, it is unclear if fragmentation can occur during this phase, and it would only produce objects within 1 AU of each other. It could possibly explain some close binaries.


This situation continues until the temperature of the protostar reaches $\sim 2,000 \text{ K}$ at which point enough energy is available to dissociate molecular hydrogen. This dissociation provides a heat sink for the protostar and it rapidly (and almost isothermally) collapses to stellar densities (approaching 1 g cm^{-3}).

After all of the molecular hydrogen has been dissociated, the (PMS) star then slowly contracts again until the central density and temperature are high enough to start hydrogen fusion and the star joins the main sequence.

It is an interesting coincidence that the opacity limit for fragmentation of $\sim 10^{-2} M_\odot$ is also the point at which the planet-brown dwarf distinction is drawn. The planet-brown dwarf limit is usually taken to be the deuterium burning limit; however, the opacity limit for fragmentation might provide a far more physical distinction between planets and brown dwarfs – planets form by core accretion, while brown dwarfs form by gravitational fragmentation. The lack of



 Fig. 5-4

The dependence of the Jeans mass with density during core collapse for a barytropic equation of state  5.2) with adiabatic index $\gamma = 5/3$ and $7/5$

objects around ten Jupiter masses is explained by the difficulty of building a planet to such large masses, and the difficulty of creating a brown dwarf right at the lower limit of gravitational fragmentation.

For more detailed descriptions of the thermodynamics of core collapse see Larson (1969) or Masunaga and Inutsuka (2000).

2.3.2 The Stages of Star Formation

The evolution of a star forming core and then a young star is divided into four stages (classes). Figure 5-5 illustrates schematically how the spectral energy distribution (SED) of a low-mass core and/or star evolves during star formation (from Lada 1999).

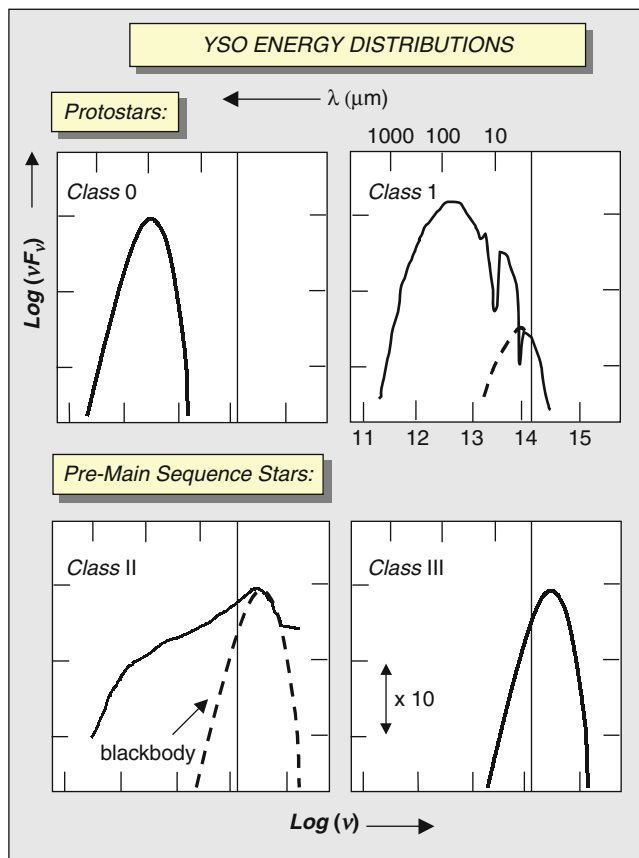


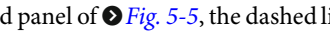
Fig. 5-5 The evolution of the spectral energy distribution of young stellar objects (YSOs) through the class 0 to class III phases of star formation (Figure from Lada 1999). Each tick on the x-axis represents an order of magnitude increase in energy emitted

Class 0. The very earliest stage of star formation is the almost free-fall collapse of the prestellar core and the formation of a class 0 object. In this stage the center of the core forms a protostar, but the majority of the mass of the core is still in the envelope. As the protostar is still heavily embedded in its natal envelope, it is optically invisible and is usually only seen as an IR point source. The emission from a class 0 core is a blackbody from cold (few 10s K) dust with a peak in the sub-mm at around $100\ \mu$ due to the low gas temperature.

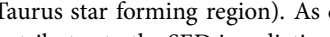
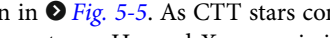
During the class 0 phase jets and outflows may also be visible. The protostar will have a strong magnetic field that is able to launch a small amount of material at high velocities along open magnetic field lines. This creates strong bipolar outflows which may be visible in a class 0 source. (The details of jet formation are extremely complex and well beyond the remit of this chapter, see the chapter on PMS stars for details).

The lifetime of the class 0 phase is extremely short as it occurs on roughly a free-fall time of the prestellar core which is $<10^5$ years.

Class I. The embedded protostar continues to accrete material from the envelope and also develop a massive disk. At the same time the protostar is heating the surrounding envelope by radiating the gravitational potential energy released by the collapse as well as potentially very significant energy input from increasingly prominent jets and outflows.

Once around half of the initial core mass has been accreted onto the protostar (and its massive disk) it enters the class I phase. Observationally the start of the class I phase is when the temperature of the envelope reaches 70 K as the protostar is still embedded and optically invisible. In the second panel of  Fig. 5-5, the dashed line at the bottom right shows the contribution of radiation directly from the star to the SED, but most of the radiation is reprocessed by the envelope with a peak at a few tens of microns.

The class I phase lasts a few $\times 10^5$ years, maybe 10^6 years, and usually it is at some point in the class I phase that the protostar collapses down to become a PMS star (this depends on the protostellar mass).

Class II/Classical T Tauri. By the end of the class I phase the bulk of the envelope has been accreted onto the (now) PMS star and its disk. The class II phase is now optically visible and is commonly known as a Classical T Tauri star (CTTs, named after the first object of its type T Tauri in the Taurus star forming region). As can be seen in the third panel of  Fig. 5-5, the dominant contributor to the SED is radiation directly from the (PMS) stellar photosphere. However, CTT stars show evidence for their disks through an IR excess where radiation from the star is reprocessed by the cool disk and reradiated creating an excess of long wavelength radiation as seen in  Fig. 5-5. As CTT stars continue to accrete material from the disk onto the star they show strong $H\alpha$ and X-ray emission as that material collides with the surface of the star in an accretion shock. CTT stars also have strong jets and magnetically driven outflows.

The class II/CTT phase lasts for a few $\times 10^6$ years until the disk is largely depleted (by a combination of accretion onto the star, planet formation, and evaporation by stellar radiation).

Class III/Weak-Lined T Tauri. Once the disk is depleted of gas the strong signatures of accretion and outflow that characterize the CTT phase also disappear. In the weak-lined T Tauri (WLT) phase only very weak spectral signatures and a slight IR excess from the disk remain. This phase is observationally difficult to distinguish from a main sequence star except that the star is rather over-luminous for its color as it has a larger radius than a main sequence star of the same mass as it is still contracting. During this phase the rest of the disk dissipates (leaving only a debris disk in some cases). Depending on the mass of the star this phase can take a few $\times 10^7$ years.

2.3.3 Different Types of Young Stars


The most common type of PMS star is a (classical or weak-lined) T Tauri star which is a typical PMS phase of stars $<2 M_{\odot}$, that is, the typical PMS phase of well over 90% of stars. But there are other classes of PMS stars.

Some $<2 M_{\odot}$ stars are classed as FU Orionis stars which are variable, with outbursts leading to increases of up to six magnitudes in a few months. Such stars are thought to be a subclass of T Tauri stars in which the accretion rate onto the star increases hugely. Typical T Tauri stars have accretion rates of $10^{-8} M_{\odot} \text{ year}^{-1}$, while FU Orionis stars are estimated to be as high as $10^{-4} M_{\odot} \text{ year}^{-1}$. With such high accretion rates FU Orionis stars cannot last very long (as a $0.1 M_{\odot}$ disk would be depleted in only 10^3 years). It may be that many T Tauri stars undergo a very short-lived FU Orionis phase at some point in their lives, or that very few do, but that they last longer and they rapidly deplete their disks.

The PMS phase of stars of $2-8 M_{\odot}$ are known as Herbig Ae/Be stars. They share most of the properties of T Tauri stars: IR excess (disks), Balmer emission, X-ray emission, and often variability.

The PMS phases of stars more massive than about $8 M_{\odot}$ are not observed as such massive stars are able to reach the main sequence while still in the deeply embedded (class 0/I) phase. The formation of massive stars will be discussed in more detail later.

2.4 The Initial Mass Function

The initial mass function (IMF) is the distribution function of individual stellar masses at birth. There are many ways of parameterizing the IMF, probably the most popular (certainly theoretically) is the Kroupa (2002) three-part power-law illustrated in  Fig. 5-6 which has the form

$$N(M) \propto \begin{cases} M^{-\alpha_1} & m_0 < M/M_{\odot} < m_1, \\ M^{-\alpha_2} & m_1 < M/M_{\odot} < m_2, \\ M^{-\alpha_3} & m_2 < M/M_{\odot} < m_3, \end{cases}$$

This parameterization usually divides the IMF into three regions.

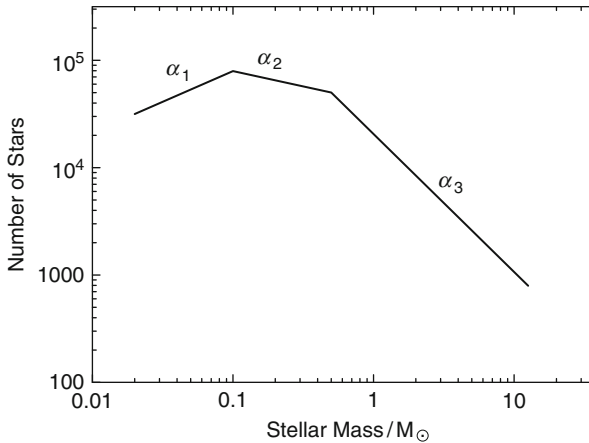
Firstly, a low-mass (substellar) regime between 0.02 and 0.08 or $0.1 M_{\odot}$ with declining numbers of stars with mass with a slope of $\alpha_1 = 0.3$.

Secondly, a roughly flat intermediate-mass region which covers M-dwarfs and the bulk of stars by a number between around 0.1 and $0.5 M_{\odot}$ with a slope of $\alpha_2 = 1.3$.

Thirdly, a high-mass regime in which the numbers of stars by mass decline with a slope of $\alpha_3 = 2.3$, also known as the Salpeter slope after the first determination of the IMF by Salpeter (1955).

It should be noted that recently Chabrier (2003) has proposed a log-normal-like parameterization which is becoming increasingly popular as well.

The number of stars by mass drops rapidly (seemingly as a power-law), and extremely high-mass stars $>10 M_{\odot}$ which become core collapse supernovae are very rare. Below $\sim 0.08 M_{\odot}$ objects are too small to burn hydrogen in their cores (brown dwarfs). While brown dwarfs are relatively numerous (one for every five or six stars), there is again a rapid decline in their numbers with decreasing mass, and they only contribute negligibly to the total mass in stars.



■ Fig. 5-6

A schematic representation of the IMF as a three-part power-law (Kroupa 2002). The three regions $-\alpha_1 = -0.3$ when $0.02 < M/M_{\odot} < 0.1$, $-\alpha_2 = -1.3$ when $0.1 < M/M_{\odot} < 0.5$, and $-\alpha_3 = -2.3$ (the Salpeter 1955 slope) when $M > 0.5 M_{\odot}$

Before proceeding it is worth emphasizing two points about the IMF.

Firstly, the IMF is the *initial* mass function which is the mass function that stars have at birth. Stellar evolution (such as the supernovae of the most massive stars) will remove massive stars over time, and dynamical evolution will tend to eject low-mass stars from star clusters. Both of these processes turn the IMF into a present day mass function (PDMF) which may be very different from the original IMF. But different PDMFs do not necessarily mean different IMFs.

For example, old globular clusters and very young star clusters have very different mass functions (e.g., old globulars completely lack stars larger than $\sim 1 M_{\odot}$ due to stellar evolution). In old globulars observations are of a population that has had 12 Gyr of stellar and dynamical evolution, while in very young clusters an unevolved IMF is seen. But simulations of the evolution of globular clusters suggest that to have the PDMF that is observed today, they must have started with an IMF very similar to that of very young star clusters (Vesperini and Heggie 1997).

Secondly, the IMF is the mass function of *individual* stars. Many stars form in binary and multiple systems and the construction of a true IMF involves correcting for these multiple systems. A G-dwarf-M-dwarf binary should contribute a single G-dwarf and a single M-dwarf to the IMF. In practice this is often impossible to achieve as many multiple systems may not be known to be multiple systems (especially if the companions are of low-mass and relatively close to the primary star) and so cannot be included as two systems. Therefore what is often presented as an IMF is often a primary star IMF (i.e., many companions are missed), or a system IMF (i.e., companion masses are merged with the primary mass).

These problems notwithstanding, the IMF is an extremely useful tool for examining star formation.

The form of the IMF suggests that the vast majority (90%) of stars are M-dwarfs, with the mean mass of a star being $\sim 0.4 M_{\odot}$, and the median mass of a star being $0.2\text{--}0.3 M_{\odot}$.

Surprisingly, the IMF seems to be remarkably universal. Wherever we observe the same basic form to the IMF is seen – few brown dwarfs, a peak at $0.2\text{--}0.3 M_{\odot}$, and a decline to higher masses with a slope of roughly 2.3 (Salpeter). Kroupa (2002) compiled a large database of IMF

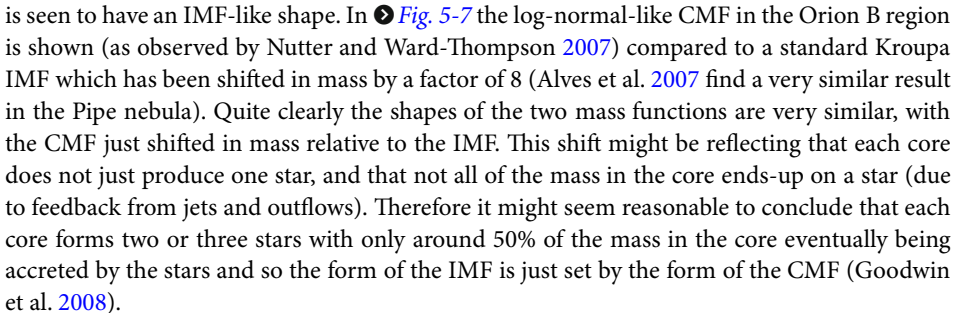
(and PDMF) determinations for clusters and regions in the Milky Way and LMC of different ages and metallicities and found no significant variations from a “standard” IMF. More recently Bastian et al. (2010) have examined many claims for nonstandard IMFs and found no strong evidence for significant deviations from a universal IMF (often while a different IMF can explain unusual observations, many other more reasonable explanations can be found such as different extinction laws).

2.4.1 The Origin of the IMF

There is no a priori reason why stars forming in $10^6 M_{\odot}$ proto-globular clusters with metallicities of only 10^{-3} solar should have the same IMF as stars forming in $10^2 M_{\odot}$ loose associations with solar metallicity. Indeed, it might be expected that the IMFs *should* be different, the physics of star formation appear to depend on the thermodynamics of the gas and the opacity limit for fragmentation which should depend on metallicity and density. So what is the origin of the IMF and why is it apparently universal? For more details the reader is directed to Bonnell et al. (2007) and Bastian et al. (2010).

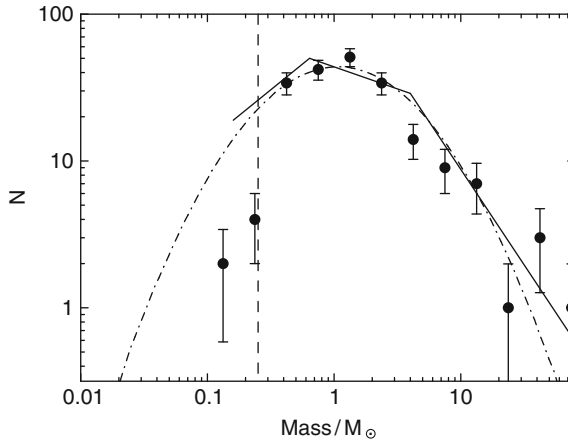
2.4.2 The IMF from Cores

Stars are seen to form in dense molecular cores. Presumably the mass of the core and the mass of the star (or stars) that form in them is related to the mass of the core; therefore, the distribution of stellar masses could well be related to the distribution of core masses. In such a scenario, the universality of the IMF may well just reflect the universality of the core mass function (CMF).

This picture is supported by observations of the CMF in different regions where the CMF is seen to have an IMF-like shape. In  Fig. 5-7 the log-normal-like CMF in the Orion B region is shown (as observed by Nutter and Ward-Thompson 2007) compared to a standard Kroupa IMF which has been shifted in mass by a factor of 8 (Alves et al. 2007 find a very similar result in the Pipe nebula). Quite clearly the shapes of the two mass functions are very similar, with the CMF just shifted in mass relative to the IMF. This shift might be reflecting that each core does not just produce one star, and that not all of the mass in the core ends-up on a star (due to feedback from jets and outflows). Therefore it might seem reasonable to conclude that each core forms two or three stars with only around 50% of the mass in the core eventually being accreted by the stars and so the form of the IMF is just set by the form of the CMF (Goodwin et al. 2008).

In many ways this may seem to be just shifting the question of the origin and universality of the IMF back one stage to the question of the origin and universality of the CMF. However, it is probably far less surprising that the CMF is universal as its origin lies in the nature of turbulence in MCs.

As described earlier, MCs are dominated by supersonic turbulence with apparently a universal power spectrum. Dense cores are formed in colliding/converging regions within MCs and the mass spectrum of cores will depend on the power spectrum of the turbulence and its strength. A core will collapse to form a star if it exceeds the local Jeans mass. The Jeans mass in a MC is $\sim 1 M_{\odot}$ at a density of about $10^{-18} \text{ g cm}^{-3}$, and so to form a prestellar core of $1 M_{\odot}$ turbulence needs to compress $1 M_{\odot}$ of gas to a density of $10^{-18} \text{ g cm}^{-3}$ – only a compression of a factor of about 10^3 or 10^4 above the average density in a MC. However, to make a



■ Fig. 5-7

Points show the core mass function of Orion B (from Nutter and Ward-Thompson 2007) compared to a Gaussian (*dot-dashed line*), and a Kroupa (2002) IMF shifted in mass by a factor of 8 (*solid line*). The *vertical dashed line* shows the completeness limits of the observations (Figure from Goodwin et al. 2008)

$0.1 M_{\odot}$ prestellar core the compression required is an extra factor of 100. Much larger cores, even though they require lower densities will be far larger in extent and expected to fragment into smaller cores unless they start at relatively higher densities (i.e., very super-Jeans).

Therefore very low-mass cores are expected to be rare as they require very high compression to reach the densities required for them to collapse, while cores of $\sim 1 M_{\odot}$ should be typical, and there should be few high-mass cores as they will fragment into smaller cores unless significantly compressed. This is exactly the form seen in observed CMFs and in the IMF (see Hennebelle and Chabrier 2008; Padoan and Nordlund 2002 for a far more detailed discussion of the creation of CMFs from turbulence). However, as is discussed later, the picture is far more complex than this.

2.4.3 Competitive Accretion

Competitive accretion (Bonnell et al. 1997, 2001, 2007) is the main competitor to core formation as the origin of the IMF. In competitive accretion stars form in cores which follow some CMF, but it is not the initial mass of the core that is important in setting the final mass of the star. Rather, cores, and later stars, continue to accrete gas as long as there is gas in the cluster to accrete (i.e., before it is blown away by feedback from massive stars).

Thus there is no such thing as a “core mass” which sets the mass of stars which form within it. The observed CMF is just the instantaneous amount of gas in dense condensations. Many of these condensations will collapse and form a star following the classical class 0 through class III evolutionary path. However, some stars will continue to accrete gas and continue to grow.

A key element of competitive accretion is that not all stars are equally successful in accreting material. More massive stars will be more successful in accreting material due to their

larger gravitational influence, but more importantly some stars will just be “lucky” in that they happen to be in regions where there is a lot of gas to accrete, while other stars will be in “unlucky” regions. Thus what becomes a massive star has no need to start as a particularly massive star. The protostar which forms the most massive star eventually can begin life in a small core.

2.4.4 The CMF Versus Competitive Accretion

Given that CMFs are observed to have a form that is very similar to the IMF it might be thought that this scenario is the most natural. However, there are a number of problems associated with the CMF-to-IMF model.

Firstly, the peak and width of the CMF should depend strongly on the Mach number of the turbulence. More energy in turbulence means stronger compression and a lower peak and more low-mass cores. Therefore the IMF should depend on the Mach number (Hennebelle and Chabrier 2008; Padoan and Nordlund 2002). MCs with Mach numbers ranging between around 5 and 20 have been observed (Ballesteros-Paredes et al. 2007), but the IMF is universal.

Secondly, it is uncertain how the observed CMF is related to the final IMF. It is often not clear which cores will collapse to form a star and which are transient density enhancements (i.e., which cores are really prestellar and which are not). In addition, cores of different masses evolve at different rates, and so any snapshot observation of a CMF is not a snapshot of the full range of cores that will eventually form stars (Clark et al. 2007). In addition, cores are often difficult to identify, and different methods can find different cores and different CMFs in the same data (Smith et al. 2008).

Indeed, it is not clear that the observed cores are the precursors of many stars. The typical cores that are observed have a size of roughly 0.1 pc. However, most stars form in dense clusters which contain (to take the example of the well-studied Orion Nebula Cluster) about 10^3 stars in a cubic pc. At such densities, the average distance between stars is the size of a core and so the filling factor of cores in the proto-Orion Nebula Cluster must have been unity which seems unlikely (Goodwin et al. 2007).

Thus the cores which form stars in dense clusters must have significantly smaller sizes, but the same mass spectrum as the more isolated cores with which CMFs are constructed. There is clearly an observational bias here, as the sizes of cores that can be observed are set by the resolution limits of the instruments used to observe them. There are no dense proto-clusters close enough to resolve many very small cores, and in more distant dense proto-clusters small cores would be unresolved. Indeed, it is found that the typical core mass increases with distance (resolution) which suggests that observers are unable to resolve smaller structures in these larger “cores.” It may be that all of the substructure cannot be resolved, even in nearby prestellar cores, and that they may contain substructure below current resolution.

It is likely that both the CMF and competitive accretion play a role in star formation and establishing the universal form of the IMF. Given the apparent link between the CMF and the IMF in diffuse regions where competitive accretion is likely to be less effective, it would appear that the CMF does largely set the form of the IMF. However, in dense environments it is difficult to imagine that cores and protostars could avoid accreting at least some of the large amounts of ambient gas around them.

2.5 The Formation of Massive Stars

The typical mass of a star of about $0.4 M_{\odot}$ is probably set by the typical mass of a core of about $1 M_{\odot}$, which is itself probably determined by the Jeans mass in a MC. However, the origin of stars significantly more massive than $1 M_{\odot}$ is very uncertain (see Zinnecker and Yorke 2007 for a detailed review of massive star formation).

There is a major theoretical problem in forming stars in excess of $\sim 10 M_{\odot}$, despite observations of many stars that appear to have masses well in excess of $100 M_{\odot}$, possibly up to $300 M_{\odot}$. As soon as a star reaches a mass of $\sim 10 M_{\odot}$ its central temperature and pressure will be high enough to begin core hydrogen burning. More massive stars evolve at a much faster rate than low-mass stars, a star $>40 M_{\odot}$ will evolve through the main sequence and post-main sequence to become a core collapse supernova in only a few Myr. And during their rapid evolution they have extremely strong stellar winds and UV fluxes. The question is then, if a star of $\sim 10 M_{\odot}$ is producing strong winds and a strong ionizing UV field, how does more material accrete onto the star to produce a star of 20, 50, or even $300 M_{\odot}$?

It is extremely difficult to observe young massive stars as they are always deeply embedded, and once they reach the main sequence, they produce an HII region. Their evolution can in some ways be tracked through the properties of the HII regions. Initially, a massive star will produce a hypercompact or ultracompact HII region which is less than 0.1 pc in size with densities of 10^4 cm^{-3} . As the HII region evolves it will increase in size and decrease in density toward a classical HII region of pc-scale as the radius of its Strömgren sphere increases. Once a massive star (or often stars) have ionized a large region, they begin to stop star formation locally. But it is in the hyper- and ultracompact phases that massive star formation is occurring and it is extremely difficult to observe the processes that are occurring.

In order to build a massive star accretion must overcome the force of radiative (UV) and mechanical (winds) feedback. Massive stars can have luminosities of $>10^5 L_{\odot}$ and mass-loss rates of $>10^{-5} M_{\odot} \text{ year}^{-1}$ which is a significant load to be overcome. By far the best way to do this seems to be via a disk. Even if feedback from the star is initially isotropic the presence of a disk means that it rapidly becomes anisotropic and escapes preferentially toward the poles allowing accretion through the disk (see Zinnecker and Yorke 2007 for details).

If it is possible to overcome feedback (and it must be as very massive stars exist), the formation of a massive star requires a large amount of gas to be available to accrete. There are two theories as to the origin of this huge reservoir of gas.

Possibly most obviously given our discussion of the core mass function is that massive stars form from massive cores (McKee and Tan 2003). Given the large amounts of feedback from a massive star, the core mass presumably needs to be significantly greater than the mass of the final star, and a $50 M_{\odot}$ star will need a $\gg 50 M_{\odot}$ core from which to form. Such massive cores are observed (as “clumps,” see above), but it is very unclear why a massive core should produce only one or two stars. A $100 M_{\odot}$ core contains many Jeans masses of material, and might be expected to form a small cluster rather than a single massive star. In addition, massive stars tend to be found in groups near the centers of clusters. For example, the Trapezium in the Orion Nebula Cluster contains four of the six most massive stars in the cluster in a region only around 0.1 pc across. But massive cores are by definition large, and so are not expected to form in the centers of clusters.

The alternative scenario for massive star formation is competitive accretion (see above). Here there is no need to form a massive core, rather massive stars begin as “normal” mass stars. A lucky few “normal” stars will sit in the deep potential where gas can be directed toward

them from which they can accrete. Thus they need no large initial reservoir of gas as the gas is channeled toward them. This has the advantage of automatically producing massive stars toward the centers of clusters where the potential is the deepest. However, it is not clear if a constant infall of gas can be maintained through the feedback from the most massive stars. Gas must accrete onto the massive star from a disk which can channel the feedback around it. If the inflow of gas into the center of the cluster is low-density then it could well be disrupted by feedback.

2.6 The Formation of Brown Dwarfs

At the other end of the mass spectrum from massive stars are very low-mass stars and brown dwarfs, and, as with massive stars, their formation is rather difficult to explain. The most obvious way in which brown dwarfs (and very low-mass stars, typically anything with a mass $< 0.1 M_{\odot}$) is in low-mass cores in a very similar way to stars. Young brown dwarfs appear very similar to young stars (Whitworth et al. 2007), and so their formation mechanism might well be exactly the same.

However, it may be difficult to form cores of a low enough mass to form brown dwarfs. To become bound a $0.1 M_{\odot}$ core must be 100 times denser than a $1 M_{\odot}$ core (due to the $\rho^{-1/2}$ dependency of the Jeans mass). This implies that very low-mass cores should only form where there is a strong shock which is able to increase the local density significantly. In particular, it implies that low-mass cores should form when the turbulence is strong (high Mach-number turbulence), and not when it is relatively weak. A problem arises in that the substellar IMF appears to be the same in all regions, even in low-density, relatively low-turbulence star forming environments such as Taurus.

In addition, very few very low-mass cores are observed. But for every very low-mass core which is able to become bound around 10^4 should be formed which just fail to reach the required densities and then disperse (Goodwin and Whitworth 2007). Therefore, huge numbers of very low-mass cores should be observed and they are not.

An alternative is to form brown dwarfs in higher-mass cores which we know to exist in significant numbers. However, the problem is that once an object has formed at around the opacity limit for fragmentation of $\sim 10^{-2} M_{\odot}$ if there is a significant amount of gas present, then it will accrete that gas (as happens in a “normal” star). Therefore some mechanism is required to stop the object accreting and to keep it at brown dwarf masses.

The first suggestion to accomplish this was the ejection scenario (Reipurth and Clarke 2001). In this scenario several very low-mass objects form at roughly the same time in a core. Systems with $N > 2$ are generally unstable and will rapidly decay and will usually eject the lowest-mass member of the system. Thus, brown dwarfs can be ejected from cores shortly after their formation and once they have left the core they have no more gas available for them to accrete.

The problem with this model is that brown dwarfs are typically ejected with speeds of a few km s^{-1} . This means that in a Myr an ejected brown dwarf could travel a few pc, and so clusters should have a halo of ejected brown dwarfs and the distribution of brown dwarfs should be different to that of stars. However, brown dwarfs and stars appear to have the same spatial distribution which suggests that violent ejection cannot be the answer (Luhman 2004).

Other possibilities have been suggested, such as forming brown dwarfs in massive extended disks around solar-type stars which can then be gently liberated (Stamatellos et al. 2007), or that brown dwarfs form as wide companions to M-dwarfs which are then disrupted

(Goodwin and Whitworth 2007). But every formation model has some problems, and the formation mechanism of brown dwarfs remains uncertain (indeed, they may form from a combination of all the processes that have been suggested).

2.7 Star Formation Efficiency

How much of its gas does a MC convert into stars? This may appear to be a simple question to answer: simply observe the total mass of gas M_g in a MC and the total mass of stars M_* , and the star formation efficiency (SFE) is then $SFE = M_*/M_g$. However, in practice this is rather difficult to do.

In order to measure the total mass of gas in a MC, there must obviously be gas present. This means that some (many? most?) of the stars will be embedded, either in their natal core or in the general cloud making them difficult to observe, therefore it is often not clear if the observed stellar mass is the total stellar mass.

That gas is still present means that star formation is ongoing. In particular more stars will form increasing the stellar mass, and the masses of stars already present will increase. If competitive accretion is occurring in the cluster the mass in stars could increase very significantly. Therefore, even if the current mass in stars is known exactly, the SFE is an instantaneous value, and not the final SFE (e.g., Evans et al. 2009). If more stars form from the same mass of gas the SFE will increase, but if more gas is channeled into the region, by turbulent flows or along filaments, for example, then the SFE could decrease.

Over whole galaxies the SFE (usually quoted as a star formation rate) can be calculated. The gas content ($\text{HI} + \text{H}_2$) of entire external galaxies can be observed (usually through 21 cm + CO). In addition, the *massive* star content of external galaxies can also be seen (either from HII regions/ $\text{H-}\alpha$ emission or naked O-stars) which, with an assumption of a standard IMF, gives the mass of young stars in that galaxy. Interestingly, the SFE in H_2 appears roughly constant in spiral galaxies at a few % (Leroy et al. 2008).

The SFE is observed to decrease with increasing scale. In prestellar cores the SFE is probably around 50% (Goodwin et al. 2008), in clumps 30–50% (Lada and Lada 2003), and in typical spiral galaxies as a whole it is usually a few percent (Leroy et al. 2008). Thus the majority of gas in a galaxy is not forming stars, and the majority of gas only ends up on a star when it reaches prestellar core scales of <0.1 pc.

3 Multiple Stars, Star Clusters, and the End of Star Formation

Stars do not form on their own. Many (possibly most) stars form as binary and multiple systems in which each core produces two or more stars on scales of a few hundred AU. In addition, most stellar systems (singles or multiples) form in star clusters of hundreds to millions of members on scales of roughly a pc.

3.1 Binary and Multiple Systems

In the discussions previously it has tended to be implicitly assumed that a single core produces a single star. However, it is known that this simplification is certainly not always true, and maybe wrong most of the time.

Observations of old main sequence stars in the solar neighborhood show that many stars are in binary and multiple systems. The numbers of stars in binaries is usually quantified by the binary fraction f_{bin}

$$f_{\text{bin}} = \frac{B + T + Q + \dots}{S + B + T + Q + \dots} \quad (5.3)$$

where S, B, T, Q, etc., are the numbers of single, binary, triple, quadruple, etc., *systems* (a single system has one star, a binary system contains two stars, etc.). (There are numerous different ways of quantifying the binary fraction, see Reipurth and Zinnecker 1993 for a detailed description of many.)

For massive $>5 M_{\odot}$ stars the binary fraction in the field appears to be unity (i.e., all massive stars are in a binary or multiple system). For solar-type stars the fraction appears to be $\sim 60\%$ (Duquennoy and Mayor 1991), falling to around 40% for M-dwarfs (Fischer and Marcy 1992; see also Lada 2006), and 15% for brown dwarfs (Burgasser et al. 2007). It should be noted though that these numbers are probably lower limits as it is generally very difficult to detect companions with a significantly lower mass than the primary star.

It seems to be impossible to make binaries in the numbers seen by dynamically combining initially single stars (Kroupa and Burkert 2001) and so the *vast* majority of the binary stars must have formed as binaries. In fact, the situation is even more extreme for young stars. **Figure 5-8** shows the fraction of solar-type stars with companions at different separations for the field (solid line), and for a selection of young stars in different regions compiled by Patience et al. (2002). There are clearly far more binaries with separations of around 100 AU amongst young stars than in the field. Indeed, the binary fraction of young solar-type stars appears to be close to 100%.

This suggests that most stars, at least $>1 M_{\odot}$ (the situation is unclear for lower-mass stars), form as binaries.

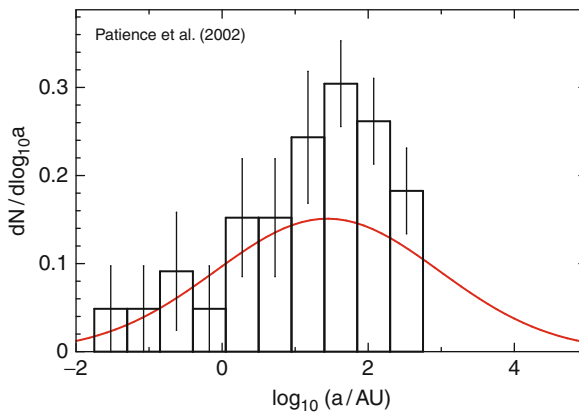


Fig. 5-8

The binary fraction with separation of field G-dwarfs (*red line* from Duquennoy and Mayor 1991), and for young (T Tauri) stars (Compiled by Patience et al. 2002)

3.1.1 Binary Formation

If binary and multiple formation is an extremely common (possibly the major) mode of star formation then it is important to ask how they form. It is thought that multiple systems form by the fragmentation of circumstellar disks during the class 0 (possibly class I) phase of star formation. Prestellar cores are observed to have at least some angular momentum (Goodman et al. 1993), probably due to small (subsonic) levels of turbulence (Burkert and Bodenheimer 2000). Low angular momentum material will form a central protostar, but high angular momentum material will form a circumstellar disk around the protostar. If this disk is massive and cool enough it will be able to fragment, forming a multiple system (see Goodwin et al. 2007 for details).

The Toomre (1964) criteria describes if a disk is unstable to gravitational fragmentation. The Toomre Q -parameter at a radius R in a disk is given by

$$Q(R) = \frac{c(R) \kappa(R)}{\pi G \Sigma(R)}$$

where c is the isothermal sound speed, κ is the epicyclic frequency, and Σ is the surface density, all at radius R . If $Q < 1$, a disk is unstable to gravitational fragmentation (the numerator is a measure of thermal and rotational support against fragmentation, the denominator is a measure of the gravitational attraction of a region).

However, in order to form a new protostar, an unstable fragment must be able to cool on a dynamical timescale (otherwise the fragment will heat as it collapses and bounce); this is known as the Gammie (2001) criterion. Therefore the cooling time t_{cool} must be less than


$$t_{\text{cool}} < \gamma t_{\text{orb}}$$

where t_{orb} is the orbital timescale at whatever radius R , and $0.5 < \gamma < 2$ is the adiabatic index (see Stamatellos and Whitworth 2008 for details).

Whitworth and Stamatellos (2006) use reasonable values for disk parameters and find that these conditions are only met if the radius is greater than a minimum radius R_{min}

$$R_{\text{min}} > 150 \left(\frac{M_{\star}}{M_{\odot}} \right)^{1/3} \text{ AU}$$

where M_{\star} is the mass of the star around which the disk is attempting to fragment. The dependence of the fragmentation radius on the central stellar mass is due to the irradiation of the disk by the central star. The more massive the central star, the higher its luminosity, and so the greater the temperature of the disk.

Interestingly, this value of ~ 150 AU is very close to the peak of the T Tauri binary separation distribution. However, it does suggest that binaries should not *form* within R_{min} . However, inspection of  Fig. 5-8 shows immediately that there are many binary systems with companions at distances $< R_{\text{min}}$. The critical point is that companions should not form below R_{min} , but they may later move to smaller radii due to energy conservation during an ejection if there are more than two bodies in the system (although this should not happen too much, Goodwin and Kroupa 2005), or due to interactions with the disk analogous to planetary migration (e.g., Lin and Papaloizou 1979).

3.1.2 Binary Destruction

Observations and theory both suggest that binaries are the major mode of star formation (at least $>0.5 M_{\odot}$, Goodwin and Kroupa 2005; Lada 2006). This raises the obvious question of why there are fewer old binaries than young binaries. At some point many young binaries with separations of >100 AU must be destroyed in order to produce the old population that observed in the solar neighborhood.

Binaries can be destroyed by dynamical interactions with single stars or other binaries. How easy it is to destroy a binary depends on two factors: the strength of an encounter, and the frequency of encounters which may destroy the binary (see Heggie 1975 and Hills 1975 for a detailed analysis).

The potential of an encounter to destroy a binary depends on the binding energy E of a binary which depends on the masses m_1 and m_2 of the components, and their separation a

$$E = -\frac{Gm_1m_2}{2a}$$

and the energy of the perturbing star which will depend on its typical mass $\langle m \rangle$, and its typical velocity σ (which will be the velocity dispersion of the system). If $|E|/\langle m \rangle \sigma^2 \gg 1$ the binary is said to be “hard” and is difficult to disrupt, if $|E|/\langle m \rangle \sigma^2 \ll 1$ the binary is “soft” and easy to disrupt.

For a typical binary in a fairly typical cluster $\langle m \rangle = m_1 = m_2 = 0.5 M_{\odot}$, and $\sigma = 2 \text{ km s}^{-1}$. This would place the hard-soft boundary at a separation of $a \sim 50$ AU. Therefore encounters would be expected to destroy binaries with separations $\gg 50$ AU, and have no effect on binaries with separations $\ll 50$ AU. In the intermediate regime around 50 AU encounters may or may not destroy binaries (this regime can only be probed by numerical simulations).

However, the encounter rate also plays a crucial role in determining binary survival. The velocity dispersion in the Galactic disk is $\sim 30 \text{ km s}^{-1}$ which gives a hard-soft boundary of only a few AU. But there are many binaries (in fact most) in the field with separations $\gg 10$ – 100 AU. While these binaries are soft they are able to survive because the encounter rate in the field is so low.

The rate R at which a star will have an encounter within a distance b depends on the number density n and velocity dispersion σ of the environment

$$R \sim 30n\sigma b^2(1 + \Omega)$$

where Ω is the Safranov number and is a measure of gravitational focusing which increases the encounter rate. For the situations that are of interest here Ω is generally <1 and may be ignored.

For an encounter with an impact parameter of $b = 500$ AU in the Galactic field with a typical number density of 1 star pc^{-3} and velocity dispersion of $\sim 30 \text{ km s}^{-1}$ ($\sim 30 \text{ pc Myr}^{-1}$), the encounter rate is roughly once every 5.5 Gyr. However, in a cluster with $10^3 \text{ star pc}^{-3}$ and $\sigma \sim 2 \text{ km s}^{-1}$, the encounter rate is once every 5.5 Myr. Therefore encounters that would be expected to destroy binaries with separations greater than about 100 AU are very rare in the field, but are expected to have occurred in even very young star clusters.

3.2 Binaries or Singles?

That so many stars form in binaries shows that binary formation is a major mode of stars formation. But is it the main mode of star formation? Most stars (around 90%) are M-dwarfs, and the binary fraction of M-dwarfs is low. A binary fraction of only 40% suggests that most M-dwarfs are single. However, the binary fraction is for *systems*. Of every 100 M-dwarf systems 40 are binary systems, and 60 are single. Of the 40 binary systems, each contains 2 M-dwarfs, meaning that by number 80/140 M-dwarfs are in a binary system, and 60/140 are singles. This now suggests that most M-dwarfs are in binary systems. However, the 40 binary systems presumably formed as binary systems and so only 40% of low-mass stars are known to form as binaries, while 60% are *currently* singles. If all the single field M-dwarfs formed as singles this tells us that single star formation is the major mode of star formation, but if many formed as binaries which were later destroyed then binary formation is the major mode of star formation. Sadly, observations of low-mass binary fractions in clusters cannot currently distinguish these possibilities.

3.3 Star Clusters

Around 75–90% of stars form in clusters of hundreds to millions of stars (Lada and Lada 2003). A particularly interesting observation is that the mass function of these star clusters appears to be $N(M) \propto M^{-2}$ (Lada 2010; Lada and Lada 2003). This is interesting because it means an equal mass of stars forms in each equal logarithmic mass interval. Therefore there are as many star forms in Taurus-like associations with mass $\sim 10^2 M_\odot$, as in Orion-like clusters with mass $\sim 10^3 M_\odot$, as in massive starburst clusters like Westerlund 1 with a mass of $\sim 10^5 M_\odot$. Therefore all masses of clusters are equally important, and there is no such thing as an “average” or “typical” star cluster.

Our mental image of star clusters is set by pictures of old globular clusters, which are remarkably spherical, smooth, and dynamically relaxed. This is not, however, how star clusters form. MCs are complex, hierarchical structures, and they form stars in the same complex structures (see ● Fig. 5-2). Star clusters also appear to form highly out-of-equilibrium, with velocities well below what would be expected for virial equilibrium (see Allison et al. 2009 and references therein).

Any clumpy, out-of-equilibrium system will rapidly attempt to reach equilibrium and it appears that star clusters rapidly collapse and smooth-out their initial clumpiness. This has two important consequences when attempting to interpret observations of clusters.

Firstly, clusters change their appearance rapidly and can appear very different on timescales of just 10^5 years. While this is extremely rapid on astrophysical timescales, it is somewhat longer than the average baseline of our astronomical observations. Therefore only a single snapshot in the evolution of a star cluster is observed. Two star clusters of different ages that appear very different might well just be very similar objects at different points in their evolution. Equally, two objects that appear very similar might be seen at different points in their evolution and will result in very different objects. In addition, the evolution of star clusters occurs far faster than our ability to accurately date star clusters (at best to about a Myr), so it is impossible to tell if two young clusters are really the same or of significantly different ages.

Secondly, our observations beyond ~ 100 pc are only two-dimensional projections of complex, three-dimensional objects. A clumpy, hierarchical cluster can look very different from

different viewing angles. This should improve with the advent of GAIA which will provide accurate distances to many objects (the 10^9 brightest objects in the sky) and also proper motions to give us detailed kinematic information. However, GAIA will not be able to observe the earliest embedded stages of star cluster formation and evolution during which much of the rapid evolution occurs.

Clusters that are initially dynamically cool and clumpy will attempt to reach an equilibrium which is virialized and smooth. Following Allison et al. (2009), if the initial virial ratio of a cluster is $Q_i = -T_i/\Omega_i$, where T_i is the initial kinetic energy and Ω_i is the initial potential energy (and $Q_i = 0.5$ is virial equilibrium), then the total energy of a cluster is

$$E = T_i + \Omega_i = (1 - Q_i)\Omega_i$$

the initial potential energy is

$$\Omega_i = -\eta_i \frac{GM^2}{R_i}$$

where M is the mass of the cluster, R_i is a characteristic radius, and η_i is a structure parameter. The cluster will virialize and erase its substructure resulting in a cluster where

$$E = \frac{\Omega_f}{2} = -\eta_f \frac{GM^2}{2R_f}$$

with a new distribution and so a new structure parameter η_f , and a new radius R_f . The degree to which the cluster has collapsed is

$$\frac{R_i}{R_f} = \frac{\eta_i}{\eta_f} 2(1 - Q_i)$$

Typical values for these parameters are an initial virial ratio of $Q_i \sim 0.3$, an initial clumpy structure parameter of $\eta_i = 1.5$, and a final smooth structure parameter $\eta_f = 0.75$ (clumpy distributions have a higher η as clumps contain more potential energy than if they were smoothed-out). This means that clusters will typically collapse by a factor of about 2.5. After collapse clusters will tend to “bounce,” rapidly increasing their radii (Allison et al. 2009).

An interesting consequence of this collapse and bounce is that the dynamical timescales of clusters will change significantly during their early life. The crossing time/dynamical time⁴ of a cluster can be calculated if the size R and the velocity dispersion σ of a cluster are known

$$t_{\text{cross}} = \frac{R}{\sigma}$$

The crossing time gives the shortest timescale on which the cluster can be expected to change, in particular the timescale on which dynamical interactions and the destruction of binaries will occur. From the crossing time the two-body relaxation time of the cluster can also be calculated, which is the timescale on which the velocities of stars will significantly change.

The size of a cluster is relatively simple to observe, but the velocity dispersion of a cluster is often unknown. To observe the velocity dispersion of a cluster either detailed spectroscopy of many stars (which may be contaminated by binary motions) or long-baseline proper motion determinations are required. Very often these are not available and so the velocity dispersion is estimated by assuming that the cluster is virialized. This is probably a very poor approximation

⁴For our purposes the terms are interchangeable.

in young clusters as it is found that when the actual velocity dispersion is known, many clusters appear to be out of virial equilibrium (Goodwin and Bastian 2006).


However, determinations of the crossing times of clusters are *instantaneous* values. If a cluster has undergone a collapse and a re-expansion, then R may have changed by a factor of several, and σ may well have changed as well (although probably less than R) and therefore the cluster may be many more crossing times older than the instantaneous value might suggest, and so the cluster may be far more dynamically evolved than one might think.

As an example, the current half-mass radius of Orion is 0.8 pc, and its mass is about $10^3 M_{\odot}$. If Orion were virialized this would suggest a velocity dispersion of $\sigma \sim 2 \text{ km s}^{-1}$, and a crossing time of about a Myr giving a dynamical age of Orion of 2 or 3 crossing times. However, Orion is not virialized, and its velocity dispersion is actually observed to be $\sigma \sim 4 \text{ km s}^{-1}$, implying a crossing time of only half the virialized value. But, if the velocity dispersion of Orion is significantly super-virial, then Orion must be expanding, and so its size must have been significantly smaller in the past as well. Therefore Orion is probably tens of crossing times old, rather than the 2–3 implied by assuming it is virialized, or the 4–6 from using the actual velocity dispersion, but only the current (larger) size.

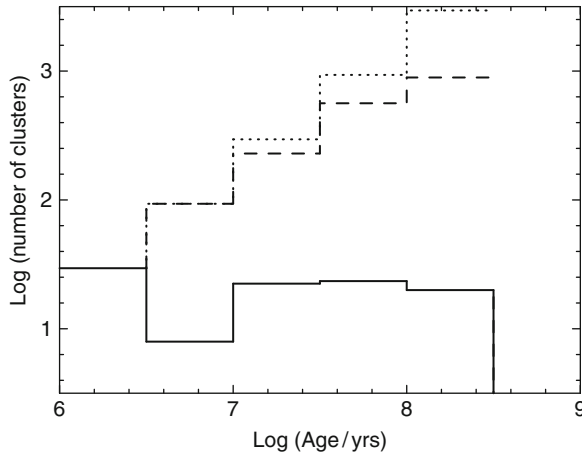
3.4 The End of Star Formation

Star formation ends when there is no (dense molecular) gas remaining from which to form stars. It is thought that the feedback of energy from massive stars – winds, radiation, and eventually supernovae – stop star formation by heating and then removing gas from the star forming region. Massive stars begin to feedback energy while they are still forming, and so the formation of the first massive star(s) marks the beginning of the end of star formation.

Stars $>30\text{--}40 M_{\odot}$ will feedback around 10^{47} J of energy in winds and UV radiation during their lifetimes, and another 10^{46} J when they become supernovae. Their prodigious energy output ionizes the gas around the stars preventing it from forming new stars and expels the gas from around them. But the expulsion of gas from a central cluster of massive stars can also trigger star formation in surrounding gas that is dense enough, and/or far enough to escape being ionized. Observations of young star clusters find that there are no star clusters older than 3–5 Myr which have gas associated with them (e.g., Lada and Lada 2003) suggesting that gas loss occurs at around this age.

It is also found that there are far fewer star clusters older than 10 Myr than would be expected from the number of young clusters we see (Lada and Lada 2003).  Figure 5-9 shows the observed numbers of clusters within 2 kpc of the Sun of different ages (solid line) against the expected numbers of clusters for a constant cluster formation rate (dotted line), and the expected number corrected for luminosity evolution (dashed line). If the cluster formation rate were constant, each equal logarithmic age bin should contain increasing numbers of clusters (as it covers a larger span of linear time).

The observations of clusters have some problems. Firstly, aging the youngest clusters is rather problematic, and so the differences between the first two bins should not be taken too seriously. Secondly, less massive clusters are expected to dissolve due to internal two-body effects and so cluster numbers will fall with time. However, the lack of older clusters compared to younger clusters is dramatic and inexplicable by either internal dynamics or errors in the observations. This tells us one of two things: (a) the cluster formation rate has not been constant over the past few hundred Myr, or (b) many (most) clusters are destroyed.



■ Fig. 5-9

The *solid line* shows the observed number of clusters with age within 2 kpc of the Sun. The *dotted line* shows the expected number of clusters if the cluster formation rate has been constant over this period and no clusters have been destroyed, the *dashed line* shows this number corrected for luminosity evolution (Adapted from Lada and Lada 2003)

The favored interpretation is that clusters are destroyed. There is no evidence that the current cluster formation rate is vastly (factors of 10) higher than in the recent past. In particular, there is absolutely no evidence that the cluster formation rate has vastly increased in the past 10–20 Myr as would be needed to explain the difference between the model and the second and third age bins.

Clusters are thought to be destroyed by two mechanisms. Most (up to 90%) clusters that form are thought to be destroyed by the rapid expulsion of gas by massive stars. And many of those clusters that do survive are destroyed by evaporating.

3.4.1 Gas Expulsion

As noted before, the star formation efficiencies of star clusters are low, only 10–30%. Therefore star clusters at the point at which massive stars begin to remove gas are gravitationally dominated by that gas. Therefore, the removal of the gas removes the largest contributor to the binding energy of the cluster (see Goodwin 2009 and references therein).

Assuming that a cluster of total mass M has formed stars with an efficiency ϵ , the stellar mass is ϵM and the gas mass is $(1 - \epsilon)M$. For a cluster of size R and velocity dispersion σ the initial energy E_i is

$$E_i = \frac{1}{2} M \sigma^2 - \frac{GM^2}{R}$$

and if it is in virial equilibrium then

$$E_i = -\frac{GM^2}{2R}.$$

If all of the gas is expelled instantaneously (in this case, this means in less than a crossing time), then R and σ for the stars do not change, but the gas mass is removed meaning that the final energy E_f is

$$E_f = \frac{1}{2} \epsilon M \sigma^2 - \epsilon^2 \frac{GM^2}{R}$$

It is easy to show that if $\epsilon < 0.5$, then the cluster is now unbound and has been destroyed by the gas expulsion.

In practice, gas expulsion is rather more complex than this as gas expulsion can be adiabatic rather than instantaneous (i.e., it takes longer than a crossing time) which somewhat alleviates its effects, and also even if it is instantaneous a much smaller bound core may be retained. But even taking these effects into account, star clusters cannot survive gas expulsion if their star formation efficiencies are less than 20–30% (Goodwin and Bastian 2006). Such star formation efficiencies are at the upper end of those observed, and so it would be expected that many star clusters would be destroyed by gas expulsion after only a few Myr in keeping with the observations of cluster numbers with age.

3.4.2 Dynamical Evolution

Even if a star cluster is able to survive gas expulsion, many will not survive for long. Two-body interactions between stars alter the velocity dispersion of the stars in a cluster driving it toward a Maxwellian. Stars in the tail of the Maxwellian with greater than the escape velocity of the cluster are able to escape, which lowers the mass of the cluster, which lowers the escape velocity, which makes escape easier. Thus a cluster will evaporate over time.

The timescale over which the velocity distribution changes significantly (and a Maxwellian is established) is the relaxation time t_{relax} given by

$$t_{\text{relax}} \sim \frac{N}{8 \ln N} t_{\text{cross}}$$

where N is the number of stars in a cluster (see Binney and Tremaine 2008 for a derivation), and clusters can survive for roughly ten relaxation times before dissolving, although this depends strongly on the strength of the external tidal field (see Spitzer 1987). Lamers et al. (2005) provide a semi-analytic formulation of the survival time of a star cluster in a galaxy.

Therefore, low-mass (low- N) clusters will not survive for long, even if they do manage to survive gas expulsion. It is no surprise that the only extremely old clusters that are observed are very massive globular clusters, as they are the only clusters that *could* have survived for a Hubble Time, even if many low-mass clusters also formed 12 Gyr ago.

3.5 Is Star Formation Universal?

The final question is that of the universality of star formation. Do all stars form in roughly the same way, or does the environment in which stars form play a crucial role? Are the stars that form in Taurus-like associations the same as those that form in very massive clusters? And do all clusters of the same mass form stars in the same way? While this might seem like a fairly simple question, the answer is far from clear.

The form of the IMF might well be expected to vary with environment and thus provide important clues as to the universality or otherwise of star formation. However, the IMF is surprisingly invariant (Bastian et al. 2010). It is unknown why the IMF is universal, but it might suggest that star formation is everywhere the same.

Are there any other clues as to the universality of star formation? Many (most?) stars form as binary and multiple systems and these provide additional information on the star formation process in a region. For example, if one region produces many wide binaries with very different masses, while another region produces fewer binaries which are all close equal-mass systems, then star formation in those two regions must have been very different, even if the IMFs are the same (Goodwin and Kouwenhoven 2009).

Observationally the binary fractions and wide binary populations of different star forming regions *are* different. For example, Taurus has a binary fraction of almost unity (Leinert et al. 1993; Patience et al. 2002), while the binary fraction of Orion is more like the field with very few wide binaries (Reipurth et al. 2007; Scally et al. 1999).

So, do different binary populations imply different star formation? The answer is maybe, but maybe not. Taurus is dynamically young and has little time to process its initial binary population so it should be close to its birth population. However, Orion is dynamically old and so the birth population may have been significantly altered (Parker et al. 2009). It is possible to model the *current* binary population of Orion as initially Taurus-like and evolved, or as initially different and less evolved (see Goodwin 2010).

Therefore the crucial question of the universality of star formation is unclear. Different regions may form stars in very different ways, or it may always be basically the same.

4 Conclusions

How stars form is an extremely complex problem. Extremely good progress has been made observationally and theoretically over the past few years, but even some quite basic issues in star formation are far from being understood.

The IMF appears universal, but its origin and why it is universal are unknown. The basics of low-mass star formation seem relatively well-understood, but initial binary fractions and distributions and the origin of binarity are unclear. High-mass star formation is very uncertain, and it is not understood how high-mass stars are able to overcome feedback to accrete to masses as high as are observed.

While the lack of understanding of many basic aspects of star formation is rather annoying, it does mean that star formation is an area that has a vast potential for new and exciting science in the coming decades.

References

- | | |
|--|---|
| <p>Allison, R. J., Goodwin, S. P., Parker, R. J., de Grijs, R., Portegies Zwart, S. F., & Kouwenhoven, M. B. N. 2009, <i>ApJ</i>, 700, L99</p> <p>Alves, J., Lombardi, M., & Lada, C. J. 2007, <i>A&A</i>, 492, 17</p> <p>Ballesteros-Paredes, J. 2006, <i>MNRAS</i>, 372, 443</p> | <p>Ballesteros-Paredes, J., Klessen, R. S., Mac Low, M.-M., & Vazquez-Semadeni, E. 2007, in <i>Protostars and Planets V</i>, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 63</p> <p>Bastian, N., Covey, K. R., & Meyer, M. R. 2010, <i>ARAA</i>, 48, 339</p> |
|--|---|

- Binney, J., & Tremaine, S. 2008, "Galactic Dynamics: Second Edition" (Princeton: Princeton University Press)
- Boldyrev, S. 2002, *ApJ*, 569, 841
- Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 1997, *MNRAS*, 285, 201
- Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 2001, *MNRAS*, 323, 785
- Bonnell, I. A., Larson, R. B., & Zinnecker, H. 2007, in *Protostars and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 149
- Burgasser, A. J., Reid, I. N., Siegler, N., Close, L., Allen, P., Lowrance, P., & Gizis, J. 2007, in *Protostars and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 427
- Burkert, A., & Bodenheimer, P. 2000, *ApJ*, 543, 822
- Burningham, B., Naylor, T., Littlefair, S. P., & Jeffries, R. D. 2005, *MNRAS*, 363, 1389
- Clark, P. C., Klessen, R. S., & Bonnell, I. A. 2007, *MNRAS*, 379, 57
- Chabrier, G. 2003, *PASP*, 115, 763
- Duquennoy, A., & Mayor, M. 1991, *A&A*, 248, 485
- Elmegreen, B. G. 2000, *ApJ*, 530, 277
- Elmegreen, B. G. 2002, *ApJ*, 577, 206
- Evans, N. J., II 1999, *ARAA*, 37, 311
- Evans, N. J., Dunham, M. M., et al. 2009, *ApJS*, 181, 321
- Fischer, D. A., & Marcy, G. W. 1992, *AJ*, 396, 178
- Gammie, C. F. 2001, *ApJ*, 55, 174
- Goodwin, S. P. 2009, *AP&SS*, 324, 259
- Goodwin, S. P. 2010, *RSPTA*, 368, 851
- Goodwin, S. P., & Bastian, N. 2006, *MNRAS*, 373, 752
- Goodwin, S. P., & Kounenhoven, M. B. N. 2009, *MNRAS*, 397, L36
- Goodwin, S. P., & Kroupa, P. 2005, *A&A*, 439, 565
- Goodwin, S. P., & Whitworth A. 2007, *A&A*, 466, 943
- Goodman, A. A., Benson, P. J., Fuller, G. A., & Myers, P. C. 1993, *ApJ*, 406, 528
- Goodwin, S. P., Kroupa, P., Goodman, A. A., & Burkert, A. 2007, in *Protostars and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 133
- Goodwin, S. P., Nutter, D., Kroupa, P., Ward-Thompson, D., & Whitworth, A. P. 2008, *A&A*, 477, 823
- Heggie, D. C. 1975, 173, 729
- Hennebelle, P., & Chabrier, G. 2008, *ApJ*, 684, 395
- Hills, J. G. 1975, *AJ*, 80, 809
- Kirk, J. M., Ward-Thompson, D., & André, P. 2005, *MNRAS*, 360, 1506
- Kroupa, P. 2002, *Science*, 295, 82
- Kroupa, P., & Burkert, A. 2001, *ApJ*, 555, 945
- Lada, C. J. 1999, in *The Origin of Stars and Planetary Systems*, eds. C. J. Lada & N. D. Kylafis (Dordrecht: Kluwer), 143
- Lada, C. J. 2006, *ApJ*, 640, L63
- Lada, C. J. 2010, *Philos Trans R Soc A*, 368, 713
- Lada, C. J., & Lada, E. A. 2003, *ARAA*, 41, 57
- Lamers, H. J. G. L. M., Gieles, M., Bastian, N., Baumgardt, H., Kharchenko, N. V., & Portegies Zwart, S. 2005, *A&A*, 441, 117
- Larson, R. B. 1969, *MNRAS*, 145, 271
- Larson, R. B. 1981, *MNRAS*, 194, 809
- Leinert, Ch., Zinnecker, H., Weitzel, N., Christou, J., Ridgway, S. T., Jameson, R., Haas, M., & Lenzen, R. 1993, *A&A*, 278, 129
- Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, *AJ*, 136, 2782
- Lin, D. N. C., & Papaloizou, J. 1979, *MNRAS*, 188, 191
- Luhman, K. L. 2004, *ApJ*, 617, L1216
- Mac Low, M.-M., & Klessen, R. S. 2004, *Rev Mod Phys*, 76, 125
- McKee, C. F., & Ostriker, E. C. 2007, *ARAA*, 45, 565
- McKee, C. F., & Tan, J. C. 2003, *ApJ*, 585, 850
- Masunaga, H., & Inutsuka, S. 2000, *ApJ*, 531, 350
- Mayne, N. J., & Naylor, T. 2008, *MNRAS*, 386, 261
- Mouschovias, T. Ch, Tassis, K., & Kunz, M. W. 2006, *ApJ*, 646
- Naylor, T. 2009, *MNRAS*, 399, 432
- Nutter, D., & Ward-Thompson, D. 2007, *MNRAS*, 374, 1413
- Padoan, P., & Nordlund, Å. 2002, *ApJ*, 576, 870
- Palla, F., Randich, S., Flaccomio, E., & Pallavicini, R. 2005, *ApJ*, 626, 49
- Parker, R. J., Goodwin, S. P., Kroupa, P., & Kounenhoven, M. B. N. 2009, *MNRAS*, 397, 1577
- Parker, R. J., Bouvier, J., Goodwin, S. P., Moraux, E., Allison, R. J., Guieu, S., & Güdel, M. 2011, *MNRAS*, 412, 2489
- Patience, J., Ghez, A. M., Reid, I. N., & Matthews, K. 2002, *AJ*, 123, 1570
- Reipurth, B., & Clarke, C. J. 2001, *AJ*, 122, 432
- Reipurth, B., & Zinnecker, H. 1993, *A&A*, 278, 81
- Reipurth, B., Guimarães, M. M., Connelley, M. S., & Bally, J. 2007, *AJ*, 134, 2272
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sclally, A., Clarke, C. J., & McCaughrean, M. J. 1999, *MNRAS*, 306, 253
- Smith, R. J., Clark, P. C., & Bonnell, I. A. 2008, *MNRAS*, 391, 1091
- Spitzer, L., Jr. 1987, *Dynamical Evolution of Globular Clusters* (Princeton: Princeton University Press)
- Stacy, A., Greif, T. H., & Bromm, V. 2011, *MNRAS*, *ApJ*, 730, L1
- Stahler, S. W., & Palla, F. 2005, *The Formation of Stars* (Berlin: Wiley-VCH)

- Stamatellos, D., & Whitworth, A. P. 2008, *A&A*, 48, 879
- Stamatellos, D., Hubber, D. A., & Whitworth, A. P. 2007, *MNRAS*, 382, 30
- Tan, J. C., Krumholz, M. R., & McKee, C. F. 2006, *ApJ*, 641, 121
- Toomre, A. 1964, *ApJ*, 139, 1217
- Vázquez-Semadeni, E., Gómez, G. C., Jappsen, A. K., Ballesteros-Paredes, J., González, R. F., & Klessen, R. S. 2007, *ApJ*, 657, 870
- Vesperini, E., & Heggie, D. C. 1997, *MNRAS*, 289, 898
- Whitworth, A. P., & Stamatellos, D. 2006, *A&A*, 458, 817
- Whitworth, A., Bate, M. R., Nordlund, Å, Reipurth, B., & Zinnecker, H. 2007 in *Protostars and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 459
- Williams, J. P., Blitz, L., & McKee, C. F. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S.S. Russell (Tucson: University of Arizona Press), 97
- Zinnecker, H., & Yorke, H. W. 2007, *ARAAS*, 45, 481

6 Young Stellar Objects and Protostellar Disks

Ana Inés Gómez de Castro

Fac. de CC Matematicas, Universidad Complutense de Madrid,
Madrid, Spain

1	<i>Introduction</i>	280
2	<i>PMS Evolution: The Star</i>	281
3	<i>PMS Evolution: The Disk</i>	288
3.1	Morphology of Protostellar Disks	290
3.1.1	The Inner Cavity Filled in with Hot Gas	290
3.1.2	Disks Are Flared	292
3.1.3	Dusty Atmospheres and Warps	292
3.2	Disk Mass	294
3.3	Composition of Protostellar Disks	295
3.3.1	Dust	295
3.3.2	Molecular Gas	296
3.3.3	Disk Chemistry	297
3.4	Disk Formation	298
3.5	Protostellar Disks and the α -Prescription Paradigm	300
3.5.1	The α -Prescription and Angular Momentum Transport by Gravitational Waves	303
3.6	Disk Structure: The Effect of Stellar Irradiation	303
3.6.1	YSOs' X-ray and UV Radiation Field	305
3.6.2	The Vertical Structure of Disks	306
3.7	Disk Stability	310
3.8	Disk Evolution	310
4	<i>PMS Evolution: The Outflow</i>	312
4.1	First Steps to a Mechanism for Outflow Generation: Disk Winds	318
4.2	The Star–Disk Interface: The Jet Engine	320
4.3	The Stellar Magnetosphere	326
4.3.1	Accretion Shocks and Funnels	327
4.3.2	General Magnetospheric Properties	327
5	<i>Some Final Thoughts and Conclusions</i>	328
	<i>Acknowledgments</i>	331
	<i>Appendix A: The Gould's Belt</i>	331
	<i>References</i>	331

Abstract: Stars form out of molecular clouds with densities of 10^4 cm^{-3} and temperatures of 10 K. Clouds fragment into molecular cores covering a broad range of masses that give rise to the formation stellar associations, clusters and single stars. High mass star formation rapidly removes the environmental gas and dust having a large impact into cloud evolution. Low mass star formation is less aggressive with the environmental cloud. Young stellar objects (YSOs) display a wide range of properties depending on their mass, age and star forming environment. Very massive YSOs generate ultracompact HII regions embedded in the cores of molecular clouds. Their key observational properties are related with the action of strong and highly photoionizing radiation fields in the surrounding molecular cloud. They evolve in time scales of few Myrs into the main sequence. Low mass YSOs progress to the main sequence at a slower pace; the gravitational collapse is slowed down by the action of the molecular cloud magnetic field and their most conspicuous observational properties are related with dissipation of the magnetic energy carried out during star formation and enhanced by shear during the star building up.

Protostellar disks are a by-product of star formation. They are formed as angular momentum reservoirs during the gravitational accretion process that gives rise to the formation of new stars. They are observed around all types of YSOs however their properties and evolution depend on the nature of the parent star/stars. The opacity of the disk and its selfshielding properties against the stellar and the environmental radiation field determine the lifetime of protostellar disks that end as debris disks and planetary systems. Disks play a key role in the generation of the powerful bipolar outflows observed during star formation. Bipolar outflows are observed from the very early stages (10^4 years) of star formation and from all the mass spectrum of YSOs.

This chapter attempts to summarize our current understanding of single star formation and the joint evolution of the star-disk system towards the main sequence.

Keywords: stars: formation; stars: pre-main-sequence, stars: variables: T Tauri, Herbig Ae/Be, stars: winds, outflows, ISM: Herbig-Haro objects, ISM: jets and outflows, accretion, accretion disks



List of Abbreviations: *HAeBe stars*; Herbig AeBe stars; *HHOs*, Herbig-Haro Objects; *ISM*, InterStellar Medium; *PDRs* PhotoDissociation Regions; *PMS*, Pre-main sequence; *TTSs*, T Tauri Stars; *YSOs*, Young Stellar Objects

1 Introduction

Stars form out of dense gas in molecular clouds. Clouds fragment into molecular cores that afterward contract gravitationally to form either single or multiple stellar systems. Fragmentation requires rapid cooling, faster than the relevant dynamical timescales of the system, and depends on the physical properties and chemical composition of matter as described in detail in ► [Chap. 4](#). Most of the stars are formed in multiple systems where the angular momentum of the original cloud is stored in a coupled spin-orbit system with a fraction of the angular momentum going into the orbits of fragments and another fraction into the collapsing fragments themselves. During star formation, the specific angular momentum decreases several orders of magnitude from $j \approx 10^{23} \text{ cm}^2 \text{ s}^{-1}$ in molecular cloud material (scales of about 1 pc) to $j \approx 10^{21} \text{ cm}^2 \text{ s}^{-1}$ in molecular cores (scales of about 0.1 pc) to reach finally $j \approx 10^{17} \text{ cm}^2 \text{ s}^{-1}$ in the young stars entering the main sequence. During this process angular momentum conservation leads to the formation of angular momentum reservoirs, the protostellar disks, that grow as the stellar embryo does it and that have their own evolutionary scales.

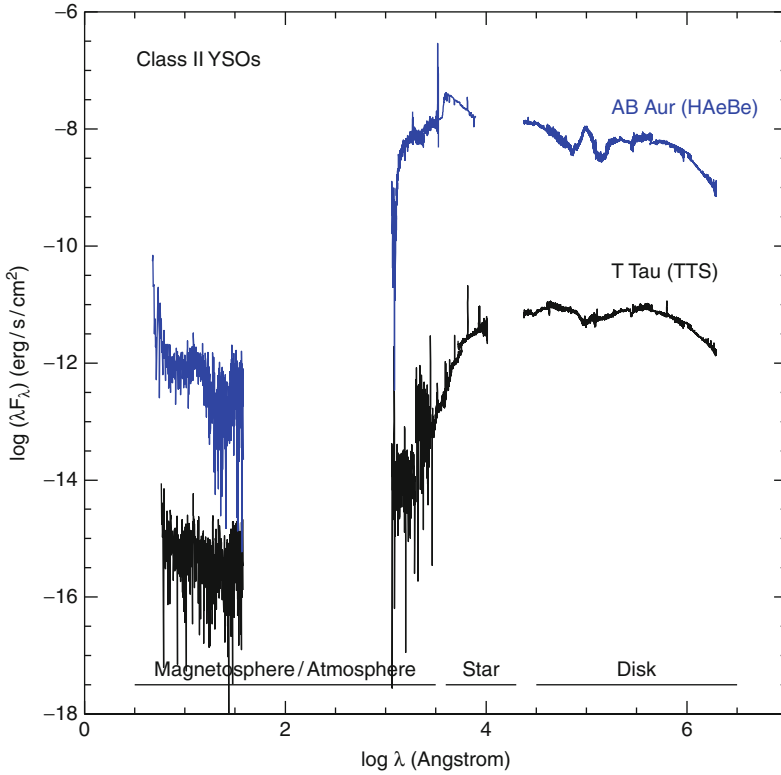
The term Young Stellar Object (YSO) is introduced to define astronomical sources in this evolutionary phase from contracting objects within opaque dust cocoons to pre-main sequence (PMS) stars close to the zero age main sequence. Three major evolutionary phases can be defined in the formation of stars:

- *Phase I: Protostellar core evolution:* The evolution of gravitationally bounded molecular cores into protostellar cores. The evolution timescale is controlled by the turbulence field in the core.
- *Phase II: Main accretion phase:* The main accretion phase begins when turbulence support is overcome and non-homologous gravitational collapse produces a differentiation between the stellar embryo and the surrounding envelope. At the end of this phase, a central source that contains about half of the mass of the core is formed. Simultaneously a large protostellar disk is formed. During this evolutionary phase YSOs are denoted as class 0 YSOs (see Froebrich 2005 for a catalog). The stellar birthline is defined by the end of this phase.
- *Phase III: PMS stellar evolution and planetary disk formation:* Hydrostatic equilibrium evolution from the stellar birthline to the main sequence. At the same time, the protostellar disk evolves from a massive structure channeling the accretion flow into a passive planetary disk. During this process the disk luminosity varies several orders of magnitude. At the first stages, the stellar atmosphere is occulted by the heavy excess produced by the accretion flow and the remnant extinction of the envelope. As these decrease, the stellar atmosphere becomes recognizable. Finally, the young planetary disk is formed and the stellar evolution is decoupled from the disk evolution. Objects in these three phases are named as Class I, II, and III YSOs, respectively. These three classes trace mainly the evolutionary state of the disk (see Herbig and Bell 1988 for a catalog).

This chapter is aimed at describing the last two evolutionary steps: the properties and evolution of protostellar disks and their interaction with the growing star. This interaction is a rich source of physics being the most conspicuous in the creation of gravitational engines able to accelerate part of the accreting matter and eject it forming large bipolar outflows. The physics of pre-main sequence evolution is very complex and competing simple scenarios have been developed to tackle the different aspects of the problem. For this reason, this chapter has been divided into three sections dealing with the evolution of the three main components of YSOs: star, disk, and outflow. This division mimics the physical systems involved and the main areas of theoretical work; however, it is worth to remark that these three components are strongly coupled through gravity, radiation and magnetic fields. In addition, the identification of the contribution from each component to the observed radiation is not always straightforward leading to uncertainties in the interpretation of the data within a given theoretical framework. The spectrum of the prototypes to low mass and intermediate mass PMS stars, T Tau and AB Aur, respectively, are displayed in  Fig. 6-1; the physical components that dominate the radiative output at each wavelength are indicated. A sketch of the structure of the inner ~ 1 AU of low mass YSOs is outlined in  Fig. 6-2.

2 PMS Evolution: The Star

YSO precursors are gravitationally bounded protostellar cores evolving in magnetized molecular clouds. Protostellar cores' lifetimes have been estimated to be $3 - 5 \times 10^5$ years (Kirk et al. 2005; Onishi et al. 2002) which are about a factor of 2–3 larger than their



■ Fig. 6-1

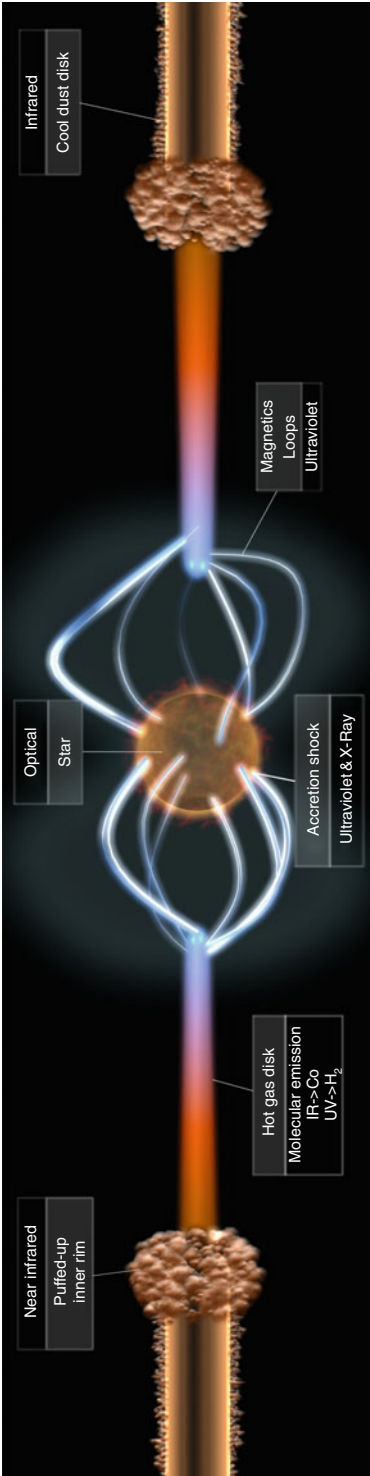
Spectral energy distribution of T Tau (*black*) and AB Aur (*blue*). The main contributors to the flux in each spectral range are marked at the bottom. T Tau flux is reduced by two orders of magnitude (-2 in the Y-axis scale) for the plot. AB Aur and T Tau are prototypes to class II YSOs of intermediate ($2M_{\odot} < M_{*} < 10M_{\odot}$) and low ($M_{*} < 2M_{\odot}$) mass, respectively

free-fall times¹ ($1.1 - 1.6 \times 10^5$ years) suggesting some degree of support during this phase. The support is most likely provided by turbulence. Molecular line emission from protostellar cores is very broad, typically $1-4 \text{ km s}^{-1}$, thus highly suprathermal for the low temperatures of the cores.² As magnetic fields pervade the InterStellar Medium (ISM) HydroMagnetic (HM) turbulence is a natural source of support. In fact, the line broadening after subtraction of the thermal component scales with the cores' mean magnetic field (Myers and Goodman 1988). Molecular cores are weakly ionized³ thus a strong coupling between ions and neutrals is required to

¹The free-fall time for a spherical core of uniform density ρ_c is $t_{ff} = (3\pi/32G\rho_c)^{1/2} = 1.08 \times 10^5 \text{ year} (n_c/10^5 \text{ cm}^{-3})^{-1/2}$, where n_c is the particle density of the core.

²Molecular cores' temperature is about 10 K and sound speed is $0.19\sqrt{T/10 \text{ K}} \text{ km s}^{-1}$.

³The high density of molecular clouds shield the cores against the environmental UV radiation. The major source of ionization is cosmic rays that produce an ionization fraction of $\chi_e \approx 4.5 \times 10^{-6} n_H^{-1/2} (\text{cm}^{-3})$ (McKee 1989).



■ Fig. 6-2

Sketch of the inner structure of a low mass ($<2M_{\odot}$) YSO. The spectral ranges where the bulk of the energy of each component is radiated are indicated. The inner border of the disk is the boundary with the stellar magnetosphere. Another important boundary is set by the dust melting point that separates the two thermal regimes of the disk: gas dominated versus dust dominated

support cores against collapse by this mean. The ion-neutral collision frequency is

$$\nu_{n,i} \sim 10^{-9} \text{ s}^{-1} \frac{\langle V \rangle}{2 \text{ km s}^{-1}} n_i (\text{cm}^{-3}) = \gamma_{n,i} n_i (\text{cm}^{-3})$$

where n_i is the ions' particle density in the core and $\langle V \rangle$ is the characteristic velocity of the gas motions and $\gamma_{n,i}$ the rate coefficient for ion-neutral collisions (Braginskii 1965). This frequency defines the ambipolar diffusion threshold. HM waves with frequencies higher than $\nu_{n,i}$ only operate on a small fraction of the mass of the core, the ionized component, and cannot support the core against collapse. HM waves with frequencies shorter than $\nu_{n,i}$ are able to support molecular cores against collapse. HM waves with frequencies $\sim \nu_{n,i}$ dissipate by the neutrals' drag force heating the cores by ambipolar diffusion. The ambipolar diffusion timescale or the timescale for momentum transfer between ions and neutrals in molecular clouds was derived by Arons and Max (1975) to be,

$$\tau_{AD} = 1.6 \times 10^5 \text{ year } \mu_{n,i} \left(\frac{P}{10^5 \text{ year}} \right)^2 \frac{x_i}{10^{-6}} \frac{\gamma_{n,i}}{10^{-9}} \frac{n(H_2)}{10^4 \text{ cm}^{-3}}$$

with $\mu_{n,i} = m_i / (m_i + m_n)$, $n(H_2)$ the density of the molecular gas, χ_i the ionization fraction (n_i / n_n), and P the characteristic timescale of the propagating mode. τ_{AD} depends on the ionization fraction of the cores and thus on the strength and penetration of high-energy (ultraviolet, X-ray) radiation. This timescale is comparable to the protostellar cores' lifetime determined from observations confirming that pseudo-hydrostatic contraction is a good approach to model the first evolutionary stages of star formation. For this reason, Bonnor–Ebert spheres are often taken as the departure point for dynamical calculations of the gravitational collapse from core to class 0 YSOs.

Hydrostatic contraction of self-gravitating spheres was first studied by Bonnor, Ebert, and McCrea in the late 1950s. Virial equilibrium in a self-gravitating molecular cloud leads to a modified equation of state of the form $pV = (M / \mu m_H) k_B T + (1/3) \epsilon_g$, where p , V , M , μm_H , and T are the cloud pressure, volume, mass, mean molecular weight, and temperature, respectively; ϵ_g is the potential energy of the self-gravitating cloud, and k_B the Boltzmann constant. The hydrostatic evolution of the gas leads to a modified Emden equation whose solution departs from the isothermal ideal gas law, that is, gravity prevents the volume to decrease smoothly as pressure rises; instead, there is a critical volume (radius) above which self-gravitation drives the object contraction. This critical radius is given by $r_c = 0.06 \text{ pc} \times (T/10 \text{ K})^{1/2} (n/10^4 \text{ cm}^{-3})^{-1/2}$ where T and n are the temperature and particle density of the self-gravitating cloud (see Bonnor 1957 for a discussion on the relation between this critical length and the Jeans length). A critical mass can also be defined, the so-called Bonnor–Ebert mass, given by $M_{B-E} = 1.21 M_\odot (T/10 \text{ K})^{3/2} (n/10^4 \text{ cm}^{-3})^{-1/2}$.

YSOs begin their life as the dynamical collapse begins. Gravitational infall is nonhomologous regardless of the choice of initial or boundary conditions; a stellar embryo rapidly develops in the center of the core; free fall depends inversely on the density and the collapse proceeds more rapidly in the center where the density is the highest. The density profile adopts the form outlined in ► Fig. 6-3 with a central plateau of uniform density (the rapidly contracting inner part of the core) and an envelope with density distribution $\rho(r) \propto r^{-p}$ with $p \sim 2$ in the outer parts of the contracting core (Larson 1969). Density gradients measured in isolated low mass protostellar cores follow closely this distribution though the slope is slightly shallower with $p \sim 1.5 - 2$ over the external $1.4 \times 10^3 - 10^4 \text{ AU}$ (Shirley et al. 2000, 2003). When the central density rises above some 10^{10} cm^{-3} the heat generated by the release of gravitational energy in the contraction is not any longer freely radiated away and compression becomes adiabatic

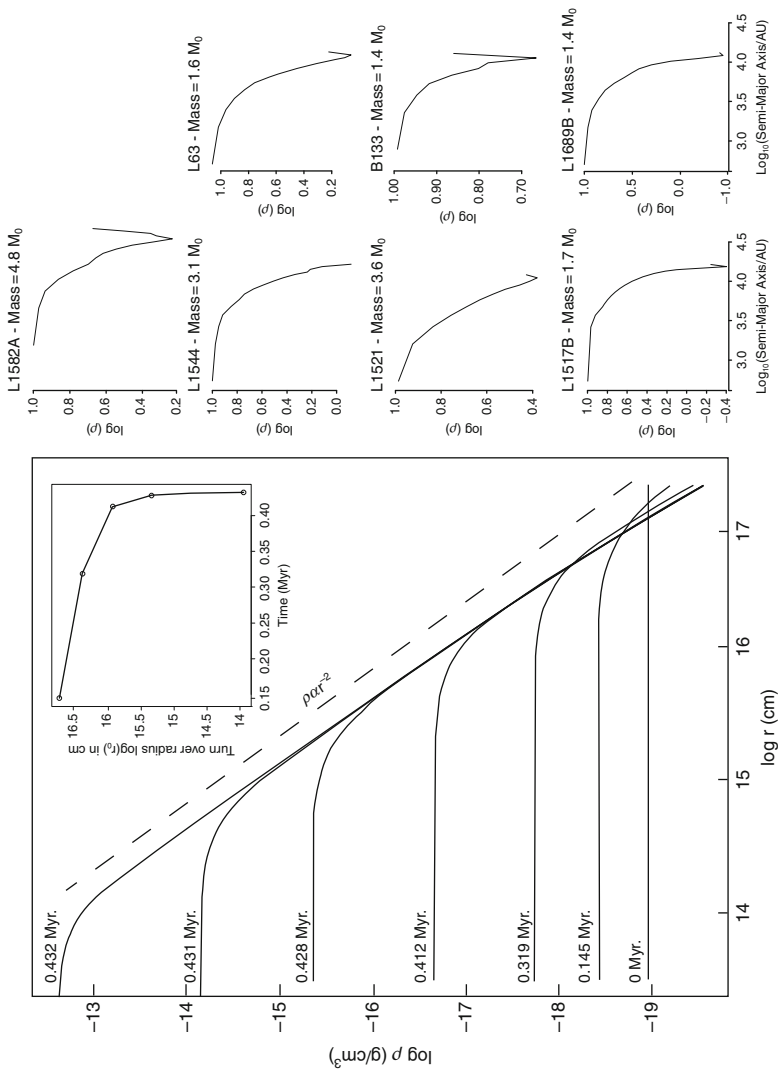



Fig. 6-3

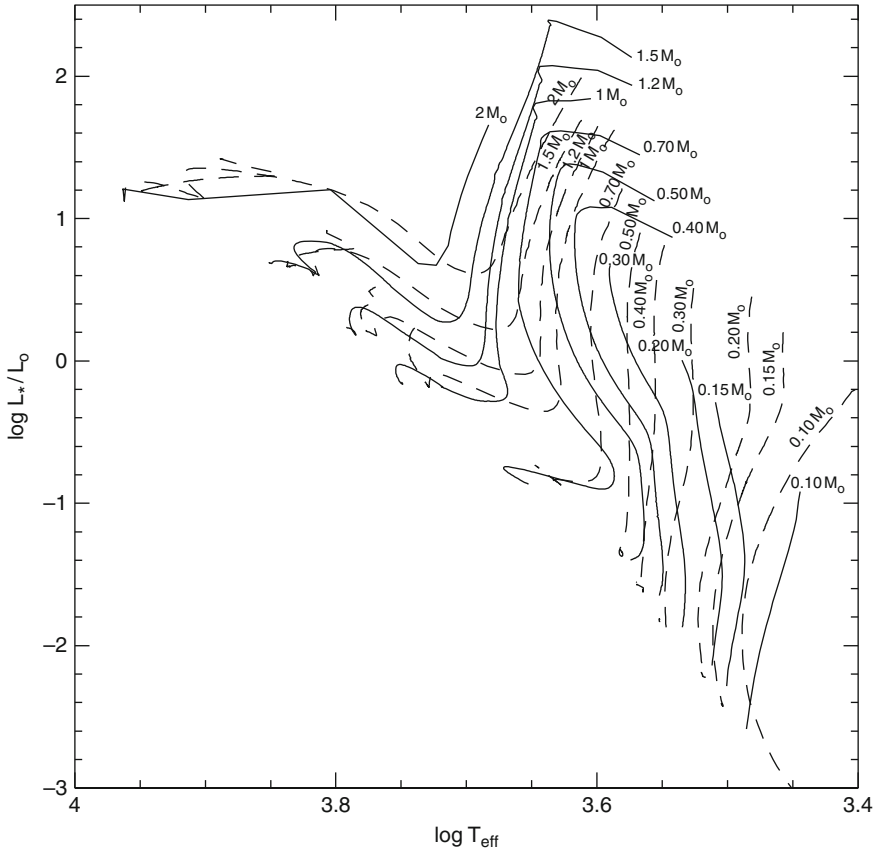
The evolution of core-envelope differentiation during the gravitational collapse is displayed in the *left panel* for a self-gravitating solar mass protostellar core with initial uniform density (from Larson 1969). The molecular core central density rises. The envelope density decreases as $\sim 1/r^2$. There is a turnover point in the density distribution that shifts to smaller radii as the collapse progresses as shown in the *upper inset*. The observed density radial distribution for the cores mapped by Kirk et al. (2005) is plotted in the *right panels*. The measured surface densities have been converted into volumetric densities assuming that the emitting cores are Bonnor–Ebert spheres with the radii derived by Kirk et al. (2005)

in the stellar embryo. The inner core contracts slowly hydrostatically while the remnant envelope keeps free-falling. Several readjustments of the protostellar core take place during this period caused mainly by the variations in the internal energy of the embryo associated with H_2 dissociation, dust processing, and the physics of the boundary layer and the shock front between the infalling external layers of the protostellar core and the growing embryo. The main accretion phase starts when the kinetic energy of the infalling material is negligible compared with its thermal energy and ends when approximately half of the total mass has fallen into the core. At this point the kinetic energy influx begins to decrease since the rise in the infall velocity ($V_{\text{ff}} = (2GM/R)^{1/2}$) cannot compensate the decrease in the density of the infalling material.

Observationally, objects during this period are named class 0 YSOs and are classified as such by either having a central compact centimeter radio source or by displaying extended and centrally peaked submillimeter continuum emission from the protostellar core. Additional criteria are a high ratio of submillimeter to bolometric luminosity ($L_{\text{submm}}/L_{\text{bol}} > 0.005$ with L_{submm} measured longward of $350 \mu\text{m}$) which basically measures the mass ratio between the embryo and the rest of the core (Andre et al. 2000). Young embryos can also be detected by means of the powerful bipolar outflows they generate. A good catalog of class 0 YSOs has been compiled by Froebrich (2005).

Accretion causes a variation of the Kelvin–Helmholtz timescale, $\tau_{K-H} = GM_*^2/R_*L_* = 3.2 \times 10^7 \text{ year} (M_*/M_\odot)^2 (R_*/R_\odot)^{-1} (L_*/L_\odot)^{-1}$, the timescale on which the star can radiate its thermal energy. τ_{K-H} increases as the star grows, and the growth rate is set by matter infall. After the main accretion phase, the temperature of the surface layers of the embryo decreases, as well as its luminosity, and the energy radiated from the surface of the embryo comes primarily from the interior instead of from the kinetic energy inflow. A convection zone develops and the star enters into the convective phase of the pre-main sequence (PMS) evolution (the so-called Hayashi track in low mass stars PMS evolution). At this point, the *stellar PMS evolution* begins. Hydrostatic equilibrium is appropriate to describe the main physics; accretion is often included as an external parameter that mildly affects the stellar structure.

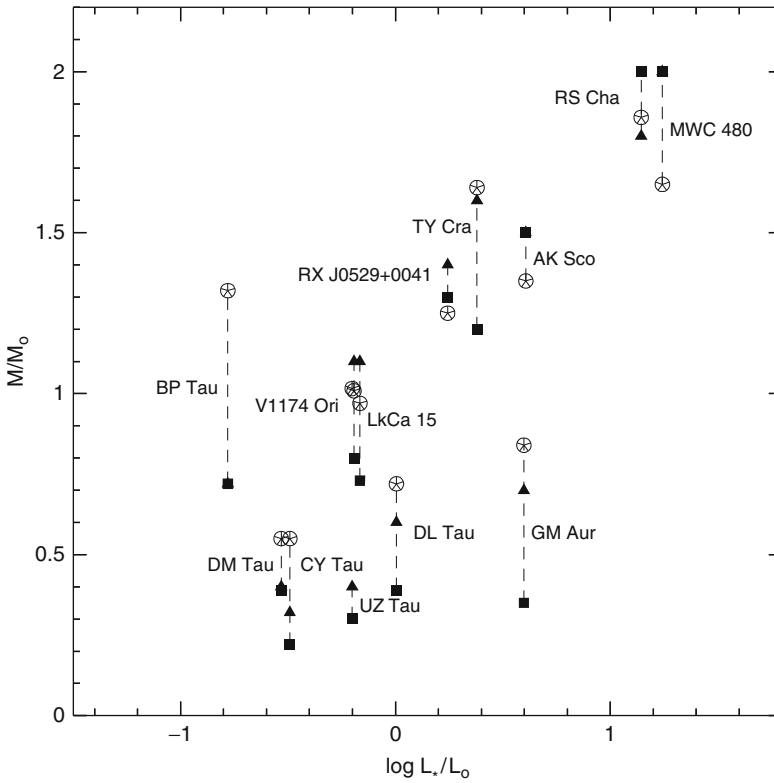
Modeling of this evolution often begins assuming central temperatures low enough to guarantee that nuclear energy generation rates are negligible relative to gravitational energy generation rates but high enough to guarantee that hydrogen and helium are ionized throughout most of the star (Iben 1965). Central-outward integrations are matched with surface-inward integrations. PMS evolutionary tracks are computed from an uncertain initial mass–radius relationship. Key ingredients are the equation of state (Mihalas et al. 1988; Pols et al. 1995; Rogers et al. 1996; Saumon et al. 1995), the opacity tables (Alexander and Ferguson 1994; Iglesias and Rogers 1996), and the reaction rates (Caughlan and Fowler 1988). Most of the models assume gray atmospheres and adopt the standard Mixing Length Theory (MLT) treatment (Boehm-Vitense 1958) in which the free parameter, α_{MLT} , measures the ratio between the mixing length and the pressure scale height; this parameter is tuned in a range between 1 and 2. D’Antona and Mazzitelli (1997) have introduced a different formalism to describe turbulent convection in the models that takes into account the energy redistribution among different size eddies within the turbulence cascade, following the Kolmogorov prescription based on Canuto and Mazzitelli (1991) treatment. Direct attempts to include the effect of accretion have been led by Siess and Forestini (1996) and Behrend and Maeder (2001). Accretion implies a flow of fresh deuterium into the star; the rapid rate of the reaction ${}^2\text{H}(p,\gamma)\rightarrow {}^3\text{He}$ (barely few hundred years) and the fast penetration of deuterium are able to influence the global stellar evolution (Siess et al. 1997). PMS evolutionary tracks from non-accreting PMS stars (D’Antona and Mazzitelli 1997) and accreting PMS stars (Siess et al. 2000) are plotted in  Fig. 6-4.



■ Fig. 6-4

Pre-main sequence evolutionary tracks are shown for low mass stars. *Continuous lines* represent the tracks obtained by D'Antona and Mazzitelli (1997) based on a turbulence cascade approach to treat energy transfer in the convective layer. Siess et al. (2002) tracks are shown with *dashed lines*. They include the effect of fresh deuterium accretion during the PMS evolution

A fundamental calibration of the PMS evolutionary tracks is difficult. The amount of sources with known dynamical masses is small: four PMS eclipsing binaries – RS Cha (Andersen 1991), TY CrA (Casey et al. 1998), V1174 Ori (Stassun et al. 2004) and RXJ0529+0041 (Covino et al. 2004), and three spectroscopic binaries – AK Sco (Andersen et al. 1989), UZ Tau A (Simon et al. 2000) and RXJ0529+1210 (Mace et al. 2009). Keplerian disks orbiting around PMS stars also provide dynamical measurements of the stellar mass. Spatially resolved CO molecular emission has been detected from nine of them (MWC 480, LkCa 15, DL Tau, GM Aur, DM Tau, CY Tau, BP Tau, GG Tau A, and UZ Tau E) by Simon et al. (2000). The mass–luminosity relationship derived from these few observations is compared with that derived from the theoretical tracks in ● Fig. 6-5. There is not a clear mass–luminosity trend and the deviations between the dynamical masses and the ones derived from evolutionary tracks fitting are about a 10–50%. This comparison can be further extended including double PMS stars and early-main sequence



■ Fig. 6-5

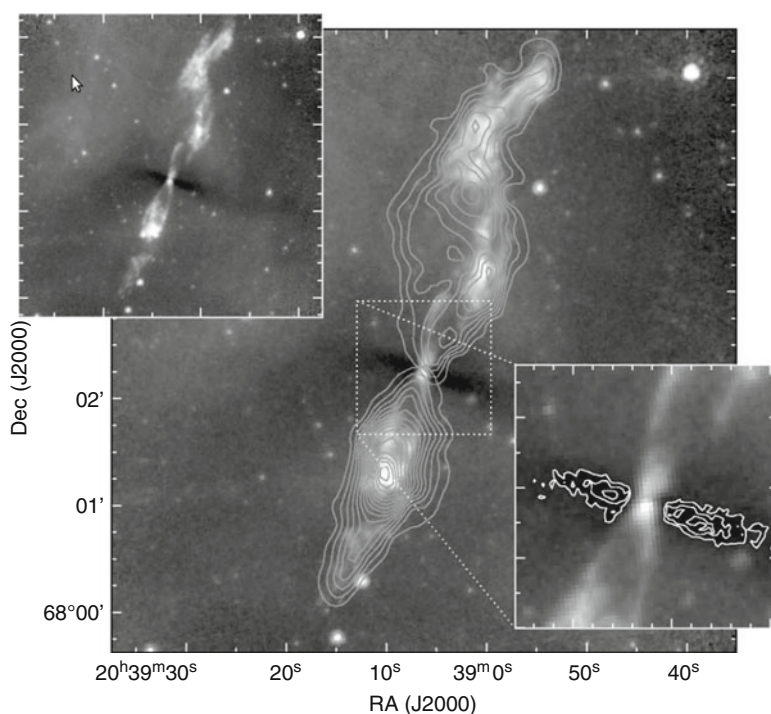
Mass–luminosity diagram for the PMS stars with known dynamical masses. Dynamical masses are marked with circled stars, masses derived from PMS tracks are marked with triangles and squares for Siess et al. (2002) and D’Antona and Mazzitelli (1997) tracks, respectively

binaries for which dynamical masses are also available. As shown by Hillenbrand and White (2004), there is generally good agreement between predicted and dynamical masses both for PMS and early main sequence stars in the mass range $1.2\text{--}2\,M_{\odot}$. Below this threshold, models underestimate the stellar masses, being the trend more significant for the lower mass end, from early M-type stars to $0.1\text{--}0.2\,M_{\odot}$ stars. At this end of the mass spectrum, the role of accretion disks in PMS evolution extends longer in time.

3 PMS Evolution: The Disk

Disks form because free fall drives rotating cores to its minimum energy configuration, where most of the mass is concentrated in the center and most of the angular momentum is stored in a distant reservoir of negligible mass, as in the Solar System. They are observed around YSOs independently of their mass and evolutionary phase with the only possible exception of early O-type stars. As early as a free-fall time, a flattened structure forms around the growing stellar embryo. In this lapse, material with negligible angular momentum has already been accreted

onto the stellar embryo and the envelope material begins to feed massively the flattened reservoir in centrifugal equilibrium around the star. About this time, a “*mechanism*” sets in the disk that allows transporting the high angular momentum material from the disk into the star, coupling stellar growth to disk evolution. This mechanism acts efficiently on the gas component and, as disks loose gas, the accretion rate decreases. In about few Myrs, the gas to dust mass density ratio falls from 100, the fiducial ISM value, to less than 4 in transitional disks (see Collins et al. 2009) and later to values $\ll 1$ in debris disks as, for instance, β Pic disk (Roberge et al. 2006). In [Fig. 6-6](#) (from Looney et al. 2007), both components of the disk (gas and dust) are displayed for a class 0 YSO: L1157. The extinction of the environmental infrared radiation by the dust in the disk produces the dark lane. The molecular (CO) gas radio emission from the disk is also plotted as a contour map in the lower left inset. Perpendicular to them, there is a prominent bipolar outflow, mainly constituted of molecular gas with two clear components: the hot shocked gas, at temperatures $\geq 2,000$ K, radiating in the infrared H_2 lines and the cold (< 100 K) entrained gas radiating at radio frequencies in the CO line. As the radial extent of disks is some thousands of AU, they are less conspicuous than the outflows that may extend over more than a parsec. As other astrophysical disks, protostellar disks are associated with the



■ Fig. 6-6

Far infrared emission/absorption of the class 0 YSO: L1157 obtained with IRAC (Looney et al. 2007). Emission is produced along the outflow by molecular radiation. Absorption of the environmental infrared radiation from the dusty disk produces a dark lane perpendicular to the outflow, as shown in the upper inset. The molecular bipolar outflow is represented in the main panel. In the bottom right panel, the CO 2-1 emission from the molecular gas disk (from Bachiller et al. 2001) is overlaid on the dusty disk

formation of powerful bipolar outflows which are often the earliest observational signature of star formation. The connection between disks and outflows is described in [Sect. 4](#).

There are two main differences between protostellar and other types of astrophysical disks such as the disks around active galactic nuclei or white dwarfs. Firstly, the thermal properties of the disk are controlled by the dust and depend strongly on the optical properties and thickness of their icy coatings as well as on the dust grains' size distribution. As a result, radiative processes are significantly more difficult to constraint than in the usual astrophysical disks made of simple atomic gas. Secondly, protostellar disks are not formed by smooth mass transfer between binary companions but as a result of gravitational collapse. Early stages, when the disk grows out of the envelope can be observed opening the way to the study of the physics of disk formation.

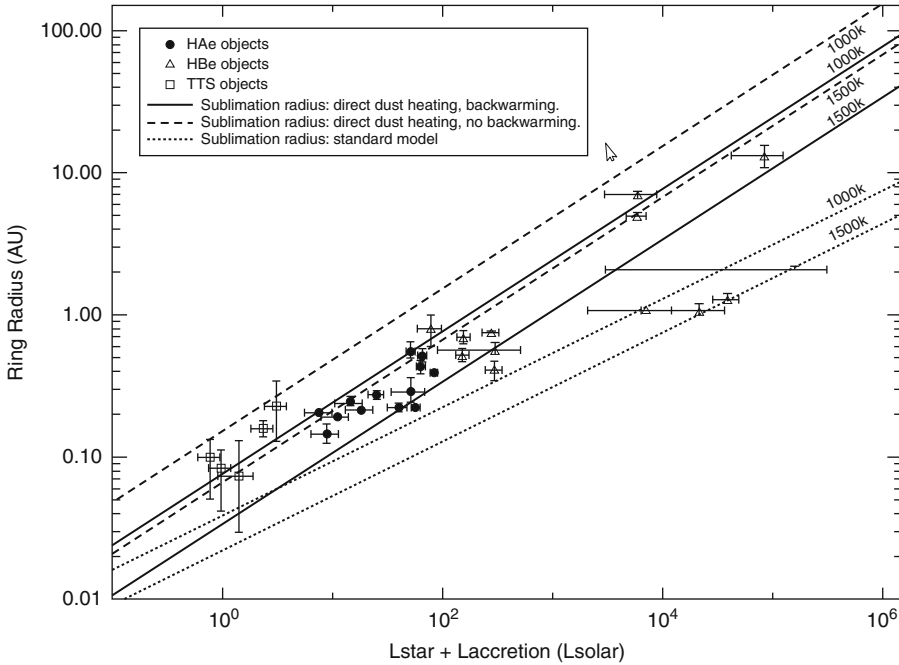
Disks are observed around YSOs at all evolutionary stages from the early class 0 stages to the late class III sources. *This sequence traces the history of accretion and disk evolution.* In this section, the current knowledge on the properties, structure, and evolution of protostellar disks is reviewed.

3.1 Morphology of Protostellar Disks

The nearest star forming regions are in a ring of radius 140 pc around the Sun. This ring marks the molecular walls of the Local Bubble in the ISM and forms part of a local kinematical and physical structure in the Galaxy known as the Gould's Belt (see Appendix A). Only very few and low mass class III YSOs are placed closer to the Sun, namely, TW Hya at 56 pc or the more evolved AB Dor at 15 pc. Thus, the distance of 140 pc sets the target angular resolution for disks' studies: 7 mas to resolve 1 AU scales or 0.07 mas to resolve the boundary between the star and the disk. Currently, the only way to achieve such a high resolution is by means of interferometry at optical and infrared wavelengths. Typical facilities have baselines of about 100 m and probe wavelengths of 1–13 μm providing spatial resolutions of 2–10 mas or 0.25–1.25 AU. These wavelengths trace hot dust emission (300–1,800 K) from the disk inner AU. The main structural properties of protostellar disks are described below.

3.1.1 The Inner Cavity Filled in with Hot Gas

The best studied targets are the Herbig Ae/Be stars (HAeBe stars). Herbig stars are YSOs in *Phase III* with masses of 2–10 solar masses that correspond to spectral types A to B. They are more luminous than the low mass ($<2M_{\odot}$) YSOs also known as T Tauri Stars or TTs. Disks show as extended ellipses due to the projection of the circular disk on the plane of the sky. At 1–4 μm , disk emission is expected to be dominated by hot dust at the disk surface that scatters the stellar radiation. In fact, there is a strong correlation between the measured disk radii (r) and the luminosity of the central stars (L_{\star}): $r \propto L_{\star}^{1/2}$ (Millan-Gabet et al. 2007) as shown in [Fig. 6-7](#). The observed disks radii are too large for the inefficient grazing incidence illumination of the disk surface. Direct illumination by the star is required instead, i.e., an inner disk cavity devoid of dust must exist around the star. Measured radii are consistent with the dust sublimation radii of relatively large dust grains ($\geq 1\mu\text{m}$) with sublimation temperatures in the range 1,000–1,500 K. The existence of this dust sublimation ring has been independently inferred from the bump in the near infrared energy distribution of Herbig stars (Natta et al.



■ Fig. 6-7

Measured sizes of H A e and TTSs objects versus central luminosity and comparison with sublimation radii for dust directly heated by the central luminosity (*solid and dashed lines*) and for the oblique heating of flaring disks (*dotted line*) after Millan-Gabet et al. (2007)

2001). The inner rim of the dusty disk is expected to be puffed-up by hydrostatic equilibrium but its vertical structure is still unclear.


The inner cavity of the disk is filled in with hot gas that is well traced by the CO fundamental ($\Delta v = 1$) ro-vibrational transition at $2.3 \mu\text{m}$. CO is an abundant molecule with high dissociation energy that can survive to temperatures of 4,000–5,000 K and is observed in nearly all CTTS (Najita et al. 2003). Many line profiles show the classic double-peaked profile produced in rotating rings. The location of the emission region can be derived assuming Keplerian rotation in a thin disk. The profile also contains information on the radial variation of the gas emission. Radial variation of the gas temperature and column density can also be constrained if the first overtone band ($\Delta v = 2$) at $4.7 \mu\text{m}$ or other transitions from the same molecule are detected. The inner gas radius peaks about 0.04 AU and it is generally smaller than the corotation radius,⁴ while the inner rim in the dust distribution is above this value. In addition, strong continuum emission has been detected in the far ultraviolet (1,300–1,700 Å) from some TTSs with bright molecular disks and inner disk gaps of few AU (Ingleby et al. 2009). This emission is likely due to energetic photoelectrons mixed into the molecular layer, providing additional evidence on the existence of a hot component in the inner disk.

⁴Corotation radius is the disk radius at which the Keplerian frequency equals the angular velocity of the star. Typical rotation velocities of YSOs range between 2 and 15 km s⁻¹; thus, the corotation radius is: $r_{\text{corot}} = 0.08 \text{ AU} (M_*/M_\odot)^{1/3} (P/8\text{d})^{2/3}$, with P the rotation period of the star.

3.1.2 Disks Are Flared


The infrared spectrum of disks depends on the geometry of the disk and the composition of the dust. “Bowl-shaped” flared disks are much more efficient at capturing stellar photons than flat disks since (1) the geometrical covering factor of the stellar radiation field is higher and (2) the radiation incidence angle is closer to the normal to the disk surface.

Disk flaring is required to explain the spectral energy distribution (SED) of TTSs and H AeBe stars. In the 1–10 μm range, SEDs can be fitted by power laws ($\nu S_\nu \propto \nu^s$) with slopes, s , that range between 0.6 and 1 for the TTSs and between -1 and 1 for H AeBe stars. These slopes are softer than those expected from standard irradiated disks ($s = 4/3$) suggesting that the outer rings of the disk, the ones that radiate at the longest wavelengths, are capturing the stellar photons more efficiently than expected, i.e., disks are flared.

The upper, atmospheric disk layers can be mapped directly by means of the 8–13 μm radiation produced by the warm dust; infrared interferometers allow to probe scales of 1–4 AU. The radius of the region radiating at 8–13 μm is larger in H AeBe stars with strong flux in the far infrared⁵ (Leinert et al. 2004) that are expected to have more flared disks. The measured radii are however, smaller than predicted by simple flared disk models (Millan-Gabet et al. 2001) suggesting an extra shadowing of the stellar radiation illuminating the puffed up inner rim of the dusty disk, as the depicted in  Fig. 6-2.

In addition, radio interferometry at millimetric wavelengths allows to explore the gas distribution as traced by the CO emission at scales >50 AU. $^{12}\text{CO}(J = 2 - 1)$ emission is produced at the disk surface and the radial distribution of CO kinetic temperature also agrees with the expectations of flared disks.

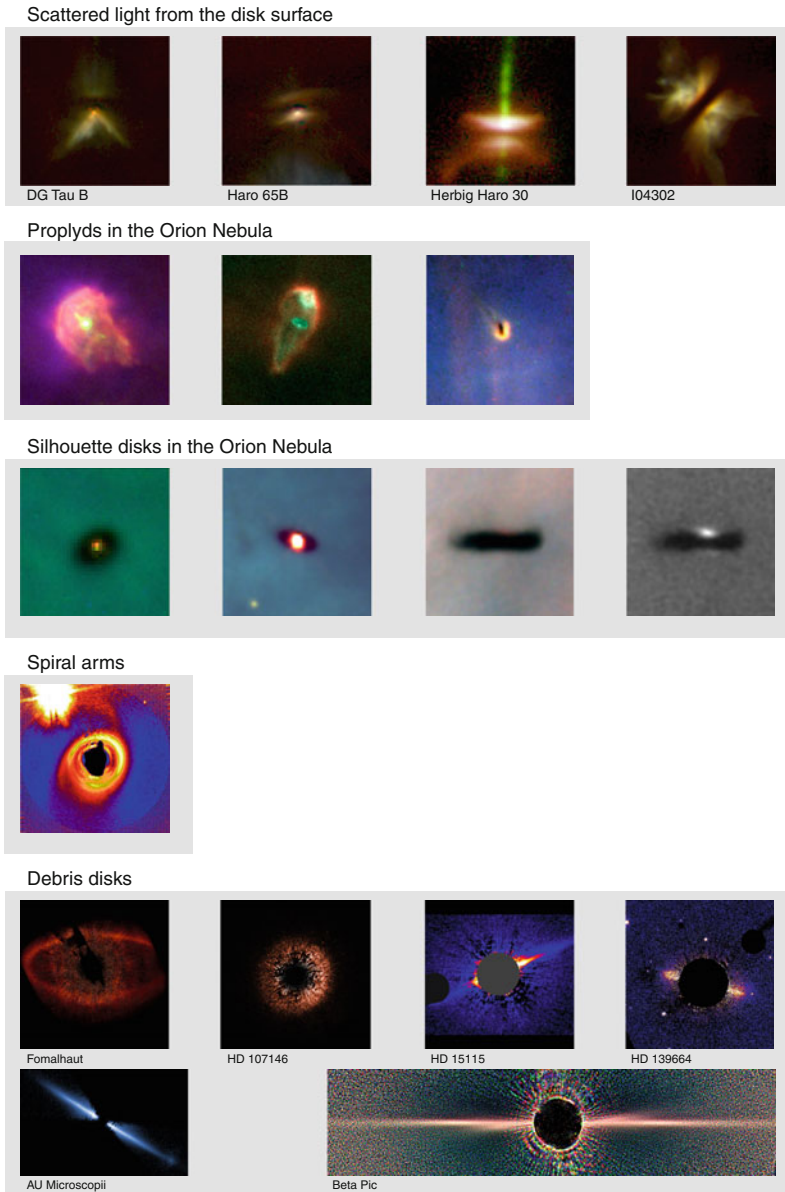
3.1.3 Dusty Atmospheres and Warps

At 0.4–1 μm scales, high resolution direct imaging is feasible with the Hubble Space Telescope (HST) and Adaptive Optics from the ground. The stellar radiation scattered by the circumstellar dust allows mapping the distribution and properties of the dust in the inner 500 AU with spatial resolution of about 100 mas (14 AU). Class I and II YSOs have thick dust disks that absorb the stellar radiation producing a dark lane in the images when disks are observed close to edge-on (see  Fig. 6-8). This lane also acts like a natural coronagraph that blocks direct stellar light providing high contrast for direct imaging. Further information is obtained from highly disturbed disks detected by their blockage of the environmental radiation, the so-called silhouette disks. Coronagraphs have also been used to block the stellar radiation and increase the dynamical range in the images; they typically block the inner, and more interesting, 0.”5 (70 AU). However, despite all the efforts, most of the YSOs disks remain undetectable.

There are about 15 edge-on systems known with outer radii⁶ ranging from 30 AU in IRAS 04325+2402 (Hartmann et al. 1999) to 620 AU in Orion 114–426 (Shuping et al. 2003). The vertical distribution of dust in the disks is derived from these images (see details in Watson et al. 2007). Disk scale heights are larger in systems with outflows like HH 30 where it reaches 6.3 AU at 50 AU from the star. Scattering dust grains are dominated by small particles that have

⁵H AeBe stars are classified into two groups according with s : Group I with $s = -1 - 0.2$ and strong far infrared excess and Group II with $s = 0.2 - 1$ and weak far infrared excess (Meeus et al. 2001).

⁶Disk may continue beyond the optical nebula



■ Fig. 6-8

High resolution images of YSOs' disks obtained with the Hubble Space Telescope. From *top to bottom*: edge-on YSOs illuminating the dusty atmosphere of the disks in a quiet molecular cloud; proplyds or disrupted YSOs moving in the photoionized environment of the Orion star forming region; silhouette disks observed against the background radiation in Orion; AB Aur disk where hints of spiral structure has been detected; debris disks, with a low gas content illuminated by the star: rings, warps, and double structures are observed (Padgett et al. 1999; Bally et al. 2000; Clampin et al. 2003; Kalas et al. 2006, 2007, 2008; Krist et al. 2005; Golimowski et al. 2006)

much higher scattering cross section at shorter wavelengths. There is also evidence of grain size growth toward the disk midplane. Remnant circumstellar dusty envelopes have been detected in several systems. Occasionally, morphological variations are detected in the images. The most intriguing ones have been observed in HH 30; the reflection nebula varies on timescales ≤ 1 year that cannot be accounted by variations in the large-scale structure of the disk but rather on the illumination from the inner 1 AU. Whether this is yet another manifestation of the variable extinction events produced at the inner rim of the disk or in disk warps (as the detected in AA Tau by O'Sullivan et al. 2005) is still under debate.

3.2 Disk Mass

Molecular emission provides direct measurements of the gas content of the disk and its velocity pattern. The gas to dust mass ratio is not constant in YSOs' disks; close to the disk midplane most of the gas is frozen out into dust grains while dust grains are melted in the surface areas (Rettig et al. 2006). On average, the dust to gas mass ratio in protostellar disks is considered to be about a hundred, similar to that observed in the ISM. For this ratio, the column densities of YSOs disks can be approximated to be $\Sigma(r) = 0.8 \text{ g cm}^{-2} (r/100 \text{ AU})^{-1.5}$ from CO interferometric measurements (Dutrey et al. 2007).

Disk rotation has been investigated in a limited amount of YSOs' disks since CO interferometry is required to resolve the disk rotation curve (Mannings and Sargent 2000; Simon et al. 2000). The sample is biased to disks with strong CO emission that are found to have a large outer radius: 200–1,000 AU. However, smaller (and less massive) disks are detected around some heavily accreting TTSSs, as BP Tau: the CO disk radius is ~ 120 AU and its mass $\sim 1.2 \times 10^{-3} M_{\odot}$ (Dutrey et al. 2003). All disks are found in Keplerian rotation with the exception of AB Aur that also shows a prominent spiral pattern in the optical and near infrared (Grady et al. 1999; Lin et al. 2006).

The main discriminating feature between disks around high and low mass PMS stars is the ratio between the disk and the stellar mass, M_d/M_* . In low mass stars $M_d/M_* < 0.1$, in massive B-type stars $M_d/M_* \leq 1$, and $M_d/M_* \gg 1$ in O-type stars. The stability and lifetime of high mass YSOs' disks is uncertain. Typically, the massive rotating structures observed in YSOs with luminosities $\geq 10^5 L_{\odot}$ have radii of 4,000–30,000 AU, masses of 60–500 M_{\odot} , and rotation speeds of a few km s^{-1} . For these parameters, the accretion rate would be about $2 \times 10^{-3} - 2 \times 10^{-2} M_{\odot} \text{ year}^{-1}$ (Zhang et al. 2005) corresponding to timescales for the whole disk depletion of 10^4 year, i.e., an order of magnitude smaller than the rotation period at the outer disk radius. This indicates that these very massive rotating structures cannot be centrifugally supported and should be considered as transient, nonequilibrium evolving structures; they probably host a stellar cluster. Most of the bona fide candidates to disks around high mass YSOs orbit about $\sim 10 M_{\odot}$ stars, with luminosities typical of B-type ZAMS stars. In these objects, the estimated lifetime of the disks is about 10^5 year that it is comparable to the free-fall time of the mother molecular core. The most convincing disk around a high mass YSO has been detected in the Orion KL region through OH maser emission: IRAS 20126+4104 (Edris et al. 2005). The temperature law in the disk seems to follow $T(r) \propto r^{-3/4}$ and the rotation to be Keplerian (Cesaroni et al. 2005).

Identifying candidate disks around very massive protostars requires a careful selection of targets and tracers, to overcome the problems related to observations of massive YSOs. The most important of these are the large distance (a few kpc i.e. about ten times farther than the Gould's

Belt) and the confusion caused by the fact that massive stars form in rich clusters. In spite of the numerous attempts, no disks have been detected about early O-type stars (see Cesaroni et al. 2007 for a detailed analysis on possible observational biases).

3.3 Composition of Protostellar Disks

Disk composition is dominated by molecular hydrogen while the thermal properties of disks are controlled by dust that acts as the main coolant. A thin layer of atomic gas is also detected at the disk atmosphere caused by the photodissociation cascade activated by the stellar UV photons.

3.3.1 Dust

Dust composition and mixture is best investigated in the infrared spectrum of the PMS stars. The strong resonances of abundant dust species both O-rich (silicates either amorphous or crystalline) and C-rich (Polycyclic Aromatic Hydrocarbons or PAHs) are observed in the 8–13 μm range. The broad 9.7 μm amorphous silicate (Si-O stretching) feature dominates this wavelength regime and evolves from absorption in young, embedded sources, to emission in optically visible stars and to complete absence in debris disk systems both for TTs and HAeBe stars (Furlan et al. 2006; Kessler-Silacci et al. 2005). The presence of PolyAromatic Hydrocarbons (PAHs) (6.2, 7.7, 8.6, 11.3 μm) has been reported in the spectra of only some TTs (Geers et al. 2006).

A typical model of dust grain composition in protostellar disks consists of a mixture of three main components: small amorphous carbon grains, large silicate grains (in two basic compounds olivine and orthopyroxene), and silicate grains coated with “dirty ice.” “Dirty ice” is mainly constituted of H_2O , CO, and other simple molecules though the exact composition is very uncertain and depends on the physical conditions and the local star forming history in the parent molecular cloud. The most abundant ice in protostellar disks is H_2O ice that plays a major role in the opacity of disks at temperatures below its vaporization threshold (Preibisch et al. 1993).

Grain growth in the dense protostellar disks modifies the size distribution with respect to that observed in the ISM⁷ increasing a_{max} up to several microns and producing variations in the power index. Disk opacity is very sensitive to the fraction of dust grains with sizes comparable to the local thermal wavelength: $a \sim \lambda = 0.29 \mu\text{m}/T(K)$. Ice and silicates behave like good dielectrics; following Mie theory, absorption and scattering efficiencies are weakly wavelength dependent if $a/\lambda > 1$ but below this value they drop rapidly to 0. As shown in ► Table 6-1, the fraction of the disk mass in silicates and water ice is about a 64% while refractory organics account for 25% of the total mass of the disk in dust. Strongly absorbing materials (with imaginary refraction index), such as iron or fractals made of highly conducting materials (e.g., graphite), represent a small fraction of the total dust mass of the disk (Pollack et al. 1994).

There are clear indications of dust evolution from the molecular cores to the protostellar disks. In addition to the coagulation of 0.1 μm to micron-sized dust grains that modifies the size distribution, dust observed in disks is crystallized resulting in an increasing fraction of crystalline silica, forsterite, and enstatite with respect to the amorphous pyroxene and olivine.

⁷The number density of grains of size, a , is given by $N(a) \propto a^{-3.5}$ with $a_{\text{min}} = 5 \times 10^{-3} \mu\text{m}$ and $a_{\text{max}} = 0.25 \mu\text{m}$ (Mathis et al. 1977).

■ Table 6-1

Dust grain species in protostellar disks^a

Species	Bulk density (g/cm ⁻³)	Mass fraction ($M_{\text{specie}}/M_{\text{disk}}$)
Metallic iron	7.87	1.26×10^{-4}
Olivine	3.40	2.51×10^{-3}
Orthopyroxene	3.49	7.33×10^{-4}
Troilite (FeS)	4.83	7.68×10^{-4}
Refractory organic	1.5	3.53×10^{-3}
Volatile organic	1.0	6.02×10^{-4}
Water ice	0.92	5.55×10^{-3}

^aAfter Table 2 in Pollack et al. 1994. About 10% of the silicates were later found to be in crystalline form (Sargent et al. 2009), see text

In HAeBe stars, silicate emission provides evidence of stellar heating of the disk atmosphere in the inner ~ 10 AU. In fact, there is a very high fraction of crystalline dust in the central 1–2 AU around HAeBe stars. This decreases toward 2–10 AU but still is significantly higher than in the ISM, even at stages as early as 1 Myr, well before the onset of planet formation (van Boekel et al. 2004). This is surprising since the temperature in these areas of the disk is well below the glass formation temperature (about 1,000 K) but could be explained by chemical models of protoplanetary disks that include effects of radial mixing and local processes in the outer disk (Gail 2004). Crystalline silicate emission has been detected from TTSSs; however, there is no evidence of a radial gradient in the mass fraction of crystalline silicates (Sargent et al. 2009).

As YSOs' disks evolve, their mid/far-infrared spectrum becomes more similar to that observed in comets. There is also a radial gradient in dust chemistry and in the grain size with large grains being more abundant in the inner disk.

3.3.2 Molecular Gas

H₂ is the main component of protostellar disks (the gas to dust mass ratio is ~ 100 and CO/H₂ $\sim 7 \times 10^{-5}$) and does not freeze-out onto grains. Though the molecule is effectively photodissociated through absorption of UV radiation in the Lyman and Werner bands (912–1,100 Å range), H₂ self-shields efficiently against photodissociation. Unfortunately, the pure rotational lines in the infrared are intrinsically weak and difficult to detect. The strong electronic transitions at ultraviolet wavelengths are pumped by the stellar UV photons and are good tracers of the disk atmosphere and innermost region.

After CO and H₂, the most abundant gas molecules are: H₂O, OH, S₂H, HCO⁺, CN, C₂H, CS, HCN, H₂CO, and DCO⁺. The sensitivity of radio-interferometers has been too low to map their spatial distribution in disks. Simple molecules such as CO, ¹³CO, HCO⁺, CN, and HCN are detected from single dish observations. Line ratios indicate that the molecular emission arises from dense (10^6 – 10^8 cm⁻³) and moderately warm ($T \sim 20$ –40 K) intermediate height regions of the disk atmosphere between the midplane and the disk upper layer, in accordance with the predictions from models of disk chemistry. Molecular gas abundances are lower in the YSOs than in the molecular prestellar cores. This is caused by gas freeze-out into dust grains in the cold midplane and molecules' photodissociation in the upper disk layers. CN is very strong in all disks. The CN/HCN abundance ratio toward the HAeBe stars is even higher than that

found in galactic photon-dominated regions, pointing out to the relevance of photodissociative processes in the upper layers of the disk. No CH_3OH has been detected to abundances 10^{-8} – 10^{-9} with respect to H_2 .

3.3.3 Disk Chemistry

Understanding disk chemistry requires to consider both gas phase and dust grain-surface processes. Dust grains are key ingredients in the disk chemistry since they control the degree of ionization in the most opaque areas of the disk; dust is an efficient electron donor for recombining ions and a sink for neutrals. At low temperatures ($T \approx 10$ K) large molecules are stationary on the grain surface and grow by reaction with mobile species like hydrogen. However, as the temperature rises, diffusion and thermal evaporation produce desorption of the light species such as hydrogen, thus braking the hydrogenation-driven chemistry on the surface. Later, heavier compounds move from the grain surface to the gas phase. Desorption processes depend exponentially on the grain temperature, gas phase abundances, and chemistry; thus chemical modeling is very sensitive to the processing of the grain mantles. Surface reactions are compiled by Hasegawa et al. (1992) and Hasegawa and Herbst (1993) into the Ohio State University (OSU) gas-grain chemical model.

Gas phase reactions and their rate coefficients are available in the UMIST RATE06 database (Woodall et al. 2007) that includes about 420 species involving 13 elements: H, He, C, N, O, F, Na, Mg, Si, P, S, Cl, and Fe. Uncertainties in the chemical reaction rates (RATE06 contains about 4,800 chemical reactions) may lead to ill-determined gas columns by as much as a factor of ~ 4 , for some species. This is mainly caused by the large uncertainties in about a hundred chemical reactions (Vasyunin et al. 2008). Species such as CO, H_3^+ , H_2O , NH_3 , N_2H^+ , and HCNH^+ seem to be rather insensitive to these uncertainties while other relevant species such as CS, CO_2 , H_2CO , CN, HCN, and HNC are sensitive to them. Further uncertainties in the chemical modeling of disks are introduced by the difficulties to identify the precise route to the formation of a given molecule, especially when dealing with large molecules that form both through gas phase and grain surface reactions. As an example, the theoretically derived abundance of HCOOCH_3 gas in class 0 YSOs (hot cores) through the gas-based formation route is roughly a third of the total abundance derived including both gas and grain routes (Garrod and Herbst 2006).



In general, chemical models of disk structure and evolution are developed with reduced sets of equations that are defined selecting only those species necessary to compute the abundances of the compounds under study. As an example, to determine the ionization degree of protostellar disks, Semenov et al. (2004) worked with a reduced network of chemical reactions selected according to a cutoff defined by the requested accuracy of the final model. The number of species included in the reduced chemical networks to study the disk ionization fraction ranged from four in the disk midplane at 1 AU from the star to some 50 at intermediate layers between the irradiated atmosphere in the opaque interior. The integration of gas phase and grain surface processes in chemical modeling of protostellar disks is hampered by the different nature of both processes. Gas phase chemistry can be simulated using rate equations; however, grain chemistry requires taking into consideration the availability of reacting particles in the grain. Some recent approaches have been taken introducing modified rate equations to grain surface reactions that include the probability of each reactant being available in the grain surface (Garrod et al. 2009).

Unfortunately, there is not yet a fully consistent disk model that solves for the chemical and dynamical structures simultaneously. As the chemistry of the disk evolves on timescales comparable to the dynamical timescale, this is a relevant source of uncertainty.

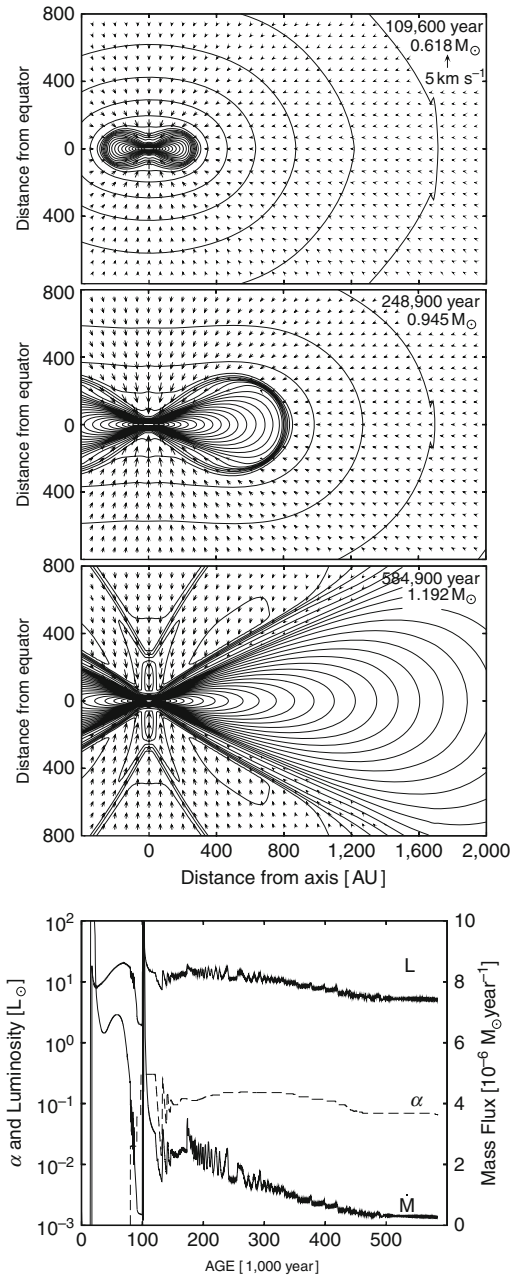
3.4 Disk Formation

During *Phase II*, accretion proceeds directly from the envelope of the molecular core onto the stellar embryo and the flattened structure that sets the seed of the disk. After roughly a free-fall time, material with negligible angular momentum has already been accreted onto the stellar embryo, the disk begins to develop and channels the mass flow onto the star.

The formation of an equilibrium disk requires that self-gravity balances gas pressure at any given point in the rotating disk. This balance is measured by the Toomre parameter,⁸ Q , that must be larger than 1 for disks to remain stable. Disturbances on scales above $\Lambda_c = G\Sigma/\kappa^2$ are stabilized by the shear setting a maximum size of the unstable regions (Lynden-Bell and Pringle 1974). At smaller scales, instabilities develop, accompanied of local heating and angular momentum and mass transport.

Numerical simulations of the early gravitational collapse and disk/star differentiation are displayed in  Fig. 6-9 from Yorke and Bodenheimer (1999). Let us use them as a departure point to describe disk formation. Departing from an initial configuration $\rho(r) \propto r^{-2}$, a total mass of $2 M_\odot$, and a specific angular momentum of $\sim 10^{21} \text{ cm}^2 \text{ s}^{-1}$, *Phase II* ends in about 10^5 years, just one free-fall time, since no hydromagnetic support is taken into account in this simulation. Afterward, the disk seed begins to grow radially. It is necessary to prescribe the radial transport of mass through the disk to the star to avoid mass accumulation in the disk, i.e., to evolve the flattened structure into an accretion disk. The pace at which matter flows onto the star is controlled by angular momentum conservation and thus, it is set by prescribing the angular momentum transport within the disk. This is often done by introducing the parameter α that scales the mass flow with the local sound speed in the disk. In this approach, transport within the disk is not a conservative process and a fraction of the gravitational energy is lost into disk heating and radiation. In the bottom panel of  Fig. 6-9, it is shown how α needs to be modified for this particular simulation to keep the disk stable (according to Toomre's criterium). Major variations need to be introduced during the first free-fall time, after that α only needs to be varied mildly. Once transport is prescribed, the evolution of the disk during *Phase II* depends on the mass of the core (the larger the mass the earlier it ends) and on the initial specific angular momentum of the core (the smaller the angular momentum, the larger the amount of low angular momentum material available for direct accretion and so the duration of this phase). After the end of *Phase II*, the disk grows radially out to several thousands of AU and accretion shocks develop as the kinetic energy of the infalling material is released. The layered structure of the protostellar disk and the central force exerted by the star favors kinetic energy dissipation in two major steps. Firstly, an external shock front forms where the diffuse core material encounters the outer vertical edge of the disk. This material is later accelerated by the stellar gravitational field reaching highly supersonic velocities giving rise to the formation of a second inner shock front. The simulations represented in the figures are based on a simple treatment of angular momentum transport and ignore the presence of magnetic fields and the effect of outflows; as a result, after a little over three times the original free-fall scale, still 40% of the core mass is in the disk and the disk-to-mass ratio is 0.8, much larger than observed. In fact, values typically inferred from observations indicate that the disk-to-star mass ratio in low mass stars at protostellar stages is about 0.1 (Beckwith et al. 1990). Notice that

⁸Rotating, thin, gaseous disks are stable against gravitational perturbations if, $Q = c_s \kappa / \pi G \Sigma > 1$, where c_s is the sound speed, κ is the epicyclic frequency that measures the shear locally in the disk, G is the gravitational constant, and Σ is the surface density of matter (Toomre 1964).



■ Fig. 6-9

Numerical simulations of the gravitational collapse of a $2M_{\odot}$ from Yorke and Bodenheimer (1999). (a) Density contours map and the velocity field are plotted at $0.56t_{\text{ff}}$, $1.28t_{\text{ff}}$, and $3.01t_{\text{ff}}$ in the upper panels. The initial angular velocity of the protostellar core is $5 \times 10^{-14} \text{ s}^{-1}$. During disk evolution, angular momentum transport is prescribed through the α formalism. (b) In the bottom inset, the evolution of α needed to keep the disk stable and accreting during the gravitational collapse is plotted, as well as the evolution of the luminosity (L) and the accretion rate (\dot{M})

if disks were to absorb directly all the angular momentum of the core, their radius would be $j^2/GM = 4.97\text{AU} \times 10^2 (j/10^{21})^2 (M/M_\odot)^{-1}$ and star formation would be a rather inefficient process since a fair fraction of the mass of the core would be kept stored in the disk. Angular momentum transport during disk formation is still an open problem.

3.5 Protostellar Disks and the α -Prescription Paradigm

The Toomre parameter is derived from pure hydrodynamical, linear theory; it cannot be applied to turbulent or magnetized disks. In 1973, Shakura and Sunyaev, showed that a simple prescription, the α -prescription, can be used to describe angular momentum transport and heating in disks. In this approach, disks are treated as a set of nested rings with surface density $\Sigma(r)$ that rotate at different angular velocities. The friction between adjacent rings generates a tangential stress $-\alpha\Sigma c_s^2$ that produces an inward flow of matter to the central object. The name of the prescription comes from the α parameter used to scale the stress with the local pressure. Each ring radiates through the disk face roughly as a blackbody with a temperature $T(r)$ given by equating the gravitational energy dissipation rate to the blackbody flux:

$$T(r) = 2,100 \text{ K} \left(\frac{\dot{M}}{10^{-8} M_\odot \text{ year}^{-1}} \right)^{1/4} \left(\frac{M_*}{M_\odot} \right)^{1/4} \left(\frac{r}{r_{in}} \right)^{-3/4} f(r)$$

where \dot{M} is the accretion rate, M_\odot and R_\odot are the solar mass and radius, respectively, and $f(r) = (1 - (r/R_*)^{-1/2})^{1/4}$. The magnitudes are scaled to typical values for TTSs: accretion rates of $10^{-8} M_\odot \text{ year}^{-1}$, stellar masses of $1 M_\odot$, and inner disk radius, r_{in} , of about $3R_\odot = 0.014 \text{ AU}$. Henceforth, according to Wien's law, disk radiation should peak at the near infrared and, the higher the accretion rate, the higher the disk temperature should be. YSO's infrared SED is well fitted within this prescription that drove to the first classification of YSOs into class I, II, and III (Adams et al. 1987). The SED slope in the range $1\text{--}100 \mu\text{m}$ varies from $0 < s \leq 1$ in class I sources, to $-2 \leq s \leq 0$ in class II sources, to $-3 < s \leq -2$ in class III sources. In class I sources, the infalling envelope still has a non-negligible contribution to the infrared SED producing an excess toward lower energies and longer wavelengths. In class II sources, SEDs depend directly on the accretion rate following the α -prescription. Finally, the SED of class III sources is dominated by the reprocessing of the stellar radiation on the disks (see also [Sect. 3.6](#)).

In spite of this success, it is still unclear why the α -prescription works in protostellar disks. The α -prescription is tied to the definition of a kinematical viscosity, ν , such that $\nu = \alpha c_s(r)H(r)$ with $H(r)$ being the maximal scale of the flow circulation cell, considered to be of the order of the disk thickness (Lynden-Bell and Pringle 1974), in a disk analogue to the mixing length theory for stellar convection. In steady accretion disks, the radial mass flow is constant and the radial distribution of matter is prescribed by the disk viscosity, ν , as:

$$\Sigma(r) = \frac{1}{\nu(r)} \frac{\dot{M}}{3\pi} f(r)^4$$

with $\Sigma(r)$ being the disk surface density. However, hydrodynamical turbulence cannot feed this viscosity since Keplerian disks are stable to the Rayleigh criterion (angular momentum increases radially outward). Magnetohydrodynamical turbulence could play this role if the disk conductivity is high enough to allow the Magneto Rotational Instability (MRI) to set in and feed the turbulence cascade. To operate, MRI requires that the disk is magnetized so the field creates a stress, $-B_r B_\phi / 4\pi$, that couples fluid elements at different radial distances from the star and forces them to exchange angular momentum (Balbus and Hawley 1991). The transfer of angular

momentum from the inner element to the outer element permits the inner element to move to lower orbits while the outer element moves to higher orbits. The magnetic tension acts like a spring that has as rest length the wavelength of the most rapidly growing mode. In principle, this scale corresponds to the length traveled by an Alfvén wave in one orbital period: $2\pi v_A(r)/\Omega$ where $\Omega(r)$ is the ring's angular velocity and $v_A = B/(4\pi\rho)^{1/2}$ is the Alfvén velocity. The finite thickness of the disk sets a vertical cutoff making the longest wavelength equal to $H(r)$, the disk thickness.

If the gravitational field is dominated by the star, the disk thickness can be readily derived as $H(r) = c_s(r)/\Omega(r)$ (or $H(r)/r = c_s(r)/(r\Omega(r))$) where $\Omega(r)$ is given by the third Kepler law⁹. As $c_s = \gamma k_B T / \mu m_H$, with γ the adiabatic index; μ the mean molecular weight, 2.4 as for a molecular H_2 disk; and m_H the mass of a hydrogen atom,

$$c_s(r) = 4.1 \text{ km s}^{-1} \left(\frac{\dot{M}}{10^{-8} M_\odot \text{ year}^{-1}} \right)^{1/8} \left(\frac{M_*}{M_\odot} \right)^{1/8} \left(\frac{r}{r_{in}} \right)^{-3/8} f(r)^{1/2}$$

and thus, a simple expression can be derived for $H(r)$ from the disk temperature law,


$$H(r) = 3.4 \times 10^9 \text{ cm} \left(\frac{\dot{M}}{10^{-8} M_\odot \text{ year}^{-1}} \right)^{1/8} \left(\frac{M_*}{M_\odot} \right)^{-3/8} \left(\frac{r}{r_{in}} \right)^{9/8} f(r)^{1/2}$$

This would set the basic scale for angular momentum transport at any given radius and thus, the efficiency of the mechanism.

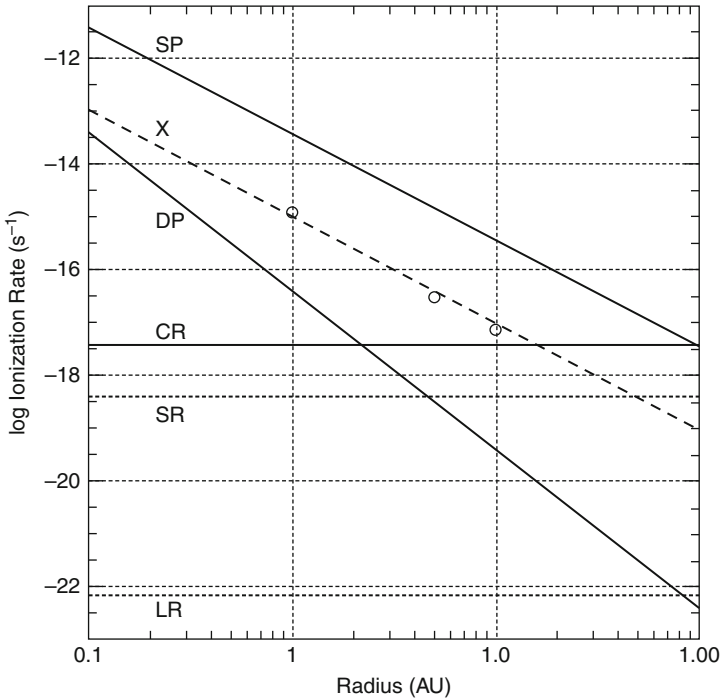
Numerical simulations have been run to track down the nonlinear evolution of the MRI-driven cascade and the energy released into disk heating. The local physics of the disk is simulated in Cartesian boxes sheared by differential rotation where both tidal and Coriolis forces are taken into account, the so-called shearing box. Regardless of the initial field geometry and strength, even for weak field seeds, the simulations show that the magnetic and kinematic fluctuations transport angular momentum at a rate proportional to the magnetic pressure thus, the magnetic stress, $-B_r B_\phi / 4\pi = \alpha_{\text{mag}} \rho v_A^2$, with $\alpha_{\text{mag}} \simeq 0.5 - 0.6$. This prescription is consistent with the α -prescription with $\alpha \sim 0.01$ (Hawley et al. 1995). The effects on the MRI evolution of the disk resistivity, the detailed analysis of the interaction between ions and neutrals in the fluid or the Hall currents have been studied in the shearing box approach (see Hawley 2009 for a recent review). Numerical simulations show that the turbulence depends on the sign of the magnetic field threading the disk and that quiescent and MRI-active regions may coexist in a complex dynamical structure that depends on the disk ionization degree.

The growth of MRI requires that the disk resistivity, η , satisfies $\eta \geq v_A H$. In turn, η depends on the electron fraction as: $\eta = 234 T^{1/2} \chi_e^{-1} \text{ cm}^2 \text{ s}^{-1}$. Thus, the MRI is approximately suppressed when,

$$\frac{v_A H}{\eta} = 7.4 \times 10^{13} \chi_e \alpha^{1/2} \left(\frac{r}{1 \text{ AU}} \right)^{3/2} \left(\frac{T}{500 \text{ K}} \right) \left(\frac{M_*}{M_\odot} \right)^{-1/2} \leq 1$$

i.e., when the electron fraction is roughly below 10^{-13} or the magnetic Reynolds number $R_M = v_A H / \eta < 1$ (Gammie 1996). This electron fraction depends on the main ionization sources and on the disk opacity to the environmental radiation. It is still unclear whether the MRI can be supported over the full disk extent or whether dead zones do exist. The high mass column of the dusty protostellar disks ($\geq 1,000 \text{ g cm}^{-2}$) and the absence of strong internal sources of ionization may cause that dead and undead zones coexist. The ionization rates expected from the various possible contributors are shown in  Fig. 6-10 from Turner and

⁹ $\Omega(r) = (GM)^{1/2} r^{-3/2} = 2\pi \text{ years}^{-1} (M/M_\odot)^{1/2} (r/1 \text{ AU})^{-3/2}$.



■ Fig. 6-10

Ionization rates at column 8 g cm^{-2} versus distance from the star from Turner and Drake (2009). The *dashed line* shows the contribution of stellar X-rays, and the *dotted lines* the decay of the short-lived (*upper*) and long-lived (*lower*) radionuclides. Among the three *solid lines*, the uppermost is for stellar energetic particles, the steepest shows the disk coronal particles, and the *horizontal solid line* is for interstellar cosmic rays. Large surface density is assumed and only radiation striking the top of the disk is included on the diagram. *Circles* indicate the results of the Monte Carlo X-ray transfer calculation by Igea and Glassgold (1999), scaled to a stellar X-ray luminosity of $2 \times 10^{30} \text{ erg s}^{-1}$; in these simulations the disk is illuminated from above – the source of energy is located at $\sim 10 R_{\odot}$ above the disk on the axis. The disk coronal particle estimate corresponds to a fraction of 10^{-4} of the energy released by mass accretion at a rate of $10^{-8} M_{\odot} \text{ year}^{-1}$

Drake (2009)

for a uniform gas column of 8 g cm^{-2} as a function of the distance to the star. Cosmic ray propagation provides uniform ionization rates across the disk that can be enhanced by as much as a factor of a thousand in the proximity of a supernova remnant. The decay of radionuclides' contribution is negligible even for short-lived species (primarily Al^{26} with half-life 0.717 Myr). The main source of ionization within the inner 100 AU is the radiation released in the high-energy processes associated with the star and the disk–star interaction region that penetrates through the disk atmosphere toward the midplane.

Thus, the ionization degree of disks depends on their vertical structure that it is still poorly constraint (see ► Sect. 3.6.2). There are, however, indications that turbulent mixing may be able to prevent the formation of dead zones (Ilgner and Nelson 2008).

3.5.1 The α -Prescription and Angular Momentum Transport by Gravitational Waves

Once massive disk-like structures form (in 1–3 free-fall times), they are especially prone to gravitational waves and instabilities. Gravitational waves can effectively transport angular momentum in a conservative manner, without dissipation, and they are a global phenomenon that propagate over the whole disk as opposite to the α -disk model that is based on the properties of turbulence, i.e., on the *local* dissipation of disturbances.

Gravitational instabilities and waves are a well-studied mean for angular momentum transport via tidal torques (see Lin and Papaloizou 1996 for a review on angular momentum transport mechanisms in disks). The simplest modes are one-armed spirals built from a progressive tilted sequence of elliptical orbits (Adams et al. 1989); however, they do not play a dominant role in YSOs' disks (Durisen et al. 2007). In general, non-axisymmetric disturbances grow as multi-armed spirals that become unstable for $Q < 1.5$ (see ► Fig. 6-11). Dissipation is driven by compression and subsequent heating/cooling of the gas that can be enhanced by shocks driving to a *local* dissipation of energy. Nonlinear wave coupling produces a wide spectrum of waves resulting in a cascade from single gravitational modes to a broad variety of wavelengths and arms. Disks evolving under self-gravity tend to keep around $Q \simeq 1$ (Laughlin and Bodenheimer 1994) and anomalous energy transport terms related with the gravitational shear can become negligible supporting the use of the α -prescription to track the radial angular momentum transport (Balbus and Papaloizou 1999). In this case a gravitational turbulence (or gravito-turbulence) α_g parameter can be defined. As an example, the tuned α -prescription used to treat gravitational collapse in ► Fig. 6-9 is set to keep Q between 1.3 and 1.5 assuming that the disk evolves toward marginal gravitational instability and maintains it during its evolution. This gravitational α or α_g needs to be submitted to rapid variations during the beginning of the gravitational collapse due to the rapid compression; however, after this period α_g stabilizes around 0.05 for a solar mass core.

The α_g parameter can also be used to evaluate the disk stability against fragmentation; gravitational waves can drive local compression that for a given disk opacity renders the disk unstable producing disk fragmentation into protoplanets or brown dwarfs (see Goodwin's chapter). In thermal equilibrium, α_g can be evaluated from the disk cooling time, τ_c , as:

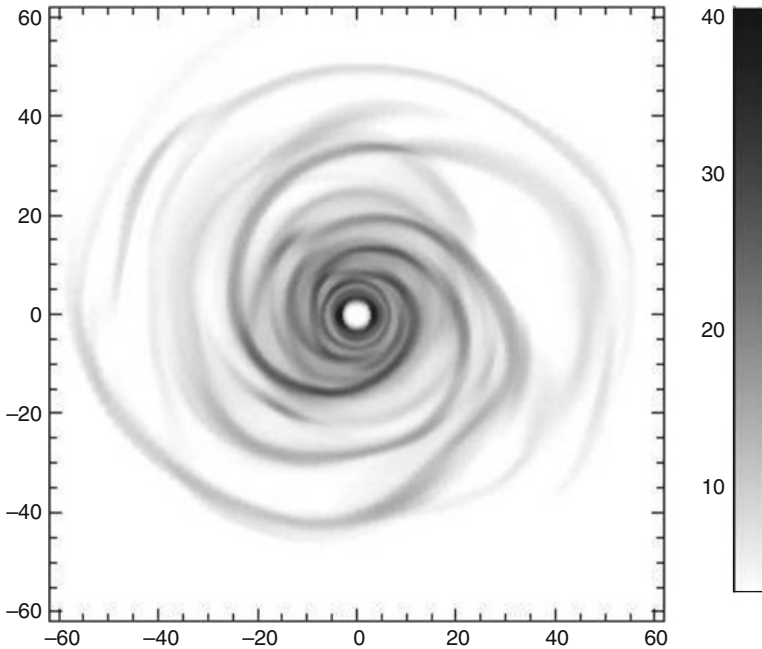
$$\alpha_g = \frac{4}{9\gamma(\gamma - 1)\Omega\tau_c} \quad (6.1)$$

(Gammie 2001). Numerical simulations by Rice et al. (2005) suggest that only if $\alpha_g < 0.06$ angular momentum transport is viable without driving to disk fragmentation.

In general, the disk dispersal timescale would be around 10^7 years if only gravitational instabilities are considered; however disk dispersal timescales derived from the observations are significantly shorter ~ 5 Myr (Haisch et al. 2001). MRI and gravitational instabilities have complementary properties and their joint action, together with the disk irradiation and outflow, could explain the main observed properties of protostellar disks.

3.6 Disk Structure: The Effect of Stellar Irradiation

X-ray and, especially, high-energy particles are able to penetrate deep in the disk releasing their energy within the disk interior. Photoionizing ultraviolet radiation is rapidly absorbed at the



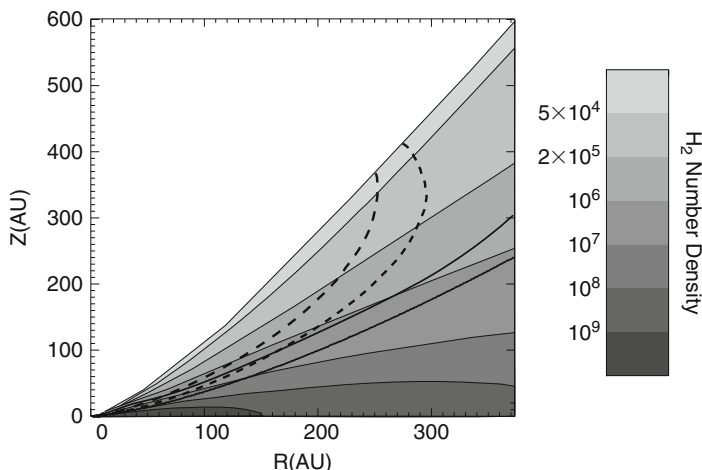
■ Fig. 6-11

Shocks and spiral waves from numerical simulations of gravitational instabilities in YSOs' disk (after Mejía et al. 2005). The *grayscale codes* the effective temperature of the gas in a disk undergoing gravitational instabilities that has reached an asymptotic state of thermal self-regulation. The mass of the disk is $0.07 M_{\odot}$ and it is orbiting around a $0.5 M_{\odot}$ star

surface by the two main compounds of disks: dust and gas. Dust grains efficiently absorb UV photons through the photoelectric effect, gas particles do it through the resonance transitions of both molecules and atoms. A PhotoDissociation Region (PDR) forms at the surface that has its own dynamics and carves the disk surface producing photoevaporative flows. The penetration of the ionizing radiation toward the disk midplane depends on the disk opacity that is dominated by dust grains at temperatures below 1,500 K (i.e., over most of the disk volume). According to the penetration of the stellar high-energy radiation protostellar disks are divided into three basic layers (see ● Fig. 6-12):

- *Dense midplane*: which is opaque to both X-ray and UV radiation and is ionized primarily by cosmic ray particles.
- *Intermediate layer*: which is opaque to UV radiation. The bottom of this layer will be defined by the height at which fractional ionization begins to grow due to the stellar X-rays photons.
- *Atmosphere*: which is transparent to the stellar UV radiation. The bottom of this layer will be defined by the height at which fractional ionization begins to grow due to the stellar UV photons.

The thickness of these layers and the location of the frontiers depends on the details of the protostellar disk model: irradiation field, dust opacity, density, and temperature laws in the disk.



■ Fig. 6-12

Vertical structure of YSOs disks. The *gray levels* mark the number density of H_2 molecules in the disk from the D'Alessio et al. (1998) model. The *thick lines* trace the scale height of the disk for a standard (nonirradiated) α disk – *bottom line* – and for an irradiated disk – *top line*. The *dotted lines* mark the penetration of the UV (1,000 Å) and optical (3,400 Å) radiation from the central source into the disk. Radiation impinging at the disk surface drops by a factor of 10^{-6} at the location of the lines: *top* (UV), *bottom* (optical) (from Van Zadelhoff et al. 2003)

Solving the full structure of the disk is thus, a complex problem since many different processes are coupled.

3.6.1 YSOs' X-ray and UV Radiation Field

The observed X-ray luminosity of YSOs is on average about $10^{30} \text{ erg s}^{-1}$ and the energy distribution is well fitted by the thermal radiation of optically thin plasmas at temperatures of $\sim 5\text{--}30 \text{ MK}$ (see i.e. Guedel 2008); for comparison the Solar X-ray luminosity is about $10^{27} \text{ erg s}^{-1}$ and the plasma temperatures 1–2 MK. In class I and II YSOs, the X-ray spectrum often shows the contribution of two unrelated components subjected to very different absorption: a soft component from a cool (1–2 MK) plasma that is barely absorbed and a much harder component (10–30 MK) subjected to about ten times higher absorption. X-ray-induced iron $\text{K}\alpha$ emission from the disk has been detected from some YSOs, tracing directly the effect of X-ray irradiation on the disk surface (Giardino et al. 2007).

The main effect of X-ray photons is to ionize atomic hydrogen that ends shielding the disk to the penetration of the X-ray radiation. Also, X-ray-induced photons produce the desorption of formaldehyde (H_2CO) from grain surfaces. Chemical evolution models show that this process is fundamental for the disk ionization (and coupling to the magnetic field) since HCO^+ becomes the dominant ion in the intermediate layer of the disk in timescales comparable to the disk evolution timescales (about 10^6 years).

The UV luminosity of YSOs is ill determined. During the early phases heavy dust extinction prevents measuring directly the UV radiation field and only indirect measurements are available based on the chemical modeling of disks. The radiative UV output from massive YSOs (O to A spectral types) can be modeled using blackbody spectral energy distributions but the UV spectrum of low mass YSOs is dominated by the extended magnetosphere/atmosphere and it is poorly constraint from the observations. Typically, line fluxes of TTSs are a factor of 50 stronger than their main sequence analogues (see Gómez de Castro 2009 and references therein). Accretion also contributes to the UV excess of TTS and HAeBe stars (see Sect. 4).

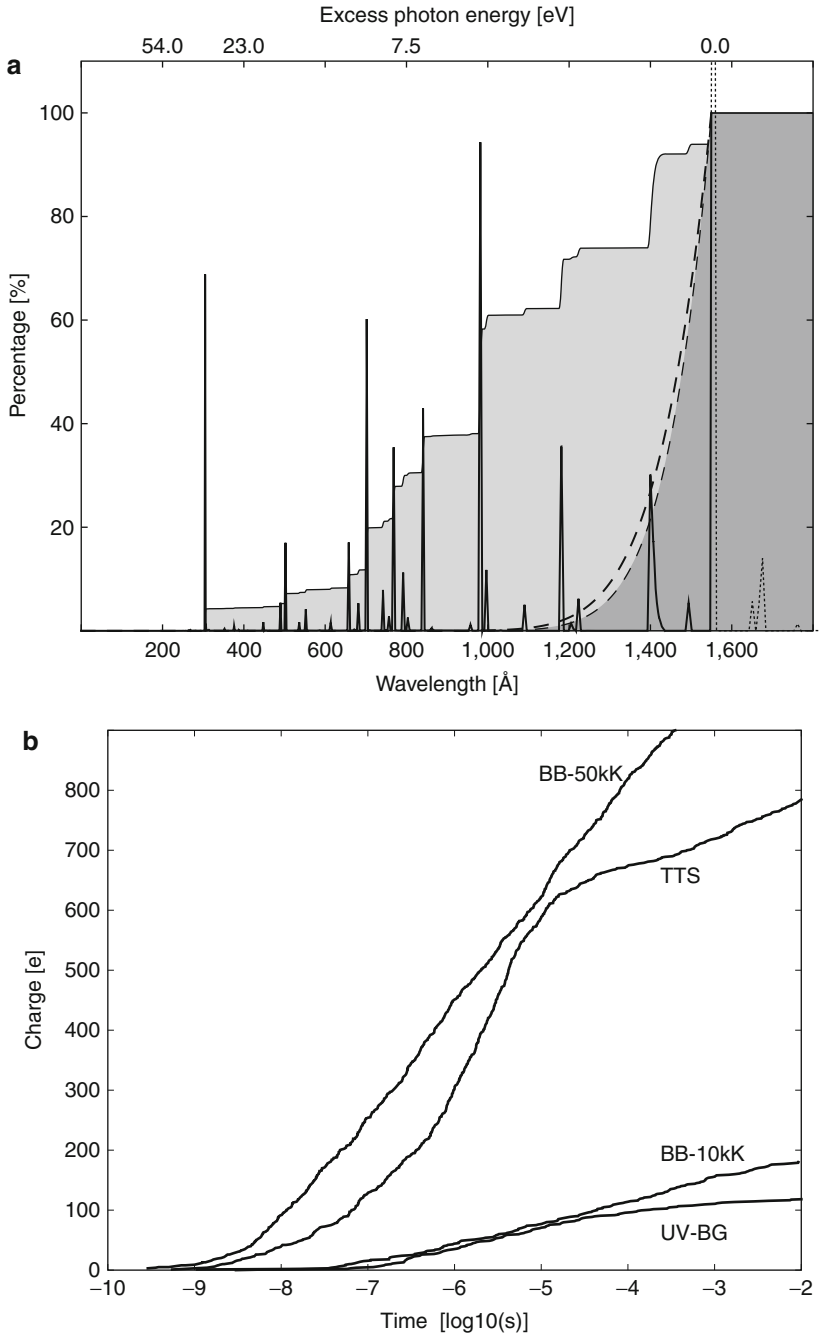
UV radiation interacts strongly with matter in the disk. H_2 dissociates through absorption in the Lyman and Werner bands (912–1,100 Å range). As H_2 is the main component of protostellar disks, its response to the UV radiation field etches the surface of the disk. Unfortunately, very few observations are available of TTSs (Herczeg et al. 2002; Wilkinson et al. 2002) and Herbig stars (Martin-Zaidi et al. 2008) in the 912–1,100 Å range. The absorption of UV photons is followed by UV and IR fluorescence that has been detected from the disk of several TTSs (Herczeg et al. 2004). Only about 15% of the photons lead to H_2 photodissociation. The neutral gas is ionized by the Lyman continuum radiation that heats it to some 10,000 K surpassing the escape velocity from the disk surface (see Sect. 3.8).

UV photons are also absorbed/scattered by the dust and dust grains get charged through the photoelectric effect releasing high-energy electrons that heat collisionally the plasma. The kinetic energy of the photoelectrons depends on the hardness of the illuminating UV spectrum since the dust absorption cross section covers a broad spectral range; dust extinction peaks at about 750 Å and declines toward X-rays (Laor and Draine 1993) with a characteristic absorption cross section in the Lyman continuum as $\sim 0.61 \times 10^{-21} \text{ cm}^2$. The photoelectric charging of the dust grains produces a population of secondary fast photoelectrons in the atmosphere of the disk whose characteristics depend on the illuminating radiation field. In Fig. 6-13a, the most characteristic ionizing spectra in the YSO's environments are plotted. Blackbodies at 50,000 and 10,000 K represent roughly the energy distribution of O-type and A-type YSOs. The fit to the galactic UV background based on early UV missions (Draine 1978) is also plot. Finally, a high-energy spectrum has been built by assuming that the X-ray radiation of the TTSs is produced by magnetic reconnection in a clumpy corona. The interaction of the X-ray photons with the dense ($\sim 10^9 \text{ cm}^{-3}$) clumps produces the UV spectrum plotted. The charging profiles of the dust grains are shown in Fig. 6-13b.

3.6.2 The Vertical Structure of Disks

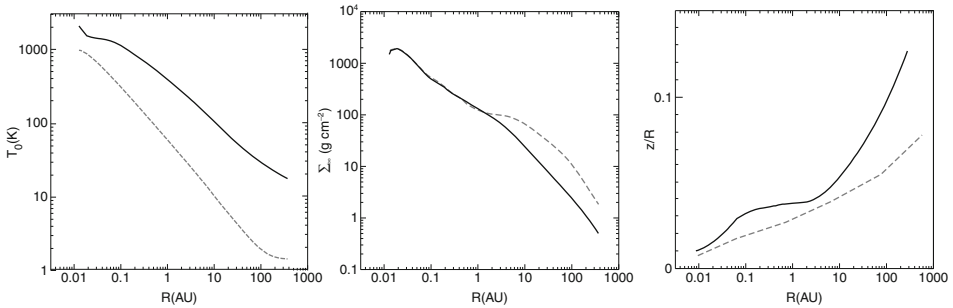
The innermost region of the YSO's disk is heavily disturbed by the interaction with the stellar magnetosphere in TTSs (see Sect. 4) and by the stellar radiation field in HAeBe stars. Further out, disks can be treated as rather simple structures. In the absence of dead zones, most of the accretion energy is expected to be released at the disk midplane that cools mainly by dust continuum radiation in the infrared range. At the same time, the disk is heated from above (from the atmosphere to the midplane) by the high-energy radiation produced in the inner region (star, magnetosphere, inner outflow).

In a first approach, D'Alessio et al. (1998) computed the vertical structure of an stationary α -disk irradiated by a TTS with effective temperature 4,000 K. Cosmic rays and radioactive decay were also considered as sources of disk ionization. At every given radius, the disk is treated as a stratified, plane-parallel atmosphere. Viscous dissipation in the α -disks occurs locally, and



■ Fig. 6-13

Response of dust grains to typical UV radiation fields in the YSO's environments. (a) UV radiation fields normalized to the same total energy of a 50,000 K blackbody (*shaded continuous line*), a 10,000 K blackbody (*dark shaded dashed line*), the UV background (*dashed line*), and the hard TTS spectrum (*continuous line*). (b) Evolution of the grain charge with time at 1 AU from the star (from Pedersen and Gómez de Castro 2011)



■ Fig. 6-14

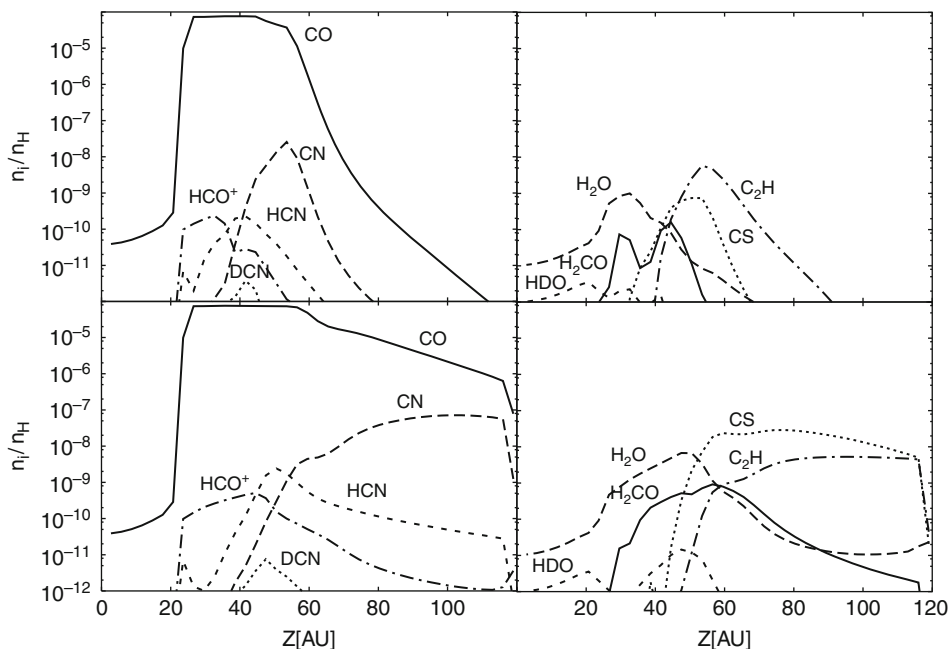
Vertical structure of an irradiated disk (solid lines) compared with a standard accretion α -disk (dashed lines) for an accretion rate of $10^{-8} M_{\odot} \text{ year}^{-1}$ onto a $0.5 M_{\odot}$ star (from D'Alessio et al. 1998). Left panel: Temperature at the disk surface, T , for an irradiated disk (solid line) and a nonirradiated disk (dotted line). Central panel: Radial distribution of surface density of an irradiated (solid line) and a nonirradiated (dashed line) disk model. Right panel: Characteristic height as a function of R

the energy rate generated per unit volume is computed assuming $\alpha = 0.01$ and an accretion rate of $10^{-8} M_{\odot} \text{ year}^{-1}$. Stellar photons are assumed to enter in the disk at grazing angles as in a flared disk.¹⁰ A fraction of the stellar radiation is scattered at the upper boundary of the disk creating a diffuse field at the disk surface; it is assumed that gas and dust are thermally coupled in the disk.

With this provision, this simple model already shows that irradiation from the central star is the dominant heating agent farther than some 5 AU. As a result the temperature distribution of the disk changes from $T(r) \propto r^{-3/4}$ to $T(r) \propto r^{-1/2}$ and the disk becomes less dense than in the purely viscous case. In this model, the disk temperature has a vertical inversion caused by the properties of the dust opacity. Dust is more opaque to the optical stellar spectrum than to the infrared radiation produced by itself to release the viscous heating. This creates a superheated layer above the disk photosphere where most of the stellar radiative energy is spent into dust heating. In ► Fig. 6-14, the temperature law, surface density, and scale height are compared for irradiated and nonirradiated accretion disks.

The D'Alessio et al. (1998) model has been extensively used as a baseline to study disk chemistry and ionization state. For instance, Van Zadelhoff et al. (2003) used the density and temperature structure of the disk derived from 1-D modeling to compute in more detail the propagation of UV radiation into the disk interior. From this, the vertical distribution of molecules is computed for different disk radii and stellar irradiation fields. As shown in ► Fig. 6-15, the vertical distribution of some molecules is very sensitive to the UV irradiation field. In the absence of a significant UV radiation field, the CO photodissociation rate becomes small and the molecule survives further in the disk atmosphere. As a result, the column densities of abundant radicals such as CN or C_2N decrease in the inner regions due to the smaller supply of atomic carbon. The chemical timescale is very short in the disk midplane since adsorption on dust grains dominates the chemistry $\sim 100 \text{ yr}$ ($10^8 \text{ cm}^{-3} / n_H$); however, in the intermediate layer steady state is

¹⁰In disk modeling, it is often assumed that the stellar photons enter the disk surface at grazing angles given by $\phi \approx h_{*}^{\text{eff}} / (r + dH/dr - H/r)$, where h_{*}^{eff} is the effective height of the stellar radiation field over the disk midplane, often taken as $0.4R_{*}$.



■ Fig. 6-15

Vertical distribution of molecules at $R = 105$ AU and $t = 1$ Myr structure of an irradiated disk (after Van Zadelhoff et al. 2003). *Top*, the disk is assumed to be irradiated by the spectrum of a TTS from 1,000 to 10,000 Å. *Bottom*, the disk is assumed to be irradiated with Draine's (1978) spectrum

reached in 10^4 – 10^5 years for many organic molecules and this timescale is significantly larger than the UV flux variability timescales (days to years).

The ionization structure of the disk has also been calculated based on this simple 1-D model (Semenov et al. 2004) who found a dead region between 2 and 20 AU where the ionization fraction is not high enough to keep the gas coupled to the magnetic field. In this area, MRI would not grow keeping the disk midplane cooler than in the standard α -model and preventing mass transport. This raised concerns on the applicability of the MRI to the cool and dense protostellar disks and drove to the development of new disk models that take into account vertical mixing between the midplane and the irradiated disk surface. Ilgner and Nelson (2008) have made use of a simplified chemical model for the gas phase chemistry to code in the ionization fraction evolution in a sheared box-based model of a magnetized accretion disk. The chemical model is based on charge exchange reactions between H_2 and Mg (Oppenheimer and Dalgarno 1974) that are well suited to follow the electron fraction in the disk (Ilgner and Nelson 2006). For a fiducial temperature of 100 K, the charge transfer reaction ($H_2^+ + Mg \rightarrow H_2 + Mg^+$) occurs at a rate that is about 1% of the H_2^+ recombination rate. As the recombination rate of Mg^+ is five orders of magnitude lower than of H_2^+ , a Mg-rich gas will sustain better a higher ionization fraction favoring MRI development. The model does not take into account the ionization of the disk surface by UV radiation and considers a Mg abundance slightly below the solar. With these provisions, the numerical calculations led to a fully turbulent disk with no dead zones.

A full disk modeling including the MRI cascade to turbulence, disk irradiation, and the microphysics of the dissipative processes is yet to be done.

3.7 Disk Stability

Some TTs are observed to suffer nova-like outburst. They are named FUOrs as the prototype of the class: FU Ori. During the outburst, the system brightens by several magnitudes (≥ 4 magnitudes in V) followed by a decade-long fading (see Abraham et al. 2004 for a compilation of these sources). The transport rate typically rises to $10^{-4} M_{\odot} \text{ year}^{-1}$ during the outburst (Hartmann and Kenyon 1996).

Disk emission dominates at all the wavelengths during outburst suggesting that YSOs' disks are prone to strong instabilities. Disk instability triggering can be intrinsic to the disk or extrinsic. Extrinsic instabilities are driven by the close passage of nearby stars or the presence of a companion in a wide, highly eccentric orbit. Accretion disks may also render unstable by thermal instabilities and frontier instabilities related with the coexistence of two transport mechanisms (MRI and gravitational waves) in the disk.

Thermal instabilities are driven by strong enhancements of the disk opacity as the temperature rises locally. The α -prescription relates the accretion rate with the local disk temperature; thus, as the temperature rises, the effective local mass flux also increases and the hot regions begin to propagate radially through the disk on a timescale $\tau_p \sim r/\alpha c_s \sim 5 \text{ year} (r/0.1 \text{ AU}) / (c_s/10 \text{ kms})^{-1} \text{ years}$ for $\alpha = 10^{-2}$. During the outburst, the ionized regions deplete on the viscous diffusion timescale, $\sim r^2/\nu \sim (\Omega/c_s) \tau_p$, that is about ten times the propagation timescale. The opacity of protostellar disks is not a smooth function of the temperature but shows strong jumps between the dust melting point and the onset of scattering by atomic hydrogen as the main source of opacity. After dust grains melting (above $\sim 1,500 \text{ K}$), the opacity drops by about four orders of magnitude; however as the temperature increases from 2,000 to 10,000 K the opacity rises by eight orders of magnitude(!) by the onset of hydrogen scattering. Thus, a variation on the accretion rate enough to drive the inner disk to the dust melting point have strong effects on the disk structure. Bell and Lin (1994) developed the first models for this process based on work done for outburst generation in accretion disks within cataclysmic variables. However, the model requires very high accretion rates and very low $\alpha (\simeq 10^{-4})$ to reproduce the observations that are unusual for protostellar disks (see Bell and Lin 1995; Turner et al. 1997 for later developments).

Another unstable regime can be built in the disk if there is a physical boundary separating an MRI-dominated accretion flow in the inner disk from a gravitational instability-dominated flow in the outer disk. Material could pile up at this border and cause the temperature in the inner disk to rise to remain gravitationally stable (to keep $Q \sim 1$). Once hot enough, the increase of the ionization fraction would onset the MRI producing an accretion outburst (Armitage et al. 2001).

3.8 Disk Evolution

Two competing scenarios have been proposed to drive disk evolution:

- *Cool evolution* driven by gravity
- *Hot evolution* that takes into account the effect of the stellar radiation field into disk photoevaporation

Cool evolution is driven by the exhaust of remnant material in the molecular core to feed the disk with fresh gas for accretion. Typical models follow the evolution of the rotating protostellar core

under self-gravity with some simplifying assumptions to make the problem tractable (Yorke and Bodenheimer 1999; Yorke and Sonnhalter 2002; Young and Evans 2005). The energy sources are identified: star (accretion and deuterium burning), disk (accretion, direct infall of gas from the envelope), and star–disk boundary layer. Radiative transfer within the dusty envelope is computed and from that, the main observables are determined. The two main observables are the so-called bolometric temperature of the YSO and its bolometric luminosity. The bolometric temperature is defined as the temperature of a blackbody with the same flux-weighted mean frequency as the source:

$$T_{\text{bol}} = 1.25 \times 10^{-11} \frac{\int_0^\infty \nu S_\nu d\nu}{\int_0^\infty S_\nu d\nu}$$


T_{bol} tracks the evolution of the SED, from the heavily weighted to the far infrared SEDs of class 0 YSOs, to the optically weighted SEDs of class III YSOs, already very close to the zero age main sequence (ZAMS). The boundaries of the YSOs' classes were redefined according to T_{bol} (Chen et al. 1995) as follows:

Class 0: $T_{\text{bol}} < 70 \text{ K}$

Class I: $70 \text{ K} \leq T_{\text{bol}} \leq 650 \text{ K}$

Class II: $650 \text{ K} < T_{\text{bol}} \leq 2,800 \text{ K}$

Class III: $T_{\text{bol}} > 2,800 \text{ K}$

The evolutionary tracks and the class boundaries are plotted in  Fig. 6-16 as well as the ZAMS (from Dunham et al. 2010). The 1,024 YSOs detected in a major infrared survey of star forming regions (Evans et al. 2009) is overplotted. The evolutionary tracks for 1 M_\odot and 3 M_\odot protostellar cores are represented for two families of models, Young and Evans (2005) and Myers et al. (1998), where accretion rates are prescribed to decrease exponentially with time. The low luminosity of class 0 sources evidence unexpectedly low accretion rates $\leq 10^{-6} M_\odot \text{ year}^{-1}$ suggesting that accretion could be episodic in nature since the very early phases.

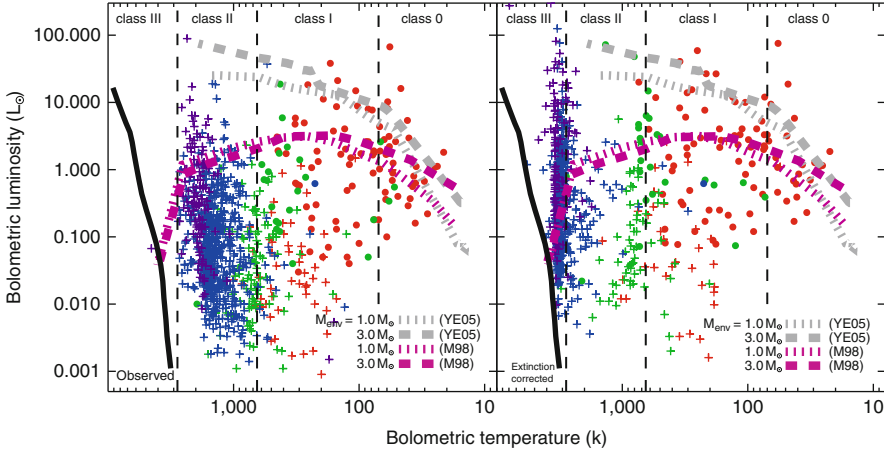
Hot evolution: the impact of the stellar radiation field into disk evolution is taken into account. UV radiation is very efficient at etching the disk surface by producing a thermal flow. The tenuous gas in the disk atmosphere is ionized and heated to temperatures of $\sim 10^4 \text{ K}$. The thermal velocity of the gas reaches typically $V_{th} \sim 11.4 \text{ km s}^{-1}$ leading to a slow photoevaporative flow at radii, r_s , where the thermal speed surpasses the escape velocity from the disk surface: $r_s = 15.6 \text{ AU} \left(\frac{M_*}{M_\odot} (V_{th}/11.4 \text{ km s}^{-1})^{-2} \right)$. When the dynamics of the process is analyzed, it is shown that the flow crosses the sonic point if $r > 0.1 r_s$ (Font et al. 2004). The mass-loss rate in photoevaporative outflows can be roughly estimated as:

$$\dot{M} \sim 4 \times 10^{-10} M_\odot \text{ year}^{-1} \left(\frac{C_{\text{Lyc}}}{10^{41}} \right)^{1/2} \left(\frac{M_*}{M_\odot} \right)^{1/2}$$

where C_{Lyc} is the count rate in the Lyman continuum (Hollenbach et al. 1994). However, as soon as hydrogen recombines, the flow itself may shield the upper layers of the disk atmosphere to the UV radiation. A measure of the opacity of the photoevaporative flow to Lyman continuum photons can be given by the parameter:

$$\Psi = \frac{\alpha_B C_{\text{Lyc}} \delta r \phi}{4\pi r^2 V_{th}^2} = 1.1 \times 10^2 \frac{C_{\text{Lyc}}}{10^{41}} \frac{\delta r}{1 \text{ AU}} \left(\frac{r}{1 \text{ AU}} \right)^{-2}$$

where α_B is the case B hydrogen recombination coefficient and δr is the thickness of the photoevaporative flow in the direction of radiation propagation (Bertoldi 1989). Significant photoevaporative flows require $\Psi \gg 1$.



■ Fig. 6-16

Cool evolution of YSOs represented in a bolometric luminosity versus bolometric temperature diagram (after Evans et al. 2009). Class I YSOs are plotted in *red*, Flat in *green*, Class II in *blue*, and Class III in *purple*. *Filled circles* indicate sources associated with envelopes as traced by millimeter continuum emission, while *plus signs* indicate sources with no such associations. The *thick lines* are model tracks for initial core masses of 1.0, and 3.0 M from Young and Evans (2005) and Myers et al. (1998) marked in *gray* and *purple*, respectively. The *vertical dashed lines* show the boundaries between classes, as defined by Chen et al. (1995). The heavy *dashed line* on the left is the ZAMS D'Antona and Mazzitelli (1997) from 0.1 to 2 M_{\odot} . Data are represented without and with extinction correction in the *left* and *right panels*, respectively

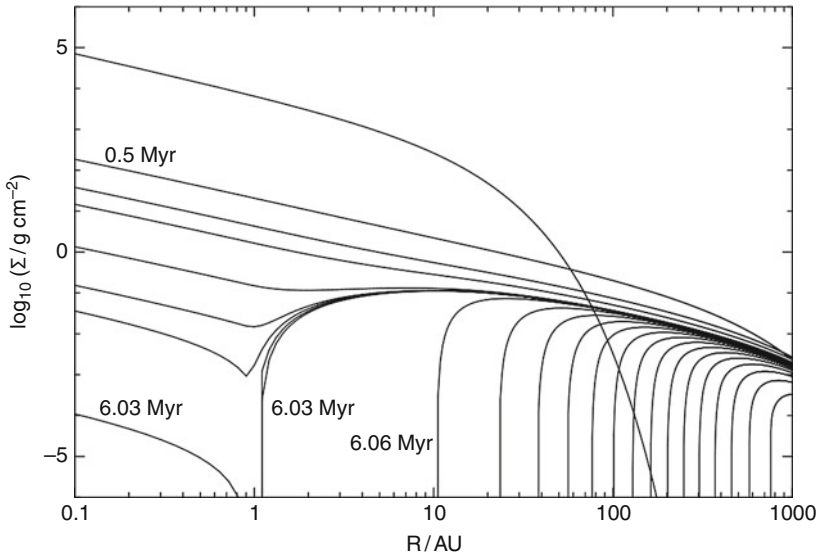
Once the irradiation field photoevaporates the inner disk, viscous spreading tends to move matter to this area accelerating disk dissipation. The evolution of disks is modeled through the diffusion equation for the disk surface density, $\Sigma(r, t)$:

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r} \frac{\partial}{\partial r} \left(r^{1/2} \frac{\partial}{\partial r} \left(\nu \Sigma r^{1/2} \right) \right) + \frac{d\Sigma_w}{dt}$$

where the vertical structure is neglected and an extra term is added to the initial Lynden-Bell and Pringle (1974) formulation to account for the impact of photoevaporative flows. As shown in ● Fig. 6-17 from Alexander et al. (2006), there is a rapid transition in the disk structure at ~ 6 Myrs. Once the inner hole is excavated the spread of the inner cavity drives to full disk photoevaporation in few 10^4 years.

4 PMS Evolution: The Outflow

Outflow is the first signature of star formation however, no theory anticipated it prior to the discovery of the bipolar molecular outflow from L1551 by Snell et al. 1980. A decade later, Andre et al. (1990) detected the VLA 1623 outflow from a Class 0 YSO, pointing out that bipolar outflows are ejected as early as during *Phase II*. The collimation and mechanical power of the



■ Fig. 6-17

Hot evolution of YSOs' disks after Alexander et al. 2006. Photoevaporation drills an inner hole in the disk in roughly 1 Myr. The curves represent the evolution of the disk surface density, from an initial value of $\sim 10^5 \text{ g cm}^{-2}$ in the inner border of the disk to a final state with $\sim 10^4 \text{ g cm}^{-2}$. Notice that in 6 Myr, the area at $R = 1 \text{ AU}$ is evacuated

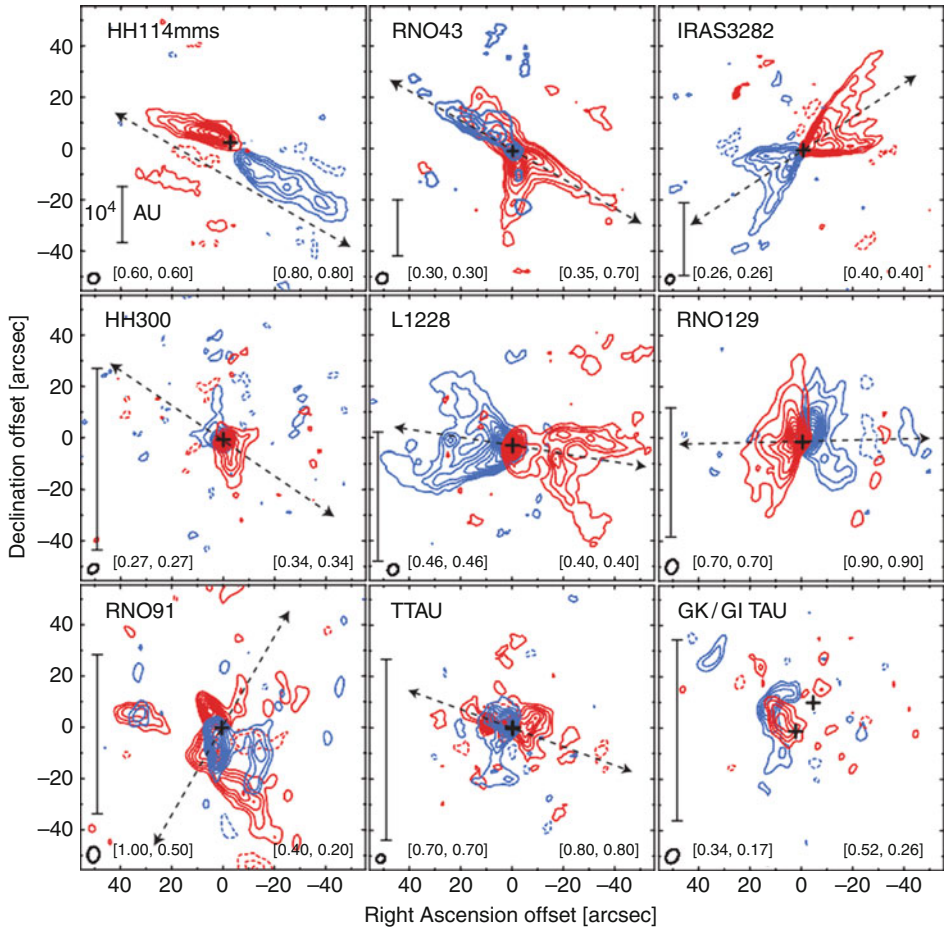
outflows decreases as stars approach the main sequence and protostellar disks evolve into debris disks as shown in ► Fig. 6-18.

The tracers of the outflows depend on the degree of evolution of the parent source. Flows from young class 0/I sources often contain molecules such as H_2 , CO, and SiO. Later on, they tend to be dominated by HI and low-ionization metals that show as prominent large-scale jets detected at optical wavelengths ([S II], [N II], or H_α lines). Mass-loss rates decline from $\sim 10^{-6} M_\odot \text{ year}^{-1}$ in the class 0 YSOs, to $\sim 10^{-8} M_\odot \text{ year}^{-1}$ in class I, and $\sim 10^{-10} M_\odot \text{ year}^{-1}$ in class II. Typically, the outflow mass-loss rate is about a 10% of the accretion rate and the velocity about few hundred km s^{-1} .

The main mechanical properties of bipolar outflows have been recently redetermined from the large data base of known CO outflows (Wu et al. 2004):

- The mass carried out by the outflows, $M_{w,\text{CO}}$, correlates strongly with the bolometric luminosities of the outflow powering source, L_{bol} . The flow mass increases with the luminosity as a power law: $\log M_{w,\text{CO}} = (-1.04 \pm 0.08) + (0.56 \pm 0.02) \log L_{\text{bol}}$.
- The mechanical force needed to accelerate the outflow ($F_{w,\text{CO}} = \dot{M}_{w,\text{CO}} V_{w,\text{CO}}$) is larger than the radiative force from the parent YSO: L_{bol}/c . There is, however, a clear correlation between mechanical and radiative power from the source: $\log F_{w,\text{CO}} = (-0.92 \pm 0.15) + (0.65 \pm 0.04) \log L_{\text{bol}}$.

These relations (see ► Fig. 6-19) indicate that accretion, the major source of energy during star formation, powers the formation of outflows. Moreover, these relations hold from low mass



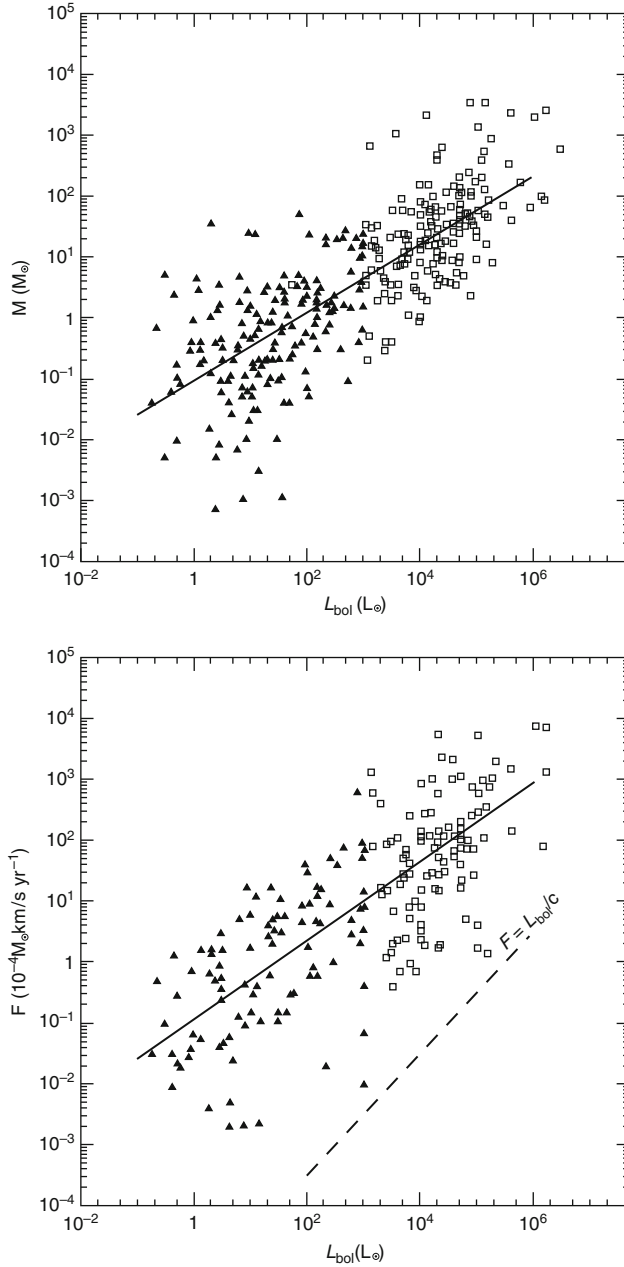
■ Fig. 6-18

Molecular outflows from Class 0 (top), I (middle), and II (bottom) sources after Arce and Sargent (2006)

to early B protostars suggesting a common accretion/outflow mechanism over a wide range of masses.

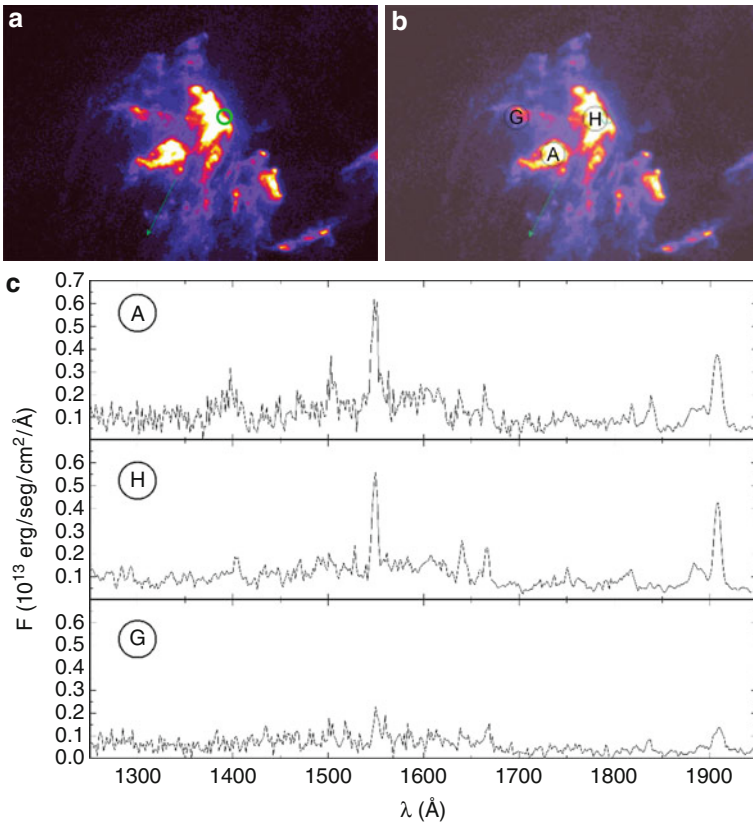
YSOs' outflows transfer momentum and entrain material from their surrounding molecular core/cloud by means of shock waves that propagate into the medium. Typical velocities are about few hundred km s^{-1} and produce prominent shock-excited nebulosities known as Herbig–Haro objects (or HHOs, see Reipurth 1999 for a catalog). HHOs are classified according to their electron temperature into two main groups: low excitation HHOs with $\log(T_e(K)) \leq 4$ and high excitation HHOs with $\log(T_e(K)) \geq 4.5$.

Low excitation HHOs are the result of mild oblique shocks mainly produced along the jet beam, as material from the surrounding cloud is entrained into the flow (see ● Fig. 6-20). HHOs' gas enters the shock front at several tens of km s^{-1} and most of the kinetic energy is damped into the excitation of the H_2 vibrational bands (in the infrared) as well as in radiation



■ Fig. 6-19

Top: CO outflow mass versus the bolometric luminosity (L_{bol}). *Bottom:* Outflow force F versus bolometric luminosity (L_{bol}) of the associated infrared sources. Low mass YSOs and high mass YSOs are represented with *filled triangles* and *open squares*, respectively. The fits given in the text are represented in the figures (Wu et al. 2004)



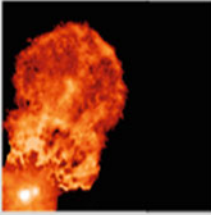
■ Fig. 6-20

(a) The high excitation HHO: HH2. The location of the X-ray emission is marked with a *green circle*; the source of the flow is located in the direction traced by the arrow. The main knots in HH2H are marked in (b) and their UV spectrum (from the archive of the International Ultraviolet Explorer) is plotted in the bottom panel (c)

from the neutral/singly ionized species ([O I] and [S II] are among the most prominent optical tracers). The electron densities are $10\text{--}10^2\text{ cm}^{-3}$ and the gas is only partially ionized with typical electron fractions of 0.1. Most of the molecular emission in the large-scale CO flows is also produced by gas entrained and accelerated in mild shocks. However, it is still unclear whether there is an additional, large scale, slow component made of rarified gas to the bipolar flow (Bachiller and Tafalla 1999).

High excitation HHOs mainly trace the bow shock or the working surface of the jet in high density environments. The kinetic energy of the jet is damped into heating producing very high temperatures that may reach $\sim 10^6$ K as observed in HH 2 (see ● Fig. 6-21). They may radiate from X-rays, in the most extreme cases, to the high excitation lines in the ultraviolet (i.e., He II or N V) and in the optical (i.e., [O III]). The electron densities are typically $10^3\text{--}10^4\text{ cm}^{-3}$ and the gas is fully ionized. However, the predictions of strong radiative shock models are unable to reproduce fully the observations (Raymond et al. 1997), though this could be caused by the poor

Binary

XZ Tau ($d=140$ pc)HH 111 ($d=470$ pc)

Class 0

HH 49/50 ($d=150$ pc)HH 47 ($d=450$ pc)HH 211 ($d=300$ pc)

Bow shock

HH 34 ($d=470$ pc)


Irradiated jet

Jet in Trifid nebula ($d=2,280$ pc)

■ Fig. 6-21

Images of YSOs' jets. Jets from binary systems as XZ Tau or HH111 show wiggles and strong episodic ejections. Jets from class 0 sources are more deeply embedded in the cloud and interact strongly with the environment. Jets in clearer environments are highly collimated and produce bow shocks at the locations where the environmental conditions change. Finally irradiated jets are observed in several regions as Orion or the Trifid allowing a more reliable determination of the ionization fractions (Bally et al. 2006; Heathcote et al. 1996; Hester and Desch 2005; Krist et al. 1997; Reipurth et al. 1999; Tappe et al. 2008)


spatial resolution of the UV spectroscopic observations and the high clumpiness of HHOs. As an example, the UV-optical monitoring of HH 29 allowed to measure the inhomogeneity of the flow (Liseau et al. 1996) driving to a two-phase model of HHOs with a hot and dense component with $T = 10^5$ K and $n_e = 10^6 \text{ cm}^{-3}$ embedded into low excitation component at $T = 10^4$ K and $n_e = 10^3 \text{ cm}^{-3}$. The estimated filling factor of the high excitation component was very small, barely $\sim 0.1\text{--}1\%$.

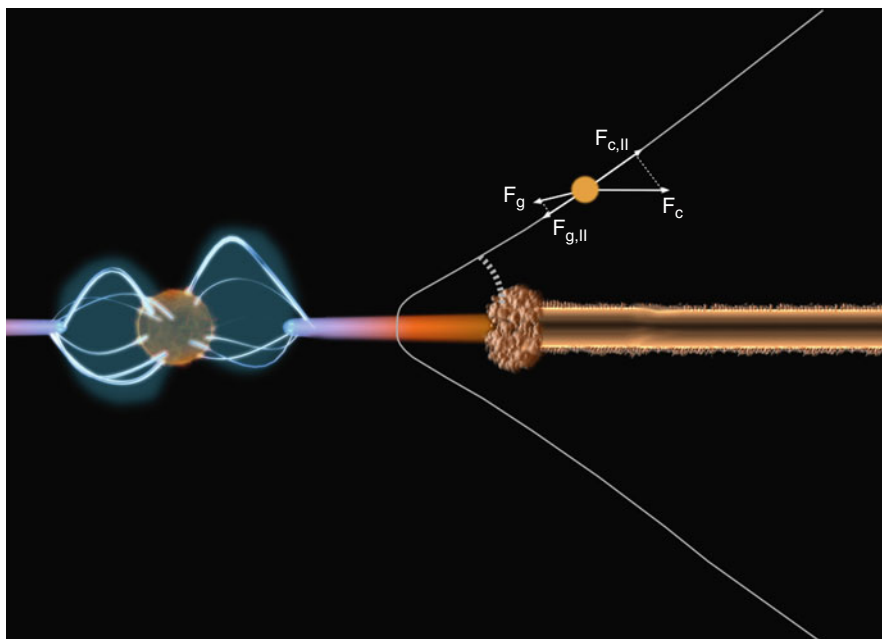
Jets as large as $1\text{--}10$ pc have been observed (i.e., the 10 pc flows detected in Orion; Reipurth et al. 1998). They provide the chance to study the midterm evolution of outflow (and accretion) on timescales of $5\text{--}50 \times 10^3$ year with a time resolution of ~ 10 years. As shown in  Fig. 6-21, episodic ejections are common to all the YSOs' outflows. In fact, the collimated jets seem to result of the overlap of several shocked structures or knots. High activity and quiescent periods alternate during the flow evolution. Precession has been detected in a few sources (i.e., Cep E, Eisloffel et al. 1996) as well as bendings and misalignments suggesting variations in the outflow ejection direction. There are also indications of outflow rotation (Coffey et al. 2004). Evidence of magnetic fields have been reported for the outflow from T Tau S (Ray et al. 1997) and from several radio sources (Curiel et al. 1993).

4.1 First Steps to a Mechanism for Outflow Generation: Disk Winds

The high kinetic energy of the outflow requires an efficient launching mechanism that acts from the beginning of *Phase II* powered by accretion. The low effective gravity of the disk, the increasing centrifugal force as the star grows, and the disk ionization by the radiation from the accretion shocks creates the ideal environment to develop centrifugally driven outflows from the disk along the magnetic field lines permeating it.

Centrifugally driven outflows were first proposed in the context of Solar System research. The Sun is able to accelerate outflows from its slowly rotating atmosphere to $\sim 400 \text{ km s}^{-1}$ at the Earth orbit. These outflows depart through open holes in the magnetic configuration of the Sun. As the solar material is magnetized and, hence, firmly attached to the field lines, it leaves the Sun speeding up along the field lines like the water drops ejected from the nozzles of a rotating sprinkler. This magnetized outflow rotates rigidly with the Sun as far as the Alfvén radius, where the flow speed surpasses the Alfvén speed, i.e., the velocity at which information is transported in the magnetized flow (Weber and Davis 1967). Afterward, the flow evolves freely at this high speed. An important aspect of centrifugally driven winds is that they transport angular momentum away from the star. As the gas is forced to corotate with the star up to the Alfvén radius, r_A , it transports a specific angular momentum, $l_A = \Omega_* r_A^2$.

Blandford and Payne (1982) developed a version of centrifugally driven MHD winds for accretion disks. In their model, each ring of the disk is treated as a separate entity that rotates with Keplerian velocity. The disk is assumed to be permeated by an external field entering the disk in grazing incidence, as shown in the  Fig. 6-22. Matter is loaded onto the lines and accelerated by the centrifugal lever arm. As the Solar MHD centrifugally driven winds, disk winds also carry a specific angular momentum for each ring given by $l_A = \kappa \Omega_{K,A} r_A^2$, where $\Omega_{K,A} r_A^2$ is the specific angular momentum at the Alfvén point and κ is a factor that depends on the precise geometry of the solution and ranges between 14 and 19 for simple self-similar solutions. Notice that the lever arm for angular momentum removal by MHD centrifugally driven winds is r_A , significantly larger than the lever arm for viscous dissipation within the α -prescription,



■ Fig. 6-22


Sketch of the inner structure of low mass YSOs. The *inner disk cavity* filled in with hot gas is constrained between the stellar magnetosphere and the puffed up dusty rim that marks the separation between the heavily irradiated disk and the cool dusty disk. Disk winds depart along field lines that enter the disk in grazing angles acting like ramp speeds. If the centrifugal force projected along the field line is larger than the gravitational force, matter leaves the disk along the field lines and it is accelerated as far as the Alfvén radius

just αH . As a result, disk winds are an efficient mean to transport angular momentum and build up rapidly transport in YSOs' disks.

Disk winds extend the nested rings set up of the α -prescription for YSOs' disks into a nested Russian-dolls scheme that includes the large-scale outflow as a set of plasma sheets anchored in each ring (see ● Fig. 6-23). The velocity of matter flowing along this plasma sheet is usually projected into the three cylindrical components: rotation around the disk axis, expansion away from the rotation axis, and collimated motion along the axis. These components and their manifestation into unresolved line profiles are shown also in the figure. The dominant kinematics at the base of the outflow is radial expansion that produces very broad double-peaked profiles, rotation is only dominant very close to the Alfvén point and produces narrower double-peaked profiles, and further up, the dominant velocity is motion along the disk axis with a velocity $\sim \Omega_{K,A} r_A$.

Ferreira and Pelletier (1993, 1995) attached for the first time disk winds to α accretion disks. An interesting outcome is that outflows can only be generated in a very narrow range of plasma

β s¹¹ with $2.84 < \beta < 3.12$ that only marginally intersect with the range where the MRI works; MRI ceases for $\beta > 3$.

The temperature of the outflow is regulated by three processes: (1) the rapid adiabatic radial expansion at the base, (2) heating by ambipolar diffusion, and (3) photoionization by the star. Ambipolar diffusion heating depends on the electron fraction of the wind. The higher the electron fraction, the smaller the heating. In YSOs' outflows, this results in an equilibrium temperature of about 10,000 K similar to that observed in the optical jets (Garcia et al. 2001; Safer 1993). However, if wind photoionization by the stellar X-ray photons is taken into account, the base of wind can reach temperatures of few thousand Kelvins (Ferro-Fontán and Gómez de Castro 2003). As shown in the  Fig. 6-24, disk winds shield effectively the disk surface, protecting it from the high-energy X-ray photons from the star. In fact, they act as a protective layer that transforms the hard energy from the star into a milder (1,000–3,000 K) radiation field before it can reach the disk surface providing a wide spread source of heating of the disk atmosphere.

Disk winds are kept collimated by the pinching stress derived from the Lorentz force $\vec{J} \times \vec{B}$, where \vec{J} is the current driven by the flow, mainly along the disk axis, and \vec{B} is the magnetic field in the flow, mainly toroidal around the disk axis. This is a rich source of possible plasma large-scale instabilities, especially as flows depart from several radii in the disk. Pelletier and Pudritz (1992) showed that if the magnetic flux in the disk goes like $r^{-1/2}$, disk winds are stable since the same current is carried at all radii.¹² Typically the mass-loss rates estimated from disk winds modeling are about 10% of the accretion rate, in good agreement with the observational estimates.

YSOs' observations show that jets are not smooth laminar flows; instead, they show variable structures and knots. Some attempts were made in the past to reproduce this behavior through disk winds. Gómez de Castro and Pudritz (1993) showed that the unresolved forbidden line emission from jets could be produced by shocks at the magnetic refocusing points of disk winds. Garcia et al. (2001) and Ferreira and Casse (2004) attempted to reproduce the large-scale properties of jets by warm disk winds. In general, the coexistence of flows at different velocities from different disk radii, the interaction with the environment, and the pinching stresses are a rich source of oblique shocks and nonlinear interactions, especially amenable to numerical simulations, that can reproduce a large wealth of the observed phenomenology (see Pudritz et al. 2007 for a recent review in the field).

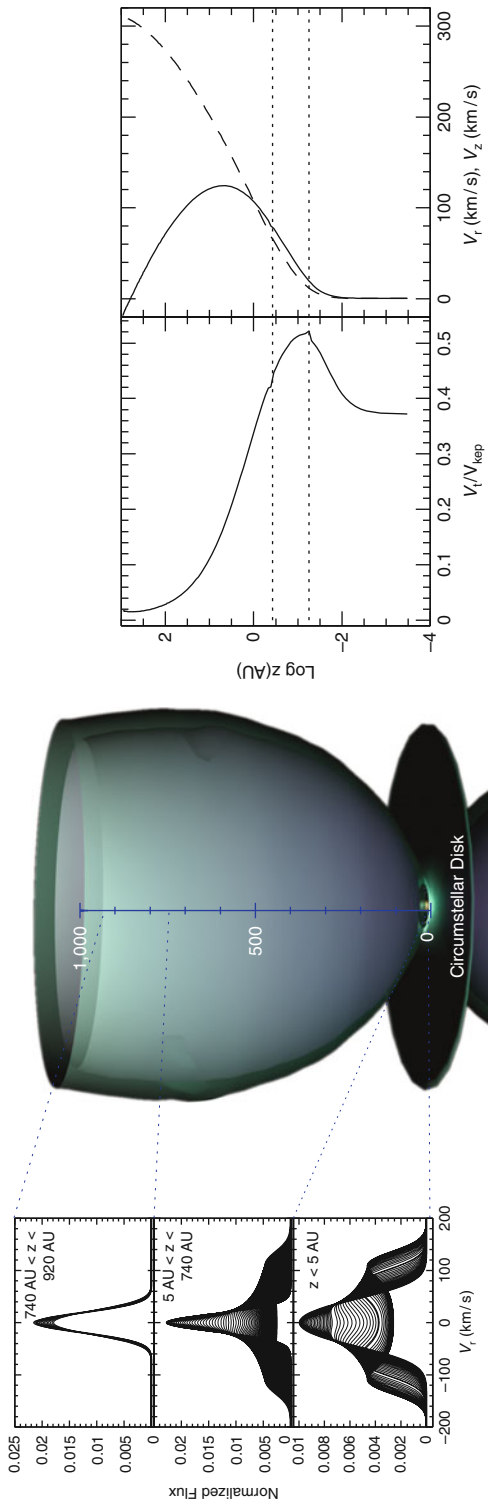
However, the observations clearly show that YSOs' bipolar outflows have a strong episodic component. The recognition that the interaction between the magnetosphere of a TTS and the disk would generate in a natural manner episodic ejections drove to the current paradigm for YSOs' outflow generation.

4.2 The Star–Disk Interface: The Jet Engine

Monitorings to derive the rotation rates of YSOs during *Phase III* revealed the existence of hot spots at temperatures of about 10^4 K on the surface of some TTSs (Bouvier 1990). This was further confirmed by UV monitorings (see Gómez de Castro 1998 for a review) raising the temperature of the spots to some 40,000 K upon the detection of the flux modulation in the CIV line. In 1991, Koenigl showed that hot spots would be naturally produced if the accretion

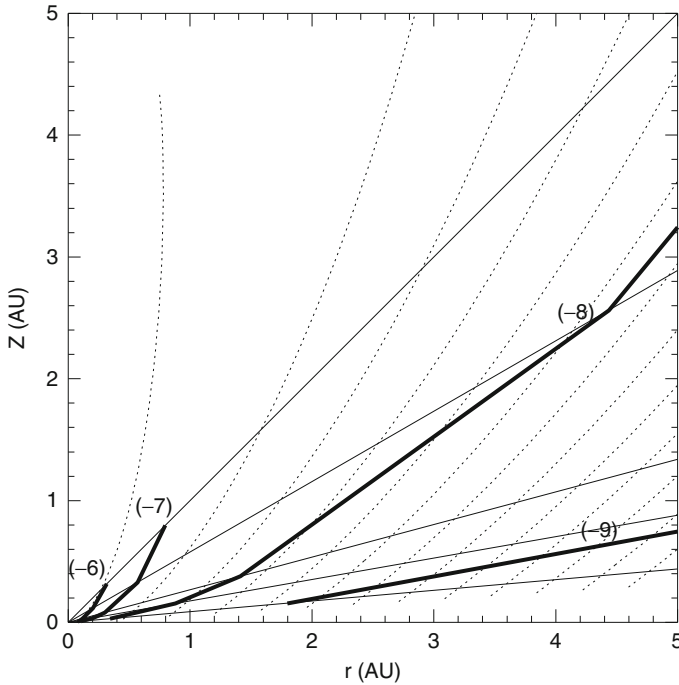
¹¹The β of plasma measures the rate between the thermal and the magnetic pressure, or, $\beta = (nk_B T)/(B^2/8\pi)$.

¹²The self-similar solution corresponds to a scaling of the magnetic flux as $\propto r^{3/4}$.



■ Fig. 6-23

The basic kinematics of MHD centrifugal winds is outlined from the simple semiempirical model of centrifugally driven MHD winds with thermal pressure of Gómez de Castro and Ferro-Fontán (2005). In the *central panel*, the nested plasma sheets are represented. To the *right*, the properties of the velocity field in the flow are described; the variation of the toroidal (V_t), radial (V_r) and axial (V_z) components with the height above the disk midplane are represented. The toroidal component is scaled with respect to the Keplerian velocity of the disk at the radius from which the wind is ejected. V_r and V_z are plotted in the rightmost inset with *solid and dashed lines*, respectively. The location of the Alfvén points is marked with *horizontal dashed lines*. In the *left panel*, line profiles generated by rings of gas in the flow, at different heights along the z -axis (the disk axis) are represented for an observed at $i = 90^\circ$ (disk seen edge-on)



■ Fig. 6-24

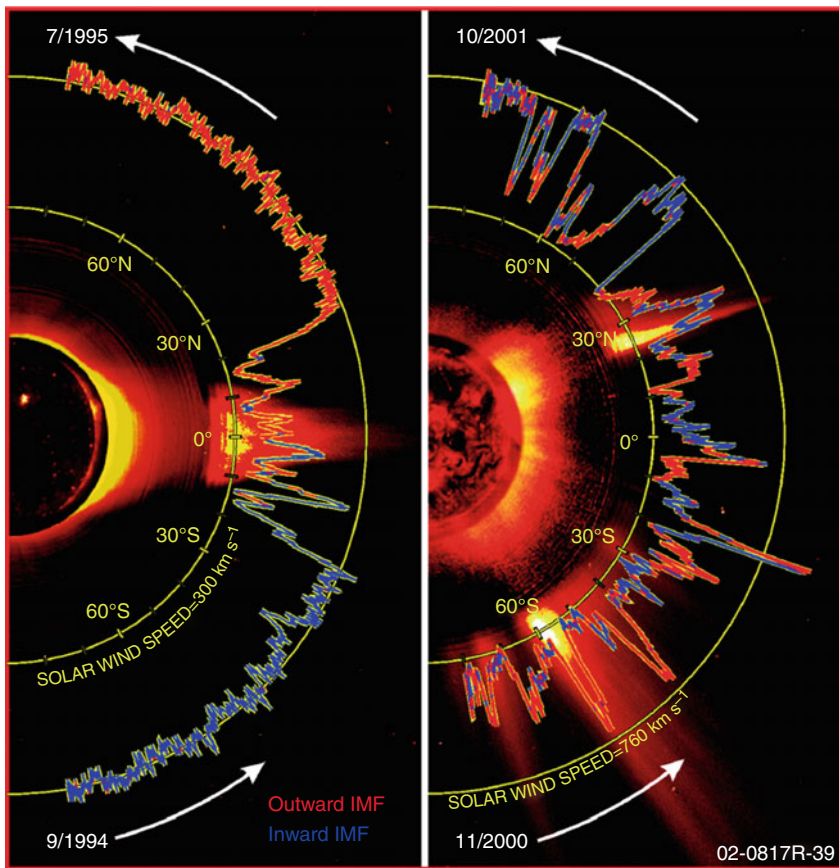
Surfaces where the temperature reaches 4,000 K due to photoionization of the disk wind by the stellar X-ray field. The *thick solid lines* represent the isotherm surfaces for accretion rates of 10^{-6} , 10^{-7} , 10^{-8} and 10^{-9} solar masses per year as marked by the numbers in brackets. The *thin dashed lines* represent the flow lines of matter in the jet. The *thin solid lines* mark the direction of colatitude $\vartheta = 45^\circ, 60^\circ, 75^\circ, 80^\circ$ and 85° (from Ferro-Fontán and Gómez de Castro 2003)

flow from the disk is channeled by the stellar magnetic field onto the star. The stellar field is assumed to be anchored in the inner part of the disk creating a sheared layer between the rigid body rotation of the star and the Keplerian rotation of the disk. This interaction has a profound influence on the star and the accretion flow but also acts as a dynamo that transforms part of the angular momentum excess in the inner disk into magnetic field amplification that self-regulates through quiescent periods of field building up and eruptions when the energy excess is released.

The basic physics is rather simple: the stellar magnetic field acts like the magneto in the dynamo that is able to drive currents by means of the Lorentz force into the material falling under the action of gravity. There are, however, a great number of uncertainties in the way the system self-regulates and also on the dependence of the engine details on the initial conditions such as the effective gravity of the star, the role of stellar radiation and magnetic field on the engine performance, and the role of the ionizing radiation produced by the engine on the evolution of the mass storage, the disk.

Going forwards on time: the Sun, itself, provides important clues to understand the physics of this gravito-magnetic engine and its evolution. At the base of the Sun convective layer, the tachocline marks the location of the shear layer between the rigid body rotation of the radiative core and the differentially rotating convective envelope. The radial gradient of the angular

velocity changes sign becoming negative for latitudes $>35^\circ$ and positive above this latitude. This latitude marks the limit of the two latitude belts where the overwhelming majority of sunspots occur. There are also some indications of the meridional flow moving equatorward below this latitude and poleward above it (see Miesch 2005). The Solar wind is (magnetic) latitude dependent during solar minimum; above $\sim 35^\circ$ is fast ($1,000 \text{ km s}^{-1}$) and thin, below $\sim 35^\circ$ is slow (300 km s^{-1}) and dense (see ► Fig. 6-25 from Ulysses data). The current paradigm for how solar dynamo operates includes: (1) field amplification in a turbulent downflow (α effect) that is pumped downward by convection and accumulated in the overshoot region and the tachocline;



■ Fig. 6-25

Solar wind observations collected by the Ulysses spacecraft during two separate polar orbits of the Sun, 6 years apart, at nearly opposite times in the solar cycle. Near solar minimum (*left*) activity is focused at low altitudes, high-speed solar wind prevails, and magnetic fields are dipolar. Near solar maximum (*right*), the solar winds are slower and more chaotic, with fluctuating magnetic fields (From NASA Solar Probe Web (solarprobe.gsfc.nasa.gov), courtesy of Southwest Research Institute and the Ulysses/SWOOPS team)

(2) field amplification and organization into toroidal flux tubes and sheets by differential rotation in the tachocline; (3) magnetic instabilities (buoyancy) drives the field to the surface, and (4) the Coriolis force acting on the rising structures to produce a latitude-dependent emergence of bipolar magnetic structures.

Going backward in time, during star formation, there is an “external tachocline” or differentially rotating region attached to the stellar surface (the outer part of the convective layer). This *external tachocline* connects the star with the accretion disk which rotates significantly more slowly than the stellar surface; rotation periods of TTs are about 7–8 days ($\Omega_* = 0.8\text{--}0.9 \text{ day}^{-1}$) while the Keplerian frequency is:

$$\Omega_k = 0.6 \text{ day}^{-1} \left(\frac{M}{M_\odot} \right)^{1/2} \left(\frac{r}{3R_\odot} \right)^{-3/2}$$

Keplerian disk corotation radius is at

$$r_{\text{corot}} = 17.5 R_\odot \left(\frac{M}{M_\odot} \right)^{1/3} \left(\frac{P}{8 \text{ days}} \right)^{2/3}$$


To avoid this large shear, the magnetosphere grows to balance the toroidal component of the flux with the angular momentum of the infalling matter (Ghosh and Lamb 1979) thus,

$$\frac{B_p B_t}{4\pi} 4\pi r^2 \Delta r \simeq \dot{M}_a r V_k$$

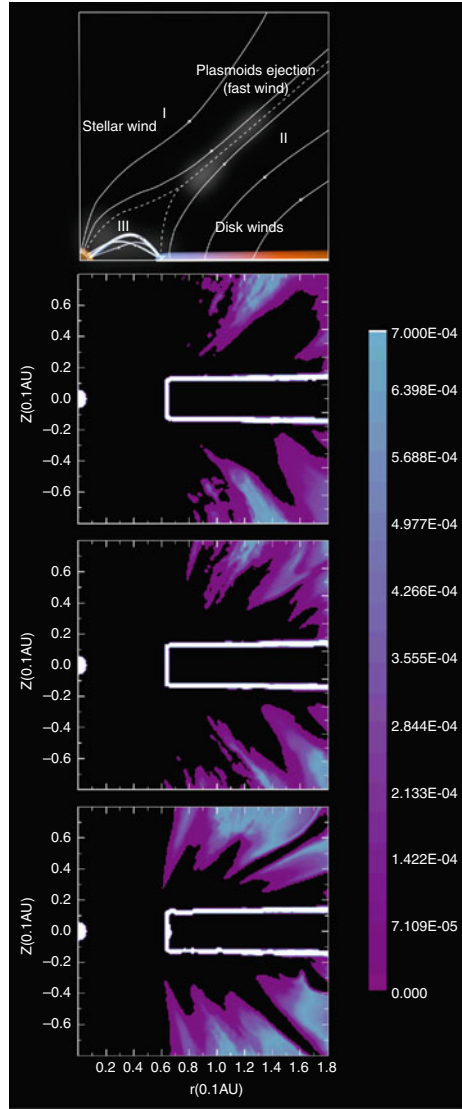
where B_p and B_t are the poloidal and toroidal components of the field, respectively, r is the magnetosphere radius, Δr is the thickness of the shear layer, \dot{M}_a is the accretion rate, and V_k is the Keplerian velocity at the magnetospheric radius. For typical T Tauri star parameters:

$$r_{\text{mag}} = 4.4 R_\odot \gamma^{2/7} \left(\frac{B_*}{1 \text{ kG}} \right)^{4/7} \left(\frac{\dot{M}}{10^{-8} M_\odot \text{ year}^{-1}} \right)^{-2/7} \left(\frac{M_*}{M_\odot} \right)^{-1/7}$$

where $\gamma^{2/7}$ is a factor about unity ($\gamma = (B_t/B_p)(\Delta r/r)$, see Lamb 1989). Notice that the main uncertainties in the physics, namely, the ratio between the toroidal and the poloidal components and the relative thickness of the *external tachocline*, are enclosed in this factor.

As in the Solar interior, the shear region is fed by turbulent, magnetized material, though in YSOs, this comes from the accretion disk. Shear amplifies the field producing a strong toroidal component and a dynamo sets in an extended stellar magnetosphere. This toroidal field and the associated magnetic pressure push the field lines outward from the disk rotation axis, inflating and opening them in a *butterfly-alike pattern* reminiscent of the helmet streamers in the solar corona. As a result, a current layer is produced between the region dominated by a hot and thin stellar wind and the area where the disk wind is ejected from as displayed in  Fig. 6-26. Magnetic field dissipation in the current layer produces high-energy radiation and particles. The magnetic link between the star and the disk is broken and reestablished continuously by magnetic reconnection as shown in the figure. The opening angle of the current layer, as well as its extent, depends on the stellar and disk fields, the accretion rate, and the ratio between the inner disk radius and the stellar rotation frequencies. Hot, pressure-driven outflows are produced from the region closer to the rotation axis while centrifugally driven flows are produced by the disk; plasmoids are ejected from the current layer generating a third outflowing episodic component.

Disk–star interaction has been investigated since the early concepts of X-winds (see Shu et al. 2000 for a review) to the numerical simulations by Goodson et al. (1997) and the last



■ Fig. 6-26

The interaction between the stellar magnetic field and the disk twists the stellar field lines due to the differential rotation. The toroidal magnetic field generated out of the poloidal flux and the associated pressure tends to push the field lines outward, inflating them, and eventually braking the magnetic link between the star and the disk (boundary between Regions I and II). Three basic regions can be defined: Region I dominated by the stellar wind, Region II dominated by the disk wind, and Region III dominated by stellar magnetospheric phenomena. The *dashed line* traces the boundaries between these three regions. The *continuous lines* indicate the topology of the field and the shadowed areas represent regions where magnetic reconnection events are likely to occur, producing high-energy radiation and particles. UV radiation released by the plasma heated in this interaction going from quiescence (*top*) to outburst (*bottom*) after Gómez de Castro and von Rekowski (2011)

results (i.e., von Rekowski and Brandenburg 2006). They show that the fundamental mechanism for disk winds formation is robust; numerical simulations with different parameters (disk/star fields) and initial conditions produce disk winds. Stellar winds are much more sensitive to the physical conditions and, especially, to the stellar field. The characteristics of the accretion flow and the winds (dominant driver, temperature, terminal velocity, density, variability) depend on the physical properties of the system such as the degree of magnetization of the disk, the characteristics of the disk dynamo, and the stellar field. During *Phase III*, the bulk of the energy losses of the engine are radiated at UV and X-ray wavelengths as in the Sun atmosphere (though a large fraction of the high energy radiation is reprocessed to longer wavelengths in the YSOs environment at early stages).

This paradigm is based (and inspired) in the Sun thus, naturally, applies to low mass YSOs with well-defined stellar fields; it is yet unclear how to adapt it to H Ae Be stars and, in general, to massive YSOs. However, Ae Herbig stars have a rich UV emission-line spectrum consistent with the presence of a chromosphere above the photosphere (Brown et al. 1996; Deleuil et al. 2005). Also, overionized species (transition region or corona-like) are observed. Magnetic field dissipation seems the most likely source for the radiative losses in the chromosphere/wind that represent 4–8% of the stellar bolometric luminosity according to Bouret and Catala (1998), although accretion flows may be a non-negligible energy source (Blondel et al. 1993). Marginal detection of magnetic fields has been reported for HD 104237 (Donati et al. 1997).

Further evidence to the presence of magnetic fields in the Ae Herbig stars atmospheres comes from the detection of azimuthal structures in the wind of AB Aur (Praderie et al. 1986). They are best interpreted resourcing, again, to the Solar wind and assuming that the very hot clumps are produced in the shocks between a “high” velocity component made by streamers of magnetically confined gas and a dense and “slow” outflow driven by radiation. In this sense, the hot clumps would be tracing a PMS analogue to the Solar Corotating Interaction Regions (CIRs). Evidence of such a class of structures has also been found in a Class III TTS: AB Dor (Gómez de Castro 2009).

Though observations suggest that fields are present in Ae Herbig stars, at least, during the first $\sim 5 \times 10^6$ years of their PMS evolution, the ultimate source of the field remains unidentified. Turbulence and rotation could set up a dynamo in the outer stellar layers. Turbulence can be generated by stellar pulsation; radial and nonradial modes have been detected with periods from some tens of minutes to hours (see Catala 2003 for a review). Also, the rotational braking produced by the strong stellar wind could induce turbulent motions below the stellar surface, forcing magnetic fields into the outer stellar layers (Lignieres et al. 1996).

Accretion is not expected to be the only driver of the outflow in these luminous sources. Radiatively driven winds are also able to produce collimated outflows provided there is a magnetic field (Rotstein and Gimenez de Castro 1996).

4.3 The Stellar Magnetosphere

The stellar magnetosphere is the key structure in the engine. Magnetic fields of few kG have been detected in the surface of the TTSs (Guenther et al. 1999; Johns-Krull et al. 1999). The surface field is not bipolar but has a rather complex structure as in the Sun (Johns-Krull et al. 2004). Higher order multipolar components fall off more rapidly with radius than the dipolar field hypothesized for the engine, making extremely difficult to track the path followed by matter


from the inner disk border to the stellar surface. In addition, the magnetosphere has its own dynamics and forcing due to the interaction with the disk.

4.3.1 Accretion Shocks and Funnels


The accretion flow has a clearly identifiable episodic component produced by the capture of material from the inner border of the disk into magnetic loops or funnels that drive it into the stellar surface. The kinetic energy of the infalling material is released into heating at the impact point. As matter infall occurs along the field lines, all the gravitational energy is damped into heating and the gas may reach temperatures as high as $\sim 10^6$ K. The dominant output radiation is produced by the photoionized preshock infalling gas radiating mainly in the UV range; the shock front itself radiates in soft X-rays (Gullbring et al. 2000; Lamzin 1998). As the density of the infalling gas column is high ($n_e \simeq 10^9 - 10^{12} \text{ cm}^{-3}$) the thickness of the radiating column is expected to be negligible compared with the stellar radius; thus, accretion shocks are observed as *hot spots* on the stellar surface. As such, they are expected to produce a rotationally modulated signal at high energies that is especially strong at the shortest wavelengths (see Petrov et al. 2001 or Bouvier et al. 2003 for optical monitorings and Gómez de Castro 1998 for UV monitorings). A soft X-ray excess is reported for actively accreting TTSs that would trace the location of the $\sim 10^6$ K shock front (Guedel and Telleschi 2007). The presence of infalling gas cloudlets is detected in the redwing of the line profiles of many species tracing warm plasma ($\text{H}\alpha$, Mg II, Ca II, etc.); however, there are very few monitorings tracking the motion of gas parcels as they fall onto the star, and when this is done, a correlation is found between accretion and outflow (RU Lup: Stempels and Piskunov 2002; RW Aur: Gómez de Castro and Verdugo 2003 and AA Tau: Bouvier et al. 2007). Four interesting features have been derived from the observations:


1. The interaction between the magnetosphere and the disk may drive to the formation of warps in the inner disk as reported for AA Tau (Bouvier et al. 2003).
2. The end points of the accretion funnels cover just a few percent of the stellar surface (Bouvier et al. 2007).
3. Accretion flows may switch on or switch off on timescales of a few years without large variations in the overall properties of the star or the disk as observed in RY Tau (Gómez de Castro 2009).
4. The hot plasma ($4 < \log T_e(K) < 5$) tracers registering the accretion shocks, as for instance the He I 10830 (Fischer et al. 2008) or the UV lines, indicate that there is an extended magnetospheric component in addition to the accretion shock emission. Only $\sim 50\%$ of the UV continuum excess is rotationally modulated (Gómez de Castro 2009).

4.3.2 General Magnetospheric Properties

Unfortunately, the characteristics of the TTSs extended magnetospheres are still escaping the diagnosis (see Hartmann 2009). Further to some qualitative properties as that outlined in the sketch in  Fig. 6-26, little is known about TTSs magnetospheres apart from having a density of about $10^9 - 10^{10} \text{ cm}^{-3}$ and an electron temperature between some few thousand Kelvin and 100,000 K. The line widths of typical atmospheric/magnetospheric tracers are about

200–300 km s⁻¹ that exceed by far what was expected from thermal broadening or from rotational broadening even if the lines are assumed to be formed in a corotating magnetosphere that extends to some 4–5 stellar radii.

The source of these large broadenings remains evasive since the very beginning of the studies of TTSs atmospheres. In the 1980s, line broadening was modeled with macro turbulence fields in a very massive outflow (Calvet et al. 1985). Later on, magnetospheric accretion was proposed to explain the large broadening of the TTSs H α lines (Muzerolle et al. 1998). Infall adds an extra radial velocity component to the rotation of the radiating gas that has an effect in the line profiles somewhat similar to that shown in  Fig. 6-23 for the outflow. As free-fall velocities are comparable to the broadening¹³, the observed profile widths can be reproduced without difficulty tuning the curvature of the plasma trajectory from the inner disk into the stellar surface. However, there must be a dependence on the inclination that has not been clearly identified on the spectrally resolved observations of magnetospheres. As of today, it is still unclear whether the broadening is produced by unresolved macroscopic flows or by magnetic waves propagating on the magnetospheric field. In general, the broadening of magnetospheric tracers does not vary significantly pointing out that the average motions are rather stable. The combined effect of funnel flows and inclined magnetic rotators as simulated by Romanova et al. (2004) is the current baseline for the numerical simulation of TTSs magnetospheres.

Magnetospheres are assumed to produce the truncation of the inner disk producing the observed holes in the molecular gas distribution (see  Sect. 3). As the disk is not unlocked from the magnetosphere, it is expected to be subjected to the propagation of the Alfvén waves, shear waves, and global alfvén oscillations driven from the interface. The inner rim of the disk is expected to be hot with temperatures of about 10⁴ K. Evidence for such a structure has been found only in RW Aur (Gómez de Castro and Verdugo 2003).

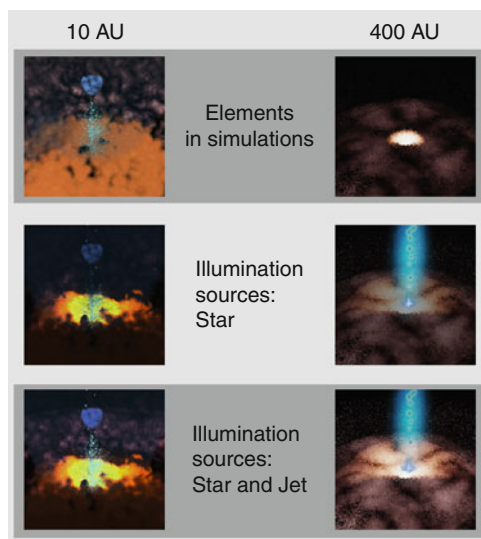
5 Some Final Thoughts and Conclusions

Star, disk, and outflow are the three components of a basic gravitational engine. The star grows out of the accretion flow and it is the most stable element of the system. After reaching a critical mass, the spheroid reaches hydrostatic equilibrium and any collision or dramatic event leading to the partial or total disruption of the disk will just reduce its final mass making the star formation process just more inefficient. Thus stars and disk ages do not always match.

Disks have a profound effect on stellar evolution since the magnetic interaction between the star and the disk controls the evolution of angular momentum during star formation. As soon as the star–disk locking is released, YSOs follow the hydrostatic contraction to the ZAMS speeding up as the radius decreases and the mass distribution in the star gets more centrally peaked. As a result, the Class III TTSs rotate faster than the Class II (see Choi and Herbst 1996) and a broad distribution of rotation periods is expected among stars of the same mass in the same molecular cloud depending on the timing of the disk unlocking.

There is, however, a link between the age of young associations as a whole and the rotation spread of their member stars. As an example, the spread of velocities observed in α Per (a 50 Myr old stellar association) is 150 km s⁻¹ for G-type stars (0.8–1.0 M_⊙) (Prosser 1992). However no

¹³Typical free-fall velocities are: $v_{\text{ff}} \approx 315 \text{ km s}^{-1} \left(\frac{M_*}{M_{\odot}} \right)^{1/2} \left(\frac{R_*}{R_{\odot}} \right)^{-1/2}$.

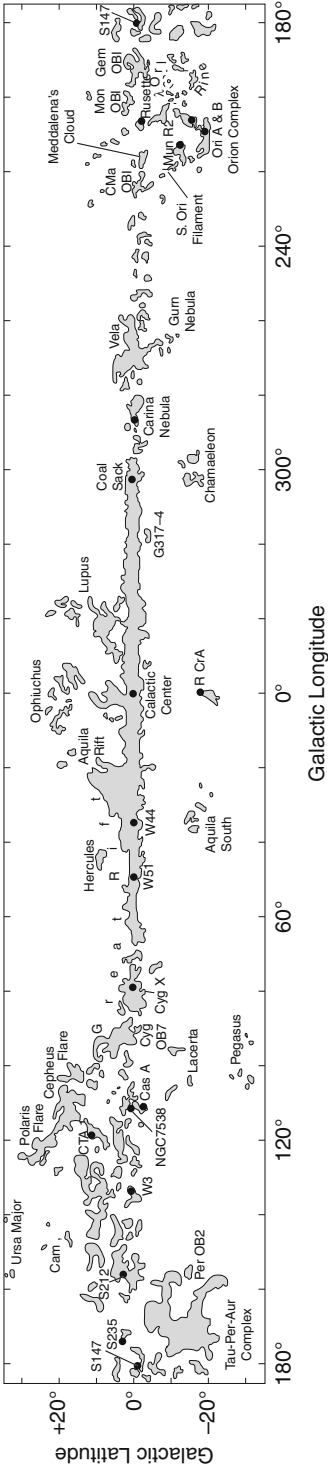


■ Fig. 6-27

Simulation of the effect of the outflow and the inflated magnetic structure on the disk irradiation. *Top*, YSOs' components; *middle*, only stellar illumination is switched on; *bottom*, full illumination, including the jet. The density in the disk is simulated by seven nested surfaces that act as boundaries to volumes filled with 10^6 particles each, up to a grand total of 7×10^6 particles in the disk. The surfaces have been selected from the final density distribution in the gravitational collapse simulations of Fig. 6-9 (bottom panel of Fig. 6-9a). The density levels correspond to $\Delta \log \rho = 0.6$. The optical performance of the particles is characterized by an albedo of $57 \pm 2\%$

spread is observed among G-type stars in the older Pleiades association (100 Myr old) while K dwarfs still display a broad dispersion in rotation velocities (Mayor and Mermilliod 1991) that is lost in the even older Hyades association (Stauffer et al. 1997). This indicates that once the disk locking is lost, a Solar-like wind develops braking the stellar rotation. As the total angular momentum of the system ought to be conserved, the angular velocity of the star decreases at a rate of roughly $I_A \dot{M}_w$, where \dot{M}_w is the wind mass-loss rate (for the Sun currently is $\sim (2-3) \times 10^{-14} M_\odot \text{year}^{-1}$). As the lever arm (the Alfvén radius) is larger for faster rotators, they brake more efficiently than the slower ones reaching an equilibrium configuration in some tens of Myrs.

Stars and the accretion engine control the irradiation of the disk and its lifetime. The base of the outflow and the reconnection at the star-disk interface is an abundant source of hard UV radiation that has a profound impact into disk evolution and chemistry. The geometrical location of the illuminating source is relevant to determine the flux that reaches the disk surface. As shown in Fig. 6-27, the effect of the outflow and the inflated magnetic structure on the disk irradiation ought to be taken into account. High latitude illumination facilitates energy absorption by the disk. The simulation shown in the figure is made with ray-tracing algorithms. The YSO's environment is simulated assuming a density law as in Yorke and Bodenheimer (1999) to model the protostellar disk. The disk is irradiated by three sources (star, extended magnetosphere, and jet). The luminosity of the star is assumed to be L_\odot , the luminosity of the extended magnetosphere to be 10% of this value, and the luminosity of the jet $0.01 L_\odot$. The system was scaled to



■ Fig. 6-28
Molecular gas in the Gould's Belt from Dame (2001)

represent a solar mass star. The effect of the jet in the disk irradiation can be drawn by comparing the two lower panels where the jet contribution to the disk illumination is switched on/off. Unfortunately, the current uncertainties on the inner structure of the engine and the properties of the magnetosphere make this kind of studies extremely model dependent.

In the coming years, the understanding of YSOs and protostellar disks evolution will suffer a major revolution due to the impact of the new facilities, now under development, for (1) high spatial resolution (1 AU scales) mapping of the molecular emission of protostellar disks and (2) dedicated monitorings of the YSOs high-energy output at UV wavelengths to map the magnetospheric structure and the inner engine.

Acknowledgments

Writing this review has been a challenge given the enormous amount of information to be summarized in a meaningful text. Detailed reviews on the specific topics addressed in this chapter can be found in the *Protostars and Planets* series. During the writing, I have enjoyed conversations with Javier López-Santiago, Ralph Pudritz, José Cernicharo, and Rafael Bachiller that I wish to acknowledge. A most rewarding experience was building the last figure in the article. The tools developed for movie and game makers are becoming increasingly and amazingly sophisticated. I would like to thank Victor Rodrigo for introducing me to this field. I wish also to thank Fátima López-Martínez for her assistance with the figures.

Appendix A: The Gould's Belt

The Gould's Belt is the local super association of star formation and has characteristics similar to a starburst. Molecular gas, early B to A stars, star forming regions, and young associations as the Pleiades or the Hyades form part of the belt. The extent of the belt is about 800 pc and the Sun is located at about 160 pc of its innermost border (see Poeppel 1997 for a review). Around the position of the Sun, there is an inner belt of radius 140 pc where the main star forming regions are located: Taurus-Aurigae, Ophiuchus, Lupus, Chamaleon, Sco-Cen, and Serpens (see ● Fig. 6-28). There are not very massive O-type stars in the belt. The origin of the Gould's Belt is unknown. The constituents of the belt share a common kinematics and are located in a plane inclined with respect to the galactic plane by 16° . Three main formation scenarios have been proposed to explain its characteristics: a supernovae explosion (Perrot and Grenier 2003), the impact of a high velocity cloud from the halo (Comerón and Torra 1994), and the formation of a corrugated galactic disk due to the onset of the Parker instability (Gómez de Castro and Pudritz 1992).

References

- | | |
|--|--|
| Abraham, P., Kospal, A., Csizmadia, Sz., et al. 2004, A&A, 428, 89 | Adams, F. C., Ruden, S. P., & Shu, F. H. 1989, ApJ, 347, 959 |
| Adams, F. C., Lada, Ch. J., & Shu, F. H. 1987, ApJ, 312, 788 | Alexander, D. R., & Ferguson, J. W. 1994, ApJ, 437, 879 |

- Alexander, R. D., Clarke, C. J., & Pringle, J. E. 2006, *MNRAS*, 369, 1, 229
- Andersen, J. 1991, *Astron Astrophys Rev*, 3, 91
- Andersen, J., Lindgren, H., Hazen, M. L. et al. 1989, *A&A*, 219, 142
- Andre, Ph., Martin-Pintado, J., Despois, D. et al. 1990, *A&A*, 236, 80
- Andre, P., Ward-Thompson, D., & Barsony, M. 2000, *Protostars and Planets IV* (Tucson: University of Arizona Press), 59
- Arce, H. G., & Sargent, A. I. 2006, *ApJ*, 646, 1070
- Armitage, P. J., Livio, M., & Pringle, J. E. 2001, *MNRAS*, 324, 705
- Arons, J., & Max, C. E. 1975, *Astrophys J Lett*, 196, L77
- Bachiller, R., & Tafalla, M. 1999, *The Origin of Stars and Planetary Systems* (Dordrecht: Kluwer), 227
- Bachiller, R., Pérez Gutiérrez, M., Kumar, M. S. N. et al. 2001, *A&A*, 372, 899
- Balbus, S. A., & Hawley, J. F. 1991, *ApJ*, 376, 214
- Balbus, S. A., & Papaloizou, J. C. B. 1999, *ApJ*, 521, 650
- Bally, J., O'Dell, C. R., & McCaughrean, M. J. 2000, *AJ*, 119, 2932
- Bally, J., Walawender, J., Luhman, K. L. et al. 2006, *AJ*, 132, 1930
- Beckwith, S. V. W., Sargent, A. I., Chini, R. S. et al. 1990, *Astronom J*, 99, 924
- Behrend, R., & Maeder, A. 2001, *A&A*, 373, 190
- Bell, K. R., & Lin, D. N. C. 1994, *ApJ*, 427, 987
- Bell, K. R., Lin, D. N. C., Hartmann, L. W. et al. 1995, *ApJ*, 444, 376
- Bertoldi, F. 1989, *ApJ*, 346, 735
- Blandford, R. D., & Payne, D. G. 1982, *MNRAS*, 199, 883
- Blondel, P. F. C., Talavera, A., & Djie, T., 1993, *A&A*, 268, 624
- Boehm-Vitense, E. 1958, *Zeitschrift fur Astrophysik*, 46, 108
- Bonnor, W. B. 1957, *MNRAS*, 117, 104
- Bouret, J. C., & Catala, C. 1998, *A&A*, 340, 163
- Bouvier, J. 1990, *Astronomical Journal*, 99, 946
- Bouvier, J., Grankin, K. N., Alencar, S. H. P. et al. 2003, *A&A*, 409, 169
- Bouvier, J., Alencar, S. H. P., Boutelier, T. et al. 2007, *A&A*, 463, 1017
- Braginskii, S. I. 1965, *Rev Plasma Phys*, 1, 205
- Brown, A., Deeney, B. D., Ayres, T. R. et al. 1996, *ApJS*, 107, 263
- Calvet, N., Basri, G., Imhoff, C. L., et al. 1985, *ApJ*, 293, 575
- Canuto, V. M., & Mazzitelli, I. 1991, *ApJ*, 370, 295
- Casey, B. W., Mathieu, R. D., & Vaz, L. P. R. 1998, *Astronom J*, 115, 1617
- Catala, C. 2003, *Magnetism and activity of the sun and stars. Proceedings of the Conference held 17–21 September, 2002 in Toulouse, France*. EAS Publication Series, Vol. 9 (Les Ulis: EDP Sciences), 325
- Caughlan, G. R., & Fowler, W. A. 1988, *Atomic Data and Nuclear Data Tables*, 40, 283
- Cesaroni, R., Neri, R., Olmi, L. et al. 2005, *A&A*, 434, 1039
- Cesaroni, R., Galli, D., Lodato, G. et al. 2007, *Protostars and Planets V*, Vol. 951 (Tucson: University of Arizona Press), 197
- Chen, H., Myers, P. C., Ladd, E. F. et al. 1995, *ApJ*, 445, 377
- Choi, P. I., & Herbst, W. 1996, *Astronom J*, 111, 283
- Clampin, M., Krist, J. E., Ardila, R. D. et al. 2003, *AJ*, 126, 389
- Coffey, D., Bacciotti, F., Woitas, J. et al. 2004, *ApJ*, 604, 758
- Collins, J. A., Shull, J., & Giroux, M. L. 2009, *ApJ*, 705, 962
- Comerón, F., & Torra, J., 1995, *A&A*, 423, 652
- Comerón, F., & Torra, J., 1994, *A&A*, 241, 57
- Covino, E., Frasca, A., & Alcal, J. M. 2004, *A&A*, 427, 637
- Curiel, S., Rodriguez, L. F., Moran, J. M. et al. 1993, *ApJ*, 415, 191
- D'Antona, F., & Mazzitelli, I. 1997, *Memorie della Societa Astronomia Italiana*, 68, 807
- D'Alessio, P., Canto, J., Calvet, N. et al. 1998, *ApJ*, 500, 411
- Dame, T. M. (2001), in *Atoms, Ions and Molecules: New Results in Spectral Line Astrophysics*, ASP Conf. Ser. (San Francisco: ASP), 16, 43
- Deleuil, M., Bouret, J.-C., Catala, C. et al. 2005, *A&A*, 429, 247
- Donati, J. F., Semel, M., Carter, B. D. et al. 1997, *MNRAS*, 291, 658
- Draine, B. T. 1978, *Astrophys J Suppl Series*, 36, 595
- Dunham, M. M., Evans, N. J., Terebey, S. et al. 2010, *ApJ*, 710, 470
- Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, in *Protostars and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 607
- Dutrey, A., Guilloteau, S., & Simon, M. 2003, *A&A*, 402, 1003
- Dutrey, A., Guilloteau, S., & Ho, P. 2007, *Protostars and Planets V*, Vol. 951 (Tucson: University of Arizona Press), 495
- Edris, K. A., Fuller, G. A., Cohen, R. J. et al. 2005, *A&A*, 434, 213
- Eisloffel, J., Smith, M. D., Davis, C. J. et al. 1996, *Astronom J*, 112, 2086
- Evans, N. J., Dunham, M. M., Jorgensen, J. K. et al. 2009, *Astrophys J Suppl Series*, 181, 321

- Ferreira, J., & Pelletier, G. 1993, *A&A*, 276, 625
- Ferreira, J., & Pelletier, G. 1995, *A&A*, 295, 807
- Ferreira, J., & Casse, F. 2004, *Astrophys Space Sci*, 292, 479
- Ferro-Fontán, C., & Gómez de Castro, A. I. 2003, *MNRAS*, 342, 427
- Fischer, W., Kwan, J., Edwards, S. et al. 2008, *ApJ*, 687, 1117
- Font, A. S., McCarthy, I. G., Johnstone, D. et al. 2004, *ApJ*, 607, 890
- Frøebrich, D. 2005, *Astrophys J Suppl Series*, 156, 169
- Furlan, E., Hartmann, L., Calvet, N. et al. 2006, *Astrophys J Suppl Series*, 165, 568
- Gail, H. P. 2004, *A&A*, 413, 571
- Gammie, Ch. F. 1996, *ApJ*, 457, 355
- Gammie, Ch. F. 2001, *ApJ*, 553, 174
- García, P. J. V., Ferreira, J., Cabrit, S. et al. 2001, *A&A*, 377, 589
- Garrod, R. T., & Herbst, E. 2006, *A&A*, 457, 927
- Garrod, R. T., Vasyunin, A. I., Semenov, D. A. et al. 2009, *Astrophys J Lett*, 700, L43
- Geers, V. C., Augereau, J. C., Pontoppidan, K. M. et al. 2006, *A&A*, 459, 545
- Ghosh, P., & Lamb, F. K. 1979, *ApJ*, 232, 259
- Giardino, G., Favata, F., Micela, G. et al. 2007, *A&A*, 463, 275
- Golimowski, D., Ardila, D. R., Krist, J. E. et al. 2006, *AJ*, 131, 3115
- Gómez de Castro, A. I. 1998, *Ultraviolet Astrophysics Beyond the IUE final Archive*, ESA-SP 413 (Noordwijk, The Netherlands), 59
- Gómez de Castro, A. I. 2009, *Astrophys Space Sci*, 320, 97
- Gómez de Castro, A. I., & Pudritz, R. E. 1992, *ApJ*, 395, 501
- Gómez de Castro, A. I., & Pudritz, R. E. 1993, *ApJ*, 409, 748
- Gómez de Castro, A. I., & Verdugo, E. 2003, *ApJ*, 597, 443
- Gómez de Castro, A. I., & Ferro-Fontán, C. 2005, *MNRAS*, 362, 569
- Gómez de Castro, A. I., & von Rekowski, B. 2011, *MNRAS*, 411, 849
- Goodson, A. P., Winglee, R. M., & Boehm, K. H. 1997, *ApJ*, 489, 199
- Grady, C. A., Woodgate, B., Bruhweiler, F. C. et al. 1999, *ApJ*, 523, L151
- Guedel, M. 2008, *Astronomische Nacht*, 329, 218
- Guedel, M., & Telleschi, A. 2007, *A&A*, 474, L25
- Gullbring, E., Calvet, N., Muzerolle, J. et al. 2000, *ApJ*, 544, 927
- Guenther, E. W., Lehman, H., Emerson, J. P. et al. 1999, *A&A*, 341, 768
- Haisch, K. E., Jr., Lada, E. A., & Lada, Ch. J. 2001, *Astronom J*, 121, 2065
- Hartmann, L. 2009, *Protostellar Jets in Context, Astrophysics and Space Science Proceedings Series* (New York: Heidelberg), 23
- Hartmann, L., & Kenyon, S. J. 1996, *ARA&A*, 34, 207
- Hartmann, L., Calvet, N., Allen, L. et al. 1999, *Astronom J*, 118, 1784
- Hasegawa, T. I., & Herbst, E. 1993, *MNRAS*, 261, 83
- Hasegawa, T. I., Herbst, E., & Leung, Ch. M. 1992, *Astrophys J Suppl Series*, 82, 167
- Hawley, J. F., Gammie, Ch. F., & Balbus, S. A. 1995, *ApJ*, 440, 742
- Hawley, J. F. 2009, *Astrophysics and Space Science*, 320, 107
- Heathcote, S., Morse, J. A., Hartigan, P. et al. 1996, *ApJ*, 112, 1150
- Hester, J. J., & Desch, S. J. 2005, *Chondrites and the Protoplanetary Disk*, eds. A. N. Krot, Edward R. D. Scott, & Bo Reipurth. *ASP Conference Series*, 341, 117
- Herbig, G. H., & Bell, K. R., 1988, *Lick Observatory Bulletin*, 1111, 90
- Herczeg, G. J., Linsky, J. L., Valenti, J. A. et al. 2002, *ApJ*, 572, 310
- Herczeg, G. J., Wood, B. E., Linsky, J. L. et al. 2004, *ApJ*, 607, 369
- Hillenbrand, L. A., & White, R. J. 2004, *ApJ*, 616, 998
- Hollenbach, D., Johnstone, D., Lizano, S. et al. 1994, *ApJ*, 428, 654
- Iben, I., Jr. 1965, *ApJ*, 141, 993
- Igea, J., & Glassgold, A. E. 1999, *ApJ*, 518, 848
- Iglesias, C. A., & Rogers, F. J. 1996, *ApJ*, 464, 943
- Ilgner, M., & Nelson, R. P. 2006, *A&A*, 445, 205
- Ilgner, M., & Nelson, R. P. 2008, *A&A*, 483, 815
- Ingleby, L., Calvet, N., Bergin, E. et al. 2009, *Astrophys J Lett*, 703, L137
- Johns-Krull, C. M., Valenti, J. A., & Koresko, C. 1999, *ApJ*, 516, 900
- Johns-Krull, C. M., Valenti, J. A. et al. 2004, *ApJ*, 617, 1204
- Kalas, P., Graham, J. R., Clampin, M. C. et al. 2006, *ApJ*, 637, L58
- Kalas, P., Fitzgerald, M. P., & Graham, J. R. 2007, *ApJ*, 661, L86
- Kalas, P., Graham, J. R., Chiang, E. et al. 2008, *Science*, 322, 1345
- Kessler-Silacci, J. E., Hillenbrand, L. A., Blake, G. A. et al. 2005, *ApJ*, 622, 404
- Kirk, J. M., Ward-Thompson, D., André, P. 2005, *MNRAS*, 360, 1506
- Koenigl, A. 1991, *ApJ*, 370, L39
- Krist, J. E., Burrows, C. J., Stapelfeldt, K. R. et al. 1997, *ApJ*, 481, 447
- Krist, J. E., Ardila, D. R., Golimowski, D. A. et al. 2005, *AJ*, 129, 1008

- Lamb, J. K., 1989, *Timing Neutron Stars* (Dordrecht: Kluwer), 649
- Lamzin, S., 1998, *Astronomy Rep*, 42, 322
- Laor, A., & Draine, B. T. 1993, *ApJ*, 402, 441
- Larson, R. B. 1969, *MNRAS*, 145, 271
- Laughlin, G., & Bodenheimer, P. 1994, *ApJ*, 436, 335
- Leinert, Ch., van Boekel, R., Waters, L. B. F. M. et al. 2004, *A&A*, 423, 537
- Lignieres, F., Catala, C., & Mangeney, A. 1996, *A&A*, 314, 465
- Lin, D. N. C., & Papaloizou, J. C. B. 1996, *ARA&A*, 34, 703
- Lin, S. Y., Ohashi, N., Lin, J. et al. 2006, *ApJ*, 645, 1297
- Looney, L. W., Tobin, J. J., & Kwon, W. 2007, *ApJ*, 670, L131
- Liseau, R., Ceccarelli, C., Larsson, B. et al. 1996, *Astronomy and Astrophysics*, 315, L181
- Lynden-Bell, D., & Pringle, J. E. 1974, *MNRAS*, 168, 603
- Mace, G. N., Prato, L., & Wasserman, L. H. 2009, *Astronom J*, 137, 3487
- Mannings, V., & Sargent, A. I. 2000, *ApJ*, 529, 391
- Martin-Zaidi, C., Deleuil, M., Le Bourlot, J. et al. 2008, *A&A*, 484, 225
- Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, *ApJ*, 217, 425
- Mayor, M., & Mermilliod, J. C. 1991, Angular momentum evolution of young stars, in *Proceedings of the NATO Advanced Research Workshop on Angular Momentum Evolution of Young Stars*, held in Noto, Sicily, Italy (Boston: Kluwer), 143
- McKee, C. F. 1989, *ApJ*, 345, 782
- Meeus, G., Waters, L. B. F. M., Bouwman, J. et al. 2001, *A&A*, 365, 476
- Mejía, A. C., Durisen, R. H., Pickett, M. K. et al. 2005, *ApJ*, 619, 1098
- Mihalas, D., Dappen, W., & Hummer, D. G. 1988, *ApJ*, 331, 815
- Miesch, M. S. 2005, *Living Reviews in Solar Physics*, 2, 1
- Millan-Gabet, R., Schloerb, F. P., & Traub, W. A. 2001, *ApJ*, 546, 358
- Millan-Gabet, R., Malbet, F., Akeson, R. et al. 2007, *Protostars and Planets V*, Vol. 951 (Tucson: University of Arizona Press), 539
- Muzerolle, J., Calvet, N., & Hartmann, L. 1998, *ApJ*, 492, 743
- Myers, P. C., & Goodman, A. A. 1988, *ApJ*, 326, L27
- Myers, P. C., Adams, F. C., Chen, H. et al. 1998, *ApJ*, 492, 703
- Najita, J., Carr, J. S., & Mathieu, R. D. 2003, *ApJ*, 589, 931
- Natta, A., Prusti, T., Neri, R., et al. 2001, *A&A*, 371, 186
- Onishi, T., Mizuno, A., Kawamura, A. et al. 2002, *ApJ*, 575, 950
- Oppenheimer, M., & Dalgarno, A. 1974, *ApJ*, 192, 29
- O'Sullivan, M., Truss, M., Walker, C. et al. 2005, *MNRAS*, 358, 632
- Padgett, D. L., Brandner, W., Stapelfeldt, K. R. et al. 1999, *AJ*, 117, 1490
- Pedersen, A., & Gómez de Castro, A. I. 2011, *ApJ*, in press
- Pelletier, G., & Pudritz, R. E. 1992, *ApJ*, 394, 117
- Perrot, C. A., & Grenier, I. A. 2003, *A&A*, 404, 519
- Petrov, P. P., Gahm, G. F., Gameiro, J. F. et al. 2001, *A&A*, 369, 993
- Poeppel, W. 1997, *Fundam Cosmic Phys*, 18, 1
- Pollack, J. B., Hollenbach, D., Beckwith, S. et al. 1994, *ApJ*, 421, 615
- Pols, O. R., Tout, C. A., & Eggleton, P. P. 1995, *MNRAS*, 274, 964
- Praderie, F., Catala, C., & Simon, T. 1986, *ApJ*, Part 1, 303, 311
- Preibisch, Th., Ossenkopf, V., Yorke, H. W., et al. 1993, *A&A*, 279, 577
- Prosser, C. F. 1992, *Astronom J*, 103, 488
- Pudritz, R. E., Ouyed, R., Fendt, Ch. et al. 2007, in *Protostar and Planets V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: University of Arizona Press), 277
- Ray, T. P., Muxlow, T. W. B., Axon, D. J. et al. 1997, *Nature*, 385, 415
- Raymond, J. C., Blair, W. P., & Long, K. S. 1997, *ApJ*, 489, 314
- Reipurth, B., Chun Yu, K., Rodriguez, L. F. et al. 1999, *Astron Astrophys*, 352, L84
- Reipurth, B., 1999, *General Catalogue of Herbig-Haro Objects*, available through the Vizier service at vizier.u-strasbg.fr
- Reipurth, B., Bally, J., Fesen, R. A. et al. 1998, *Nature*, 396(6709), 343
- Rettig, T., Brittain, S., Simon, T. et al. 2006, *ApJ*, 646, 342
- Rice, W. K. M., Lodato, G., & Armitage, P. J. 2005, *MNRAS*, 364, L56
- Roberge, A., Feldman, P. D., Weinberg, A. J. et al. 2006, *Nature*, 441, 724
- Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, *ApJ*, 456, 902
- Romanova, M. M., Ustyugova, G. V., Koldoba, A. V. et al. 2004, *ApJ*, 610, 920
- Rotstein, N., & Gimenez de Castro, C. G. 1996, *ApJ*, 464, 859
- Safier, P. N. 1993, *ApJ*, 408, 148
- Sargent, B., Forrest, W. J., Tayrien, C. et al. 2009, *ApJ*, 690, 1193
- Saumon, D., Chabrier, G., & van Horn, H. M. 1995, *Astrophys J Suppl Series*, 99, 713

- Semenov, D., Wiebe, D., & Henning, Th. 2004, *A&A*, 417, 93
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shirley, Y. L., Evans, N. J., II., Rawlings, J. M. C. et al. 2000, *Astrophys J Suppl Series*, 131, 249
- Shirley, Y. L., Evans, N. J., II., Young, K. E. et al. 2003, *Astrophys J Suppl Series*, 149, 375
- Shu, F., Najita, J., Shang, H. et al. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S. S. Russell (Tucson: University of Arizona Press), 789
- Shuping, R. Y., Bally, J., & Morris, M. 2003, *ApJ*, 587, L109
- Siess, L., Forestine, M., & Dougados, C., 1997, *A&A*, 324, 556
- Siess, L., Dufour, E., & Forestini, M. 2000, *A&A*, 358, 593
- Simon, M., Dutrey, A., & Guilloteau, S. 2000, *ApJ*, 545, 1034
- Snell, R. L., Loren, R. B., & Plambeck, R. L. 1980, *ApJ*, 329, L17
- Stassun, K. G., Mathieu, R. D., & Vaz, L. P. R. 2004, *Astrophys J Suppl Series*, 151, 357
- Stauffer, J. R., Balachandran, S. C., Krishnamurthi, A. et al. 1997, *ApJ*, 475, 604
- Stempels, H. C., & Piskunov, N. 2002, *A&A*, 391, 595
- Tappe, A., Lada, C. J., Black, J. H. et al. 2008, *ApJ*, 680, L118
- Toomre, A. 1964, *ApJ*, 139, 1217
- Turner, N. J., & Drake, J. F. 2009, *ApJ*, 703, 2152
- Turner, N. J. J., Bodenheimer, P., & Bell, K. R. 1997, *ApJ*, 480, 754
- van Boekel, R., Waters, L. B. F. M., Dominik, C. et al. 2004, *A&A*, 418, 177
- Van Zadelhoff, G.-J., Aikawa, Y., & Hogerheijde, M. R. 2003, *A&A*, 397, 789
- Vasyunin, A. I., Semenov, D., Henning, Th. et al. 2008, *ApJ*, 672, 629
- Von Rekowski, B., & Brandenburg, A. 2006, *Astronomische Nachrichten*, 327, 53
- Watson, A. M., Stapelfeldt, K. R., & Wood, K. 2007, *Protostars and Planets V*, Vol. 951 (Tucson: University of Arizona Press), 523
- Weber, E. J., & Davis, L., Jr. 1967, *ApJ*, 148, 217
- Wilkinson, E., Harper, G. M., Brown, A. et al. 2002, *Astronom J*, 124, 1077
- Woodall, J., Agúndez, M., Markwick-Kemper, A. J. et al. 2007, *A&A*, 466, 1197
- Wu, Y., Wei, Y., Zhao, M. et al. 2004, *A&A*, 426, 503
- Yorke, H. W., & Bodenheimer, P. 1999, *ApJ*, 525, 330
- Yorke, H. W., & Sonnhalter, C. 2002, *ApJ*, 569, 846
- Young, Ch. H., & Evans, N. J., II. 2005, *ApJ*, 627, 293
- Zhang, Q., Hunter, T. R., Barnd, J. et al. 2005, *ApJ*, 635, 864

7 Brown Dwarfs

I. Neill Reid

Space Telescope Science Institute, Office of Public Outreach,
Baltimore, MD, USA

1	<i>Introduction</i>	338
2	<i>Formation and Evolution</i>	340
2.1	The Hydrogen-Burning Mass Limit	340
2.2	Formation Mechanisms	342
2.3	Brown Dwarf Evolution	344
2.4	Brown Dwarfs and the Mass Function	347
3	<i>Observational Properties of Brown Dwarfs</i>	353
3.1	Spectroscopic Properties	354
3.2	Photometric Properties	357
3.3	Observations of Young Brown Dwarfs	365
3.4	Lithium Absorption and Other Mass-Dependent Spectroscopic Phenomena	366
3.5	Metallicity and Ultracool Subdwarfs	368
3.6	Brown Dwarf Kinematics and Brown Dwarf Rotation	369
3.7	Chromospheric and Coronal Activity in Brown Dwarfs	371
4	<i>Brown Dwarf Atmospheres</i>	379
4.1	Atmosphere Structure	379
4.2	Dust and Clouds in Brown Dwarf Atmospheres	380
4.3	Y Dwarfs	381
5	<i>Brown Dwarfs as Companions</i>	383
5.1	Brown Dwarf Companions to Main-Sequence Stars	385
5.2	Brown Dwarf Binaries	389
5.3	Ultracool Binaries in Context	392
6	<i>Summary and Conclusions</i>	393
7	<i>Further Reading</i>	393
	<i>Acknowledgments</i>	394
	<i>References</i>	394

Abstract: Brown dwarfs are celestial objects that form in the same manner as stars, but which, due to their low mass, fail to generate sufficient energy during the initial collapse to achieve central temperatures that are capable of igniting hydrogen fusion. Lacking a long-lived energy source, brown dwarfs cool and fade to extremely low-luminosities on timescales that are short by astronomical standards. After a long drought, many examples have been discovered over the past 15 years, both as isolated objects in the field and as companions of known nearby stars. Most have extremely low surface temperatures, complex molecule-dominated atmospheres, and populate the new spectral classes L and T. This chapter summarizes our current knowledge of the basic characteristics of these low-mass objects and gives brief consideration to likely developments in the near future.

Keywords: Binary system, Brown dwarf, Brown dwarf - bolometric corrections, Brown dwarf - colors, Brown dwarf - disks, Brown dwarf - luminosities, Brown dwarf - spots, Brown dwarf desert, Brown dwarf evolution, Brown dwarf formation, Brown dwarf radius, Brown dwarf rotation, Brown dwarf survey, Brown dwarf weather, Brown dwarfs - atmospheres, Brown dwarfs - radio emission, Brown dwarfs - temperatures, Brown dwarfs - X-ray emission, Chromospheric activity, Dark matter, Deuterium fusion, Dust, Gl 229B, Hydrogen-burning limit, Initial mass function, Lithium fusion, Local mass density, Luminosity function, Mass-luminosity relation, Methane, Spectral classification, Ultracool dwarf, Ultracool pulsar

1 Introduction

The likely existence of star-like objects with insufficient mass to trigger hydrogen fusion was first postulated in the early 1960s (Kumar 1962). Originally tagged as “black dwarfs,” these substellar dwarfs resisted discovery for more than a quarter of a century, and were represented by a mere smattering of much-debated candidates until the mid-1990s. During that interim period, brown dwarfs acquired their name (Tarter 1974), and, for a limited time, were in vogue as a potential candidate for dark matter.

Prior to their discovery, the broad characteristics of brown dwarfs were anticipated theoretically, from models, and observationally, from measurements of low-mass stars. The models (e.g., Grossman et al. 1974) predicted that the mass-limit for hydrogen burning fell at $\approx 0.075 M_{\odot}$, corresponding to luminosities less than $10^{-3} L_{\odot}$ and temperatures below $\approx 2,400$ K. At those temperatures, the bulk of the radiation is emitted at infrared wavelengths, as emphasized by observations of the archetypical late-type dwarf, VB 10 (van Biesbroeck 1944). With a photospheric temperature of $\approx 2,500$ K, VB 10 has a luminosity $L \sim 3 \times 10^{-4} L_{\odot}$, corresponding to a bolometric magnitude of ~ 13.5 , or $\delta M_{\text{bol}} \sim 9$ magnitudes relative to the Sun. The absolute visual magnitude is $M_V = 18.65$, almost a million times fainter than the Sun at those wavelengths, but the K-band ($2.2 \mu\text{m}$) magnitude is $M_K \sim 10$, or ~ 500 times fainter than the Sun. The only wide-field surveys available in the 1970s and 1980s were based on photographic plates, which are most sensitive at blue and optical wavelengths. Thus, searches for field brown dwarfs were limited to sources within a few parsecs of the Sun.

Given those technical constraints, it is not surprising that the first brown dwarfs were detected as companions to low-luminosity stars already known to lie within the Solar Neighborhood, rather than in the field. The key discovery was Gl 229B (Nakajima et al. 1995), the low-luminosity companion of an early-type M dwarf within 6 parsecs of the Sun. Prior to its discovery, a handful of other candidates had been identified (see 🔗 Sect. 5.1) but none was

unambiguously substellar.¹ In contrast, Gl 229B has a very distinctive spectral energy distribution, with strong near-infrared absorption due to methane, leading to extremely blue near-infrared colors.² Those features indicate a photospheric temperature close to 1,000 K, substantially cooler than any star. Gl 229B provided unequivocal evidence of the existence of substellar brown dwarfs.

Advances in detector technology in the 1980s and early 1990s led to the development of large-format arrays, sensitive at optical and near-infrared wavelengths. The latter developments enabled initiation in the mid-1990s of the first deep, wide-field surveys at near-infrared wavelengths, the southern Deep Near-Infrared Survey (DENIS; Epchtein et al. 1994) and the 2-Micron All-Sky Survey (2MASS; Skrutskie et al. 2006). At optical wavelengths, the Sloan Digital Sky Survey (SDSS; York et al. 2000) covered a quarter of the sky to flux levels 10 to 100 times fainter than the photographic Palomar and UK Schmidt surveys.

Combined, data from these new sky surveys have led to an avalanche of discoveries. There are now several thousand dwarfs known that are as cool or cooler than VB 10, including $\approx 1,000$ dwarfs that are sufficiently cool that they required the establishment of two new spectral types, L and T. The expression *ultracool* dwarfs has been coined to describe all dwarfs with spectral types M7 or later. A quick note on terminology is in order here: All save the youngest brown dwarfs are ultracool dwarfs, but a significant number of ultracool dwarfs, at least at spectral type M, are very low-mass stars. From the observer's perspective, it is often impossible to distinguish whether a particular dwarf is above or below the hydrogen-burning limit. Thus, throughout this chapter, the noun "dwarf" is often prefaced with the neutral adjective "ultracool," reserving "brown" for unambiguous situations.

This inherent ambiguity in observational properties raises an important issue with regard to how brown dwarfs fit within the broader astrophysical context of stars and extrasolar planets. Brown dwarfs and stars are labeled separately based on their internal energy sources. However, the physical laws governing energy generation are entirely independent of those governing fragmentation, collapse, and accretion within gaseous clouds. Thus, while brown dwarfs receive particular attention in the literature, largely because of their ability to disappear, it is generally more appropriate that they are viewed as the low-mass extreme of a continuous distribution. On the other hand, while giant extrasolar planets are likely to have apparent characteristics that resemble cool brown dwarfs, planets form within a distinct environment, circumstellar disks. To borrow from life science, giant planets and brown dwarfs are more appropriately viewed as separate species with shared characteristics.

The overwhelming majority of ultracool dwarfs are isolated objects in the general field, with minorities drawn from clusters and as companions to main-sequence or evolved stars. The brightest ultracool dwarfs have been scrutinized through spectrophotometry, mapping the spectral energy distributions; high angular resolution imaging, to search for resolved luminous companions; and detailed spectroscopy at optical and near-infrared wavelengths, probing chemical compositions, chromospheric activity, and searching for velocity variations characteristic of close companions. Theoreticians have built on those results to model atmospheric structure, opacity sources, and dust formation and destruction. At the same time, deep imaging of young clusters and associations at far-red and near-infrared wavelengths has illuminated

¹For a more extensive historical discussion of brown dwarf discoveries, see Reid and Hawley (2005), Chapter 6.

²It is probably fair to say that these colors came as a surprise to most observers, who were generally focused on searching for objects that were extremely red at all wavelengths. This should not have been a surprise, since Tsuji (1964) predicted exactly these characteristics almost three decades before the discovery of Gl 229B.

both the early stages of formation, including the presence of disks and jets, and the overall frequency of brown dwarfs relative to stars.

This chapter provides an introduction to the nature and characteristics of brown dwarfs. At the appropriate junctures, parallels are drawn and connections made with higher-mass stars and lower-mass planetary objects. Given the page limits, it is not possible to cover all aspects of brown dwarf research within this single chapter. Cross-references are provided to selected articles from the literature and, where appropriate, to other chapters in this volume. The final section lists suggestions for further reading of review articles and textbooks.

2 Formation and Evolution

2.1 The Hydrogen-Burning Mass Limit

The absence of long-lived hydrogen fusion is the defining characteristic that sets brown dwarfs apart from stars. Stars and brown dwarfs form by gaseous collapse within molecular clouds. During the collapse, potential energy is released, leading to an increase in the core temperature, T_C . At the same time, the central density, ρ_C , rises. If T_C is driven above $T_{\text{crit}} \sim 3 \times 10^6$ K (the exact value is pressure dependent), then hydrogen fusion is initiated, providing a long-lived energy source that establishes hydrostatic equilibrium and supports the star against further collapse.

The energy injected into the core to raise the central temperature is a product of the total mass and the distance collapsed. Consequently, low-mass protostars have to undergo a higher degree of collapse than high-mass protostars. As a guide, in the perfect gas régime,

$$\begin{aligned} \frac{GM^2}{R} &\sim NkT, \text{ and} \\ \frac{M^2}{R} &\sim \text{constant}, \text{ so} \\ \langle \rho \rangle &\propto M^{-2} \end{aligned}$$

Thus, a $0.1 M_\odot$ dwarfs must achieve core densities that are a factor of a hundred higher than those present within solar-type G dwarfs before hydrogen fusion can be initiated. The existence of a lower mass limit to the hydrogen-burning main sequence stems from the onset of degeneracy in the core.

As core densities increase, particles are placed in closer proximity. Electron degeneracy pressure is a consequence of the Pauli exclusion principle, which states that two fermions cannot simultaneously occupy the same quantum state. That pressure is governed by the following relation

$$P_d = \frac{h^2}{20m_e m_p^{5/3}} \left(\frac{3}{\pi} \right)^{2/3} \left(\frac{\rho}{\mu_e} \right)^{5/3}$$

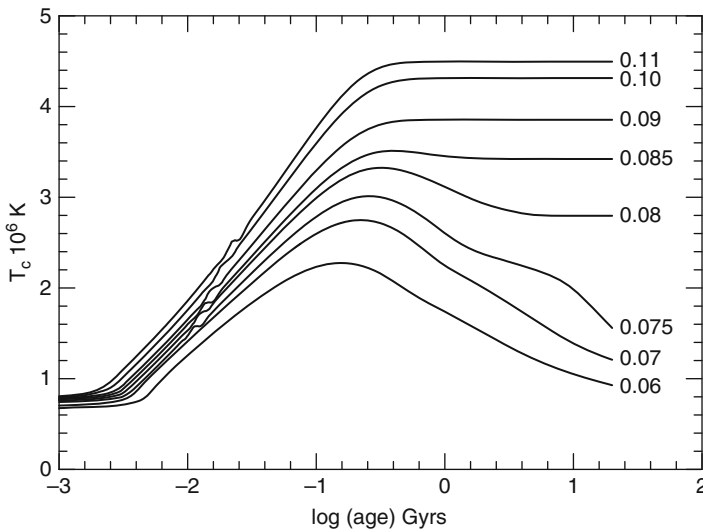
Where m_e and m_p are the mass of the electron and proton, h is Plancks constant, ρ is the density, and $\mu_e = \frac{N_e}{N_p}$ is the ratio of electron number to proton number. As ρ increases, more energy must be devoted into overcoming this pressure rather than transformed to thermal energy and contributed to heating the core. There are two major consequences: First, as originally pointed out by Kumar (1963), degeneracy can prevent the central temperature rising above the

critical level for hydrogen burning in low-mass dwarfs; second, degenerate incompressibility prevents low-mass dwarfs contracting to radii smaller than $\sim 0.1R_{\odot}$, comparable to the size of Jupiter.

Evolution: The detailed evolution of a low-mass dwarf depends critically on its mass, M . However, the stellar/brown dwarf division is shaded in gray. There are four scenarios to consider, as illustrated in [Fig. 7-1](#), which uses theoretical models from Burrows et al. (1993). Taken in order of decreasing mass, they are as follows:

- The central temperature rises above the critical threshold for hydrogen fusion, $T_{H,crit}$, and the core is nondegenerate. The temperature remains above the threshold and fusion is maintained. The star settles onto the main sequence and evolves with a lifetime exceeding 10^{12} years.
- The central temperature peaks above $T_{H,crit}$, but core degeneracy is sufficient that T_C is reduced below that peak, although not to the extent that it falls below the $T_{H,crit}$. The object achieves a main-sequence configuration, albeit at lower luminosities than the initial values.
- The central temperature rises above $T_{H,crit}$ and initiates fusion, but degeneracy within the core is sufficient to drop T_C below that value. As a result, these objects achieve stellar status for a limited time and eventually cool into oblivion as brown dwarfs.
- The central temperature never rises above $T_{H,crit}$; hydrogen fusion is never established, and the object “cools like a rock” (Burrows and Liebert 1993).

Scenarios 1 and 2 produce stable hydrogen-burning stars; the fourth scenario leads to a brown dwarf; and the third scenario corresponds to a transition object, with a mass balanced



■ Fig. 7-1

The time evolution of the core temperature in low-mass dwarfs with masses between 0.11 and $0.06 M_{\odot}$, as predicted by the Burrows et al. (1993) models. Dwarfs with masses below $0.075 M_{\odot}$ fail to achieve long-lived hydrogen burning, and $\sim 0.075 M_{\odot}$ transition objects are able to sustain H-fusion for a few Gyrs

on the threshold of the hydrogen-burning limit. Kumar originally placed that limit as $M_{BDcrit} \sim 0.07 M_{\odot}$ (or $\sim 73 M_{Jup}$). More recent models, such as those illustrated in [Fig. 7-1](#), predict that a solar-metallicity object with a mass of $\sim 0.078 M_{\odot}$ will meet scenario 2 and eventually settle onto the main sequence as a low-luminosity, but long-lived, late-type star; a $\sim 0.075 M_{\odot}$ dwarf will become a transition object; and the brown dwarf limit is $M_{BDcrit} \sim 0.073 M_{\odot}$.

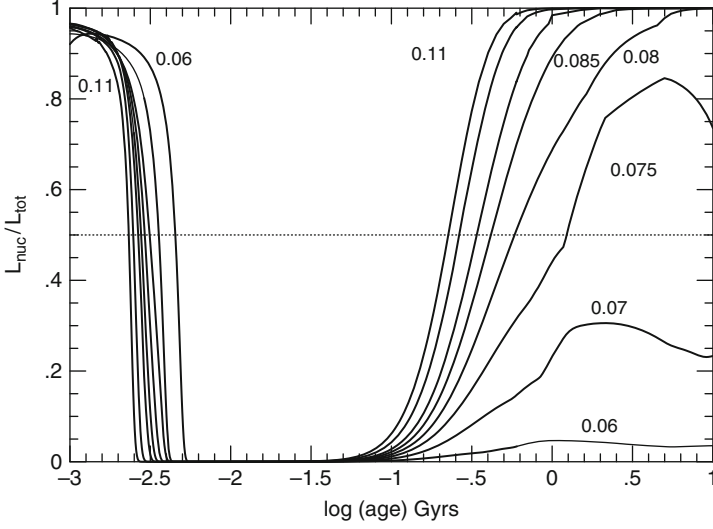
Chemical abundance plays a secondary role in defining the mass threshold that separates brown dwarfs and stars. Higher helium abundance, Y , leads to higher mean molecular weight, smaller radii, and higher T_C and ρ_C for the same mass; consequently, the critical mass for H-burning decreases as Y increases. Lower metallicity dwarfs have lower atmospheric opacities, which produce shallower temperature gradients and higher luminosities (less efficient energy retention) and correspondingly lower T_C . As a result, decreasing $[M/H]$ leads to higher brown dwarf mass limits, and higher luminosities and temperatures at the H-burning limit. As an example, at zero metallicity, $M_{BDcrit} \sim 0.092 M_{\odot}$, and the H-burning limit lies at $T_{eff} \sim 3,600$ K. In comparison, the same models predict effective temperatures of $\sim 1,700$ K for solar abundance brown dwarfs at the H-burning limit (Burrows et al. 2001).

Other fusion sources: Hydrogen is not the only source of fusion energy in low-mass dwarfs. Trace quantities of primordial lithium and deuterium can add to the energy budget if the core temperature rises sufficiently. Lithium fusion is triggered at temperatures above $T_{Li,crit} \sim 2.5 \times 10^6$ K, with the element burning to form helium (one of the later steps in the extended P–P chain, see [Chap. 6](#)). This temperature is achieved in the cores of solar-metallicity dwarfs with masses exceeding $\sim 0.06 M_{\odot}$. Late-type dwarfs are fully convective, so all material is exposed to those temperatures. Primordial lithium has a low abundance, $[Li]=3.3$ on a logarithmic scale where $[H]=12.0$, and the available fuel is quickly consumed: An $0.075 M_{\odot}$ dwarf is expected to deplete lithium fully in $\sim 10^8$ years, by which time it has cooled to photospheric temperature of $\sim 2,800$ K, corresponding to spectral type M5. This behavior permits the lithium test (Rebolo et al. 1992; [Sect. 3.4](#)), one of the few observational methods of unambiguously identifying (low mass) brown dwarfs. Ultracool dwarfs that exhibit Li I absorption at $6,708 \text{ \AA}$ have at most partially depleted the primordial lithium. Therefore, they have achieved core temperatures exceeding $T_{Li,crit}$ for a limited period of time, at most, and must have masses below $\sim 0.06 M_{\odot}$.

Deuterium fusion is triggered at temperatures as low as $\sim 4 \times 10^5$ K, and consequently occurs in lithium-rich brown dwarfs. Indeed, the mass threshold for deuterium burning, $\sim 0.012 M_{\odot}$ (at solar metallicities), has been proposed as the lower boundary of the brown dwarf régime. In principle, one might envisage an observational test comparable to the lithium test that could distinguish these very low-mass dwarfs, but deuterium has only a few weak spectroscopic signatures at mid-infrared wavelengths, making any observations exceedingly challenging (Chabrier et al. 2000). The primordial deuterium abundance is $[D] \sim -4.7$, or 17 orders of magnitude lower than hydrogen. Thus, as [Fig. 7-2](#) shows, deuterium provides only a brief spike in energy production during the earliest stages of brown dwarf formation.

2.2 Formation Mechanisms

Star formation is a complex process (see [Chap. 6](#) of this volume). This section focuses on brown dwarf formation which, in conventional theories, is characterized as the low-mass tail of a process of collapse and fragmentation that spans at least four orders of magnitude in mass.



■ Fig. 7-2

The fractional luminosity contributed by nuclear fusion reactions in low-mass dwarfs. As in [Fig. 7-1](#), the models are from Burrows et al. (1993). The initial energy burst is contributed by deuterium and lithium fusion, persisting longest in the lowest-mass model

The Jeans mass is a key parameter in many of those theories, representing the mass where gravitational forces in a cloud core just overcome the internal pressure, precipitating collapse. The Jeans mass is given by

$$M_{\text{Jeans}} = \left(\frac{5kT}{G\mu m_H} \right)^{3/2} \left(\frac{3}{4\pi\rho} \right)^{1/2}$$

where T and ρ are the average temperature and density of the gas cloud, m_H is the mass of the hydrogen atom, and μ is the mean molecular weight. This can be rewritten as

$$M_{\text{Jeans}} = 400(T^3/n)^{1/2} \approx 0.35 \left(\frac{T}{10} \right)^{3/2} \left(\frac{P}{10^6 k} \right)^{-1/2} M_{\odot}$$

where n is the number density in molecules cm^{-3} , P is the core pressure, and k the Boltzmann constant.

The Jeans mass sets the lowest-mass star-like object that can form in an ideal gas cloud. Typical temperatures for giant molecular clouds are close to 10 K, while pressures range from $\sim 10^5 k$ for low-density star-forming regions, such as Taurus, to $\sim 10^7 k$ for the cores of giant molecular clouds (GMCs). These values correspond to Jeans masses from ~ 1.1 to $\sim 0.1 M_{\odot}$. In contrast, brown dwarfs with masses as low as $\sim 0.005 M_{\odot}$ ($\sim 6 M_{\text{Jup}}$) are known to exist ([Sects. 3.3](#) and [5.2](#)). Only a fraction, perhaps as little as 10%, of the material in the cloud core contributes to the star or brown dwarf that is the final product of the formation process. Even so, this suggests that low-mass brown dwarfs require cold, very high density environments that are rarely observed in GMCs.

The Jeans mass is an idealized concept, and real GMCs are subject to shocks, magnetic fields, turbulence, and ordered motions due to stellar winds and cloud–cloud collisions that can lead to significant density inhomogeneities. However, the large numbers of brown dwarfs discovered in the field have led some to explore alternative formation scenarios. In particular, brown dwarfs (and, by extension) low-mass stars have been identified as interrupted stars, embryos ejected part way through the formation process.

The ejection hypothesis is founded on theories of star formation through competitive accretion (Bonnell et al. 1997). Those theories envisage stars forming in small groups, with two to four stars per cloud core. Each star within that group is competing for material from the same reservoir, but they are not competing on an equal footing (see [Sect. 5.1](#)). Gravity follows the principle that to those that have, more shall be given, and, as a consequence, the most massive object quickly grows to dominate the group, to the extent it may acquire a circumstellar disk of sufficient mass that they can support formation of brown dwarf mass objects. Dynamical interactions within the group can lead to rapid ejection of the lowest-mass object, which is therefore no longer able to accrete significant mass. If those ejected objects have masses below the hydrogen-burning limit, then they will evolve to form brown dwarfs (Bate et al. 2002).

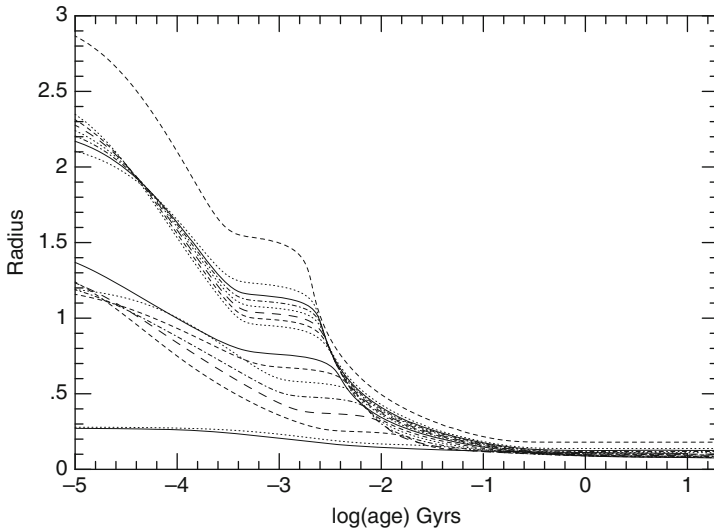
This model makes several predictions: First, the velocity dispersion of brown dwarfs should be higher than stars within the cluster; second, the brown dwarf binary fraction should be extremely low (<5%), since wider binaries are expected to be disrupted dynamically; and, third, brown dwarfs are expected to possess highly truncated circumstellar disks. None of these predictions are supported by observations: Brown dwarfs continue to reside within clusters, so any dynamical accelerations must be limited to less than $1\text{--}2\text{ km s}^{-1}$; the brown dwarf binary fraction is closer to 20% ([Sect. 5.2](#)); and observations of brown dwarfs in young clusters reveal infrared excess radiation, suggestive of the presence of cool material in a surrounding disk, and H α emission, indicating ongoing gas accretion. Thus, while the embryo ejection model is attractive in that it avoids the Jeans mass problem for low-mass stars and brown dwarfs, it has little direct observational support.

2.3 Brown Dwarf Evolution

The fundamental parameter governing brown dwarf evolution is mass. Mass determines the energy generated in the initial collapse, sets the maximum possible central temperature, and hence defines whether any nuclear fusion occurs. Brown dwarf radii are set primarily by degeneracy ([Sect. 2.1](#)); consequently, all brown dwarfs, regardless of mass, have similar radii, and similar surface area. This has two consequences: First, irrespective of mass, all brown dwarfs follow similar evolutionary tracks in the (L , T_{eff}) HR diagram; second, since the rate of energy depletion is essentially independent of mass, higher-mass brown dwarfs, with a larger store of energy, evolve and cool at a slower rate than lower-mass brown dwarfs. These broad characteristics have been codified by Burrows et al. (2001) in the following approximate relations:

$$\begin{aligned} L &\propto \tau^{-1.3} M^{2.64} \\ T_{\text{eff}} &\propto \tau^{-0.32} M^{0.83} \\ R &\propto g^{-0.18} T_{\text{eff}}^{0.11} \approx R_{\text{Jup}} \end{aligned}$$

where τ is age in Gyrs, and g is the surface gravity.



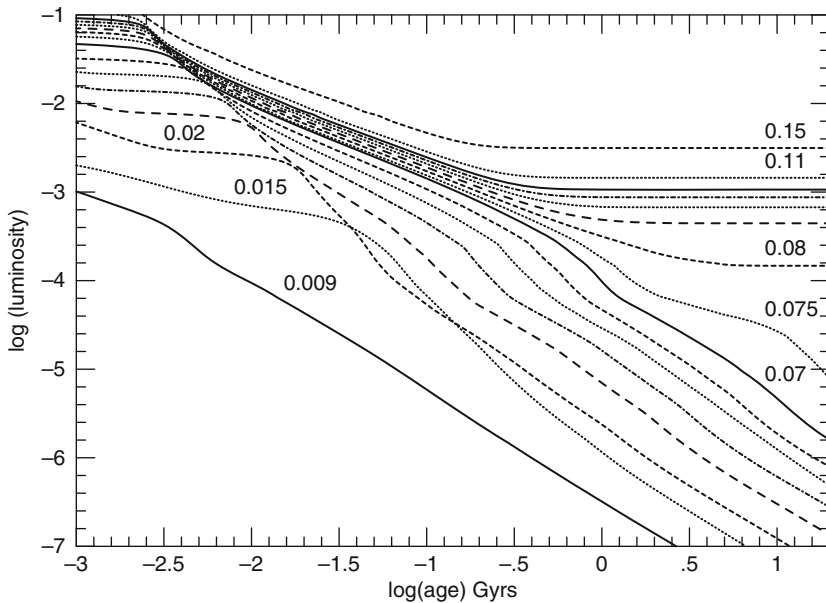
■ Fig. 7-3

Time evolution of radius in low-mass dwarfs. The models range in mass from $0.15 M_{\odot}$ (top to bottom), and are drawn from the calculations by Burrows et al. (1993, 1997)

A more detailed picture of brown dwarf evolution is given by ► Figs. 7-3–7-5 which show the predicted time variation in radius, luminosity, and temperature, and in ► Fig. 7-6, which plots the corresponding evolutionary tracks in the HR diagram. The models span a range of masses, from $0.15 M_{\odot}$ low-mass stars through the brown dwarf regime to $0.009 M_{\odot}$ ($\sim 10 M_{\text{Jup}}$) planetary-mass objects, and show the rapid convergence in radius. The initial radii, at $\tau \sim 100,000$ years, range from 0.3 to $2.4 R_{\odot}$, but within ten million years all dwarfs have radii between 0.15 and $0.3 R_{\odot}$, and the dispersion is less than 25% at $\tau = 1$ Gyr.

► Figure 7-3 shows that contraction pauses at ages of a few million years for masses exceeding $0.012 M_{\odot}$. This stems from additional energy provided by fusion of lithium and deuterium (see ► Fig. 7-2). There is a corresponding standstill in luminosity and temperature (► Figs. 7-4 and ► 7-5). Indeed, ► Fig. 7-5 shows that the surface temperature for higher-mass dwarfs ($M > 0.02 M_{\odot}$) is actually rising at ages between 1 and ~ 10 million years. Thus, the significant decline in luminosity at those ages is driven by contraction and decrease in surface area. As dwarfs age beyond 1 Gyr, the bifurcation between hydrogen fusion-supported stars and brown dwarfs becomes apparent. Stars settle onto the main-sequence at constant temperature and luminosity; transition objects, with masses in the range $0.074 < \frac{M}{M_{\odot}} < 0.078$, achieve temporary core fusion that slows the descent before degeneracy drops the core temperature below $T_{H,\text{crit}}$; and brown dwarfs fade and cool monotonically.

As the surface temperature changes, the chemical species present in the atmosphere also change, leading to evolution of the electromagnetic energy distribution, the photospheric spectrum, and spectral type. In the last context, the approximate temperature boundaries between spectral classes are marked on ► Fig. 7-5. Clearly, all brown dwarfs are classed initially as mid-type M dwarfs, but as they age and cool they evolve through late-M and L to T. ► Figure 7-6 reinforces this conclusion, demonstrating how the evolutionary tracks for different masses run

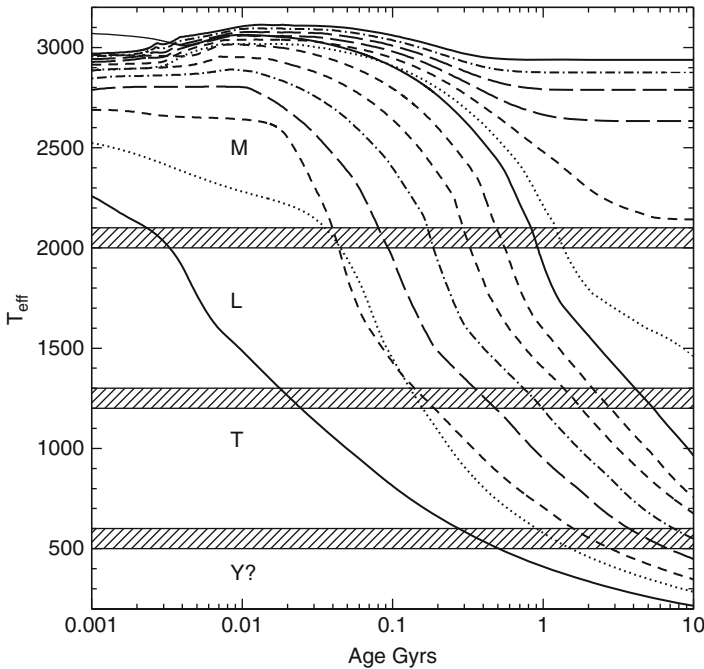


■ Fig. 7-4

Luminosity evolution of brown dwarfs: The evolutionary tracks are from the model calculations by Burrows et al. (1993, 1997). The models are labeled, and masses range from $0.15 M_{\odot}$ (uppermost dashed line) to $0.009 M_{\odot}$ (lowest solid line). The lowest-mass dwarf is not capable of initiating deuterium burning, and therefore fades more rapidly over the initial 10^8 years. The higher-mass objects (stars) achieve equilibrium, and constant luminosity, after ~ 300 Myrs; the $0.075 M_{\odot}$ and $0.070 M_{\odot}$ tracks represent transition objects that are capable of sustaining hydrogen fusion for no more than a few billion years

almost parallel in the HR diagram. This figure highlights a major complication in studying brown dwarfs: There is no simple mass-luminosity (or mass-temperature) relation. In main sequence stars, the energy generated is dependent directly on the total mass, with $L \propto M^{4.5}$ for $M > 1 M_{\odot}$ and $L \propto M^{2.5}$ at lower masses (Smith 1983). As a consequence, the mass of an individual star can be inferred if there are measurements of its apparent magnitude and distance, and hence its luminosity. In contrast, an age measurement is essential to break the inherent degeneracy in the brown dwarf tracks plotted in ● Fig. 7-6, and derive a mass estimate for an individual object.

Two further broad conclusions can be drawn from ● Fig. 7-5 that inform any analysis of brown dwarf demographics. First, regardless of mass, the rate of evolution scales with the temperature; as a result, brown dwarfs spend least time as M dwarfs, a relatively short period as an L dwarf, and the bulk of their visible lifetime as T (and then Y) dwarfs. Second, higher-mass objects evolve more slowly, so they have longer residence times within each spectral type, and they therefore dominate in any sample drawn from the field. Thus, a volume-limited sample of late-type M dwarfs is comprised predominantly of low-mass stars, while the majority of late-type L dwarfs in the same volume are high-mass brown dwarfs, with masses in the range $0.06\text{--}0.072 M_{\odot}$. The latter prediction is confirmed by the scarcity of L dwarfs with detectable lithium absorption.



■ Fig. 7-5

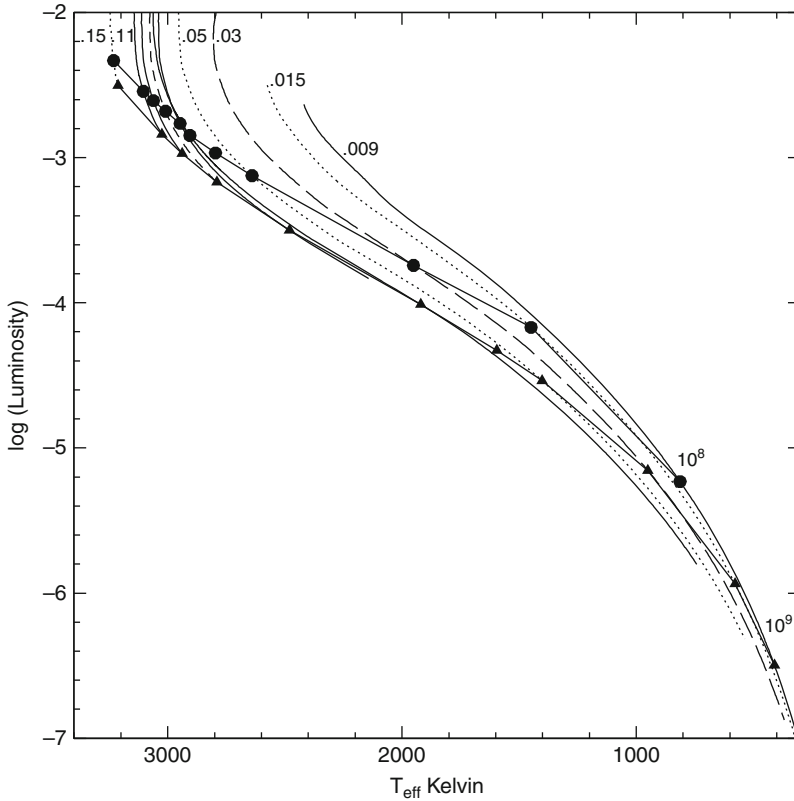
Temperature evolution of brown dwarfs – the models are those shown in [Fig. 7-4](#). The horizontal hatched lines mark the approximate temperature ranges that correspond to the transitions between spectral types M, L, T and (more speculatively) the yet-to-be discovered class Y

2.4 Brown Dwarfs and the Mass Function

Individual brown dwarfs are of very low mass and each makes a correspondingly small contribution to the global mass density. However, intermediate-age and old brown dwarfs can achieve mass-to-light ratios exceeding 10^4 , and, collectively, have the potential to make major contributions to the mass density without producing a visible signature. As a consequence, brown dwarfs and, to a lesser extent, low-mass stars have long been regarded as dark matter candidates. The crucial parameter for determining their viability in this respect is the number of dwarfs as a function of mass, the mass function.

The concept of dark matter first appeared in the astronomical literature (albeit, largely unnoticed) in the 1930s with Fritz Zwicky's observations of velocity dispersions and luminosities of galaxy clusters. The first widespread recognition of the potential presence of substantial quantities of nonluminous material came with Oort's (1932) identification of "missing mass" (actually, missing light) in the local Galactic disk. A comparison between estimates of the local mass density derived from a census of local constituents (gas, dust, and stars), and from the gravitational field inferred from the vertical motions of disk stars at moderate heights above the Plane implied that visible matter accounted for less than half the gravitating material in Solar Neighborhood.

Recent analyses have shown that the Oort-limit missing mass stems primarily from the complex nature of the Galactic disk, notably the dominance of thick disk stars at $|z| > 1$ kpc,



■ Fig. 7-6

Evolutionary tracks for very low-mass dwarfs in the HR diagram. As in previous figures, the data are from the Burrows et al. (1993, 1997) model calculations. As in [Figs. 7-3–7-5](#), the models span the mass range $0.15 M_{\odot}$ to $0.09 M_{\odot}$. Lower-mass dwarfs are more luminous at a given temperature due to their larger radii. Solid points mark the 100-Myr isochrone and triangles the 1-Gyr isochrone

rather than dark matter in the disk (Cr     et al. 1998). Nonetheless, Oort’s conclusions permeated Galactic structure investigations until at least the early 1980s and stimulated considerable interest in determining the local number density of red dwarf stars. The apparent deficit led to numerous investigations of the stellar mass function at low masses that continued through the 1980s and 1990s (see Reid and Hawley 2005, [Chapter 8](#)). Observations focused on two tactical approaches: A complete census of the stellar constituents in the immediate Solar Neighborhood, within 8–10 parsecs of the Sun (e.g., Reid and Gizis 1997; Reid et al. 2008b); or exhaustive surveys for low-mass members of nearby open star clusters, such as the Hyades and Pleiades (e.g., Bouvier et al. 2008; Stauffer et al. 2007).

The results from these investigations have converged to give a consistent picture of the initial mass function (IMF). Introduced by Salpeter (1953), the IMF describes how gas redistributes itself as stars and brown dwarfs within a star-forming region. This fundamental parameter has received considerable observational and theoretical attention in the succeeding half century.

The IMF is often characterized as a power-law,

$$\Psi(M) = \frac{dN}{dM} \propto M^{-\alpha}$$

and can also be cast in logarithmic units

$$\xi(M) = \frac{dN}{d\log M} \propto M^{-\alpha+1}$$

It is important to be clear on which formalism is adopted in a given analysis. The power-law is described as “steep” for large values of α (i.e., rapidly decreasing numbers of stars with increasing mass) and “flat” for $\alpha = 1$. An $\alpha = 1$ IMF has the interesting characteristic that each decade of mass ($0.01\text{--}0.1 M_{\odot}$, $0.1\text{--}1 M_{\odot}$, etc.) contributes the same total mass to the local mass density.

Observations show that the IMF increases steeply with decreasing mass for high-mass stars with a power-law slope close to Salpeters original index, $\alpha = 2.3$, at masses $M > 2 M_{\odot}$. The IMF flattens at masses below $\sim 1 M_{\odot}$, and can be matched as a power-law with $\alpha \sim 1$ (Kroupa 2002). These results demonstrate that low-mass stars are not present in sufficient numbers to make a significant contribution to dark matter (Reid and Gizis 1997).

As an alternative to a power-law, the stellar IMF can be represented in the log-normal form originally proposed by Miller and Scalo (1979),

$$\xi(M) = \text{const.} \exp - \frac{\log M - \log M_c}{2\sigma^2}$$

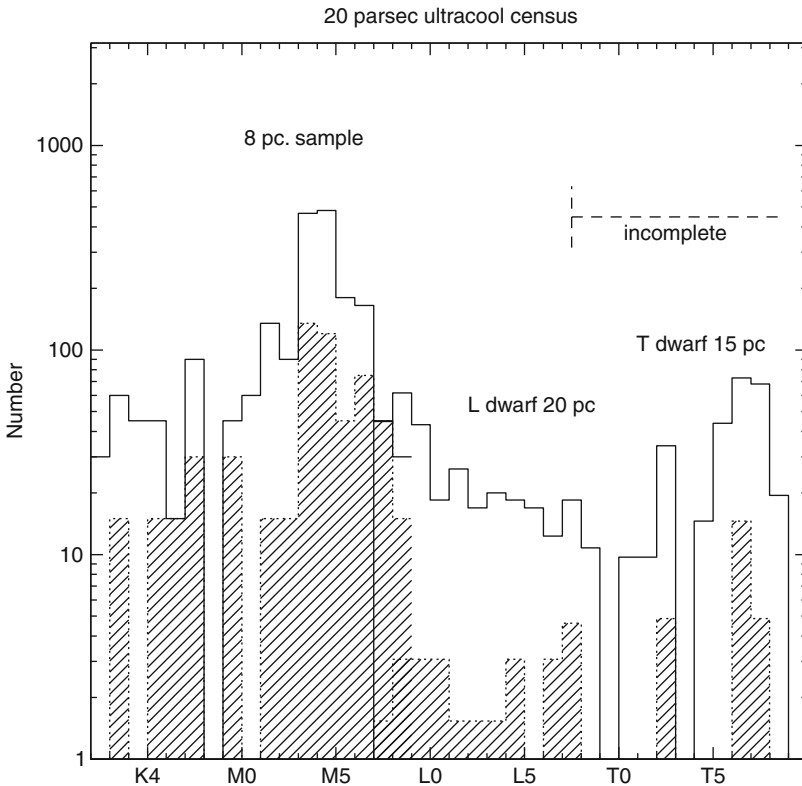
where M_c is the characteristic mass and σ^2 in the variance in $\log M$. These values define the peak and the width of the mass distribution. The log-normal distribution is often grafted onto a Salpeter-slope power-law at high masses. Chabrier (2003) favors $M_c \sim 0.22 M_{\odot}$ and $\sigma = 0.57$. It is important to emphasize that both the power-law and the log-normal distributions are fitting functions, and, as yet, neither is tied to a physical model for the star formation process. Nonetheless, these two mathematical representations have different implications with regard to the expected number-mass distribution when extrapolated below the hydrogen-burning limit into the brown dwarf domain.

Main-sequence stars follow (metallicity-dependent) mass-luminosity relations, allowing masses to be estimated on a star-by-star basis. The same situation does not hold for brown dwarfs, as outlined in [Sect. 2.3](#). Strategies for estimating the parent mass function need to adapt accordingly. In the case of analyses of star clusters and young associations, the observed parameters (magnitudes, colors) are matched against theoretical models for the appropriate age and metallicity. Ideally, one would like to concentrate on the youngest systems, with ages less than 10–20 Myrs, since brown dwarfs are more luminous at young ages, and observations can therefore reach lower masses. The complications, however, are that ages for those clusters have uncertainties of at least 20%, models are less reliable at very young ages, and individual objects can still be surrounded by detritus from the star-forming region and cocooned in circumstellar disks that can influence the measured parameters (see [Sect. 3.3](#)). Moreover, the nearest substantial star-forming regions lie at distances of ~ 150 parsecs, complicating the detection of low-luminosity members and, particularly, binary companion at even modest separations.

Theoretical models are more reliable for intermediate-age systems such as the Pleiades (~ 120 Myrs) and the Hyades (~ 600 Myrs), but dynamical relaxation and energy equipartition within the cluster lead to significant mass segregation, distributing lower-mass stars and brown dwarfs toward the cluster outskirts where they are more vulnerable to stripping. As a result, the

present-day mass functions of those clusters can be significantly depleted at low masses relative to the initial mass function. These issues are discussed in more detail by Bastian et al. (2010), who provide an extensive review of analyses of the IMF in a wide range of environments.

Brown dwarfs and stars in the Solar Neighborhood have ages that span the full lifetime of the Galactic disk, and many have migrated to this locale from widely different initial Galactic radii (Reid et al. 2002). As a result, the local population can be taken as characteristic of much broader swathe of the disk. Figure 7-7 shows the luminosity function of nearby red dwarfs, plotting numbers as a function of spectral type. The parent sample is complete statistically for spectral types K through M to mid-type L dwarfs, but likely underestimates numbers at later types. The latest L-type dwarfs and early-T dwarfs have colors that overlap with those of main-sequence stars (Figs. 7-15 and 7-17), while the low intrinsic luminosities of late-type T



■ Fig. 7-7

The red dwarf luminosity function, plotting the predicted number of stars and brown dwarfs as a function of spectral type for a volume of radius 20 parsecs, centered on the Sun. The results are based on a census of the immediate Solar Neighborhood (Reid et al. 2008b), combining statistics for M dwarfs within 8 parsecs, L dwarfs within 20 parsecs and T dwarfs within 15 parsecs. The hatched histogram shows the contribution from secondary components, and the solid histogram shows the combined total of single dwarfs, primaries and secondaries. The observational sample becomes incomplete close to the L/T transition

dwarfs render them detectable within a very limited volume. Nonetheless, the overall form of the luminosity function is clear, with the densities peaking at mid-type M dwarfs (M3, M4), dropping sharply to a minimum at late-L and rising through T0 to beyond T5.

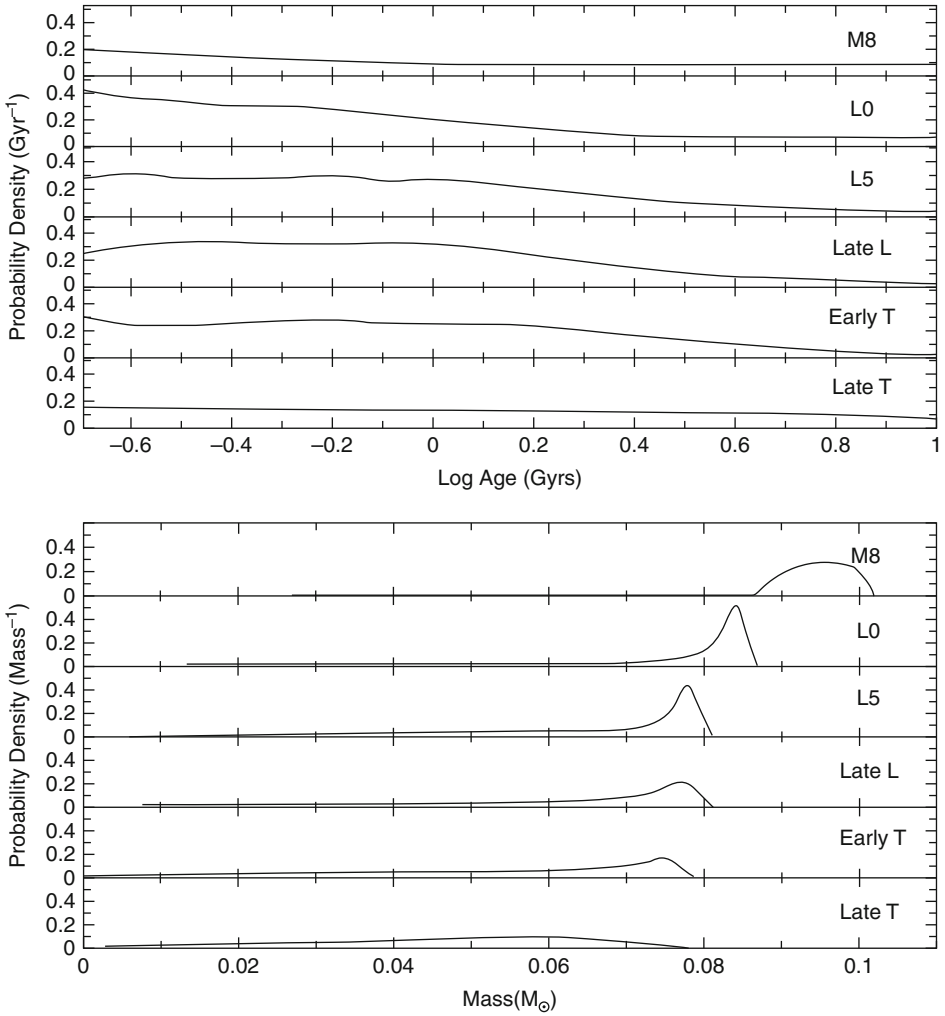
The morphology of the luminosity function reflects how spectral types map onto the underlying mass distribution. Qualitatively, the distribution can be understood in the following terms:

- Spectral types M3 to M8 are dominated by stars. The mass-luminosity relation steepens beyond \sim M4, leading to later spectral types sampling a smaller mass range and a corresponding decline in numbers.
- Early-type L dwarfs are a mix of long-lived, very low-mass stars and brown dwarfs. The latter spend a relatively short time ($< \text{few} \times 10^8$ yrs) in this temperature range (► Fig. 7-5).
- Late-type L dwarfs and early-type T dwarfs are exclusively brown dwarfs (► Fig. 7-7). The relative numbers therefore drop further, since the residence times at the appropriate temperatures are significantly less than 1 Gyr for even the most massive brown dwarfs.
- Number densities increase through mid-type and late-type T dwarfs since the rate of evolution slows, and brown dwarfs of all masses start to pile up at those temperatures (► Fig. 7-5).

Brown dwarfs in the field lack individual age estimates; consequently, statistical analysis must be applied to derive constraints on the mass function from data such as those plotted in ► Fig. 7-7. Coupling an assumed star formation history with evolutionary models allows one to define probability distributions for mass and age in the temperature range that correspond to different spectral types. ► Figure 7-8 shows predictions for models that assume a constant star formation history in the disk over the last 10 Gyrs. This model predicts relatively narrow mass distributions within each range of spectral types, and the resultant age distributions depend on the lifetimes of dwarfs of those masses. Thus, in this model, M dwarfs are dominated by stellar masses, and therefore have a flat age distribution that matches the assumed star formation history. Brown dwarfs make an increasing contribution at later types, although longer-lived transition objects still make the largest contribution even in later-type L dwarfs.

Turning this analysis around, one can use these mass and age distributions to predict the expected luminosity function as a function of the mass distribution, and match those predictions against the observations. Initial analyses focused on the dark matter issue, and whether there evidence for steepening mass function at brown dwarf masses. The conclusion was that the observational data could be matched by a power-law mass function with $1 < \alpha < 1.5$, consistent with extrapolating the stellar mass function below the hydrogen-burning limit, but eliminating brown dwarfs as likely dark matter (Reid et al. 1999). Subsequent investigations, based on more extensive samples, suggest that the mass function is significantly flatter. ► Figure 7-9 matches model predictions against the Solar Neighborhood luminosity function, plotted against J-band ($1.2 \mu\text{m}$) absolute magnitudes; adopting a power-law representation, the data suggest $\alpha < 0.4$ in the substellar régime. Similarly, analysis of T dwarf statistics from SDSS results in an inferred local number density of $0.007 \pm 0.003 \text{ pc}^{-3}$ (Metchev et al. 2008). That sample is likely to be incomplete to some degree, so this represents a lower limit. Nonetheless, the derived value is consistent with a power-law index of $\alpha \sim 0$, again suggesting that the mass function flattens at very low masses.

Observations of star clusters lead to similar conclusions. ► Figure 7-10 shows a compilation of results for systems including the IC 348 and the Orion Nebula clusters (both $\tau \sim 1\text{--}2$ Myrs), the much sparser λ Orionis association ($\tau \sim 5$ Myrs) and the more mature Pleiades cluster

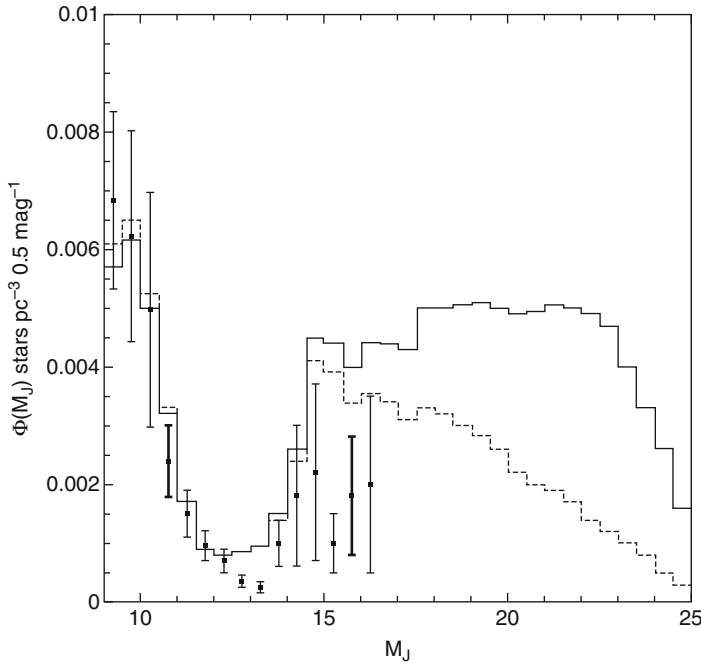


■ Fig. 7-8

The expected age and mass distribution of ultracool dwarfs in the general Galactic field as a function of spectral type. The predictions assume a constant birthrate, a flat mass function ($\alpha = 1$) and an age of 10 Gyrs for the Galactic disk, and use the Burrows et al. (1993, 1997) models to predict the temperature/spectral-type distribution (From Allen et al. 2005)

($\tau \sim 120$ Myrs). Despite the range in ages and environmental properties, these clusters have mass functions with broadly similar characteristics, including a significant downturn in number densities below $0.1\text{--}0.15 M_{\odot}$. The data are well matched by either a broken power-law with $\alpha \sim 0.5$ for $M < 0.11 M_{\odot}$, or by a log-normal distribution (e.g., Chabrier 2003). By and large, there is little evidence that these characteristics vary in a systematic way with the star-forming environment (Bastian et al. 2010).

The observational consensus is therefore that the IMF is flat ($\alpha \sim 1$) for low-mass stars ($0.11 < \frac{M}{M_{\odot}} < 0.9$) and decreasing ($\alpha < 0.5$) at lower masses. As yet, a consensus theoretical



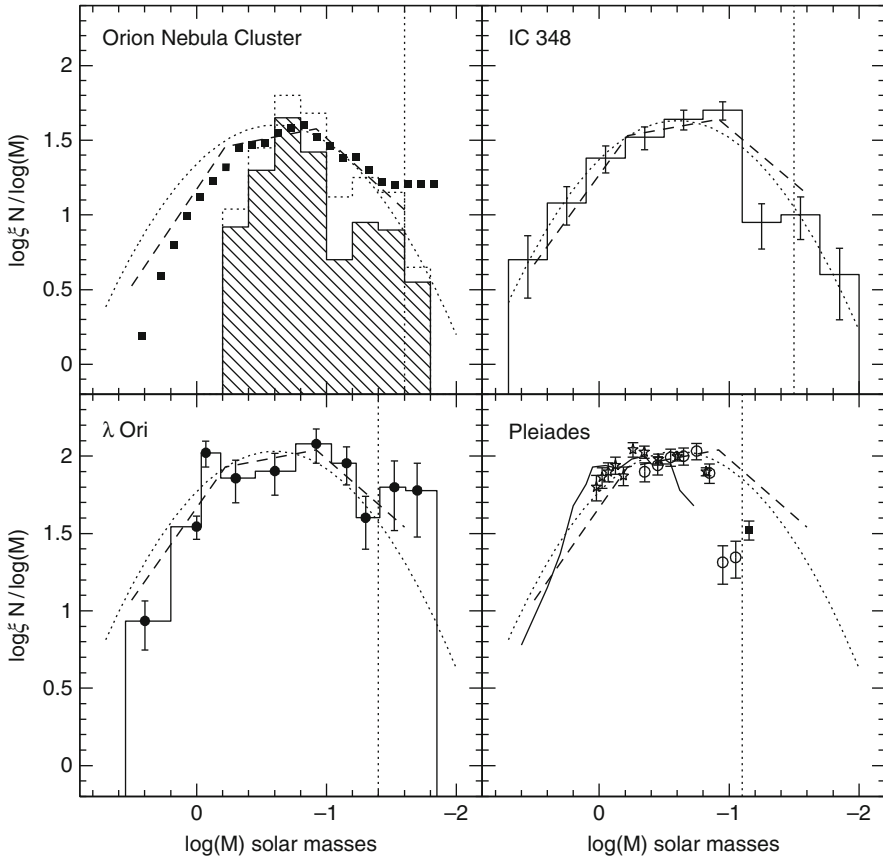
■ Fig. 7-9

A comparison between the J-band ($1.2\ \mu\text{m}$) luminosity function defined by local ultracool dwarfs (solid points with error bars) and theoretical predictions based on a constant birthrate and a power-law mass function for $\alpha = 0.8$ (solid line) or $\alpha = 0.4$ (dotted line). While the observational data are likely to be incomplete at the faintest magnitudes, the data appear to favor a relatively flat mass function (From Allen et al. 2005)

explanation for the overall morphology of the IMF has yet to emerge. Nor is there consensus on whether there is a formal low-mass cutoff to the IMF, although observations of individual objects, such as the wide binary 2M1207-39B (► Sect. 5.2), show that brown dwarfs can form with masses as low as $\sim 0.005\ M_{\odot}$. However, even if brown dwarf formation spans almost two decades in mass, and the IMF is extrapolated to sub-Jovian mass, the net result is that brown dwarfs are less common than stars by number, and contribute less than 10% of the mass density due to stars with masses between 0.1 and $1.0\ M_{\odot}$. Brown dwarfs have now been eliminated as viable candidates for dark matter.

3 Observational Properties of Brown Dwarfs

Almost two decades have elapsed since the discovery of the first brown dwarf. Those years have not only seen the identification of several thousand ultracool dwarfs, many likely of substellar mass, but also compilations of the first local census and extensive observations over a wide range of wavelengths of the brightest representatives. This section summarizes the key observational characteristics of those dwarfs.



■ Fig. 7-10

Mass functions for a representative sample of clusters and associations, ranging from the τ 5 Myr λ Orionis association to the ~ 120 Myr Pleiades cluster. In each case, the dotted vertical line marks the approximate completeness limit of the survey; the dashed line shows the three-component power-law fit to the ONC data; and the dash-dotted line plots the log-normal mass function favored by Chabrier (2003) (see Reid and Hawley 2005, Figs. 9.14 and 9.18 for further details)

It is important to recognize that, while brown dwarfs are defined as a class by their mass, their appearance depends primarily on their effective temperature. Thus, a 75 Jupiter-mass brown dwarf with a photospheric temperature of 3,000 K is observationally remarkably similar to a $\sim 0.2 M_{\odot}$ M5 star, and a 1,600 K brown dwarf will be characterized as spectral type L4 regardless of whether the mass is $5 M_{\text{Jup}}$ or $60 M_{\text{Jup}}$, although there are subtle spectral differences in both cases.

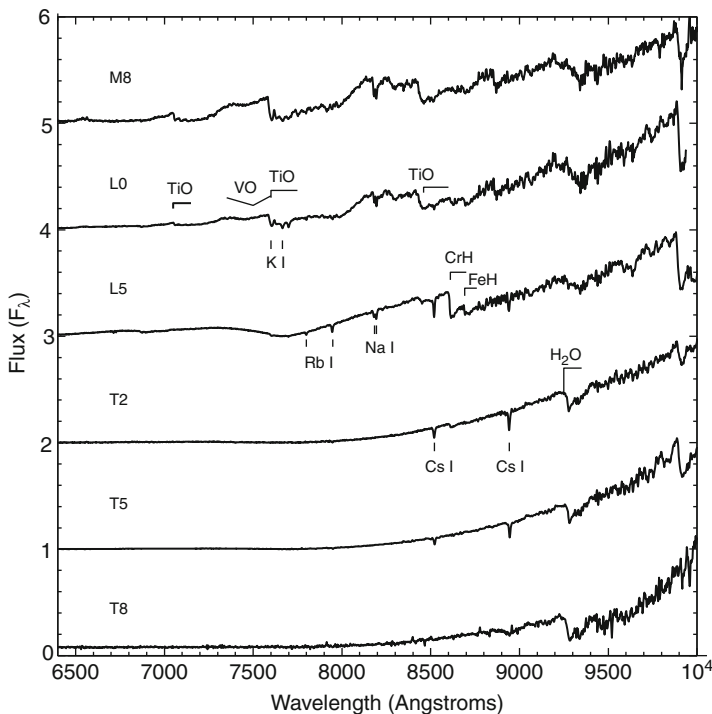
3.1 Spectroscopic Properties

Spectral types provide astronomers with a shorthand method of describing the overall characteristics of a particular star in a manner that is immediately understandable by other

astronomers. The MKK spectral sequence, OBAFGKM, runs from high to low temperature. Within that sequence, spectral types are defined in a qualitative, rather than quantitative, manner, assigned by comparison with standard stars. In the case of classes K and M, classifications are based on the appearance in the far-red of the optical spectrum, at wavelengths from 6,000 to 10,000 Å. For more than 50 years, M dwarfs represented the lowest extreme of the hydrogen-burning main sequence, but the discovery of GD 165B and a few other dwarfs demonstrated that an extension to the classical sequence was required.

Optical spectral classification: M dwarf spectra are dominated by molecular absorption, notably TiO, although metal hydrides such as CaH, MgH, and FeH are also prominent and VO absorption appears in the latest types. The initial optical/red spectra obtained of GD 165B, the L dwarf archetype, show strong hydride absorption, but, confusingly, only weak TiO. As more sources were identified and spectroscopic observations were acquired, clear patterns emerged in the appearance and disappearance of individual features. These systematic changes permitted the definition of two new spectral classes, type L and T. The gross characteristics are illustrated in [Fig. 7-11](#), which shows spectra at far-red optical wavelengths of representative dwarfs from spectral type M8 to T8.

Qualitatively, the prime distinction between spectral classes M and L is the diminishing strength of TiO absorption. VO persists among the earliest-type L dwarfs, but decreases rapidly



■ Fig. 7-11

Optical spectra of M, L, and T-type brown dwarfs. The effective temperature ranges from ~2,400 K at M8 through L5 to ~1,600 K to ~900 K at T5. The most prominent atomic and molecular features are labeled

in strength, disappearing by type L2. Metal hydrides are the dominant molecular features, including CaH, CrH, and FeH in the far red, and MgH shortward of 6,000 Å supplemented by water absorption near 9,400 Å in the later L dwarfs. In addition, resonance lines due to alkaline atoms, K I, Cs I, Rb I, and the Na I D lines, become increasingly prominent in the spectrum. The behavior of the D lines and K I 7,665/7,699 Å doublets are particularly notable. Pressure broadening within an increasingly transparent atmosphere leads to their forming features with full width of almost 1,000 Å in the latest-type L dwarfs. These trends continue in far-red spectra of T dwarfs, with the metal hydride and atomic absorption features weakening as H₂O absorption strengthens from T0 through T5.

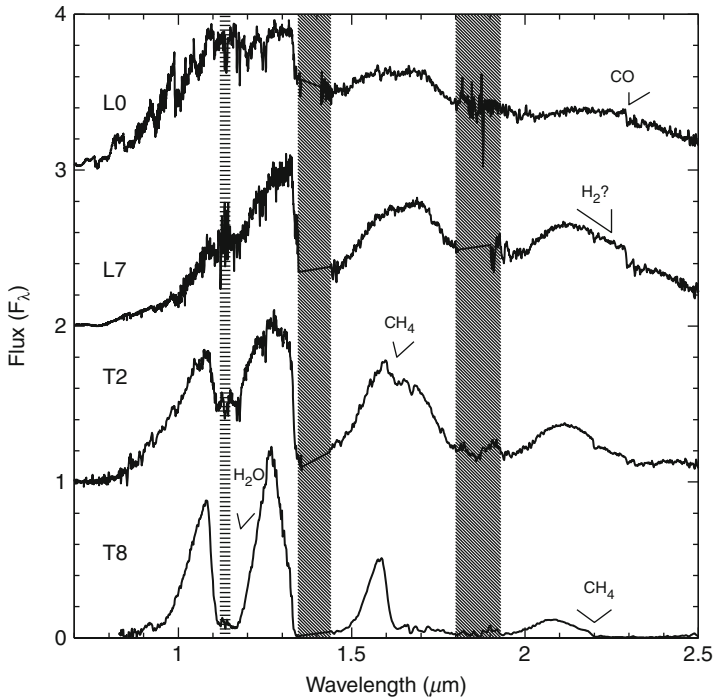
The spectral-type sequence can be quantified using a number of narrowband indices, designed to measure the relative strengths of the more prominent features. A full discussion of the definition of those indicators is given by Kirkpatrick et al. (1999) and Kirkpatrick (2005). The physical causes for the changes in spectral appearance are discussed in ► Sect. 4.

Infrared spectral classification: Spectral types for M and L dwarfs are generally derived based on the appearance at far-red optical wavelengths. The key discriminator between spectral classes L and T, however, lies at near-infrared wavelengths. L dwarfs, like M dwarfs, exhibit broad absorption bands at 1.4, 1.8, and 2.4 μm due to H₂O. The latter coincide with water bands in the terrestrial atmosphere, but the higher temperatures and pressures in the stellar atmospheres broaden the wings, and the full line profiles can be mapped from high, dry sites, such as Mauna Kea. Other absorption features are present, including features due to CO at 2.3 μm, and FeH and atomic lines (Na I, K I, Fe I, Mg I) in the 1.1–1.2 μm region. The hydrides and atomic features mirror the behavior of those species at optical wavelengths, weakening in later-type L dwarfs. Collisionally induced H₂ absorption appears at 2.2 μm in the latest-type L dwarfs, flattening the flux distribution shortward of the CO feature (► Fig. 7-12).

The most pronounced change in the near-infrared spectrum occurs with the appearance of overtone bands of methane absorption at 1.6 and 2.2 μm. Their appearance marks the transition from type L to T, and they rapidly grow in strength from T0 to ~T6, where they saturate. At the same time, the H₂O band at 1.1 μm strengthens. That band provides a cleaner determination of spectral type for late T dwarfs (see Burgasser et al. 2001, for a thorough description of near-infrared classification criteria).

Spectral features due to molecular species are also present at mid-infrared wavelengths, a region that can be observed imperfectly through the atmosphere even from high, dry sites such as Mauna Kea or Paranal. Nonetheless, ground-based observations through the 3 μm window demonstrated that the CH₄ 3.3 μm fundamental band first becomes apparent in mid-type L dwarfs. IRAS, the Infrared Astronomical Satellite, and ISO, the Infrared Space observatory, opened up these wavelengths in the 1980s and 1990s, but neither satellite had sufficient sensitivity to detect the relatively few ultracool dwarfs known at the time. Widespread photometric observations of such objects became possible with the Spitzer observatory, launched in 2002, although only the brighter L and T dwarfs proved accessible to spectroscopy.

► Figure 7-13 presents low-resolution 5.5–14.5 μm spectra of L and T dwarfs obtained with the Infrared Spectrograph (IRS) on Spitzer. This wavelength range includes absorption features due to H₂O, the ν₂ fundamental band at 6.27 μm and the 2ν₂ – ν₂ band at 6.42 μm, which combine to give an absorption feature at ~6.5 μm; the ν₄ CH₄ band at 7.65 μm, which first becomes evident in mid-type L dwarfs; and the ν₂ fundamental band of NH₃, which strengthens in late-type T dwarfs. All these features are characteristic of temperatures significantly cooler than the formal photospheric temperatures for the target ultracool dwarfs, and therefore provide insight into the properties of the atmosphere at higher levels. Spitzer spectra of mid- and late-type L



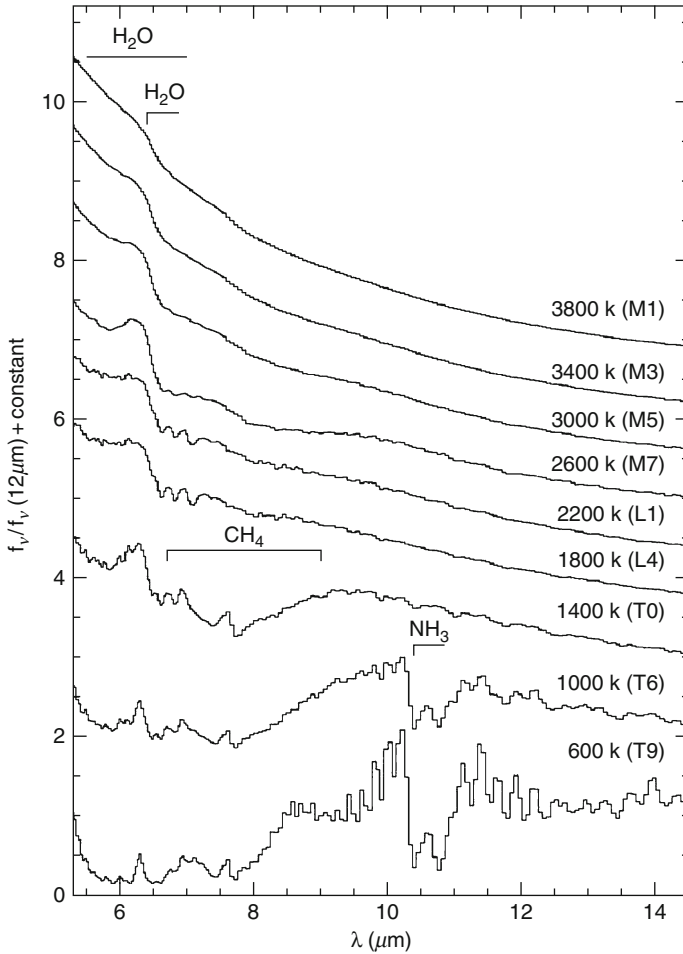
■ Fig. 7-12

Near-infrared spectra of L- and T-type brown dwarfs (Adapted from Reid and Hawley 2005). As in ► Fig. 7-11, the most prominent spectral features are labeled. Note, in particular, the onset of methane absorption bands that define spectral type T

dwarfs also show evidence for broad absorption at $10\ \mu\text{m}$ that might be attributed to silicates and dust formation in those cool atmospheres. These observations give pointers to the likely appearance of dwarfs cooler than type T, the Y dwarfs discussed in ► Sect. 4.3.

3.2 Photometric Properties

Spectral types provide one means of characterizing brown dwarfs; broadband photometry provides another. ► Figure 7-14 shows how the standard photometric passbands sample the spectral energy distribution of late-type dwarfs. The spectral energy distribution peaks strongly near $1.2\ \mu\text{m}$, demonstrating why near-infrared surveys were crucial in searching for cool dwarfs. Integrating the flux distribution, less than 0.5% of the total energy emitted by an M8 dwarf emerges shortward of $7,000\text{\AA}$ and that fraction decreases sharply at later spectral types. The near-infrared photometric passbands are chosen to avoid H_2O absorption bands in the terrestrial atmosphere, but intersect with the broader steam bands in M and L dwarf spectral energy distributions.

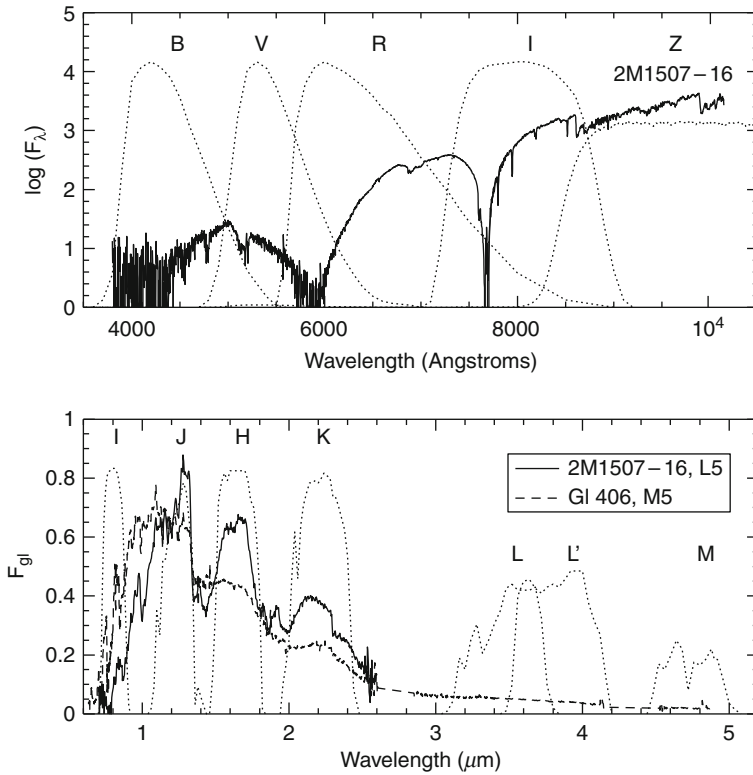


■ Fig. 7-13

Low-resolution mid-infrared spectra of L and T dwarfs obtained with the Spitzer Infrared Spectrograph; the most prominent spectral features are identified (From Cushing et al. 2006)

The evolution in broadband colors of ultracool dwarfs is driven primarily by the changing temperature and corresponding spectral evolution. Figure 7-15 presents representative far-red optical and near-infrared color-magnitude diagrams for M, L, and T dwarfs with trigonometric parallax measurements. Optical/far-red colors generally become monotonically redder with later spectral type for M dwarfs, reflecting the overall steepening of the spectral energy distribution at those wavelengths. There is a noticeable steepening in optical color-magnitude relations at M7. Similar behavior is seen in colors that span far-red to near-infrared wavelengths (e.g., (I–J)), and there is significantly higher dispersion in the T dwarf colour-magnitude distribution.

The near-infrared (J–K) colors show the most dramatic variations. In M0–M7 dwarfs, the J and K passbands both lie longward of the peak spectral energy distribution and are relatively

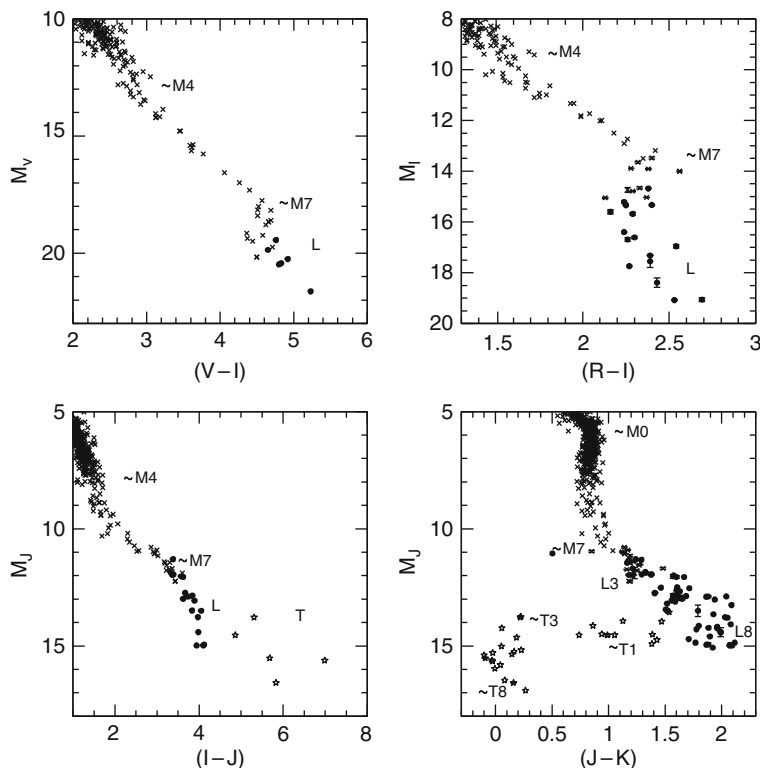


■ Fig. 7-14

Bandpasses of the main optical and near-infrared filters matched against spectroscopic data for the L5 dwarf, 2M1507-1627 (*in both panels*) and the M6 dwarf, Gl 406 or Wolf 359 (*lower panel only*) (Adapted from Reid and Hawley (2005))

unaffected by absorption features, with the result that those dwarfs all have (J–K) ~ 0.8 magnitudes. As the energy peak moves through the J band, an increased proportion of the total flux shifts to the K band, and (J–K) becomes redder with later spectral type, albeit with increasing dispersion for mid- and late-type L dwarfs. That trend reverses near the L/T transition as methane absorption eats into the flux emitted within the H and K bands. Consequently, early-type T dwarfs have near-infrared colors that are comparable with main-sequence FGK stars in the JHK plane, rendering their detection extremely difficult without optical/far-red data to supplement near-infrared colors. As the (I–J) color-magnitude relation shows, the latter data need to have high sensitivity if surveys are to be extended beyond a few parsecs from the Sun. This is the key factor that has enabled the identification of so many T dwarfs from SDSS data.

The first generation of surveys for ultracool dwarfs focused primarily on selecting objects with extreme colors. Any survey that uses color selection as a criterion runs the risk of introducing bias in the color distribution of the final sample; indeed, this is inevitable, unless the targets have colors that are distinct from any contaminant. Such is not the case for ultracool dwarfs. Fortunately, SDSS included both photometric and spectroscopic surveys, and the spectroscopic data provide a means of testing for biases in the photometric selection. Schmidt et al. (2010) have

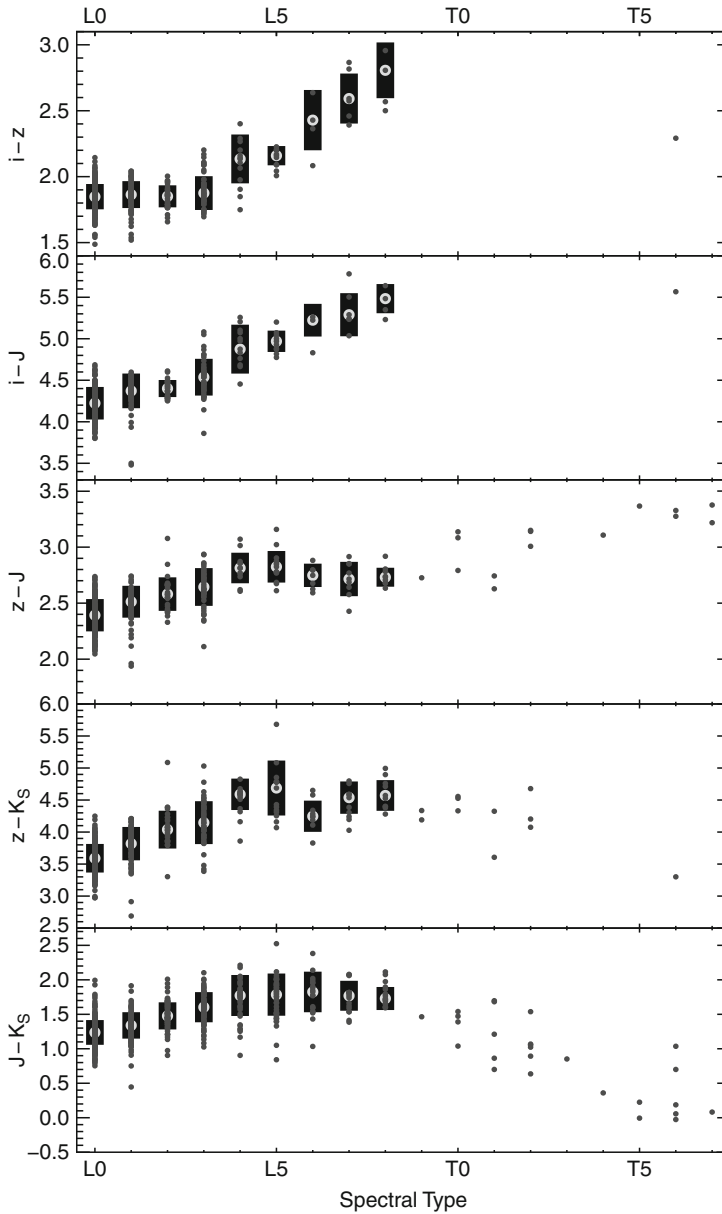


■ Fig. 7-15

Representative color-magnitude diagrams for late-type stars and ultracool dwarfs. In each figure, M dwarfs are plotted as crosses, L dwarfs as solid points, and T dwarfs as 5-point stars. Optical photometry of T dwarfs is scarce, even in the I passband

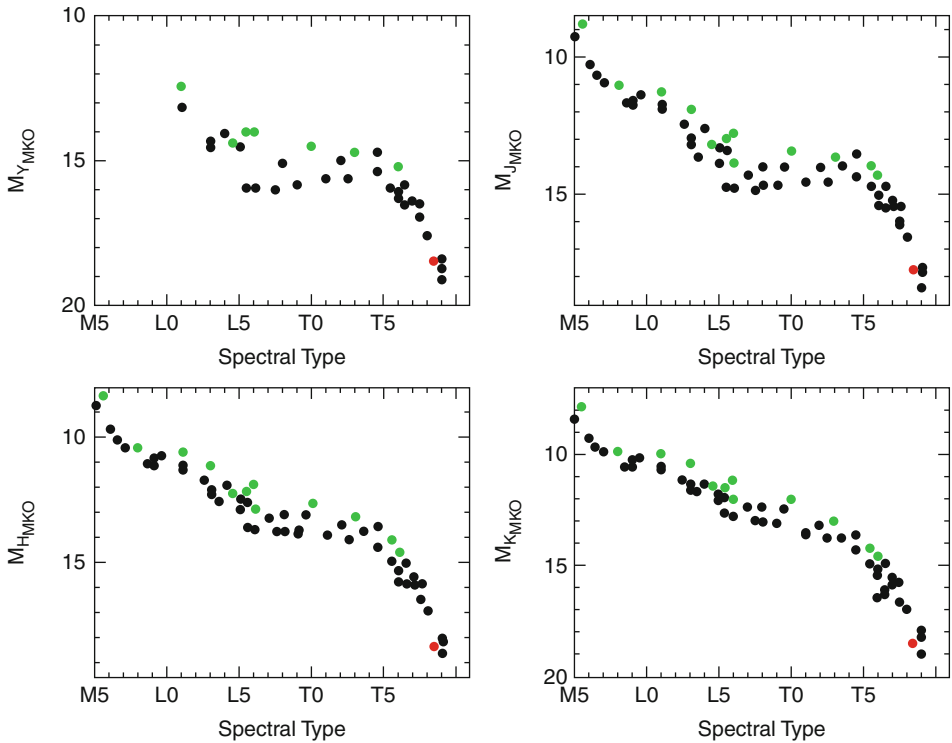
analyzed the *iz*JK color distributions of ultracool dwarfs identified within the SDSS (the Sloan *z* passband has an effective wavelength of $\sim 9,000\text{\AA}$), and the results are shown in [Fig. 7-16](#). These data show a broader spread in color, particularly to bluer (J–K) colors for L dwarfs, than in photometrically selected samples. However, current indications are that this selection bias has a relatively small impact on quantitative calculations, such as the Solar Neighborhood luminosity function plotted in [Figs. 7-7](#) and [7-8](#). Nonetheless, further investigations of the implications of these and other potential biases clearly need to be undertaken.

Absolute luminosities can also be estimated for individual objects based on their spectral types. M dwarfs span such a wide range of properties within a given spectral subtype that this method offers a poor means of estimating absolute magnitudes and distances. However, [Fig. 7-17](#) shows that the near-infrared absolute magnitudes, particularly M_J , are well correlated with spectral types, at least for L dwarfs. Indeed, the color distributions plotted in [Figs. 7-16](#) demonstrate that, unless one has I-band (or i-band) photometry, spectral types offer the most effective means of estimating absolute magnitudes for L and T dwarfs. However, the relations are not single-valued, with distinct flattening in both J and K near the L/T transition. The M_J -spectral-type distribution even suggests a reversal, with $\sim T2$ dwarfs brighter than



■ Fig. 7-16

Color distributions of ultracool dwarfs as a function of spectral type for a range of optical and near-infrared colors. The data are taken from the extensive SDSS-based survey by Schmidt et al. (2010). Individual dwarfs are plotted as dots; the gray circles mark the median colors, and the solid bars the one-sigma dispersions for each spectral type. Note that the $(z-J)$ and $(z-K)$ colors essentially saturate at L3, and therefore offer very limited constraints on spectral type and absolute magnitude; indeed, the $(z-K)$ color follows $(J-K)$ in showing a reversal for mid- and late-type T dwarfs, as methane absorption becomes prominent in the K-band



■ Fig. 7-17

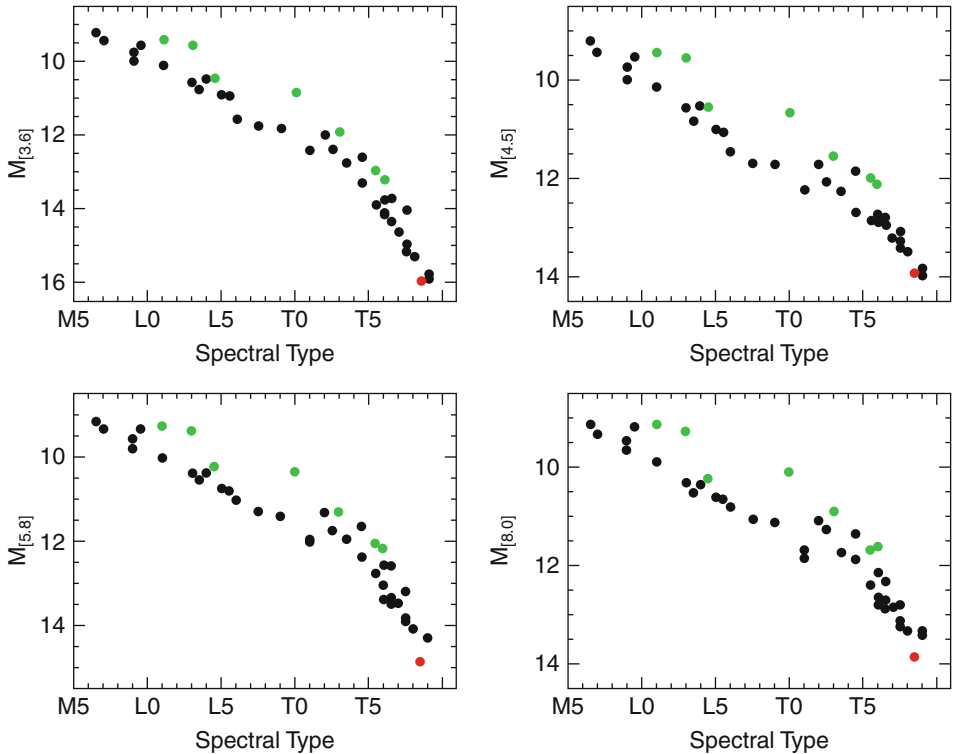
The absolute-magnitude/spectral-type relations at near-infrared wavelengths for M, L, and T dwarfs (From Leggett et al. 2010); lighter points are known binary systems. All passbands show a distinct flattening through the L/T transition, with a suggestion of an upturn in M_J ; that is, early-type T dwarfs appear to be brighter than late-type L dwarfs at $1.2\ \mu\text{m}$. Unrecognized binaries contribute to this feature, but, as discussed in the text, are unlikely to account fully for the feature

the latest-type L dwarfs. This behavior must at least partly reflect the onset of methane absorption, which reduces the luminosity at K and potentially redistributes flux from the H and K windows to the J passband.

A secondary effect, which may contribute significantly to the apparent flattening in M_K , is unrecognized binarity. As discussed further in [Sect. 5](#), low-mass binaries tend to have components with near-equal masses. Both evolutionary models and the luminosity function ([Fig. 7-7](#)) indicate that there is rapid evolution through the L/T transition. That evolution is accompanied by radical changes in the spectral energy distribution at near-infrared wavelengths. Brown dwarf binaries are therefore likely to go through a phase where the two components have similar J-band luminosities, but the more massive component has spectral type late-L while the lower-mass component has evolved through the L/T transition. If those components are unresolved, the composite spectrum will be approximately consistent with an early-type T dwarf, albeit with somewhat unusual bandstrengths, but the measured near-infrared absolute fluxes are enhanced compared with a single dwarf. Consequently, if the parallax is known and the combined data are plotted in the HR diagram, the system will appear to be overluminous, particularly in M_J . Observational support for this hypothesis comes from

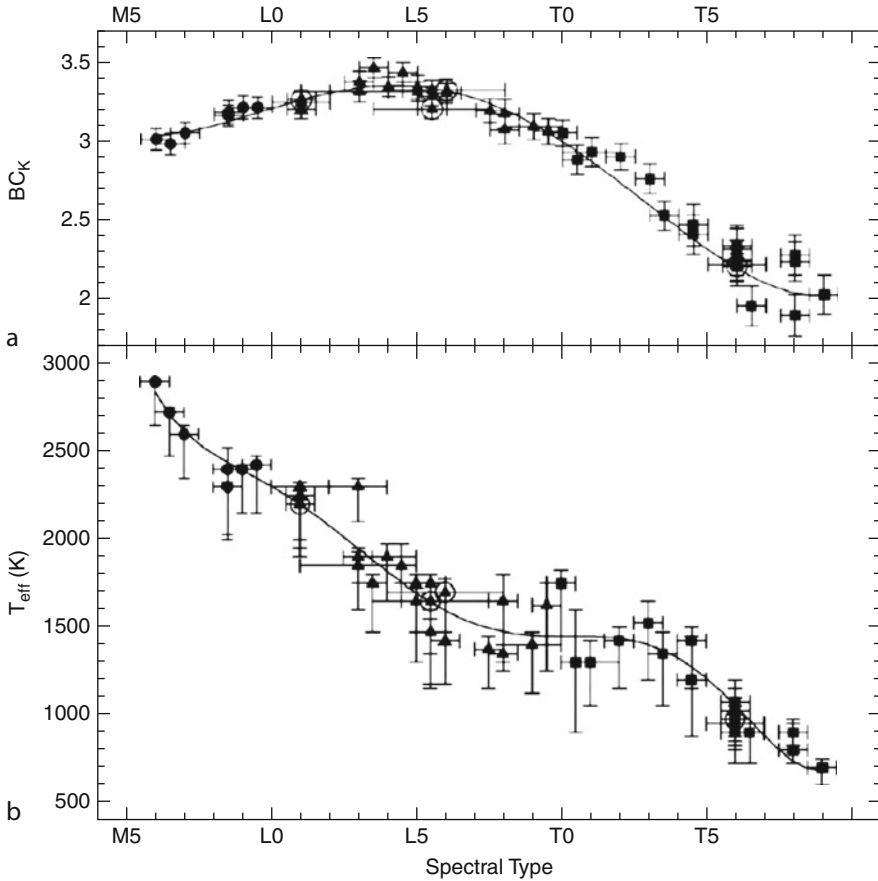
indications of significantly higher binary fractions at spectral types L8–T4, and binarity likely accounts for almost all of the effect in M_K (Burgasser 2007). However, photometry of resolved binary components suggests that even removing unresolved systems leaves a residual “bump” in the mean relation in the M_J /spectral-type plane.

Photometric data provide a reliable means of estimating bolometric luminosities and effective temperatures. In the case of ultracool dwarfs, observations are required that extend to mid-infrared wavelengths. Ground-based observations longward of $3\ \mu\text{m}$ are possible for only the brightest ultracool dwarfs. In general, (K–L) colors ($[2.5\text{--}3.8]\ \mu\text{m}$) increase almost linearly with spectral type from ~ 0 at M0 to ~ 2.0 magnitudes at T5; (L–M) colors ($[3.8\text{--}4.75]\ \mu\text{m}$) trend blueward from ~ 0 at M0 to ~ -0.5 mag. at L5, but reverse to reach ~ 0.5 mag. for the latest T dwarfs. Initial ground-based data have been supplemented by extensive mid-infrared observations with Spitzer, permitting calibration of absolute magnitude/spectral-type relations at mid-infrared wavelengths. Those relations show a distinct steepening as one moves from spectral L to T, reflecting the onset of significant absorption at mid-infrared wavelengths (► Fig. 7-18). Combined, the near- and mid-infrared data allow bolometric corrections to be derived for a representative set of late-M, L, and T dwarfs (► Fig. 7-19).



■ Fig. 7-18

The absolute-magnitude/spectral-type relations at mid-infrared wavelengths for M, L, and T dwarfs (From Leggett et al. 2010); as in ► Fig. 7-17, lighter points are known binary systems. In each case, the data indicate a significant change in slope after the L/T transition



■ Fig. 7-19

Bolometric corrections and effective temperatures for ultracool dwarfs as a function of spectral type (From Golimowski et al. 2004). The solid lines plot fourth-order and sixth-order polynomial fits. Note the small range in temperature spanned by late-L and early-type T dwarfs

If the absolute bolometric magnitude is known, the effective temperature can be derived using

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

where σ is Stefan's constant. In the case of M dwarfs, radii have been measured from eclipsing binaries, such as YY Gem (M0) and CM Dra (M4), and through interferometry. The derived temperatures span a range of $\sim 1,400$ K, from $\sim 3,800$ K M0 dwarfs to 2,300–2,400 K ultracool M9s. Spectral types L and T extend the sequence to lower temperatures. Deriving quantitative temperature estimate for ultracool dwarfs is simplified by the fact that they show little variation in radius with mass range, except at very young ages (► Fig. 7-3). All dwarfs plotted in ► Figs. 7-17 and ► 7-19 are drawn from the field, and can therefore be expected to have ages of a few Gyrs (► Fig. 7-8). Theoretical models can be used to provide estimates of their radii. ► Figure 7-19 shows the resulting effective temperature/spectral-type relation.

The inferred temperatures range from $\sim 2,200$ K at L0 to $\sim 1,400$ K at the L/T transition, with the latest-type T dwarfs assigned temperatures below 700 K. The physical causes for the changes in spectral appearance over this temperature range are discussed in [Sect. 4](#).

As in the M_J /spectral-type plane, the temperature/spectral-type distribution exhibits a pronounced flattening from $\sim L5$ to $\sim T4$. This is not coincidental: The inferred temperatures scale with the bolometric luminosity, and the near-infrared is the dominant contributor to the total flux ([Fig. 7-14](#); thus, similar near-infrared absolute magnitudes ([Fig. 7-17](#)) lead to similar temperature estimates. The observations suggest that L5 to T4 dwarfs have temperatures in the range 1,250–1,500 K. Theoretical models predict monotonic temperature evolution through this temperature range ([Fig. 7-5](#)). Thus, these results strongly suggest that the onset of methane absorption, and evolution through the L/T transition, is a relatively rapid process.

There is an important caveat to bear in mind in interpreting [Fig. 7-19](#). In deriving the effective temperature, one is making the implicit assumption that the spectral energy distribution, and hence the measured luminosity, is generated at an approximately uniform depth within the atmosphere. This is unlikely to be the case for at least T and Y dwarfs, as discussed further in [Sect. 4.1](#). Consequently, the effective temperature scale is likely to be more figurative than literal for the coolest dwarfs.

To date, surveys for ultracool dwarfs have relied on ground-based technology, progressing from red-sensitive photographic plates to search for late-type M dwarfs, through near-infrared sky surveys to search for L dwarfs, to combinations of deep, far-red CCD data and near-infrared surveys to search for T dwarfs. Each step required increased sensitivity to probe sources of fainter intrinsic luminosity. Deep near-infrared ground-based surveys, such as the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007), are currently underway to find even cooler dwarfs, sub-500K “Y dwarfs,” which are expected to be characterized by ammonia absorption (see [Sect. 4.3](#)). Those surveys are just starting to reveal interesting sources, with close to a dozen dwarfs classified as T8 or T9. Most recently, UKIDSS has resulted in the identification of several dwarfs with spectral types later than T7. Those include the nearest brown dwarf to the Sun, with a preliminary spectral type estimated as T10, a temperature close to 500 K and a distance of ~ 2.9 parsecs (Lucas et al. 2010).

The future, however, may well lie with space-based mid-infrared surveys. Analyses of Spitzer data (Leggett et al. 2010) and predictions based on theoretical models (Saumon and Marley 2008) suggest a number of color combinations that might prove effective in finding very cool dwarfs. In particular, the $([3.6\text{--}4.5] \mu\text{m})/([4.5\text{--}5.8] \mu\text{m})$ two-color plane may act as an effective means of segregating extremely cool dwarfs from other sources. Observations show that late-type T dwarfs, which are extremely red in the former color, but blue in the latter. The Wide-field Infrared Survey Explorer (WISE), launched in December 2009, was designed to survey the whole sky at those wavelengths, and may prove responsible for the first identification of a room-temperature (300 K) brown dwarf.

3.3 Observations of Young Brown Dwarfs

Youthful brown dwarfs are scarce in the immediate vicinity of the Sun, as might be inferred from [Fig. 7-8](#). However, within the last decade it has become apparent that a number of relatively sparse groups and associations of young stars lie within 50 to 100 parsecs of the Sun (Zuckerman and Song 2004). Those include the TW Hydrae Association ($d \sim 50\text{--}100$ parsecs, $\tau \sim 8$ Myrs), β Pictoris moving group ($\sim 20\text{--}40$ parsecs, ~ 12 Myr), Tucana/Horologium Association

($\sim 40\text{--}60$ parsecs, ~ 30 Myr), and AB Doradus moving group ($\sim 20\text{--}50$ parsecs, ~ 50 Myr). Several nearby late-M and L dwarfs have been identified as likely members of these groups (Kirkpatrick et al. 2008), and provide crucial insight into the properties of such systems. In particular, these dwarfs serve as templates for the gravity-dependent features described in the following section.

At larger distances lie younger, active star-forming regions, such as Chamaeleon ($d \sim 160$ parsecs, $\tau \sim 1$ Myrs), the Scorpius-Centaurus Association (~ 150 parsecs, $\sim 5\text{--}15$ Myrs), ρ Ophiuchi (~ 140 parsecs, ~ 0.3 Myrs), and the low-density Taurus cloud (~ 140 parsecs, ~ 5 Myrs). The Orion Nebula Cluster, at a distance of ~ 400 parsecs, is the nearest dense star-forming cluster. Those associations and clusters are all sufficiently young that the hydrogen-burning limit falls within spectral type M; as a guide, the stellar/brown dwarf boundary is generally placed at spectral type M5/M6 in the Orion Nebula Cluster. Deep imaging and spectroscopic surveys have identified numerous later-type dwarfs in these regions, including ultracool M dwarfs that are brown dwarfs with inferred masses as low as $0.01 M_{\odot}$ the youngest regions, such as ρ Ophiuchi. Those substellar-mass dwarfs have broad characteristics that closely match those of older, stellar-mass dwarfs of comparable spectral type. As an example, chromospheric and coronal activity is prevalent among late-type M dwarfs (see [Sect. 3.7](#)). Observations of M5–M8 brown dwarfs in very young clusters (e.g., Orion: Preibisch and Feigelson 2005) find that they have X-ray properties that are indistinguishable from field dwarfs of similar type. This strongly suggests that temperature, rather than mass, is the controlling factor governing such properties.

Young brown dwarfs differ from field dwarfs in one important respect: Like higher-mass T Tauri stars, they show clear evidence for the presence of circumstellar disks. Photometric observations reveal infrared excess radiation consistent with the presence of cool dust, and Spitzer mid-infrared spectroscopy confirms the presence of silicates. Moreover, optical spectroscopy reveals emission lines that are strongly suggestive of gas accretion (see Pascucci et al. 2009, for a comparative review of stellar and brown dwarf disks). This raises the extremely interesting possibility of planet formation around substellar-mass brown dwarfs. Clearly, the disk mass is sufficiently low that there is little potential for forming gas giants. However, models suggest that there should be sufficient material to form terrestrial planets (Payne and Lodato 2007). This is clearly an extremely interesting area that is likely to receive substantial attention in the near future.

3.4 Lithium Absorption and Other Mass-Dependent Spectroscopic Phenomena

Brown dwarf spectral evolution is driven largely by changes in temperature. There are, however, a few features that are sensitive to other physical parameters. The presence of lithium absorption at $6,107 \text{ \AA}$ offers a straightforward means of segregating lower-mass brown dwarfs (with an important caveat, outlined later in this section). As outlined in [Sect. 2.1](#), models indicate that primordial lithium remains undepleted in solar-metallicity dwarfs with masses below $M_{\text{Li}} \sim 0.055 M_{\odot}$. This mass limit rises with decreasing metallicity, reaching ~ 0.06 at $[\text{M}/\text{H}] = -1$, since metal-poor dwarfs have higher densities and therefore higher degeneracy and lower central temperatures at a given mass.

Dwarfs with masses exceeding M_{Li} can sustain lithium fusion, and therefore deplete lithium at a rate that increases with increasing mass. As a guide, a $\sim 0.2 M_{\odot}$ star depletes $\sim 50\%$ of the primordial lithium in ~ 25 Myrs, while a $0.1 M_{\odot}$ star requires ~ 70 Myrs, a $0.08 M_{\odot}$

star ~ 130 Myrs and a $0.07 M_{\odot}$ brown dwarf ~ 250 Myrs. At these junctures, all of these dwarfs have temperatures $T_{\text{eff}} > 2,800$ K, corresponding to spectral types earlier than M5. Thus, a field dwarf with spectral type later than $\approx M5$ and substantial lithium absorption meets the lithium test (Rebolo et al. 1992), and can be identified as a brown dwarf with a likely mass $M < 0.06 M_{\odot}$. Similarly, this behavior can be applied to derive age estimates for relatively young star clusters by identifying the location of the lithium depletion boundary along the main sequence: the spectral type/absolute magnitude that marks the boundary between cluster members with and without detectable lithium absorption (Burke et al. 2004).

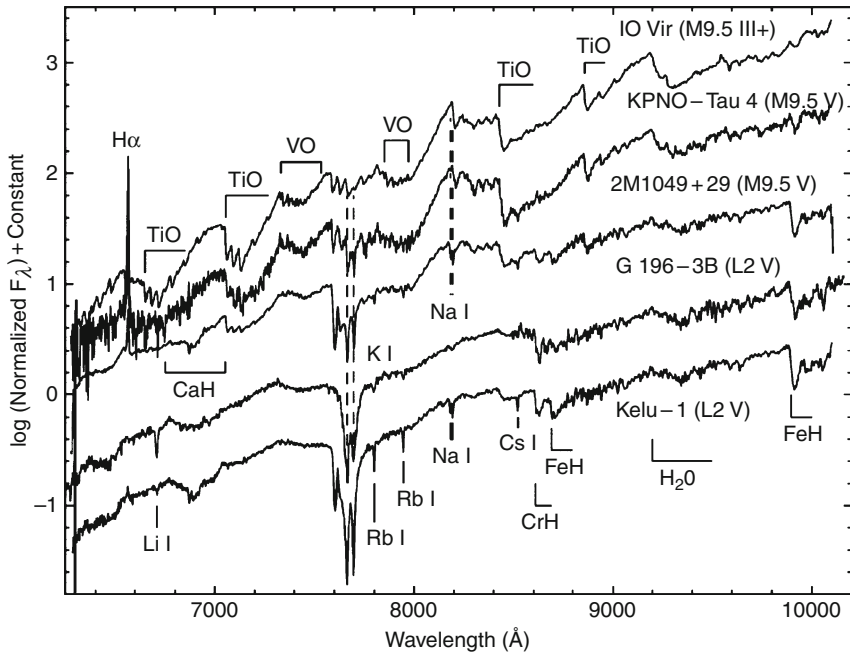
Brown dwarf masses can also be constrained to some extent through measurement of surface gravity. As a reference, mid-type M dwarf stars have surface gravities of $\log g \sim 5$, or $g \sim 10^5 \text{ g cm}^{-2} \text{ s}^{-1}$. Electron degeneracy constrains mature brown dwarfs to have radii within $\sim 10\%$ of Jupiter's radius (Sect. 3.4); thus, low-mass brown dwarfs can have surface gravities more than an order of magnitude lower than transition objects at the same temperatures. Moreover, young brown dwarfs with ages less than ~ 100 Myrs are still in the process of contraction, and have radii of $3\text{--}5 R_{\text{Jup}}$ (see Fig. 7-3). Thus, while a mature $0.075 M_{\odot}$ dwarf has a surface gravity $\log g > 5.2$, young $0.01 M_{\odot}$ ($\sim 11 M_{\text{Jup}}$) brown dwarfs can have gravities less $\log g \sim 3.5$.

Differences in surface gravity at this level have observational consequences. A comparison between spectra of dwarfs and giants provides insight into the likely differences at spectral class M. Ultracool companions of young field dwarfs and the low-mass members of nearby young ($\tau < 100$ Myr) associations (Sect. 3.3) have allowed coverage to be extended to spectral class L. At optical wavelengths, the most significant effects are weaker absorption features due to the alkaline metals (Na I, K I, Cs I, Rb I, Li I) and the metal hydride bands (CaH, FeH, CrH), but stronger absorption features due to metal oxides (VO, TiO) (Fig. 7-20). Similar effects are present at near-infrared wavelengths, and appear to be coupled with strengthening of the water bands in dwarfs later than $\sim L2$.

Metal hydrides are absent from giants because they require high pressures for their formation, and the lower pressures in lower gravity atmospheres inhibit formation. Similar circumstances likely apply in low-gravity brown dwarfs. The atomic alkaline lines are heavily pressure broadened in ultracool dwarfs, and they also become weaker at lower gravity. These effects are likely enhanced by the presence of stronger oxide absorption, which results in higher atmospheric opacities and effectively raises the physical level of the photosphere. Interestingly, lithium is one of the alkaline species affected and, as a result, can be difficult to detect in brown dwarfs younger than ~ 100 Myrs (Kirkpatrick et al. 2008).

Photometric observations of moderately young M and L dwarfs show that there are also effects in broadband colors. Low-gravity L dwarfs, for example, have redder (J–K) colors than the average for a given spectral type. It is not yet clear whether this reflects depressed flux in the J-band or enhanced emission at K. Note also that dust in L dwarfs can lead to redder (J–K) colors. Thus, a red (J–K) color appears to be a necessary, but not sufficient, condition for identifying low-gravity field dwarfs. Low-resolution spectrophotometry also suggests a significant change in the energy distribution in the H-band in L and T dwarfs, with the overall profile becoming more peaked at lower gravities. Again, the origin of that effect remains unclear at present.

Efforts are currently underway to quantify these effects using an appropriate set of narrowband spectral indices. As a first step, a broad classification scheme is being introduced that will designate “normal” ultracool dwarfs as class α , moderately low-gravity dwarfs as class β , and extremely low-gravity dwarfs as γ (Cruz et al. 2009). As yet, this proposal is still under consideration by the broader community.



■ Fig. 7-20

Gravity-dependent spectral variations in late-M and early-L dwarfs (From Reid and Hawley 2005). The figure plots far-red spectral data for three M9.5 objects, including a giant (IO Vir), a young brown dwarf (KPNO Tau 4), and a field dwarf (2M0149+29), and for two L2 dwarfs, G196-3B, the companion of an active M dwarf whose age is estimated as ~ 300 Myrs, and the binary system Kelu 1

3.5 Metallicity and Ultracool Subdwarfs

The overwhelming majority of brown dwarfs identified in the immediate Solar Neighborhood are drawn from the Galactic disk population. These dwarfs therefore have chemical compositions that are within a factor of 2 of solar abundance. However, a limited number of ultracool dwarfs have been identified that have characteristics suggestive of significantly lower metallicities. Moreover, these dwarfs tend to have higher space motions relative to the Sun, as might be expected for metal-poor members of the thick disk or Galactic halo.

The lower opacities present in metal-poor dwarfs produce higher internal pressures and lower central temperatures at a given mass. As a result, the hydrogen-burning mass limit increases with decreasing metallicity (see Sect. 3.1). At the same time, the lower atmospheric opacities lead to higher photospheric temperatures, higher luminosities, and bluer colors. Thus, while the subdwarf sequence appears to lie below the main sequence in the HR diagram, constant mass lines actually run up and to the left from solar abundance reference points.

Metal-poor K and M-type subdwarfs have been known for many years. The original K subdwarf is Kapteyn's star (Gl 191), catalogued by Jacobus Kapteyn in 1898. Their spectra are characterized by enhanced absorption by metal hydrides (MgH, CaH, FeH) and weaker oxide absorption (TiO, VO) at optical and near-infrared wavelengths, coupled with enhanced absorption due to collisionally induced H_2 that suppresses the flux in the H- and K-bands,

and correspondingly bluer near-infrared colors. These late-type subdwarfs are characterized as either intermediate subdwarfs, identified with the prefix “sd” (e.g., sdM6) and estimated to have average metallicities $[m/H] \sim -1.3$, or extreme subdwarfs, identified as esdM dwarfs and with average metallicities closer to $[m/H] \sim -2$.

In recent years, observations have been extended to significantly cooler dwarfs, including L-type subdwarfs. Those new discoveries were enabled by more sensitive proper motion surveys that utilized digital scans of photographic plate material, pushing the detection limit to fainter apparent magnitudes than was possible with hand-and-eye surveys (Lépine and Shara 2005). Those data can be matched against 2MASS to determine the optical-infrared colors, and segregate candidate subdwarfs. Similarly, cool subdwarfs can be selected using photometric criteria that combine optical data from SDSS and near-infrared data from 2MASS.

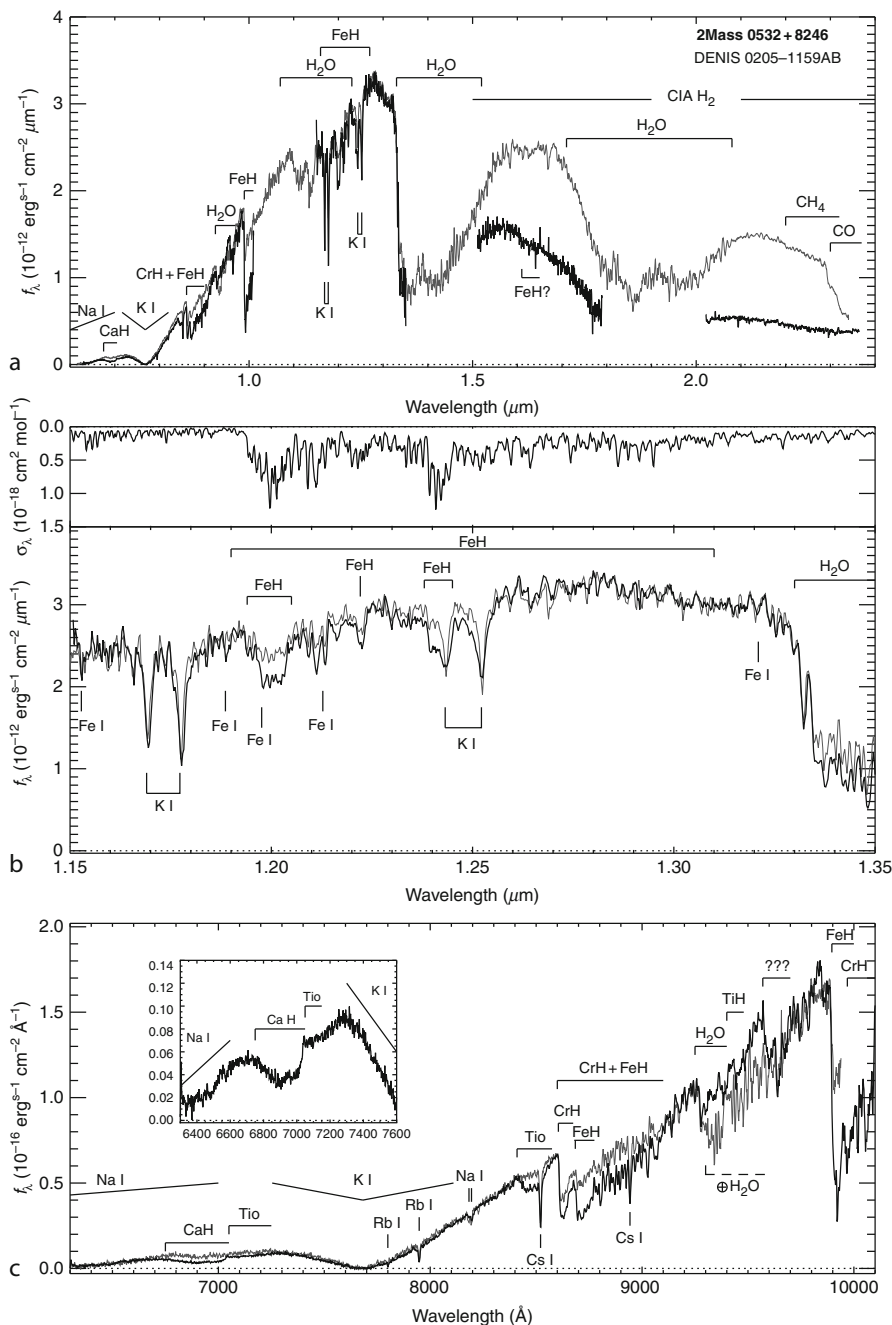
Spectroscopic observations of the resulting sample show that, like sdM dwarfs, sdL dwarfs have enhanced metal hydride absorption bands at optical wavelengths, and they also possess the broad atomic alkaline absorption features present in L dwarfs. Unlike L dwarf, however, sdLs continue to show weak TiO absorption at $\lambda < 1 \mu\text{m}$, indicating the continued presence of that compound in gaseous form. The near-infrared spectral energy distributions are radically different from L dwarfs with similar optical spectra (► Fig. 7-22). The spectra are almost featureless in the H and K bands, and the overall flux levels at those wavelengths are severely depressed due to strong collisionally induced H_2 absorption. There is also no evidence for any absorption due to the CO band at $2.3 \mu\text{m}$. Consequently, these dwarfs have photometric properties that are qualitatively similar to T dwarfs, with steep spectra and red colors shortward of $\sim 1.2 \mu\text{m}$ (T dwarfs are redder), but extremely blue near-infrared colors, $(J-H) \sim (H-K) \sim 0.0$. Indeed, several ultracool subdwarfs were uncovered in the course of T dwarf-focused surveys.

The spectroscopic characteristics of L-type subdwarfs are well illustrated in 2M0532+82, the latest-type subdwarf yet discovered at type sdL7 (► Fig. 7-21, Burgasser et al. 2003). With an estimated distance of 20 parsecs, this dwarf has a velocity of $\sim 320 \text{ km s}^{-1}$ relative to the Sun and is clearly a member of the Galactic halo. As such, it has a likely age of 11–13 Gyrs. The temperature remains uncertain, but is probably close to 2,000 K. The mass is estimated as less than $0.08 M_{\odot}$, placing 2M0532+82 very close to the hydrogen-burning limit for metallicities of $[m/H] \sim -1$. Thus, 2M0532+82 may well be the first known halo brown dwarf.

3.6 Brown Dwarf Kinematics and Brown Dwarf Rotation

The overwhelming majority of stars and brown dwarfs in the Galactic disk are born in dense clusters like the Orion Nebula (Lada and Lada 2003). The internal velocity dispersion within those clusters is extremely low ($\sim 0.2 \text{ km s}^{-1}$), but, over time, clusters disperse as the individual members acquire energy through gravitational interactions with massive objects in the Galactic disk. The increase in peculiar velocities leads to a statistical correlation between the velocity dispersion of a group of objects drawn from the Galactic field and the average age of that group. The velocity distribution of field dwarfs is usually modeled as either a single Gaussian or as a sum of Gaussians. In general terms, the overall velocity dispersion increases at $\tau^{1/3}$, where τ is age (Wielen 1977).

Computing space motions requires measurements of proper motion, distance, and radial velocity. In recent years, it has become possible to obtain those data for increasingly larger samples of late-M, L, and T dwarfs in the immediate Solar Neighborhood. Initial investigations suggested that ultracool dwarfs had velocity distributions that were similar to those measured



■ Fig. 7-21
Optical and near-infrared spectra of the coolest known subdwarf, 2M0532, spectral type sdL7, matched against data for the L5 dwarf Denis 0205-1159AB. The two spectra are relatively similar shortward of $1.2\ \mu\text{m}$, but diverge substantially at longer wavelengths, with 2M0532 essentially featureless in the H and K bands (From Burgasser et al. 2003)

for chromospherically active M dwarfs (dMe stars), significantly lower than the overall population of local M dwarfs. This would imply a statistical age in the range 1 to 4 Gyrs. However, those studies were limited to samples of less than 50 dwarfs. More recent analyses, based on samples that are an order of magnitude larger, show that field L dwarfs have kinematics that are indistinguishable from M dwarfs in the Solar Neighborhood (Schmidt et al. 2010). This result is consistent with the expectation that local field samples of ultracool dwarfs are dominated by the longer-lived low-mass stars and high-mass brown dwarfs.

The high-resolution spectra used to measure radial velocities also provide information on the rotational velocities of ultracool dwarfs. At stellar masses, observations of members of young clusters show that most stars have high initial rotation velocities ($>10 \text{ km s}^{-1}$). Those velocities decline fairly rapidly with time, as the stellar magnetic field interacts with circumstellar material, and angular momentum drag leads to magnetic braking. Observations of nearby early- and mid-type M dwarfs show that most have rotation velocities of a few km s^{-1} . Such is not the case for ultracool dwarfs, where the average rotation velocity is closer to 20 km s^{-1} and can exceed 80 km s^{-1} . These observations strongly suggest a significantly longer timescale for rotational damping, perhaps stemming for less efficient magnetic braking due to a higher degree of confinement of the magnetic field. This hypothesis is also relevant to observations of magnetically driven activity, as discussed in the [● Sect. 3.7](#).

3.7 Chromospheric and Coronal Activity in Brown Dwarfs

Magnetic fields on the surface of stars make their presence known in a variety of ways, including chromospheric line and continuum emission at optical and ultraviolet wavelengths, radio emission, X-ray radiation, spots, and flares. These phenomena are referred to collectively as stellar activity, and are the subject of [● Chap. 4](#) in this volume. Stellar activity is driven by an internal dynamo. In solar-type stars, the magnetic field is believed to be generated by an $\alpha\Omega$ shell dynamo in the boundary layer between the radiative core and the convective envelope (Babcock 1961; Parker 1955). This mechanism persists through K and early- M-type dwarfs, but stars become fully convective at a spectral type of $\approx M3/M4$, removing the physical basis for a shell dynamo. This prompted expectations of a corresponding change in the magnetic properties at that spectral type, but no such change is observed. However, there is growing evidence for a significant change in both the observed properties and the underlying atmospheric structure in ultracool dwarfs with spectral types M8 and later.

Activity manifests itself in two forms in stars and brown dwarfs: At an approximately constant level, generally referred to as *quiescence*, monitored through line emission and, to a lesser extent, photometric variability; and in bursts of activity, known as transients or flares, marked by strongly enhanced flux levels, particularly at high energies (blue, ultraviolet, and X-ray wavelengths).

Quiescent activity: M dwarf atmospheres are generally believed to have overall structures that are similar to those of solar-type stars. Observations of the Sun (see [● Chap. 6](#) in this volume) show that while atmospheric density drops monotonically above the photosphere, the temperature profile reverses $\sim 600 \text{ km}$ above the visible surface, and rises sharply thereafter, reaching $10,000 \text{ K}$ at a height of $2,000 \text{ km}$ and 10^6 K at $5,000 \text{ km}$. The inner regions above the reversal are the chromosphere, and the outer, very low-density, very high-temperature gas forms the corona, with a narrow transition region lying between. The optical emission line cores in

alkaline elements (Ca, K, Na) and the hydrogen Balmer series are generated by the $\sim 10^4$ K gas in the chromosphere.

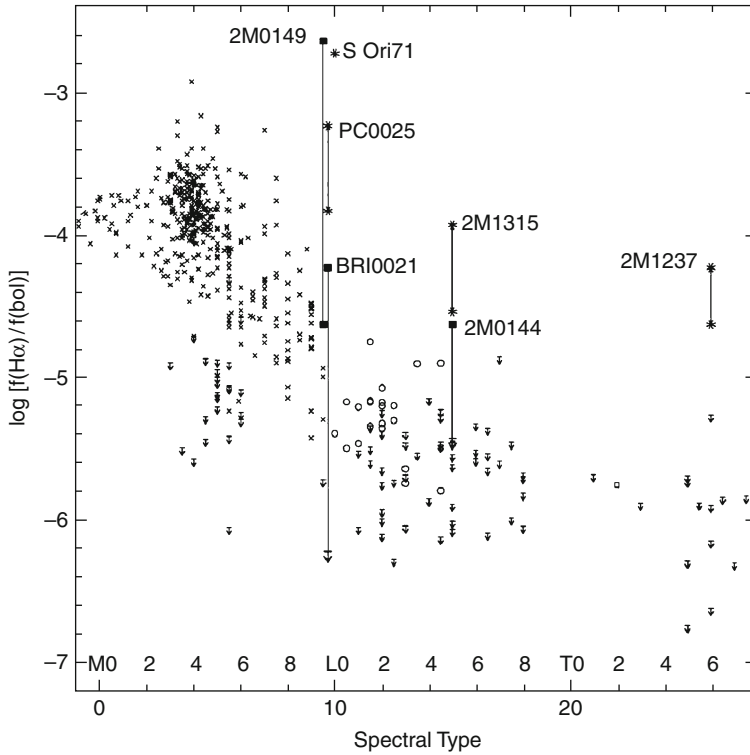
Activity in late-type dwarfs is generally tracked using Balmer line emission, specifically H α emission, rather than the Ca II H and K emission employed to survey late-G and K dwarfs. Recent years have seen extensive surveys of activity in late-type dwarfs. These surveys usually measure equivalent widths, a parameter that depends on the contrast between the emission line and the continuum. Since the continuum flux at 6,500 Å decreases with decreasing temperature, lower emission fluxes are detectable in cooler stars and these raw measurements are not suitable for a comparison of activity across a range of spectral types. Consequently, the overall activity level is generally quantified by measuring $F_{\alpha}/F_{\text{bol}}$, the ratio between the H α line flux and the measured bolometric flux (Reid et al. 1995). This uses the H α emission as a proxy to measure the proportion of a cool dwarf's energy budget that is devoted to chromospheric activity. This measurement is independent of spectral type.

► *Figure 7-22* shows the distribution of the H α activity index as a function of spectral type (Liebert et al. 2003). The mean level of activity is approximately constant at $\frac{L_{\alpha}}{L_{\text{bol}}} \sim 10^{-3.9}$ for spectral types M0 to \sim M6, although the dispersion clearly increases toward the latter types. Some of the scatter is likely due to intrinsic variability, since repeated observations indicate that most active stars exhibit night-to-night variability by up to a factor of 2. The overall frequency of active dMe dwarfs, defined as dwarfs with H α emission exceeding an equivalent width of 1 Å, rises from \sim 1% at M0 through 25–30% at M4 to over 60% at M6 and close to 100% at M8. Interestingly, observations of M dwarfs at significant distances above the Galactic Plane reveal lower dMe/dM ratios at these spectral types, with the active fraction peaking closer to 75% at M8 (West et al. 2008).

Observations of late-type dwarfs in young open clusters (α Persei, IC 2391, Pleiades, Praesepe, Hyades and M67) show that activity is correlated with age, in the sense that emission becomes apparent at earlier spectral types in younger clusters. Thus, late-type K dwarfs in the \sim 120 Myr Pleiades exhibit H α in emission, while emission appears at \sim M2 in the 600-Myr Hyades and closer to M4 in the \sim 4 Gyr-old M67. This indicates that later-type, lower-mass dwarfs have longer lifetimes as active stars than earlier-type, higher-mass stars, accounting for the higher fraction of activity at later spectral types among field dwarfs. The level of activity, however, does not show a corresponding correlation: M4 dwarfs in the Pleiades have similar H α equivalent widths to M4 dwarfs in the Hyades. As noted in ► *Sect. 3.2*, this similarity in level of activity at a given spectral type extends to very young brown dwarfs, indicating that activity is tied to atmospheric structure (i.e., temperature) rather than mass.

The correlation between the presence of activity and age likely accounts for the decreased proportion of active dwarfs at large distances above the Plane. Younger stars have smaller velocity dispersions, and therefore tend to concentrate near the Plane, leading to a growing preponderance of older, less active stars at increasing height. From the brown dwarf perspective, even the most massive substellar dwarfs have evolved to spectral types later than M5 by an age of 200 Myrs. Consequently, all M-type brown dwarfs are expected to exhibit H α emission.

It is clear from ► *Fig. 7-22* that there is a precipitous drop in activity as one moves beyond spectral type M7 into the ultracool régime. The mean flux ratio drops by an order of magnitude or more, and the activity fraction drops through \sim 60% at M9 to less than 30% at L2. Emission has been detected in a very limited number of later-type L and T dwarfs. There are a few very active brown dwarfs, notable examples being the peculiar M9 dwarf, PC0025+0447, and the T5 dwarf, 2MASS 1237+6526, both of which have maintained H α emission at a near-constant



■ Fig. 7-22

Activity in ultracool dwarfs, as measured by the presence of detectable $H\alpha$ emission. There is a substantial drop in the average level of activity as one moves from spectral type M7 to L0, a characteristic that is mirrored by the behavior at X-ray wavelengths. Individual objects of particular interest are identified, including the unusual M9 dwarf PC0025+0447 and the active T dwarf 2M1237+6526 (From Liebert et al. 2003)

level for several years, but they are the exception rather than the rule. The potential physical mechanisms driving this behavior are discussed further below.

Quiescent activity in cool dwarfs has also been detected at X-ray wavelengths; indeed, the initial detections with the Einstein satellite triggered the realization that magnetic fields persisted beyond the M3/M4 boundary, where stars become fully convective (Rosner 1980). This high-energy radiation is generated by the extremely hot gas in the stellar coronae and generally correlates well with the mean chromospheric activity. Subsequent X-ray satellites, notably XMM, ROSAT, and Chandra Observatory, have extended observations to later spectral types and lower luminosity dwarfs.

As with chromospheric indicators, the activity level across different spectral types can be compared using the flux ratio L_X/L_{bol} . The overall behavior is similar to that seen in $H\alpha$, with an average flux ratio $\frac{L_X}{L_{\text{bol}}} \sim 10^{-3}$ between spectral types M0 and \approx M6 (i.e., ten times the typical $H\alpha$ flux), coupled with a significant decline in activity among ultracool dwarfs. As a result, only the nearest ultracool M dwarfs are sufficiently bright to be detected at X-ray wavelengths,

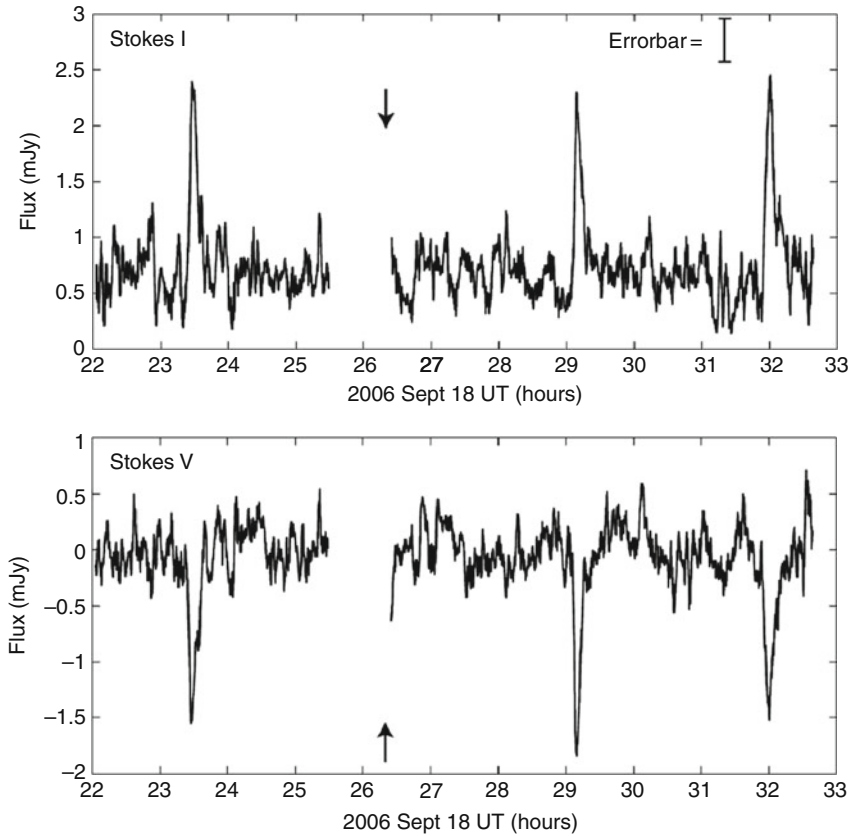
even with the Chandra Observatory. These include a number of young brown dwarfs (e.g., Stelzer et al. 2006), all with spectral types in the range M6 to M9; their X-ray fluxes are consistent with stellar-mass M dwarfs, suggesting that atmospheric properties such as temperature govern activity, rather than mass. A number of L dwarfs are known to flare at X-ray wavelengths, but to date quiescent emission has been detected from only one L dwarf, the binary system Kelu 1AB (L2/L3.5). No T dwarfs have been detected at X-ray wavelengths.

Moving to the opposite end of the electromagnetic spectrum, a number of the nearest K and M dwarfs have been known as radio emitters for more than 30 years (Gudel and Benz 1993). Observations are at GHz frequencies, notably 4.88 and 8.44 GHz, and a very small fraction of total luminosity is emitted at those wavelengths: Flux ratios of $\frac{L_R}{L_X} \sim 10^{-15.5}$ Hz are typical for K and M dwarfs. Clearly, observations are limited to the immediate Solar Neighborhood, but the Arecibo and VLA telescopes have been used recently to extend observations to a few of the nearest ultracool dwarfs. Close to a dozen have been detected, and their luminosities are 10^3 – 10^5 higher than those predicted by the canonical relation for mid-type M dwarfs (Berger 2006). The detections include the archetypical fast rotator BRI 0021 (M9.5), the brown dwarf LP 944-20 (M9.5), 2MASS 0523-14 (L2.5), and 2M0036+1820 (L3.5). In several cases, a significant component of the radiation is circularly polarized. Several dwarfs exhibit systematic variations with periodicities that are consistent with the likely rotation period (see further below).

Spots and pulsations: Sunspots form as opposite-polarity pairs where magnetic flux tubes break through the solar photosphere, producing large active regions with reduced temperature, and lower flux levels. Star spots are common in young, active G and K dwarfs, and can introduce periodic photometric modulations of a few percent as they rotate in and out of the line of sight (see ► Chap. 4 of this volume for further discussion). Young, active brown dwarfs are likely to show similar behavior, and, consequently, one might expect significant photometric modulations. Brown dwarfs have typical rotation velocities between 10 and 80 km s⁻¹. For a radius of 0.1 R_\odot , this corresponds to rotation periods between ~12 h and ~100 min. A number of photometric campaigns have been carried out with the aim of detecting variations on this timescale, but so far they have registered limited success. Variability has been detected, but repeated observations show that observations lose phase after only a few days. This suggests that the variability stems from relatively short-lived phenomena, and the suspicion is that the observations are detecting changes in cloud cover (brown dwarf “weather,” akin to Jupiter’s Great Red Spot) rather than magnetically related activity. These results are discussed further in ► Sect. 4.2.

One of most interesting developments in recent years is the detection of periodic variations in the radio emission from a small number of ultracool dwarfs. In some cases, such as the M8.5 dwarf 2M1835+32 (► Fig. 7-23), the dwarf exhibits strong, short-lived pulses of highly circularly polarized radiation. Other dwarfs, such as 2M0036+1820 (L35), exhibit variations that are broader in scale and are detected in both polarized and unpolarized light. In both cases, the variability is on a timescale of a few hours, matching the expected rotational periods of ultracool dwarfs. The behavior is generally consistent with beamed radiation, presumably associated with the magnetic poles, sweeping across our line of sight as the ultracool dwarf rotates on its axis. This is directly analogous to classical neutron star pulsars; indeed, these objects have been termed “ultracool pulsars.”

Flares: Stellar flares are an extreme example of magnetic activity that occurs on a subset of low-mass stars and brown dwarfs. During a classical flare, the blue and ultraviolet flux increases by orders of magnitude, and this variability is accompanied by substantial brightening at X-ray and/or radio wavelengths. Individual flares evolve on extremely rapid timescales, with the dwarf brightening by 5–7 magnitudes at blue and UV wavelengths in a matter of seconds, and decaying



■ Fig. 7-23

Periodic fluctuations in the radio emission from the nearby M8.5 dwarf, 2M 1835+32. The observations reveal strong periodic pulses ($P \sim 2.84$ h) with a very strong circularly polarized component in addition to continuous quiescent emission (From Hallinan et al. 2008). The upper plot shows the light-curve in intensity (Stokes parameter I) and the lower plot shows the temporal variation in circular polarization (Stokes parameter V). The magnetic field strength is estimated as 3 KG

to its quiescent level over a matter of minutes to hours. The continuum flux is generally consistent with a 9,000–10,000 K blackbody (Hawley and Fisher 1992), and is accompanied with flaring at X-ray wavelengths. These variations are accompanied by strengthened emission line fluxes in the Balmer lines, Ca II H and K and, in many cases, He I and He II.

Flares in M dwarfs are generally believed to originate through mechanisms that are similar to those operating on the Sun, although there are significant differences in behavior. In the Sun, the energy is released through reconnection between magnetic flux tubes, producing a current sheet. That sheet provides a means of rapidly accelerating electrons along magnetic field lines into the lower atmosphere where collisions with ions produce hard X-rays via nonthermal bremsstrahlung, enhanced activity through heating of the chromosphere, and white-light continuum due to hydrogen recombination. Evaporated chromospheric material leads to higher

coronal densities, and direct coronal heating by electrons, which, in turn, leads to soft X-ray production. The corona and the chromosphere cool as the electrons dissipate, and the star returns to its quiescent state. Flares in late-type dwarfs differ from solar flares in showing much stronger production of white-light continuum, and evidence for other continuum processes in addition to hydrogen recombination.

The best studied flare stars are mid-type dMe dwarfs, such as EV Lac and AD Leo (Hawley et al. 1995); these are stars, not brown dwarfs. However, optical flares have also been observed in a number of ultracool dwarfs, including VB 10 (M8), BRI 0021 (M9.5), 2MASS0149+2956 (M9.5), and 2MASS0144-0716 (L5). Those flares can attain extremely high luminosities, sometimes approaching the bolometric flux during quiescence (e.g., [Fig. 7-24](#), Liebert et al. 1999). Flares have also been observed at X-ray and radio wavelengths on ultracool dwarfs, including a handful of brown dwarfs (e.g., LP 944-20). As with quiescent emission, the overall characteristics of individual objects are tied to their temperature, not the mass: An M9 dwarf does not care whether it falls above or below the hydrogen-burning limit.

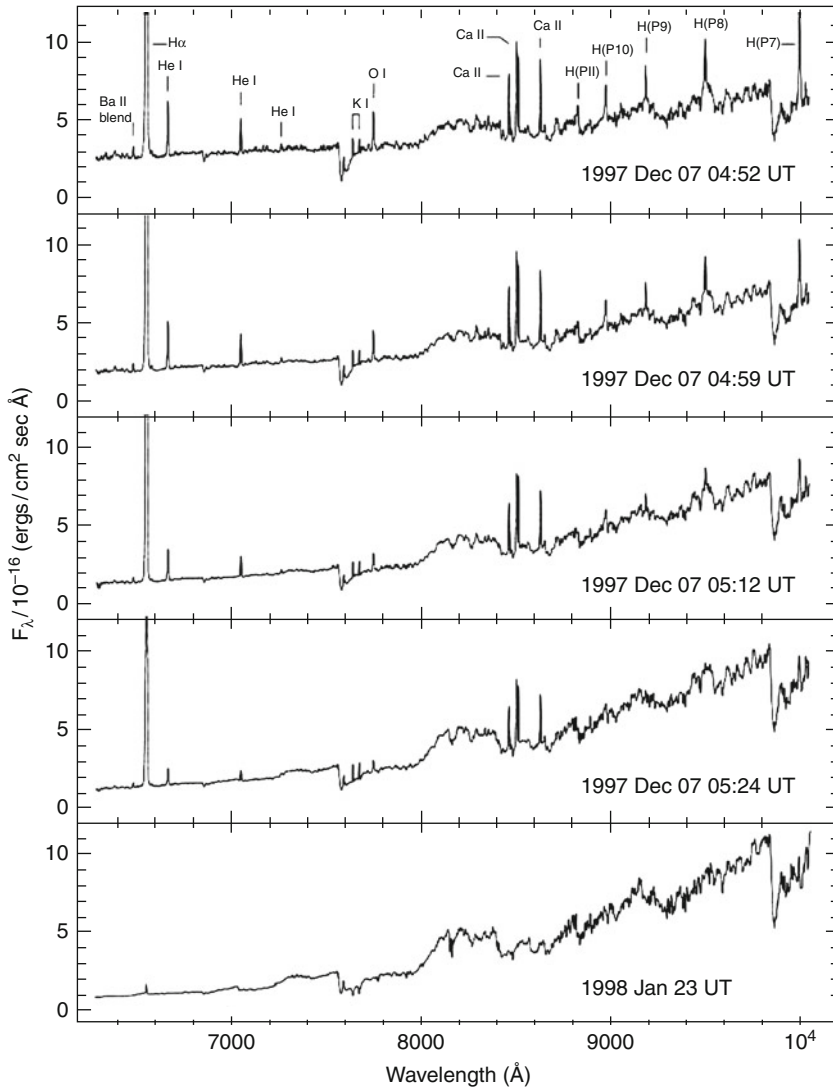
Combined, those observations are among the strongest evidence that magnetic activity remains present in those extremely cool dwarfs, which often lack quiescent H α emission and are not detected at X-ray wavelengths. Repeated observations of active ultracool M dwarfs indicate a duty cycle of $\sim 5\%$; that is, those dwarfs spend approximately 5% of the time in outburst. As yet, there are insufficient observations of multiple flares in L or T dwarfs to provide a reasonable estimate of the frequency of flaring, and whether that frequency varies with spectral type.

Magnetic dynamos: Thirty years ago, theoretical models were predicting a significant change in magnetic activity as dwarfs made the transition to full convection at spectral types M3/M4. That prediction was not supported by observations (Rosner 1980). It is now becoming apparent from X-ray, UV, emission line, and radio data that there is a significant change in the observational characteristics of low-mass dwarfs, but that change occurs at spectral type \approx M8.

Early-type M dwarfs, whether stellar or substellar, have the same internal structure as solar-type stars, with a convective envelope surrounding a radiative core. An initial poloidal field is subject to shear at the base of the convective envelope, generating a toroidal field (the Ω effect) which is itself modified by small-scale motions (the α effect) to general new poloidal fields. This $\alpha\Omega$ dynamo maintains quiescent levels of activity in those stars. Both effects are a consequence of stellar rotation, and, as a result, there is a correlation between magnetic field strength (and the level of activity) and rotation in early-type M dwarfs. This correlation naturally leads to a correlation between age and activity for these stars, since the rotational velocities of single stars decreases with age as angular momentum is lost through interactions between the global magnetic field and circumstellar gas. Stellar rotation can be maintained against angular momentum loss in close binary systems, where tidal locking matches the individual stellar rotation rates to the orbital period. This can lead to enhanced magnetic activity.

The $\alpha\Omega$ mechanism breaks down at spectral type \approx M4 when dwarfs become fully convective. The radiative core, and the convective/radiative boundary, disappear. The magnetic field in later-type dwarfs is believed to be generated by a turbulent dynamo (Browning 2008), driven by currents generated by shear produced from convective motions. In consequence, there is no clear correlation between stellar rotation and the level of activity in later-type M dwarfs. There have been suggestions that active dwarfs require a threshold rotational velocity, although observations place that threshold at a few kms^{-1} .

Activity and global properties: Surveys of field dMe stars show that they can exhibit significantly different bandstrengths than inactive M dwarfs (Reid and Hawley 2005, [Chapter 5](#)). Further evidence of global effects prompted by enhanced activity is presented by the eclipsing



■ Fig. 7-24

The progress of an optical flare on the M9.5 dwarf, 2M0149+2956 (From Liebert et al. 1999). The top four spectra show the rapid decline in strength of the emission line fluxes and the enhanced continuum emission at shorter wavelengths; the lowest spectrum shows the star in quiescence. It is estimated that the total luminosity increased by $\sim 10\%$ at the peak of the flare activity

binary, 2MASS0535-0546. This binary comprises a pair of brown dwarfs in the Orion Nebula Cluster. The two dwarfs have dynamically measured masses of 56 ± 4 and $36 \pm 3 M_{\text{Jup}}$, spectral types of $\sim M7$ and radii of 0.67 ± 0.03 and $0.49 \pm 0.02 R_{\odot}$ respectively, all consistent with membership of the ~ 1 Myr-old Orion Nebula Cluster. Contrary to expectations, the lower-mass dwarf has the higher surface temperature. Detailed observations show that the higher-mass dwarf is

rapidly rotating and is significantly more active, and the current hypothesis is that strong magnetic fields inhibit convection on this dwarf, leading to a cooler surface temperature. The system is too young to have achieved tidal locking, but over time one would expect a closer balance in activity levels as the rotation rates of the two brown dwarfs converge.

Activity in ultracool dwarfs: Magnetic fields are clearly still present in the very low-mass stars and brown dwarfs that populate the ultracool régime, even if they are manifested in a much lower level of quiescent activity and less frequent flares. Direct estimates are available for a few dwarfs, either through measuring Zeeman splitting of the FeH Wing-Ford bands (Reiners and Basri 2007) or radio emission, and they indicate field strengths of a few kiloGauss. Those values are comparable with measurements of classical field stars such as AD Leo (~ 4 kG) and EV Lac (~ 3.8 kG). What, then, leads to the significant change in observational properties between spectral types M6 and M9?

Theoretical analyses are just starting to probe this question. Initial investigations suggest that the effect is tied to the conditions within the upper atmosphere rather than reflecting substantial changes in the interior structure. In brief, the behavior is ascribed to the increasingly neutral state of the atmosphere. The combination of a low ionization fraction and high density leads to strong resistivity and efficient diffusion of magnetic fields, making it difficult for the atmosphere to sustain significant field strength at the surface (Mohanty et al. 2002).

Radio observations are further illuminating the nature of activity in these very cool dwarfs. In mid-type M dwarfs, emission at these wavelengths is largely ascribed to gyrosynchrotron radiation, generated by electrons spiraling around magnetic field lines within the stellar coronae. However, the highly directional and strongly polarized nature of the radio emission from ultracool dwarfs points to generation through a more coherent process. Current analyses (Hallinan et al. 2008) suggest that electron cyclotron maser (ECM) instability (Treumann 2006) is a strong candidate. The process requires a population inversion in the electron distribution in a magnetized plasma that has high field strength and/or low electron density. Within the Solar System, the ECM instability is known to be responsible for generating radio emission from planets with magnetic fields, and is also responsible for the Jovian and Saturnian auroral glows.

As noted above, the periodic variations observed at radio wavelengths are explained naturally if the emission is generated within a restricted region that moves in and out of the line of sight as the ultracool dwarf rotates on its axis. The presence of strong, narrow pulses indicates that the radiation is beamed. The magnetic poles are obvious locations for these active regions, and there is some evidence for dipolar structure in mid-type M dwarfs. The ECM instability produces emission that has 100% circular polarization. Polarized radiation is detected intermittently from dMe flare stars, such as AD Leo, and the ECM instability may make a minor contribution to radio emission from those stars, although gyrosynchrotron dominates.

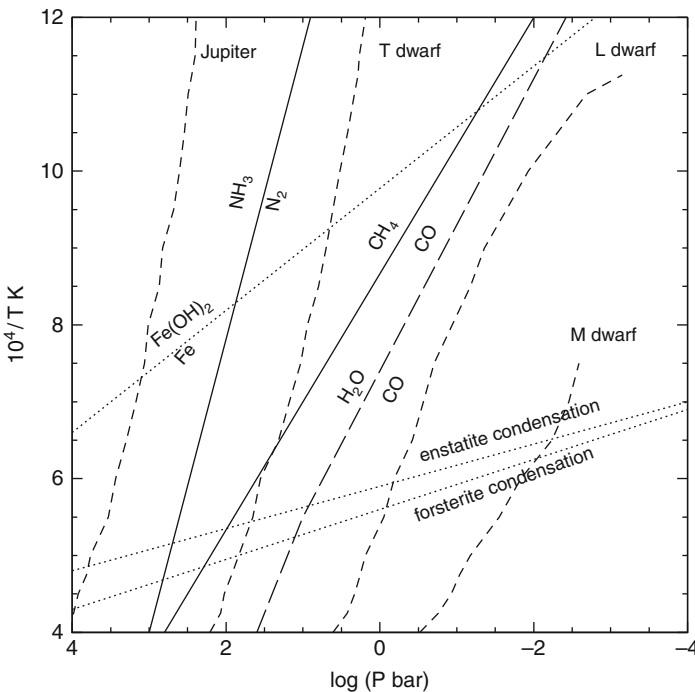
Circumstances appear to be reversed in cooler dwarfs, with the ECM instability as the dominant source. Most observations of ultracool dwarfs are consistent with partially polarized radiation, but close to 100% polarized light has been observed during outburst in the M8.5 dwarf TVLM 513-46546. The unpolarized radio emission present at other times may stem either from secondary scattering that processes the ECM radiation or from lower-level gyrosynchrotron coronal radiation. In either case, radio observations provide an intriguing link between brown dwarfs and the likely characteristics of giant exoplanets. Further observations, and more detailed modeling, are required to further elucidate how magnetic fields affect the properties of these very low-mass objects.

4 Brown Dwarf Atmospheres

4.1 Atmosphere Structure

Cool dwarfs have complex atmospheres with a significant complement of molecular species. The relative distribution of elements between various species, and their consequent spectral prominence, depends on gas-phase chemistry and the temperatures and pressure structure of the atmosphere. Thus, most carbon is sequestered in CO at high temperatures and low pressures, but resides in CH₄ (methane) at lower temperatures and higher pressures. Similarly, oxygen is predominantly in CO at high temperatures and low pressures, but transitions to H₂O (steam) at low temperatures and high pressures, while nitrogen moves from N₂ to NH₃ (ammonia). Oxygen is approximately twice as abundant as carbon, so water vapor remains an important opacity source even when all of the atmospheric carbon is locked up in CO.

Schematic temperature/pressure profiles for a late-type M dwarf, a mid-type L dwarf and a Gl229B-like T dwarf are shown in [Fig. 7-25](#). The photospheric temperatures are ~2,500 K, ~1,500 K, and ~1,000 K, respectively. A representation of the Jovian atmospheric profile is also shown for comparison. At a given temperature, the pressures increase for the lower mass



■ Fig. 7-25

Schematic pressure-temperature profiles for late-type dwarfs. The solid line shows where CO and CH₄ are expected to have equal abundance, and the long-dashed line shows where the same condition holds for CO and H₂O. The dashed lines mark the approximate regions where solid dust particles can condense within the atmosphere (Based on data from Lodders and Fegley 2002)

objects. This partly reflects the greater physical depth sampled within each object, but also the greater impact of degeneracy pressure within brown dwarfs. Superimposed on this diagram are lines that indicate the equal-abundance boundaries for CO/H₂O, CO/CH₄, and N₂/NH₃ in solar-metallicity dwarfs. Those boundaries map the physical basis for the transition from CO-dominated to methane-dominated atmospheres that account for the observed transition from spectral type L to T. Moreover, they illustrate the complexity within individual atmospheres. As an example, methane is associated with T dwarfs. However, the temperature/pressure relation for L dwarfs indicates that methane predominates in the upper atmosphere, and the 3.3 μm CH₄ feature has indeed been detected in L4/L5 dwarfs.

This highlights the non-gray nature of L and, particularly, T dwarf atmospheres. In mid-type L dwarfs at 3.3 μm , the photosphere (at optical depth $\tau = 1$) lies at a relatively shallow physical depth within the atmosphere. In contrast, at optical wavelengths, as discussed further in [◆ Sect. 4.2](#), dust condensation has eliminated the dominant opacity sources at optical wavelengths, leading to high transparency, and a substantially deeper photosphere. Structure within the atmosphere, such as banding or patchy cloud cover, may further complicate the issue, adding spatial variations. The net result is that atmospheric transparency is both wavelength and spatially dependent, leading to substantial variation in the physical location of the photosphere within the atmosphere; that is, the emergent spectrum samples a range of different physical depths at different wavelengths.

The strong wavelength dependence of photospheric depth in ultracool dwarfs warns against simplistic interpretation of global parameters, notably effective temperature. As defined *via* Stefans law, the effective temperature carries the implicit assumption that the measured luminosity originates from uniform depth in the atmosphere. Thus, while the derived value of T_{eff} can be taken as an approximate measurement of surface temperature for solar-type stars, this parameter is more akin to a weighted average of atmospheric properties in T dwarfs.

4.2 Dust and Clouds in Brown Dwarf Atmospheres

Dust formation plays an important role in the spectral evolution of brown dwarfs. TiO and VO absorption bands weaken in the latest-type M dwarfs. Tsuji originally suggested that these spectral changes might reflect the formation of solid condensates (dust) within cool dwarf atmospheres (Tsuji et al. 1996). [◆ Figure 7-25](#) includes the approximate condensation thresholds for several common condensates, including enstatite (MgSiO₃), forsterite (Mg₂SiO₄), and iron hydroxide (Fe(OH)₂). The formation of those condensates, together with corundum (Al₂O₃), perovskite (CaTiO₃), and solid vanadium oxide, removes metallic elements from the gaseous state and reduces the atmospheric opacity.

The detailed mechanism(s) for formation and growth of dust grains in cool dwarfs and the resulting cloud structure are still subjects of extensive theoretical investigations. However, there is general agreement in the overall process. Once formed, grains settle gravitationally to lower layers within the atmosphere and form cloud-like structures. The lower boundary to the cloud layer is set by the point within the atmosphere where the partial pressure of the condensates exceeds their saturation vapor pressure. Condensates that sink below this level reach higher temperatures, and evaporate, but are replaced by material borne upward on convective updrafts.

As the temperature decreases through late-type M dwarfs, dust formation becomes more efficient, and, in consequence, the atmospheric transparency increases, particularly at optical wavelengths. The net result is depletion and eventual elimination of absorption due to TiO and VO, and the transition from spectral type M to type L. In contrast, metal hydride lines

strengthen, as in low metallicity halo subdwarfs, since the available metals are preferentially found in these species in higher pressure atmospheres. At the same time, the increased atmospheric transparency pushes the photosphere deeper within the atmosphere in the optical, resulting in extremely broad alkali lines, such as the Na D lines and K I 7,600 Å doublet, due to the increased column through the atmosphere and substantial pressure broadening at $\tau = 1$.


Dust also plays a significant role in determining the spectral energy distribution at longer wavelengths in late-M and L dwarfs. Energy absorbed at shorter wavelengths heats the dust, leading to reradiation at longer wavelengths and backwarming in the upper atmosphere. Backwarming dissociates H₂O in late-type M and L dwarfs, producing shallower steam absorption bands at near-infrared wavelengths. The details of these processes depend crucially on where dust resides within the atmosphere, whether it forms clouds, and on the covering factors of those clouds. L dwarfs exhibit substantial dispersion in their near-infrared and mid-infrared colors at a given spectral type, and this probably reflects differences in the detailed dust distribution within their atmospheres.

There is general consensus that the cloud layer lies deeper within the atmosphere in T dwarfs. As a consequence, dust should play a lesser role in influencing their spectral energy distribution and photometric properties. However, the detailed behavior depends on the exact structure (depth, thickness, and covering factor) adopted for the cloud layer within the atmosphere model, and different models predict 10–20% bandstrength differences in spectral features at near- and mid-infrared wavelengths. These predictions have been constrained to some extent by observations from Spitzer, but a more thorough understanding must await higher resolution, higher signal-to-noise spectroscopy with other large telescopes, particularly the James Webb Space Telescope.

Brown dwarfs have substantial rotational velocities. Consequently, the coolest T dwarfs might be expected to develop spatial structure within their atmosphere, possibly including the latitudinal banding and large-scale cloud features that are evident in Solar System gas giants, notably Jupiter. Such features might give rise to systematic photometric variations as they are carried in and out of the line of sight by the brown dwarfs rapid rotation. As with starspots, signatures of chromospheric activity, searching for such features requires coordinated photometric monitoring campaigns, but in this case observations need to be made at near- and mid-infrared wavelengths where the effects of clouds are most prominent. A few investigations have been conducted to date, and short-lived photometric variability at the ~0.5% level has been detected in a handful of late-L and early-T dwarfs (e.g., Artigau et al. 2009; Morales-Calderon et al. 2006), while spectroscopic monitoring suggests that some features (including FeH, VO, K I, and the water bands) may vary in strength by 2–10% (Bailer-Jones 2008).

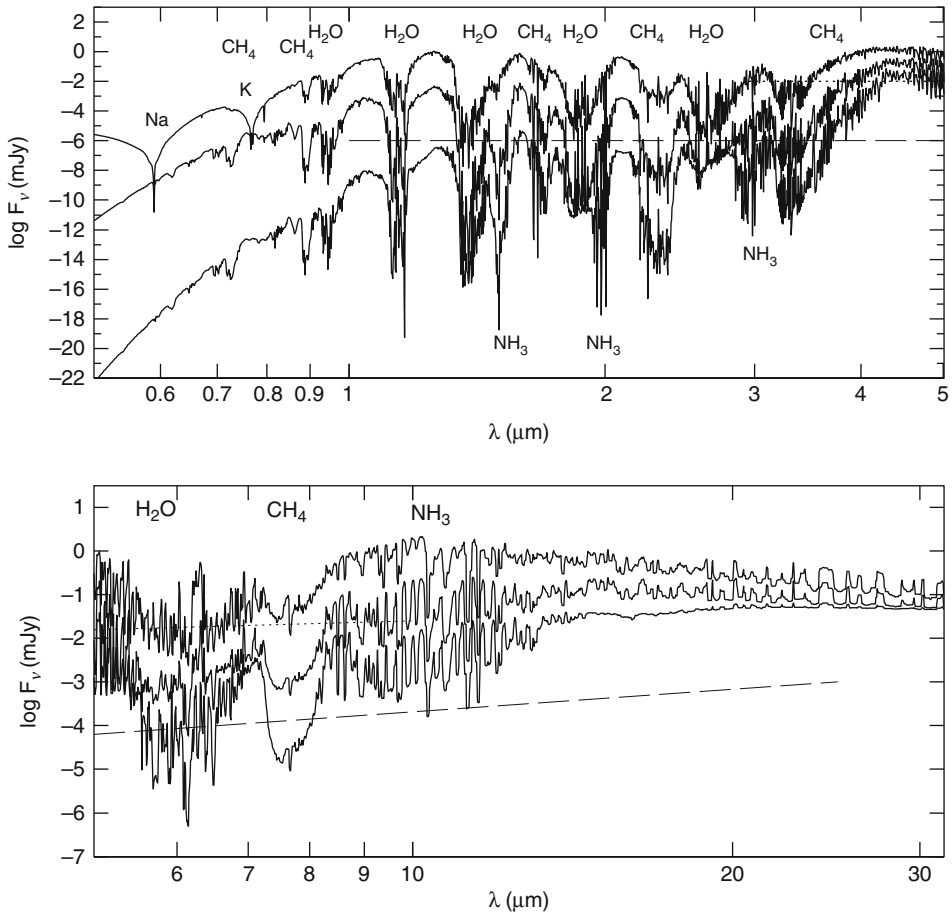
The observed modulations have periodicities of a few hours, consistent with rotation, and may well be a consequence of variations in cloud cover within the atmosphere (brown dwarf weather). Further observations, spanning longer time series for individual dwarfs and a wider sample of targets, are required to probe the spectral characteristics and lifetimes of these features. In principle, these observations offer the potential to probe the temperature structure of brown dwarf atmospheres, and perhaps provide insight into the intrinsic properties of gas giant exoplanet atmospheres.

4.3 Y Dwarfs

The coolest T dwarfs, and hence the coolest brown dwarfs, currently known have effective temperatures that are estimated as below 700 K (Burgasser et al. 2006a). However,  Fig. 7-5 shows

that low-mass brown dwarfs are predicted to achieve much lower temperatures at ages of a few Gyrs, and modeling of the likely brown dwarf luminosity functions (► [Fig. 7-9](#)) suggests that “room-temperature” 300 K dwarfs should be common in the Solar Neighborhood. Such objects have extremely low luminosities, less than $5 \times 10^{-7} L_{\odot}$, and none have yet been detected. However, theoretical models have been used to predict their likely photometric and spectroscopic properties.

► [Figure 7-26](#) shows the predicted spectral energy distributions for brown dwarfs at temperatures of ~600, 425, and 225 K. As the temperature drops, the broad atomic features due to alkaline metals weaken and disappear, and features due to new molecular species appear. In



■ Fig. 7-26

Model spectra of a $5 M_{\text{Jup}}$ brown dwarf at ages of 0.1, 0.3 and 3 Gyrs; the corresponding effective temperatures are ~600, 425, and 225 K. In each case, the flux levels are scaled to a distance of 10 parsecs, and the principal spectral features are identified. The upper (*dotted*) lines show the detection limits of Spitzer’s Infrared Imaging Array camera and the Infrared Spectrograph; the lower (*dashed*) lines mark the expected sensitivity limits for the James Webb Space Telescope (Based on models from Burrows et al. [2003](#))

particular, absorption bands due to ammonia (NH_3) become evident at temperatures below 500 K, particularly at 1.5, 1.95, and 2.95 μm . These bands are likely to become the spectral signature for spectral class Y. At the same time, methane absorption strengthens at 2.2, 3.3, and 7.8 μm , and the H_2O steam bands continue to be prominent features at near- and mid-infrared wavelengths, notably at 1.5 μm .

The overall spectral energy distributions steepen sharply at optical wavelengths with decreasing temperature. At 600 K, the flux decrement between the H-band (1.6 μm) and R-band (0.65 μm) is $\sim 10^4$; by 225 K, the offset has risen to $\sim 10^{10}$. The numerous molecular absorption features tend to concentrate the emitted energy within a few relatively transparent windows, notably the JHK bands at near-infrared wavelengths, and at 4 μm and 10 μm in the mid-infrared. The peak energy in those windows is relatively flat at 600 K, but the shift in the overall energy distribution at cooler temperatures pushes the peak flux distribution into the M-band.

The spectra plotted in [Fig. 7-26](#) are all scaled to the same distance (10 parsecs) and emphasize the low luminosities of the coolest brown dwarfs. As a guide, the peak flux in the H-band decreases by a factor of $\sim 10^6$ as the effective temperature drops from 600 K to 226 K; this corresponds to a decrease from $M_H \sim 17$ to $M_H \sim 32$. The flux decrement is less in the M-band, but still corresponds to close to a factor of 10. Thus, these models suggest that high sensitivity wide-field surveys at 5 μm are likely to be a powerful means of finding old, cool brown dwarfs.

Ground-based surveys in the mid-infrared are rendered extremely difficult by the poor transparency and highly variable behavior of the terrestrial atmosphere at those wavelengths. Space-based surveys avoid those issues. The Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) carried out such a survey, imaging the full sky in several passbands between 3 and 24 μm . Although a relatively small-aperture telescope, WISE is capable of detecting room-temperature brown dwarfs within a few parsecs of the Sun. Examples of Y dwarfs may also turn up in focused searches for cool companions to known nearby stars, a search technique discussed further in [Sect. 5](#). As of mid-2011, both approaches have been successful in identifying dwarfs with effective temperatures cooler than 400 K and spectra consistent with those expected for Y dwarfs. Those objects and further examples will be studied in detail by the James Webb Space Telescope, scheduled for launch in 2018, which is equipped with cameras and spectrographs that cover near- and mid-infrared wavelengths.

5 Brown Dwarfs as Companions

Binarity is a key characteristic of stellar and substellar systems, as discussed in detail in [Chap. 16](#) of this volume. Properties such as the overall frequency and the distributions of mass ratio and semimajor axis can give crucial insight into the star formation process, and into the subsequent evolution of the circumstellar environment. In the case of brown dwarf companions, higher-mass, main-sequence primaries can supply crucial, and otherwise unobtainable, constraints on the space motions, composition, age, and mass of these cool systems. Brown dwarf binaries provide vital information on the origins of very low-mass systems.

Finding low-mass companions: A wide variety of complementary techniques have been employed in searching for low-luminosity, brown dwarf binaries. Direct imaging is most effective for systems with large separations, particularly when coupled with either astrometric or

multicolor photometric information. The initial forays into this area were based on photographic data, notably the Lowell Observatory proper motion survey by H. Giclas and his coworkers, and on Luytens extensive analyses of plates taken by the 48-inch Schmidt at Palomar. Those observations are not well suited to detecting extremely cool brown dwarfs, although the nearest early-type L dwarfs are visible on the red-sensitive Schmidt plates. More recent investigations have focused on CCD data from the Sloan Digital Sky Survey and near-infrared imaging from the 2-Micron All-Sky Survey.

Direct imaging from wide-field surveys is limited to detecting binary systems with angular separations exceeding 2–3 arc seconds, more in the case of binaries with bright main-sequence components. Ultracool L and T dwarfs are readily observable to distances of 20–30 parsecs, so this resolution corresponds to linear separations of tens of Astronomical Units (AU). Direct imaging at smaller separations (~ 0.1 arc seconds) is possible either using the Hubble Space Telescope at optical and near-infrared wavelengths, or with large ground-based telescopes employing adaptive optics (AO) techniques at near-infrared wavelengths. Those observations push the linear resolution to a few AU for the nearer systems.

Completeness limits for direct imaging surveys can be specified in a relatively straightforward fashion as magnitude limits, usually with some radial variation close to the primary. Binary systems at smaller separations are detectable through a number of indirect techniques that also offer lower completeness. Photometric catalogues can be used to identify systems whose components have very different temperatures by searching for objects with anomalous colors. White dwarf/brown dwarf systems, for example, are blue at optical/ultraviolet wavelengths, but extremely red at far-red and near-infrared wavelengths. This technique is useful only for systems with similar luminosities, but was responsible for the discovery of the first L dwarf, GD 165B (as discussed further below). Time series photometry is required to identify eclipsing binaries. Such systems are rare, but offer a treasure trove of information.

Radial velocity monitoring is a powerful method for searching for close binary companions. Recent years have seen extensive use of this technique to search for very low-mass (specifically, planetary mass) companions of the nearest stars. Such systems, with components of unequal luminosity, are identified as single-lined binaries (SB1), where the spectral line of the primary show periodic velocity variations with amplitude $K_1 \propto \frac{M_2 \sin i}{M_1^2 p^{1/3}}$. The period is measured directly, and one can infer the mass of the primary component from its luminosity and spectral type. However, the mass of the companion is determined uniquely only if the orbital inclination is known. That parameter is known for eclipsing systems, where the orbit must be aligned close to $i = 90^\circ$.

Inclinations can be derived for SB1 systems if an astrometric orbit can be measured for the primary star. This was the method employed by Friedrich Bessel to infer the existence of Sirius B in the 1840s, but the technique has a chequered history in searches for low-mass companions. In particular, planetary-mass companions have come and gone around nearby late-type dwarfs, such as Barnard's star, as astrometric analyses have first identified periodic wobbles, and then resolved such variations as systematic instrumental effects. Reliable astrometric data for low-mass systems is scarce and, to be fully effective, must be obtained at the microarcsecond level to detect brown dwarf companions at modest separations. Such measurements are coming within range of potential future ground- and space-based instrumentation.

If components have similar luminosities, then the system usually forms a double-lined spectroscopic binary (SB2) with spectral features due to both evident in the combined spectrum. (Exceptions are systems with white dwarf components with featureless spectra.) In double-lined systems, the relative velocity amplitudes of the two components scale inversely with the

component mass ratio (i.e., the lower-mass component has higher velocity amplitude). As with single-lined systems, transforming those mass ratios to absolute masses requires determination of the orbital inclination.

At present, most high-precision spectroscopic surveys are still limited to optical wavelengths, reducing the potential for identifying close binaries with low-mass primaries. Stability is the crucial factor in obtaining the meter/second resolution that is required for detecting planetary-mass companions. This is ensured at optical wavelengths by imprinting reference lines on the spectrum via an iodine cell. Considerable effort is being invested in extending that technique to near-infrared wavelengths, closer to the peak flux of red dwarf primaries. A further complication for ultracool dwarfs is their rapid rotation, which lowers the effective resolution and limits the ability to detect very low-mass companions.

Finally, one other technique has the potential to identify brown dwarfs in binary systems: gravitational microlensing. This occurs when a foreground object, the lens, aligns closely with a background source. As the lens moves across the line of sight, the light from the background source is focused gravitationally and amplified. The source brightens equally at all wavelengths, peaks in brightness when lens and source are at closest separation, and subsequently declines as the angular separation increases. If the lens has a binary companion at an appropriate orientation, then that companion will also amplify the background source, producing a secondary spike in the light curve. Microlensing events are almost always one-off events, with the lens often never observed separately from the background source. Consequently, it is usually impossible to obtain direct information on the foreground object. The duration of the event depends on the mass of the foreground lens and the transverse motion. The latter parameter can be estimated if one assumes membership of a particular stellar population (typically, the Galactic disk). Thus, as data are accrued for more and more events, statistical analysis allows one to infer the broad properties of those objects, including the likely frequency of brown dwarf and planetary-mass companions. At the present time, this technique is starting to provide useful constraints on planetary system architecture, but has had almost no impact on brown dwarf astronomy.

The following sections describe how these different techniques have been combined to provide information on brown dwarfs in binary systems.

5.1 Brown Dwarf Companions to Main-Sequence Stars

Main-sequence binaries with low-mass companions have played a crucial role in the discovery and characterization of ultracool dwarfs in general, and brown dwarfs in particular. Low-luminosity dwarfs are detectable, even at infrared wavelengths, within a limited volume. The surface density of isolated dwarfs is low: There are $\sim 2,000$ M dwarfs within 20 parsecs of the Sun, or one per ~ 20 square degrees; there are approximately 100 L dwarfs, or one per ~ 400 square degrees. Those statistics have emerged only within the last few years, but the conceptual problem has been clear for decades. One means of circumventing the issue is to search for companions to brighter stars that are already known to be in the immediate vicinity of the Sun. This is the “under the lamppost” search technique: A certain proportion of stars are binaries; therefore, rather than searching the full local volume, concentrate on areas where the presence of a known nearby star heightens the probability of finding a much lower luminosity star or brown dwarf.

A brief history: The earliest major discovery was the prototype ultracool dwarf, van Biesbroeck’s star, also known as Gl 752B and VB 10. Discovered in 1944 by the Belgian

astronomer George-Achille van Biesbroeck as a wide, common proper motion companion of an M3 dwarf within 5 parsecs of the Sun, this M8 dwarf was the first ultracool dwarf (albeit a designation that was not invented for another 50 years), and held the record as the lowest luminosity known star for more than 40 years. In the 1960s, van Biesbroeck added 20 additional cool companions to nearby stars, including VB 8, an M7 dwarf, and VB 7, an extremely cool white dwarf.

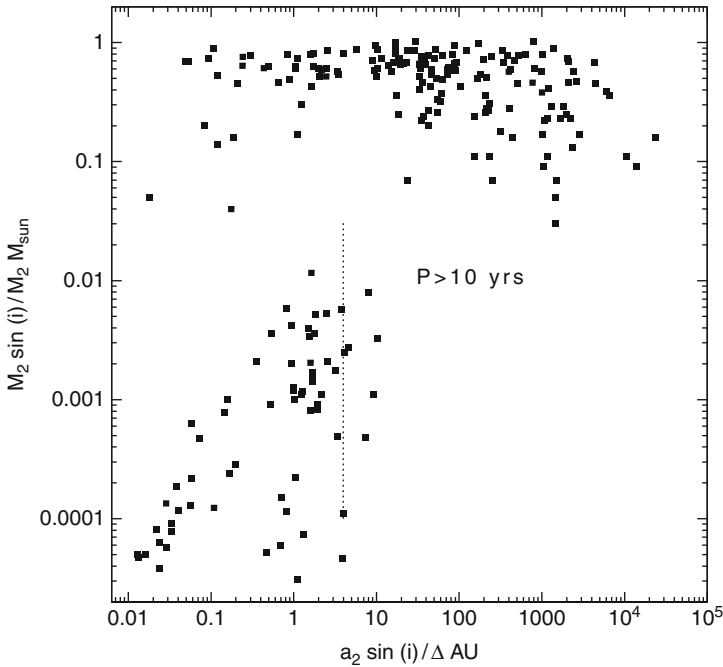
Three highly interesting objects emerged from companion searches in the late 1980s and early 1990s. First, a photometric survey of white dwarfs turned up two sources with significant excess radiation at infrared wavelengths: G 39-28, and GD 165. In the former case, further observations showed that the excess was produced by cool dust, perhaps remnants of a planetary system, but imaging of the latter star revealed a cool, extremely red resolved companion at a separation of ~ 50 AU (Becklin and Zuckerman 1988). The companions optical spectrum is clearly unusual for an M dwarf, but initial interpretations were confused by the potential for pollution by mass lost from the higher-mass star during its red giant and asymptotic giant evolutionary phases. GD 165B is now recognized as an L4 dwarf, with a mass very close to the hydrogen-burning limit.

Almost contemporaneously, radial velocity surveys led to the identification of a very low-mass companion to HD 114762, a mildly metal-poor F-type star at a distance of ~ 40 parsecs (Latham et al. 1989). This discovery continues to highlight the ambiguities of such detections. The formal mass estimate is $\sim 0.011 M_{\odot}$, well below the hydrogen-burning limit, but, as a single-lined, unresolved system with unknown inclination, it remains unclear whether the companion is a low-mass star, a brown dwarf, or a high-mass exoplanet. The latest analyses suggest that the observed rotational velocity of the F-type primary is most consistent with an inclination of $15\text{--}20^{\circ}$, implying a companion mass of $\sim 0.045 M_{\odot}$, but the data are consistent with a mass anywhere between $\sim 0.025 M_{\odot}$ and $\sim 0.08 M_{\odot}$.

As discussed in [Sect. 1](#), the dénouement came in the early 1990s, when a coronagraphic survey of stars in the immediate Solar Neighborhood revealed an extremely low-luminosity companion to a nondescript M0.5 dwarf, Gl 229, lying ~ 5 parsecs distant. The companion, Gl 229B, is 13 magnitudes fainter than the primary in the $\sim 7,500$ Å I-band, placing it firmly within the brown dwarf régime (Nakajima et al. 1995). Crucially, spectroscopy revealed strong methane absorption at near-infrared wavelengths, more reminiscent of gas giant planets than cool stars. Gl 229B is, of course, the archetypical T dwarf, with a temperature of ~ 950 K and a mass between 0.03 and $0.06 M_{\odot}$. Thus, the first ultracool dwarf (VB 10), the first L dwarf (GD 165B) and the first T dwarf (Gl229B) were all discovered because they are companions to previously known nearby stars.

By a curious coincidence, the discoveries of Gl 229B and of the first extrasolar planet, 51 Pegb, were announced at the same scientific meeting in late 1995. The latter discovery energized the few radial velocity exoplanet surveys then underway, and stimulated a number of new surveys. Observations have tended to concentrate on solar-type stars for obvious reasons and, as a result, reliable statistics are now in place on the companion mass distribution for those stars. A key feature that emerged early on in those surveys was the scarcity of brown dwarfs within 10 AU of the primary star, characterized as the “brown dwarf desert.” With the passage of time, and the acquisition of complementary radial velocity and imaging data for statistically well-defined samples, it becomes clear that this characterization oversimplifies the situation.

Statistics: [Figure 7-27](#) shows the companion mass as a function of separation for solar-type stars (spectral types late-F to early-K) within 25 parsecs of the Sun. This is essentially complete for brown dwarfs and low-mass stars, since almost all these stars are included in exoplanet radial



■ Fig. 7-27

The mass/separation diagram for known stellar, brown dwarf, and planetary companions of the volume-complete sample of 479 solar-type stars ($4 < M_v < 7$) within 25 parsecs of the Sun. Results from radial velocity surveys have not been corrected for assumed inclinations. Masses for wide companions are estimated based on their photometric properties. Notice that the sample includes no more than a handful of companions with $0.01 < \frac{M}{M_\odot} < 0.5$ and $\Delta < 10\text{ AU}$ – the brown dwarf desert extends through the M dwarf régime (Adapted from Reid and Hawley 2005)

velocity surveys. No correction for orbital inclination has been applied to the mass estimates derived from radial velocity measurements. Masses for wide companions are inferred from their photometric properties. The brown dwarf desert is strikingly obvious at separations less than 10 AU, with a significant number of very low-mass ($M < 0.013 M_\odot$) exoplanetary companions, but only two brown dwarfs. However, it is also evident that the scarcity of low-mass companions extends to $\sim 0.5 M_\odot$; that is, the *brown dwarf desert* is accompanied by an *M-dwarf desert* and even a *K-dwarf desert* (Burgasser et al. 2007). Turning this around, it is clear that there is a strong bias toward near-equal mass companions at separations less than $\sim 50\text{ AU}$ in solar-type binary systems. Lower-mass companions appear at wider separations, and an analysis of the companion mass function shows much greater consistency with the field mass function at those separations.

It is likely that ► Fig. 7-27 is providing key insight in binary formation. The preference for equal-mass systems at small separation suggests that competitive accretion quickly disrupts potential low-mass ratio ($q = \frac{M_2}{M_1}$) binaries during the early stages of star formation; in brief, big stars do not let little stars form next to them. The low-mass exoplanets present at small separations ($\Delta < 10\text{ AU}$) undoubtedly form within the circumstellar disk. Similar patterns are emerging for lower-mass primaries, as discussed further in ► Sect. 5.2.

Indeed, [Fig. 7-27](#) may be offering a means of segregating brown dwarfs and jupiter-like exoplanets. Both objects form by accretion, but within different environments, and therefore have different internal structures and compositions. Those intrinsic differences result in very subtle differences in the resultant spectrophotometric properties. Consequently, brown dwarf and planets are essentially indistinguishable through direct observations. [Figure 7-27](#), however, suggests that they may be distinguishable indirectly, through their occupying near-separate spatial domains, with a maximum separation for gas giant planets set by the characteristics of the protoplanetary disk, and a minimum separation for brown dwarfs set by the gravitational field of the primary star. One may conjecture that the boundary between these two domains lies at radii between 50 and 100 AU in solar mass stars, scaling to smaller separations in lower-mass stars with smaller protostellar disks and, perhaps, lower-mass planets. In higher-mass stars, with more massive disks, the boundary likely lies at larger radii, and formation of exoplanets with masses well into the brown dwarf régime may be possible. As an example, the HR 8799 planetary system that was recently resolved through round-based imaging has an $\sim A5$ primary and super-Jovian planetary companions with masses that may be in the range $15\text{--}20 M_{\text{Jup}}$.

Statistical analyses indicate that, in contrast to the inner desert, brown dwarfs (and low-mass stars) are not uncommon ($\sim 5\%$) as companions to main-sequence stars (Metchev and Hillenbrand 2009). This is useful, since the higher-mass primaries offer a means of better characterizing the properties of ultracool dwarfs, whose complex spectra continue to present strong challenges to theoretical models. It is reasonable to assume that binary components are coeval, and that they are formed from essentially identical materials. Thus, metallicities and ages derived for the main-sequence primary are also likely to be valid for the companion.

Several star/brown dwarf systems can be highlighted as of particular interest:

- Gl 569Bab*: In the late 1980s, near-infrared observations revealed a late-type M dwarf companion to this nearby mid-type M dwarf. The primary is chromospherically active, with an age estimated as ~ 300 Myrs. A decade later, AO-assisted imaging resolved the companion into close M8.5 and M9 components with an orbital period of ~ 2.5 years. Dynamical mass estimates for the two components are $0.055\text{--}0.08$ and $0.05\text{--}0.07 M_{\odot}$.
- ϵ IndiBab*: At a distance of only 3.6 parsecs, these are the second closest brown dwarfs currently known. Originally identified from photographic material as a common proper motion companion lying ~ 7 arc minutes from the K5 primary, high-resolution imaging has resolved the system into a pair of T dwarfs, spectral types T1 and T6. The system is estimated to have an age in the range $0.8\text{--}2$ Gyrs, leading to mass estimates of 0.05 and $0.02 M_{\odot}$ for the two brown dwarfs. With an orbital period of ~ 15 years, dynamical mass measurements will be possible by 2020.
- HD 114762B*: Already discussed here with reference to the $M_2 \sin i \sim 0.011 M_{\odot}$ radial velocity companion, subsequent AO-assisted imaging resulted in the identification of a faint, red companion lying ~ 3.3 arc seconds from the primary. This system is metal-poor: The primary has a well-determined metallicity of $[\text{Fe}/\text{H}] = -0.7$, and has an estimated age of 11 Gyrs. It is likely that the companion lies slightly above the hydrogen-burning limit, but this ultracool dwarf provides crucial quantitative insight into the spectral evolution of ultracool dwarfs at subsolar metallicities.

TW Hya 5B: The TW Hydrae group is an association of very young ($\tau < 8$ Myr) stars lying at distances of 50–60 parsecs from the Sun. TW Hya 5A, or CD-33 7795, is an early-type M dwarf binary whose components likely have masses of $0.6\text{--}0.7 M_{\odot}$. TW Hya 5B is an M8 dwarf, lying ~ 100 AU from the primary with a likely mass of $\sim 0.02 M_{\odot}$. This young system serves as benchmark for the early stages of evolution of low-mass brown dwarfs.

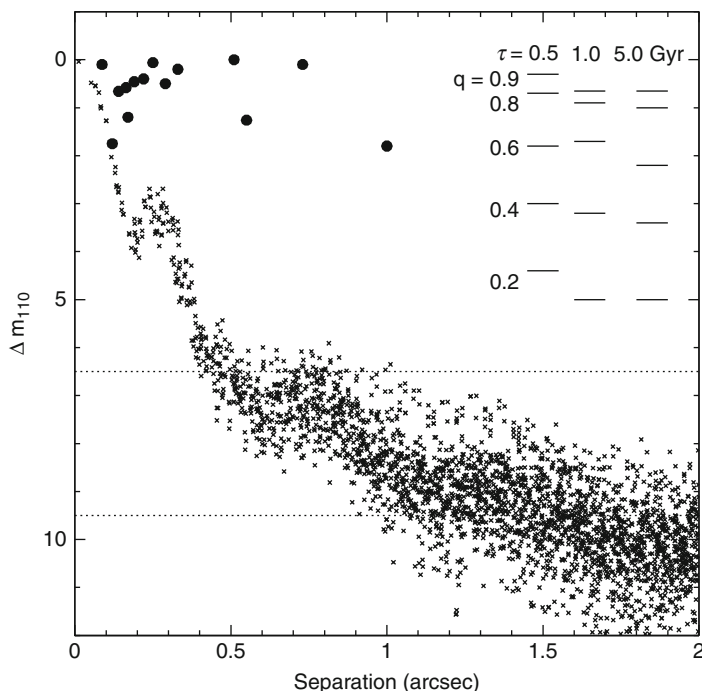
5.2 Brown Dwarf Binaries

The mass limit for the detection of secondary components in binary systems depends on the properties of the primary star. The higher the mass of the primary, the lower the reflex motion due to a companion of fixed mass at a given separation, and the higher precision required in radial velocity surveys. Similarly, the higher the primary mass, the higher the luminosity and the higher the flux limit for detection of a companion at small angular separations. Given these constraints, considerable effort has been devoted in recent years to pushing surveys to the lowest possible masses by searching for faint companions to nearby ultracool dwarfs.

Recent years have seen the compilation of extensive catalogues of ultracool M, L, and T dwarfs from the SDSS, DENIS, and 2 MASS surveys. Companion searches have focused on the nearest such dwarfs, particularly those within 20 parsecs of the Sun. Radial velocity surveys remain relatively restricted in scope, complicated both by the need for observations at near-infrared wavelengths for the cooler sources, and by the tendency for L and T dwarfs to have substantial rotational velocities, reducing the sensitivity to orbital motions. A few single-lined binaries have been discovered, all late-type M dwarfs. The only double-lined system identified to date is the young eclipsing binary, 2MASS0535-0546 (see [Sect. 3.7](#)), a member of the Orion Nebula Cluster.

High-resolution imaging surveys of local ultracool dwarfs are much more extensive in scope. Near-infrared HST data are available for over 70 L dwarfs and 25 T dwarfs within 20 parsecs of the Sun (Reid et al. 2008a; Burgasser et al. 2006b), and those data are supplemented by ground-based near-infrared AO imaging of comparable resolution. As discussed in [Sect. 2.4](#), ultracool dwarfs in the field are likely to have masses close to the hydrogen-burning limit and typical ages of several Gyrs; consequently, any very low-mass brown dwarf companions in those systems are likely to have faded to extremely low luminosities and become correspondingly difficult to detect. In an attempt to deal with those issues, extensive surveys using HST and ground-based AO have also been undertaken of low-mass members of nearby open clusters, including the Hyades, Pleiades, and α Persei, and of the nearest star formation regions, including the Scorpius-Centaurus, ρ Ophiuchi, and Chamaeleon associations. Very low-mass brown dwarfs are brighter and correspondingly easier to detect in these younger systems, but, at distances of 40–150 parsecs, current observations cannot probe linear separations smaller than 5–15 AU.

Combining the results from these various surveys, $\sim 15\%$ of nearby L and T dwarfs are resolved as binaries with separation exceeding ~ 3 AU. The available radial velocity data suggest that no more than 5–10% are spectroscopic binaries, implying an overall binary fraction $< 25\%$ (Burgasser et al. 2007). This is significantly lower than the $\sim 70\%$ determined for solar-type stars, and lower even than the $\sim 35\%$ cited for M dwarfs. Moreover, there is a clear tendency toward near-equal luminosity, implying near-equal mass, among the ultracool systems. The latter result is illustrated in [Fig. 7-28](#), which shows data for L dwarf binaries within 20 parsecs of the Sun,

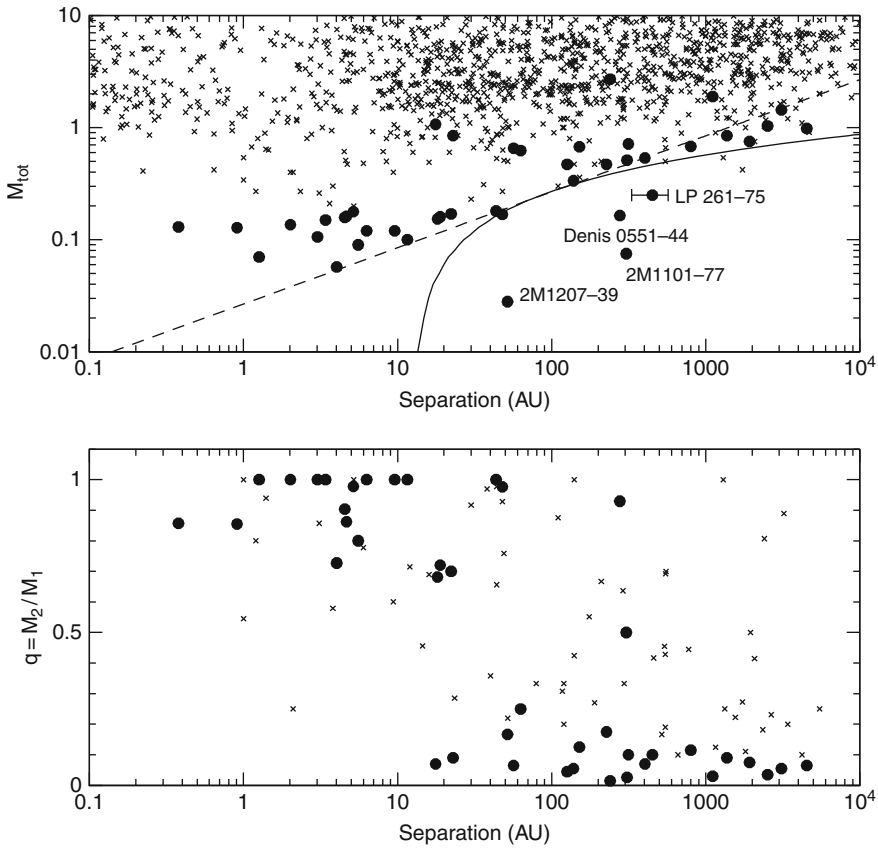


■ Fig. 7-28

The HST NICMOS J-band point-spread function, in relative magnitudes, matched against the peak brightness of known companions of ultracool dwarfs plotted at the appropriate angular separation. The lower dotted line marks the sky brightness level, and the upper line the formal detection limit at wide separation (> 1 arc second); the point-spread function of the primary sets brighter constraints at smaller separations. It is clear that there is a substantial area of discovery space that is accessible, but unoccupied (Adapted from Reid et al. 2008a)


plotting the relative magnitudes of the companions as a function of separation from the primary. The figure includes the near-infrared (J-band) point-spread function for HST imaging, and, as a reference, the flux ratios predicted for brown dwarf binaries with a range of ages (τ) and mass ratios (q). The latter indicate that these observations are capable of detecting binaries with $q \sim 0.5$ over a wide range of separations for brown dwarfs with age 5 Gyrs; that is, even if the local brown dwarf population is dominated by $0.07 M_{\odot}$ dwarfs, companions with masses as low as $0.035 M_{\odot}$ should be readily detectable. Their absence suggests that such high-mass ratio systems are rare.

The tendency toward equal-mass systems may have important implications for the apparent spectral evolution of brown dwarf binaries in key regions of the HR diagram. As discussed in [Sect. 3.2](#), the M_J and M_K /spectral-type relations exhibit significant flattening in the vicinity of the L/T transition. Observations show that this is at least partly due to the presence of binaries comprising of a late-L and an early-type T component. However, statistical analyses show that the underlying binary fraction is consistent with that estimated for early-L and late-T dwarfs (Burgasser 2007).



■ Fig. 7-29

Total system masses and mass ratios as a function of separation: crosses mark stellar binaries; solid points are systems with ultracool companions; the solid and dashed lines in the mass/separation diagram plot separation limits that scale exponentially with M_{tot} and with M_{tot}^2 , respectively (Adapted from Burgasser et al. 2007)

The overwhelming majority of ultracool binaries known in the immediate Solar Neighborhood have angular separations that correspond to linear separations, a , less than ~ 15 AU, with the distribution peaking in the range 3–10 AU. Wider systems are known, but primarily in star-forming associations, such as Sco-Cen and Taurus, and among the youngest members of the field population, including DENIS 0551-4434AB, GQ Lupi, and 2M1207-39AB (discussed further below).  [Figure 7-29](#) puts these results for brown dwarf binaries in the broader stellar context. There is a clear scaling between the maximum separation of binary components, a_{max} , and the total system mass, M_t . At high masses, this appears to scale as M_{tot}^2 , but at $\sim 0.15 M_{\odot}$ the relation steepens and can be represented by an exponential dependence. Computing the binding energy for these systems, $E_{\text{bind}} \propto M_1 M_2 a^{-1}$, the widest brown dwarf binaries in the Solar Neighborhood have binding energies almost an order of magnitude higher than wide stellar binaries. There are some indications of a distinct transition in the low-energy limit for systems with total mass exceeding $0.3 M_{\odot}$, matching the transition between the exponential and M_{tot}^2 a_{max} relation. The origins of this behavior remain uncertain. Nonetheless, it seems likely that



dynamical disruption is responsible for modifying the initial separation distribution of very low-mass binaries to give the sharply truncated distribution evident among local field dwarfs.

One of the young brown dwarf binary systems, 2M1207-39ABm deserves particular mention. This system is a member of the TW Hydrae association, whose age is estimated as ~ 8 Myrs. The primary is spectral type M8 and has a mass in the range $0.025\text{--}0.045 M_{\odot}$. The secondary is 8 magnitudes fainter, has an inferred mass of $4\text{--}6 M_{\text{Jup}}$, corresponding to a mass ratio $q \sim 0.2$, and is occasionally described as a planetary companion. The separation between the two components is ~ 40 AU, placing the secondary well beyond the likely extent of the primary's protoplanetary disk, and beyond the extent of any plausible planet formation. Thus, describing 2M1207B (and its ilk) in planetary terms is more than a little misleading. Nonetheless, 2M1207-39B remains of considerable interest as the lowest-mass brown dwarf yet discovered, and, consequently, an important indicator that fragmentation or accretion processes within molecular clouds are capable of producing star-like objects with masses of $\sim 0.005 M_{\odot}$ (Lodato et al. 2005).

5.3 Ultracool Binaries in Context

Pulling together the observations summarized in the previous sections, there are three key properties that are associated with ultracool binary systems:

- The overall binary fraction is lower than 25%. This maintains the trend of decreasing multiplicity with decreasing primary mass that is evident at stellar masses.
- There is a strong tendency toward near-equal mass systems among low-mass binary systems, both above and below substellar boundary. This is consistent with the mass ratio distribution measured for close binaries at higher masses.
- Very low-mass binaries in the field are generally confined to separations less than 10–15 AU, although wider systems are occasionally found in very young groups and associations. This extends an overall trend evident among stellar binaries toward decreasing maximum separation with lower total system mass.

These properties can be accounted for in a self-consistent manner within the broader scheme of stellar binarity, where brown dwarf binaries represent the low-mass extreme of a continuum that stretches through solar-type stars to massive OB binary systems. Qualitatively, one might expect the number of binary systems to depend on the range of possible semi-major axes.  [Figure 7-29](#) shows that the maximum separation, a_{max} , for stable systems varies with the total system mass, probably reflecting the gravitational binding energy. Consequently, one might infer that higher-mass primaries have a higher probability of retaining companions, and therefore the overall binary fraction should decrease with the mass of the primary. At the same time, data for solar-type stars show that equal-mass systems tend to dominate at small separations, leading to the K/M/brown dwarf desert ( [Fig. 7-27](#)). Indeed, one might consider dividing binaries into two classes: near-equal mass close systems; and wide systems, where the companion mass function is much closer to the IMF. In solar-type stars, “close” and “wide” binaries are represented in almost equal proportions. But as a_{max} decreases with decreasing mass, one can envisage the “close” systems coming to dominate the sample, leading to an overwhelming number of equal-mass systems at very low masses. Of course, it is important to emphasize that this is a highly qualitative scenario, and, as yet, there is no full understanding of the physical mechanisms that underlie binary formation and govern the dynamical evolution of systems.

6 Summary and Conclusions

Over the past 20 years, brown dwarfs have evolved from a much sought after theoretical concept to an integral part of observational astrophysics. Several hundred unequivocal examples of isolated or binary substellar mass objects are now known, ranging from $\sim 3,000$ K mid-type M dwarfs through 2,100–1,300 K L dwarfs to T dwarfs with effective temperatures as low as 600 K. Extensive multiwavelength observations of this growing compendium of ultracool dwarfs have defined the spectrophotometric characteristics of these low-temperature objects, and allowed theoretical astrophysicists to refine their models of the atmospheres and interior structure. Detailed observations at X-ray, optical, and radio wavelengths show that these systems retain magnetic fields with strengths that are comparable to mid-type M dwarfs, but the activity manifests itself in a manner that is more consistent with Solar System gas giant planets than low-mass stars. Brown dwarfs are constituents of binary systems, both as (usually wide) companions of more massive primaries and as near-equal mass, small-separation ultracool systems. In numbers, surveys of the local field and nearby young clusters and associations give results that are in broad consensus, showing that the mass function is flat or declining in the substellar regime. Brown dwarfs are not a viable source of substantive quantities of dark matter.

As the number of ultracool dwarfs catalogued by high-sensitivity, wide-field surveys has increased, brown dwarf astronomy has transitioned from discovery to characterization. Larger samples, drawn from a variety of survey methods, provide better estimates of the range of properties of individual systems, circumventing biases that can be introduced by particular search techniques. The resultant datasets allow a thorough exploration of correlations between different parameters, such as near-infrared colors and the presence of dust or variations in metallicity. At the same time, these more extensive datasets open the possibility of finding unusual ultracool dwarfs, whose properties might provide insight on rare or short-lived evolutionary phenomena.

The next step in brown dwarf exploration will be extending coverage to substantially cooler systems, the much-anticipated Y dwarfs. Several deep imaging surveys are currently underway that offer prospects of achieving this goal, and the first plausible candidates are starting to emerge. However, following historical precedent, one might anticipate that the first unambiguous example will be identified through an entirely different avenue, perhaps even as a wide companion to a known nearby star.

7 Further Reading

Reid and Hawley (2005) remains the only textbook that focuses exclusively on low-mass celestial objects, from M dwarf stars through brown dwarfs to extrasolar planets. That text provides a thorough summary of the consensus knowledge of brown dwarf characteristics at that time, and places those systems in the broader context of the astrophysics of low-mass stars and the Galactic structure parameters of the various stellar populations in the Milky Way. More detailed information on specific areas, and more recent advances in our understanding, are covered by a number of useful review articles. The review article by Burrows et al. (2001) provides a thorough evaluation of the evolution and interior structure of low-mass dwarfs. Kirkpatrick's (2005) annual reviews article gives an excellent overview of the spectroscopic properties of cool dwarfs, while Burgasser et al. (2007) discuss binarism in ultracool dwarfs. Kirkpatrick and Burgasser have reviewed the spectroscopic properties of L and T dwarfs, respectively, in

Gray and Corbally's (2009) textbook on spectral classification. Chabrier (2003) and Bastian et al. (2010) have reviewed determinations of the initial mass function in a wide variety of environments, and provide extensive discussion of the functional form through the substellar limit.

Acknowledgments

Thanks to Rachel Osten, Sarah Schmidt, Daniel Apai and, particularly Suzanne Hawley for reading and commenting on an earlier draft of this manuscript.

References

- Allen, P. R., Koerner, D. W., Reid, I. N., & Trilling, D. E. 2005, *ApJ*, 625, 385
- Artigau, E., Bouchard, S., Doyon, R., & Lafreniere, D. 2009, *ApJL*, 701, 1534
- babcock, H. W. 1961, *ApJ*, 133, 572
- Bailer-Jones, C. A. L. 2009, *MNRAS*, 348, 1145
- Bastian, N., Covey, K. R., & Meyer, M. 2010, *ARAA*, 48, 339
- Bate, M. R., Bonnell, I. A., & Bromm, V. 2002, *MNRAS*, 332, L65
- Becklin, E. E., & Zuckerman, B. 1988, *Nature*, 336, 656
- Berger, E. 2006, *ApJ*, 648, 629
- Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 1997, *MNRAS*, 285, 201
- Bouvier, J., et al. 2008, *A&A*, 481, 661
- Browning, M. K. 2008, *ApJ*, 676, 1262
- Burgasser, A. J. 2007, *ApJ*, 659, 655
- Burgasser, A. J., Kirkpatrick, J. D., Brown, M. E., Reid, I. N., Liebert, J., Matthews, K. Y., Burrows, A., Gizis, J. E., Dahn, C. C., Monet, D. G., Cutri, R. M., Skrutskie, M. F. 2001, *ApJ*, 564, 421
- Burgasser, A. J., Kirkpatrick, J. D., Burrows, A. S., Liebert, J., Reid, I. N., Gizis, J. E., McGovern, M. R., Prato, L., & McLean, I. S. 2003, *ApJ*, 592, 1186
- Burgasser, A. J., Burrows, A., & Kirkpatrick, J. D. 2006a, *ApJ*, 639, 1095
- Burgasser, A. J., Kirkpatrick, J. D., Cruz, K. L., Reid, I. N., Leggett, S. K., Liebert, J., Burrows, A., & Brown, M. E., 2006b, *ApJS*, 166, 585
- Burgasser, A. J., Reid, I. N., Siegler, N., Close, L. M., Allen, P., Lowrance, P. J., & Gizis, J. E. 2007, in *Planets and Protostars V*, eds. B. Reipurth, D. Jewitt, & K. Keil (Tucson: Univ. Arizona Press), 427
- Burke, C. J., Pinsonneault, M. H. & Sills, A. 2004, *ApJ*, 604, 272
- Burrows, A., & Liebert, J. 1993, *Rev Mod Phys*, 65, 301
- Burrows, A., Hubbard, W. B., Saumon, D., & Lunine, J. I. 1993, *ApJ*, 406, 158
- Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, *RvMP*, 73, 719
- Burrows, A., Marley, M., Hubbard, W. B., Lunine, J. I., Guillot, T., & Saumon, D. 1997, *ApJ*, 491, 856
- Burrows, A., Sudarsky, D., & Lunine, J. I. 2003, *ApJ*, 596, 587
- Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, *ApJL*, 542, L119
- Chabrier, G. 2003, *PASP*, 115, 763
- Cr    , M., Chereul, E., Bienaym  , O., & Pichon, C. 1998, *A&A*, 329, 920
- Cruz, K. L., Kirkpatrick, J. D., & Burgasser, A. J. 2009, *AJ*, 137, 3345
- Cushing, M. C., Roellig, T. L., Marley, M. S., Saumon, D., Leggett, S. K., Kirkpatrick, J. D., Wilson, J. C., Sloan, G. C., Mainzer, A. K., Van Cleve, J. E., & Houck, J. R. 2006, *ApJ* 648, 614
- Epchtein, N., et al. 1994, in *Science with Astronomical Near-Infrared Surveys*, Vol. 217 (Dordrecht/Boston: Kluwer), 3
- Golimowski, D. A., Leggett, S. K., Marley, M., Fan, X., Geballe, T. R., Knapp, G. R., Vrba, F. J., Henden, A. A., Luginbuhl, C. B., Guetter, H. H., Munn, J. A., Canzian, B., Zheng, W., Tsvetanov, Z. I., Chiu, K., Glazebrook, K., Hoversten, E. A., Schneider, D. P., & Brinkmann, J. 2004, *AJ*, 127, 3517
- Gray, R. O., & Corbally, C. J. 2009, *Stellar Spectral Classification*, eds. O. G. Richard & J. C. Christoper (Princeton: Princeton University Press)
- Grossman, A. S., Hays, D., & Graboske, H. C. 1974, *A&A*, 30, 95
- G    l, M., & Benz, A. O. 1993, *ApJ*, 405, L63
- Hallinan, G., Antonova, A., Doyle, J. G., Bourke, S., Lane, C., & Golden, A. 2008, *ApJ*, 684, 644
- Hawley, S. L. et al. 1995, *ApJ*, 453, 464

- Hawley, S. L. & Fisher, G. H. 1992, *ApJS*, 78, 565
- Kirkpatrick, J. D., Reid, I. N., Liebert, J., Cutri, R., Nelson, B., Beichman, C. A., Dahn, C. C., Monet, D. G., Skrutskie, M. F., & Gizis, J. E. 1999, *ApJ*, 519, 802
- Kirkpatrick, J. D. 2005, *ARAA*, 43, 195
- Kirkpatrick, J. D., et al. 2008, *ApJ*, 689, 1295
- Kroupa, P. 2002, *Science*, 295, 82
- Kumar, S. S. 1962, *AJ*, 67, 579
- Kumar, S. S. 1963, *ApJ*, 137, 1121
- Lada, C. J. & Lada, E. A. 2003, *ARAA*, 41, 57
- Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M., & Burki, G. 1989, *Nature*, 339, 38
- Lawrence, A., et al. 2007, *MNRAS*, 379, 1599
- Leggett, S. K., et al. 2010, *ApJ*, 710, 1627
- Lépine, S., & Shara, M. M. 2005, *AJ*, 129, 1483
- Liebert, J., Kirkpatrick, J. D., Reid, I. N. & Fisher, M. D. 1999, *ApJ*, 519, 345
- Liebert, J., Kirkpatrick, J. D., Cruz, K. L., Reid, I. N., Burgasser, A. J., Tinney, C. G. & Gizis, J. E. 2003, *AJ*, 125, 343
- Lodato, G., Delgado-Donate, E., & Clarke, C. J. 2005, *MNRAS*, 364, L91
- Lodders, K., & Fegley, B. 2002, *Icarus*, 155, 393
- Lucas, P. W. et al. 2010, *MNRAS*, 408, 56
- Metchev, S. A., & Hillenbrand, L. A. 2009, *ApJS*, 181, 62
- Metchev, S. A., Kirkpatrick, J. D., Berriman, G. B., &Looper, D. 2008, *ApJ*, 676, 1281
- Miller, G. E., & Scalo, J. M. 1979, *ApJS*, 41, 513
- Mohanty, S., Basri, G., Shu, F., Allard, F., & Chabrier, G. 2002, *ApJ*, 571, 469
- Morales-Calderon, M., et al. 2006, *ApJ*, 635, 1454
- Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., Golimowski, D. A., Matthews, K., & Durrance, S. T. 1995, *Nature*, 378, 463
- Oort, J. H., 1932, *BAN*, 6, 249
- Parker, E. N. 1955, *ApJ*, 122, 293
- Pascucci, I., Apai, D., Luhman, L., Henning, T., Bouwman, J., Meyer, M. R., Lahius, F., & Natta, A. 2009, *ApJ*, 696, 143
- Payne, M. J., & Lodato, G. 2007, *MNRAS*, 381, 1597
- Preibisch, T., & Feigelson, E. D. 2005, *ApJS*, 160, 390
- Rebolo, R., Martin, E. L., & Magazzu, A. 1992, *ApJL*, 389, L83
- Reid, I. N., & Gizis, J. E. 1997, *AJ*, 113, 2246
- Reid, I. N., Gizis, J. E., & Hawley, S. L. 2002, *AJ*, 123, 3356
- Reid, I. N., Hawley, S. L., & Mateo, M. 1995, *MNRAS*, 272, 828
- Reid, I. N., & Hawley, S. L. 2005, *New Light on Dark Stars*, (2nd ed.; New York: Praxis/Springer)
- Reid, I. N., Cruz, K. C., Burgasser, A. J., & Liu, M. 2008a, *AJ*, 135, 580
- Reid, I. N., Cruz, K. C., Kirkpatrick, J. D., Allen, P. R., Mungall, F., Liebert, J., Lowrance, P., & Sweet, A. 2008b, *AJ*, 136, 1290
- Reid, I. N., Kirkpatrick, J. D., Liebert, J., Burrows, A., Gizis, J. E., Burgasser, A. J., Dahn, C. C., Monet, D., Cutri, R. M., Beichman, C. A., & Skrutskie, M. F. 1999, *ApJ*, 521, 613
- Reiners, A., & Basri, G. 2007, *ApJ*, 656, 1121
- Rosner, R. 1980, in *Cool Stars, Stellar Systems and the Sun*, *SAO Special Report*, 389, 79
- Salpeter, E. E. 1956, *ApJ*, 121, 161
- Saumon, D., & Marley, M. S. 2008, *ApJ*, 689, 1327
- Schmidt, S. J., West, A. A., Hawley, S. L., & Pineda, J. S. 2010, *AJ*, 139, 1808
- Skrutskie, M. F., et al. 2006, *AJ*, 131, 1163
- Smith, R. C. 1983, *Observatory*, 103, 29
- Stauffer, J. R., et al. 2007, *ApJs*, 172, 663
- Stelzer, B., Micela, G., Flaccomio, E., Neuhauser, R., & Jayawardhana, R. 2006, *A&A*, 448, 293
- Tarter, J. C. 1974, Ph.D. thesis, Univ. of California, Berkeley
- Treumann, R. 2006, *A&A Rev.*, 13, 229
- Tsuji, T. 1964, *Ann. Tokyo Obs. Ser. II* 9, 1
- Tsuji, T., Ohnaka, K., Aoki, W., & Nakajima, T. 1996, *A&A*, 308, L29
- van Biesbroeck, G. 1944, *AJ*, 51, 61
- West, A. A., Hawley, S. L., Bochanski, J. J., Covey, K. R., Reid, I. N., Dhital, S., Hilton, E. J., & Masuda, M. 2008, *AJ*, 135, 785
- Wielen, R. 1977, *A&A*, 60, 263
- Wright, E. L. et al. 2010, *AJ*, 140, 1868
- York, D. G., et al. 2000, *AJ*, 120, 1579
- Zuckerman, B., & Zong, I. 2004, *ARAA*, 42, 685

8 Evolution of Solar and Intermediate-Mass Stars

Falk Herwig


Department of Physics and Astronomy, University of Victoria,
Victoria, BC, Canada

1	<i>Introduction</i>	398
2	<i>The Physics for Stellar Evolution</i>	400
3	<i>The Evolution of a $3 M_{\odot}$ Star</i>	405
3.1	Pre-main Sequence to the End of He-Core Burning	407
3.2	Asymptotic Giant Branch Evolution	412
3.2.1	Thermal-Pulse AGB	414
3.2.2	Nucleosynthesis	421
3.2.3	Post-AGB Evolution	430
4	<i>Evolution as a Function of Initial Mass</i>	432
5	<i>Evolution at Low Metallicity and the Dynamical Star</i>	438
5.1	Rotation	438
5.2	Dynamic Interaction of Nuclear Burning with Convection	439
	<i>Acknowledgments</i>	442
	<i>References</i>	442

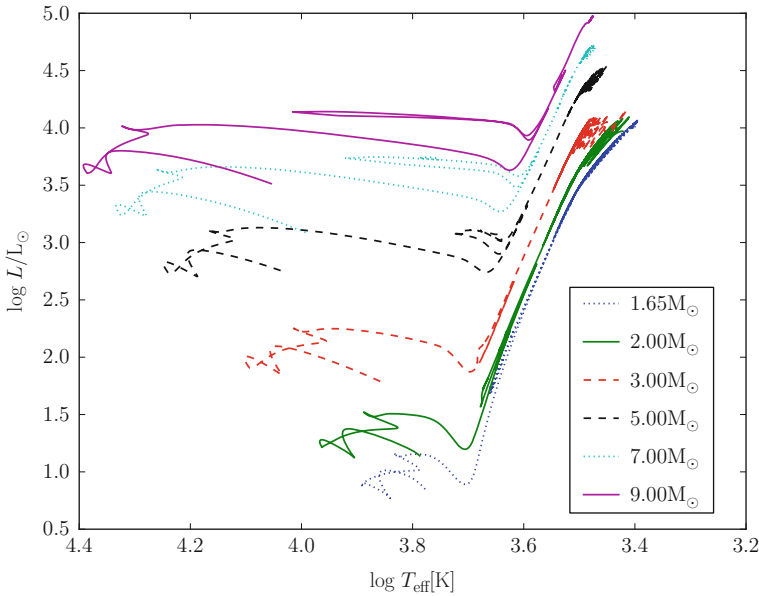
Abstract: A brief discussion of the physics and numerical techniques for stellar evolution of solar and intermediate-mass stars is followed by a detailed description of the evolution of a $3 M_{\odot}$ star from the main sequence through the red giant branch, He-core burning, and on to the thermal-pulse AGB evolution. He-shell flashes and the associated mixing as well as the nucleosynthesis are discussed, including the *s* process. Post-AGB evolution including post-AGB He-shell flashes are presented. Variations of the evolution as a function of mass, the initial–final mass relation, and lifetimes are summarized. Peculiarities of intermediate-mass AGB stars, hot-bottom burning, the super-AGB stars, as well as low-mass stellar evolution through the He-core flash are described in detail. The evolution of low- and intermediate-mass stars is presented. The final chapter covers some aspects of the dynamical star that are not well described in the spherically symmetric framework of stellar evolution.

1 Introduction

There are different ways in which stellar evolution can be grouped as a function of initial mass. The most obvious distinction is in terms of the final outcome. At solar metallicity stars more massive than $\sim 9 M_{\odot}$ will eventually explode as a core-collapse supernova, while stars with initially smaller mass will end their lives as white dwarfs. At the low-mass end, stars only a little less massive than the sun will have lifetimes exceeding the age of the universe. For example, a star with initially $0.8 M_{\odot}$ and a metal content of $Z = 0.01$ would ignite He only after $\sim 2.2 \times 10^{10}$ yr. While the detailed main-sequence evolution of such stars is important, for example, for the determination of ages in globular clusters, such low-mass stars would not yet have evolved through advanced stages and eventually into white dwarfs. However, especially these advanced stages, including the evolution following He-core burning, on the asymptotic giant branch (AGB) and through the He-shell flashes (or thermal pulses, to be used synonymously), are responsible for an important aspect of stellar evolution: the formation of the elements through nucleosynthesis.

The solar and intermediate-mass stars can therefore be defined as those which end their lives as white dwarfs. When studying the evolution of stars in this mass range, there may be different aspects to be interested in, which may be related to specific ways in which stars can be investigated through observations. For example, stars as members of populations and their appearance in the Hertzsprung–Russell diagram (HRD) or color-magnitude diagram are investigated through photometry or spectrophotometry. Such a diagram is shown in  Fig. 8-1 for a set of solar and intermediate-mass stars, and the features of these evolutionary tracks will be discussed in detail in this chapter. This area has now expanded into using abundance tracers to characterize stellar populations (e.g., Venn et al. 2004, for the case of dwarf spheroidal galaxies), aided enormously through the now common place use of multi-object, fiber optics spectroscopy. Another important avenue to investigate the physics and evolution of stars is asteroseismology, and new space missions, like NASA's Kepler mission, or previously the Canadian MOST satellite, provide data of unprecedented quality that allow probing directly the otherwise inaccessible mechanical properties of stellar interiors. A case of particular importance is of course the sun, which is very well studied through helioseismology.

Finally, stars reveal their internal workings through the surface abundances of elements, and in some cases isotopes, that can be measured spectroscopically. While stars, at least in the solar neighborhood, have formed out of material that is very similar in its composition compared to



■ Fig. 8-1

Hertzsprung–Russell diagram (HRD) for a range of masses metallicity $Z = 0.02$. Most of the pre-main sequence has been omitted for clarity. One-dimensional models during the very advanced thermal-pulse AGB phase show erratic behavior of the surface parameters due to pulsations that is not shown here for clarity; see text for discussion

the solar abundance distribution, many stars are known with very nonsolar abundance patterns. In most cases, these nonsolar abundance patterns are the result of internal nucleosynthesis processes in the deep stellar interior in combination with mixing processes that bring the nuclear burning products to the stellar surface. In some cases, observed abundance patterns are due to external contamination in a binary star system.

Binary stellar evolution cannot only lead to such a chemical pollution of a companion, but many binary stars will interact at some point in their lives. The evolution of such interacting binaries, for example, through a common envelope evolution (e.g., Ricker and Taam 2008; Passy et al. 2012) will lead to the formation of several types of very important systems, often involving at least one compact member. For example, the progenitor evolution of the important class of thermonuclear supernova (SN Type Ia) involves at least one binary interaction of this type. The impact of binarity on stellar evolution is treated elsewhere in [Chap. 13](#).

The different ways of investigating stars through observations are complemented by theoretical investigation of stellar physics. The combination of theories and models of the many microscopic and macroscopic processes, that affect the evolution of stars, into computer simulations leads to quantitative models of stellar evolution. These are, under certain simplifying assumptions, as, for example, the assumption of spherical symmetry, able to quantitatively follow the global evolution of stars from their birth on the pre-main sequence through a series of nuclear burning stages and thermal transition phases all the way to their final white dwarf star destination.

These stellar evolution models are continuously improved, then confronted with observational properties of stars described above, and then again improved. This process, while easy to spell out in principle, is not always straight forward nor immediately successful. For example, helioseismology shows that the radiative, that is, non-convective interior of the sun rotates as a solid body, while the outer convectively unstable layers rotate differentially (Thompson et al. 2003). This is at odds with some models of the effect that rotation has in stellar evolution simulations, which assume rather effective transport of angular momentum in convection zones, enforcing near-solid body rotation there, while allowing for differential rotation in the inner layers of the sun that are not mixed through convection.

While the different observational approaches to investigate the evolution and nucleosynthesis in stars are mostly covered in other chapters of this book, this chapter describes the evolution and nucleosynthesis of solar and intermediate-mass stars as it presents itself according to the state of the art of global simulations of stars through computer simulations, that is, through stellar evolution computations. With all its limitations in mind (see previous paragraph for an example), a very detailed and in many ways probably realistic picture of the overall evolution, nucleosynthesis processes and yields, and final outcomes has emerged. And as the various aspects of this picture of stellar evolution will be presented, the many unsolved and open questions will be pointed out as well.

This chapter is organized as follows. In ► Sect. 2, the physics and to a limited extend the computational ingredients needed to construct realistic models of stars and their evolution, including the nuclear physics for the energy production and synthesis of the elements, are reviewed. Then, the evolution of a $3 M_{\odot}$ star will be described from the main sequence all the way to the post-AGB phase (► Sect. 3). In ► Sect. 4, differences in the evolution of lower and higher mass stars are presented. The final section covers briefly evolution effects due to rotation and dynamical interactions between nuclear energy release and convective mixing, that is not well described within the usual one-dimensional stellar evolution framework assuming strict spherical symmetry (► Sect. 5).

2 The Physics for Stellar Evolution

The basic ingredients for stellar evolution models are known for many decades now, and the material is covered in great detail in textbooks, such as Cox and Giuli (1968), Kippenhahn and Weigert (1990), and Hansen et al. (2004), while Clayton (1968), Rolfs and Rodney (1988), and Iliadis (2007) are geared more toward nucleosynthesis and nuclear astrophysics. Habing and Olofsson (2004) deal specifically with all aspects of AGB stars.

Stellar evolution codes have been constructed since the first computers became powerful enough in the 1960s (e.g., Henyey et al. 1964; Hofmeister et al. 1964). Over time, many groups have built sophisticated simulation codes, including more detailed and realistic physics data and models and advancing the simulation capabilities to address stellar evolution in the most extreme conditions. A recent and comprehensive description of a modern multipurpose stellar evolution code was given by Paxton et al. (2011) for the MESA code, which can be downloaded – together with detailed instructions and examples – at <http://mesa.sourceforge.net>. That paper gives a complete technical account of the input physics and numerical techniques used in stellar evolution simulations today. However, the state of the MESA code described there did not include many of its more advanced capabilities yet, as, for example, modules that attempt to

account for the effect of rotation and magnetic fields. These aspects of stellar physics have been reviewed recently by Maeder and Meynet (2012) and will not be covered much in this chapter. With so many excellent resources available, this section will only attempt to provide a basic review of the most important concepts.

The conservation laws of mass, momentum, and energy can be written down in the following way (e.g., Clarke and Carswell 2003):

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0 \\ \rho \frac{\partial \mathbf{u}}{\partial t} + \rho (\mathbf{u} \cdot \nabla) \mathbf{u} &= -\nabla p + \rho \mathbf{g} \\ \frac{\partial E}{\partial t} + \nabla \cdot [(E + p) \mathbf{u}] &= \rho \dot{Q} + \rho \frac{\partial \Psi}{\partial t} \end{aligned} \quad (8.1)$$

where \mathbf{u} is the velocity, ρ the density, p the pressure, and the gravitational acceleration \mathbf{g} is given by the gravitational potential Ψ ($\mathbf{g} = -\nabla \Psi$). The total energy per unit volume is $E = \rho \left(\frac{1}{2} u^2 + \Psi + \epsilon \right)$ and consists of the kinetic, potential, and internal energy. In each case, the left side of the equations represent the time variation of a quantity and the spatial transfer, while the right side contains sources and sinks. For the momentum, equation sources are the gravitational acceleration and the pressure force, for example, of the ideal gas. In the energy, equation Q represents all sources (e.g., from nuclear reactions) or sinks (e.g., from neutrino losses).

Solving this set of conservation laws numerically with high enough spatial and time resolution, and complemented with appropriate microphysics, allows in principle to simulate all possible behaviors of stars, except the effects of magnetic fields. However, in practice, this is not possible just like that. The reason is the vastly different time scales (as well as spatial scales) that interact in stars. These are the nuclear time scale, the thermal time scale, and the dynamical time scale, and although these are very basic concepts, they are so important that it is justified to briefly review them.

The dynamical time scale indicates how fast pressure fluctuations propagate and can be approximated by $t_{\text{dyn}} \sim 1/\sqrt{G\bar{\rho}}$, where G is the gravitational constant and $\bar{\rho}$ is the average density of the star or region under consideration. For the sun, $t_{\text{dyn}} \sim 1\text{h}$. Another way to think of the dynamical time scale is to consider a star that suddenly loses its pressure support from the equation of state, like in a core-collapse supernova. The star will respond by collapsing on the dynamical time scale. Sound waves propagate through the star on the dynamical time scale. Other processes that would be considered dynamical, as, for example, the advection flows in convectively unstable parts of the star, are not proceeding on this dynamic time scale, but, in this example, on the convective turnover time scale. The thermal time scale is broadly speaking the time on which differences in internal energy or temperature equilibrate, and it is given by $t_{\text{therm}} \sim \epsilon/L$, where ϵ is again the internal energy of a region of interest and L is the associated luminosity. The thermal time scale of the initial cooling of a pre-main-sequence star to the zero-age main sequence, when H-burning first starts, is called the Kelvin–Helmholtz time scale, and it is for the sun $1.5 \times 10^6\text{ yr}$. Finally, the fraction of nuclear fuel (e.g., H in the core of main-sequence stars) that provides energy ϵ_{nuc} to the star to temporarily halt further gravitational contraction will be depleted on the nuclear time scale $t_{\text{nuc}} = \epsilon_{\text{nuc}}/L$, which is of the order to 10^{10} yr for the sun. This global nuclear time scale determines the long-term evolution of stars in thermal and hydrostatic equilibrium. It needs to be distinguished from a nuclear reaction time scale of a particular reaction that may be compared with dynamical processes like convection.

The evolution time scale of solar and intermediate-mass stars is either the thermal or the nuclear time scale, depending on their evolutionary stage.

The evolution of stars is a typical example of a multi-physics, multi-scale problem where the properties of different types of physics on vastly different scales interact in a nonlinear way. **Equation 8.1** allows for any pressure fluctuations, that is, sound waves, to propagate through the star. In particular, in the stellar interior, the sound speed ($v_s \sim \sqrt{P/\rho}$ for an ideal gas) is very high, and the Courant–Friedrich–Levy condition imposes in that case very short time steps (of the order $\Delta t = \Delta x/v_s$, where Δx is the chosen spatial grid size). In convective He-shell flashes, to be discussed in detail in **Sect. 3.2**, for example, the sound speed is $\sim 1,000$ km/s, and in present-day simulations that solve **(8.1)**, explicitly this implies $\Delta t \sim 0.01$ s. However, the convective turnover time scale is of the order 1,000s while the evolutionary time (in this case, the thermal time scale) is ~ 100 yr. Similar arguments can be made for all other phases of stellar evolution, and also, the fact that some effort is going into implicit solution schemes does not alter the conclusion that it is not possible, and will not be possible in the foreseeable future, to simulate the entire evolution of stars by solving **(8.1)**. This seems to be almost a trivial statement. However, it is important to keep in mind that the picture of the evolution of stars and their nucleosynthesis, that will be presented in subsequent sections, is based on simplifications of **(8.1)** and that there are a number of unresolved properties of stars that these simplified equations cannot accurately describe. Among these unresolved properties are many mixing processes, and these in turn are often essential for the quantitatively reliable description of nuclear production in stars.

In order to solve the conservation laws over the long thermal and nuclear time scales, **(8.1)** is solved implicitly assuming spherical symmetry and, in the Lagrangian coordinate m , the enclosed mass. The stellar evolution equations are therefore (e.g., Kippenhahn and Weigert 1990):

$$\begin{aligned} \frac{\partial r}{\partial m} &= \frac{1}{4\pi r^2 \rho} \\ \frac{\partial P}{\partial m} &= -\frac{Gm}{4\pi r^4} - \frac{1}{4\pi r^2 \rho} \frac{\partial^2 r}{\partial t} \\ \frac{\partial l}{\partial m} &= \sum_i \epsilon_i \end{aligned} \tag{8.2}$$

where r is the radius, l is the luminosity, and the ϵ_i stand for the various energy sources and sinks (see the chapter on stellar structure for more details). The implicit solution of these equations allows to advance an evolution sequence numerically with time steps of the order of a fraction of $\min(t_{\text{nuc}}, t_{\text{therm}})$. Often, the inertial term in the momentum equation is ignored since stars evolve in almost perfect hydrostatic equilibrium ($t_{\text{dyn}} \ll \min(t_{\text{nuc}}, t_{\text{therm}})$). Indeed, whenever the inclusion of the inertial term is really necessary, the star is probably also violating the assumption of spherical symmetry. Most dynamic motions of fluid elements in stars are not spherically symmetric but turbulent. These issues will be discussed more in **Sect. 5**. Here, it is only important to fully appreciate that in real stars, there are relevant processes at work that are not described by **(8.2)**.

In order to solve **(8.2)**, the expression for the temperature gradient $\partial T/\partial m$ has to reflect the dominant mode of energy transport. In solar and intermediate-mass stars, three modes of energy transport are relevant. In convective layers, the temperature gradient is adiabatic, at least in the stellar interior. In the envelope where convection may be inefficient, or in convectively

stable layers, the temperature gradient is given by optically thick, gray radiation diffusion:

$$\left. \frac{\partial T}{\partial m} \right|_{\text{rad}} = -\frac{3}{64\pi^2 a c} \frac{\kappa l}{r^4 T^3} \quad (8.3)$$

where κ is the Rosseland mean opacity or the electron conduction coefficient, a is the radiation-density constant, and c the speed of light. In the electron-degenerate cores (the white dwarfs to be) of solar and intermediate-mass stars in the advanced evolution stages, electron conduction dominates the energy transport.

The detailed microphysics entering modern stellar evolution calculations is described elsewhere. As an example, the options available in the MESA code are documented in Paxton et al. (2011). Radiative opacities depend on the assumed composition mix. Standard tables are available for solar or solar-scaled compositions. An important improvement was the introduction of tables with more general abundance mixes, in particular with enhanced C and O abundances (Iglesias and Rogers 1996). These allow a more realistic treatment of advanced stages of stellar evolution when the C and O abundance change due to nuclear burning and mixing processes. One of the results of the adoption of these more realistic opacity tables is the greater difficulty of computing advanced AGB models. The large opacity of C and O enriched AGB envelopes in high-mass loss phases at the tip of the AGB makes surface layers only weakly bound, and in this phase, the spherically symmetric simulation is at or past its application limit. Another important advancement has been the introduction of chemistry-dependent low-temperature molecular opacities (Marigo 2002; Cristallo et al. 2007; Weiss and Ferguson 2009; Marigo and Aringer 2009). In particular, when AGB stars become C-rich, the added sources of molecular opacity in this chemistry regime make the stars ~ 200 – 300 K cooler and probably boost mass loss through enhanced dust formation (Sect. 3.2).

The equation of state of solar and intermediate-mass stars is dominated by either the ideal gas equation of state (including partial ionization where applicable) and the pressure from the degenerate electron gas in the cores in advanced evolution stages (Figs. 8-3 and 8-6). Modern EOS routines combine tabulated data from different specialized sources that take nonideal effects appropriately into account (e.g., Paxton et al. 2011, for references).

The terms ϵ_i in (8.2) contain neutrino losses, for example, through plasmon decay and bremsstrahlung, that are important in the high-density environments of the degenerate cores in the advanced evolution stages of solar and intermediate-mass stars. In addition to this energy sink, the energy generation ϵ_{nuc} from nuclear reactions needs to be taken into account, and these are simultaneously responsible for composition changes:

$$\left. \frac{d\mathbf{X}_j}{dt} \right|_{\text{burn}} = \hat{F}_j \cdot \mathbf{X}_j \quad (8.4)$$

where \mathbf{X}_j is a vector containing the mass fractions of all species in zone j and the operator \hat{F}_j represents all reactions through which these species can interact. Instead of in terms of mass fraction X , the network equations are rather written in terms of the number density N_m of species m by collecting all production and destruction terms of reactions of the type $k + l \rightarrow m + n$:

$$\frac{dN_m}{dt} = N_k N_l < \sigma v >_{kl,m} - N_m N_n < \sigma v >_{mn,o} + \dots + N_i \lambda_{i,m} - N_j \lambda_{m,j} \quad (8.5)$$

where $\langle \sigma v \rangle$ is the product of the cross section and the relative velocity in the center-of-mass system averaged over the appropriate distribution function and λ is the rate for β decays. Reaction rates can be obtained, for example, from (Angulo et al. 1999, http://pntpm.ulb.ac.be/Nacre/nacre_d.htm). The number density is expressed in terms of a number fraction or mole fraction $Y = X/A$, with A the atomic mass number, by $N = Y\rho N_A$ where $N_A = 1/M_u$ is the Avogadro number and M_u is the atomic mass number. For each reaction, an amount of energy Q is released, which is partly in the form of neutrinos and therefore lost. The energy released from ΔN reactions is therefore $\epsilon_{\text{nuc}} = \Delta N Q = \Delta Y \rho N_A Q$.

Typically, in a stellar evolution simulation, only a comparatively small network needs to be considered. It only needs to be large enough to take into account the reactions that are important for the energy generation. For stars that end their lives as CO-white dwarfs, this can be accomplished with approximately 15 species. A post-processing simulations can then take care of the many trace elements that are produced in stars but do not feed back into the thermal evolution, such as the s -process elements (► Sect. 3.2).

In addition to nuclear reactions, abundance changes are due to mixing processes, which are often taken into account in a time-dependent manner through a diffusion equation in the Lagrangian coordinates:

$$\left. \frac{dX_i}{dt} \right|_{\text{mix}} = \frac{\partial}{\partial m} \left[(4\pi r^2 \rho)^2 D(m) \frac{dX_i}{dm} \right] \quad (8.6)$$

where X_i is the radially dependent mass fraction of species i and D is a radially varying diffusion coefficient that reflects the different types of mixing in stars that the calculations take into account. The use of the diffusion equation is just a numerically convenient mathematical tool, whereas the processes treated in this way often have nothing to do with diffusion. Most mixing in stars is a result of convection, which is not a diffusion but an advection process. However, in practice, the mixing efficiency inside the convection zone is so high that one can even assume instantaneous mixing and the details of the mixing algorithm do not matter. But there are important exceptions, as, for example, when dealing with mixing due to rotation-induced instabilities (Maeder and Meynet 2012). It is also not clear if the partial mixing at convective boundaries is correctly described by a diffusion process. Finally, in situations where reactions are occurring fast compared to the convective turn-over and large-scale feedback of energy release from nuclear burning into the convective flow can be expected, the diffusion approximation is also probably not applicable (Herwig et al. (2011), see ► Sect. 5 for details).

Macroscopic processes that are responsible for mixing in single stars can be divided into two categories according to the time scale on which the driving mechanism operates. An efficient convection zone (e.g., in the deep stellar interior) has an adiabatic temperature stratification. Convection occurs in stars when energy transport through radiation alone is not efficient enough and the radiative temperature gradient (► 8.3) is larger than the adiabatic gradient. As can be seen from the nominator of (► 8.3), convective instability is triggered by a large luminosity, for example, from a burning shell in the interiors of stars during advanced phases of evolution, or because of large opacity, typically in the cool outer envelopes of stars, for example, during the red and asymptotic giant branch phase. Convective velocities are of the order of several km/s and lead to advective mixing that is fast compared to the evolutionary time scale of the star. Convection will also trigger mixing at or near convective boundaries, through overshooting, entrainment, or the excitation of internal waves in the stable layers neighboring the convection zones. These convection-induced mixing processes are already slower than

proper convective mixing. The three-dimensional and turbulent nature of convection can only be captured eventually by highly resolved numerical solutions of (☛ 8.1), while stellar evolution (☛ 8.2) needs to be supplemented with an approximate model of convection that is based on the available radially and temporal averaged, local values for the state variables. The simplistic model that is still used most often is the mixing-length theory (Böhm-Vitense 1958), and it has been pointed out that “one does not therefore expect quantitative results derived on the basis of this theory to have high accuracy or reliability” (Cox and Giuli 1968). Work to improve this situation is much needed and underway (e.g., Xiong et al. 1997; Canuto et al. 1996; Arnett et al. 2010) and supplemented by increasingly more realistic 3D simulations of convection (e.g., Porter and Woodward 2000; Browning et al. 2004; Freytag and Höfner 2008; Brun and Palacios 2009).

Even slower than that are mixing processes related to both thermal and abundance diffusivity or secular processes. These include thermohaline mixing and semi-convection where in the first case a thermally stable stratification is becoming unstable due to a destabilizing gradient of the mean molecular weight, or the other way around in the latter case. Despite renewed research effort (e.g., Charbonnel and Zahn 2007; Denissenkov 2010; Denissenkov and Merryfield 2011; Canuto 2011), these double-diffusive instabilities and the associated amount of mixing in stars are quantitatively not well understood, which represents another source of uncertainty in stellar evolution simulations. Rotation will induce mixing, for example, through meridional circulations or shear-driven flows (Endal and Sofia 1976; Pinsonneault 1997; Langer et al. 1999; Maeder and Meynet 2012). These slow mixing processes may be associated with velocities of the order of cm/s. Significant uncertainties exist in the quantitative treatment of mixing processes.

The physics ingredients of stellar evolution calculations briefly reviewed so far need to be complemented with a numerical solution scheme. Since the early days of stellar evolution, the solution scheme of choice has been a fully implicit Newton–Raphson scheme (Henyey et al. 1964), which is also adopted by the MESA code (Paxton et al. 2011). The three operators structure (☛ 8.2), mixing (☛ 8.6), and nuclear burn (☛ 8.4) may or may not be coupled in different combinations. Formally, the operators can be solved separately if the associated time scales are largely different. For solar and intermediate-mass stars, this is most of the times the case. Nevertheless, it is found that the fully coupled solution, that is adopted, for example, in the MESA code, seems to provide greater numerical efficiency and robustness. There are many additional numerical details, for example, concerning adaptive mesh refinement and boundary conditions that need to be addressed. At the end, a modern stellar evolution code can follow the entire evolution from the pre-main sequence to the white dwarf stars in a day or two on a present-day multi-core workstation. However, the difficulties of simulating the very advanced last thermal pulses on the AGB, as mentioned already above, remain (cf. ☛ Sect. 3.2).

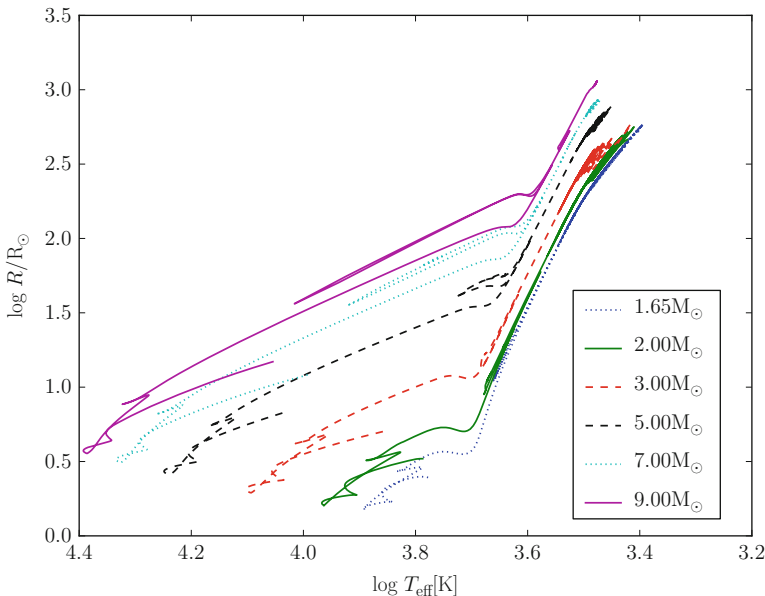
3 The Evolution of a $3 M_{\odot}$ Star

This section will describe the evolution of a $3 M_{\odot}$ star according to the results of stellar evolution computations. This mass is representative of the lower-mass regime of CO-white dwarf progenitors which will experience the third dredge-up during the thermal-pulse AGB (☛ Sect. 3.2.1) and become C-rich and produce *s*-process elements. However, this mass avoids the complication of the He-core flash of lower initial mass, and it also avoids the effect of hot-bottom burning and C-ignition that is characteristic for more massive AGB stars. These additional evolutionary

options as a function of initial mass will be discussed in [Sect. 4](#). Also, the evolution described in this section is for metallicity $Z = 0.02$, which is similar or maybe slightly higher compared to the sun's metallicity (Asplund et al. 2009). The evolution of stars with lower metal content has some peculiarities that will be presented in [Sect. 5](#).

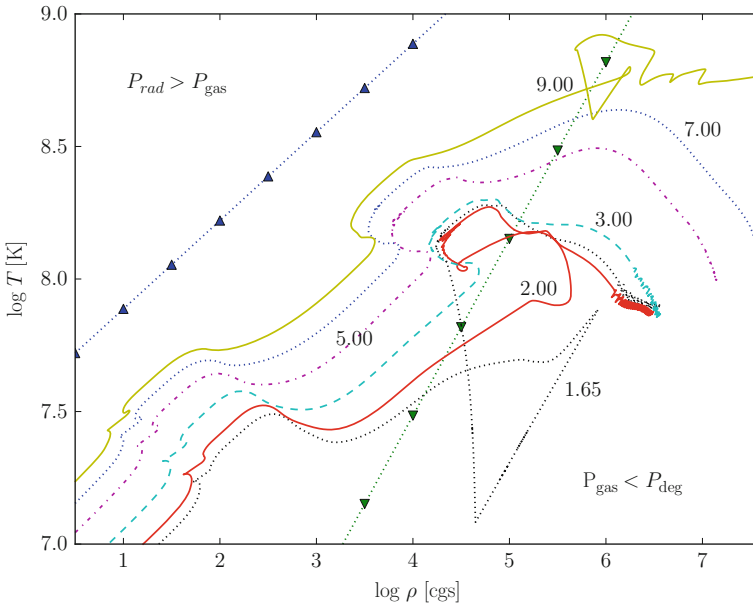
The evolution in the HRD is shown in [Fig. 8-1](#). The $3 M_{\odot}$ star evolves from the pre-main sequence toward the zero-age main sequence at $\log T_{\text{eff}} = 4.093$ and $\log L = 1.905$. Here it has its smallest radius and highest surface gravity, and the overall evolutionary trend for the star from hereon all the way to the tip of the AGB is to become cooler, more luminous, and, because $L = 4\pi R^2 \sigma T_{\text{eff}}^4$ (with σ the Stefan-Boltzmann constant), also larger, as is shown in [Fig. 8-2](#).

Stellar evolution can be thought of as a continuous process of gravitationally induced collapse, temporarily halted by the ignition of and energy release from nuclear burning processes. The evolution of the central $\rho_c - T_c$ is shown in [Fig. 8-3](#). In solar and intermediate-mass star evolution, advanced burning stages will be ignited off-center, due to the effect of neutrino cooling in the highest-density central locations. Nevertheless, the evolution of some key species involved in nuclear processes for energy generation in the center of the star is very instructive and demonstrates the temperature dependence of the different burning processes. Such a diagram is shown in [Fig. 8-4](#) for the highest mass case discussed in this chapter, $9 M_{\odot}$ (because only the highest mass eventually ignites C in the center). However, the ignition temperatures for the various nuclear burning processes are only weakly dependent on the density, and the corresponding $3 M_{\odot}$ diagram would look similar.



■ Fig. 8-2

Same as [Fig. 8-1](#) but for the stellar radius. $\log T_{\text{eff}} - \log R$ diagram for a range of masses with metallicity $Z = 0.02$



■ Fig. 8-3

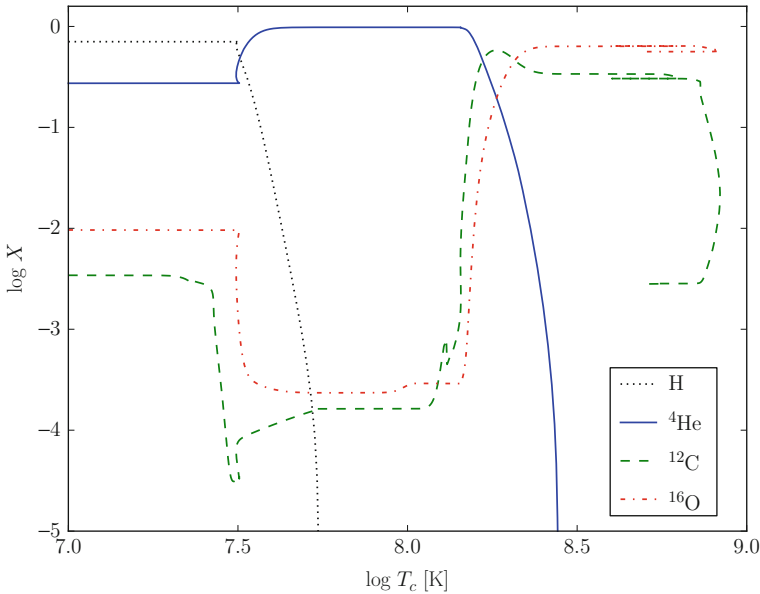
Evolution of the central density and temperature for the tracks shown in [Fig. 8-1](#), plus a 7 and 9 M_{\odot} case. *Straight lines with symbols divide areas dominated by radiation pressure, gas pressure, and the area dominated by the equation of state of degenerate electrons*

Abundances profiles of for key species of the reaction network are given for representative evolution stages of the internal evolution in [Fig. 8-5](#) and are complemented by profiles in the $\rho_c - T_c$ -diagram ([Fig. 8-6](#)).

Furthermore, the so-called Kippenhahn diagram shows the time evolution of the mass coordinate of convection zones as well as in this case the mass coordinate of the H-free core and the He-free core (and for the highest mass cases also the ^{12}C -free core). All three mass coordinates are defined in this chapter as the smallest mass coordinate where the mass fraction of the respective species is $X < 10^{-4}$. These lines correspond either to the location of the nuclear burning shell or, in the case of a dredge-up process, to the location of the bottom of the convection zone that causes the dredge-up. A Kippenhahn diagram for the 3 M_{\odot} case is shown in [Fig. 8-7](#).

3.1 Pre-main Sequence to the End of He-Core Burning

Stars form on the pre-main sequence in the cool and luminous part of the HRD. Because they are cool, the opacity is high and they are throughout convectively unstable. For that reason, they are well described by a polytrope with the index $n = 3/2$ of an adiabat. During this phase without significant nuclear energy sources, the star contracts on the short Kelvin–Helmholtz time scale ([Sect. 2](#)). This short-lived convection therefore only shows up as a vertical line at $\log(t_{\text{final}} - t) = 8.68$ in [Fig. 8-7](#).

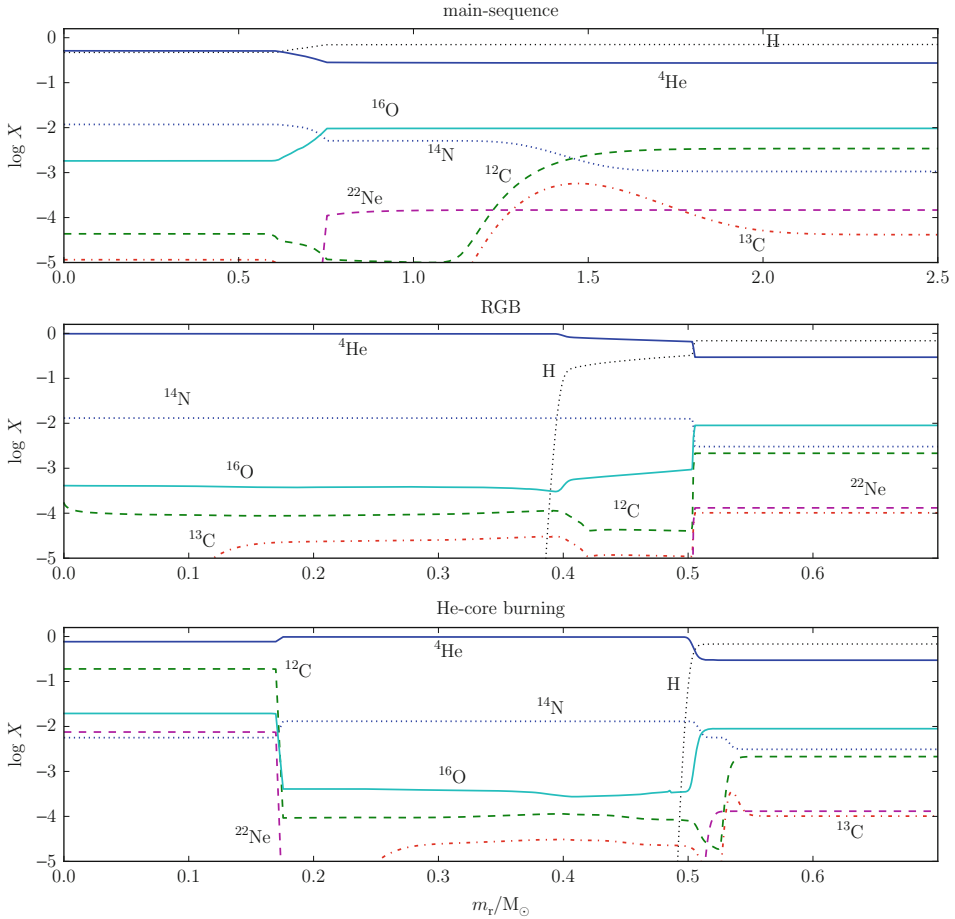


■ Fig. 8-4

Evolution of the central mass fractions of key isotopes that trace the main nucleosynthesis processes, as a function of the central temperature for the $9 M_{\odot}$, $Z = 0.02$ stellar evolution track. The major nuclear burning stages evident in the abundance variations cause prominent features in the $\rho_c - T_c$ diagram (► Fig. 8-3). This super-asymptotic giant branch star is right at the border to the massive star regime, and hence, C-ignition causes the ^{12}C abundance to decrease at the highest temperature (see ► Sect. 4)

► Figure 8-4 shows that core H-burning takes place at $\log T = 7.5$ in the $9 M_{\odot}$ stellar model. For the $3 M_{\odot}$ star H burns already at $\log T = 7.4$ and $\log \rho = 1.6$. The ignition of H in the center causes a characteristic feature in the $\rho_c - T_c$ -diagram (► Fig. 8-3). The central evolution toward higher ρ and T is halted during the major core burning phases. The $\rho_c - T_c$ profile (► Fig. 8-6) which also features markers for mass coordinates shows that the innermost $0.65 M_{\odot}$ are relatively compact and have only small dynamic range in density and temperature.

In these higher-mass stars, H-burning proceeds via the CNO cycle that depends on the availability of C, N, and O isotopes as a catalyst. The basic cycle involves p -captures on ^{12}C , ^{13}C , ^{14}N , and the $^{15}\text{N}(p, \alpha)^{12}\text{C}$ reaction that closes the cycle. The relative abundances of the CNO isotopes are reshuffled during CNO cycling so that ^{14}N , which has the largest Coulomb barrier and therefore the smallest p -capture cross section, has the highest abundance in CNO-cycle H-burning ashes (Arnould et al. 1999). This can be seen on the left of the top panel in ► Fig. 8-5 that shows the central region of the $3 M_{\odot}$ during the main sequence. The initial (solar-scaled) abundance distribution can be seen on the right side of that panel, which shows the unprocessed and at this point also still unmixed surface composition. There, ^{16}O dominates the CNO isotopes. The workings of the CNO cycle can also be seen in ► Fig. 8-13, which shows a H-burning situation in a later phase of the evolution. However, the CNO cycle is the same, and the profiles show the signatures of the various reactions, being activated from lower to



■ Fig. 8-5

Abundance profiles for the same evolutionary stages as shown in ► Fig. 8-6 (except the AGB profiles, which will be described in detail later, e.g., ► Fig. 8-14). The evolutionary stages shown at times $\log t_{\text{final}} - t = 8.45, 8.04, 7.91$ (top to bottom) on the scale used in ► Fig. 8-7

higher temperatures from the right (outer layers) to the left (layers further inward). The effect of H-burning via the CNO cycle is to turn most ^{16}O and ^{12}C into ^{14}N . This will be important, as shown in later sections, for the s process.

Nuclear reaction rates have a high temperature dependence, which means that nuclear burning is confined only to the most central region where the temperature peaks. Despite that, the abundance profiles are flat in the central region of the main-sequence star (top panel in ► Fig. 8-5), because that core region is convectively unstable and therefore well mixed. As shown in the Kippenhahn diagram (► Fig. 8-7, $\log(t_{\text{final}} - t) = 8.68\text{--}8.07$), the convection zone is decreasing in mass extent. This explains the shallow slope of the abundance profile of species like H, He, and ^{16}O (► Fig. 8-5) between $m_r = 0.61$ and $0.75 M_\odot$. The slope is preserving the central abundances as a function of time according to the decreasing convective core.

While nuclear energy generation during the main sequence is restricted to the central region, more nucleosynthesis with consequences that are eventually observable is taking place in layers further out in the star. These layers are convectively stable (radiative), and in the stellar evolution models no mixing takes place. Therefore, even slow, nonequilibrium nucleosynthesis can build up abundance features. In particular, around $m_r \sim 1.5 M_\odot$, a layer with enhanced ^{13}C results from the activation of the first reactions in the CN cycle: $^{12}\text{C}(p, \gamma)^{13}\text{N}(^+\beta)^{13}\text{C}(p, \gamma)^{14}\text{N}$. During the subsequent red giant branch (RGB) or first giant branch phase, when a deep convective envelope emerges, this ^{13}C -bump will be mixed throughout the envelope (middle panel, [Fig. 8-5](#)) and to the surface where it leads to an observed decrease of the $^{12}\text{C}/^{13}\text{C}$ abundance. This first convective mixing of nuclear processed material to the surface due to a deepening envelope convection zone is called the first dredge-up. The second ([Figs. 8-23](#) and [8-24](#)) and third ([Figs. 8-9](#) and [8-10](#)) dredge-up are encountered after the end of He-core burning and after the end of He-shell flashes, respectively.

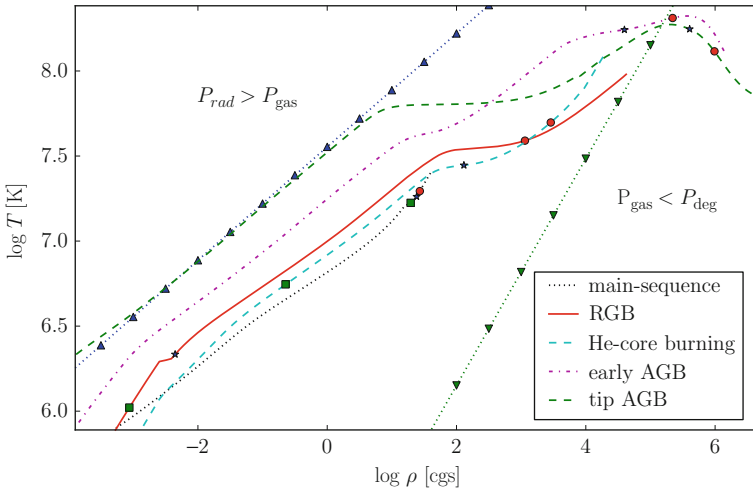
When H in the center is exhausted, the main-sequence phase ends, which in the HRD is the case at $\log T_{\text{eff}} = 3.956$ and $\log L = 2.142$. A characteristic switchback feature marks the terminal age main sequence and the transition from core H-burning to H-shell burning. The location of the H-shell is now tracked by the line for the H-free core in the Kippenhahn diagram ([Fig. 8-7](#)). The core, not supported anymore by nuclear burning, contracts; H-burning starts in the shell, and the outer layers expand. The model star evolves at almost constant luminosity from the tip of the main sequence to the bottom of the RGB at $\log T_{\text{eff}} = 3.698$ (adopted here to be at the local minimum of L at $\log L = 1.177$), and this readjustment on the thermal time scale lasts 3.684×10^6 yr. At this stage, the H-burning shell is located at $m_r = 0.305 M_\odot$. Because of the decreasing temperature in the envelope, the opacity increases and the outer layers become unstable against convection ([8.3](#)). At the bottom of the RGB, the bottom of the convective envelope has reached $m_r = 2.617 M_\odot$ ([Fig. 8-7](#), at $\log(t_{\text{final}} - t) \sim 8.049$).

The star now evolves on the RGB, with an inert He core, a H-burning shell, and further out a large convective envelope. The structure has dramatically changed compared to the main-sequence configuration, as can be seen by comparing the $\rho_c - T_c$ profiles ([Fig. 8-6](#)). The dynamic range between the three marked mass coordinates has significantly increased. The profiles shown in [Figs. 8-5](#) and [8-6](#) are taken when the stellar evolution track has almost reached $\log L = 2.602$ which is very close to the tip of the RGB at $\log L = 2.666$ for the $3 M_\odot$ track. He-core burning is about to start, and in fact, a tiny uptick of the ^{12}C abundance can already be noticed in the center (middle panel, [Fig. 8-5](#)). ^{13}C is already gone because of the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ reaction.

While the $0.35 M_\odot$ mark is in the He core just below the H-burning shell at $0.393 M_\odot$, the $0.5 M_\odot$ mark is located just below the bottom of the convective envelope, while the $0.65 M_\odot$ mark is well inside the convective envelope. The location of these markers on the $\rho_c - T_c$ profiles demonstrates the large density and temperature contrast between the core and the envelope in giant stars. This contrast is even more extreme in the AGB stars.

The H-burning shell is relatively shallow, compared to lower-mass RGB stars, with a gently sloping H-profile between the H-burning shell and the bottom of the convective envelope. The latter is characterized by steep abundance gradients and flat profiles above. In the layer beneath the convective envelope, ^{22}Ne is destroyed by H-burning via the Ne-Na cycle.

The evolution from the bottom to the tip of the RGB (i.e., from the minimum to the maximum in L) takes 7.99×10^7 yr. The RGB profile shown in [Figs. 8-5](#) and [8-6](#) is at 5.91×10^6 yr before the RGB tip, when the He core has reached a mass of $0.400 M_\odot$ and He ignites in the core. In the $9 M_\odot$ star, this happens at $\log T \sim 8.17$ ([Fig. 8-4](#)) while in the $3 M_\odot$ case, He-core burning starts at $\log T \sim 8.06$. The core becomes convectively unstable because of



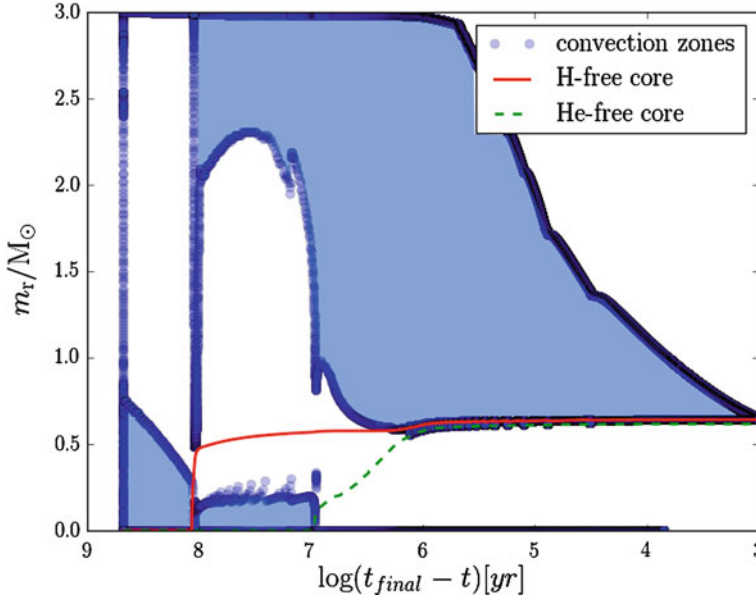
■ Fig. 8-6

Density-temperature profiles for the $3 M_{\odot}$ stellar evolution track for the evolutionary stages shown at times $\log t_{\text{final}} - t = 8.45, 8.04, 7.91, 6.23$, and 3.38 (in the order in which the lines appear in the legend) on the scale used in ► Fig. 8-7. The marks along each profile indicate the enclosed mass 0.35 (red circle), 0.5 (blue star), and $0.65 M_{\odot}$ (green square). Straight lines with triangle up and down marks divide areas dominated by radiation pressure, gas pressure, and the area dominated by the equation of state of degenerate electrons

the large energy generation from He burning. In the next 7.01×10^6 yr, the model star evolves to below $100 L_{\odot}$. This is a typical luminosity for the He-core burning phase which lasts 1.02×10^8 yr and ends at $\log(t_{\text{final}} - t) 6.967$ in ► Fig. 8-7.

The ignition of He produces again a characteristic feature in the $\rho_c - T_c$ evolution, similar to when H ignited. Compared to the RGB, the density has decreased in the center, and the temperature has increased. The star is again more in a dwarflike configuration with a much smaller range in temperature and density from the center to the surface, compared to the giant configuration. The Kippenhahn diagram shows that the convective envelope has receded and the bottom of the convection zone reaches a maximum mass coordinate of $2.30 M_{\odot}$. The abundance profile (bottom panel, ► Fig. 8-5) shows how ^{12}C is forming from the triple- α reaction. Several other He-burning reactions are now activated in the core, including $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ as well as the α -capture chain $^{14}\text{N}(\alpha, \gamma)^{18}\text{F}(\beta^+)^{18}\text{O}(\alpha, \gamma)^{22}\text{Ne}$. In massive stars, this reaction chain transforms ^{14}N from the CNO cycle into the neutron source ^{22}Ne ($^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$), and in those stars, this ^{22}Ne plays an important role for the weak s process. However, in solar and intermediate-mass stars, this core material will be eventually buried in the white dwarf. The relevant nuclear production site for the s process is during the AGB stage.

During the He-core burning phase, H burning continues in the shell. The H-free core reaches $0.57 M_{\odot}$ at the end of He-core burning, and throughout this phase, the luminosity from H burning exceeds the He-burning luminosity for this initial mass. Contrary to the core convection zone during H-core burning, the convective core during He-core burning grows. In this particular situation, the radiative and adiabatic gradients outside the proper core convection region have about the same value within a mass shell of mass $\leq 0.05 M_{\odot}$ due to opposing trends impacting the radiative gradient with an outward decreasing luminosity and increasing



■ Fig. 8-7

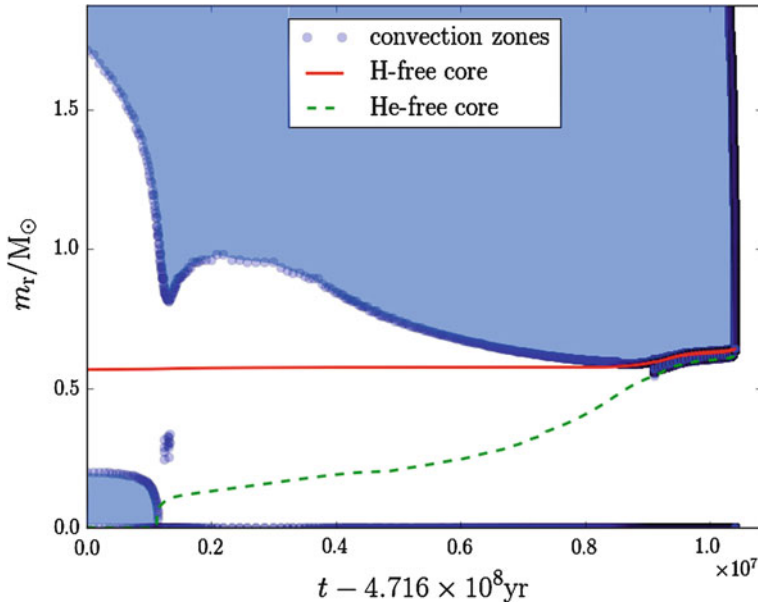
Kippenhahn diagram of $M_{\text{ZAMS}} = 3 M_{\odot}$ stellar evolution calculation with $Z = 0.02$ from the pre-main sequence to the end of the TP-AGB evolution when all envelope mass is lost. The position of convection zones and the mass coordinate below which the star is H- and He-free is shown as a function of reverse logarithmic time. In this way, the abscissa represents the logarithm of the time left until the end of the AGB evolution, which is reached when almost all the envelope mass is lost and the star will leave the giant branch and evolves into the central star of a planetary nebula (► Sect. 3.2)

opacity. This is a case where secular instabilities (in this case semi-convection) that were discussed in ► Sect. 2 play a role. This introduces an uncertainty for post-He-core burning core masses. When the He in the center is completely transformed into ^{12}C and ^{16}O , He-core burning is finished.

3.2 Asymptotic Giant Branch Evolution

The events after the end of He-core burning are very similar to those at the end of the main-sequence evolution. The core resumes contraction while He-burning now switches to a shell. As a response to the contracting core, the envelope expands again, and the envelope convection deepens again. This and the subsequent evolution along the early AGB up to the thermal pulses are shown in ► Fig. 8-8. In stars below $M_{\text{ZAMS}} \lesssim 3.5 M_{\odot}$, the bottom of the convective envelope will not penetrate into the H-free core (when this does happen, it is referred to as the second dredge-up, ► Figs. 8-23, and ► 8-24).

In the $\rho_c - T_c$ diagram, both the evolution of the central conditions (► Fig. 8-3) as well as the profile for the early AGB and even more so the tip AGB star (► Fig. 8-6) take a turn into



■ Fig. 8-8

Kippenhahn diagram of $M_{\text{ZAMS}} = 3 M_{\odot}$ evolution from the end of core He burning through the early AGB to the end of the thermal-pulse AGB. The wiggles at the *bottom* of the convective envelope at $\sim 0.6 M_{\odot}$ on the *right* end of the diagram are the thermal pulses. On the time scale of this diagram, the envelope is lost through winds at an instant (cf. ● Fig. 8-7)

the regime dominated by the electron degenerate equation of state. The temperature decreases toward the center and in the center as function of time. Neutrino losses extract so much energy from the central region so that C burning, which would be the next nuclear burning process, cannot ignite. What will become eventually the white dwarf, an inert CO core supported through the pressure from degenerate electrons, starts to form after the end of He-core burning.

Contrary to the He-core burning phase, the He-burning luminosity from the He shell now by far exceeds the H-shell luminosity during the early AGB. In fact, the main thing that happens during this transition phase is the catching up (in mass coordinate) of the He-shell with the H-shell. Once the two shells are close to each other in mass, they start to interfere with each other, which leads to the He-shell flash or thermal-pulse phenomenon. As can be seen from the $\rho - T$ profiles (● Fig. 8-6), the model star has turned again into a giant configuration, and the envelope continues to expand as the core contracts. The phenomenon, that in stars with nuclear shells the envelope does the opposite of what the core does, is also known as the mirror principle (Kippenhahn and Weigert 1990). For the early AGB star, the $0.65 M_{\odot}$ mass coordinate is already at such low density and temperature that it falls out of the range of the diagram. More quantitative details on the early AGB evolution can be found in Forestini and Charbonnel (1997).




3.2.1 Thermal-Pulse AGB

Once the mass coordinate of the He-free core has approached the location of the H-free core to within $\sim 0.04 M_{\odot}$, the conditions for He-shell flashes are given. The He-burning shell is feeding on the ashes of the H-burning shell. For both to be burning stable in a steady-state manner, the He shell would have to process the same amount of mass that the H shell is producing as ashes. For this to happen, the conditions would have to be fine-tuned in an unlikely way, and stellar evolution simulations have not found any cases with simultaneously stable H and He burning.

He-Shell Instability

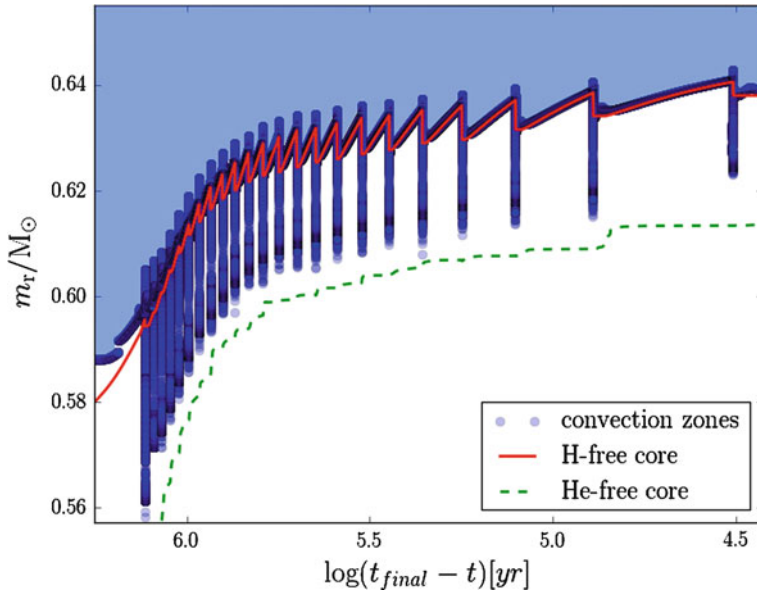
Instead, once the He shell has approached the H-burning shell in mass coordinate toward the end of the early AGB, H burning will take over the lead in energy production. The H-burning shell will support a lot of the pressure from the overlying envelope mass, and the temperature in the layer below does not allow full-blown He burning. He continues to be accreted from the H shell above and is accumulating. The region of the He shell is contracting and compressed due to the lack of energy production. The He shell is getting geometrically thinner, considering that the shell region is defined as the region between the absence of fuel at the bottom and the absence of high enough temperature at the top. Depending on the mass, this region will also become increasingly degenerate. In the $3 M_{\odot}$ evolution track, the thin-shell instability will eventually trigger the He-shell flash. The basic idea is that with very little volume in the shell, at the onset of He-burning via the triple- α reaction with a very-high-temperature exponent, an increase of the energy generation cannot immediately lead to an expansion via PdV work. Locally and temporarily, the pressure is effectively decoupled from the temperature, which results in a thermonuclear runaway. A complete and more rigorous analysis taking also into the account the effect of degeneracy is given by Yoon et al. (2004).

The He-shell flash instability plays a central role in the final evolution of solar and intermediate-mass stars. They cause periodic convective mixing of the intershell, which is the region between the He and H shell. And they cause the third dredge-up mixing that brings the bottom of the convective envelope down into the intershell and, like the previous dredge-up episodes, mixes processed material from the intershell to the surface. The thermal pulses (the terms thermal pulse and He-shell flash are used synonymously) also trigger mixing episodes that are critical for the activation of the neutron source reactions $^{13}\text{C}(\alpha, n)^{16}\text{O}$ and $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$. These processes will now be described in some detail, while an in-depth coverage of thermal-pulse AGB evolution has been given by Herwig (2005).

He-shell flashes are recurrent.  Figure 8-9 shows the complete series of flashes for the $3 M_{\odot}$ evolution track, while  Fig. 8-10 provides details of the 12th thermal and the subsequent interpulse and 13th thermal pulse. The interpulse time, the time between thermal pulses, depends on the core mass (Forestini and Charbonnel 1997). For the advanced thermal pulses shown in  Fig. 8-10, our simulation has an interpulse time of $t_{\text{ip}} \sim 6 \times 10^4$ yr. Models with higher core mass have shorter interpulse time.

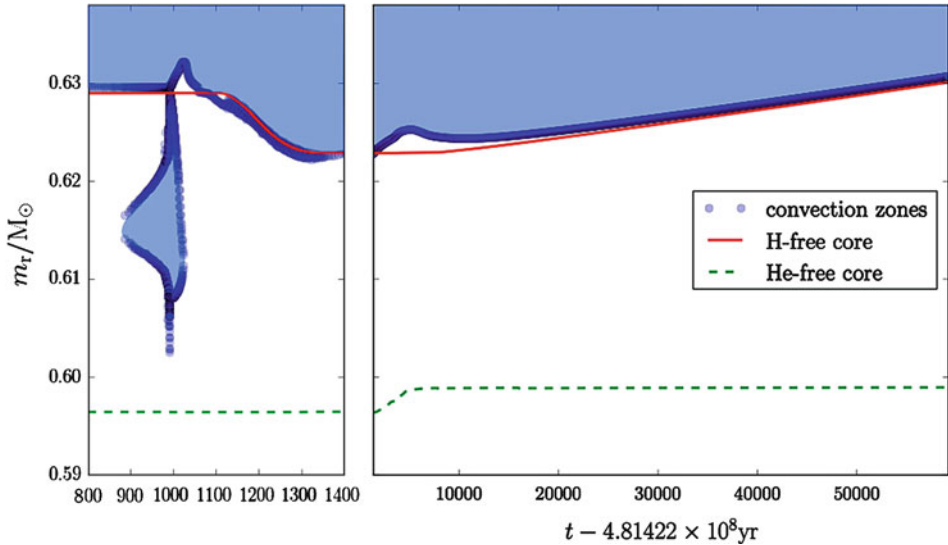
Other global quantities of the AGB thermal-pulse evolution are also strongly correlated with the core mass, such as the luminosity (Paczynski 1970; Herwig et al. 1998; Marigo 2000) or the third dredge-up (Karakas et al. 2002). For example, Blöcker (1993, cited from Herwig et al. 1998) gives the following core-mass luminosity relationship:

$$L/L_{\odot} = 62,200(M_{\text{H}}/M_{\odot} - 0.487) . \quad (8.7)$$



■ Fig. 8-9

Zoom-in of Kippenhahn diagram [Fig. 8-7](#), showing the sequence of thermal pulses. Sections where the mass coordinate of the H-free core increases correspond to H-shell burning phases, while a decrease of the H-free core indicates mixing of processed material from the intershell region (between the H- and He-free core) into the envelope via the third dredge-up



■ Fig. 8-10

Kippenhahn diagram of the 12th thermal pulse (*left panel*) and the following interpulse phase (*right panel*). Note the different time scale for the two panels. The thermal pulses are very short compared to the interpulse phase

This so-called asymptotic behavior has motivated the construction of synthetic AGB models. These do not solve the full set of stellar evolution equations (► 8.2) but describe the properties of full stellar models through fitting formula of key quantities, such as the temperature in nuclear processing regions (e.g., Renzini and Voli 1981; Groenewegen and de Jong 1993; Marigo 2001).

Mass Loss

The total number of thermal pulses in an AGB sequence will be determined by the assumed mass loss. Mass loss will depend on the stellar mass, the temperature, the luminosity, the surface chemistry, and the envelope pulsations. It is the result of a complex interplay between mechanical driving of the circumstellar envelope through stellar pulsations, the formation of molecules and dust in the hydrodynamic environment of the stellar surface as well as the interaction of radiation with all involved constituents, and the transfer of moment between dust and gas (Willson 2000). Recent self-consistent modeling efforts are quite promising (e.g., Wachter et al. 2002; Mattsson et al. 2010). Observational determination of mass loss rates is indirect and therefore difficult (Jura and Kleinmann 1989, 1992; Groenewegen and de Jong 1998; van Loon et al. 1999; Olofsson et al. 2002). AGB stars will start the thermal-pulse phase with an O-rich surface composition (M star) left over from the initial composition. During the advanced thermal pulses, due to the dredge-up of primary C from the He-shell, lower mass AGB stars, like the $3 M_{\odot}$ case, star will eventually become C-rich (see below and ► Fig. 8-12). The lower-mass M stars have to be distinguished from the much more luminous massive AGB stars that are O-rich because of hot-bottom burning (► Sect. 4). Larger and more homogeneous observational samples are now obtained, and the emerging picture (Olofsson 2003; van Loon 2005; De Beck et al. 2010), in very broad terms, is that the low-luminosity O-rich stars (and S stars which have a C/O ratio around unity) have mass loss rates in the range $-7.5 < \log \dot{M}/M_{\odot} \lesssim -6$, whereas the mass loss rate of C-stars are systematically higher, $-6.2 \lesssim \log \dot{M}/M_{\odot} \lesssim -4.8$. The massive O-rich AGB stars have on average even higher mass loss in the range $-5.5 \lesssim \log \dot{M}/M_{\odot} \lesssim -4.2$.

For the $3 M_{\odot}$ stellar model, it is therefore expected that only a small fraction of the mass will be lost in the O-rich phase, while the C/O ratio gradually increases with each third dredge-up event (► Fig. 8-12) until the higher mass loss rate for C-rich stars kicks in. With an assumed average mass loss rate of $\log \dot{M}/M_{\odot} \sim -5.5$ out of the range for C-rich stars, the envelope mass of $M_{\text{env}} \sim 2.3 M_{\odot}$ of the $M_{\text{ZAMS}} = 3 M_{\odot}$ models star will be lost in $\sim 7 \times 10^5$ yr. With the interpulse time t_{ip} from ► Fig. 8-10, one can expect of the order of one dozen thermal pulses in the C-rich phase before all the envelope mass is lost, and the star will evolve off the AGB. In fact, the 12th thermal pulse shown in ► Fig. 8-10 corresponds to the thermal pulse after which the envelope composition becomes C-rich, and therefore, a $3 M_{\odot}$ star will have of the order 20–25 thermal pulses. The situation would be very similar for a $2 M_{\odot}$ case.

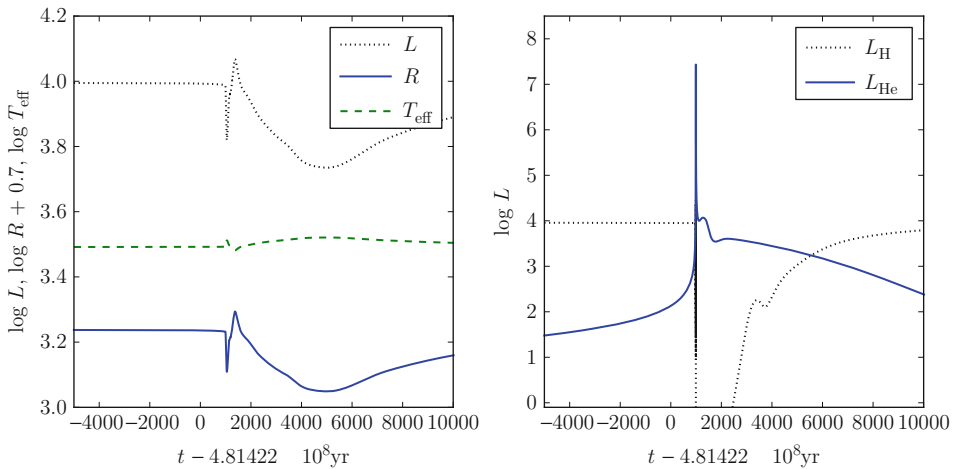
Although observational and theoretical investigations of mass loss of AGB stars are making swift progress, the differences in adopted mass loss rates in stellar evolution simulations of different groups are still one of the major sources of uncertainty. Mass loss rate prescriptions used in stellar evolution simulations, if they are not entirely based on simulation based data, should take advantage of the well-established correlation of mass loss and pulsation period, as pointed out by Vassiliadis and Wood (1993). These authors, and approximately at the same time Blöcker (1995a), presented a large set of detailed AGB thermal-pulse stellar evolution calculations, followed all the way from the main sequence, through a large number of thermal pulses, and including mass loss. Such stellar evolution grid sets have been frequently updated now, with improved microphysics and more elaborate treatment of nucleosynthesis

(e.g., Forestini and Charbonnel 1997; Weiss and Ferguson 2009; Karakas 2010) and a lot of work that has specifically focused on lower metallicity, see Sect. 5.

In one of the previous paragraphs, the *primary* production of C was mentioned. In this context, a species is produced in a primary mode if it can be made only out of H and He (most of which in the solar abundance mix is still from the Big Bang) or, in other words, if the production does not directly depend on the presence of metals (anything not H or He) from a previous generation of stars. *s*-process elements depend on the presence of Fe from previous star generations as a seed and are therefore a secondary production, even if the neutron source may be primary (see below).

He-Shell Flash Mixing and Third Dredge-Up

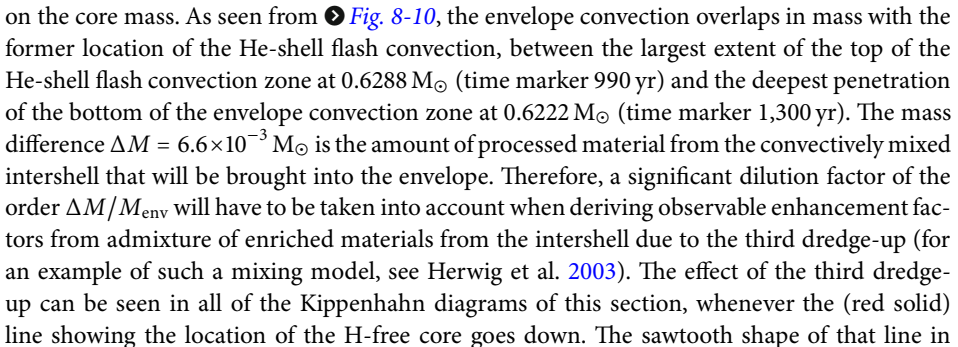
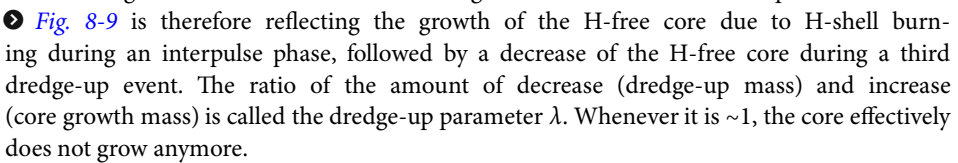
He-shell flashes follow a sequence of events that is demonstrated in Fig. 8-10 where the evolution of convection zones is shown, and Fig. 8-11 shows the evolution of the H- and He-burning luminosity and the surface parameters. The He-shell flash starts during the preceding quiescent intershell phase when the stellar luminosity is almost entirely provided by the H shell. He is accreted onto the CO core, and eventually, the He-burning luminosity starts to increase. This leads via the instabilities described above into a thermonuclear runaway, and the He-burning luminosity rapidly increases to 10^7 – $10^8 L_{\odot}$ (Fig. 8-11). Due to this high local luminosity, the region just above the shell becomes convectively unstable (8.3). The mass extent of the He-shell flash convection zone grows both inward and outward until the entire intershell region is convectively unstable and therefore well mixed (left panel, Fig. 8-10). Most of the energy will go into expansion work of the intershell layers and immediately above. The layer in which the H-shell is located will be lifted and cools, and the temperature decreases below the limit for H-burning. The H-burning luminosity drops to insignificant values (Fig. 8-11). Not only the H-shell will extinguish as a result of the lifting and cooling of those outer layers but also the opacity will increase. Then, again, the radiative gradient will increase making the

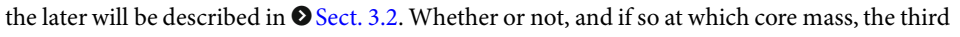



■ Fig. 8-11

Left panel: Surface parameters R , L , and T_{eff} before, during, and after the 12th He-shell flash of the $3 M_{\odot}$, $Z = 0.02$ stellar model track; **right panel:** luminosities of H burning and He burning. The time axis is the same as in Fig. 8-10

layer convectively unstable, and the mass coordinate of the bottom of the convective envelope will decrease (at $\sim -0.625 M_{\odot}$ and time mark $\sim 1,200$ yr). This is the third dredge-up.

Both the formation and evolution of the He-shell flash convection and the dredge-up episode afterward proceed on the local thermal time scale of 150–200 yr, which will depend on the core mass. As seen from  Fig. 8-10, the envelope convection overlaps in mass with the former location of the He-shell flash convection, between the largest extent of the top of the He-shell flash convection zone at $0.6288 M_{\odot}$ (time marker 990 yr) and the deepest penetration of the bottom of the envelope convection zone at $0.6222 M_{\odot}$ (time marker 1,300 yr). The mass difference $\Delta M = 6.6 \times 10^{-3} M_{\odot}$ is the amount of processed material from the convectively mixed intershell that will be brought into the envelope. Therefore, a significant dilution factor of the order $\Delta M / M_{\text{env}}$ will have to be taken into account when deriving observable enhancement factors from admixture of enriched materials from the intershell due to the third dredge-up (for an example of such a mixing model, see Herwig et al. 2003). The effect of the third dredge-up can be seen in all of the Kippenhahn diagrams of this section, whenever the (red solid) line showing the location of the H-free core goes down. The sawtooth shape of that line in  Fig. 8-9 is therefore reflecting the growth of the H-free core due to H-shell burning during an interpulse phase, followed by a decrease of the H-free core during a third dredge-up event. The ratio of the amount of decrease (dredge-up mass) and increase (core growth mass) is called the dredge-up parameter λ . Whenever it is ~ 1 , the core effectively does not grow anymore.

The third dredge-up will bring to the envelope all materials that are found in the He-shell flash convection zone. These are first of all primary ^{12}C from the triple- α He burning, possibly some ^{16}O depending on the mixing assumptions at the bottom of the He-shell flash convection zone (Herwig 2000), as well as any *s*-process elements. The nucleosynthesis and formation of the later will be described in  Sect. 3.2. Whether or not, and if so at which core mass, the third dredge-up would occur was debated, since early models (e.g., Iben 1981; Vassiliadis and Wood 1993; Blöcker 1995a) did not obtain this dredge-up mixing in simulations. However, it became clear that in order to reproduce this obviously very important feature of thermal-pulse AGB calculations, some kind of nonstandard convective boundary mixing across the Schwarzschild condition (which sets the convective boundary where the adiabatic and radiative gradient are equal) was needed (Lattanzio 1986; Hollowell and Iben 1988; Lattanzio 1989; Frost and Lattanzio 1996; Herwig et al. 1997; Mowlavi 1999; Herwig 2000). This issue is now settled, and all contemporary simulations obtain amounts of third dredge-up that are broadly consistent with observational constraints.

Such constraints come from a combination of observations. In particular, the synthetic models by Marigo et al. (1996, and subsequent work) that initially incorporated the third dredge-up as a free parameter have been tremendously useful to establish when and how efficiently the third dredge-up has to operate. Constraints include the luminosity function of C stars (Marigo et al. 1999), the star counts of O- and C-rich stars in clusters or extragalactic populations, the observed abundance of AGB stars (e.g., Fig. 6 in Herwig 2005), planetary nebulae and post-AGB stars, as well as additional constraints from nucleosynthesis discussed in  Sect. 3.2.

The efficiency of the third dredge-up will also depend on nuclear physics input. For example, the downward revision of the $^{14}\text{N}(p, \gamma)^{15}\text{O}$ rate leads to weaker H-shell burning, larger peak He-shell flash luminosities, and more efficient third dredge-up (Herwig and Austin 2004; Herwig et al. 2006a).

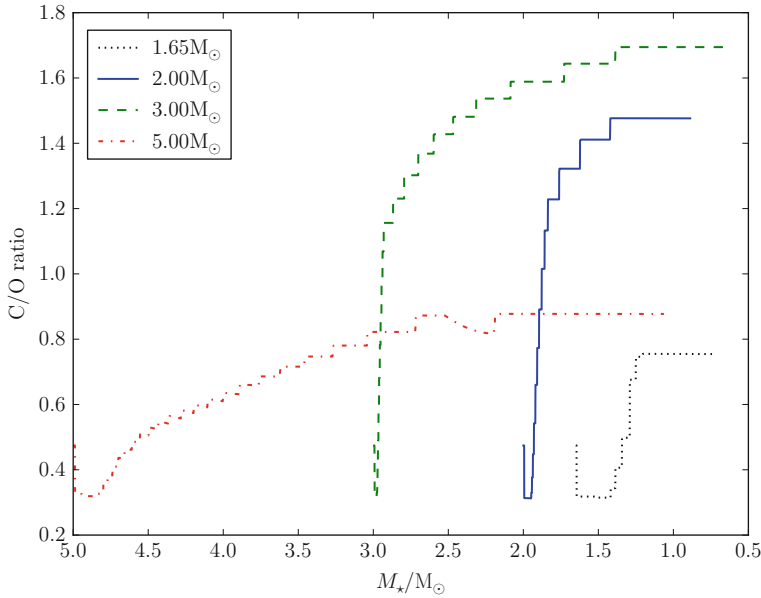
Nonstandard mixing during and at the end of the third dredge-up is also required for the formation of a layer enriched in ^{13}C which serves as the neutron source for the s process, to be further discussed in [Sect. 3.2](#). It has been proposed that convective boundary mixing needs to be taken into account as well at other convective boundaries, as, for example, at the bottom of the He-shell flash convection zone (Herwig et al. 1997; Herwig 2000; Miller Bertolami et al. 2006; Weiss and Ferguson 2009). Such mixing would have wide ranging implications for the thermal-pulse strength, the dredge-up efficiency, the s process in the He-shell flash convection zone, and the subsequent radiative ^{13}C -driven s process. Such mixing leads to larger ^{12}C and ^{16}O abundances in the intershell, in agreement with observations of H-deficient post-AGB stars (Werner and Herwig 2006). If correct, this would imply that AGB stars would dredge-up primary ^{16}O and, possibly limited to lower metallicity environments, may be a contributor of ^{16}O to the galactic chemical evolution. This type of mixing may also imply larger peak temperatures for the ^{22}Ne -fuelled s -process contribution. Lugaro et al. (2003) have argued that this may lead to isotopic ratio predictions, for example, for Zr, that could be incompatible with observations from presolar grains. This question needs to be revisited, considering possible modifications to the details of the mixing algorithm as well as the latest nuclear reaction rate data, in particular for the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron source reaction. Eventually, highly resolved hydrodynamic simulations of the He-shell flash convection may settle this question (for early attempts, see Herwig et al. 2006b, 2007).

Returning to [Fig. 8-11](#), eventually the thermonuclear runaway will be quenched, and after the peak luminosity has passed, the He shell will settle for a while into a stable burning mode. The star is for a couple 1,000 yr a He-burner, and He burning takes place radiatively. While during the thermal pulse He burning is dominated by the triple- α process, the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction dominates the radiative He burning (cf. $m_r \sim 0.6 M_\odot$ in [Fig. 8-14](#) where the ^{16}O abundance exceeds the ^{12}C abundance). After the He-shell flash, the layers above the H shell, the intershell region, and the region of the H-free core contract again, and the temperature rises. As much as H-shell burning is rejuvenated, the He burning decreases. Finally, about 8,000–10,000 yr into the new interpulse period the AGB star is back in the H-shell-supported single-shell burning mode, and the cycle starts over again.

Remarkably, despite the enormous peak He-shell flash luminosity released into the stellar interior during a thermal pulse, the effect on the surface parameters is not too dramatic ([Fig. 8-11](#)). Newer models with third dredge-up show the radius peak triggered by the thermal pulse. This may be relevant for binary evolution considerations. A radius peak may trigger a common envelope evolution to occur preferentially at thermal-pulse cycle phase zero, that is, immediately following the thermal pulse.

Molecules, Dust, and Grains

One of the most important results of the third dredge-up is the enrichment of the envelope and surface with the nuclear products of nuclear shell burning. The evolution of the C/O ratio ([Fig. 8-12](#)) is important for the chemistry of molecular opacities at the surface, which in turn effects the temperature and the mass loss conditions discussed above. The chemistry at the surface will also determine the type of dust that can form. Molecular and dust chemistry are fundamentally different in C- and O-rich environments. The reason is the high dissociation energy (11.1 eV) of carbon monoxide. Thus, coming out of the star, the gas will first form CO until all of either O (in a C-rich environment) or C (if $\text{C/O} < 1$) is consumed (Gustafsson and Höfner 2003). On the surface of C-stars, only molecules without O can form, for example, CN, which



■ Fig. 8-12

Evolution of the C/O ratio (by number) as a function of stellar mass. Since the stellar mass decreases with time the C/O evolution corresponds to a time, sequence from left to right

has been identified as a particularly important source of opacity (see above and ▶ Sect. 2). This dichotomy in chemistry between first the O-rich and later the C-rich thermal-pulse evolution is responsible for the difference in mass loss efficiency discussed above.

The evolution of the C/O ratio by number is shown in ▶ Fig. 8-12. The 3 M_\odot case shows that the C/O ratio is increasing initially almost at constant stellar mass, reflecting the fact that during the O-rich phase of the evolution, the mass loss is smaller than in the following C-rich phase. According to the assumptions made for this calculation, the mass starts to decrease significantly only after a certain C excess has been reached. Still, further thermal pulses trigger dredge-up events that further increases the C/O ratio. However, since not too many thermal pulses can occur once the star is C-rich (because of the higher mass loss, see discussion above), there is naturally a limit on the maximum C/O ratio that can be obtained. (This limit also comes from the C/O ratio in the He-shell flash convection, which in turn depends on the boundary mixing assumptions at the bottom of the He-shell flash convection zone; see discussion in previous section on He-shell flash mixing.)

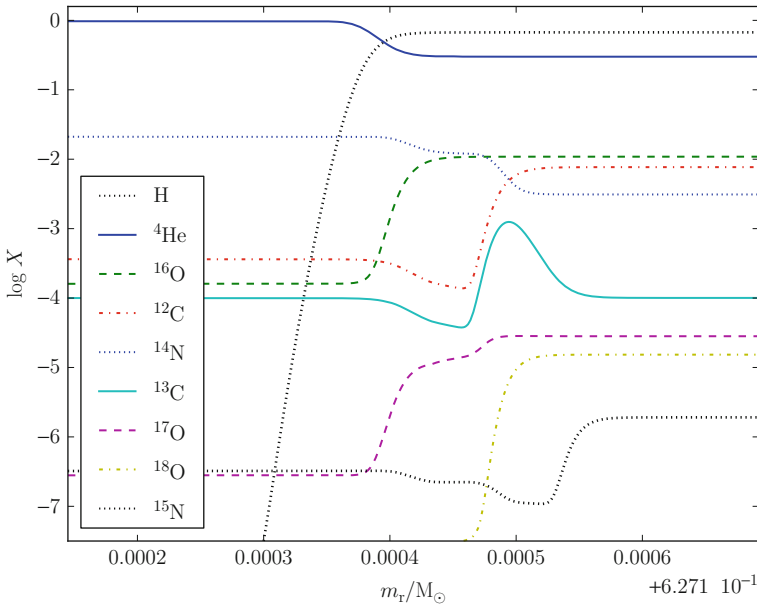
The transformation into a C-rich, high-mass loss star has another intriguing consequence, the formation of C-rich dust grains, the SiC grains. These are among the most abundant presolar grains that are extracted from meteorites (Zinner 1998). Presolar grains have formed in the cool and dense outflows of individual stars. They survived the condition in the interstellar medium until they were included unhomogenized into meteoritic material as part of the solar system formation. These grains are analyzed with great precision in the laboratory. They carry the signature of the particular nucleosynthetic processes inside individual AGB stars (or other sources,

depending on where they formed). This still young field is covered in the monograph *Stardust from meteorites. An introduction to presolar grains* by Lugaro (2005). It provides fascinating constraints on the nucleosynthesis processes in AGB stars, which will be covered next.

3.2.2 Nucleosynthesis

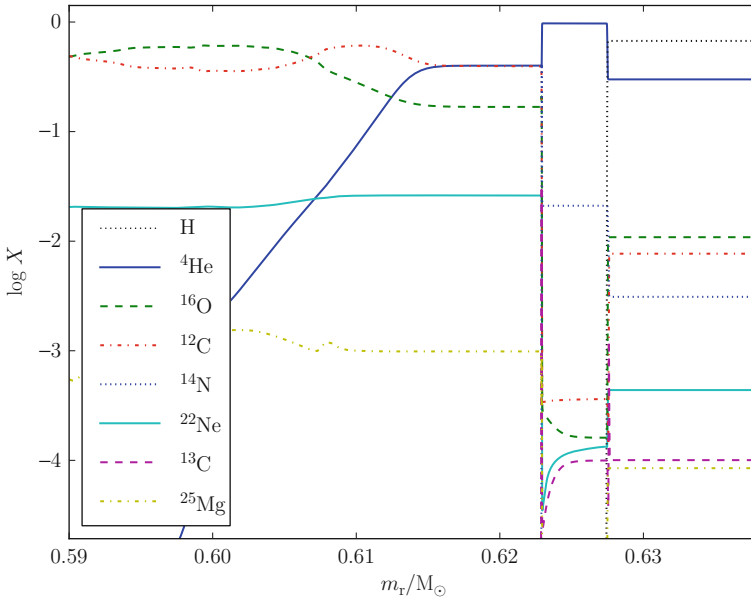
The events through the He-shell flash are described here once more, starting again with the quiescent H-shell burning before the 12th thermal pulse (● Fig. 8-10), however, now focusing on the nucleosynthesis. As the H shell burns through unprocessed material, it leaves behind its burning products (● Fig. 8-13). These are of course He, to be accreted by the He shell, but also, like during H-core and H-shell burning during the pre-AGB phase, most of the C and O isotopes will be transformed into ^{14}N (see discussion in ● Sect. 3.1). While before only the C and O from the initial composition could be transformed to ^{14}N , now also the primary C (and possibly O, see discussion in section He-flash mixing) from previous He-shell flashes, that was mixed into the envelope due to the third dredge-up, can be turned into ^{14}N . As discussed below, this is important for the s process.

The profiles over a larger mass range are shown in ● Fig. 8-14. The left side of the figure shows the upper parts of the CO core. The He shell is located around $0.605\text{--}0.615\text{ }M_{\odot}$. The first flat profile between 0.615 and $0.622\text{ }M_{\odot}$ contains the layer formally well mixed by the previous He-shell flash convection zone. The He-rich layer above, ($0.623\text{--}0.627\text{ }M_{\odot}$) contains the ashes



■ Fig. 8-13

Profiles of H, ^4He and the CNO isotopes in the H-burning shell during the interpulse period at time $t = 41,011\text{ yr}$ on the time scale of ● Fig. 8-10. The order of species in the legend corresponds to the abundance of the species at the right edge of the diagram, from top to bottom




■ Fig. 8-14

Profiles of key species from the CO core through the intershell region and the H-shell burning ashes and to the bottom of the convective envelope. As in ● Fig. 8-13, the point in time of this profile is $t = 41,011$ yr on the time scale of ● Fig. 8-10. The mass range is the same as in the Kippenhahn diagram ● Fig. 8-10. The order of species in the legend corresponds to the abundance of the species at the right edge of the diagram, from top to bottom

of the H-shell burning during the present interpulse phase, and above the abundance profile of the bottom layers of the convective envelope can be seen. The variation of the $^{12}\text{C}/^{16}\text{O}$ ratio has already been pointed out, and it is a signature of the T dependence of the ratio of the triple- α to $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ rate.

Neutron Sources for the s Process

Another important feature of this diagram (● Fig. 8-14) is the significant enhancement of the ^{22}Ne and ^{25}Mg abundance in the intershell region. Because the thermal pulses discussed here have a dredge-up parameter $\lambda \sim 1$ (see above), the abundance profiles after the following thermal pulse would look almost identical. The high ^{22}Ne and ^{25}Mg abundance are the result of the nucleosynthesis in the He-shell flash convection zone. The large amount of primary ^{14}N in the H-burning ashes is completely mixed into the next He-shell flash convection zone (cf. ● Fig. 8-10) where it is burned into ^{22}Ne by the same chain of two α capture already discussed in ● Sect. 3.1. In stars as massive as this $3 M_{\odot}$ case, the peak temperature at the bottom of the He-shell flash convection zone is high enough to activate the reaction $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$. This reaction is one of the two main neutron source reactions, and it leads to s-process nucleosynthesis in the He-shell flash convection zone that will be discussed below. The ^{25}Mg is only a tracer of that reaction, and the elevated abundance in the intershell indicates that neutrons have been released here during previous cycles.




But there is another and even more important neutron source in the thermal-pulse AGB stars, as discovered by Straniero et al. (1995). This neutron source is the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ reaction. However, in standard stellar evolution models, there is no appreciable source of ^{13}C , unlike the situation for ^{22}Ne which, as already mentioned, forms quite naturally even in a primary manner out of the ^{14}N from the ashes of the H shell that burns through third dredge-up C and O enriched envelope material. Therefore, it is assumed that additional mixing processes, beyond what the mixing-length theory and the Schwarzschild condition account for, must be at work. If the location of these additional mixing processes is the bottom boundary of the convective envelope during the third dredge-up (time marker 1,300 yr, $m_r = 0.622 M_\odot$,  Fig. 8-10, left panel), and if it mixes a sufficient amount of H from the envelope into the ^{12}C -rich intershell, then everything else will fall in place. The necessity to have some mixing at that convective boundary was already mentioned in the context of obtaining the third dredge-up (see above), and here such mixing is also needed.

The ^{13}C Pocket

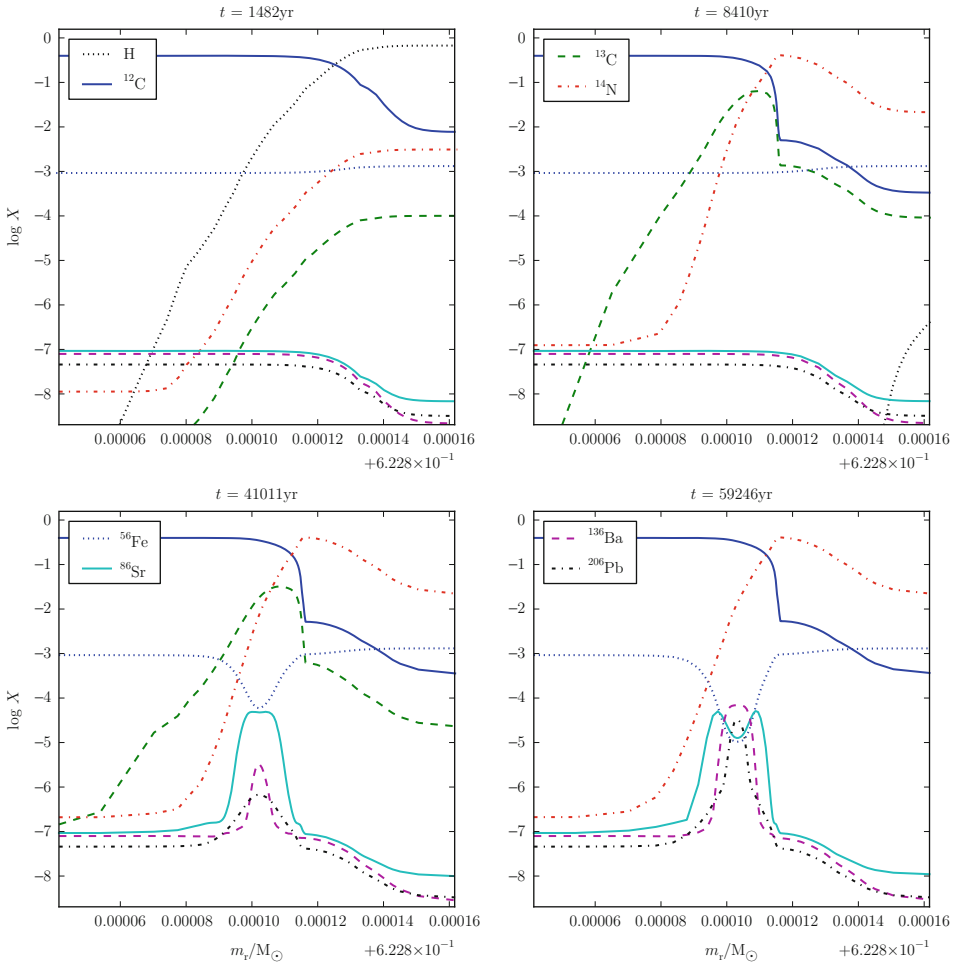
The models discussed here include a simple model for convective boundary mixing that was first introduced in stellar evolution models by Herwig et al. (1997). After the diffusion coefficient D is calculated inside the convectively unstable zone from the mixing-length theory, the diffusion coefficient for mixing is extended into the formally stable layer by an exponentially decreasing function of the distance from the formal convective boundary:

$$D_{\text{OV}} = D_0 \exp\left(-\frac{2|r - r_0|}{f H_P}\right) \quad (8.8)$$

where H_P is the pressure scale height and D_0 is the mixing-length diffusion coefficient, both taken at the radius r_0 close to the boundary. In this model, f is a free parameter that has to be calibrated for each type of convective boundary either semiempirically through observations or through hydrodynamic simulations. Lugaro et al. (2003) have used this framework with $f = 0.128$ to generate a mixing zone that is of the required order of magnitude size to reproduce observed overabundance levels of s -process enriched AGB stars (see also Herwig et al. 2003). A more in-depth discussion of the partial-mixing zone has been given by Straniero et al. (2009) with applications in Cristallo et al. (2011).

The formation of the partial-mixing zone during the dredge-up due to the mixing model ( 8.8), and the subsequent nucleosynthesis in the ^{13}C -pocket is illustrated in  Fig. 8-15. The top-left panel shows the formation of the partially mixed layer across the bottom boundary of the envelope convection zone shortly after that boundary has reached the smallest mass coordinate at the end of the third dredge-up phase. The mixing model forces protons below the Schwarzschild boundary, and as a result, the abundances are transitioning smoothly at the envelope-core interface over a mass region with width $\Delta M_{\text{PM}} \sim 2 - 3 \times 10^{-5} M_\odot$ (which is probably a bit smaller than what is required to reproduce observed s -process overabundances quantitatively). The width of this partial-mixing layer, of course, is the direct consequence of the choice of f in  8.8.

Once contraction resumes after the He-shell flash (see discussion above), the temperature in the partial-mixing zone gradually increases (Fig. 6, Herwig et al. 1999) and H burning resumes by creating a pocket of ^{13}C and ^{14}N (upper-right panel, note the tail of the H profile at the very right end of the plot). Then, for a long time, nothing happens, until the temperature reaches $T \sim 10^9$ K to activate the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ reaction. Peak neutron densities around $N_n \sim 10^7 \text{ cm}^{-3}$



■ Fig. 8-15

Profiles of key species showing the formation and evolution of the ^{13}C -pocket during the interpulse phase, the primary site of the main s process. The times of the four frames correspond to the time scale of [Fig. 8-10](#). The mass range is zoomed in around the envelope–core interface at the end of the third dredge-up (*upper-left panel*) when a simple mixing model generates a partial-mixing zone containing both H and ^{12}C . A few thousand years later, ^{13}C has formed (*top-right panel*), and again, a few ten thousand years later, neutrons are released, and copious amounts of trans-iron elements form (*bottom-left*), leading eventually to the formation of Pb (*bottom-right panel*). Fe is the seed for the s process

are maintained for several 1,000 up to 10^4 yr, and the neutron exposure is $\tau \sim 0.2 - 0.4 \text{ mbarn}^{-1}$. In simulations with a partial-mixing model the peak neutron exposure is proportional to the intershell ^{12}C abundance, which in turn is proportional to the mixing efficiency at the convective

bottom boundary of the He-shell flash convection zone (Herwig et al. 1999; Lugaro et al. 2003; Herwig 2005). The neutrons are captured by ^{56}Fe and trans-iron elements form.

The s -process abundance distribution is characterized by three distinct peaks corresponding to the low n -capture cross sections at the neutron magic-number closed shells at $n = 50, 82,$ and 126 . In the lower-left panel, ^{86}Sr , representing the first s -process peak, increases first as the nuclear production through continued neutron captures on seed nuclei arrives at the $N = 50$ barrier. The second peak elements, represented by ^{136}Ba are at this point enhanced much less, and the third peak species, represented by ^{206}Pb are only starting to be produced. The s -process seed, ^{56}Fe is significantly depleted in the ^{13}C pocket.

The final panel (bottom right) shows the situation immediately before the nuclear products will be mixed into the next He-shell flash convection zone. Now, the ^{86}Sr profile itself has a little dip in its peak, because it has served as a seed for higher mass species. The network flux has reached now Pb, and species of all three s -process peaks are enhanced by approximately the same amount, that is, almost 3 dex compared to the already enriched intershell abundance of trans-iron elements as a result of previous thermal-pulse cycles and $\gtrsim 4$ dex compared to the envelope abundance of the s -process elements (at the first thermal pulse, the intershell abundance would have a level less than the present envelope abundance). This is the result of the radiative s process driven by the ^{13}C pocket.


When mixed into the He-shell flash convection zone, the content of the ^{13}C pocket will first be diluted with the ratio of the mass in the partial-mixing layer to the mass in the He-shell flash convection zone $\Delta M_{\text{PM}}/\Delta M_{\text{HeCZ}} = 2.1 \times 10^{-5}/0.021 = 1/1,000$. Then, the freshly produced heavy elements, mixed with the heavy elements already present from previous thermal pulses will be exposed to the ^{22}Ne neutron source, the origin of which was described above. Although the He-shell flash convection lasts for about 150 yr (see discussion above), the temperature is only for a short fraction high enough ($\log T > 2.7$) to activate the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron source. Again, depending on the convective boundary mixing assumptions, and on the core mass, the duration of the high- T phase varies from less than 1 yr to a few years, and the peak neutron density reaches $N_n \sim 10^9 - 10^{11} \text{ cm}^{-3}$ which implies neutron exposures in the range $\tau \sim 0.01 - 0.90 \text{ mbarn}^{-1}$. Presently, models with larger convective boundary mixing predict higher peak temperatures, longer high- T periods, and consequently higher neutron densities. Lugaro et al. (2003) have pointed out that such models may be incompatible with measurements of isotopic ratios, for example, the $^{96}\text{Zr}/^{94}\text{Zr}$ ratio, in presolar grains, which are sensitive to s -process branchings, in this case at ^{95}Zr . Clearly, these issues need further investigations, but the interconnections between the details of the s process with aspects of the dynamical star (► Sect. 5) are intriguing.


After the end of the exposure in the He-shell flash convection zone, the s -process enriched material will be mixed into the envelope, again with a large dilution factor $\Delta M/M_{\text{env}}$ (see discussion in section *He-shell Flash Mixing and Third Dredge-Up* ► Sect. 3.2.1).

The present generation of models, such as those illustrated here, or those by Cristallo et al. (2011), include the mixing processes into the stellar evolution code, and at least any feedback that such mixing may have back to the stellar structure would be taken into account. An example for such feedback can be encountered in simulations of stars with very-low metallicity, where the mixing of protons into the ^{12}C -rich intershell leads to a hot dredge-up (Goriely and Siess 2004; Herwig 2004) with poorly understood quantitative consequences for the nuclear yield predictions and final stellar evolution outcomes. In any case, the properties and realistic modeling of the mixing for the ^{13}C pocket remains an open problem.

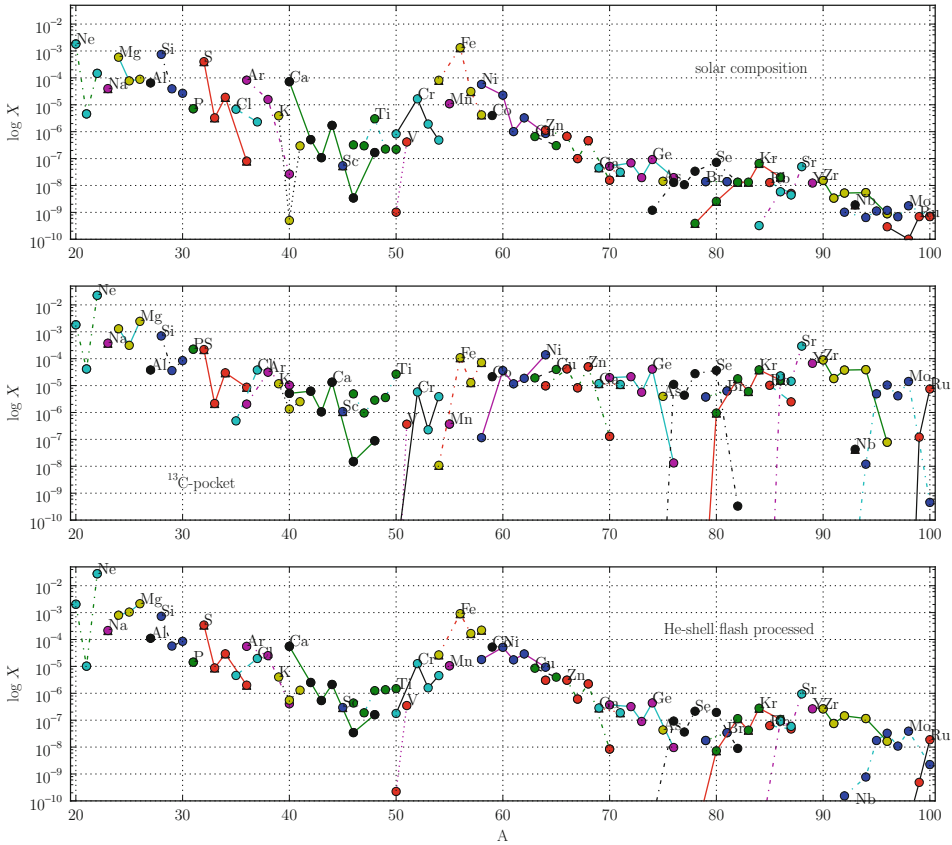
s-Process Nucleosynthesis

Much of what is known today about the *s* process quantitatively has been discovered through a parameterized post-processing method, in which the required amount of ^{13}C was directly inserted into the post-processing calculation. This approach was pioneered by Gallino et al. (1998), and followed by many others, including Busso et al. (2001), and Travaglio et al. (2004) with a focus on galactic chemical evolution, and Bisterzo et al. (2010) with an emphasis on low-metallicity stars. Käppeler et al. (1989) and Busso et al. (1999), and more recently Käppeler et al. (2011) provide a detailed discussion not only on simulation results but also on the nuclear physics involved (incl. discussions on *s*-process branchings) and on comparisons with and constraints from observations.

Here it is not the intention to give a comprehensive overview of the results covered in an extensive body of literature (see above for examples). Instead, some features of the workings of the *s* process are highlighted which may help to understand that body of literature better.  Figs. 8-16–8-18 provide abundance distributions for three atomic mass ranges (so that the entire distribution from Ne to Pb is covered). In each figure, the top panel is the initial solar-scaled abundance distribution with which the 3 M_{\odot} , $Z = 0.02$ model star was started. The middle panel represents the production of the ^{13}C pocket while the bottom panel shows the distribution after dilution and processing via the ^{22}Ne neutron source in the He-shell flash convection zone (cf. figure caption). Plotted are only stable isotopes, and the abundance distribution has not been decayed. Careful inspection of these plots reveals a number of interesting aspects about the different components of the heavy element nucleosynthesis, in fact more than can be discussed here. The abundances in these diagrams are the result of theoretical stellar simulations and of the nuclear physics network of isotopes and rates used. Therefore, limitations in our present understanding on stellar physics and nuclear physics may affect the accuracy of such predictions. For instance, final abundances at branching points located along the *s*-process path are very sensitive to the uncertainties of the nuclear rates involved. Furthermore, the treatment of ground and isomeric states of isotopes that are not fully thermalized at stellar conditions is still not clear (e.g., Kr85 is a branching point in the *s*-process path, and its ground and isomeric state are not thermalized at typical *s*-process conditions in low-mass AGB stars and should therefore be treated separately, e.g., Ward and Fowler 1980).

Just for the purpose of this discussion, the ^{13}C pocket frame will be referred to as CP, and the post-He shell flash distribution will be referred to as HF. Starting with , as discussed before, the ^{22}Ne and ^{25}Mg abundances are enhanced in HF, as well as ^{26}Mg due to the $^{22}\text{Ne}(\alpha, \gamma)$ reaction competing with the $^{22}\text{Ne}(\alpha, n)$ branch. Most species up to iron are insignificantly modified, keeping in mind the two large dilution factors that need to be applied before the nuclear processed material gets to the surface. In the solar system distribution, most of the *s*-process abundances between Fe and the first *s*-process peak around Sr, Y, and Zr are produced by the weak *s* process in massive stars, where a nonnegligible contribution is also supposed to come from AGB stars, in particular from massive AGB stars.

Skippping therefore right through to the first peak isotopes, ^{84}Sr is a *p*-process element which cannot be reached by neutron captures in the *s* process. It is therefore only destroyed in CP and HF. However, the other three Sr isotopes show in both panels the same relative ratios as in the solar distribution. The same is true for the ratios of ^{88}Sr , ^{89}Y , and ^{90}Zr , indicating their dominantly *s*-process origin. For Zr, the solar isotopic ratios for the mass numbers 90, 91, and 92 are well reproduced by CP. However, ^{96}Zr is produced ~ 2 dex less in CP compared to the solar-scaled distribution. The reason for that is the branching at ^{95}Zr where the β decay with a terrestrial half life of 64d competes with a neutron capture. In the ^{13}C pocket, the neutron

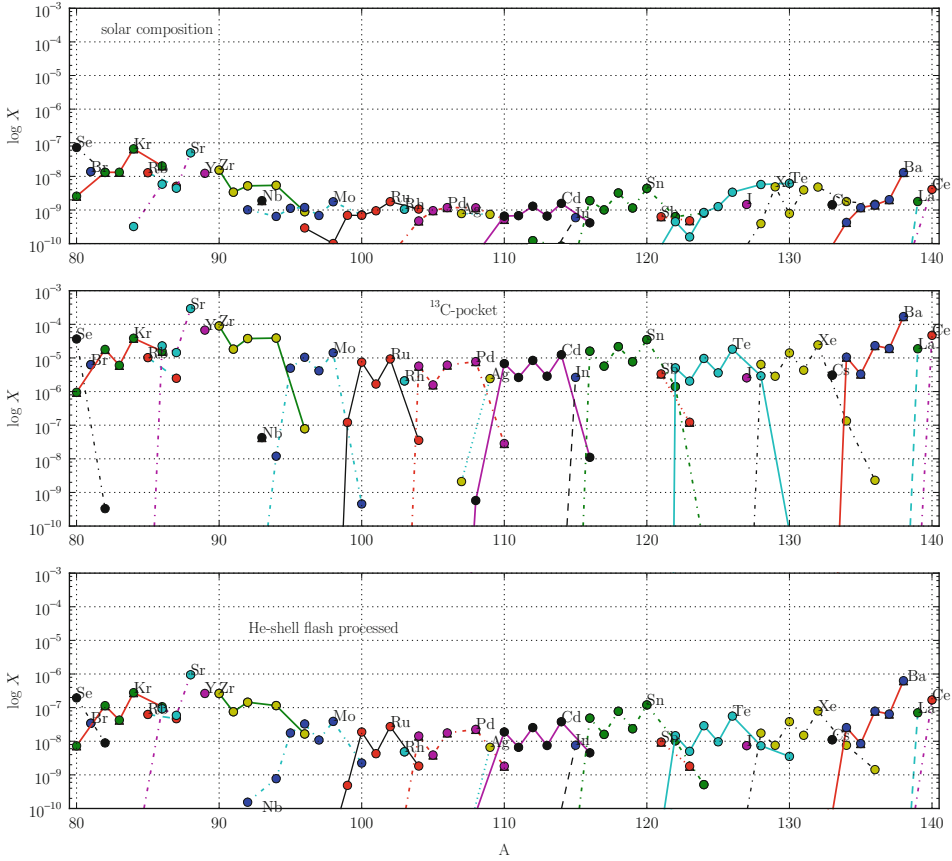


■ Fig. 8-16

Abundance distribution (mass fraction) for the solar composition (*top panel*) and in comparison the abundance distribution averaged over the ^{13}C pocket at the end of the interpulse (*bottom-right panel*, see Fig. 8-15), just before its nuclear production will be engulfed into the next He-shell flash convection zone (see Fig. 8-10, $t = 59,246$ yr), and finally averaged over the mass region of the He-shell flash convection zone, after the pulse is finished. The *bottom panel* therefore shows the abundance distribution in the intershell region, part of which will be mixed to the envelope during the immediately following third dredge-up. The atomic mass ranges from the light elements that are not significantly effected by the s process through Fe which is the seed of the s process to just past the first s -process peak around Sr, Y, and Zr

density is too low to open the branching and all ^{95}Zr will decay to ^{95}Mo . However, as the HF panel shows, in the high neutron density environment of the He-shell flash convection zone, at least in the $3 M_{\odot}$ case considered here, the branching is in fact open, and ^{96}Zr is produced. In fact, after processing in the HF, the ratios of all the Zr isotopes seem to reproduce the solar ratios.

Zr is an example that demonstrates how the bulk of the s -process production takes place in the ^{13}C pocket, while certain isotopic rearrangements are taken care of by the ^{22}Ne neutron source in the He-shell flash convection zone. The sensitivity of such branchings to the exact



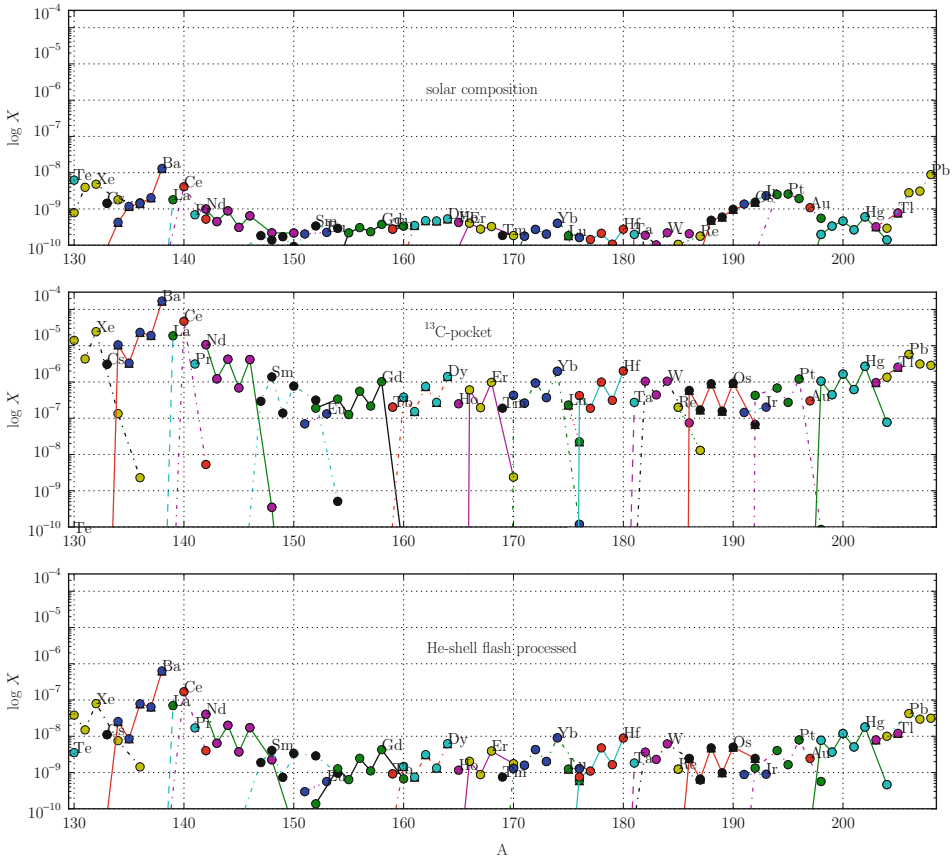
■ Fig. 8-17

Same as ► Fig. 8-16, for the atomic mass range from the first s -process peak to past the second s -process peak around Ba and La

details of the thermodynamic and hydrodynamic conditions in the convection zone may provide eventually a powerful diagnostic tool for constraining mixing physics, given that other physics ingredients, such as the nuclear physics, are known to sufficient accuracy.

In this discussion, the main emphasis is on s -process signatures in the solar-scaled initial abundance distribution. However, only about half of all trans-iron elements are made in the s process (Arlandini et al. 1999). The other half comes from the r process, and it should be expected that there are numerous patterns that are not the result of the nucleosynthesis in $3 M_{\odot}$ stars. In addition, the solar abundance distribution is the superposition of many sources as a result of galactic chemical evolution, including s -process production from a range of initial masses.

The r process is taking place in some kind of explosive environment in which very large neutron densities ($N_n \gtrsim 10^{20} \text{ cm}^{-3}$) can be reached (Meyer 1994; Sneden et al. 2008). Sites that have been suggested for the r process include neutron-star mergers (Freiburghaus et al. 1999) or supernova fallback (Fryer et al. 2006). Simplified, in the r process the very large neutron



■ Fig. 8-18

Same as ► Fig. 8-16, for the atomic mass range from the second *s*-process peak to Pb

densities first drive the network flux out to the neutron drip line from where nuclei decay back to the valley of stability.

Moving on to ► Fig. 8-17, ^{101}Ru and ^{104}Ru are not produced either in CP or sufficiently in HF. Both of these are predominantly produced by the *r* process. This is an example of how a pronounced odd-even effect from the *s* process in one stellar model will be complemented by other sources and processes. Such aspects of galactic chemical evolution have been addressed, for example, by Timmes et al. (1995) and Travaglio et al. (2004).

Sn is another example that seems to be well reproduced by the *s*-process prediction from mass number 116 to 121. Again, ^{112}Sn is a *p*-process isotope, while ^{124}Sn would come from the *r* process. Similar arguments would indicate that ^{122}Te , ^{123}Te , and ^{124}Te are primarily produced by *s* process. In fact, these are *s*-only isotopes which are shielded by stable species on the same isobar on the neutron-rich side. Therefore, they cannot be produced by the *r* process. However, this is not the case for the more massive Te isotopes, and their production is dominated by the *r* process. It is interesting to note though that some although insignificant amount of ^{130}Te is produced in HF, indicating again some branching activity. Also, for Xe and Ba, the ratio of

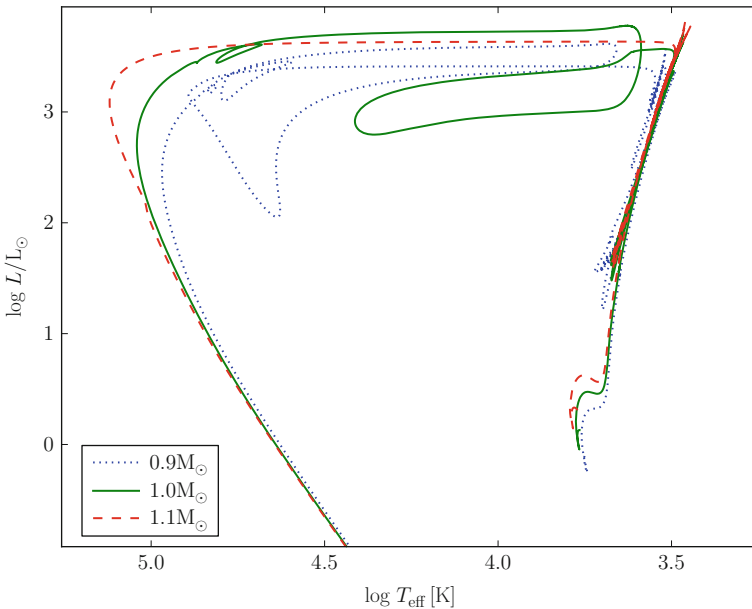
the s -only isotopes in the solar abundance distribution is well reproduced in CP and HF. For completeness, the corresponding plot for the mass range between the second and the third peak is provided (● Fig. 8-18). It reveals upon inspection, aided by a chart of nuclei (<http://www.nndc.bnl.gov/chart>) similar details about the production mode of each isotope.

3.2.3 Post-AGB Evolution

Once the AGB star has lost all but a small fraction of the original envelope, the star will evolve away from the giant solution, initially on a thermal time scale, at constant luminosity to high temperatures. The unprocessed envelope mass that is left on the surface depends on the core mass and is a result of stellar evolution simulations of this transition phase. The envelope mass is approximately $10^{-4} M_{\odot}$ for a core mass of $\sim 0.6 M_{\odot}$. For the massive cores, this will decrease by approximately an order of magnitude while the less massive He-white dwarfs will have envelope masses larger than $10^{-3} M_{\odot}$ (Schönberner 1981; Driebe et al. 1998).

The post-AGB evolution is shown for three stellar evolution tracks with masses around $1 M_{\odot}$ as part of an HRD diagram showing the entire evolution from main sequence to white dwarf, in ● Fig. 8-19.

The horizontal evolution from cool to hot at constant luminosity corresponds to the decrease of the envelope mass due to mass loss and due to nuclear burning. In fact, the temperature of the star along the constant luminosity cooling track is directly correlated with the envelope. The luminosity follows the core-mass luminosity relationship (● Sect. 3.2.1) and is



■ Fig. 8-19

HRD for stars with masses around $1 M_{\odot}$ and $Z = 0.01$, including the post-AGB and white dwarf evolution. The white dwarf masses are $M_f = 0.525, 0.543$, and $0.554 M_{\odot}$ for $M_{\text{ZAMS}} = 0.9, 1.0$, and $1.1 M_{\odot}$. The $0.9 M_{\odot}$ track arrives on the white dwarf cooling-down track after 1.47×10^{10} yr

therefore set. For a $1 M_{\odot}$ post-AGB star (which would have started its evolution with $M_{\text{ZAMS}} \sim 6 M_{\odot}$, see [Fig. 8-21](#)), one may expect a nuclear burning rate of $\dot{M}_{\text{H-shell}} \sim 4 \times 10^{-7} M_{\odot}/\text{yr}$ while a star with $0.65 M_{\odot}$ will have a lower burning rate of $\sim 1.4 \times 10^{-7} M_{\odot}/\text{yr}$.

Mass Loss

The time scale on which the post-AGB star evolves to the hottest point may depend therefore on the mass loss, if it is larger than $\dot{M}_{\text{H-burn}}$. Higher mass loss implies a shorter post-AGB evolution time scale (Schönberner 1981; Blöcker 1995b). AGB mass loss decreases steeply with increasing stellar temperature because dust formation is less efficient at higher temperature, and because the material at the surface is gravitationally more bound when the radius is smaller (and the escape velocity is higher). For temperatures above $\sim 5,000$ K, the mass loss according to formulae that are intended to capture the mass loss mechanism of AGB stars (Arndt et al. 1997; van Loon et al. 2005) falls below $\dot{M}_{\text{H-burn}}$. Hotter stars have radiation-driven wind mass loss $\dot{M}_{\text{rad-wind}}$ (Pauldrach et al. 1988, see also chapter on winds in this volume). For lower-mass post-AGB stars, such as the $0.65 M_{\odot}$ example given above, $\dot{M}_{\text{rad-wind}} < \dot{M}_{\text{H-burn}}$. Therefore, if the wind-driven mass loss theory is correct, the post-AGB evolution time will be uniquely determined due to the mass of the post-AGB star. For higher mass cases, such as the $1.0 M_{\odot}$ case, $\dot{M}_{\text{rad-wind}} \gtrsim \dot{M}_{\text{H-burn}}$, and in this case, the transition time from the AGB to the hot “knee” where the track turns around into the WD cooling track will depend on the mass loss as well.

Note that the correlation between temperature and envelope mass implies that H-deficient (or H-free) central stars of planetary nebulae of type PG 1159 or [WC] (Werner and Herwig 2006) or white dwarf of type DQ or DZ (see chapter on white dwarfs in this volume) cannot form due to enhanced mass loss that expells the remaining H-rich envelope. Such an enhanced mass loss would simply accelerate the post-AGB evolution, and the resulting envelope mass of the pre-WD remains the same. The envelope mass of white dwarfs can now be probed by asteroseismology.

Post-AGB He-Shell Flashes

Instead, the main mechanism that causes H deficiency in post-AGB stars and young white dwarfs is believed to be He-shell flashes on the post-AGB (Herwig 2001b; Werner and Herwig 2006). As discussed in [Sect. 3.2](#), He-shell flashes occur in regular intervals. The evolution from the AGB to the post-AGB and eventually to the WD cooling track is comparable to a fraction of the order 10% of the interpulse time, depending on mass. H-shell burning will cease when the envelope mass is too low to support H-burning temperatures at the mass coordinate where the H-free core begins. At that point, the star evolves off the horizontal post-AGB track and enters the white dwarf cooling track. This hottest feature of the post-AGB evolution is also referred to as the knee.

He burning takes place in deeper layers. While the star is evolving along the post-AGB track, the thermal-pulse clock continues to tick in the He-shell region. Clearly, if the post-AGB star has evolved onto the proper WD cooling track, no nuclear process can take place anymore. However, the AGB star statistically happens to leave the giant configuration at a time in the thermal-pulse cycle before the He-shell flash that is comparable to the evolution time to the WD cooling track. In that case, the next He-shell flash will take place during the post-AGB evolution (Fujimoto 1977; Schönberner 1981; Iben et al. 1983; Iben and MacDonald 1995; Herwig et al. 1999).

Both the $0.9 M_{\odot}$ and the $1.0 M_{\odot}$ tracks in [Fig. 8-19](#) show the characteristic loops associated with post-AGB He-shell flashes, which usually cause a temporary evolution back to the AGB regime in the HRD. For this reason, this phenomenon is also called *born-again* evolution. Several born-again stars, such as Sakurai’s object (e.g., Duerbeck and Benetti 1996; Hajduk et al.

2005), FG Sge (Blöcker and Schönberner 1997; Jeffery and Schönberner 2006), or V605 Aq (e.g., Clayton et al. 2007; Hinkle et al. 2008), are known.

Two cases need to be distinguished. If the He-shell flash occurs during the high-luminosity horizontal post-AGB evolution before the knee, the H-burning shell is still on. A high entropy barrier prevents the mixing of the H-rich envelope into the He-shell flash convection zone. However, the stellar surface will become H-poor later during the born-again phase due to the emerging envelope convection and mass loss. This case is also called the late thermal pulse (LTP). If the He-shell flash occurs later, after the star has evolved past the knee, the H-shell is extinct, and the He-shell flash convection zone will engulf the H-rich envelope. This is called a very-late thermal pulse (VLTP). An example of such a case is Sakurai's object (Herwig 2001a; Lawlor and MacDonald 2003). Sakurai's object has detailed surface abundance observations of trans-iron elements (Asplund et al. 1999) and can be used to guide and constrain models and simulations of the physics of H combustion, that is, the interaction of rapid nuclear energy release from the $^{12}\text{C}(p, \gamma)^{13}\text{N}$ reaction with hydrodynamic and turbulent convection flow (Herwig et al. 2011 and ► Sect. 5).

Very similar events are known from post-RGB evolution, where flash mixing of H from the remaining envelope into a post-RGB He-core flash has been associated with horizontal branch features in globular clusters (e.g., Brown et al. 2001).

Abundance Determinations from Post-AGB Stars

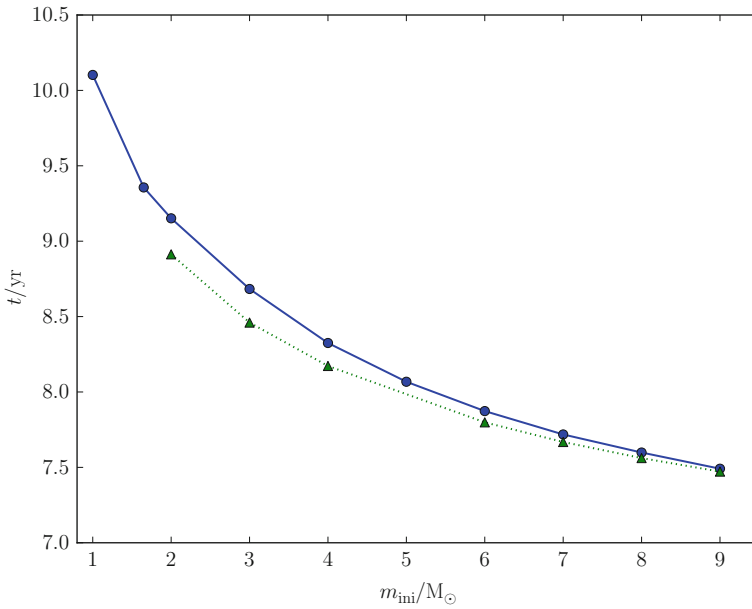
When post-AGB stars evolve along the high- L horizontal track, they evolve through a range of temperatures and can be spectroscopically observed (van Winckel 2003). As a result, detailed abundance information has been obtained, in particular of many heavy elements that have formed through the s process during the progenitor AGB evolution. The abundances of post-AGB stars represent the final enrichment of the envelope with s -process elements that have formed cumulatively over many thermal pulses. Any future generation of stellar evolution and nucleosynthesis simulations needs to be confronted with this type of data.

4 Evolution as a Function of Initial Mass

Most of the basic processes and concepts of solar and intermediate-mass stars were covered by the description of the evolution of the $3 M_{\odot}$ star. This chapter covers how the evolution depends on mass and introduces some processes, like hot-bottom burning that are unique to stars of other masses.

H- and He-Core Burning

Stellar evolution times depend significantly on the initial mass. Higher mass stars evolve faster than lower mass stars, as illustrated in ► Fig. 8-20. All evolutionary phases after He-core burning are short compared to H- and He-core burning. One of the greatest sources of uncertainty concerning stellar lifetimes remains the efficiency of core overshooting. The reproduction of the observed width of the main sequence requires some convective overshooting to extend the main sequence to higher luminosities (Schaller et al. 1992). Such overshooting will prolong H-core burning lifetimes. Rotation will also broaden the main sequence. At masses below $M_{\text{ZAMS}} \sim 1.25 M_{\odot}$, H-burning will be radiative. In the transition region between radiative cores



■ Fig. 8-20

Stellar age versus initial mass for $Z = 0.02$ (solid line with circles) and $Z = 6 \times 10^{-4}$ (dashed line with triangles)

and extended convective cores, the overshooting efficiency, which is often based on parameterizations in terms of the pressure scale height at the convection boundary, needs to be carefully ramped up in order to avoid artificially enlarged convective cores (VandenBerg et al. 2006).

Lower mass stars, like the sun, will have convective envelopes during the main-sequence evolution while more massive main-sequence stars have radiative envelopes. Overlaying opacities and profiles of stars with different masses in the $\rho - T$ diagram (such as in Fig. 3 in Paxton et al. 2011) demonstrate that convective stability or instability of envelope and core regions as a function mass can be understood in terms of the variation of opacities as a result of atomic physics in terms of ρ and T , and its effect on stability via the radiative gradient (► 8.3).

A major source of uncertainty during He-core burning is again mixing. Here, the complicated situation of semi-convection, that was introduced in ► Sect. 2, implies that He-core burning lifetimes and subsequently the core masses at the first thermal pulse for a given initial mass are still uncertain. This is particularly unfortunate because the thermal-pulse AGB evolution, as, for example, the onset and efficiency of the third dredge-up does depend sensitively on the core mass.

The Initial–Final Mass Relation

The initial–final mass relation relates the initial, main-sequence mass with the mass of the white dwarf that is the evolutionary end point of solar and intermediate-mass stars (Weidemann 2000). Observational constraints seem to indicate a significant spread in WD masses for a given initial mass (Catalán et al. 2008). Several processes could be responsible for that, including WD

mergers or other binary interactions, or differential mass loss, for example, as a result of stellar rotation.

The importance of this relation derives from the fact that it relates two other important relations: the initial mass function (Kroupa 2002) and the white dwarf mass distribution (e.g., Liebert et al. 2005). Ignoring the effect of ongoing star formation, the initial–final mass relation maps the initial mass function to the white dwarf mass distribution. An initial–final mass relationship that creates a consistent picture between the IMF and the white-dwarf mass distribution is a powerful constraint on evolution simulations of solar and intermediate-mass stars.

A theoretical initial–final mass relation is shown in **Fig. 8-21** for three metallicities. According to the initial mass function, most stars ($\geq 85\%$) have low-mass $\lesssim 2.5 M_{\odot}$. The white dwarf mass distribution is sharply peaked around $0.59 M_{\odot}$. That means that stars with initially around or below $2.5 M_{\odot}$ have to be projected by the MiMf relation to the WD mass distribution peak. Such considerations have been performed quantitatively by Ferrario et al. (2005). They find the best fit to the white dwarf mass distribution with an MiMf relation with double curvature shown in **Fig. 8-21** for comparison. The MiMf relation is in this way an important global

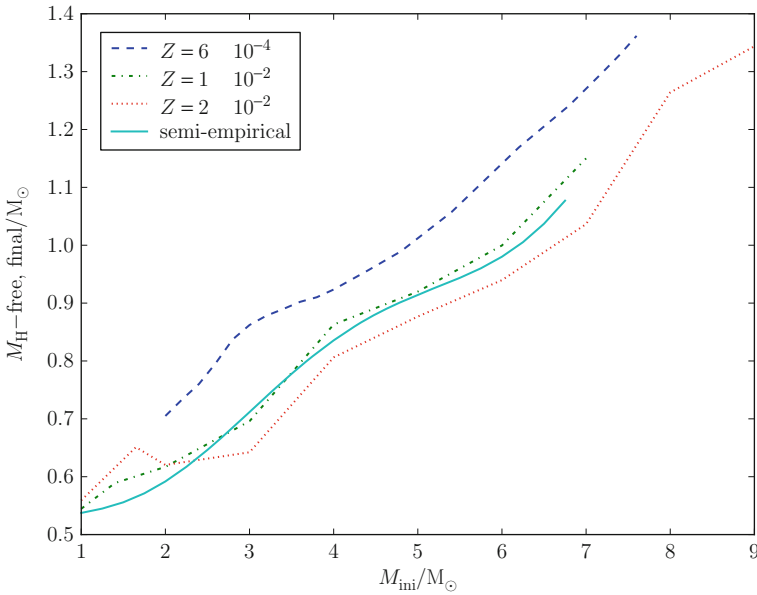


Fig. 8-21

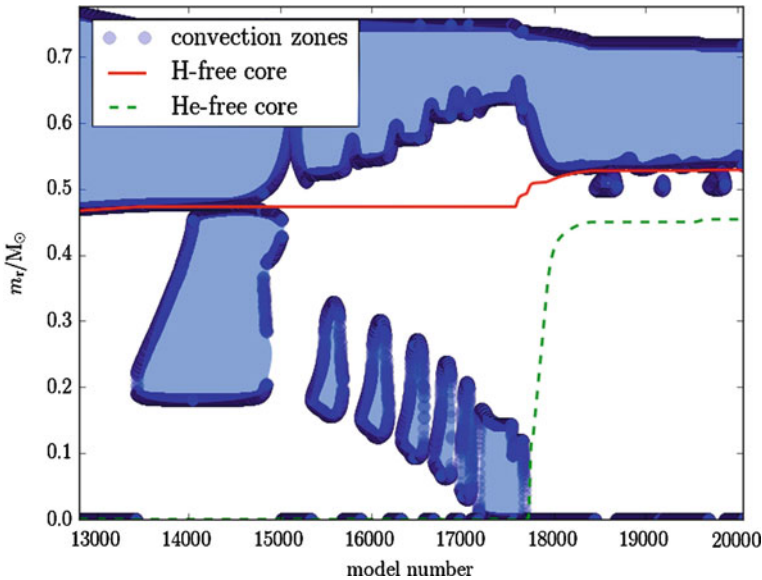
H-free core mass at the last computed thermal pulse as a function of initial mass. Each relation is based on a grid of sequences that have been computed to the tip of the AGB when all mass is lost or close enough to the point that very little change of the core mass is expected before the model stars evolve to become white dwarfs. Therefore, the shown relations constitute theoretical initial to final mass relations for the given metallicities. The line with the label *semiempirical* represents the curvature MiMf that Ferrario et al. (2005) found to best reproduce the white dwarf mass distribution from the Palomar-Green survey

constraint on the evolution models. The lower mass models cannot have a lot of core growth, with implications for mass loss and third dredge-up efficiency.

Low-Mass Stars with He-Core Flash and Those Which Become C-Rich AGB Stars

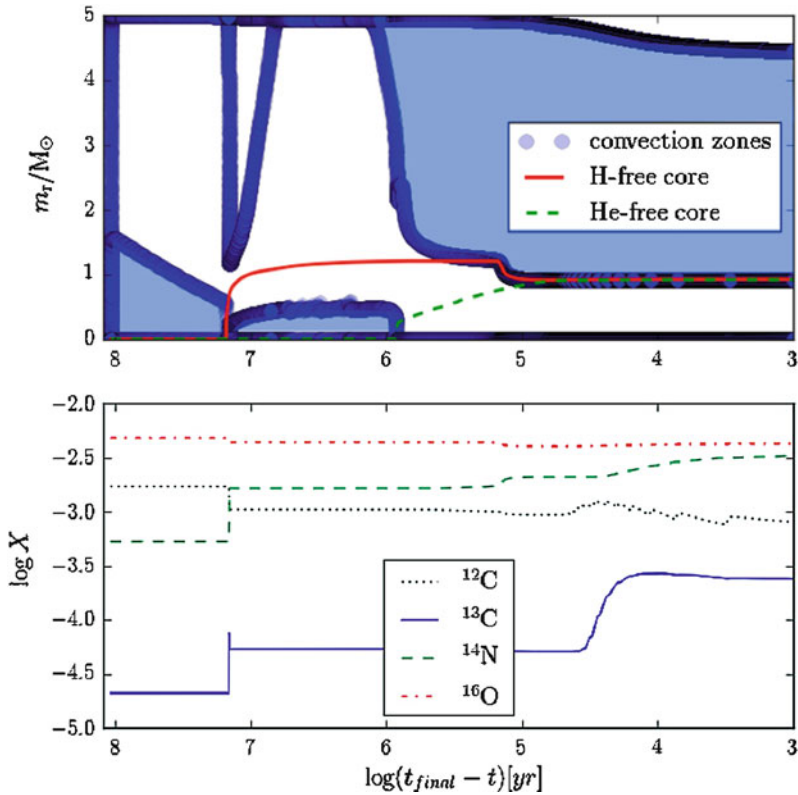
Some of the structure in the MiMf relation may be related to initial mass limits of different burning types. Stars with $M_{\text{ZAMS}} \lesssim 1.5 M_{\odot}$ will ignite He-core burning under degenerate conditions (Serenelli and Weiss 2005). The characteristic triangle-shaped feature of the $1.65 M_{\odot}$ track in the $\rho_c - T_c$ (► Fig. 8-3) signals the expansion of the core at constant degeneracy (parallel to the $P_{\text{gas}} = P_{\text{deg}}$ line) has a result of the energy input from the flash, followed by heating at almost constant pressure (sharp upward evolution at $\log \rho_c \sim 4.5$) toward the location of He-core burning. A Kippenhahn diagram of a stellar evolution simulation through the He-core flash is shown in ► Fig. 8-22. Using the model number instead of time on the abscissa allows to show both the short-lived He-core flash, followed by multiple smaller post-flashes, the He-core burning (model $\sim 17,100$ – $17,900$), the early AGB evolution, as well as the first two thermal pulses.

The $1.65 M_{\odot}$, $Z = 0.02$ track does not show significant third dredge-up because the envelope mass at the first thermal pulse is already relatively small. The stellar model therefore never becomes C-rich (► Fig. 8-12) and does not experience the associated enhanced mass loss. The $2 M_{\odot}$ model has efficient dredge-up, which implies quick increase of the C/O ratio and reduced core growth compared to a sequence with less efficient third dredge-up. Once enhanced mass-loss for C-rich surface composition is activated further core growth will be limited.



■ Fig. 8-22

Kippenhahn diagram (as a function of model number) of the He-core flash evolution, He-core burning phase, and following early and thermal-pulse AGB evolution for a $M_{\text{ZAMS}} = 1 M_{\odot}$, $Z = 0.01$ stellar evolution calculation. The HRD of this track is shown in ► Fig. 8-19



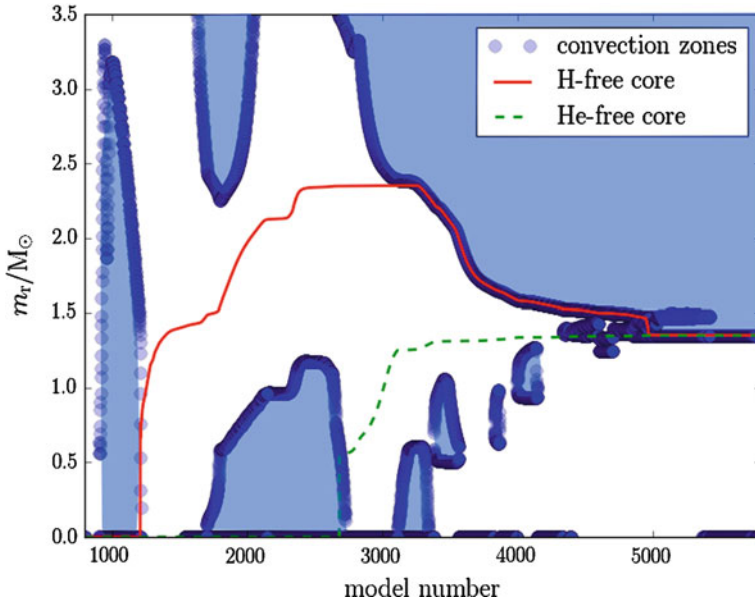
■ Fig. 8-23

Kippenhahn diagram (*top panel*) and surface abundance evolution for a $M_{\text{ZAMS}} = 5 M_\odot$, $Z = 0.01$ stellar evolution calculation. For this mass, the first ($\log(t_{\text{final}} - t) = 7.1$), second ($\log(t_{\text{final}} - t) = 5.1$), and the third ($\log(t_{\text{final}} - t) < 7.24$) dredge-ups are present, plus hot-bottom burning is modifying the CNO isotopes during the interpulse phases between He-shell flashes

Massive AGB Stars with Hot-Bottom Burning That Become CO White Dwarfs

At a core mass $\gtrsim 0.8 M_\odot$, hot-bottom burning sets in and the slope of the MiMf relation changes again. Hot-bottom burning is the colorful name for a situation when the bottom of the convective envelope reaches into the outer layers of the H-burning shell (Boothroyd et al. 1995). As a result, the envelope abundances, in particular of the CNO isotopes, change during the interpulse phase (► Fig. 8-23). Due to the convectively added fuel, the H-burning shell generates luminosity that is larger than expected from the core-mass luminosity relation (see above and Bloeker and Schönberner 1991).

► Figure 8-23 shows the surface abundance evolution for the entire evolution of a $5 M_\odot$, $Z = 0.01$ stellar evolution track. The evolution up to the end of He-core burning is similar to that described in detail for the $3 M_\odot$ case. The effect of the first dredge-up can be seen at $\log(t_{\text{final}} - t) = 7.1$. At the time mark 5.1, the H-free core decreases significantly, and some ^{14}N and ^4He (not shown) are mixed into the envelope. This is the second dredge-up which is only found in the more massive AGB stars (► Figs. 8-23 and ► 8-24). Once the thermal-pulse AGB starts, the CNO isotopes show the effect of both third dredge-up and H burning at the



■ Fig. 8-24

Same as [Fig. 8-22](#), but showing the ignition of C in after He-core burning in the $M_{\text{ZAMS}} = 9 M_{\odot}$, $Z = 0.02$ stellar evolution calculation. The central abundance evolution of this track was shown in HRD of this track is shown in [Fig. 8-4](#)

bottom of the envelope. ^{12}C is mixed up and its abundance increases a little bit, only to decrease during the subsequent interpulse by processing into ^{14}N . As a result, the $5 M_{\odot}$ stellar model will not become C-rich. The C/O ratio evolution for this case ([Fig. 8-12](#)) first shows initially simply small increases of the C/O ratio because, although $\lambda \sim 1$, the total mass that is dredge-up, compared to the envelope mass is simply rather small. The intershell mass decreases with increasing core mass. Only during the last thermal-pulse cycle is a reduction due to hot-bottom burning evident (around $M_{\star} \sim 2.5 M_{\odot}$).

Characteristic for hot-bottom burning is a $^{12}\text{C}/^{13}\text{C}$ ratio close to the (T-dependent) equilibrium value for H burning. In the case shown here, hot-bottom burning is not efficient enough to destroy O. This is however possible in more massive cases and at lower metal content.

The reduced slope of the MiMf in this mass range may be explained by a combination of effects. In hot-bottom burning, part of the fuel is brought into the H-burning shell convectively from the envelope above, while the H-burning ashes are brought back into the envelope. This reduces the amount of He accreted on the CO core and slows the core growth. Models in this mass range also show the largest dredge-up efficiency, further limiting core growth during the thermal-pulse phase.

Progenitors of ONeMg WDs or Electron-Capture Core-Collapse Supernova

If the H-free core exceeds $\sim 1.05 M_{\odot}$, C is ignited and O, Ne, and Mg are produced. These are the super-AGB stars (García-Berro and Iben 1994; Siess 2010). The limiting masses for this evolutionary outcome are important for supernova progenitor determinations. Only white dwarfs that form a CO core can explode as a thermonuclear supernova of type Ia. The limiting initial mass for CO core formation is $M_{\text{ZAMS}} = 7 M_{\odot}$ for $Z = 0.02$ which can be read

off (Fig. 8-21). Stars with larger initial mass are not supernova type Ia progenitors. Since the largest core mass with which a CO-white dwarf can form from single star evolution is $1.05 M_{\odot}$, the white dwarf will somehow have to gain mass to reach Chandrasekhar mass. This can either happen through accretion and shell burning from a binary companion or via a double degenerate merger. If the initial mass is larger than $M_{\text{ZAMS}} = 9 M_{\odot}$, the star will ignite C as well as following burning stages under non- or only weakly-degenerate conditions, and the outcome is likely a core-collapse supernova.

For stellar models that just make it over the $\sim 1.05 M_{\odot}$ core-mass limit (after the second dredge-up), C-ignition takes place off-center in a shell and just like the He-core flash under degenerate conditions. In fact, in the $\rho_c - T_c$ diagram (Fig. 8-3), a similar triangle-shaped feature can be observed for the $9 M_{\odot}$ track that again signals the C-shell ignition. Whether or not these C-shells always burn all the way to the center is not yet clear. In the case shown in Fig. 8-22, C-ignition takes place in the core, followed by a series of shell flashes that work their way out to the He-free core-mass boundary. During the same time, significant second dredge-up can be seen in the outer layers. The H-free core is reduced from $\sim 2.3 M_{\odot}$ to just below the Chandrasekhar mass. Around model number 4,500, a small He-shell flash convection zone emerges which gives the second dredge-up a final boost. This feature has been observed in many super-AGB simulations, and it is also called dredge-out (Ritossa et al. 1999; Siess 2007).

Super-AGB stars could in principle end their lives in supernova explosions if the H-free core can grow to the Chandrasekhar mass. Whether or not this can happen depends primarily on the mass loss of the supergiants, as well as on the dredge-up efficiency (Poelarends et al. 2008). The latter depends, again, on the efficiency of convective boundary mixing. If the mass loss is large, the envelope of the super-AGB star is lost quickly, and the core, which is maximally growing at the H-shell burning rate, does not have time to reach Chandrasekhar mass. If convective boundary mixing is efficient, the dredge-up parameter may be $\lambda \sim 1$, and the core does not grow, no matter how low the mass loss is. If the super-AGB star reaches the Chandrasekhar mass, it will start collapse via electron captures on Mg and Ne. The details of the subsequent supernova explosion are presently investigated.

If the super-AGB star does not explode, it will end its life as a massive ONe white dwarf. Numerous ONe nova have been found (Shore et al. 2003), and for those, obviously, ONe white dwarf must be able to form. Also, the WD mass distribution is now better populated with objects around $1.2\text{--}1.3 M_{\odot}$. These could very well be ONe white dwarf.

5 Evolution at Low Metallicity and the Dynamical Star

In Sect. 2, some limitations of the spherically symmetric stellar evolution assumptions were described. In this section, the status of modeling rotation in AGB stars as well as the phenomenon of dynamic interactions between rapid nuclear energy release via the $^{12}\text{C}(p, \gamma)^{13}\text{N}$ reaction rate and convective flow will be described.

5.1 Rotation

Rotation will change the evolution of stars in at least three ways. When stars rotate, equipotential and isothermal surfaces in the star may not coincide anymore. Therefore, the direction in which

the temperature gradient points is not always parallel to the direction where gravity points. This leads to global thermal and structural imbalances and deviations from strict spherical symmetry that give rise to large-scale meridional circulations which induces mixing. Mixing will also result, for example, from shear flows in regions of large angular velocity gradients (Maeder and Meynet 2000).

In intermediate-mass stars of low metallicity, these rotation-induced mixing processes are predicted to lift primary ^{12}C and ^{16}O out of the He-burning core (Meynet and Maeder 2002). This material can further mix through the H-burning shell, where most of it will be processed into ^{14}N (see [Fig. 8-14](#) for an example of a H-burning shell in an AGB star). According to these models, intermediate-mass stars with even very-low initial metal content would arrive on the thermal-pulse AGB with an envelope CNO abundance that is only 1 dex below solar. This would have implications for, and in fact may prevent, the occurrence of the H-ingestion events in low-metallicity thermal-pulse AGB stars.

Rotation, as well as other mixing mechanisms, have been investigated as sources for extra-mixing below that envelope convection zone that is required to explain the abundance evolution along the RGB and (early-)AGB (e.g., Denissenkov and Vandenberg 2003; Palacios et al. 2006; Charbonnel and Lagarde 2010).

Rotation will effect the way the s process operates in thermal-pulse AGB stars (Langer et al. 1999; Herwig et al. 2003; Siess et al. 2004). The current models do not induce enough mixing at the end of the third dredge-up to account quantitatively for the size of the ^{13}C pocket (see [Sect. 3.2.2](#)). In addition, present models of rotating AGB stars predict too much mixing in the ^{13}C pocket which would lead to a poisoning of the s -process neutron source via the admixture of ^{14}N and the $^{14}\text{N}(n, p)^{14}\text{C}$ reaction. This large amount of mixing at the core-envelope interface may be the result of a too large angular velocity gradient resulting from inefficient angular momentum transport during the previous evolution. The fact that observed white dwarfs rotate much slower than the cores of rotating tip-AGB star models (Suijs et al. 2008) points in the same direction of missing angular momentum transport mechanisms in present models of rotating stars. However, a significantly reduced amount of rotation-induced mixing at the core-envelope boundary may in fact help to account for the observed spread in neutron exposure for stars of a given initial mass and metallicity (Herwig 2005).

5.2 Dynamic Interaction of Nuclear Burning with Convection

The most important difference for the evolution of stars at very low or even zero metal content is the reduced availability of CNO material that either prevents or limits the efficiency of the CNO-cycle mode of H burning. As a result, to make better use of what is available, the CNO cycle will operate at higher temperature to recycle the available CNO material more often.

If the metal content is indeed zero (Pop. III), then H burning will proceed via the PP chains. At higher masses, this will require such large temperatures that even during H-core burning at some point trace amounts of ^{12}C appear from triple- α reaction.

When H burns at higher temperature, it is closer, in terms of entropy, to the conditions in the He shell. One of the themes of stellar evolution at extremely low metal content ($Z \lesssim 10^{-4}$) is frequent encounters of H-combustion events. These are all events in which in one way or another protons get mixed with primary ^{12}C at convective velocities. Such instances have been reported in stellar evolution calculations of solar and intermediate-mass stars of very-low metal content for many decades, for example, in the He-core flash during which H-rich material from

the unprocessed envelope above is ingested into the ^{12}C -rich layer underneath (Fujimoto et al. 2000; Weiss et al. 2004). A similar scenario are the AGB He-shell flashes (Iwamoto et al. 2004; Campbell and Lattanzio 2008), where again H ingestion in the He-shell flash has been found. Closely related phenomena have been observed during the third dredge-up in metal-poor AGB stars, where the convective boundary mixing induced to facilitate dredge-up and ^{13}C causes a vicious feedback from protons reacting with ^{12}C on the mixing time scale. This hot dredge-up is quantitatively not understood and limits the accuracy of simulation predictions at these low metallicities (Herwig 2004; Goriely and Siess 2004).

All these events are characterized by an interaction of the rapid nuclear energy release by $^{12}\text{C}(p, \gamma)^{13}\text{N}$ as well as the β -decay of ^{13}N and the hydrodynamics and turbulence of the convective flow. The regime can be identified in terms of the Damköhler number (Dimotakis 2005):

$$D_\alpha = \frac{\tau_{\text{mix}}}{\tau_{\text{react}}} \quad (8.9)$$

where τ_{mix} represents the mixing time scale, which can be estimated from mixing-length theory in the convection zone as the convective turnover time scale. τ_{react} is the reaction time scale, and can be estimated for each reaction, for example, for the $^{12}\text{C}(p, \gamma)$ reaction by

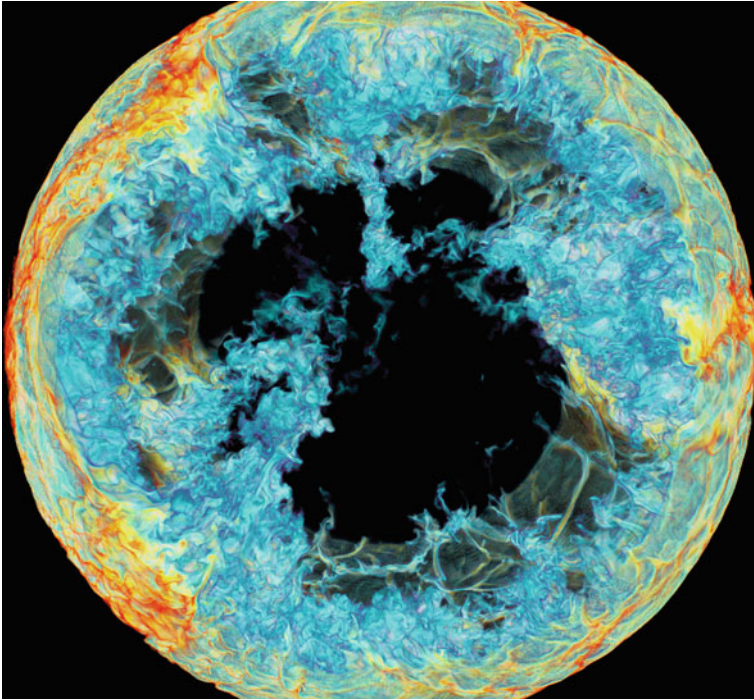
$$\tau_{^{12}\text{C}}(p) = \frac{12}{X(^{12}\text{C})\rho N_A <\sigma v>} \quad (8.10)$$

where 12 is the mass number of ^{12}C , X is the mass fraction of ^{12}C , ρ density, and $N_A <\sigma v>$ is the Avogadro number times the reaction rate, as, for example, tabulated for many charged particle reactions by Angulo et al. (1999).

If $D_\alpha \ll 1$, then nuclear burning is slow compared to mixing, and the convection zone is always well mixed. This would be, for example, the situation during H-core burning. However, if $D_\alpha \sim 1$ burning and advective flows are on the same time scale and will effect each other if the amount of energy released in the reactions is dynamically relevant. The reaction rate increases strongly with temperature. At $T \sim 1.5 \times 10^8 \text{ K}$ and $\rho \sim 10^3 \text{ g/cm}^3$, $\tau_{\text{mix}} \sim \tau_{\text{react}} \sim 15 \text{ min}$. The temperature corresponds to a location approximately halfway between the location where the H ingestion takes place ($T < 10^8 \text{ K}$) and where the He burning takes place ($T \sim 3 \times 10^8 \text{ K}$).

The main problem of these H-combustion events is that it is unlikely that they can be correctly described with the mixing model (mixing-length theory) presently used in many spherical stellar evolution codes (see discussion in ► Sect. 2). Mixing-length theory uses time and spatial averages over times and scales that are much longer/larger than the scales of significant nuclear energy release on the convective turnover time scale. The key question, that spherically symmetric simulations cannot answer, is how distributed and inhomogeneous the nuclear energy release is on large scales. The more inhomogeneous the burn is due to inhomogeneous distribution of H-fuel enriched downdrafts, the less spherically symmetric is the situation. Another key question concerns the width of the burning flame that depends on the velocity distribution of H-fuel enriched downdrafts.

In the previous section, the H-ingestion evolution of very-late thermal-pulse stars was described. It is triggered by a He-shell flash that happens after the post-AGB star has evolved around the knee, but before, it has cooled down so much that it cannot support nuclear burn anymore at all. These post-AGB H-ingestion flash stars perform the same type of event as their very-low-metallicity counterparts. The advantage of the post-AGB stars is that for some of them, most notably Sakurai's object, excellent observational information of the real-time evolution



■ Fig. 8-25

Concentration (fractional volume) of H-rich material being entrained and ingested into the He-shell flash convection zone of a pre-white dwarf (born-again) model at a relatively early time (278 min). This 3D simulation was performed on a 768^3 grid with Paul Woodward's PPMstar code. It aims to reproduce the conditions of the H-ingestion flash that occurred in post-AGB star Sakurai's object. The stratification is almost identical to stellar evolution model 4908 from Herwig et al. (2011). The color scale is chosen so that concentration levels above $\sim 10^{-3.5}$ are transparent. Yellow corresponds to concentration $\sim 10^{-4}$, white $\sim 10^{-5}$, and light blue $\sim 10^{-6}$.

and surface abundance changes is available. These can be used to validate multi-physics simulations of the nucleosynthesis of the H-combustion phenomenon. Such a program is presently underway (Herwig et al. 2011).

A snapshot from a hydrodynamic simulation of the entrainment and subsequent nuclear burning of H into the He-shell flash convection zone is shown in ► Fig. 8-25. The simulation aims to reproduce specifically the conditions of Sakurai's object when it erupted in 1996. One of the most obvious but also most important differences between 3D and 1D simulations of such phases is that in 1D, the H-rich material is represented by one number for the entire 4π sphere according to the assumption of spherical symmetry. Contrary to that, the 3D simulation shown in ► Fig. 8-25 reveals large-scale inhomogeneities during this early stage of the H entrainment. The energy generation follows the composition distribution and they both will be patchy and spotty. While 1D simulations predict that the energy should be released in a very well-defined and thin layer, the 3D picture suggests that a more regional, variable, and time-dependent nuclear burning mode is possible under certain circumstances. The spherically symmetric burn

of the 1D simulations would predict that the energy release from H-combustion would quickly prohibit mixing from the outer parts of the convection zone to the hottest layers where neutrons can be released. Investigations based on hydrodynamic simulations such as those shown in [Fig. 8-25](#) will eventually clarify the properties of H combustion in stellar evolution. This will greatly improve the fidelity of simulations and yield prediction of stars with very low or even zero metal content that were formed in the early universe.

Acknowledgments

The author would like to thank the NuGrid collaboration that has enabled the detailed nucleosynthesis simulations discussed in this chapter, and Bill Paxton and the MESA council for creating and supporting the MESA stellar evolution code, as well as Paul Woodward and the members of the LCSE at the University of Minnesota, with whom the author has collaborated substantially to create a new generation of stellar hydrodynamics simulations. The author would like to thank Chris Fryer from Los Alamos National Laboratory and the New Mexico Consortium, Los Alamos, NM for their hospitality during the time when this chapter was finished.

References

- Angulo, C., Arnould, M., & Rayet, M. et al. 1999, Nucl. Phys. A, 656, 3, NACRE compilation
- Arlandini, C., Käppeler, F., Wisshak, K., Gallino, R., Lugaro, M., Busso, M., & Straniero, O. 1999, ApJ, 525, 886
- Arndt, T., Fleischer, A., & Sedlmayr, E. 1997, A&A, 327, 614
- Arnett, D., Meakin, C., & Young, P. A. 2010, ApJ, 710, 1619
- Arnould, M., Goriely, S., & Jorissen, A. 1999, A&A, 347, 572
- Asplund, M., Lambert, D. L., Kipper, T., Pollacco, D., & Shetrone, M. D. 1999, A&A, 343, 507
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
- Bisterzo, S., Gallino, R., Straniero, O., Cristallo, S., & Käppeler, F. 2010, MNRAS, 404, 1529
- Blöcker, T. 1993, PhD thesis, Universität Kiel
- Blöcker, T. 1995a, A&A, 297, 727
- Blöcker, T. 1995b, A&A, 299, 755
- Blöcker, T., & Schönberner, D. 1991, A&A, 244, L43
- Blöcker, T., & Schönberner, D. 1997, A&A 324, 991
- Böhm-Vitense, E. 1958, Z. Astrophys., 46, 108
- Boothroyd, A. I., Sackmann, I.-J., & Wasserburg, G. J. 1995, ApJL, 442, L21
- Brown, T. M., Sweigart, A. V., Lanz, T., Landsman, W. B., & Hubeny, I. 2001, ApJ, 562, 368
- Browning, M. K., Brun, A. S., & Toomre, J. 2004, ApJ, 601, 512
- Brun, A. S., & Palacios, A. 2009, ApJ, 702, 1078
- Busso, M., Gallino, R., & Wasserburg, G. J. 1999, ARA&A, 37, 239
- Busso, M., Gallino, R., Lambert, D. L., Travaglio, C., & Smith, V. V. 2001, ApJ, 557, 802
- Campbell, S. W., & Lattanzio, J. C. 2008, A&A, 490, 769
- Canuto, V. M. 2011, A&A, 528, A77
- Canuto, V. M., Goldman, I., & Mazzitelli, I. 1996, ApJ, 473, 550
- Catalán, S., Isern, J., García-Berro, E., & Ribas, I. 2008, MNRAS, 387, 1693
- Charbonnel, C., & Lagarde, N. 2010, A&A, 522, A10
- Charbonnel, C., & Zahn, J. 2007, A&A, 467, L15
- Clarke, C. J., & Carswell, R. F. 2003, Principles of Astrophysical Fluid Dynamics (Cambridge: Cambridge University Press)
- Clayton, D. D. 1968, Principles of Stellar Evolution and Nucleosynthesis (Chicago: University of Chicago Press)
- Clayton, G. C., Geballe, T. R., Herwig, F., Fryer, C., & Asplund, M. 2007, ApJ, 662, 1220
- Cox, J. P., & Giuli, R. T. 1968, Principles of Stellar Structure (New York: Gordon and Breach)
- Cristallo, S., Straniero, O., Lederer, M. T., & Aringer, B. 2007, ApJ, 667, 489
- Cristallo, S., Piersanti, L., Straniero, O., Gallino, R., Domínguez, I., Abia, C., Di Rico, G., Quintini, M., & Bisterzo, S. 2011, ApJS, 197, 17

- De Beck, E., Decin, L., de Koter, A., Justtanont, K., Verhoelst, T., Kemper, F., & Menten, K. M. 2010, *A&A*, 523, A18
- Denissenkov, P. A. 2010, *ApJ*, 723, 563
- Denissenkov, P. A., & Merryfield, W. J. 2011, *ApJL*, 727, L8
- Denissenkov, P. A., & VandenBerg, D. A. 2003, *ApJ*, 593, 509
- Dimotakis, P. E. 2005, *Annu. Rev. Fluid Mech.*, 37, 329
- Driebe, T., Schönberner, D., Böcker, T., & Herwig, F. 1998, *A&A*, 339, 123
- Duerbeck, H. W., & Benetti, S. 1996, *ApJ Lett.*, 468, L111
- Endal, A. S., & Sofia, S. 1976, *ApJ*, 210, 184
- Ferrario, L., Wickramasinghe, D., Liebert, J., & Williams, K. A. 2005, *MNRAS*, 361, 1131
- Forestini, M., & Charbonnel, C. 1997, *A&AS*, 123, 241
- Freiburghaus, C., Rosswog, S., & Thielemann, F.-K. 1999, *ApJL*, 525, L121
- Freytag, B., & Höfner, S. 2008, *A&A*, 483, 571
- Frost, C. A., & Lattanzio, J. C. 1996, *ApJ*, 473, 383
- Fryer, C. L., Herwig, F., Hungerford, A., & Timmes, F. X. 2006, *ApJ Lett.*, 646, L131
- Fujimoto, M. Y. 1977, *PASJ*, 29, 331
- Fujimoto, M. Y., Ikeda, Y., & Iben, I., Jr. 2000, *ApJ Lett.*, 529, L25
- Gallino, R., Arlandini, C., Busso, M., Lugaro, M., Travaglio, C., Straniero, O., Chieffi, A., & Limongi, M. 1998, *ApJ*, 497, 388
- García-Berro, E., & Iben, I., Jr. 1994, *ApJ*, 434, 306
- Gorieli, S., & Siess, L. 2004, *A&A*, 421, L25
- Groenewegen, M. A. T., & de Jong, T. 1993, *A&A*, 267, 410
- Groenewegen, M. A. T., & de Jong, T. 1998, *A&A*, 337, 797
- Gustafsson, B., & Höfner, S. 2003, in *Astronomy and Astrophysics Library, Asymptotic Giant Branch Stars*, ed. H. J. Habing & H. Olofsson (New York/Berlin: Springer)
- Habing, H. J., & Olofsson, H. (eds.) 2004, *Astronomy and Astrophysics Library, Asymptotic Giant Branch Stars* (New York/Berlin: Springer)
- Hajduk, M., Zijlstra, A. A., Herwig, F., van Hoof, P. A. M., Kerber, F., Kimeswenger, S., Pollacco, D. L., Evans, A., López, J. A., Bryce, M., Eyres, S. P. S., & Matsuura, M. 2005, *Science*, 308, 231
- Hansen, C. J., Kawaler, S. D., & Trimble, V., ed. 2004, *Stellar Interiors : Physical Principles, Structure, and Evolution* (Springer-Verlag)
- Heney, L. G., Forbes, J. E., & Gould, N. L. 1964, *ApJ*, 139, 306
- Herwig, F. 2000, *A&A*, 360, 952
- Herwig, F. 2001a, *ApJ Lett.*, 554, L71
- Herwig, F. 2001b, *APSS*, 275, 15
- Herwig, F. 2004, *ApJ*, 605, 425
- Herwig, F. 2005, *ARAA*, 43, 435
- Herwig, F., & Austin, S. M. 2004, *ApJ Lett.*, 613, L73
- Herwig, F., Böcker, T., Schönberner, D., & El Eid, M. F. 1997, *A&A*, 324, L81
- Herwig, F., Schönberner, D., & Böcker, T. 1998, *A&A*, 340, L43
- Herwig, F., Böcker, T., Langer, N., & Driebe, T. 1999, *A&A*, 349, L5
- Herwig, F., Langer, N., & Lugaro, M. 2003, *ApJ*, 593, 1056
- Herwig, F., Austin, S. M., & Lattanzio, J. C. 2006a, *Phys. Rev. C.*, 73, 025802
- Herwig, F., Freytag, B., Hueckstaedt, R. M., & Timmes, F. X. 2006b, *ApJ*, 642, 1057
- Herwig, F., Freytag, B., Fuchs, T., Hansen, J. P., Hueckstaedt, R. M., Porter, D. H., Timmes, F. X., & Woodward, P. R. 2007, in *Astronomical Society of the Pacific Conf. Ser.* 378, *Why Galaxies Care About AGB Stars: Their Importance as Actors and Probes*, ed. F. Kerschbaum, C. Charbonnel, & R. F. Wing, 43 (San Francisco, CA: ASP)
- Herwig, F., Pignatari, M., Woodward, P. R., Porter, D. H., Rockefeller, G., Fryer, C. L., Bennett, M., & Hirschi, R. 2011, *ApJ*, 727, 89
- Hinkle, K. H., Lebzelter, T., Joyce, R. R., Ridgway, S., Close, L., Hron, J., & Andre, K. 2008, *A&A*, 479, 817
- Hofmeister, E., Kippenhahn, R., & Weigert, A. 1964, *Z. Astrophys.*, 59, 215
- Hollowell, D., & Iben, I., Jr. 1988, *ApJ*, 333, L25
- Iben, I., Jr. 1981, *ApJ*, 246, 278
- Iben, I., Jr., & MacDonald, J. 1995, in *LNP No. 443, White Dwarfs*, ed. D. Koester & K. Werner (Heidelberg: Springer), 48
- Iben, I., Jr., Kaler, J. B., Truran, J. W., & Renzini, A. 1983, *ApJ*, 264, 605
- Iglesias, C. A., & Rogers, F. J. 1996, *ApJ*, 464, 943
- Iliadis, C. 2007, *Nuclear Physics of Stars* (Weinheim: Wiley)
- Iwamoto, N., Kajino, T., Mathews, G. J., Fujimoto, M. Y., & Aoki, W. 2004, *ApJ*, 602, 378
- Jeffery, C. S., & Schönberner, D. 2006, *A&A*, 459, 885
- Jura, M., & Kleinmann, S. G. 1989, *ApJ*, 341, 359
- Jura, M., & Kleinmann, S. G. 1992, *ApJS*, 83, 329
- Käppeler, F., Beer, H., & Wisshak, K. 1989, *Rep. Prog. Phys.*, 52, 945
- Käppeler, F., Gallino, R., Bisterzo, S., & Aoki, W. 2011, *Rev. Mod. Phys.*, 83, 157
- Karakas, A. I. 2010, *MNRAS*, 403, 1413
- Karakas, A. I., Lattanzio, J. C., & Pols, O. R. 2002, *PASA*, 19, 515
- Kippenhahn, R., & Weigert, A. 1990, *Stellar Structure and Evolution* (Berlin: Springer)
- Kroupa, P. 2002, *Science*, 295, 82

- Langer, N., Heger, A., Wellstein, S., & Herwig, F. 1999, *A&A*, 346, L37
- Lattanzio, J. C. 1986, *ApJ*, 311, 708
- Lattanzio, J. C. 1989, *ApJ*, 344, L25
- Lawlor, T. M., & MacDonald, J. 2003, *ApJ*, 583, 913
- Liebert, J., Bergeron, P., & Holberg, J. B. 2005, *ApJS*, 156, 47
- Lugaro, M. 2005, *World Scientific Series in Astronomy and Astrophysics, Stardust from Meteorites. An Introduction to Presolar Grains*, Vol. 9 (London/Singapore: World Scientific)
- Lugaro, M., Herwig, F., Lattanzio, J. C., Gallino, R., & Straniero, O. 2003, *ApJ*, 586, 1305
- Maeder, A., & Meynet, G. 2000, *ARA&A*, 38, 143
- Maeder, A., & Meynet, G. 2012, *Rev. Mod. Phys.*, 84, 25
- Marigo, P. 2000, *A&A*, 360, 617
- Marigo, P. 2001, *A&A*, 370, 194
- Marigo, P. 2002, *A&A*, 387, 507
- Marigo, P., & Aringer, B. 2009, *A&A*, 508, 1539
- Marigo, P., Bressan, A., & Chiosi, C. 1996, *A&A*, 313, 545
- Marigo, P., Girardi, L., & Bressan, A. 1999, *A&A*, 344, 123
- Mattsson, L., Wahlin, R., & Höfner, S. 2010, *A&A*, 509, A14
- Meyer, B. S. 1994, *ARA&A*, 32, 153
- Meynet, G., & Maeder, A. 2002, *A&A*, 390, 561
- Miller Bertolami, M. M., Althaus, L. G., Serenelli, A. M., & Panei, J. A. 2006, *A&A*, 449, 313
- Mowlavi, N. 1999, *A&A*, 344, 617
- Olofsson, H. 2003, in *Astronomy and Astrophysics library, Asymptotic Giant Branch Stars*, ed. H. J. Habing & H. Olofsson (New York, Berlin: Springer)
- Olofsson, H., González Delgado, D., Kerschbaum, F., & Schöier, F. L. 2002, *A&A*, 391, 1053
- Paczynski, B. 1970, *Acta Astron.*, 20, 47
- Palacios, A., Charbonnel, C., Talon, S., & Siess, L. 2006, *A&A*, 453, 261
- Passy, J.-C., De Marco, O., Fryer, C. L., Herwig, F., Diehl, S., Oishi, J. S., Mac Low, M.-M., Bryan, G. L., & Rockefeller, G. 2012, *ApJ*, 744, 52
- Pauldrach, A., Puls, J., Kudritzki, R. P., Mendez, R. H., & Heap, S. R. 1988, *A&A*, 207, 123
- Paxton, B., Bildsten, L., Dotter, A., Herwig, F., Lesaffre, P., & Timmes, F. 2011, *ApJS*, 192, 3
- Pinsonneault, M. 1997, *ARA&A*, 35, 557
- Poelarends, A. J. T., Herwig, F., Langer, N., & Heger, A. 2008, *apj*, 675, 614
- Porter, D. H., & Woodward, P. R. 2000, *ApJS*, 127, 159
- Renzini, A., & Voli, M. 1981, *A&A*, 94, 175
- Ricker, P. M., & Taam, R. E. 2008, *ApJL*, 672, L41
- Ritossa, C., García-Berro, E., & Iben, I. J. 1999, *ApJ*, 515, 381
- Rolfs, C. E., & Rodney, W. S. 1988, *Cauldrons in the Cosmos: Nuclear Astrophysics* (Chicago: University of Chicago Press)
- Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, *A&AS*, 96, 269
- Schönberner, D. 1981, *A&A*, 103, 119
- Serenelli, A., & Weiss, A. 2005, *A&A*, 442, 1041
- Shore, S. N., Schwarz, G., Bond, H. E., Downes, R. A., Starrfield, S., Evans, A., Gehrz, R. D., Hauschildt, P. H., Krautter, J., & Woodward, C. E. 2003, *AJ*, 125, 1507
- Siess, L. 2007, *A&A*, 476, 893
- Siess, L. 2010, *A&A*, 512, A10+
- Siess, L., Goriely, S., & Langer, N. 2004, *A&A*, 415, 1089
- Snedden, C., Cowan, J. J., & Gallino, R. 2008, *ARAA*, 46, 241
- Straniero, O., Gallino, R., Busso, M., Chieffi, A., Raiteri, C. M., Salaris, M., & Limongi, M. 1995, *ApJ*, 440, L85
- Straniero, O., Cristallo, S., & Gallino, R. 2009, *PASA*, 26, 133
- Suijs, M. P. L., Langer, N., Poelarends, A.-J., Yoon, S.-C., Heger, A., & Herwig, F. 2008, *A&A*, 481, L87
- Thompson, M. J., Christensen-Dalsgaard, J., Miesch, M. S., & Toomre, J. 2003, *ARAA*, 41, 599
- Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995, *APJS*, 98, 617
- Travaglio, C., Gallino, R., Arnone, E., Cowan, J., Jordan, F., & Sneden, C. 2004, *ApJ*, 601, 864
- van Loon, J. T. 2005, *Stellar Evolution at Low Metallicity: Mass Loss, Explosions, Cosmology*. ASP Conf. Ser. 353, Proceedings of the Conference Held 15-19 August, 2005 in Tartu, Estonia, (San Francisco, CA: ASP) p. 211. *ArXiv Astrophysics e-prints*
- van Loon, J. T., Groenewegen, M. A. T., de Koter, A., Trams, N. R., Waters, L. B. F. M., Zijlstra, A. A., Whitelock, P. A., & Loup, C. 1999, *A&A*, 351, 559
- van Loon, J. T., Cioni, M.-R. L., Zijlstra, A. A., & Loup, C. 2005, *A&A*, 438, 273
- van Winckel, H. 2003, *ARAA*, 41, 391
- VandenBerg, D. A., Bergbusch, P. A., & Dowler, P. D. 2006, *ApJS*, 162, 375
- Vassiliadis, E., & Wood, P. 1993, *ApJ*, 413, 641
- Venn, K. A., Irwin, M., Shetrone, M. D., Tout, C. A., Hill, V., & Tolstoy, E. 2004, *AJ*, 128, 1177
- Wachter, A., Schröder, K.-P., Winters, J. M., Arndt, T. U., & Sedlmayr, E. 2002, *A&A*, 384, 452
- Ward, R. A., & Fowler, W. A. 1980, *ApJ*, 238, 266
- Weidemann, V. 2000, *A&A*, 363, 647
- Weiss, A., & Ferguson, J. W. 2009, *A&A*, 508, 1343

- Weiss, A., Schlattl, H., Salaris, M., & Cassisi, S. 2004, *A&A*, 422, 217
- Werner, K., & Herwig, F. 2006, *PASP*, 118, 183
- Willson, L. A. 2000, *ARAA*, 38, 573
- Xiong, D. R., Cheng, Q. L., & Deng, L. 1997, *ApJS*, 108, 529
- Yoon, S.-C., Langer, N., & van der Sluis, M. 2004, *A&A*, 425, 207
- Zinner, E. 1998, *Ann. Rev. Earth Planet. Sci.*, 26, 147

9 **The Evolution of High-Mass Stars**

*Geraldine J. Peters*¹ · *Raphael Hirschi*^{2,3}

¹Space Sciences Center and Department of Physics and Astronomy,
University of Southern California, Los Angeles, CA, USA

²Astrophysics Group, EPSAM Institute, Keele University,
Keele, UK

³Kavli IPMU, University of Tokyo, Kashiwa, Chiba, Japan

1	<i>Introduction</i>	448
2	<i>The Stars of the Upper Main Sequence</i>	452
2.1	Main-Sequence O and Early B Stars	454
2.2	The Wolf–Rayet Stars	455
2.3	The Luminous Blue Variables and Early-Type Supergiants	456
3	<i>Contemporary Calculations of Evolutionary Tracks</i>	457
3.1	Stellar Evolution Models	457
3.1.1	Stellar Structure Equations	457
3.1.2	Rotation and Magnetic Fields	459
3.1.3	Other Input Physics	464
3.2	Grids of Evolutionary Tracks and Stellar Populations	464
3.2.1	Impact of Rotation	464
3.2.2	Impact of Mass Loss	466
3.2.3	Impact of Metallicity	466
3.2.4	Grids of Models and Stellar Populations	468
4	<i>Observation Confronts Theory</i>	468
4.1	Surface Abundance of Light Elements in Massive Stars	470
4.2	The Be Stars	472
4.3	Post-Main-Sequence Evolution	473
4.3.1	Blue and Red Supergiants	474
4.3.2	Wolf–Rayet (WR) Stars	478
4.4	Modeling Uncertainties and Fate of Massive Stars	481
	<i>Acknowledgment</i>	481
	<i>References</i>	482

Abstract: The evolution of stars more massive than $8 M_{\odot}$ is discussed in this chapter. On the main sequence, these stars have spectral types of B2 or earlier, but depending on their mass can evolve into red supergiants, blue supergiants, Cepheids, Wolf–Rayet stars, Of stars, or luminous blue variables before ending their evolution as core collapse supernovae and neutron stars or black holes. The chapter begins with a general discussion of the energy production in the interior of a massive star as it evolves. The main fusion reactions that generate the star’s energy are listed. Some observed properties of the O and early B main-sequence stars and their evolved products are discussed including the best determinations of their masses. The computation of contemporary evolutionary tracks that include stellar rotation and magnetic fields is detailed. The equations of stellar structure including those for energy conservation, momentum transfer, mass conservation, and energy transport are listed. The discussion includes the meridional circulation in the interior of a rotating massive star and its effect on the transport of nuclear-processed material to the surface and the impact of rotation, mass loss, and metallicity on the evolutionary tracks. Recent evolutionary tracks from the Geneva group are presented. Finally the newest evolutionary tracks and the surface abundances predicted by the calculations are compared with recent observations.

Keywords: Stars: Abundances, Stars: Early-type, Stars: Evolution, Stars: Magnetic fields, Stars: Massive, Stars: Rotation, Stars: Supergiants, Stars: Supernovae: general, Stars: Wolf–Rayet

1 Introduction

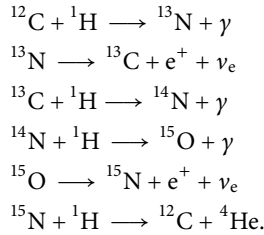
Stars more massive than $8 M_{\odot}$ may be quite rare,¹ but they supply most of the ionizing radiation in a galaxy, typically show many types of spectroscopic and photometric variability, display some of the most striking spectra, and will eventually end their evolution as core collapse supernovae, neutron stars or black holes, and some gamma-ray bursters. This chapter discusses the evolution of these stars from the main sequence to their final products. We consider only single-star evolution even though more than half of the massive stars are most likely formed in close binary systems and mass transfer between the components stars and systemic mass loss will produce profound effects on their post-main-sequence lives and the chemical evolution of a galaxy (Vanbeveren 2001; Van Rensbergen et al. 2010, 2011). In this section, a basic overview of the interior evolution of a massive star is presented. In the following sections, we discuss the observed properties of the stars that are found in the upper portion of the HR diagram, the computation of evolutionary tracks that include rotation and magnetic fields, and a discussion on the extent to which the contemporary calculations agree with the observations.

The first set of theoretical evolutionary tracks for massive stars that agreed well with the observed color-magnitude diagrams of star clusters were computed in the 1960s (and references therein Iben 1967; Iben and Renzini 1983). The computations were carried through for a nonrotating star without the influence of a magnetic field. A new generation of stellar models continues to be produced by the Geneva group (Schaller et al. 1992; Maeder and Meynet 2000a; Ekström et al. 2012, <http://obswww.unige.ch/Recherche/evol/-Database->) that include

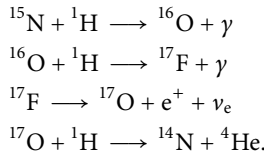
¹The ratio of the number of $20 M_{\odot}$ stars to one solar mass stars in the Milky Way galaxy is about 10^{-5} ; for every $100 M_{\odot}$ star, there are a more than a million suns (Massey 2003).

improved energy generation and neutrino loss rates, better opacities in the interior and optically thick winds of Wolf–Rayet stars, and the effects of rotation.

The energy generated in a massive star's convective core is produced by the basic thermonuclear reactions listed below (see Clayton 1983 for information on the full gamut of reactions). More recent refinements can be found in Pagel (2009) and the papers from the Geneva group that can be found on the website listed above. During the main-sequence phase, hydrogen is converted to helium through the carbon, nitrogen, and oxygen (CNO) cycle and its branch the CNO bi-cycle. If the total mass of the end products is less than the initial ones, the differential proton mass is converted to energy via the familiar Einstein equation $\Delta E = \Delta mc^2$. For massive stars, the energy generation exceeds $10^{10} \text{ MeV g}^{-1} \text{ s}^{-1}$ (Pagel 2009). The process of building the elements in the Periodic Table from other elements through thermonuclear fusion is called *nucleosynthesis*. For the CNO cycle, most of the time the reactions proceed as follows²:



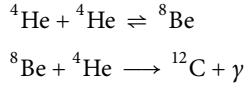
C, N, and O serve as catalysts to convert four hydrogen nuclei into one helium nucleus. γ is the gamma radiation in the MeV domain that is generated. The abundance of C, N, and O does not change. But 0.04% of the time, the last reaction takes another path and gives the CNO bi-cycle:



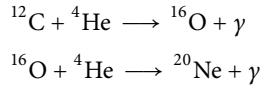
Similar to the first branch, hydrogen is converted to helium, but over time each ${}^{12}\text{C}$ catalyst is destroyed and replaced by a ${}^{14}\text{N}$ nucleus. The convective core becomes rich in helium and nitrogen but depleted in carbon. As the hydrogen fuel becomes substantially depleted, the energy production in the core decreases and the mean molecular weight increases causing a pressure decrease in the core. The core is no longer in hydrostatic equilibrium and an overall contraction ensues. Energy production eventually moves to a hydrogen-rich shell that surrounds the hydrogen-depleted core. The contraction raises the density and temperature of the core (through the release of gravitational energy), and when the core temperature reaches about 10^8 K , helium fusion can commence through the triple-alpha process (cf. Clayton 1983; Pagel 2009, for the details surrounding this remarkable set of reactions). The energy generation by the triple-alpha process and the advanced nucleosynthesis processes listed below are very temperature and density dependent and can raise the energy output from the core by an order

²Thermonuclear reaction chains are frequently written in a compact form showing the by-products in parentheses (e.g., the CNO bi-cycle is $\text{O}^{16}(\text{p}, \gamma)\text{F}^{17}(\beta^+ \nu_e)\text{O}^{17}(\text{p}, \alpha)\text{N}^{14}$, where p, β^+ , ν_e , and α are the proton, positron, neutrino, and helium nucleus.)

of magnitude or more when the threshold conditions that allow the thermonuclear reactions to operate are achieved.

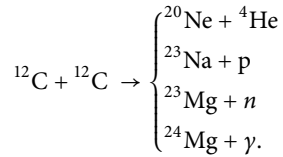


After sufficient ${}^{12}\text{C}$ has been produced, carbon can then go on to make oxygen and neon.

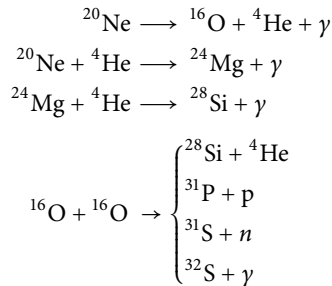


Unlike the stars of low and intermediate mass, the condensed cores of massive stars do not develop a degenerate structure, and degeneracy pressure is not a major source of support for the overlying envelope. Rather, a steady contraction sets in that ultimately leads to the synthesis of the light and Fe group elements as less massive nuclei are fused into heavier species. Each successive fusion chain is fed by the ashes left behind by the earlier one and each set of reactions ceases when the seed nuclei are consumed.³ The stars less massive than $40 M_{\odot}$ will pass through blue/red supergiant and Cepheid phases as they evolve (see [Sects. 3](#) and [4](#)), but for the most massive stars mass loss prevents the star from reaching the coolest part of the HR Diagram.

Continued core contraction past He burning,⁴ raises the central temperature to about 5×10^8 K in massive stars and produces additional light elements from the fusion of carbon. Ignition is in a nondegenerate core. In the later stages plasma neutrino loss becomes significant producing a cooling effect in the core, and the evolution greatly accelerates with each new phase. Some examples of advanced nucleosynthesis are:



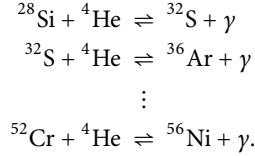
Further contraction leads to a prevailing core temperature of around $1.5 - 2.0 \times 10^9$ K and the onset of neon and oxygen fusion.



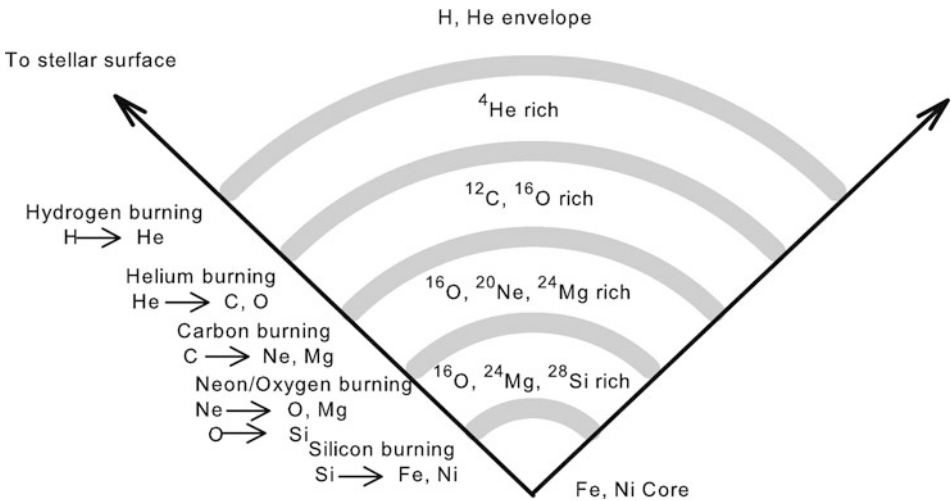
³In the jargon of nucleosynthesis, the term “ash” is used to mean the nuclear product of a series of thermonuclear reactions.

⁴Historically such thermonuclear reactions have been called “burning” as the fusion reactions consume a nuclear “fuel,” but this is a misnomer as the term implies that the reactions are chemical rather than nuclear. Nevertheless, the term is occasionally used in this chapter.

At this stage, the energy production proceeds at a very high rate, and the nuclear fuel in the core is exhausted in a short time. As the core continues to contract and its ambient temperature exceeds 3×10^9 K, silicon and other elements of intermediate mass can undergo an α -capture process to produce the Fe group nuclei through reactions similar to these⁵:



The chains above ultimately yield to the production of ${}^{56}\text{Ni}$, ${}^{56}\text{Fe}$, and the Fe group elements (Sc, Ti, V, Cr, Mn, Fe, Co, Ni). The core of the star has now taken on an onion-shell appearance (► Fig. 9-1) with an Fe–Ni center surrounded by thin shells that are the sites of Si, O, Ne, C, He, and H burning. Each shell has a temperature that is favorable to its prevailing fusion reactions. The reactions feed on the ash left behind by the fusion in the shell above it. Most of the thermonuclear reactions listed above are exothermic and thus provide energy to the stellar interior. But the Fe group elements fall at the peak of a plot of binding energy per nucleon (Clayton 1983; Pagel 2009), and ${}^{56}\text{Fe}$ is the most stable of all. Consequently, less energy is released per nucleon at this stage, so the lifetime of the thermonuclear reaction is shorter. Any further attempt to synthesize the Fe peak species is endothermic and precipitously cools the core. Energy loss from



■ Fig. 9-1

Schematic representation of the core region of a massive star prior to the collapse that results in a supernova. Energy is produced in a series of thin shells (gray regions) undergoing nuclear “burning.” The structure is not drawn to scale. Regions between the active shells contain the material for the fusion reactions below them but are not hot enough to enable them to take place. For a $25 M_{\odot}$ star, the above figure includes only the inner 30% of the star’s mass. The values of $\log (T)$ range from about 8.0 at the He-burning shell to 10.0 at the center (The illustration is adapted from data in Woosley and Weaver 1986)

⁵ ${}^4\text{He}$ comes from the photo-disintegration of ${}^{28}\text{Si}$.

neutrino production accelerates. The result is a collapse of the core and the production of a type Ib, Ic, or II supernova. It is interesting to compare the lifetimes of the hierarchy of thermonuclear fusion reactions discussed above. For a $20 M_{\odot}$ star, the main-sequence lifetime (H burning) is about 10^7 years, core He burning lasts 10^6 years, C burning is completed in 300 years, O burning in 200 days, and Si burning is less than 2 days. The final stages of core nuclear fusion occur in less than 1 year!

2 The Stars of the Upper Main Sequence

The stars considered in this chapter begin as main-sequence stars of B2V or earlier (Harmanec 1988). It is difficult to determine reliable masses for the most massive stars because we do not observe a normal photosphere. Their spectra are dominated by features formed in optically thick, dynamic atmospheres and often circumstellar material. Stellar parameters for the massive eclipsing binaries with the most accurately determined masses and radii are given in [Table 9-1](#). These systems have spectral lines that are primarily photospheric.⁶ Their masses and radii are accurate to about 2%. The most massive star with well-determined stellar parameters, V3903 Sgr (O7 V), has a mass of $27.3 M_{\odot}$, though theoretical calculations suggest that the most massive stars main-sequence stars (O2 V) are $120 M_{\odot}$ or slightly more. The highest mass determined so far is for the primary in NGC 3603-A1, a system with two WN6ha emission line stars with masses of 116 ± 31 and $89 \pm 16 M_{\odot}$ (Schnurr et al. 2008).

The evolved products of OB main-sequence stars include blue, yellow, and red supergiants (BSGs, YSGs, RSGs), Cepheids, Of stars, Wolf-Rayet (WR) stars, and luminous blue variables (LBVs). The domains in the HR diagram occupied by these objects and the locations of the well-calibrated stars listed in [Table 9-1](#) are shown in [Fig. 9-2](#). With the overlay of the evolutionary tracks from the Geneva group, it is immediately evident which mass ranges produce the various evolved stars discussed in this chapter. Stars with masses less than $25 M_{\odot}$ will evolve to the red supergiant phase. Those with masses $\leq 15 M_{\odot}$ will pass through the Cepheid instability strip. The LBVs end their evolution before they reach the red supergiant stage and are less luminous than the Humphreys–Davidson limit (Humphreys and Davidson 1994) where the star's luminosity approaches the Eddington limit.⁷

To illustrate the range of spectral characteristics that the massive stars discussed in this chapter display, we show representative far-ultraviolet (FUV) spectra of selected main-sequence and evolved objects in [Fig. 9-3](#). The FUV contains numerous resonance lines of abundant species, many of which are formed in the star's wind or circumstellar (CS) material, lines from ions of carbon and nitrogen that can confirm or refute the presence of mixing of nuclear-processed material from the interior, and the FUV is also the region where these objects radiate most of their flux. More information on the nature of these stars is presented in the three subsections below.

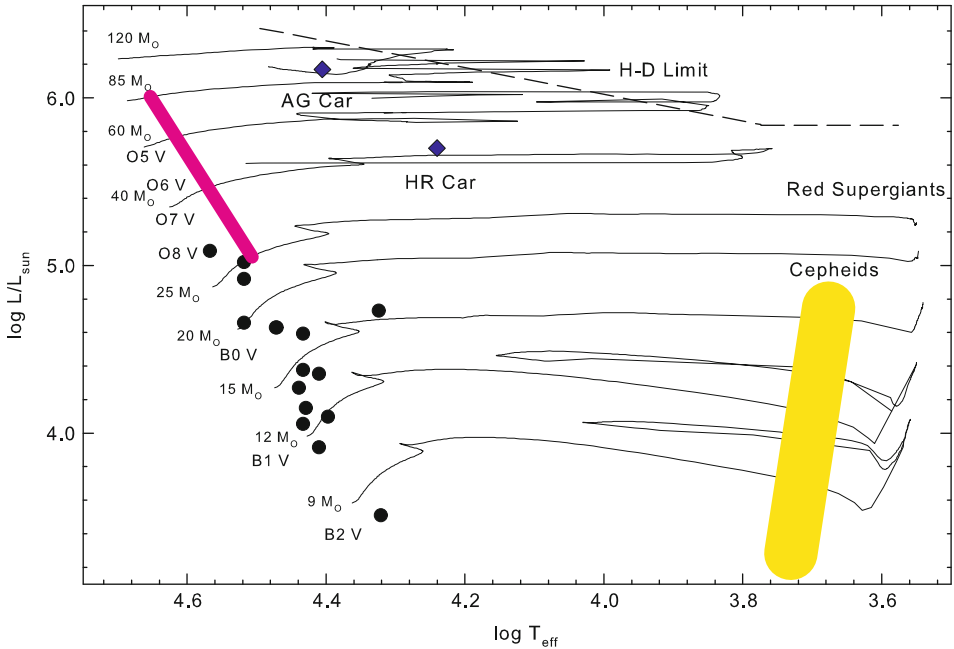
⁶The far ultraviolet spectra of B stars as cool as B0.5 IV (e.g., HR 1887) reveal violet asymmetries in the resonance lines from abundant ions such as Si IV, and such features most certainly contain a weak wind component.

⁷The Eddington limit is the maximum radiative luminosity a star can reach and still be in hydrostatic equilibrium. $L_{\text{Ed}} = 4\pi cGM/\kappa$, where c is the speed of light, G is the gravitational constant, M is the star's mass, and κ the mean opacity in the surface layers, which can include electron scattering and a temperature-dependent component due to lines from the Fe group elements.

■ Table 9-1
Early-type eclipsing binaries with well-determined masses

System	Period	V	Spectral types	Mass (M_{\odot})	Radius (R_{\odot})	$\log(T_{\text{eff}})$	$\log(L/L_{\odot})$	References
V3903 Sgr	1.744	7.27	O7 V	27.27 ± 0.55	8.088 ± 0.086	4.580 ± 0.021	5.087 ± 0.029	Vaz et al. (1997)
EM Car	3.414	8.38	O9 V	19.01 ± 0.44	6.125 ± 0.060	4.531 ± 0.021	4.658 ± 0.032	Andersen and Clausen (1989)
			O8 V	22.89 ± 0.32	9.35 ± 0.17	4.531 ± 0.026	5.020 ± 0.10	
			O8 V	21.43 ± 0.33	8.34 ± 0.14	4.531 ± 0.026	4.920 ± 0.10	
Y Cyg	2.996	7.32	O9.8 V	17.57 ± 0.27	5.93 ± 0.07	4.545 ± 0.007		Simon et al. (1994)
			O9.8 V	17.04 ± 0.26	5.78 ± 0.07	4.534 ± 0.007		
V478 Cyg	2.881	8.63	O9.5 V	16.67 ± 0.45	7.423 ± 0.079	4.484 ± 0.015	4.630 ± 0.06	Popper and Hill (1991)
			O9.5 V	16.31 ± 0.35	7.423 ± 0.079	4.485 ± 0.015	4.630 ± 0.06	Popper and Etzel (1981)
AH Cep	1.775	6.81	B0.5 V	15.2 ± 0.2	6.38 ± 0.11	4.476 ± 0.029		Holmgren et al. (1990)
			B0.5 V	13.6 ± 0.2	5.86 ± 0.13	4.456 ± 0.015		
V578 Mon	2.408	8.54	B1 V:	14.54 ± 0.08	5.23 ± 0.06	4.477 ± 0.007		Hensberge et al. (2000)
			B1 V:	10.29 ± 0.06	4.32 ± 0.07	4.422 ± 0.007		Pavlovski and Hensberge (2005)
V453 Cyg	3.890	8.29	B0.4 IV	14.36 ± 0.20	8.551 ± 0.055	4.446 ± 0.006	4.593 ± 0.025	Southworth et al. (2004)
			B0.7 IV	11.11 ± 0.13	5.489 ± 0.063	4.410 ± 0.008	4.098 ± 0.034	Pavlovski and Southworth (2009)
CW Cep	2.729	7.59	B0.5 V	13.52 ± 0.39	5.685 ± 0.130	4.452 ± 0.016	4.270 ± 0.06	Clausen and Gimenez (1991)
			B0.5 V	12.08 ± 0.29	5.177 ± 0.129	4.442 ± 0.016	4.150 ± 0.07	Andersen (1991)
V380 Cyg	12.426	5.68	B1.5 III	13.13 ± 0.24	16.22 ± 0.26	4.337 ± 0.006	4.730 ± 0.03	Pavlovski et al. (2009)
			B2 V	7.779 ± 0.095	4.060 ± 0.084	4.334 ± 0.011	3.510 ± 0.04	
DW Car	1.328	9.68	B1 V	11.34 ± 0.12	4.558 ± 0.045	4.446 ± 0.016	4.055 ± 0.063	Southworth and Clausen (2007)
			B1 V	10.63 ± 0.14	4.297 ± 0.055	4.423 ± 0.016	3.915 ± 0.067	Clausen et al. (2007)
QX Car	4.478	6.64	B2 V	9.267 ± 0.122	4.289 ± 0.091	4.446 ± 0.016	4.377 ± 0.009	Andersen et al. (1983)
			B2 V	8.480 ± 0.122	4.051 ± 0.091	4.423 ± 0.016	4.354 ± 0.010	

This compilation from Southworth (<http://www.astro.keele.ac.uk/~jkt/debcats/>) represents an update from Andersen (1991)



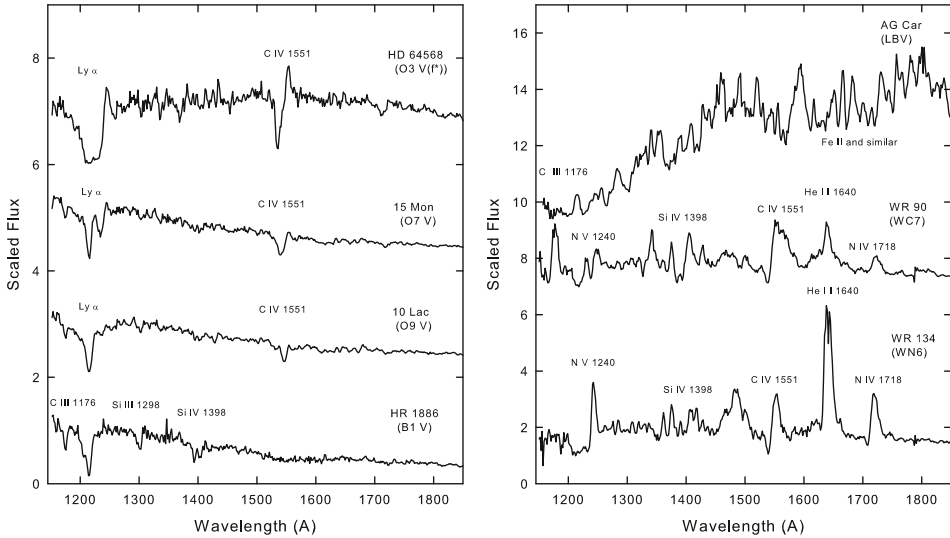
■ Fig. 9-2

The upper HR diagram showing the locations of the massive stars discussed in this chapter and their evolved counterparts. Main-sequence spectral types are indicated. The stars with well-determined masses listed in [Table 9-1](#) are shown with *filled circles*. Also shown (*filled diamonds*) are the locations of two well-studied LBVs, AG Car, and HR Car, from Groh et al. (2009). The Humphreys–Davidson limit is shown (*dashed line*) as is the Cepheid instability strip. The *thick, diagonal line* to the right of the upper main sequence shows the range of stars that likely develop into the Wolf–Rayet objects. The evolutionary tracks shown in the background are from the Geneva group’s website in 2011

2.1 Main-Sequence O and Early B Stars

The spectra of the nonevolved early B stars are distinguished in both the optical and UV by prominent lines of H, He I, and Si III. Fe III and other doubly ionized species from the Fe group pervade the FUV/NUV spectrum. Furthermore C III, N III, and Si IV lines are strong in the FUV for objects near B0. O II features are prominent in the optical spectra of the B1–B2 stars. Early B stars tend to have moderate–high rates of rotation and thus typically display rotationally broadened lines that render it difficult to determine the elemental abundances in their atmospheres and sometimes the radial velocity curves for binary systems. It is unclear whether the ultrasharp-lined B stars that are so useful for abundance investigations are intrinsically slow rotators or just viewed pole-on. Violet asymmetries in the FUV resonance lines reveal weak winds in B0–B0.5 stars. Be stars are rapidly rotating B stars which develop CS disks that are variable on time scales of minutes to decades.⁸ There remains debate on whether these objects

⁸See Porter and Rivinius (2003) for historical information on these interesting objects and Neiner (2011) for a recent update on their properties, evolution, and activity.



■ Fig. 9-3

The FUV spectra of a sample of massive main-sequence stars and some of their evolved products. Data are low-resolution images (6 \AA resolution) from the *International Ultraviolet Explorer*. Main-sequence stars are shown in the *left panel* and evolved objects in the *right panel*. Some of the more prominent spectral features are labeled

are rotating at or near their critical velocities. The earliest Be stars will evolve like the stars discussed in this chapter and provide a good test for stellar evolution calculations that include the effects of rotation.

In the O stars, wind effects steadily increase with earlier spectral type and by O5 the atmosphere is entirely dynamic (see ► Fig. 9-3). The earliest spectral type O2 was defined by Walborn et al. (2002). All of the stars at the upper end of the main sequence show emission in one or more spectral features. Oe stars with He II $\lambda 4686$ and the N III triplet $\lambda\lambda 4634\text{--}42$ in emission are historically designated as Of stars. These are very young massive stars rapidly evolving from their short life on the main sequence. For detailed information on the optical and FUV spectral characteristics of the main-sequence O and B stars, as well as their evolved counterparts, the reader is referred to the book by Gray and Corbally (2009). The X-ray properties of a large number of OB stars in the Carina star formation complex can be found in Nazé et al. (2011) and Gagné et al. (2011), and other articles in the 2011 special issue of the *Astrophysical Journal Supplement* on observations of the Carina region with the *Chandra* X-ray Observatory. From 60 single O stars, Gagné et al. determined that $\log(L_X/L_{\text{BOL}}) = -7.26 \pm 0.21$. The observed properties of the massive star populations in Local Group galaxies are detailed in a review by Massey (2003).

2.2 The Wolf–Rayet Stars

The Wolf–Rayet stars are the evolved products of O stars. The literature on these objects is extensive, and the reader is referred to the review by Crowther (2007). WR spectra are characterized

by strong broad emission lines (● Fig. 9-3) that can be classified into two basic sequences. The spectra of one group are dominated by lines of He and N (the WN stars), while the other shows lines of He, C, and O (the WC stars). Each branch is further classified into subtypes (WN3–WN8, WC4–WC9) based upon ionization criteria (Gray and Corbally 2009) where the smaller subtype means higher-ionization species dominate. The stars which show high-ionization emission lines are called “early” types and those with less ionization are designated “late” types and are designated as WNE, WCE, WNL, or WCL. Within the past 20 years, another branch that shows lines of O VI, O V, and O IV, designated WO (subtypes WO1–WO5), has been identified but these stars appear to be just higher-ionization versions of the WC objects (Crowther et al. 1998). Sometimes, a hybrid WN/C star is observed, but such objects are probably WN stars with an enhanced carbon abundance (Massey and Grove 1989). The spectra of WR stars are usually pure emission except for occasional P-Cygni absorption components formed in the star’s wind. Normal photospheric absorption lines are not observed in the WR stars.

The masses of WR stars range from 10 to 25 M_{\odot} , though these are objects that have undergone substantial mass loss from the time of their formation as O stars. Their spectral features are formed entirely in an optically thick wind, and the dominant chemical species observed depends on which by-products from the nucleosynthesis in the star’s interior are currently being circulated up to the surface. The WN stars have lost their H envelopes and are now exposing the N-rich material from the CNO cycle. Further mass loss produces a WC/WO star as the carbon-rich material from the triple-alpha process is now exposed. The original mass of the star appears to be important in determining which type of WR star is produced. The longer a star survives, the more of the star’s interior material we will observe. Crowther (2007) suggests that O stars more massive than 75 M_{\odot} will evolve as:

$$\text{O} \rightarrow \text{WN(H-rich)} \rightarrow \text{LBV} \rightarrow \text{WN(H-poor)} \rightarrow \text{WC} \rightarrow \text{SN Ic.}$$

If the initial mass is in the range 40–75 M_{\odot} :

$$\text{O} \rightarrow \text{LBV} \rightarrow \text{WN(H-poor)} \rightarrow \text{WC} \rightarrow \text{SN Ic.}$$

O stars from 25–40 M_{\odot} will evolve as:

$$\text{O} \rightarrow \text{LBV/RSG} \rightarrow \text{WN(H-poor)} \rightarrow \text{WC} \rightarrow \text{SN Ib.}$$

2.3 The Luminous Blue Variables and Early-Type Supergiants

Luminous blue variables, also known as Hubble–Sandage variables or S Dor stars, are evolved O stars observed at the top region of the HR diagram. This class includes familiar objects such as η Car and P Cyg. Other well-studied LBVs are AG Car, HR Car, and the Pistol star. Less than 20 LBVs have been studied in any detail. At one time, η Car and the Pistol star may have been the most massive stars in our galaxy. Properties of this class of objects are discussed by Smith (2011), Smith et al. (2010), Humphreys and Davidson (1994), and Crowther (2007).

The hallmark of LBVs is their discrete outbursts (Smith 2011; Gray and Corbally 2009). Some such as η Car and P Cyg have displayed giant eruptive events in recent history where it is estimated that 0.1–10 M_{\odot} of material may have been ejected. Since the luminosities of LBVs are very close to the Eddington limit, the eruptions may have been triggered when the star’s flux exceeded their Eddington luminosity. But the origin of the LBV eruptions remains unknown.

The radiation pressure alone appears not to be capable of producing such an abrupt mass loss. Perhaps, magnetic fields are important.

LBVs bear some similarity to WR stars. Their spectra are H poor and primarily composed of emission lines formed in an optically thick wind. They are probably the progenitors to the WN stars that are deficient in hydrogen (Crowther 2007). Due to the rarity of LBVs, some researchers have suggested that they represent a brief transitory evolutionary phase between main-sequence O stars and the core He-burning WR stars, while others caution that a large number of LBVs have not yet been identified because their defining characteristic, the eruptions, are infrequent (Smith 2011). According to Smith, the original masses of some LBVs could be as low as 20–25 M_{\odot} , in which case they have lost significant mass during an earlier RSG phase. But many LBVs probably were the most massive O stars ever formed. Observations of LBVs are incredibly important for gaining insight into how the very massive stars evolve. They can give us information on the extent of mass loss, wind variability and structure (including degree of clumpiness), binarity, and presence of magnetic fields and pulsations.

LBVs fall in the same domain in the HR diagram as the O, B, and A supergiants. Characteristics of the latter group of variable stars are given in Gray and Corbally (2009) and Howarth and Prinja (1989). Wind variability in the B supergiants is varied and substantial (Prinja et al. 2002). $H\alpha$ is typically in emission in the late B and A supergiants. Most of these objects probably originated from late O stars.

3 Contemporary Calculations of Evolutionary Tracks

3.1 Stellar Evolution Models

Stellar evolution models require a wide range of input physics ranging from nuclear reaction rates to mass loss prescriptions. In this section, we review the basic equations that govern the structure and evolution of stars as well as some of the key input physics with a special emphasis on rotation.

3.1.1 Stellar Structure Equations

There are four equations describing the evolution of the structure of stars: the mass, momentum, and energy conservation equations and the energy transport equations, which we recall below. In addition, the equations for the evolution of chemical abundances must be followed. These equations are discussed in the next section. In the Geneva stellar evolution code (GENEC, see Eggenberger et al. 2007), which we base our presentation on in this section, the problem is treated in one dimension (1D), and the equations of the evolution of chemical elements abundances are calculated separately from the structure equations, as in the original version of Kippenhahn and Weigert (Kippenhahn et al. 1967; Kippenhahn and Weigert 1990). In GENEC, rotation is included and spherical symmetry is no longer assumed. The effective gravity (sum of the centrifugal force and gravity) can in fact no longer be derived from a potential, and the case is said to be nonconservative. The problem can still be treated in 1D by assuming that the angular velocity is constant on isobars. This assumes that there is a strong horizontal (along isobars) turbulence which enforces constant angular velocity on isobars (Zahn 1992). The case

is referred to as “shellular” rotation and using reasonable simplifications described in Meynet and Maeder (1997), the usual set of four structure equations (as used for nonrotating stellar models) can be recovered:

- Energy conservation:

$$\frac{\partial L_P}{\partial M_P} = \epsilon_{\text{nucl}} - \epsilon_v + \epsilon_{\text{grav}} = \epsilon_{\text{nucl}} - \epsilon_v - c_P \frac{\partial \bar{T}}{\partial t} + \frac{\delta}{\bar{\rho}} \frac{\partial P}{\partial t}, \quad (9.1)$$

where L_P is the luminosity, M_P the Lagrangian mass coordinate, and ϵ_{nucl} , ϵ_v , and ϵ_{grav} are the energy generation rates per unit mass for nuclear reactions, neutrinos, and gravitational energy changes due to contraction or expansion, respectively. T is the temperature, c_P the specific heat at constant pressure, t the time, P the pressure, ρ the density, and $\delta = -\partial \ln \rho / \partial \ln T$.

- Momentum equation:

$$\frac{\partial P}{\partial M_P} = -\frac{GM_P}{4\pi r_P^4} f_P, \quad (9.2)$$

where r_P is the radius of the shell enclosing mass M_P and G the gravitational constant. f_P is defined below.

- Mass conservation (continuity equation):

$$\frac{\partial r_P}{\partial M_P} = \frac{1}{4\pi r_P^2 \bar{\rho}}, \quad (9.3)$$

- Energy transport equation:

$$\frac{\partial \ln \bar{T}}{\partial M_P} = -\frac{GM_P}{4\pi r_P^4} f_P \min \left[\nabla_{\text{ad}}, \nabla_{\text{rad}} \frac{f_T}{f_P} \right], \quad (9.4)$$

where

$$\begin{aligned} \nabla_{\text{ad}} &= \left(\frac{\partial \ln \bar{T}}{\partial \ln P} \right)_{\text{ad}} = \frac{P\delta}{\bar{T}\bar{\rho}c_P} \quad (\text{convective zones}), \\ \nabla_{\text{rad}} &= \frac{3}{64\pi\sigma G} \frac{\kappa L_P P}{M_P \bar{T}^4} \quad (\text{radiative zones}), \end{aligned}$$

where κ is the opacity and σ is the Stefan–Boltzmann constant.

$$\begin{aligned} f_P &= \frac{4\pi r_P^4}{GM_P S_P} \frac{1}{< g^{-1} >}, \\ f_T &= \left(\frac{4\pi r_P^2}{S_P} \right)^2 \frac{1}{< g > < g^{-1} >}, \end{aligned}$$

$< x >$ is x -averaged on an isobaric surface, \bar{x} is x -averaged in the volume separating two successive isobars, and the index P refers to the isobar with a pressure equal to P . g is the effective gravity and S_P is the surface of the isobar (see Meynet and Maeder 1997 for additional details). The implementation of the structure equations into other stellar evolution codes are presented for example in Paxton et al. (2011) and Chieffi et al. (1998).

3.1.2 Rotation and Magnetic Fields

The physics of rotation included in stellar evolution codes has been developed extensively over the last 20 years. A recent review of this development can be found in Maeder and Meynet (2012). The effects induced by rotation can be divided into three categories:

1. *Hydrostatic effects*: The centrifugal force changes the hydrostatic equilibrium of the star. The star becomes oblate and the equations describing the stellar structure must be modified as described above.
2. *Mass loss enhancement and anisotropy*: Mass loss depends on the opacity and the effective gravity (sum of gravity and centrifugal force) at the surface. The larger the opacity, the larger the mass loss. The higher the effective gravity, the higher the radiative flux (von Zeipel 1924) and effective temperature. Rotation, via the centrifugal force, reduces the surface effective gravity at the equator compared to the pole. As a result, the radiative flux of the star is larger at the pole than at the equator. In massive hot stars, since the opacity is dominated by the temperature-independent electron scattering, rotation enhances mass loss at the pole. If the opacity increases when the temperature decreases (cooler stars), mass loss can be enhanced at the equator.

For rotating models, the mass-loss rates can be obtained by applying a correction factor to the radiative mass-loss rate as described in Maeder and Meynet (2000b):

$$\dot{M}(\Omega) = F_{\Omega} \cdot \dot{M}(\Omega = 0) = F_{\Omega} \cdot \dot{M}_{\text{rad}}$$

$$\text{with } F_{\Omega} = \frac{(1 - \Gamma)^{\frac{1}{\alpha} - 1}}{\left[1 - \frac{\Omega^2}{2\pi G \rho_m} - \Gamma\right]^{\frac{1}{\alpha} - 1}}, \quad (9.5)$$

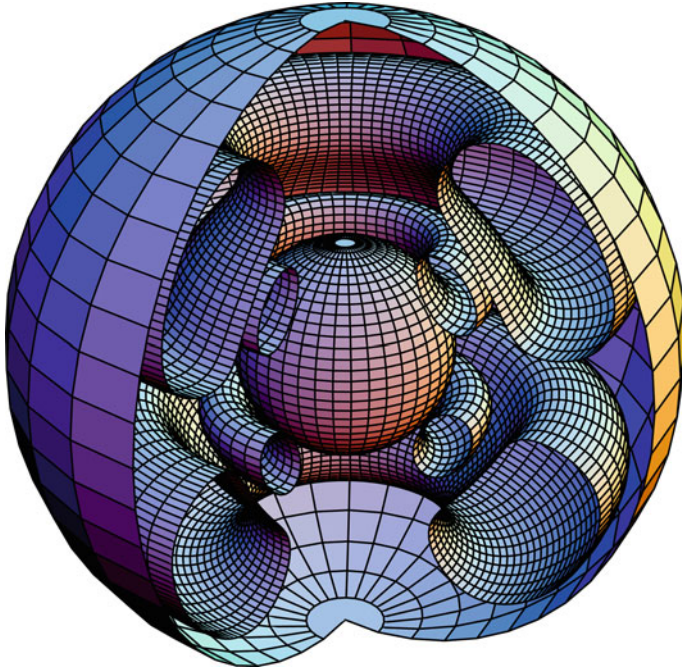
where Ω represents the angular velocity of the star, $\Gamma = L/L_{\text{Ed}} = \kappa L/(4\pi cGM)$ is the Eddington factor (κ is the electron-scattering opacity), and α is the force multiplier parameter depending on T_{eff} .

For mass-loss rates, $\dot{M}(\Omega = 0)$, the following prescriptions are commonly used: Vink et al. (2001) for radiatively driven wind of O-type stars, Nugis and Lamers (2000) for Wolf-Rayet stars, and de Jager et al. (1988) for cooler stars not covered by the other two prescriptions and for which dust and pulsation could play a role in the driving of the wind.

3. *Rotation driven instabilities*: The main rotation driven instabilities are horizontal turbulence, meridional circulation, and dynamical and secular shear (see Maeder 2009 for a comprehensive description of rotation-induced instabilities).

Horizontal turbulence corresponds to turbulence along the isobars. If this turbulence is strong, rotation is constant on isobars and the situation is usually referred to as “shellular rotation” (Zahn 1992). The horizontal turbulence is expected to be stronger than the vertical turbulence because there is no restoring buoyancy force along isobars (see Maeder 2003 for recent development on this topic).

Meridional circulation, also referred to as Eddington–Sweet circulation, arises from the local breakdown of radiative equilibrium in rotating stars. This is due to the fact that surfaces of constant temperature do not coincide with surfaces of constant pressure. Indeed, since rotation elongates isobars at the equator, the temperature on the same isobar is lower at the equator than at the pole. This induces large-scale circulation of matter, in which matter usually rises at the pole and descends at the equator (see Fig. 9-4). In this situation, angular momentum is transported inward. It is however also possible for the circulation to go in the reverse direction,



■ Fig. 9-4

Streamlines of meridional circulation in a rotating $20 M_{\odot}$ model with solar metallicity and $v_{\text{ini}} = 300 \text{ km s}^{-1}$ at the beginning of the H-burning phase. The streamlines are in the meridian plane. In the upper hemisphere on the right section, matter is rotating counterclockwise along the outer stream line and clockwise along the inner one. The outer sphere is the star's surface and has a radius equal to $5.2 R_{\odot}$. The inner sphere shows the outer boundary of the convective core, which has a radius of $1.7 R_{\odot}$ (Illustration from Meynet and Maeder 2002)

and in this second case, angular momentum is transported outward. Circulation corresponds to an advective process, which is different from diffusion because the latter can only erode gradients. Advection can either build or erode angular velocity gradients (see Maeder and Zahn 1998 for more details).

Dynamical shear occurs when the excess energy contained in differentially rotating layers is larger than the work that needs to be done to overcome the buoyancy force. The criterion for stability against dynamical shear instability is the Richardson criterion:

$$Ri = \frac{N^2}{(\partial U / \partial z)^2} > \frac{1}{4} = Ri_c, \quad (9.6)$$

where U is the horizontal velocity, z the vertical coordinate, and N^2 the Brunt–Väisälä frequency.

The critical value of the Richardson criterion, $Ri_c = 1/4$, corresponds to the situation where the excess kinetic energy contained in the differentially rotating layers is equal to the work done against the restoring force of the density gradient (also called buoyancy force). It is therefore used by most authors as the limit for the occurrence of the dynamical shear. However, recent

studies by Canuto (2002) show that turbulence may occur as long as $Ri \lesssim Ri_c \sim 1$. This critical value is consistent with numerical simulations done by Brüggén and Hillebrandt (2001) where they find shear mixing for values of Ri greater than $1/4$ (up to about 1.5).

Different dynamical shear diffusion coefficients, D , can be found in the literature. The one used in GENEC is as follows:

$$D = \frac{1}{3} \nu l = \frac{1}{3} \frac{\nu}{l} l^2 = \frac{1}{3} r \frac{d\Omega}{dr} \Delta r^2 = \frac{1}{3} r \Delta\Omega \Delta r, \quad (9.7)$$

where r is the mean radius of the zone where the instability occurs, $\Delta\Omega$ is the variation of Ω over this zone, and Δr is the extent of the zone. The zone is the reunion of consecutive shells where $Ri < Ri_c$ (see Hirschi et al. 2004 for more details and references).

If the differential rotation is not strong enough to induce dynamical shear, it can still induce the secular shear instability when thermal turbulence reduces the effect of the buoyancy force. The secular shear instability occurs therefore on the thermal time scale, which is much longer than the dynamical one. Note that the way the inhibiting effect of the molecular weight (μ) gradients on secular shear is taken into account impacts strongly the efficiency of the shear. In some work, the inhibiting effect of μ -gradients is so strong that secular shear is suppressed below a certain threshold value of differential rotation (Heger et al. 2000). In other work (Maeder 1997), thermal instabilities and horizontal turbulence reduce the inhibiting effect of the μ -gradients. As a result, shear is not suppressed below a threshold value of differential rotation but only decreased when μ -gradients are present.

There are other minor instabilities induced by rotation: the GSF instability (Goldreich and Schubert 1967; Fricke 1968; Hirschi and Maeder 2010), the ABCD instability (Knobloch and Spruit 1983; Heger et al. 2000), and the Solberg–Høiland instability (Kippenhahn and Weigert 1990). The GSF instability is induced by axisymmetric perturbations. The ABCD instability is a kind of horizontal convection. Finally, Solberg–Høiland stability criterion is the criterion that should be used instead of the Ledoux or Schwarzschild criterion in rotating stars. However, including the dynamical shear instability also takes into account the Solberg–Høiland instability (Hirschi et al. 2004).

Transport of Angular Momentum

For shellular rotation, the equation of transport of angular momentum (Zahn 1992) in the vertical direction is (in lagrangian coordinates) as follows:

$$\rho \frac{d}{dt} (r^2 \Omega)_{M_r} = \frac{1}{5r^2} \frac{\partial}{\partial r} (\rho r^4 \Omega U(r)) + \frac{1}{r^2} \frac{\partial}{\partial r} \left(\rho D r^4 \frac{\partial \Omega}{\partial r} \right), \quad (9.8)$$

where $\Omega(r)$ is the mean angular velocity at level r , $U(r)$ the vertical component of the meridional circulation velocity, and D the diffusion coefficient due to the sum of the various turbulent diffusion processes (convection, shears, and other rotation-induced instabilities apart from meridional circulation). Note that angular momentum is conserved in the case of contraction or expansion. The first term on the right-hand side, corresponding to meridional circulation, is an *advective* term. The second term on the right-hand side, which corresponds to the diffusion processes, is a *diffusive* term. The correct treatment of advection is very costly numerically because (9.8) is a fourth order equation (the expression of $U(r)$ contains third order derivatives of Ω , see Zahn 1992). This is why some research groups treat meridional circulation in a diffusive way (see, e.g., Heger et al. 2000) with the risk of transporting angular momentum in the wrong direction (in the case meridional circulation builds gradients).

Transport of Chemical Species

The transport of chemical elements is also governed by a diffusion–advection equation like (9.8). However, if the horizontal component of the turbulent diffusion is large, the vertical advection of the elements (and not that of the angular momentum) can be treated as a simple diffusion (Chaboyer and Zahn 1992) with a diffusion coefficient D_{eff} ,

$$D_{\text{eff}} = \frac{|rU(r)|^2}{30D_h}, \quad (9.9)$$

where D_h is the coefficient of horizontal turbulence (Zahn 1992). Equation 9.9 expresses that the vertical advection of chemical elements is severely inhibited by the strong horizontal turbulence characterized by D_h . The change of the mass fraction X_i of the chemical species i is simply

$$\left(\frac{dX_i}{dt}\right)_{M_r} = \left(\frac{\partial}{\partial M_r}\right)_t \left[(4\pi r^2 \rho)^2 D_{\text{mix}} \left(\frac{\partial X_i}{\partial M_r}\right)_t \right] + \left(\frac{dX_i}{dt}\right)_{\text{nuclear}}, \quad (9.10)$$

where the second term on the right accounts for composition changes due to nuclear reactions. The coefficient D_{mix} is the sum $D_{\text{mix}} = D + D_{\text{eff}}$, where D is the term appearing in (9.8) and D_{eff} accounts for the combined effect of advection and horizontal turbulence.

Rotation and Magnetic Fields

Spectropolarimetric surveys have obtained evidence for the presence of magnetic field at the surface of OB stars (see the recent review by Walder et al. 2011 and references therein). The origin of these magnetic fields is still unknown. They might be fossil fields (the spectral characteristics of Ofp stars are indicative of organized magnetic fields, most likely of a fossil origin) or fields produced through a dynamo mechanism.

The central question for the evolution of massive stars is whether a dynamo is at work in internal radiative zones. This could have far-reaching consequences concerning the mixing of the elements and the loss of angular momentum. In particular, the interaction between rotation and magnetic fields in the stellar interior strongly affects the angular momentum retained in the core and thus the initial rotation rate of pulsars and which massive stars could end their lives as long and soft gamma-ray bursters (see the discussion in Sect. 6 of Georgy et al. 2012 and references therein).

The interplay between rotation and magnetic field has been studied in stellar evolution calculations using the Tayler–Spruit dynamo (Spruit 2002; Maeder and Meynet 2005). Some numerical simulations confirm the existence of a magnetic instability; however, the existence of the dynamo is still controversial (Braithwaite 2006; Zahn et al. 2007).

The Tayler–Spruit dynamo is based on the fact that a purely toroidal field $B_\varphi(r, \vartheta)$, even very weak, in a stable stratified star is unstable on an Alfvén timescale $1/\omega_A$. This is the first magnetic instability to appear. It is nonaxisymmetric of type $m = 1$ (Spruit 2002), occurs under a wide range of conditions, and is characterized by a low threshold and a short growth time. In a rotating star, the instability is also present, however the growth rate σ_B of the instability is, if $\omega_A \ll \Omega$,

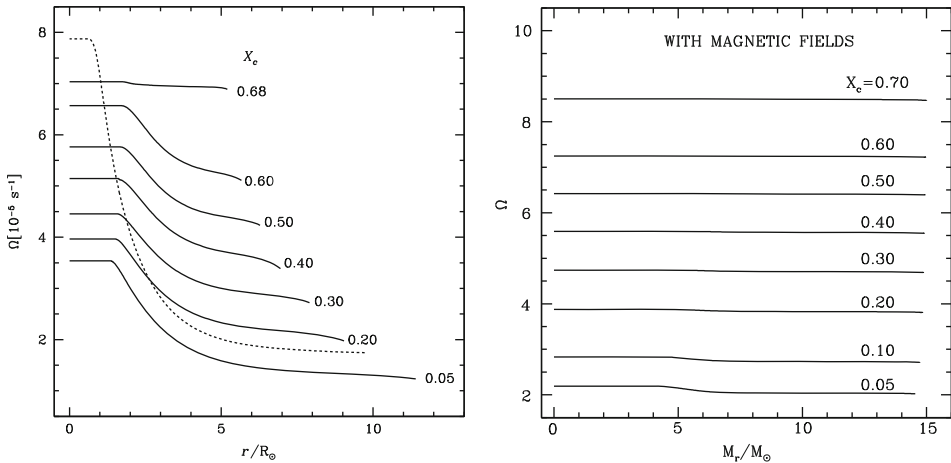
$$\sigma_B = \frac{\omega_A^2}{\Omega}, \quad (9.11)$$

instead of the Alfvén frequency ω_A , because the growth rate of the instability is reduced by the Coriolis force (Spruit 2002). One usually has the following ordering of the different frequencies,

$N \gg \Omega \gg \omega_A$. In the Sun, one has $N \approx 10^{-3} \text{ s}^{-1}$, $\Omega = 3 \times 10^{-6} \text{ s}^{-1}$ and a field of 1 kG would give an Alfvén frequency as low as $\omega_A = 4 \times 10^{-9} \text{ s}^{-1}$ (where N^2 is the Brunt–Väisälä frequency).

This theory enables us to establish the two quantities in which we are mainly interested for stellar evolution: the magnetic viscosity ν , which expresses the mechanical coupling due to the magnetic field \vec{B} , and the magnetic diffusivity η , which expresses the transport by a magnetic instability and thus also the damping of the instability. The parameter η also expresses the vertical transport of the chemical elements and enters (9.10), while the viscosity ν determines the vertical transport of the angular momentum by the magnetic field and enters the second term on the right-hand side of (9.8).

► Figure 9-5 shows the differences in the internal Ω -profiles during the evolution of a $20 M_\odot$ star with and without a magnetic field created by the Tayler–Spruit dynamo. Without a magnetic field, the star has a significant differential rotation, while Ω is almost constant when a magnetic field created by the dynamo is present. It is not perfectly constant, otherwise there would be no dynamo. In fact, the rotation rapidly adjusts itself to the minimum differential rotation necessary to sustain the dynamo. One could then assume that the mixing of chemical elements is suppressed by magnetic fields. This is, however, not the case since the interplay between magnetic fields and the meridional circulation tend to lead to more mixing in models including magnetic fields compared to models not including magnetic fields (Maeder and Meynet 2005). Fast-rotating models of GRB progenitors calculated by Yoon et al. (2006) also experience a strong chemical internal mixing leading to the stars undergoing quasi-chemical homogeneous evolution. The study of the interaction between rotation and magnetic fields is still under development, and the next 10 years will certainly provide new insights on this important topic.



■ Fig. 9-5

Left: evolution of the angular velocity Ω as a function of the distance to the center in a $20 M_\odot$ star with $v_{\text{ini}} = 300 \text{ km s}^{-1}$. X_c is the hydrogen mass fraction at the center. The *dotted line* shows the profile when the He core contracts at the end of the H-burning phase. **Right:** rotation profiles at various stages of evolution (labeled by the central H content X_c) of a $15 M_\odot$ model with $X = 0.705$, $Z = 0.02$, an initial velocity of 300 km s^{-1} and a magnetic field from the Tayler–Spruit dynamo (From Maeder and Meynet 2005)


3.1.3 Other Input Physics


The other key input physics that are essential for the computation of stellar evolution models are nuclear reactions, mass loss prescriptions (discussed above), the equation of state, opacities, and neutrino losses. Stellar evolution codes are now able to include a larger and more flexible nuclear reaction network (see, e.g., Frischknecht et al. 2010 for a description of the implementation of a flexible network in GENEC). Nuclear physics and other inputs are described for other codes, for example in Paxton et al. (2011) and Chieffi et al. (1998).


3.2 Grids of Evolutionary Tracks and Stellar Populations

The evolution of a star is primarily determined by its initial mass. Rotation, metallicity, and mass loss also play an important role for the evolution of massive stars. Before presenting the evolution of the entire range of massive stars, we first describe the effects due to the other main parameters: rotation, metallicity, and mass loss.

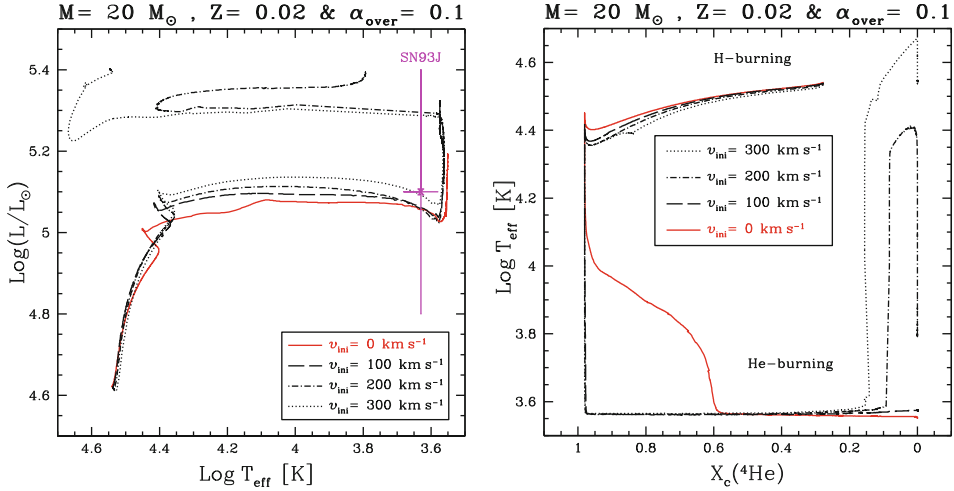
3.2.1 Impact of Rotation

As explained above, rotation affects the structure and evolution of massive stars in three ways: hydrostatic effects, enhanced mass loss, and internal mixing of chemical elements. The last effect is the most important. In order to see the impact of rotation-induced mixing, we can look at the evolution of $20 M_{\odot}$ models without rotation and with initial rotation velocities of 100, 200, and 300 km s^{-1} , with 300 km s^{-1} representing the average initial rotation rate for massive stars (Hirschi et al. 2004). The tracks of the $20 M_{\odot}$ models are presented in  Fig. 9-6.

 Figure 9-6 (left) shows the evolutionary tracks of the four different $20 M_{\odot}$ models in the HR diagram. The nonrotating model ends up as a red supergiant (RSG) like the corresponding model calculated by other groups (see Heger and Langer 2000; Limongi et al. 2000). However, the rotating models show very interesting features. Although the 100 km s^{-1} model remains a RSG, the 200 km s^{-1} model undergoes a blue loop to yellow-red supergiant, whereas the 300 km s^{-1} model ends up as a blue supergiant (BSG). Thus, rotation may have a strong impact on the nature of the supernova progenitor (red, blue supergiant, or even Wolf-Rayet star) and thus on some observed characteristics of the supernova explosion.

When does a massive star evolve back to the blue after a RSG phase and why?  Figure 9-6 (right) shows the evolution of $\log(T_{\text{eff}})$ versus the central helium abundance, $X_c(^4\text{He})$. $X_c(^4\text{He})$ increases during the main sequence, then decreases during He burning, and finally is equal to zero during the post He-burning evolution. The figure shows that:

- For the nonrotating model, He burning starts when the star crosses the HR diagram ($\log(T_{\text{eff}}) \sim 4$) and the star only reaches the RSG stage halfway through He burning. Finally, the star luminosity rises during shell He burning.
- For the $v_{\text{ini}} = 300 \text{ km s}^{-1}$ model, the star is more luminous and becomes a RSG before He-burning ignition. These two factors favor higher mass-loss rates, and the star loses most of its hydrogen envelope before He burning is finished. Thus, the star evolves toward the zone of the HR diagram where homogeneous helium stars are found, i.e., in the blue part of the HR diagram. We can see that the evolutionary track still evolves during shell He burning. The higher luminosity also leads to more massive cores and rotating stars behave



■ Fig. 9-6

Left: HR-diagram for $20 M_{\odot}$ models: *solid, dashed, dotted-dashed, and dotted lines* correspond to $v_{\text{ini}} = 0, 100, 200$, and 300 km s^{-1} , respectively. We also indicate the position of the progenitor of SN193J. **Right:** T_{eff} versus central helium mass fraction for the same $20 M_{\odot}$ models (Figure taken from Hirschi et al. 2004)

as more massive nonrotating stars. As an example, a $20 M_{\odot}$ rotating stars will have core masses similar to a nonrotating stars with an initial mass roughly equal to $25 M_{\odot}$. On the other hand, rotation-induced mixing enhances the total lifetime of stars.

- The 100 km s^{-1} model luminosity is lower than for the 300 km s^{-1} model, and therefore less mass is lost during He burning and the burning ends before the hydrogen envelope is removed. The star therefore remains a RSG. The 200 km s^{-1} model evolution is similar to the 300 km s^{-1} model, but the extent of its blue loop is smaller. At the end of He burning for the 200 km s^{-1} model, $\log(T_{\text{eff}}) = 4.28$ and the star becomes redder before C burning starts.

Although the models discussed here are for solar metallicity, one can note that the behaviors of the models with v_{ini} between 200 and 300 km s^{-1} are reminiscent of the evolution of the progenitor of SN1987A. Let us recall that this supernova had a blue progenitor which evolved from a RSG stage (see, e.g., the review by Arnett et al. 1989). In Fig. 9-6, we also indicate the position of the progenitor of SN 193J. SN 193J probably belongs to a binary system (Podsiadlowski et al. 1993). Nevertheless, it has common properties with our $v_{\text{ini}} = 200 \text{ km s}^{-1}$ $20 M_{\odot}$ model: the stellar model and the progenitor of SN 193J have approximately the same metallicity, they have a similar position in the HR diagram taking into account the uncertainties, and they both have a small hydrogen rich envelope, making possible a change from type II to type Ib some time after the explosion.

Rotation-induced mixing also brings H-burning products from the core to the surface of the star, whereas in nonrotating models no enrichment takes place before the star reaches the RSG stage. In particular, rotating models predict a surface enrichment in nitrogen and helium. Rotation-induced mixing is questioned by recent observational campaigns (see e.g., Hunter et al. 2007), but the evidence is not yet conclusive and more work is required on this topic.

The enrichment in elements like nitrogen is accompanied by the depletion of fragile elements like boron. Although observations of boron are a challenging task, combining constraints from both nitrogen enhancements and boron depletion will help improve models in the future. This point is further discussed in [Sect. 4.1](#).

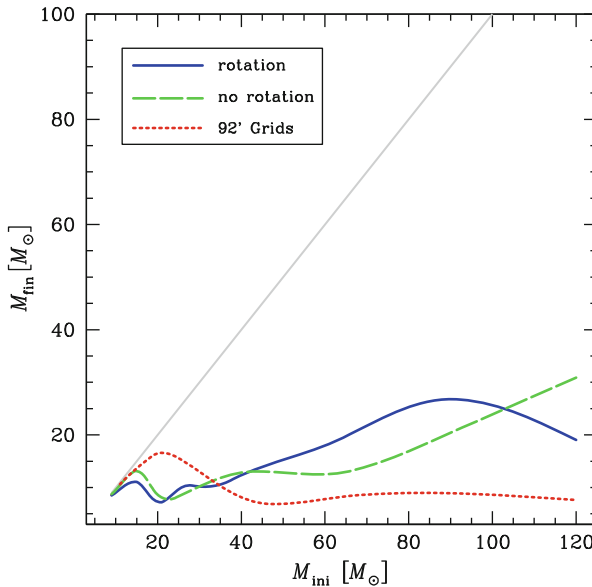
3.2.2 Impact of Mass Loss

The luminosity, L , of a star increases faster than its mass, M . For stars less massive than about $20 M_{\odot}$, $L \propto M^3$ and for more massive stars, the relationship reduces to $L \propto M$ for the most massive stars. The much higher luminosity of massive stars leads to strong radiation driven winds, which peel off the stars during its life.

► [Figure 9-7](#) compares the final versus initial mass relations. Apart from stars with masses less than $15 M_{\odot}$, we see that the final mass is significantly lower than the initial mass demonstrating the importance of mass loss for massive stars.

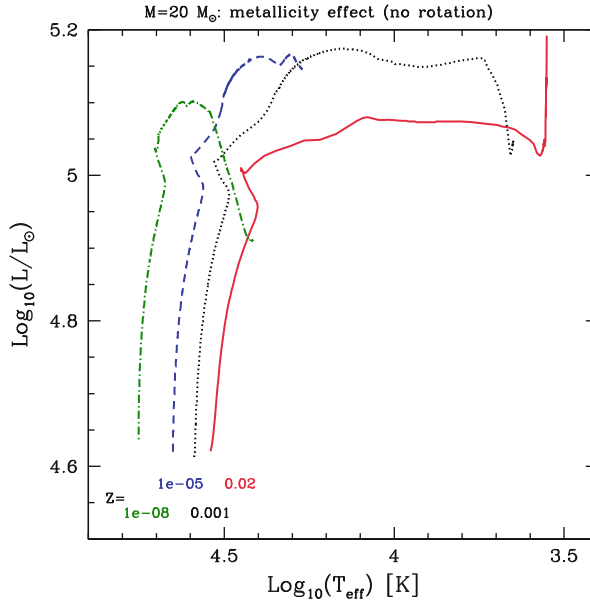
3.2.3 Impact of Metallicity

The effects of low metallicity on the evolution of massive stars are the following. A lower metallicity implies a lower luminosity which leads to slightly smaller convective cores. A lower metallicity implies lower opacity and lower mass losses (as long as the chemical composition has



■ Fig. 9-7

Final mass versus initial mass for models from 9 to $120 M_{\odot}$. Comparison between Ekström et al. (2012) nonrotating models (dashed green line) and rotating models (solid blue line), and the earlier grid of models from Schaller et al. (1992) (dotted red line). The gray line corresponds to a hypothetical case without mass loss ($M_{\text{fin}} = M_{\text{ini}}$) (Figure taken from Ekström et al. 2012)



■ Fig. 9-8


Evolution in the HR diagrams of nonrotating $20 M_{\odot}$ models at various metallicities (Figure taken from Hirschi 2007)

not been changed by burning or mixing in the part of the star one considers). Thus, at the start of the evolution, lower metallicity stars are more compact. This can be seen in the Hertzsprung–Russell (HR) diagram (► Fig. 9-8) where the lower metallicity models have bluer tracks during the main sequence. The lower metallicity models also have a harder time reaching the red supergiant (RSG) stage. The nonrotating model at $Z = 10^{-3}$ becomes a RSG only during shell He burning, and the lower metallicity nonrotating models never reach the RSG stage. As long as the metallicity is above about $Z = 10^{-10}$, no other significant differences have been found in nonrotating models. Below this metallicity and for metal-free stars, the CNO cycle cannot operate at the start of H burning. At the end of its formation, the star therefore contracts until it starts He burning because the pp chains cannot balance the effect of the gravitational force. Once enough carbon and oxygen are produced, the CNO cycle can operate and the star behaves like stars with $Z > 10^{-10}$ for the rest of the main sequence. Shell H burning still differs between $Z > 10^{-10}$ and metal-free stars.

Mass-loss rates generally depend on the metallicity of the star so we expect low Z stars to lose less mass than solar Z stars. This is especially the case for the wind of hot stars (O-type and WR stars, see, e.g., Vink et al. 2001). However, winds of cool stars (LBV and RSG) might not have such a strong metallicity dependence. There might be other phenomena leading to mass loss. In rotating stars, if the mass loss is small, angular momentum loss is also small and this angular momentum can accumulate near the surface of the star. Stars may thus rotate critically after some time, meaning that the centrifugal force balances gravity at the equator. Such phenomenon may cause low Z stars to lose significant amount of mass (Hirschi 2007), and it is important to better understand mass loss and its metallicity dependence in order to better predict the evolution of massive stars.

3.2.4 Grids of Models and Stellar Populations

Now that we have discussed the impact of the individual key parameters affecting the evolution of massive stars, let us describe the general evolution as a function of initial mass at solar metallicity.

Evolutionary tracks for a large grid of rotating models with masses ranging from 0.8 to $120 M_{\odot}$ (Ekström et al. 2012) are presented in  Fig. 9-9 (see Brott et al. 2011 for another grid of main-sequence models). Globally, for massive stars ($M > 8 M_{\odot}$), we can split stars in three mass ranges.

- For $M \gtrsim 50 M_{\odot}$: the high mass-loss rates remove enough mass so that stars lose their envelopes on the main-sequence or in the blue supergiant stage as LBVs. The stars never become red supergiants. Stars in this mass range may die as type Ib or Ic supernovae leaving a black hole as a remnant or collapse to a black hole without any supernova depending on the energy of the supernova engine.
- For $50 M_{\odot} \gtrsim M \gtrsim 20 M_{\odot}$: stars only lose a fraction of their envelopes on the main sequence. They further evolve to the red supergiant stage, where mass loss is sufficient to remove their envelope. They become bare cores and are observed as WR stars. Stars in this mass range probably die as SNIb (SNIc) leaving a black hole as a remnant or, as their more massive counterparts collapse to a black hole without undergoing a supernova.
- For $M \lesssim 20 M_{\odot}$: the stars still experience mass loss; however, it is not sufficient to alter the global evolution. The mass loss and rotation may nevertheless still modify the lifetimes and the chemical compositions. Stars in this mass range die as SN II (IIL or IIP) and form a neutron star.

The mass limits are uncertain and depend on metallicity Z . At different Z , some sequences may be absent. The last stage indicated before supernovae (SN) are usually reached near the end of central He burning. After this stage, the stellar envelopes do not evolve further and their properties determine the nature of the SN progenitors.

4 Observation Confronts Theory

There are many observational tests that can allow us to constrain stellar evolution models. In this chapter, we focus on the surface abundances in massive main-sequence stars, the properties and number of stars in the different post-main-sequence evolutionary stages (e.g., RSG and WR), and the fate of stars. To begin, let us summarize the evolutionary paths taken by single stars in various mass ranges with solar metallicity. Meynet et al. (2011) propose the following scenario:

$M > 90 M_{\odot}$: O – Of – WNL – (WNE) – WCL – WCE – SN(SNIbc/BH/SNIIn)? (PCSN/Hypernova)

$60\text{--}90 M_{\odot}$: O – Of/WNL \leftrightarrow LBV – WNL(H poor) – WCL-E – SN (SNIbc/BH/SNIIn)?

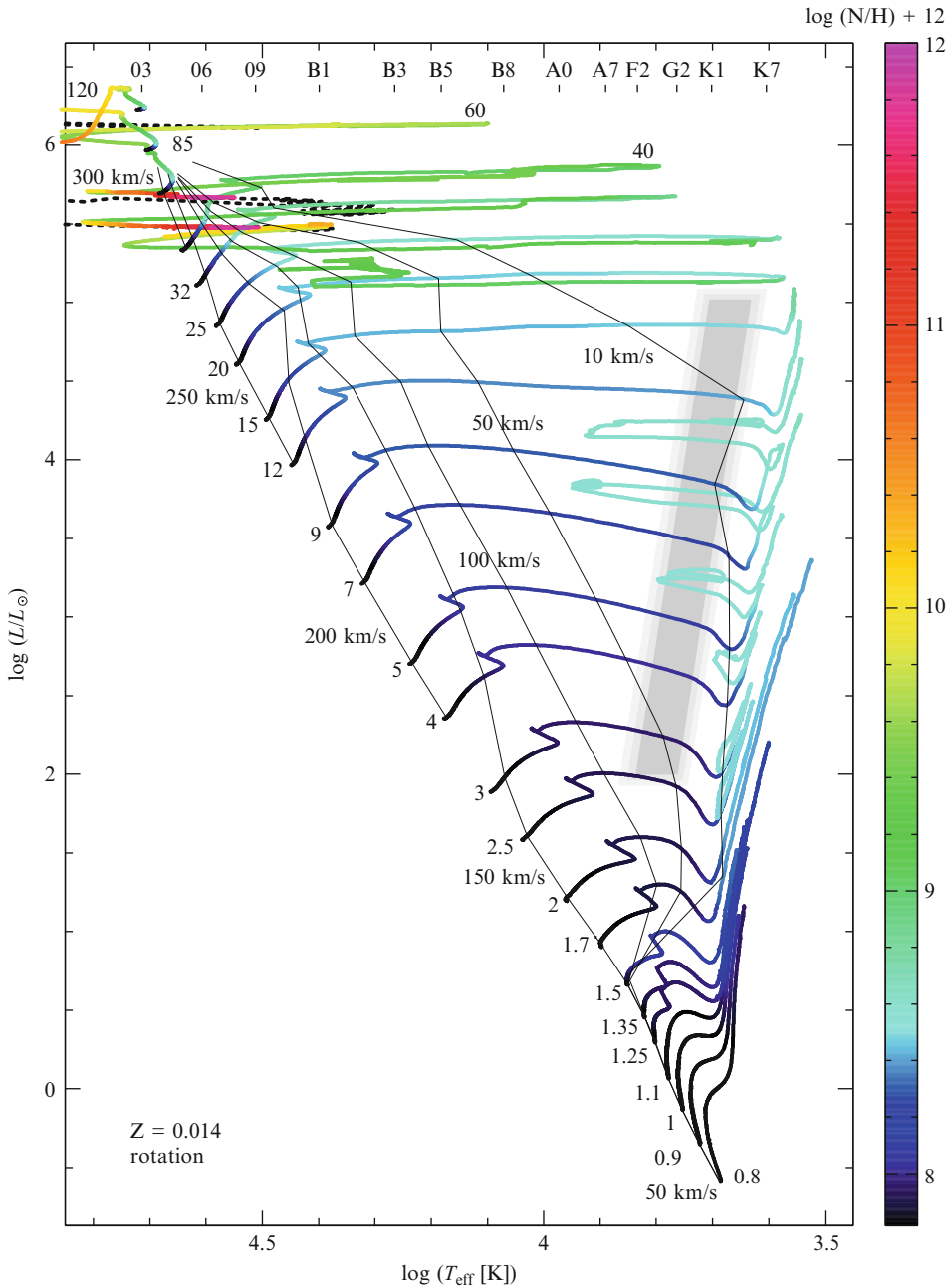
$40\text{--}60 M_{\odot}$: O – BSG – LBV \leftrightarrow WNL – (WNE) – WCL-E – SN (SNIb)
– WCL-E – WO – SN (SNIc)

$30\text{--}40 M_{\odot}$: O – BSG – RSG – WNE – WCE – SN (SNIb)

OH/IR \leftrightarrow LBV ?

$20\text{--}30 M_{\odot}$: O – (BSG) – RSG – BSG (blue loop) – RSG – SN (SNIb, SNIIL)

$10\text{--}20 M_{\odot}$: O – RSG – (Cepheid loop, $M < 15 M_{\odot}$) – RSG – SN (SNIIP)



■ Fig. 9-9

HR diagram for rotating models at solar metallicity (taken from Ekström et al. 2012, also presenting a HRD for nonrotating models). Lines of isovelocities are drawn through the diagram. The *color scale* indicates the surface number abundance of nitrogen in a log scale where the abundance of hydrogen is 12. Once the star has become a WNE type Wolf-Rayet star, the tracks are drawn with *black dotted lines*. The *gray shaded area* represents the Cepheid instability strip

The arrows \Leftrightarrow indicate that the star moves back and forth motions between the two stages. The paths and sequences depend on metallicity, Z , and rotation.

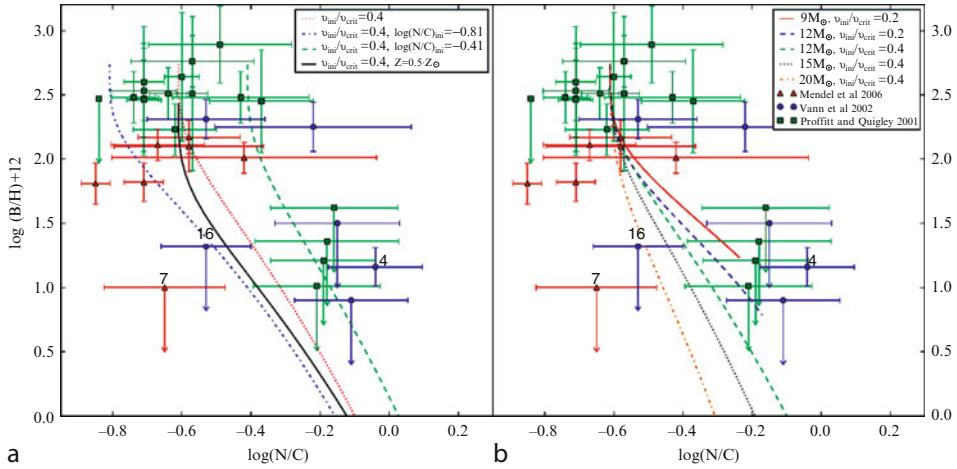
We can classify the different scenarios in three families: below about $30 M_{\odot}$, massive stars pass through a red supergiant stage; above about $40 M_{\odot}$, stars go through an LBV and WR phase; and in the intermediate mass range, between 30 and $40 M_{\odot}$, stars can possibly go through the three phases, RSG, LBV, and WR.

4.1 Surface Abundance of Light Elements in Massive Stars

Observations of light element abundance in main-sequence stars can test the treatment of transport of angular momentum and chemical species inside stellar models. Indeed, internal mixing, in particular due to rotation, predicts that the surface composition should be enhanced in ^{14}N and ^4He if mixing takes place in the radiative layer between the convective core and the surface of a massive star. New observational data of late O- and B-type stars from the VLT-FLAMES survey (Evans et al. 2005, 2006; Hunter et al. 2007; Trundle et al. 2007) lead to an intense discussion about how well the models with rotation can explain the observed nitrogen surface abundances (Hunter et al. 2008, 2009; Maeder et al. 2009) and about whether or not binaries are necessary to explain some groups of the observed O- and B-type stars (Langer et al. 2008). Light elements and in particular boron can constrain the mixing induced by rotation and help distinguish between single stars and interacting binaries (Brott et al. 2009). Boron is destroyed at relatively low temperatures ($\approx 6 \times 10^6$ K) where the CNO cycles are not yet efficient. Therefore, modest mixing due to rotation leads to a depletion of light elements at the surface without considerable nitrogen enrichment. This effect can not be explained by mass transfer in a binary system, since there, the accreted material is depleted in boron and enriched in nitrogen (Fliegner et al. 1996).

Boron is produced in spallation process of CNO atoms in the interstellar medium (ISM) by galactic cosmic rays. In massive stars, boron is only destroyed. It is the only light element out of Li, Be, and B which is observed at the surface of OB stars. Despite the difficulties of measuring boron surface abundances, an increasing number of boron abundances from O- and early B-type stars have recently been published (Mendel et al. 2006; Venn et al. 2002; Proffitt and Quigley 2001; Proffitt et al. 1999). The comparison in Mendel et al. (2006) of observational data with the models of Heger and Langer (2000) shows a good agreement with the exception of three stars (HD 30836, HD 36591, HD205021). The strong boron depletion in these young stars raises the question if the efficiency of surface mixing due to rotation is stronger or if there are other mixing processes at work. We reexamined this question because the Geneva stellar evolution code GENEC includes the effect of rotation in a different way compared with the codes used in previous works. The most important difference comes from the fact that in GENEC, the transport of the angular momentum is properly accounted for as an advection process and not as a diffusion process.

► **Figure 9-10** shows the observations of the B/H versus N/C ratio, as well as the evolutionary tracks of rotating models calculated with the Geneva code (see Frischknecht et al. 2010 for details). This diagram tests whether the concomitant changes of boron and nitrogen are reproduced by the rotating stellar models. In ► **Fig. 9-10a**, we plotted models with $v_{\text{ini}}/v_{\text{crit}} = 0.4$ to show the typical B-N/C relation. We see that these models could explain a large fraction of the observations, but they correspond to an average velocity on the main sequence of about 180 km s^{-1} , while the observational sample mainly consists of slow rotators ($v \sin i \leq 50 \text{ km s}^{-1}$). On the other hand, in panel b of ► **Fig. 9-10**, models with lower velocities are shown. We see



■ Fig. 9-10

Boron versus nitrogen over carbon for the observations and in (a) $12 M_{\odot}$ models with different initial compositions and in (b) models with different masses and low-to-intermediate velocities to show the possible spread in main-sequence evolution. The labeled stars correspond to objects that challenge the present theoretical predictions (see Frischknecht et al. 2010 for details and references)

that they would provide a good fit to the points with boron abundances between 0.9 and 1.5 and N/C ratios of the order of -0.2 . These models have difficulty explaining stars that are more depleted in boron and more enriched in nitrogen.

The two stars 7 and 16 (see ► Fig. 9-10) present no or little surface nitrogen enrichment and are strongly boron depleted. The range of values spanned by the different initial mass and initial velocity models barely explain these values therefore, these stars challenge rotating stellar models even though the difference in N/C is within 2σ of the models at the upper limit. These two stars were also found to be a problem for single-star models by other authors (Morel et al. 2008; Mendel et al. 2006; Venn et al. 2002). At this stage, it is also interesting to note that star 7, which cannot be reproduced by these models, is a spectroscopic binary with a period of 9.5 days (Pourbaix et al. 2004), so it might well be that this star was slowed down by spin-orbit coupling in the course of its main-sequence evolution. Such a braking mechanism would also slow down or even stop the N-enrichment process. Apart from these few difficult cases, the above comparisons are very encouraging on the whole and support the predictions of the rotating models.



The impact of rotation-induced mixing becomes stronger as the initial metallicity (Z) of the star decreases. This is particularly true for the abundance of so-called secondary elements like nitrogen. Secondary elements are produced from elements present in the cloud out of which the star formed. In the case of nitrogen, it is produced at the expense of carbon and oxygen in the CNO cycle. The production of secondary elements is proportional to the initial metallicity, and thus at very low Z , their production becomes negligible. In rotating models, however, rotation induces the mixing of carbon and oxygen from the helium-burning core to the hydrogen-burning shell, thus enabling a primary production of nitrogen (Meynet et al. 2006; Hirschi 2007). When yields from these rotating models are used as input in chemical evolution

models, a nice fit of the N/O in very metal poor halo stars (see Spite et al. 2005) is obtained (Chiappini et al. 2006). This provides a strong support for the occurrence of rotation-induced mixing at low Z and for our models. The primary nitrogen production in rotating low- Z stellar models is accompanied by the primary production of other isotopes like ^{13}C , and especially ^{22}Ne , which is the neutron source for s -process in massive stars (e.g. Käppeler et al. 2011 and references therein). Frischknecht et al. (2012) have calculated full stellar evolution models including the effects of rotation and a large nuclear reaction network and found that rotation boosts the s -process at low Z . The s -process in rotating massive stars thus contributes significantly to the production of elements up to strontium above $[\text{Fe}/\text{H}] \simeq -2$, the main limiting factor being the iron seeds at low Z . The production of elements up to barium is sensitive to the initial rotation rate, and we generally obtain $[\text{Sr}/\text{Ba}] > 0$ for the massive low- Z rotating stars. Finally, rotation-induced mixing is the key ingredient in progenitor models of long and soft gamma-ray bursts that undergo quasi-chemically homogeneous evolution (Yoon et al. 2006).

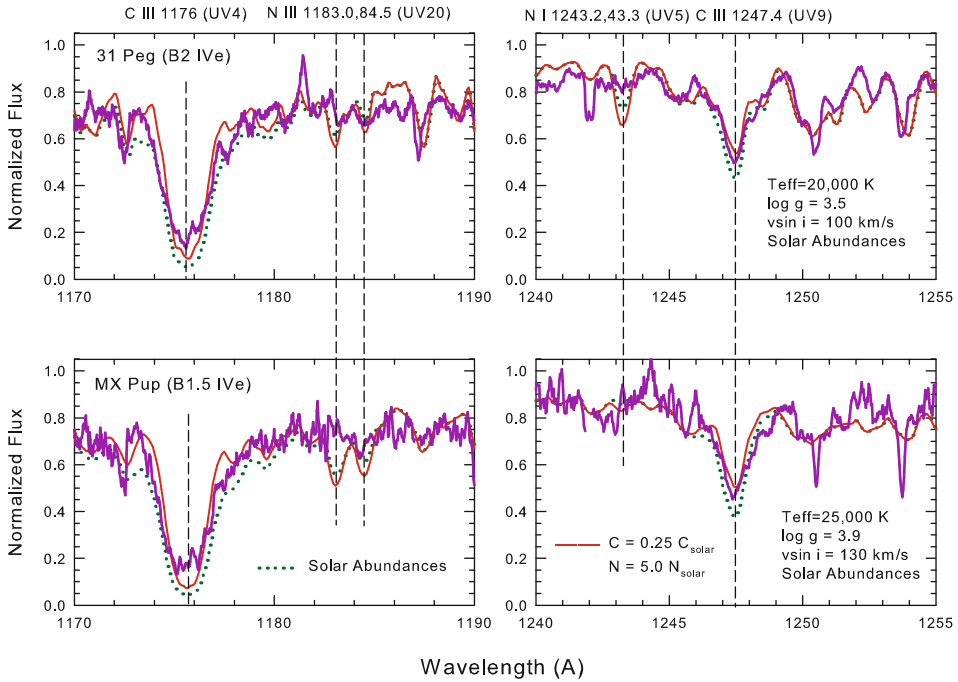
4.2 The Be Stars

It has been 145 years since the first Be star was discovered,⁹ but there is still no good explanation for the mass loss, disk formation, and variability in these stars. Following a paper by Townsend et al. (2004), there has been renewed interest in the possibility that Be stars could be rotating at their critical velocities, a central idea put forth by Struve (1931) in a model that prevailed until the 1980s when new spacecraft and ground-based observations revealed that the Be phenomenon is quite complex and offered no evidence for critical rotation (see the review by Porter and Rivinius 2003 for background on the classical Be stars). Challenges to Townsend et al. have been swift and two recent papers critically examining the effect of gravity darkening on B star parameters (Frémat et al. 2005) and a statistical investigation (Cranmer 2005) again bring the expected maximum V/V_{cr} for Be stars back to the previous consensus value of ≤ 0.90 .

Abundance studies can provide a good *independent* test for critical rotation in Be stars. As discussed above, stellar evolution models by Maeder and Meynet (2000b) and Heger and Langer (2000) predict a significant enhancement in the photospheric N abundances in very rapidly rotating OB stars due to mixing of CNO-processed material from their interiors. The higher the rotation, the more mixing. But the analyses of Be star spectra using standard spectrum synthesis techniques are confronted with some challenges that are not encountered in abundance studies of sharp-lined, nonemission B stars, including the treatment of blended, rotationally broadened lines and correction for disk emission and possible shell absorption (which is a major problem in the analysis of optical spectra). The FUV spectral region is an ideal place to look for N enhancement and C depletion in the photospheres of early Be stars as it contains many strong lines of C II, C III, N II, and N III that are prominent even if the star's projected rotational velocity ($V \sin i$) is large or the abundances are low.

FUV spectra of early Be stars obtained with the *IUE* spacecraft have recently been analyzed with the Hubeny/Lanz NLTE codes TLUSTY/SYNSEX (Hubeny 1988; Hubeny and Lanz 1995) and the Lanz and Hubeny (2007) model atmospheres for B stars (Peters 2011). Results from two stars are presented in  Fig. 9-11. Nitrogen enrichment in the B2 stars would be revealed by a stronger N I at $\lambda 1243 \text{ \AA}$, while in objects hotter than B1, the N III doublet at $\lambda \lambda 1183.0, 1184.5 \text{ \AA}$ would be prominent. From the spectra shown in  Fig. 9-11 it is clear that the presence of an

⁹ γ Cas, (Sechhi 1867).



■ Fig. 9-11

Comparison between the observed FUV spectra of two early Be stars, 31 Peg and MX Pup, (*thick line*) and model spectra computed with the stellar parameters indicated on the plot. Model spectra are for solar abundances (*dotted line*) and elevated nitrogen but reduced carbon (*thin line*). The key analysis lines are labeled at the top (from Peters 2011)

elevated nitrogen abundance is not indicated.¹⁰ In fact, often the observed nitrogen lines are weaker than predicted even if solar abundances prevail. The error in the N abundance is about 0.3–0.4 dex. Given the errors, it appears that the N abundance in the Be stars is normal or at most enhanced by a factor of 2–3. None of the program Be stars show clear evidence for an elevated nitrogen abundance. The expected mixing is apparently suppressed. Either the models need improvement or more likely this study supports the results from Frémat et al. and Cranmer mentioned above that Be stars are not critical rotators.


4.3 Post-Main-Sequence Evolution

As outlined in ► Sect. 1 at the end of hydrogen burning, there is no more fuel to compensate for gravity and the energy losses due to radiation. The core of the star thus contracts until the temperature is high enough to ignite helium in the center. As gravitational energy is suddenly released in this contraction, the envelope is forced to expand until the star becomes transparent



¹⁰The carbon lines were typically weaker than those computed with the Lanz–Hubeny model atmospheres and solar abundances. This could mean a reduced carbon abundance or emission filling from the circumstellar disk.

enough to let this additional energy escape from the star and enter the blue or red supergiant stage. The end of hydrogen burning thus induces a significant change in the position of stars in the HR diagram and leads to the gap between the main sequence and giants, clearly marking the end of the main sequence. The width of the main sequence in the HR diagram depends on the convective core boundary and overshooting beyond the strict Schwarzschild limit that is usually fixed in stellar evolution models in order to reproduce the observed width of the main sequence (see, e.g., Ekström et al. 2012).

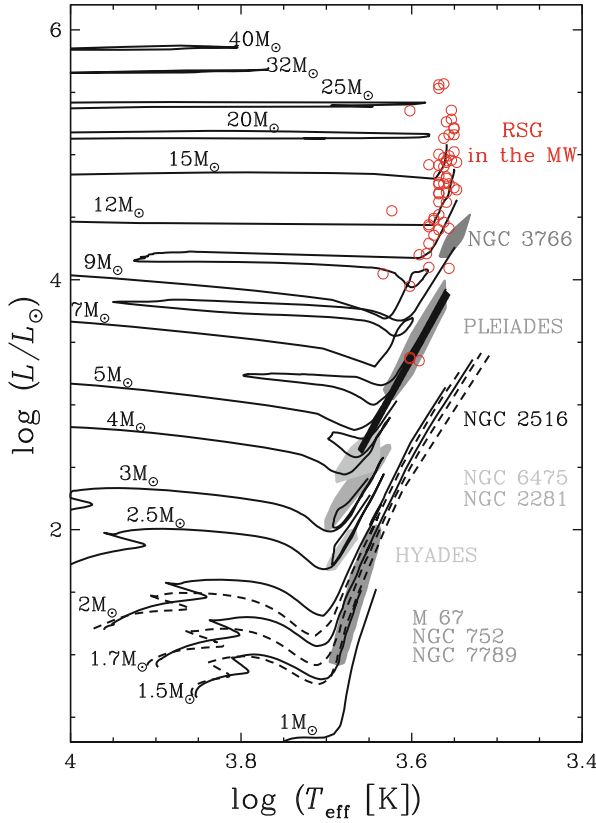
4.3.1 Blue and Red Supergiants

In the same way that the main-sequence width tests convective core overshooting, the position of stars in the red (super-)giant phase tests the treatment of convection in the envelope of stars. The outer convective zone is usually treated according to the mixing length theory, with a solar calibrated value for the mixing-length parameter of the low-mass stars ($\alpha_{\text{MLT}} \equiv \ell/H_p = 1.6467$ in GENEC, where ℓ is the mixing length and H_p is the pressure scale height). For more massive stars ($M > 1.25 M_\odot$) the difference in the equation of state implies a slightly lower value for this parameter ($\alpha_{\text{MLT}} = 1.6$ in GENEC). In  Fig. 9-12, we see that rotating models calculated with GENEC using these values well reproduce the positions of the red giant branch and of the red supergiants, supporting the choices made for the value of the mixing length.

If a star becomes a red supergiant at the very beginning of the core He-burning phase, its red supergiant lifetime will be longer than the one of a star that begins to transform its central helium in the blue part of the HR diagram and thus enters its red supergiant phase in a more advanced evolutionary stage. Thus, red supergiant lifetimes depend on the way the star evolves into the red part of the HR diagram after the main-sequence phase. This redward evolution depends in turn on the extension of the intermediate convective zone associated with the H-burning shell. Let us recall that in a convective zone the density gradients are shallower. As a consequence, the larger the intermediate convective zone, the more compact the star and thus a blue position in the HR diagram is favored. Alternatively, when this intermediate convective zone decreases in mass or even disappears, this favors the inflation of the star and a rapid evolution into the red part of the HR diagram. Mass loss (see Meynet 1993) and (rotational) mixing (Maeder and Meynet 2001) both reduce the extension of the intermediate convective zone and therefore favor an early entrance into the RSG stage during the core He-burning phase.

Let us illustrate this by looking at  Fig. 9-13, where the structures of solar metallicity models for two $20 M_\odot$ stars are shown. The models show the presence of an intermediate convective zone associated with the H-burning shell at the beginning of the core He-burning phase. This intermediate convective zone remains present for a much longer period in the nonrotating model than in the rotating one. Also, the total mass of the nonrotating model is slightly larger than the total mass of the rotating one at the same stage. One expects thus that, at the end of the main-sequence phase, the nonrotating model will cross the HR gap more slowly. This is exactly what happens as can be seen in  Fig. 9-6.

More important but similar effects are observed in metal poor models. It has been shown by Maeder and Meynet (2001) that rotation may be the reason for the large number of RSGs observed in the SMC cluster NGC 330. Also, rotation may be a key ingredient for explaining the B/R ratio estimated in the galaxy Sextans A from HST imaging by Dohm-Palmer and Skillman (2002). They find that even if the nonrotating stellar models of Schaller et al. (1992) reproduce well the shape of the variation of the B/R ratio as a function of age, the observed ratio is lower

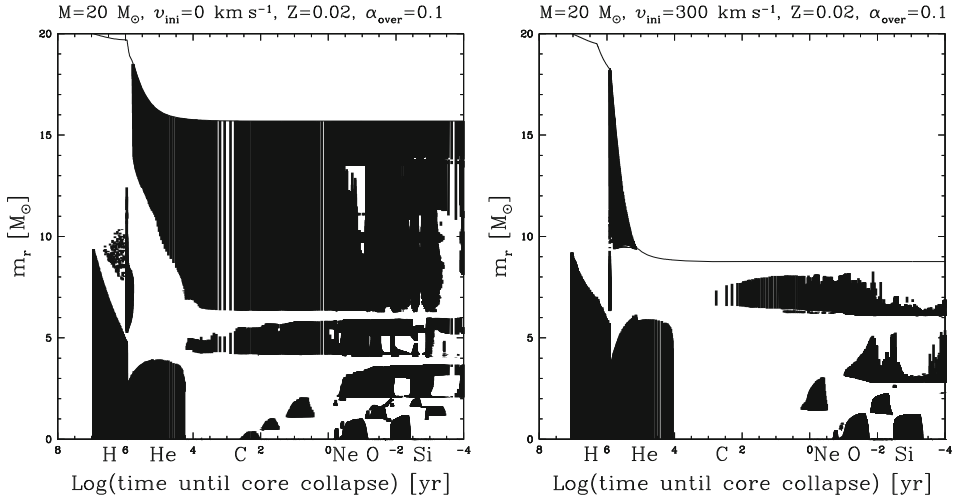


■ Fig. 9-12

Evolutionary tracks for rotating models in the *red part* of the HR diagram. In the low-mass range, a few nonrotating tracks (*dashed lines*) are shown. The *gray shaded* areas indicate the observations in clusters and associations as given in Maeder and Meynet (1989) as well as the position of Galactic red supergiants obtained by Levesque et al. (2005, *red circles*)

than the model by a factor 2. Rotating models such as those presented in Maeder and Meynet (2001) allow one to reproduce the observed value. The link between rotation and red supergiant is also indirectly supported by the observed positive correlation between the population of Be stars and the RSG populations (see Fig. 1b in Meynet et al. 2007).

Once a red supergiant, the star can remain in this state until it explodes in a type IIP core collapse supernova. But it is not necessarily the case. The star can also evolve back to the blue part of the HR diagram and end its life as a blue supergiant or even, in some probably rare cases (see below), as a Wolf–Rayet star. The important factor governing the blueward evolution from the red supergiant stage is the mass fraction occupied by the helium core (Giannone 1967). Typically, when the mass fraction of the He core becomes greater than about 60–70% of the total actual mass, the star evolves back to the blue part of the HR diagram. This is what happens in the rotating 20 M_{\odot} stellar model with $v_{\text{ini}} = 300 \text{ km s}^{-1}$, in which the core He-burning core



■ Fig. 9-13

Evolution as a function of the remaining time until core collapse of the total mass and of the masses of the convective regions (*in black*) for a nonrotating (*left*) and rotating (*right*) solar metallicity $20 M_\odot$ models (Plots taken from Hirschi et al. (2004))

occupies more than 65% of the total mass. The nonrotating model has a core that occupies only about 25% of the total mass and remains in the red part of the HR diagram.

As it was the case for the first crossing of the HR gap, this second crossing of the HR diagram depends on mass loss (during the red supergiant stage) and mixing (during the previous evolutionary phases). Strong mass loss during the red supergiant phase favors a bluemward evolution (Salasnich et al. 1999; Vanbeveren et al. 2007; Yoon and Cantiello 2010) since it makes the mass fraction of the core larger. Strong mixing during the previous phases also makes larger cores and thus favors an evolution from the red to the blue supergiant phase (Hirschi et al. 2004). Thus, we see that the lifetime of the star in the red supergiant stage, as well as the possibility for the star to explode as a type II SN in that stage, depends heavily on mass loss and mixing.

Let us now recall a few observations which are relevant in that context:

- The number ratio of blue to red supergiants in clusters with ages between 6.8 and 7.2 (in logarithm of the age in years), or with masses at the turn off between about 12 and $30 M_\odot$, increases when the metallicity increases (Meylan and Maeder 1983; Eggenberger et al. 2002). This observed trend is exactly the opposite of what is expected from standard grids of models (see the discussion in Langer and Maeder 1995).
- An important number ratio is the relative number of RSGs to WRs. Massey (2003, see his Fig. 12) shows that this ratio decreases with increasing metallicity by a factor of 100 over a range of 0.9 dex in metallicity. The strong decrease with the metallicity indicates that the mass range for the progenitors of WRs (RSGs) increases (decreases) when the metallicity increases.

These two important features emphasize the importance of metallicity in shaping the blue and red supergiants and Wolf–Rayet populations. Mass loss triggered by radiation increases with the metallicity (Kudritzki and Puls 2000), and this is probably one of the main factors responsible for the above observed trends. For instance, when the metallicity increases, stars enter at an earlier phase into the WR phase (making the duration of that phase longer). Also, stars of lower initial masses can become WR at the end of their lifetime. These two factors are exactly what is required to explain the trend of the relative number of RSGs and WRs with metallicity.

For the B/R ratio however, this is not the case! Indeed, when the metallicity increases, mass-loss rates increase at least during the main-sequence phase, and thus, RSGs are favored at high metallicity, which is the contrary of what is observed. Thus, the increase of the mass-loss rate with the metallicity, expected for hot stars, cannot be the cause of the observed B/R trend. This is in fact a counteracting effect.

Now, we have mentioned that an increase of the mass-loss rates *during the RSG phase* would shorten the RSGs lifetime and lead to the formation of a blue supergiant, or even a WR star in some cases. At the moment, mass-loss rates used in current stellar models (de Jager et al. 1988) are too weak for producing this kind of evolution for stars in the mass range between 12 and 25–30 M_{\odot} (at least for standard nonrotating models). However, the uncertainties are large; actually, the determinations of the mass loss during the RSG phase is still more difficult than in the blue part of the HR diagram due in part to the presence of dust and to various instabilities active in red supergiant atmospheres (e.g., convection becomes supersonic and turbulent pressure can no longer be ignored). An illustration of the difficulty comes from the determinations of red supergiant mass-loss rates by van Loon et al. (2005). Their study is based on the analysis of optical spectra of a sample of dust-enshrouded red giants in the LMC, complemented with spectroscopic and infrared photometric data from the literature. Comparison with galactic AGB stars and red supergiants shows excellent agreement for dust-enshrouded objects, but not for optically bright ones. Dust-enshrouded objects show mass-loss rates which are greater by a factor 3–50 than those deduced from optically bright ones! In this context, the questions of which stars do become dust enshrouded, at which stage, and for how long, become critical to make correct prediction of the mass lost by stellar winds. One can also note a very interesting point deduced from the study by van Loon et al. (2005), that for dust-enshrouded objects, mass loss appears to be independent of the metallicity! On the other hand, the formation and duration of the dust enshrouded stage may be metallicity dependent! Thus, one sees here that improvements are needed in order to clarify the situation. Let us just mention that stronger mass loss at high metallicity during the red supergiant phase would go in the right direction for explaining the observed B/R trend with metallicity. Indeed, an increase of the mass-loss rate during the red supergiant phase at high Z favors blueward evolution and thus reduces the RSG lifetimes (see Meynet et al. 2011 for additional details).

In the models of Schaller et al. (1992), the ratio of the blue to the red lifetime (red being defined as the time spent with $\log(T_{\text{eff}})$ below 3.65) for the 20 M_{\odot} is 0.2, while for the non-rotating and rotating model of Ekström et al. (2012) it is 0.7 and 1.5, respectively. The same numbers for the 25 M_{\odot} models are 0.10 in Schaller et al. (1992), and 1.6 (nonrotating) and 5.3 (rotating) in Ekström et al. (2012). Thus, in that mass range, the B/R ratio is greatly enhanced. Whether it is sufficient to explain the B/R ratio as observed in clusters remains to be checked. Note that at solar metallicity, the observational values estimated by Eggenberger et al. (2002) are between 1.2 and 4, well in agreement with the 20–25 M_{\odot} rotating models from Ekström et al. (2012).

Recent studies of the red and yellow supergiants identified in the Large Magellanic Cloud (Neugent et al. 2012) and M33 (Drouot et al. 2012) show excellent agreement between their locations in the HRD and the newest evolutionary tracks from the Geneva group. Finally, if mass loss during the red supergiant phase is strong enough to remove the entire hydrogen-rich envelope, the star becomes a Wolf–Rayet star. As can be seen in [Fig. 9-12](#), both the models of Ekström et al. (2012) and the observations of Levesque et al. (2005) agree on an upper mass limit for red supergiants around $25 M_{\odot}$. Above this limit, mass loss is strong enough to prevent the star from reaching the red supergiant phase.

4.3.2 Wolf–Rayet (WR) Stars

Wolf–Rayet (WR) stars (see [Sect. 2.2](#)) are characterized by four important observed features:

1. They are associated with young massive star regions.
2. They present broad and strong emission lines.
3. They are hot ($\log(T_{\text{eff}}) \gtrsim 4$) and luminous stars ($\log(L/L_{\odot}) > 5.0$).
4. The chemical composition of their surface shows signs of H-burning (WN-type) and/or He-burning processes (WNC, WC, WO-type).

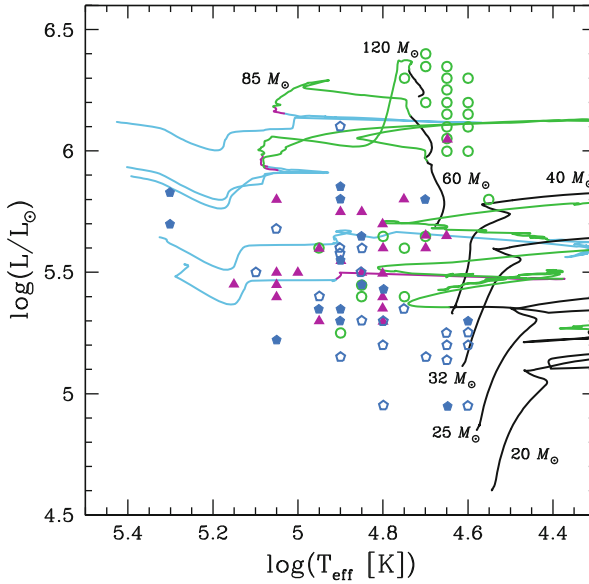
These features can well be explained if WR are massive evolved stars whose surface composition has been changed by mass loss and/or internal mixing. The mass loss can be due either to stellar winds and/or to the loss of the envelope through a Roche-lobe overflow (RLOF) in a close-binary system.

In order to compare models to observations, a classification of stellar models based on a synthetic model atmosphere spectrum would be the best method. While some attempts in that direction have already been made (Schaerer and de Koter 1997), this is by far not yet a standard procedure. Instead, the following criteria to determine the type of the star at a given time are commonly used (Meynet and Maeder 2003):

- A stellar model with $\log(T_{\text{eff}}) > 4.0$ and a surface hydrogen mass fraction $X_{\text{H}} < 0.3$ is considered as a WR star.¹¹
- A stellar model with $\log(T_{\text{eff}}) > 4.5$ which is not a WR star is an O-type star.
- A WR stellar model with a surface hydrogen mass fraction $X_{\text{H}} > 10^{-5}$ is a WNL star.
- A WR stellar model without hydrogen and with a carbon surface abundance inferior to the nitrogen abundance is a WNE star.
- A WR stellar model without hydrogen and with a carbon surface abundance superior to the nitrogen abundance is a WC or a WO star. If the surface ratio $\frac{\text{C}+\text{O}}{\text{He}}$ (in number) is less than one, the star is classified as a WC; otherwise, it is classified as a WO.
- A WR stellar model without hydrogen and with surface abundances in nitrogen and carbon such as $0.1 < \frac{X_{\text{C}}}{X_{\text{N}}} < 10$ is a WNC star.

The O-type, WNL, WNE, WC, and WO phases are exclusive. The WNC phase lies between both the WNE and WC phases.

¹¹The duration of the corresponding WR phase is not very sensitive to the choice of this limit if it remains in the range 0.3–0.4 as mentioned by Meynet and Maeder (2003).



■ Fig. 9-14

Positions of observed WN and WC stars in the HRD as given by Hamann et al. (2006) and Sander et al. (2012), respectively. The empty circles are WNL stars, and the full triangles are WNE. The WC stars are represented by pentagons, filled when the distance is known and empty when it is unknown. Evolutionary tracks of rotating models calculated by Georgy et al. (2012) are superposed

Georgy et al. (2012) have compared the observed position of some galactic WN (Hamann et al. 2006) and WC (Sander et al. 2012) stars with rotating single-star models (see ► Fig. 9-14). The luminosity of the most luminous WNL star in the sample of Hamann et al. (2006) (not shown in ► Fig. 9-14) would require an initial mass of the order of $300 M_{\odot}$. Crowther et al. (2010) have recently claimed to have observed such massive objects. However, WR 25 (or HD 93162), the most luminous star, is a binary star with a spectroscopically determined orbital period of 207.8 d (Raassen et al. 2003). The second most luminous star, WR 22, is an eclipsing binary. Rauw et al. (1996) estimated a lower limit for the actual mass for this star to be $72 M_{\odot}$. Due to their binary status, these two stars are not plotted in ► Fig. 9-14. Looking at ► Fig. 9-14, a few points deserve to be emphasized:



- The WNL, WNE, and WC present some kind of decreasing luminosity sequence. There are of course overlaps between the different WR types; however, just by eye one can notice immediately from ► Fig. 9-14 that the averaged luminosity of each class decreases from the WNL to the WC stars. The WR stars with the highest luminosity are thus the WNL stars (the three upper points are also those with the highest abundances in hydrogen, with values between 0.35 and 0.53), and the star with the lowest luminosity is a WC star.
- The range of effective temperatures (taken here at the surface of the hydrostatic core, i.e., not accounting for the optical thickness of the wind) of the WNL stars is relatively narrow (between 4.6 and 4.8) above a luminosity equal to about 5.8. Below, it widens and extends from 4.6 up to slightly more than 5.

- The ranges of effective temperatures covered by the WNE and WC are similar and extend toward higher temperatures than the range of WNL stars. The difference in temperature ranges between the H-rich and the H-poor stars reflects the strong dependency of the opacity on the quantity of hydrogen.

Before comparing these observations with the evolutionary tracks, let us try to explain the observed features listed above in a general theoretical framework. We focus here on the luminosity leaving aside the question of the effective temperature. The reason for that is that the effective temperature depends considerably on the mass-loss rate used and on the physics of the outer layers, while luminosity is quite tightly related to the total mass of the star and its internal physics. In that respect, it is a more fundamental quantity to compare with models for the interior of stars.

The decrease in luminosity when passing from the WNL to the WC stars might be interpreted in two ways. To obtain a WC star from a given initial mass star, the star must lose more mass than what is required to obtain a WNL star, implying that on average WC stars have lower actual masses and thus also lower luminosities. Another way to interpret this feature (although not incompatible with the previous one) would be to assume that the most massive stars produce the high-luminosity WNL stars (which will then evolve into lower-luminosity WNE and WC type stars), while the less massive stars produce the WC stars. In the framework of the single-star scenario, this last explanation requires very strong mass loss during the RSG phase of stars with initial masses around $15 M_{\odot}$. Such high mass loss could be due either to some physical processes originating in the envelope of the RSGs, while in the binary channel, it could be due to a RLOF process occurring in a close-binary system. Whatever process is invoked, it should not produce any long WNL or WNE phases at such a low luminosity range since these stars are not observed.

Comparing the above observed positions with the rotating single star models we note the following features:

- The group of 7 very luminous stars ($\log(L/L_{\odot}) \geq 6.25$) all have a mass fraction of hydrogen superior to 10%, except in one case where the mass fraction is estimated to be 0.05. In addition, no WNE and WC are observed in the luminosity range covered by these stars. Is there any explanation for these two features? The reason why only H-rich WN stars are observed in this high luminosity range gives support to the idea that WR stars form through a combination of mass loss and radiative-zone mixing, rather than only through mass loss as in the nonrotating models. Models without rotation or lower initial mass models, which enter into the WR phase only after the end of the main-sequence phase, do not show any long “H-rich” periods and thus cannot account for the most luminous H-rich WNL stars. The absence of WNE and WC stars in this luminosity range may be explained by the fact, already mentioned above, that WNE and WC stars are expected to be less massive and therefore less luminous than WNL stars because more mass has to be removed from the star to reach those stages and/or because they are produced from stars with lower initial masses.
-  *Figure 9-14* shows that the evolutionary tracks cover the region where the WNL stars are observed. The extension in luminosity of the WNE stars is also well reproduced, somewhat supporting the present single-star models for explaining these populations.
- A serious difficulty comes from the low observed luminosity of some WC stars. Our present tracks predict a lower luminosity limit for the WC stars of 5.35 (in $\log(L/L_{\odot})$), while the lowest luminosity plotted in  *Fig. 9-14* is around 4.9 according to the revised spectral

analyses of galactic WC stars by Sander et al. (2012). The three most likely scenarios to explain the observed low-luminosity (and thus low-mass) WC stars are the following:

- Stars with initial masses between 15 and 20 M_{\odot} lose much more mass than in the models during their RSG stage.
- Stars with initial masses above about 25 M_{\odot} lose much more mass than in the models during the WR stage.
- The low-mass WC stars are produced in close-binary systems through RLOF. This process can indeed produce smaller final masses as illustrated, for instance, by the recent models of Yoon et al. (2010).

4.4 Modeling Uncertainties and Fate of Massive Stars

As discussed above, many observed features of massive stars are reproduced by models of rotating stars. There are, however, several uncertainties and issues that require more work. A major uncertainty is mixing at convective boundaries. 3D simulations are necessary to improve on current treatment based on the mixing-length theory of Böhm-Vitense (1958) and reduce this uncertainty that can affect possible shell overlaps in the advanced stages. Arnett and Meakin have started this challenging task with promising developments (Arnett and Meakin 2011; Meakin and Arnett 2007). Another major question is the exact role of binarity in the evolution and fate of massive stars. As mentioned above, binary stars are one of the most likely origins for the low-luminosity WC stars. Finally, the interaction between rotation and magnetic field is still debated and not fully understood (see ► Sect. 3.1.2). This interaction is the key to determine the final angular momentum in the collapsing core of massive stars and the role of rotation and magnetic fields in core collapse supernovae. Without magnetic fields, current rotating models predict presupernova cores that rotate too fast (Hirschi et al. 2005). Models including the effects of magnetic fields rotate slower but still faster than long period pulsars and especially still rotate too fast compared to observed white dwarfs (Suijs et al. 2008). The reader is referred to Georgy et al. (2012) for an extended comparison between predicted and observed fate of massive stars.

Stellar evolution is an established field but there are still major open questions, uncertainties, and issues that need to be addressed. In the last few decades, the implementation of the physics of rotation into stellar evolution models has greatly improved the model predictions, and the latest models can much better reproduce many observables: surface enrichment of main-sequence stars, number ratios, and properties of red, yellow, and blue supergiants as well as Wolf–Rayet stars. There are nevertheless still many improvements needed for our understanding of massive stars in the following areas: mass loss in the LBV phase and cool stars, interaction between rotation and magnetic fields, multi-D effects associated with convective boundary mixing, and binary interactions. The field of stellar evolution is thus expected to remain an exciting and evolving field for years to come.

Acknowledgment

GJP is grateful for the support from several NASA grants including NNX07AF89G, NNX07AH56G, and NNX10AD66G and the Women in Science and Engineering Program

(WiSE) at USC. RH acknowledges support from the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and from ESF-EuroGENESIS program. RH also expresses his gratitude to the groups of Professors Maeder and Meynet in Geneva and Professor Thielemann in Basel for very fruitful collaborations and support and for the use of results and figures in this chapter.

References

- Andersen, J. 1991, *A&A Rev*, 3, 91
- Anderson, J., & Clausen, J. V. 1989, *A&A*, 213, 183
- Andersen, J., Clausen, J. V., Nordström, B., & Reipurth, B. 1983, *A&A*, 121, 271
- Arnett, W. D., & Meakin, C. 2011, *ApJ*, 741, 33
- Arnett, W. D., Bahcall, J. N., Kirshner, R. P., & Woosley, S. E. 1989, *ARA&A*, 27, 629
- Böhm-Vitense, E. 1958, *ZAp*, 46, 108
- Braithwaite, J. 2006, *A&A*, 449, 451
- Brott, I., Hunter, I., de Koter, A., Langer, N., Lennon, D., & Dufton, P. 2009, *Commun. Asteroseismol.*, 158, 55
- Brott, I., et al. 2011, *A&A*, 530, A115
- Brüggen, M., & Hillebrandt, W. 2001, *MNRAS*, 323, 56
- Canuto, V. M. 2002, *A&A*, 384, 1119
- Chaboyer, B., & Zahn, J.-P. 1992, *A&A*, 253, 173
- Chiappini, C., Hirschi, R., Meynet, G., Ekström, S., Maeder, A., & Matteucci, F. 2006, *A&A*, 449, L27
- Chieffi, A., Limongi, M., & Straniero, O. 1998, *ApJ*, 502, 737
- Clausen, J. V., & Gimenez, A. 1991, *A&A*, 241, 98
- Clausen, J. V., Helt, B. E., Gimenez, A., Vaz, L. P. R., Garcia, J. M., Olsen, E. H., & Southworth, J. 2007, *A&A*, 461, 1065
- Clayton, D. D. 1983, *Principles of Stellar Evolution and Nucleosynthesis*, (2nd ed.; Chicago, IL: University of Chicago Press)
- Cranmer, S. R. 2005, *ApJ*, 634, 585
- Crowther, P. A. 2007, *ARA&A*, 45, 177
- Crowther, P. A., De Marco, O., & Barlow, M. J. 1998, *MNRAS*, 296, 367
- Crowther, P. A., Schnurr, O., Hirschi, R., Yusof, N., Parker, R. J., Goodwin, S. P., & Kassim, H. A. 2010, *MNRAS*, 408, 731
- de Jager, C., Nieuwenhuijzen, H., & van der Hucht, K. A. 1988, *A&AS*, 72, 259
- Dohm-Palmer, R. C., & Skillman, E. D. 2002, *AJ*, 123, 1433
- Drout, M. R., Massey, P., & Meynet, G. 2012, *ApJ*, 750, 97
- Eggenberger, P., Meynet, G., & Maeder, A. 2002, *A&A*, 386, 576
- Eggenberger, P., Meynet, G., Maeder, A., Hirschi, R., Charbonnel, C., Talon, S., & Ekström, S. 2007, *Ap&SS*, 263
- Ekström, S., et al. 2012, *A&A*, 537, A146
- Evans, C. J., et al. 2005, *A&A*, 437, 467
- Evans, C. J., Lennon, D. J., Smartt, S. J., & Trundle, C. 2006, *A&A*, 456, 623
- Fliegner, J., Langer, N., & Venn, K. A. 1996, *A&A*, 308, L13
- Frémat, Y., Zorec, J., Hubert, A.-M., & Floquet, M. 2005, *â*, 440, 305
- Fricke, K. 1968, *Z Astrophys*, 68, 317
- Frischknecht, U., Hirschi, R., Meynet, G., Ekström, S., Georgy, C., Rauscher, T., Winteler, C., & Thielemann, F.-K. 2010, *A&A*, 522, A39
- Frischknecht, U., Hirschi, R., & Thielemann, F.-K. 2012, *A&A*, 538, L2
- Gagné, M., Fehon, G., Savoy, M. R., & Cohen, D. H., et al. 2011, *ApJS*, 194, A5
- Georgy, C., Ekström, S., Meynet, G., Massey, P., Levesque, E. M., Hirschi, R., Eggenberger, P., & Maeder, A. 2012, *A&A* in press, arXiv:1203.5243
- Giannone, P. 1967, *ZAp*, 65, 226
- Goldreich, P., & Schubert, G. 1967, *ApJ*, 150, 571
- Gray, R. O., & Corbally, C. J. 2009, *Stellar Spectral Classification* (Princeton: Princeton University Press)
- Groh, J. H., Damiani, A., & Hillier, D. J., et al. 2009, *ApJ*, L25
- Hamann, W.-R., Gräfener, G., & Liermann, A. 2006, *A&A*, 457, 1015
- Harmanec, P. 1988, *Bull Astron Inst Cz*, 39, 329
- Heger, A., & Langer, N. 2000, *ApJ*, 544, 1016
- Heger, A., Langer, N., & Woosley, S. E. 2000, *ApJ*, 528, 368
- Hensberge, H., Pavlovski, K., & Verschuere, W. 2000, *A&A*, 358, 553
- Hirschi, R. 2007, *A&A*, 461, 571
- Hirschi, R., & Maeder, A. 2010, *A&A*, 519, A16
- Hirschi, R., Meynet, G., & Maeder, A. 2004, *A&A*, 425, 649
- Hirschi, R., Meynet, G., & Maeder, A. 2005, *A&A*, 443, 581
- Holmgren, D. E., Hill, G., & Fisher, W. 1990, *A&A*, 236, 409

- Howarth, I. D., & Prinja, R. K. 1989, *ApJS*, 69, 527
- Hubeny, I. 1988, *Comput Phys Commun*, 52, 103
- Hubeny, I., & Lanz, T. 1995, *ApJ*, 439, 875
- Humphreys, R. M., & Davidson, K. 1994, *PASP*, 106, 1025
- Hunter, I., et al. 2007, *A&A*, 466, 277
- Hunter, I., et al. 2008, *ApJ*, 676,
- Hunter, I., et al. 2009, *A&A*, 496, 841
- Iben, I., Jr. 1967, *ARA&A*, 5, 571
- Iben, I., Jr., & Renzini, A. 1983, *ARA&A*, 21, 271
- Käppeler, F., Gallino, R., Bisterzo, S., & Aoki, W. 2011, *Rev Mod Phys*, 83, 157
- Kippenhahn, R., & Weigert, A. 1990, *Stellar Structure and Evolution* (Berlin: Springer)
- Kippenhahn, R., Weigert, A., & Hofmeister, E. 1967, in *Methods in Computational Physics*, Vol. 7, ed. B. Alder, S. Fernbach, & M. Rotenberg (New York, NY/London, UK: Academic press)
- Knobloch, E., & Spruit, H. C. 1983, *A&A*, 125, 59
- Kudritzki, R.-P., & Puls, J. 2000, *ARA&A*, 38, 613
- Langer, N., & Maeder, A. 1995, *A&A*, 295, 685
- Langer, N., Cantiello, M., Yoon, S. C., Hunter, I., Brott, I., Lennon, D., de Mink, S., & Verheijdt, M. 2008, in *Proc. IAU Symp. 250, Massive Stars as Cosmic Engines*, ed. F. Bresolin, P.A. Crowther & J. Puls (Cambridge: Cambridge University Press), 167
- Lanz, T., & Hubeny, I. 2007, *ApJS*, 169, 83
- Levesque, E. M., Massey, P., Olsen, K. A. G., Plez, B., Josselin, E., Maeder, A., & Meynet, G. 2005, *ApJ*, 628, 973
- Limongi, M., Straniero, O., & Chieffi, A. 2000, *ApJS*, 129, 625
- Maeder, A. 1997, *A&A*, 321, 134
- Maeder, A. 2003, *A&A*, 399, 263
- Maeder, A. 2009, *Physics, Formation, and Evolution of Rotating Stars*, (Berlin: Springer)
- Maeder, A., & Meynet, G. 1989, *A&A*, 210, 155
- Maeder, A., & Meynet, G. 2000a, *ARA&A*, 38, 143
- Maeder, A., & Meynet, G. 2000b, *A&A*, 361, 159
- Maeder, A., & Meynet, G. 2001, *A&A*, 373, 555
- Maeder, A., & Meynet, G. 2005, *A&A*, 440, 1041
- Maeder, A., & Meynet, G. 2012, *Rev Mod Phys*, 84, 25
- Maeder, A., & Zahn, J. 1998, *A&A*, 334, 1000
- Maeder, A., Meynet, G., Ekström, S., & Georgy, C. 2009, *Commun Asteroseismol*, 158, 72
- Massey, P. 2003, *ARA&A*, 41, 15
- Massey, P., & Grove, K. 1989, *ApJ*, 344, 870
- Meakin, C. A., & Arnett, D. 2007, *ApJ*, 667, 448
- Mendel, J. T., Venn, K. A., Proffitt, C. R., Brooks, A. M., & Lambert, D. L. 2006, *ApJ*, 640, 1039
- Meylan, G., & Maeder, A. 1983, *A&A*, 124, 84
- Meynet, G. 1993, in *The Feedback of Chemical Evolution on the Stellar Content of Galaxies*, ed. D. M. Alloin, & G. Stasińska (Meudon: L'Observatoire de Paris), 40
- Meynet, G., & Maeder, A. 1997, *A&A*, 321, 465
- Meynet, G., & Maeder, A. 2002, *A&A*, 390, 561
- Meynet, G., & Maeder, A. 2003, *A&A*, 404, 975
- Meynet, G., Eggenberger, P., & Maeder, A. 2007, in *Proc. IAU Symp. 241, Stellar Populations as Building Blocks of Galaxies*, ed. A. Vazdekis & R. F. Peletier (Cambridge: Cambridge University Press), 13
- Meynet, G., Ekström, S., & Maeder, A. 2006, *A&A*, 447, 623
- Meynet, G., Georgy, C., Hirschi, R., Maeder, A., Massey, P., Przybilla, N., & Nieva, M.-F. 2011, *Bull Soc R Sci Liege*, 80, 266
- Morel, T., Hubrig, S., & Briquet, M. 2008, *A&A*, 481, 453
- Nazé, Y., Broos, P. S., Oskinova, L., & Townsley, L. K., et al. 2011, *ApJS*, 194, A7
- Neiner, C. 2011, in *Proc. IAU Symp. 272, Active OB Stars: Structure, Evolution, Mass Loss, and Critical Limits*, ed. C. Neiner, G. Wade, G., Meynet, & G. Peters (Cambridge, UK: Cambridge University Press)
- Neugent, K. F., Massey, P., Skiff, B., & Meynet, G. 2012, *ApJ*, 749, A177
- Nugis, T., & Lamers, H. J. G. L. M. 2000, *A&A*, 360, 227
- Pagel, B. E. J. 2009, *Nucleosynthesis and Chemical Evolution of Galaxies* (Cambridge: Cambridge University Press)
- Pavlovski, K., & Hensberge, H. 2005, *A&A*, 439, 309
- Pavlovski, K., & Southworth, J. 2009, *MNRAS*, 394, 1519
- Pavlovski, K., Tamajo, E., Koubský, P., Southworth, J., Yang, S., & Kolbas, V. 2009, *MNRAS*, 400, 791
- Paxton, B., Bildsten, L., Dotter, A., Herwig, F., Lesaffre, P., & Timmes, F. 2011, *ApJS*, 192, 3
- Peters, G. J. 2011, in *Proc. IAU Symp. 272, Active OB Stars: Structure, Evolution, Mass Loss, and Critical Limits*, ed. C. Neiner, G. Wade, G. Meynet, & G. Peters (Cambridge: Cambridge University Press), 101
- Podsiadlowski, P., Hsu, J. J. L., Joss, P. C., & Ross, R. R. 1993, *Nature*, 364, 509
- Porter, J. M., & Rivinius, T. 2003, *PASP*, 115, 1153
- Popper, D. M., & Etzel, P. B. 1981, *AJ*, 86, 102
- Popper, D. M., & Hill, G. 1991, *AJ*, 101, 600
- Pourbaix, D., et al. 2004, *A&A*, 424, 727
- Prinja, R. D., Massa, D., & Fullerton, A. W. 2002, *ā*, 388, 587
- Proffitt, C. R., & Quigley, M. F. 2001, *ApJ*, 548, 429
- Proffitt, C. R., Jönsson, P., Litzen, U., Pickering, J. C., & Wahlgren, G. M. 1999, *ApJ*, 516, 342
- Raassen, A. J. J., van der Hucht, K. A., Mewe, R., Antokhin, I. I., Rauw, G., Vreux, J.-M., Schmutz, W., & Güdel, M. 2003, *A&A*, 402, 653

- Rauw, G., Vreux, J.-M., Gosset, E., Hutsemekers, D., Magain, P., & Rochowicz, K. 1996, *A&A*, 306, 771
- Salasnich, B., Bressan, A., & Chiosi, C. 1999, *A&A*, 342, 131
- Sander, A., Hamann, W.-R., & Todt, H. 2012, *ArXiv e-prints*:1201.6354
- Schaerer, D., & de Koter, A. 1997, *A&A*, 322, 598
- Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, *A&AS*, 96, 269
- Schnurr, O., Casoli, J., Chené, A.-N., Moffat, A. F. J., & St-Louis, N. 2008, *MNRAS*, 389, L38
- Secchi, A. 1867, *Astron Nachr*, 68, 63
- Simon, K. P., Sturm, E., & Fiedler, A. 1994, *A&A*, 292, 507
- Smith, N. 2011, in *Proc. IAU Symp. 272, Active OB Stars: Structure, Evolution, Mass Loss, and Critical Limits*, ed. C. Neiner, G. Wade, G. Meynet, G. Peters (Cambridge: Cambridge University Press), 571
- Smith, N., Miller, A., Li, W., & Filippenko, A. V., et al. 2010, *AJ*, 139, 1451
- Southworth, J., & Clausen, J. V. 2007, *A&A*, 461, 1077
- Southworth, J., Maxted, P. F. L., & Smalley, B. 2004, *MNRAS*, 351, 1277
- Spite, M., et al. 2005, *A&A*, 430, 655
- Spruit, H. C. 2002, *A&A*, 381, 923
- Struve, O. 1931, *ApJ*, 73, 94
- Suijs, M. P. L., Langer, N., Poelarends, A.-J., Yoon, S.-C., Heger, A., & Herwig, F. 2008, *A&A*, 481, L87
- Townsend, R. H. D., Owocki, S. P., & Howarth, I. D. 2004, *MNRAS*, 350, 189
- Trundle, C., Dufton, P. L., Hunter, I., Evans, C. J., Lennon, D. J., Smartt, S. J., & Ryans, R. S. I. 2007, *A&A*, 471, 625
- Vanbeveren, D. 2001, in *The Influence of Binaries on Stellar Population Studies*, ed. D. Vanbeveren (Dordrecht: Kluwer)
- Vanbeveren, D., Van Bever, J., & Belkus, H. 2007, *ApJ*, 662, L107
- van Loon, J. T., Cioni, M.-R. L., Zijlstra, A. A., & Loup, C. 2005, *A&A*, 438, 273
- Van Rensbergen, W., De Greve, J. P., Mennekens, N., Jansen, K., & De Loore, C. 2010, *A&A*, 510, A13
- Van Rensbergen, W., De Greve, J. P., Mennekens, N., Jansen, K., & De Loore, C. 2011, *A&A*, 528, A16
- Vaz, L. P. R., Cunha, N. C. S., Vieira, E. F., & Myrrha, M. L. M. 1997, *A&A*, 327, 1094
- Venn, K. A., Brooks, A. M., Lambert, D. L., Lemke, M., Langer, N., Lennon, D. J., & Keenan, F. P. 2002, *ApJ*, 565, 571
- Vink, J. S., de Koter, A., & Lamers, H. J. G. L. M. 2001, *A&A*, 369, 574
- von Zeipel, H. 1924, *MNRAS*, 84, 665
- Walborn, N. R., Howarth, I. D., & Lennon, D. J., et al. 2002, *AJ*, 123, 2754
- Walder, R., Folini, D., & Meynet, G. 2011, *Space Sci Rev*, 125
- Woosley, S. E., & Weaver, T. A. 1986, *ARA&A*, 24, 205
- Yoon, S.-C., & Cantiello, M. 2010, *ApJ*, 717, L62
- Yoon, S.-C., Langer, N., & Norman, C. 2006, *A&A*, 460, 199
- Yoon, S.-C., Woosley, S. E., & Langer, N. 2010, *ApJ*, 725, 940
- Zahn, J.-P. 1992, *A&A*, 265, 115
- Zahn, J., Brun, A. S., & Mathis, S. 2007, *A&A*, 474, 145

10 Stellar Activity

Isabella Pagano

INAF, Catania Astrophysical Observatory, Catania, Italy

1	<i>Introduction</i>	488
2	<i>Objects Showing Magnetic Activity</i>	490
3	<i>Stellar Chromospheres</i>	492
3.1	Ca II H&K and Wilson–Bappu Relation	492
3.2	The Activity Indexes	495
3.2.1	The Vaughan–Preston Gap	497
3.3	Chromospheric Activity Versus Rotation	498
4	<i>From the Chromospheres to the Transition Regions</i>	498
4.1	The Linsky–Haisch Dividing Line	500
5	<i>Active Regions</i>	503
5.1	Starspots and Faculae	503
5.1.1	Methods to Derive Information on Starspots	503
5.1.2	Starspots’ Main Parameters	505
5.1.3	Starspots in Solar-Like Stars	507
5.2	Chromospheric <i>Plages</i>	509
5.2.1	Direct Imaging of the Chromosphere	509
5.2.2	Spectral Imaging	509
5.3	Chromospheric Versus Photospheric Activity	512
6	<i>Transition-Region Lines as Probe for the Plasma Dynamics</i>	512
6.1	UV Line Shift Versus Temperature of Line Formation	513
6.2	Broad Wings in TR Lines	515
6.3	UV Coronal Forbidden Lines	517
7	<i>Stellar Coronae</i>	518
7.1	X-Ray Emission from Stellar Coronae	518
7.1.1	Saturation and Supersaturation	520
7.1.2	Variability of Coronal Emission and Coronal Cycles	521
7.1.3	Coronal Abundances	523
7.1.4	Density and Temperature Structure of Stellar Coronae	524
7.1.5	Coronal Versus Chromospheric and TR Emission	527
7.2	Radio Emission from Stellar Coronae	528
7.2.1	X-Ray Versus Radio Emission	530

8 *Solar-Like Stellar Winds* 531

9 *Stellar Activity Versus Age* 532

10 *Flares* 536

10.1 Flaring Stars 537

10.2 Flare Characteristics 537

10.3 Flare Classification 539

10.4 The Flare “Standard Model” and the Neupert Effect 539

10.5 Signature of Coronal Mass Ejections 542

10.6 Flares and Coronal Heating 542

10.7 Superflares 543

11 *Stellar Magnetic Activity and Extrasolar Planets* 543

11.1 Stellar Magnetic Activity as Source of Noise 544

11.2 Star Planet Interaction 545

11.3 Stellar Activity Effects on Exoplanet Atmospheres 545

Acknowledgments 547

References 547

Abstract: The atmospheres of late-type stars are site of strong magnetic fields responsible for their nonradiative heating. The stellar magnetic activity phenomena are in many cases analogous to those observed on the Sun, but very different scenarios from the solar case are also observed. Magnetic fields drive hot plasma upward, where it radiates in faculae and plagues, leaving dark spots in the photosphere at the foot points of the flux tubes. Occasionally, magnetic fields reconnect and produce magnetic free energy and particles, in form of flares, and coronal mass ejections. Avalanches of low-energy flares – nanoflares – are considered the most viable sources of heating and sustaining of the outer atmospheric layers: chromosphere, transition region, and corona. Magnetic activity is produced by some form of dynamo at work in the stellar interior, whose basic ingredient is the coexistence of a convective zone under the stellar surface and the presence of a differential rotation regime. The whole magnetic activity phenomenology is highly variable in time. Active regions grow and evolve in days to weeks, but are also site of transient phenomena – e.g., flares – evolving on time scales of few seconds to several hours. The global level of magnetic activity changes in time following cycles as it happens for the Sun and decreases with age. The time scale of stellar variability, the amount of involved energy, and the quality of the involved photons are the key observables that can be used as tools for understanding the observed phenomena. This chapter is dedicated to the illustration of the main diagnostics for stellar activity, the characteristics of active regions, and the properties of the chromospheres, transition regions, coronae, flares, and solar-like stellar winds. Finally, we discuss how stellar activity affects exoplanets detection and characterization and the impact of stellar magnetic activity in the formation and evolution of planetary atmospheres.

Keywords: Radio continuum: stars; Stars: abundances; Stars: activity; (Stars:) binaries: general; (Stars:) brown dwarfs; Stars: chromospheres; Stars: coronae; Stars: evolution; Stars: flare; Stars: late-type; Stars: low-mass; Stars: magnetic field; (Stars): planetary systems; Stars: pre-main sequence; Stars: solar-type; (Stars:) starspots; Stars: winds, outflows; Ultraviolet: stars; X-rays: stars

List of Abbreviations: *ANS*, Astronomical Netherlands Satellite; *ASCA*, Advanced Spacecraft for Cosmology Astrophysics; *BDs*, Brown Dwarfs; *CCD*, Charged Coupled Device; *CME*, Coronal Mass Ejection; *COS*, Cosmic Origins Spectrograph; *COUP*, Chandra Orion Ultra-deep Project; *CTTS*, Classical T-Tauri stars; *DEM*, Differential Emission Measure; *DI*, Doppler Imaging; *DIT*, Doppler Imaging Technique; *EUV*, Extreme Ultraviolet; *EUVE*, Extreme Ultraviolet Explorer; *FIP*, First Ionization Potential; *FUSE*, Far Ultraviolet Spectroscopic Explorer; *FUV*, Far Ultraviolet; *GALEX*, Galaxy Evolution Explorer; *HST*, Hubble Space Telescope; *IRT*, Infrared Triplet; *ISM*, Interstellar Medium; *IUE*, International Ultraviolet Observatory; *MM*, Mounder Minimum; *MWO*, Mt. Wilson Observatory; *NUV*, Near Ultraviolet; *ONC*, Orion Nebula Cluster; *PMS*, Pre-Main Sequence; *PRD*, Partial Redistribution; *PSPC*, Position Sensitive Proportional Counters; *ROSAT*, Roentgen Satellite; *SOHO*, Solar and Heliospheric Observatory; *SPI*, Star-Planet Interaction; *SPMI*, Star-Planet Magnetic Interaction; *STIS*, Space Telescope Imaging Spectrograph; *SUMER*, Solar Ultraviolet Measurements of Emitted Radiation; *TR*, Transition Region; *TTs*, T-Tauri Stars; *UV*, Ultraviolet; *VLA*, Very Large Array; *VLBA*, Very Long Baseline Array; *VLBI*, Very Long Baseline Interferometry; *VUV*, Vacuum Ultraviolet; *WB*, Wilson-Bappu (effect); *WFC*, Wide-Field Camera; *WTTS*, Weak-Line T-Tauri stars; *ZAMS*, Zero-Age Main Sequence; *ZDI*, Zeeman-Doppler Imaging

1 Introduction

Stellar activity is a collective name used to group the variability phenomena observed in the outer atmospheres of late-type stars mainly due to the presence of highly structured magnetic fields emerging from the convective envelope. Being the Sun the only star whose magnetic activity phenomena can be truly observed in great spatial detail, the interpretation of stellar activity is generally based on the “solar-stellar connection” paradigm; i.e., most of the variability phenomena observed in late-type stars are decoded using the solar analogous phenomena and their interpretation as reference tools. This approach has been successful in emphasizing not only the real analogies with the Sun but also to highlight the differences with respect to the solar case, i.e., those situations that cannot be understood in the framework of the “solar-stellar connection” but deserve to take into account different processes and scenarios. On the other hand, the investigation on stellar magnetic activity provides us with information useful to better model phenomena observed on the Sun, because any good theory developed for the solar case must survive the comparison with real data when stellar mass, radius, rotational velocity, internal structure, evolutionary state, and/or other stellar parameters change.

Since stellar activity is observed in stars having surface convection zones, the origin of the activity phenomena has been initially attributed to dissipating acoustic waves arising from the convective layer beneath the photosphere (e.g., Schwarzschild 1948). However, it was rapidly understood that the bulk of the observed activity phenomena are due to the dissipation of magnetic energy when magnetic fields, present on the surface of these stars, evolve from complex to simple topologies with a lower content of stored energy. For this reason, *magnetic activity* is often used as a synonym of *stellar activity*.

Magnetic activity is produced by a physical mechanism known as dynamo, whose basic ingredient is the coexistence of a convective zone under the stellar surface and the presence of a differential rotation regime. As shown by Parker (1977), the coupling between differential rotation and turbulent convection produces the intensification of the magnetic field and its periodic inversion. It is widely accepted that dynamo in the Sun is an interface dynamo, requiring a radiative core beneath the convective envelope ($\alpha - \Omega$ dynamo, cf. e.g., Spiegel and Zahn 1992). In fully convective stars and brown dwarfs, a distributed dynamo (α^2 dynamo, e.g., Chabrier and Küker 2006) is invoked to explain the presence of magnetic activity.

Our knowledge of stellar structure indicates main-sequence stars less massive than about 1.5 solar masses to have convective zone beneath the photosphere, and in fact, stellar activity is observed in stars of spectral type later than middle A (Simon et al. 2002). A description of how dynamo works and how the continuous regeneration of poloidal and toroidal magnetic fields takes place is beyond the scope of the present chapter. We instead focus here on the effects of the dynamo on the external stellar atmospheres: photosphere, chromosphere, transition region, and corona; i.e., those layers where the temperature increases from the minimum reached at the top of the photosphere to million degrees in the corona, while plasma density continuously decreases. The dynamo generates magnetic flux tubes; these buoyantly rise through the convection zone, penetrate the stellar surface, and become braided and twisted by surface velocity fields.¹ The magnetic field drives plasma motion inside the flux tubes, so that hot plasma is

¹When the ratio of gas pressure to magnetic pressure, $\beta = 8\pi p/B^2$, is $\beta > 1$, then the gas pressure dominates over the magnetic pressure, and the magnetic field is dominated by the plasma motion. This is the case in the photosphere at the base of the field lines. Instead, $\beta \ll 1$ occurs in the mid-corona, where plasma is dominated by the magnetic field. In the upper corona, it becomes again $\beta > 1$ (Gary 2001).

transported upward, where it radiates in faculae and plages, while the foot points of the flux tubes in the photosphere appear dark (spots). Occasionally, magnetic fields reconnect and produce magnetic free energy and particles, i.e., coronal mass ejections (CMEs), flares, and micro- and nanoflares, the last eventually heating and sustaining of the outer atmospheric layers. In short, the magnetic loop structures, arising from the surface upward to the external atmospheres, are the cause of the zoo of spots, faculae, flares, and CMEs, i.e., all the different magnetic features we observe with a great detail on the Sun. Each observed feature is characterized by its own temporal evolution and energetic, both in quantity and quality. As a consequence, the time scale of stellar variability, the amount of involved energy, and the quality of the involved photons can be used as tools for understanding the observed phenomena. Magnetic fields are also responsible for slowing down cool stars because winds magnetically coupled to the stellar surface carry away angular momentum (Schatzman 1962).

Active regions generated by the emergence of magnetic flux have different characteristics depending from the layer where we look in the stellar atmosphere. *Starspots* and *faculae* are detectable in the photosphere, with increasing contribution from the latter going toward the chromosphere. Actually, on the Sun, variations in the optical continuum are dominated by photospheric cool spots and warm faculae but more than 60% of the bolometric variations over the solar cycle are produced at wavelengths below 400 nm (e.g., Krivova et al. 2006), where the bright regions in the chromosphere and transition regions, called *plages* (French origin word), are the main contributors to variability. These are normally only detectable in the monochromatic light of spectral lines, such as H- α , Ca II H&K, or Mg II h&k.

The whole “magnetic activity” phenomenology is highly variable in time. Active regions grow and evolve in days to weeks, but are also site of transient phenomena, evolving on time scales of few seconds to several hours, like flares, microflares, and CMEs. The global level of magnetic activity changes in time following cycles as it happens for the Sun. Hence, depending upon the peculiar aspect, one can be interested to study time scales ranging from seconds to years.

The covering factor of active regions, as well as the energy output of stellar flares, can be orders of magnitude more intense in stars than on the Sun. This is particularly true for very fast rotators such as young stars and stars locked in close-binary systems. As discussed later in this chapter, the ultimate result of magnetic activity on stellar evolution is the stellar spin down driven by the loss of angular momentum carried out by the coronal stellar winds. As the star spins down, the level of its magnetic activity decreases, in a self-attenuation feedback loop.

Useful data to study magnetic activity come from the whole electromagnetic spectrum, from the radio wavelengths, where we measure the gyrosynchrotron radiation arising from the quiescent and flaring corona, to the optical region, where important signatures are the Balmer lines, the Ca II IRT, and the H&K lines, that are powerful diagnostic of the chromosphere. The UV region is of paramount relevance, being rich of emission lines whose formation temperature goes from about 7,000 to million degrees, thus sampling the different layers of the atmosphere, from the chromosphere to the transition region and corona. Finally, X-rays provide the best tool to study the coronal thermal plasma.

If the dynamo is the primary engine that produces magnetic flux tubes, the details of the mechanisms heating the outer atmosphere are still unclear. Models either rely on the dissipation of electric currents or on the damping of a variety of waves traveling through the magnetic field (Narain and Ulmschneider 1990). Among these, Alfvén waves are considered a viable heating agent because of their capability to transport through the chromosphere sufficient energy to heat and maintain a corona. Recently, Antolin and Shibata (2010) show a strong evidence for the existence of energetic Alfvén waves in the solar corona establishing several constraints on

Alfvén wave heating as a coronal heating mechanism, especially for active region loops. To summarize, magnetic fields have a role in coronal heating both via the direct liberation of magnetic energy by reconnection and through dissipation of Alfvén waves.

The presence of an emission core in the Ca II K line of Arcturus is the first report on stellar magnetic activity in the literature (Eberhard and Schwarzschild 1913). These authors recognized the solar-like nature of the observed phenomenon; however, the interest in investigating magnetic activity in other stars than the Sun arose only several years later, when Olin C. Wilson started a project to monitor Ca II H&K lines at Mt. Wilson Observatory in 1950.

The Wilson's project (e.g., Baliunas et al. 1998) remains one of the most conspicuous among many observing programs providing bulks of fundamental data for stellar magnetic activity studies. I recall also (i) the Lowell and Fairborn observatories projects (e.g., Radick et al. 1998; Messina and Guinan 2002, 2003; Lockwood et al. 2007), aimed at the detection of photospheric and chromospheric variability, and (ii) the many photometric and spectroscopic monitoring projects carried out in several European observatories to get maps of spotted photospheres (e.g., Rodonò 1992; Rodonò and Cutispoto 1994; Lanza et al. 1998; Rodonò et al. 2000; Messina 2008; Vida et al. 2007; Strassmeier 2009) and of magnetic field patterns (e.g., Donati et al. 2003a; Donati et al. 2003b; Donati et al. 2007). Recently, space observations from MOST, CoRoT, and Kepler are providing very accurate optical variability data with high cadence and months duration (e.g., Walker et al. 2007, 2008; Lanza et al. 2009a, b, 2010; Brown et al. 2011). Photometric and spectroscopic patrol of stellar flares from the ground have been carried out since the 1960s in many places (cf., Gershberg 2005 and references therein). Finally, active regions and flares in the upper chromosphere, transition regions and coronae, as well as solar-like winds have been studied thanks to UV and EUV data from, e.g., IUE, HST, GALEX, EUVE, and FUSE; X-ray data from, e.g., Einstein, Rosat, ASCA, Beppo-SAX, Chandra, and XMM-Newton; and radio observations from VLA, VLBI, VLBA, and single dishes.

In ► Sect. 2, we introduce the typical objects showing magnetic activity. In ► Sects. 3 and ► 4, the observational signatures of chromosphere and transition region are discussed. Properties of active regions are given in ► Sect. 5, while in ► Sect. 6, the plasma dynamic is analyzed. The emission from stellar coronae is illustrated in ► Sect. 7, while in ► Sect. 8, the solar-like stellar winds are discussed. ► Section 9 is dedicated to the variation of stellar activity with stellar age. Flares are the subject of ► Sect. 10. Finally, the relevance of magnetic activity in the exoplanets research is given in ► Sect. 11.

We refer the reader to the chapter “The Sun as a star” of Mason and Del Zanna in this volume for detailed description of the magnetic activity phenomena as observed on the Sun.

2 Objects Showing Magnetic Activity

Stellar magnetic activity has been detected in all late-type stars cooler than $T_{\text{eff}} \sim 8,300$ K, i.e., later than A4V (Simon et al. 2002). For typical characteristics, these can be grouped as following:

- Sunlike stars
- M-type dwarfs
- Rapidly rotating stars, i.e., stars in close binaries (e.g., BY Dra, RS CVn, FK Com) and young PMS stars
- Red giants and supergiants

Most of these stars have broadband optical flux rotationally modulated because of the presence of starspots in their photosphere. Their optical spectrum is characterized by variable Ca II H&K and Balmer emission-line cores formed in the chromosphere. All apart red giants and supergiants have transition regions and coronae that produce emission lines in the EUV-UV, thermal X-ray, and nonthermal radio emission.

Sunlike stars are a very broad class including from late F to late G type dwarfs and subgiants. Following Cayrel de Strobel and Bentolila (1989), solar-like stars are usually classified as (i) *Solar analogs*: these are pop I dwarfs having main properties similar to those of the Sun but can be of different age, and (ii) *solar twins*: these have all physical parameters almost identical to those of the Sun.

Active *M-type dwarfs* include the so-called flare stars or UV Cet-type from the name of the class prototype. These are red dwarfs, typically dKe-dMe, showing sudden luminosity increase observable from the optical to the radio wavelengths. In the optical spectrum, the increase can be 1–7 mag peeking the U region. Flares last from few minutes to few hours. Ca II H&K and Balmer emission lines, that are filled in or in emission also during quiescence, become very strong emission features during flares. Chromospheres apparently exist also in very-low-activity M stars, according to semiempirical models by Mauas et al. (1997). On the other hand, X-ray and radio flares have been observed from fully convective late M dwarfs and brown dwarfs, suggesting a nonsolar dynamo mechanism at work in these object.

Rapidly rotating stars: several type of binaries show high level of magnetic activity maintained thanks to their forced spin-orbit coupling. Among the active binaries, we list (i) *BY Dra-type* stars which are binary systems made by dKe-dMe stars. Their optical light curve have period from a fraction to a hundreds of days and amplitude from few hundredths to half a magnitude. Flares are typically observed also in these stars; (ii) *RS CVn-type* variables which are detached binaries made of a F-G V-IV plus a K IV-III, both stars moderately or very active. Period of variability is from a day to tens of days. Spots have been identified also close to the pole, having covering factor up to 50% and evolving in weeks to months; (iii) *W Uma-type* variables: these are eclipsing contact binaries made by late-type star having period between 5 and 20 h ($v \sin i \sim 100\text{--}200 \text{ km s}^{-1}$). They constitute about the 95% of eclipsing binaries close to the Sun. W Uma are classified as A-type, which have components of similar temperature, or W-type which have the primary cooler, possibly because the presence of dark spots on the surface (e.g., see Barnes et al. 2004); (iv) *FK Com-type* stars which are G-K III rapidly rotating single stars showing very strong magnetic activity. Their evolutionary state has been proposed to be the late evolution of a contact binaries (W Uma-type), with the coalescence of the original stars. However, this scenario is not consistent with the presence of Li lines in their spectra, that is generally interpreted as signature of youth. Fekel and Balachandran (1993), suggested that convective motions penetrate to the core carrying Lithium and angular momentum from the inside to the surface. From the observational point of view, the properties of FK Com-type stars are not different from the RS CVn-type ones; (v) *Algols* which are eclipsing binaries with the primary being a main-sequence hot star (B–F) and the secondary a cold subgiant (G–K) that has filled in its Roche lobe. Thanks to the rapid rotation, Algols show magnetic activity signatures as RS CVns (White and Marshall 1983; Stewart et al. 1989; Umana et al. 1989). For contrast reasons, photospheric starspots are detectable only during primary eclipses, when the hot component is hidden. Because of tidal synchronization, this implies that starspots are observed only on one hemisphere of Algols secondary stars.


Among the rapid rotators are the pre-main-sequence low-mass *T Tauri stars* (TTS) which show signatures of magnetic activities often mixed with accretion phenomena, with circumstellar gas and dust disk surviving from the formation process. TTSs have large starspots and show flares and accretion process signatures. General reviews on TTS have been given by Bertout (1989), Appenzeller and Mundt (1989), and Petrov (2003), while for their coronal activity, we refer to Feigelson et al. (2007).

Red giants and supergiants, where a low-speed cool wind is detected with a mass-loss rate $>10^{-9} \text{ M}_{\odot} \text{ year}^{-1}$, have chromospheres but do not show coronal signatures. However, yellow supergiants, the ones at the bottom limit of the wind zone, show both winds and coronal signatures (Ayres et al. 1981a); these are called *Hybrids*.

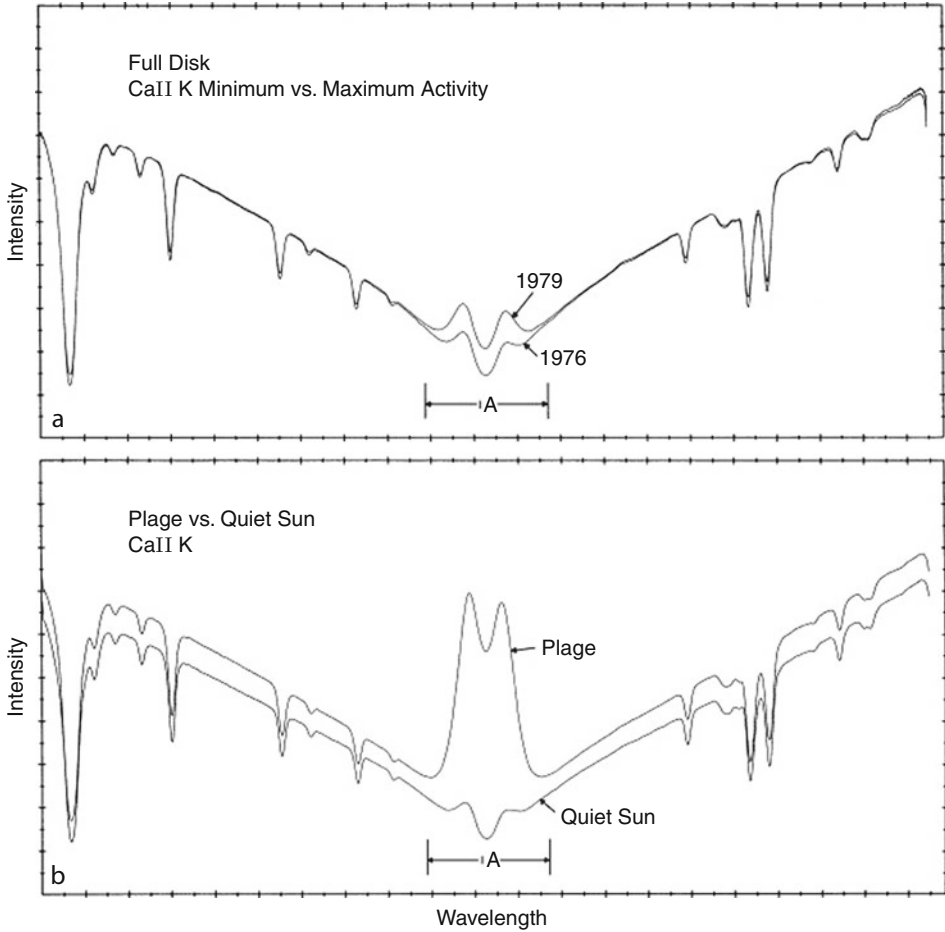
3 Stellar Chromospheres

Main diagnostics of stellar chromospheres are optically thick Ca II H&K lines, Ca II IRT, H-alpha, whose core is formed in the chromosphere, and the many UV lines, e.g., O I, C I, and Fe II, with formation temperature at around 10^4 K .

3.1 Ca II H&K and Wilson–Bappu Relation

The inversion of the temperature gradient beyond the photosphere determines the filling of the line core of resonance lines, the strongest of which in the optical spectrum are those of singly ionized calcium at 392.3 and 396.8 nm, i.e., the Ca II H&K lines. In late-type stars, these lines are very opaque; the core forms in the chromosphere and can be used as a signature of the chromospheric properties (cf. Linsky and Avrett 1970 for an early but very accurate review on the properties of the Ca II H&K lines and theories for its detailed formation and behavior). The Ca II K line profiles of the full solar disk at two epochs, maximum and minimum of the solar cycle, respectively, of a solar plage, and of the quiet Sun, obtained at the same distance from the solar limb, are shown in  Fig. 10-1; it is apparent that the emission component of the Ca II H&K lines is very small in full disk or in the spectrum of quiet Sun than it is in the spectrum of a plage. This behavior anticipates that Ca II H&K lines are a powerful diagnostic of stellar activity being very sensible to active region filling factors. As firstly annotated by Hale and Ellerman (1904), the minimum of the broad absorption line, produced in the photosphere, is called (H)K1, with the indexes *b* and *r* indicating the blue and red sides with respect to the central emission; the emission core, that originates in the chromosphere, is called (H)K2, with the suffixes *b* and *r* for the blue and red sides with respect to the central absorption; the latter, often referred to as *central reversal* and annotated as (H)K3, occurs because photons produced in the deeper chromosphere are removed from the line of sight in the upper layers, where the source function decreases.

Observations of stellar Ca II H&K spectra are largely used to identify the level of chromospheric activity in late-type stars and its variability. Pioneering studies making use of high-resolution stellar spectra which have been analyzed by using the Sun as template show several intriguing aspects: for increasing stellar activity level, the separation of K1 minima increases and the distance between the K2 maxima decreases (Ayres et al. 1981a; Pasquini et al. 1990); however, the FWHM of the emission core is independent from the activity level,



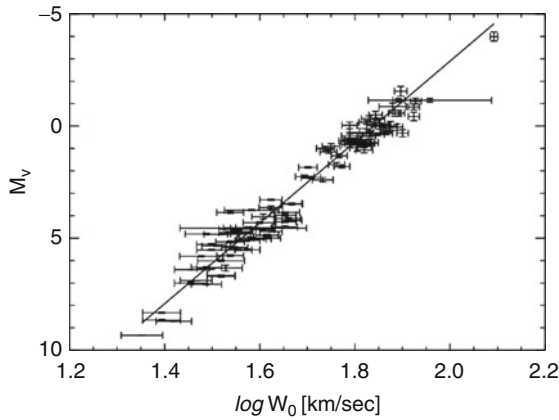
■ Fig. 10-1

Ca II K line profiles of the full solar disk at minimum and maximum of the solar activity (*top panel*), of a solar plage, and of the quiet Sun obtained at the same distance from the solar limb (*bottom panel*) (Elaborated from White and Livingston 1981)

while it is strongly correlated with the absolute visual magnitude of the star. The last relation is called Wilson–Bappu (hereafter WB) effect, from the authors who first noticed and published such a remarkable empirical correlation. WB effect works over about 15 stellar magnitudes (Wilson and Vainu Bappu 1957; Linsky 1999 for an excellent review). Originally, the WB effect was expressed by the following relation:

$$M_V = 27.59 - 14.94 \log W_0 \quad (10.1)$$

where M_V is the absolute visual luminosity, and W_0 (km s^{-1}) was the separation between the outer edges of the Ca II K emission features on the photographic spectrograms, a quantity close to the full width at half maximum of the emission core. Several studies have redetermined this



■ Fig. 10-2

Wilson-Bappu effect relation as obtained by Pace et al. (2003) by using Ca II K high-resolution spectra of a sample of 119 nearby stars. The error bars represent standard errors in both the coordinates

relation from different data sets. Pace et al. (2003) have derived the following relation:

$$M_V = 33.2 - 18.0 \log W_0 \quad (10.2)$$

where W_0 (km s^{-1}) is measured with a well-defined quantitative criterion from high-resolution and high S/N CCD spectra of 119 nearby stars, whose absolute visual magnitudes are taken from the Hipparcos database (► Fig. 10-2). The root mean square error of relation (► 10.2) – 0.6 mag – is mainly due to measurement errors, intrinsic variability of W_0 , and possible dependence on the metallicity that, according to the authors, is clearly noticeable for metallicities below $(\text{Fe}/\text{H}) \sim -0.4$.

Although the WB relation does not provide accurate distances to single stars, it can be confidently used to determine distances to clusters and aggregates, provided that a sufficient number of stars can be observed and metallicities (Fe/H) are not lower than -0.4 (Pace et al. 2003).

WB-like relations between width and luminosity have been measured for other chromospheric emission lines, such as the Mg II h&k resonance lines at 280.3 and 279.6 nm. For instance, Cassatella et al. (2001) determined the WB relation for the Mg II k line at 279.6 nm, based on a selection of 230 stars observed by both the IUE and Hipparcos satellites, covering a wide range of spectral types (F to M) and absolute visual magnitudes ($-5.4 \leq M_V \leq 9.0$); despite the coefficients of the WB relationship critically depend upon the choice of the method of analysis, these authors have not found any dependence of the WB relation from stellar metallicity and effective temperature.

The increase in chromosphere thickness with decreasing gravity in a hydrostatic equilibrium atmosphere appears to be the most convincing explanation for the WB effect, except perhaps for the most luminous stars. The dependence from gravity of the width of chromospheric resonance lines is reproduced by stellar atmosphere models when the partial redistribution (PRD) formulation of the radiative transfer equation is considered (cf. Linsky 1999, and references therein).


3.2 The Activity Indexes


The discovery of the WB effect gave input to the project at Mt. Wilson Observatory consisting of a systematic long-term spectro-photometric monitoring of the Ca II H&K lines in late-type stars. This project led to the discovery of stellar activity cycles and to the measurement of stellar rotation rates in slow rotators, thanks to the modulation of the visibility of chromospheric *plages* due to the stellar rotation. The Mt. Wilson Observatory (MWO) program was started by Olin Wilson in 1966 with observations of low main-sequence solar-like stars scheduled monthly until about 1980 at the 100 in. telescope. With the dedication of a 60 in. telescope to the project, the observations became almost nightly and a sample of post-main-sequence stars was included later among the monitored targets. Over 400,000 observations of more than 2,000 stars have been collected till the end of the project in 2003, allowing to study the chromospheric variability over time scales from days to tens of years. The instrumental setup at Mt. Wilson led to the definition of the “activity index” S :

$$S = \alpha \frac{H + K}{V + R} \quad (10.3)$$

where H and K are two narrow (0.1 nm^2) passbands centered at the Ca II H&K line centers, respectively, while V and R are two 2.0-nm-wide passbands centered at 390.1 and 400.1 nm, respectively, to sample the photospheric flux. The quantity α is a calibration factor (cf. Baliunas et al. (1995) for a summary of the Mt. Wilson program-specific measurements and results). Baliunas et al. (1998) show that of the whole MWO sample:

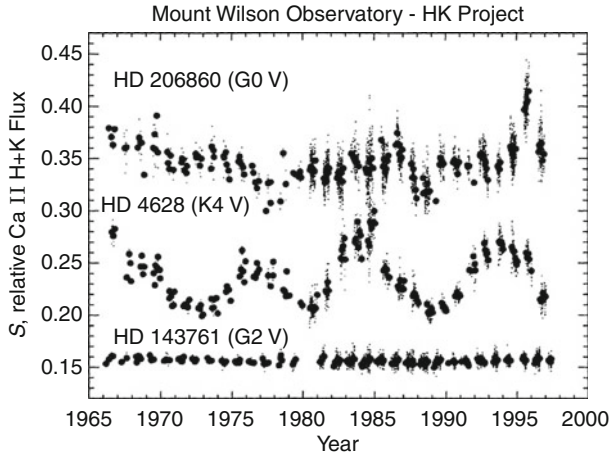
1. About 60% have periodic, or apparently periodic, variation in the S index, without any obvious dependence from spectral types or age. Activity cycles can be longer than the 30 years of the MWO project, with the older stars never showing cycles shorter than 7 years. Some of the monitored targets show two independent periodicities.
2. About 25% are variable but no obvious periodicity has been found. Most of the targets in this group are young, very active rapidly rotating stars.
3. About 15% are constant with time, possibly in a Maunder-Minimum-like state. These have been dubbed also as “flat activity stars.”

Example of the above class of objects are shown in  Fig. 10-3.

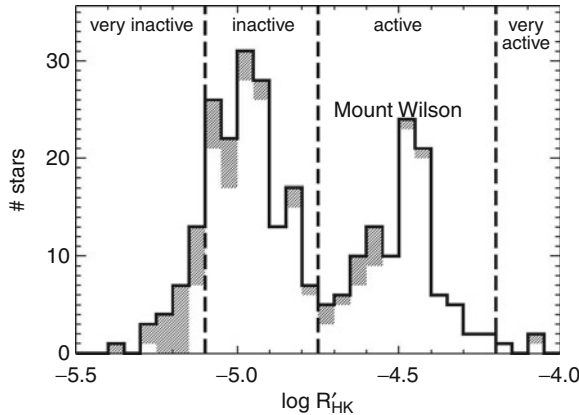
Data from MWO have been used to search for dwarf stars in flat activity state, analogous to the low-level activity period of the Sun in the seventeenth century, known as Maunder minimum. For instance, Saar and Baliunas (1992) estimated that 10–15% of the solar-like stars in the MWO sample were detected in states of low activity. Wright (2004) demonstrated that most of these objects were actually slightly evolved stars, which have lower levels of activity than main-sequence stars (cf.,  Fig. 10-4). The typical criterion used to define a star as an MM candidate is it has $\log R_{HK} < -5.1$. According to Saar (2006), this limit has been defined including in the sample subgiants, so that it is possible that solar-like stars in an MM phase do not have a so low value of $\log R_{HK}$. Moreover, the minimum R_{HK} observed in dwarfs resulted to be a decreasing function of stellar metallicity (see Hall et al. 2009). Taking this effect into account, a few couple of targets are now considered to be MM candidates. For a complete review of stellar magnetic cycles, see Lanza (2010).

S is a dimensionless index, dependent on the stellar color, with S increasing toward later spectral type as a consequence of the decrease of the photospheric continuum to which the

²H&K are measured in 0.2-nm-wide bands in case of giant stars whose line core is wide for the WB effect.



■ Fig. 10-3
Example of chromospheric variable (*top and middle panels*) and non-variables (*bottom panel*) objects detected by the MWO project (From Baliunas et al. 1998)



■ Fig. 10-4
Histogram of activity levels for stars with $0.55 < B - V < 0.9$ in the Mt. Wilson sample. The shaded regions represent stars for which $M_V > 1$, which is a conservative limit for identifying evolved stars that should exclude all but the most metal-rich main-sequence stars. Almost all of the “very inactive” stars have $M_V > 1$ and are subgiants and not main-sequence stars in a Maunder minimum (From Wright 2004)

denominator of relation (10.3) is proportional. For the S-flux calibration, using flux-calibrated spectra, see Cincunegui et al. (2007).

A more physical quantity to measure the activity index is $R_{HK} = \frac{F_{HK}}{\sigma T_{\text{eff}}^4}$, where F_{HK} is the total flux per cm^2 at the stellar surface in the H and K passbands. R_{HK} is related to the

activity index S by $R_{\text{HK}} = C_{B-V} S$, with C_{B-V} a color-dependent correction factor (cf. Cincunegui et al. 2007). Eventually, to highlight the pure chromospheric contribution, R'_{HK} was proposed as

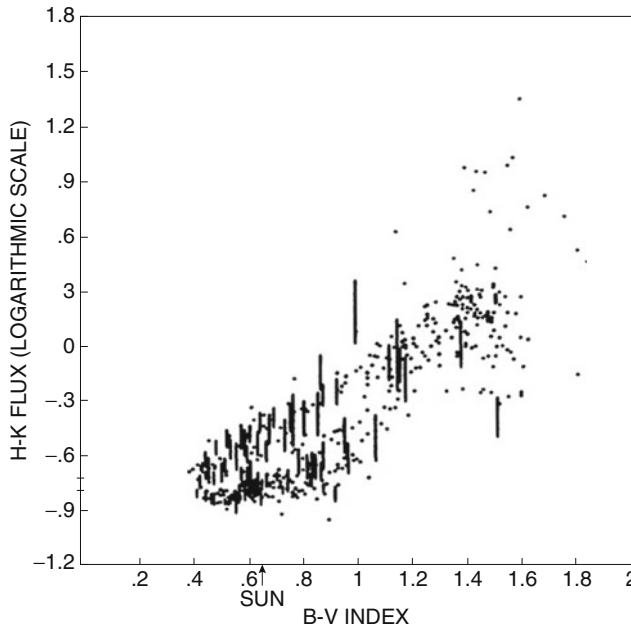
$$R'_{\text{HK}} = R_{\text{HK}} - R_{\text{phot}}. \quad (10.4)$$

R'_{HK} essentially gives the fraction of a star's bolometric luminosity radiated as chromospheric H and K emission; it is typically expressed in log units and ranges from about -4.4 to -5.1 for very active to highly inactive stars.

We will make use of this quantity in the coming sections.

3.2.1 The Vaughan-Preston Gap

Typical S -values range from ~ 0.1 to 1.0 or more (for dMe stars). For each color, a large spread in the S index is found, as shown in [Fig. 10-5](#). Eventually, for a given spectral class, both stars with low and high chromospheric activity can be found. Vaughan and Preston (1980) show that there are relatively large numbers of stars in the $B - V$ range 0.45 – 1.0 with either strong or weak chromospheric emission but relatively few with moderate emission. This discontinuity in the distribution of S versus $B - V$ has been dubbed Vaughan-Preston gap. Some authors questioned the statistical significance of such a gap (e.g., Hartmann et al. 1984); however, the bimodal behavior has been confirmed by Henry et al. (1996) who surveyed the Ca II H&K core



■ Fig. 10-5

The MWO H&K flux, expressed as $\log S$ index as function of $B - V$. For each spectral class, there is a wide spread of activity indexes. For a subsample of targets, vertical bars show the intrinsic variability (From Vaughan 1983)

of about 800 late-type stars in the southern hemisphere, hence with an entirely different instrumental setup and stellar sample. Several hypotheses have been put forward to explain the origin of the Vaughan–Preston gap. Since chromospheric activity decreases with increasing stellar age, for instance, the Vaughan–Preston gap could represent a change in the nature of activity at a certain stellar age. Similarly, it could be interpreted as due to a fluctuation in the local stellar birthrate and therefore as originated by a slowly varying star formation rate. Alternatively, since activity is closely related to the stellar dynamo generating the surface magnetic fields, the Vaughan–Preston gap could be due either to a rapid spin down at some point of the stellar evolution or to an abrupt change in the efficiency of the dynamo for a given rotation period (Durney et al. 1981; Böhm-Vitense 2007). Finally, the Vaughan–Preston gap could not be due to a discontinuity in the chromospheric heating but to a change in the response of the Ca II emission to this heating.

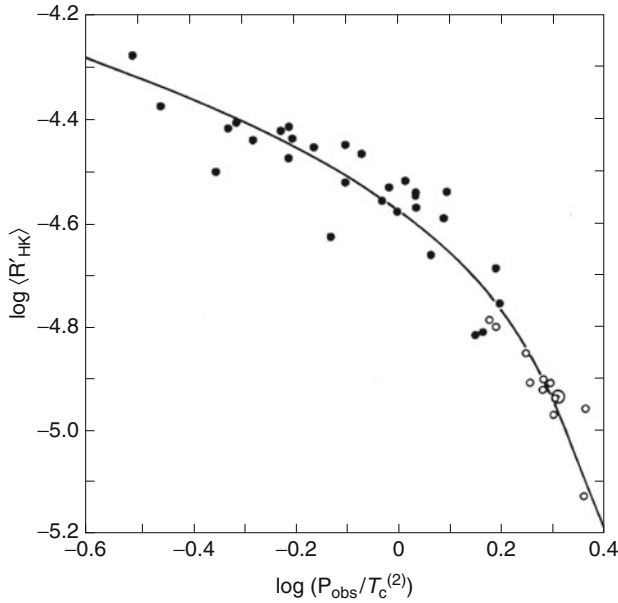
3.3 Chromospheric Activity Versus Rotation

Mt. Wilson observations taken after 1980, when nightly monitoring series have been secured, have been used to identify rotation periods, and in a few cases also differential rotation. Noyes et al. (1984) shown that R'_{HK} is correlated with the stellar rotation period P_{rot} , with the Ca II emission decreasing with increasing rotational period. Being the core of Ca II H&K lines a proxy for the efficiency of the stellar dynamo, the correlation with P_{rot} is not surprising, since qualitatively the dynamo is more effective for fast rotator. A more tight relation has been found between R'_{HK} and the Rossby number $R_0 = P_{\text{rot}}/\tau_c$, where τ_c is the convective turnover time computed at the base of the convective zone assuming a ratio $\alpha = 2$ of the mixing length to scale height (Noyes et al. 1984). These results suggest a more intimate connection between the activity index and the dynamo mechanism.

The relation shown in [Fig. 10-6](#) suggests that rotational periods can be inferred from a measurement of the chromospheric emission. Being rotational periods linked to stellar age (Skumanich 1972), in principle, a measurement of the level of chromospheric activity could yield the precious information on stellar age. As we will discuss later, many other diagnostics – e.g., X-ray flux – lead to the conclusion that magnetic activity level changes during the stellar life. We will return on this topic in [Sect. 9](#).

4 From the Chromospheres to the Transition Regions

Spectral lines falling in the UV (900–3,200 Å) and EUV (100–900 Å) bands have formation temperature ranging from few 10^4 K to few 10^6 K. These spectral ranges include the CO and H₂ molecule bands, the former formed in the so-called COmosphere (see Ayres 2002) and the latter due to fluorescent pumping from high-temperature lines; the C I and O I lines which form at typical chromospheric temperatures; the C II–IV, N IV, O III–V, Si II–IV lines which form in the transition region; and some high-temperature lines, e.g., Fe XII, and Fe XXI, which form in the corona. An inventory of stellar UV emission lines detected by FUSE and HST in late-type magnetically active stars is shown in [Fig. 10-7](#). An example of a high-resolution (2.6 km s^{-1}) UV spectrum of a solar-type star is the one of α Cen A (G2 V) acquired by the Space Telescope Imaging Spectrograph (STIS) on HST and shown in [Fig. 10-8](#). A total of 662 emission

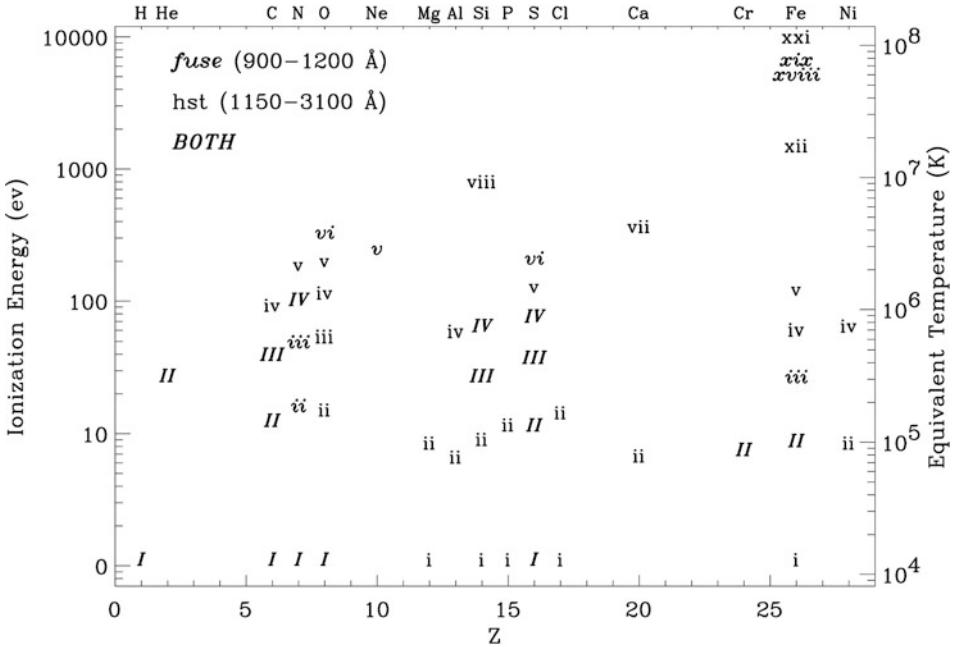


■ Fig. 10-6

Correlation between R'_{HK} and P_{rot}/τ_c (Rossby number), where τ_c is the convective turnover time computed at the base of the convective zone assuming a ratio $\alpha = 2$ of the mixing length to scale height. Closed and open circles represent “young” and “old” stars, respectively; the Sun position is also indicated (From Noyes et al. 1984)

features – of which 77 due to blends of two or more lines, 71 due to unidentified transitions, and 514 identified as due to single emission lines – have been measured in this spectrum. More than 65% of the 671 identified emission lines are due to Si I, Fe II, and C I (see Pagano et al. 2004, for the whole line list). Several intersystem lines are present in the UV spectrum, including the O IV UV 0.01 intercombination multiplet $2s^2 2p^2 P_J^0 - 2s 2p^2 {}^4P_J$ that are diagnostics of electron density (cf. Del Zanna et al. 2002; Brage et al. 1996, and references therein), the N IV line at 1,486 Å, and the O III line at 1,666 Å.

Moreover, it is on fitting of the FUV continuum that rely the semiempirical models of the solar chromosphere (e.g., Vernazza et al. 1981; Avrett and Loeser 2008; Fontenla et al. 2009). Quantitative analyses of stellar FUV continua have made feasible by the high throughput and very low detector background of the Cosmic Origins Spectrograph (COS) on the Hubble Space Telescope. Linsky et al. (2012) measured the continuum flux of 11 solar-mass stars within 136 pc, finding a clear trend of increasing continuum brightness temperature at all FUV wavelengths with decreasing rotational period, which provides an important measure of magnetic heating rates in stellar chromospheres. Comparison with semiempirical solar flux models shows that the most rapidly rotating solar-mass stars have FUV continuum brightness temperatures similar to the brightest faculae seen on the Sun. The thermal structure of the brightest solar faculae therefore provides a first-order estimate of the thermal structure and heating rate for the most rapidly rotating solar-mass stars.

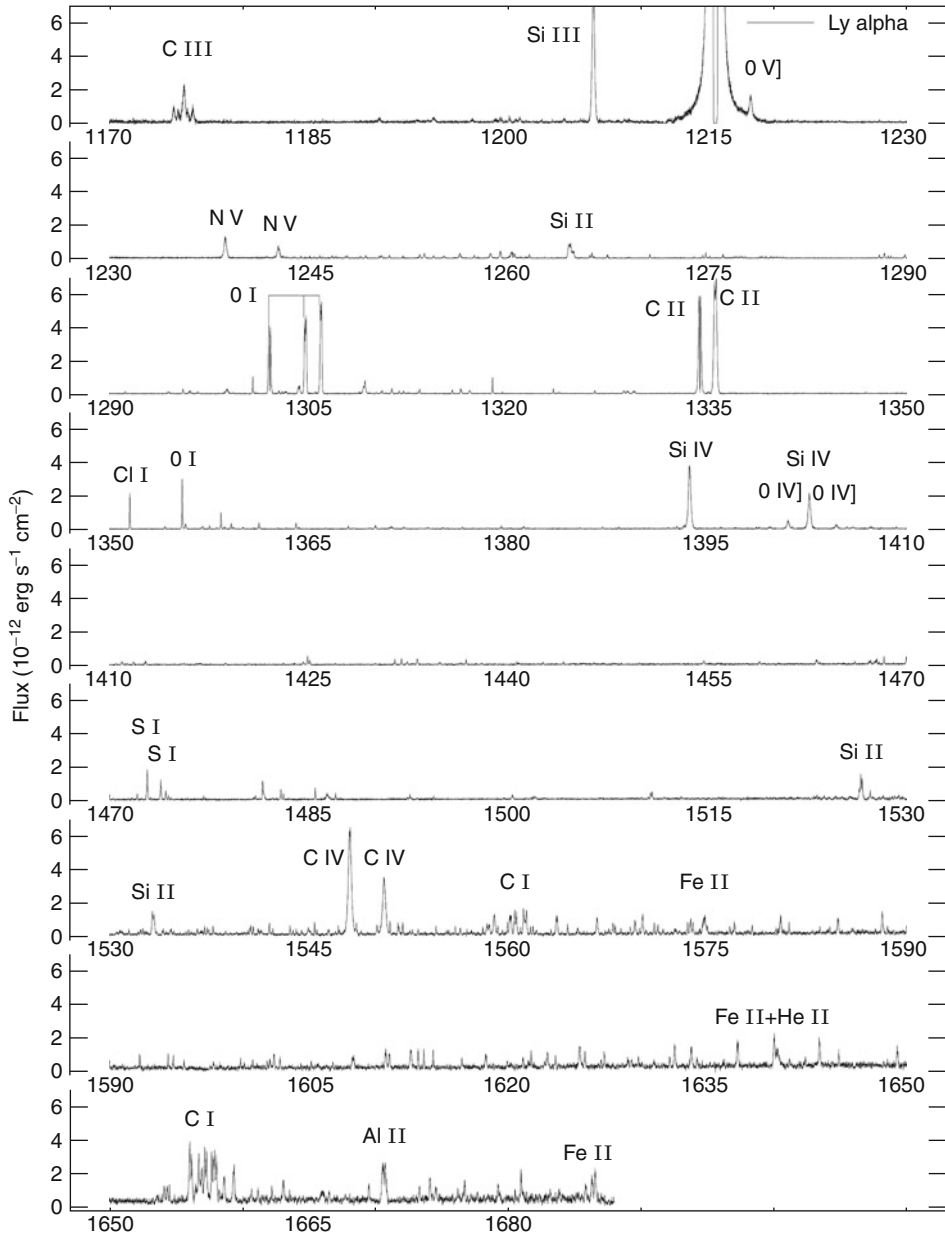


■ Fig. 10-7
Inventory of stellar emission lines detected in UV spectra of magnetically active dwarf stars. Ionization energy (eV) and equivalent temperature (K) are plotted against atomic number, with the elements identified along the top axis (From Redfield et al. 2002)

Catalogs of UV spectra of cool stars published by several authors are listed by Pitts and Imhoff (1989). A catalog of UV spectra of cool active stars acquired by STIS on *HST*, homogeneously processed with state-of-art reduction algorithms, is included in the *StarCAT* database published by Ayres (2010).

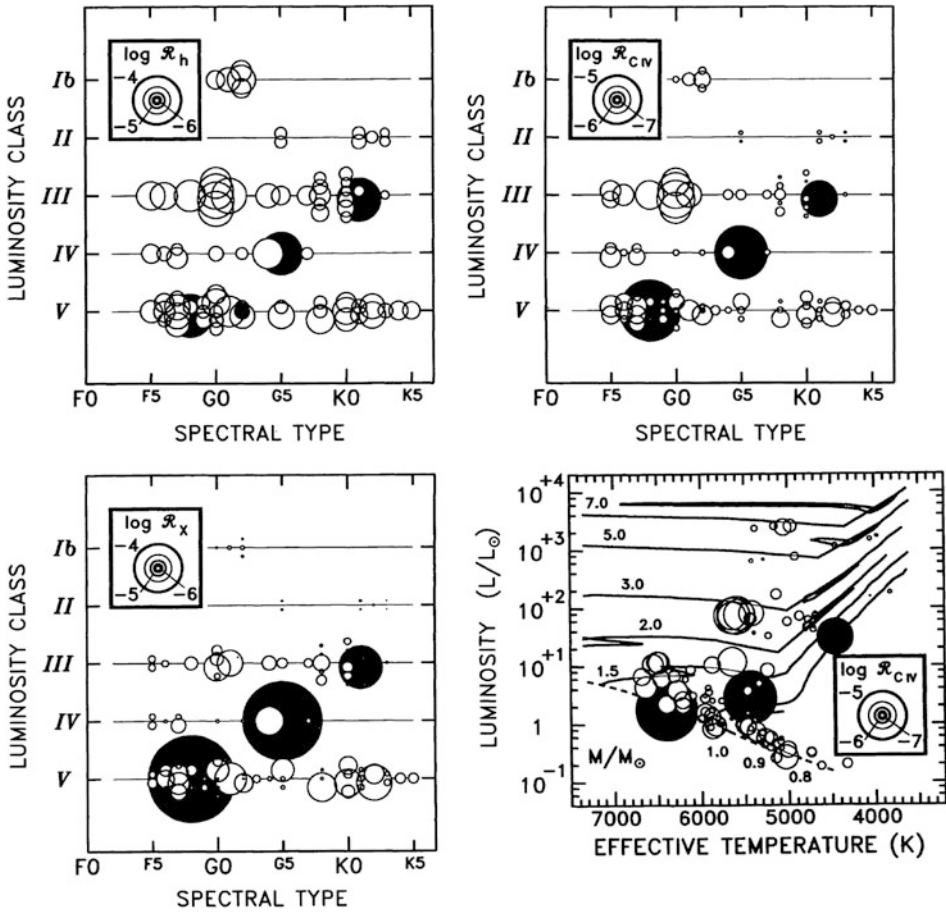
4.1 The Linsky–Haisch Dividing Line

As early as in the first year of operation of the *IUE* satellite, Linsky and Haisch (1979) recognized the presence of two distinct groups in the HR diagram; a *solar-type* group showing spectral lines formed at temperatures indicative of chromospheres, transition regions, and by implication unseen coronae, and a *non-solar-type* group showing lines formed at temperatures not hotter than 10,000–20,000 K, indicative of chromospheres only. In fact, as shown in ● Fig. 10-9, soft X-ray emissions from hot coronae ($T \geq 10^6$ K) are rarely detected to the right of the so-called “Linsky–Haisch dividing line” that bisects the giant branch near spectral type K1 III (see also Ayres et al. 1981a; Haisch and Simon 1982). Linsky and Haisch (1979) suggested the onset of large stellar winds as the origin of the absence of corona. The lack of a corona beyond the dividing line, has been ascribed by Antiochos et al. (1986) to the low surface gravity of red giants which could trigger thermal instabilities having the effect to prevent solar-like hot coronal magnetic loops to exist. Rosner et al. (1995) suggested that the dividing line can be due to a different



■ Fig. 10-8

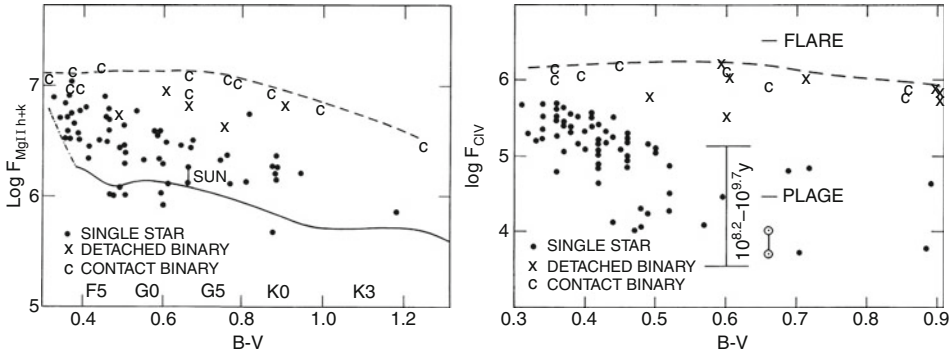
The UV spectrum of α Cen A acquired by *HST/STIS* at resolution of 2.6 km s^{-1} (E140H) between 1,140 and 1,670 Å. Main emission lines are labelled (From Pagano et al. 2004)



■ Fig. 10-9

Intensity of Mg II h, C IV, and soft X-ray for a sample of cool stars versus spectral type and luminosity class (top and bottom-left panels) and intensity of C IV in the HR diagram (bottom-right panel). The dot size is proportional to the intensity, expressed by $R_x = f_x/f_{bol}$ (From Ayres et al. 1995)

nature of the stellar dynamo; the dynamo at work on the blue-ward side of the giant tracks in the Hertzsprung–Russell (H-R) diagram is capable to produce mainly closed magnetic fields, while on the red-ward side, it produces mainly open magnetic fields. If such a topological transition occurs, then these stars naturally segregate into two classes: those having hot coronae on the blue-ward side and those having massive cool winds on the red-ward side, thus leading naturally to the so-called dividing line. The earlier reports of the absence of red giant coronae were questioned by the detection of C IV emission and soft X-ray from γ Dra (K5 III) (Brown et al. 1994) and of coronal proxies (e.g., Si IV with $\log T = 4.8$) in other noncoronal red giants, at levels only a few times lower than typical coronal K0 giants (Ayres et al. 1995). A search for coronal emission by means of *Chandra* X-ray satellite made by Ayres et al. (2003) led to the conclusion that the noncoronal giants Arcturus (K1 III) and Aldebaran (K5 III) have luminosities in the 0.2–2 keV band about two orders of magnitude less than the Sun, despite their surface areas are



■ Fig. 10-10

The Mg II at 2,800 Å and C IV 1,540 Å emission-line fluxes in $\text{erg s}^{-1} \text{cm}^{-2}$ versus $B - V$ (From Vilhu and Walter 1987)

more than 100 times the solar one. Ayres et al. (1997) suggest that coronal loops of inactive red giants like Arcturus could be very compact and merely embedded in an absorbing environment which could significantly attenuate any intrinsic X-ray emissions.

Vilhu and Walter (1987) have measured the fluxes of cool stars of different level of activities in the chromospheric Mg II doublet at 2,800 Å and in the transition region C IV doublet 1,540 Å (cf. ► Fig. 10-10), showing that the chromospheric and transition region emissions do not depend on the photospheric temperature and that span from a low limit (dubbed “basal flux”), defined by the pure acoustic heating of the chromosphere, and an upper one defined by the heating working in a saturation regime, as clearly shown also by X-ray observations, as we discuss later in ► Sect. 7.

5 Active Regions

5.1 Starspots and Faculae

A recent review on stellar spots has been given by Strassmeier (2009), to whom we refer the reader also for a complete compilation of previous reviews and a list of dedicated conferences on the matter. Spots on the stellar surfaces – like sunspots – are characterized by an effective temperature lower than the effective temperature of the unperturbed photosphere and are responsible for a decrease of the disk-integrated flux both in the bolometric band and in the optical passbands. On the contrary, bright faculae, produce an increase of the disk-integrated fluxes. Hence, the time variability of the photospheric flux mainly depends on the intrinsic evolution of starspots and faculae and on the modulation of their visibility due to stellar rotation.

5.1.1 Methods to Derive Information on Starspots

Direct stellar images of active stars are not obtainable at moment. Gilliland and Dupree (1996) used HST to get UV images of the chromosphere of the supergiant star α Ori (Betelgeuse) with

about 50 mas resolution. Imaging the closest solar-like dwarf stars – with enough resolution to record small active regions and crudely detect the supergranulation network – is feasible with a spatial resolution of about a tenth of a milliarcsecond, i.e., 500 times better than the best delivered by HST in the UV. Extending the sample to dwarfs within 100 pc would require microarcsec imaging. Hence, all we know about starspots is indirectly derived by means of the signatures they leave on the optical light curves and on the spectra. A complete description of the observational tools and diagnostic techniques to study starspots is given by Berdyugina (2005).

The *Doppler imaging technique (DIT)* is very robust in deriving maps from high-resolution spectral line profiles of rapidly rotating stars. A cool spot creates a distortion in the spectral lines at a wavelength location corresponding to the spot's longitudinal location on the stellar surface. The shape of this distortion is merely a one-dimensional projection of the spot feature. As the star rotates, we see the spot feature from a different aspect and thus get a different projection. Moreover, the velocity amplitude of the spot-generated feature depends on the spot latitude. Using a series of these projections, it is possible to derive information about the size, shape, and location of the spot feature, thus obtaining two-dimensional images of inhomogeneous stellar surfaces. The method, initially developed to map chemical peculiarities on the surface of Ap stars, was lately proposed to map the surface of V 711 Tau by Vogt and Penrod (1983). The reader interested in the DIT assumptions, the used regularization methods and their pro and contra are referred to Berdyugina (2005), Thomas and Weiss (2008), and references therein. About 80 cool stars have been studied with the Doppler imaging technique up to date (Strassmeier 2009). About half of them are single stars and half are components of close binaries. For 20% of them, DIT maps have been obtained at several epochs.

The DIT requires that the rotational broadening of a line profile is significantly larger than the local line profile at a single point on the stellar surface. Hence, slowly rotating late-type stars have line profiles not large enough to allow the search of the Doppler footprints of active regions. Assuming that the active regions are stable during their transit across the disk as the star rotates, the modulation of the visibility of the active regions generates the variability of the optical flux.

Spot models of photometric light curves are used to obtain maps of the filling factor of the active regions over the surface of a star. Earlier models were based on two circular spots whose position and size were adjusted in order to fit the light curve rotational modulation (see, e.g., Rodonò et al. 1986). The same kind of approach, but using three active regions containing cool spots and bright faculae in a fixed proportion, was applied by Lanza et al. (2003, 2004) to fit the total and the spectral solar irradiance variations using the SoHo/Virgo data of the Sun seen as a star. The position of the three model active regions and the variation of their total area along a significant fraction of activity cycle 23 turned out to be in good agreement with those of the largest observed active regions and complexes of activity dominating the solar irradiance variations. However, the longitudinal distribution of the observed active regions close to the solar maximum was too much complex to be described with only three active regions, and this resulted in a poor agreement between the model and the observations. The same is true for the most active stars. For instance, the surface maps obtained from the light curve modeling of the K1 subgiant component of the RS CVn system HR 1099 can be compared with the much more detailed Doppler imaging maps revealing that the spot models based on two or a few spots are generally not suited to reconstruct the complex distribution of surface brightness inhomogeneities characterizing a real active star (Vogt et al. 1999). On the other hand, the application of regularization techniques (e.g., the maximum entropy and Tikhonov algorithms)

to a continuous distribution of spots showed a significantly better agreement between the maps based on wide-band photometric data and those obtained by Doppler imaging. In particular, the longitude distribution of the spotted area and its total variation versus time can be safely reconstructed by means of those more sophisticated light curve inversion techniques, as it was shown by Lanza et al. (2006, 2007).

The *Zeeman-Doppler imaging* (ZDI) method is similar to the DIT, but stellar magnetic fields instead of surface temperature inhomogeneities are detected. ZDI is based on the search for the rotational modulated Zeeman signatures generated by the magnetic fields in the spectral line profiles and on their polarization state. Moderate to ultrafast rotators are preferred targets; in fact, for slow rotators, the net circular polarization signal in spectral lines is negligible because contributions of regions of opposite field polarity would be mutually cancelled. Since Zeeman signatures in atomic lines due to starspots are expected to be extremely small, multiline cross-correlation algorithms, like the “least squares deconvolution” (Donati et al. 1997), are used to extract and enhance the quality of individual polarized line profiles and to increase the Zeeman signature. By applying inversion techniques, similar to those used for Doppler imaging to all four Stokes parameters, the temperature and magnetic field vector can be recovered over the stellar surface. Discussion and references on ZDI results are provided by Strassmeier (2009).

5.1.2 Starspots’ Main Parameters

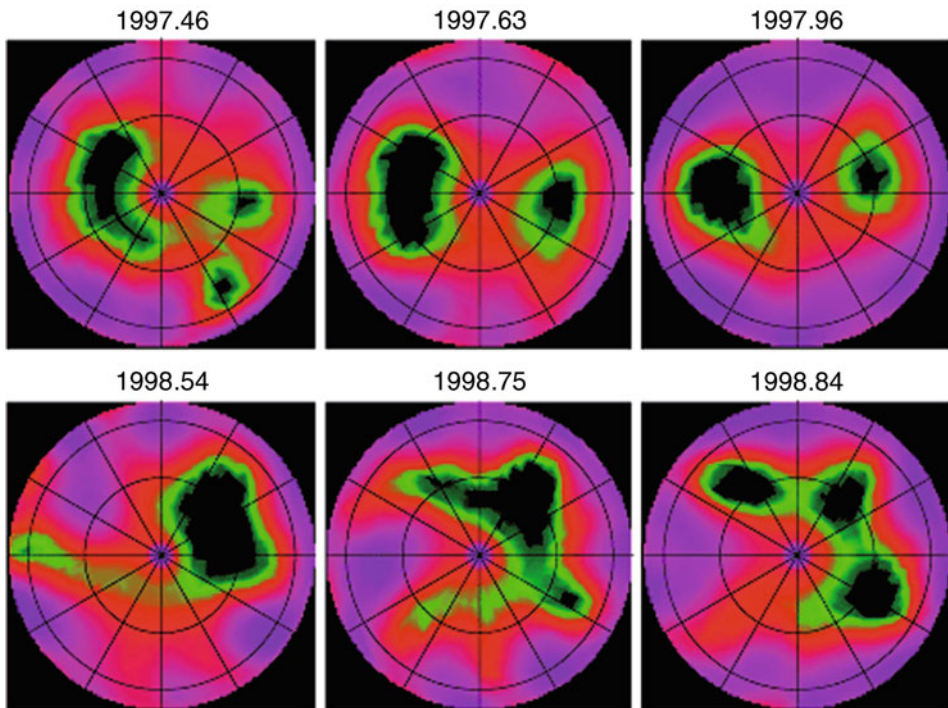
Size and Temperature. While in solar twins, starspots determine rotation modulated variations in the optical band very hard to detect, the most active stars have spotted areas covering up to 20% of the entire stellar surface or about 40% of the stellar disk. By molecular bands modeling, even higher filling factors-up to 55% of the stellar disk-have been detected, (see, e.g., the study of the very active RS CVn-type star II Peg O’Neal et al. 1996, 1998). It is worth noting that uniformly distributed spots do not produce rotational modulated signal and that depending upon the inclination axis with respect the line of sight, also, Doppler imaging may not easily reveal the absolute spot coverage. Moreover, there is a fundamental trade-off between spot area and spot temperature derived by light curve spot models. As shown by Lanza et al. (2004), even in the case light curves are available in different passbands, the spots and faculae temperature cannot be unambiguously derived except for the rare cases in which a very simple and highly stable active region configuration dominates. Also, with the Doppler imaging technique, the temperature scale remains model dependent. TiO bands at 886 nm have been used to measure starspots temperatures. For instance, Neff et al. (1995) and O’Neal et al. (1996) obtained an evaluation of the spot temperature, with an estimated error of 100–200 K, and of the surface filling factor, for a number of active stars. The method, however, is effective for stars with spot temperatures lower than 3,500 K, but it fails for stars with effective temperature higher than 5,000 K.

A very effective method in detecting low variations of the disk-averaged temperature in active stars is the one based on line-depth ratio of lines having different temperature sensitivity – e.g., the V I 625.18 nm and Fe I 625.26 nm lines. The spot temperature in this way can be determined within 5–15 K. Moreover, by combining the line-depth ratio technique with simultaneous photometry, it is possible to disentangle the spot area from the spot temperature (cf. Catalano et al. 2002). Berdyugina (2005) shows that starspot contrast with respect to the unperturbed photosphere decreases with spectral type, regardless the luminosity class.

The typical temperature difference between spots and stellar quiet photosphere goes from about 2,000 K in G0 stars to 200 K in M4 stars.

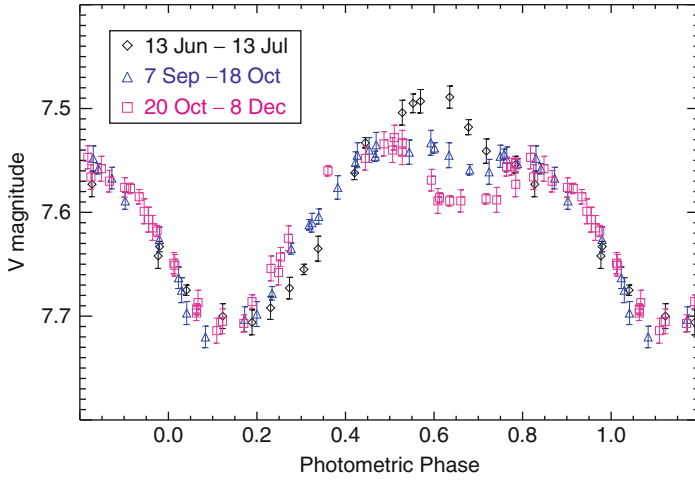
Location. Starspot latitudes can be directly recovered from DIT, that has found the presence of polar spots in about 55% of modeled targets, a clear not solar-like scenario. Also very different than the solar case is the detection of starspots migration toward the poles and the existence of high-latitude rings of toroidal field (e.g., Catala et al. 2007; Donati et al. 2003b) shown by ZDI. On the other hand, the concentration of magnetic flux in rings of opposite polarity around the pole of the active stars has been theoretically explained by models accounting for the dynamo mechanism working as in the Sun but with a bipolar magnetic field emergence rate 30 times higher than in the solar case (Schrijver and Title 2001). Another observed characteristic is the presence of permanent active longitudes, often 180° apart. These are sites where spots preferably develop and evolve. Active longitudes show in some cases migration path – e.g., AB Dor (Järvinen et al. 2005) and CoRoT2a (Lanza et al. 2009b). The level of spot activity in each active longitude has been observed to cyclically change: this is known as a “flip-flop effect”; an example is shown in Fig. 10-11 (Berdyugina et al. 1999). The period for such a change is from a few years to a decade.

Cycles and patterns. Berdyugina and Usoskin (2003) claim the presence of a flip-flop cycle on the Sun with period of 3.8 and 3.65 year in the northern and southern hemispheres, respectively. This period is about 1/3 of the 11-year sunspot cycle and agrees with the results obtained for



■ Fig. 10-11

Flip-flop effect observed on the RS CVn star II Peg (From Berdyugina et al. 1999)



■ Fig. 10-12

Light curves of the RS CVn-type star II Peg (K2 IV) acquired during a semester in 1995. Evolution of starspots' configuration is clearly apparent on time scale of weeks to months (Data extracted from the database of optical light curves of active stars at Catania Astrophysical Observatory)

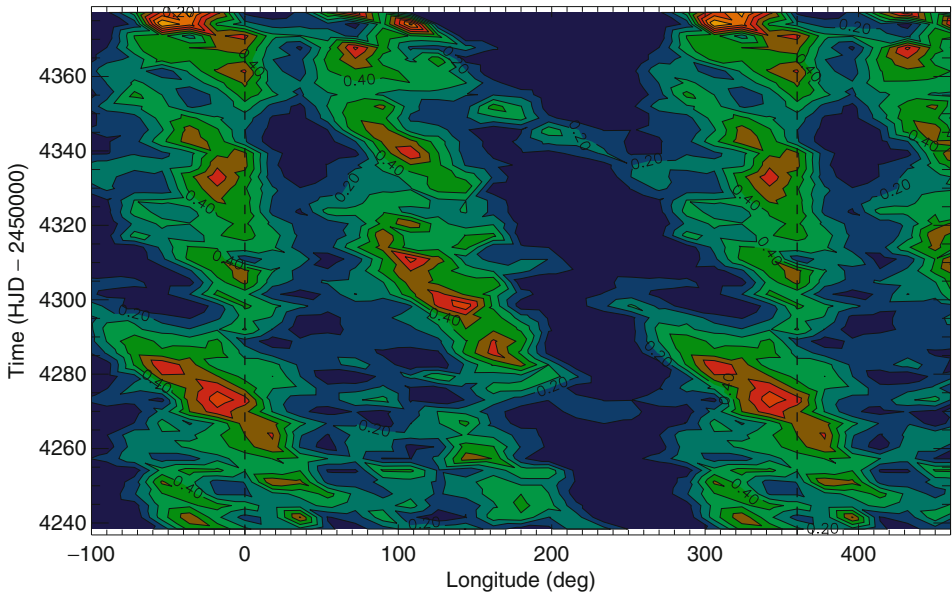
young solar analogues. The difference between flip-flop cycle lengths in the north and south is significant and produces an oscillating effect in the north–south asymmetry on a century time scale. Although active longitudes endure for a long time, active regions they consist of evolve in size (cf. the case of CoRoT-2a discussed later). Typically, starspots evolve on time scales of tens of days, as can be clearly observed in ● Fig. 10-12 where the evolution of the optical light curve of II Peg in 6 months is shown.

5.1.3 Starspots in Solar-Like Stars

The detection of starspots on low active stars like the Sun is a very recent achievement. On the Sun in fact, the largest spots at maximum of solar cycles imply variations of its optical flux of the order of 10^{-3} . From the ground in the past decades we were able to detect amplitudes of a few hundredth of magnitudes, and only recently, thanks to progresses in CCD observing techniques, we are capable to detect variations of few 10^{-3} . The very low amplitude signatures of spots on low active stars benefit from space observations and have been in the recent years detected, thanks to experiments like MOST, CoRoT, and Kepler, build to detect the weak signatures of star oscillations or planetary transits. For instance, spot modeling techniques have been applied to high-precision photometric time series obtained by the CoRoT satellite providing for the first time maps of photospheric magnetic activity in solar-like stars that have a level of activity comparable to the Sun. As an example, the case of CoRoT-2a is very illustrative; this G7 V star hosts a hot-Jupiter planet that orbits it in 1.743 d (cf. Alonso et al. 2008). The star, a good proxy for the Sun at an age of approximately 0.5 Gyr, has been continuously observed by CoRoT for 142 days with unprecedented photometric precision. The CoRoT optical light curve shows a star about 20 times more active than the Sun at its maximum. Lanza et al. (2009b) derived

the evolution of the star photospheric map (cf. [Fig. 10-13](#)) showing that this is dominated by two active longitudes initially on opposite hemispheres. In the time the star rotates with a period of about $4.52 d$, the active region separation changes by $\sim 80^\circ$ during the time span of the observations. From this variation, a relative amplitude of the surface differential rotation $\Delta\Omega/\Omega \sim 0.007$ is estimated. Moreover, the maps show that individual spots form within the active longitudes and show an angular velocity $\sim 1\%$ lower than that of the active area. This effect can be due to a progressive dynamical decoupling of the magnetic flux tubes from the deep layers once spots are formed. The overall scenario derived by the Lanza et al. (2009b) analysis has been confirmed by independent analysis by Huber et al. (2010). Active regions last about 55 days, while lifetime of spots inside them is of about 20–30 days. The total spotted area shows a cyclic oscillation with a period of 28.9 ± 4.3 day, which is close to ten times the synodic period of the planet as seen by the rotating active longitudes, suggesting star planet interaction at work, as confirmed by the analysis performed by Pagano et al. (2009), who find that the variance of the stellar light curve is modulated in phase with the planet orbital period.

Accurate photometry from space has therefore opened a new window on solar-analog variability, being these studies not possible from the ground or through mapping techniques based on high-resolution spectroscopy.



■ Fig. 10-13

Evolution of the photospheric map of the young solar analog CoRoT-2a, a G7 V star. Time is measured in HJD since 2,450,000. The contour levels are separated by $\Delta f = 0.1f_{\max}$ with light yellow indicating the maximum covering factor and dark blue the minimum. Two active longitudes are clearly detected, having slightly different migration patterns. Individual spots form within the active longitudes and show an angular velocity $\sim 1\%$ lower than that of the active area (From Lanza et al. 2009b)

5.2 Chromospheric Plages

Our knowledge of stellar magnetic cycles is mainly based on observations of starspots, i.e., sites of concentrated strong magnetic fields. However, an important fraction of the global magnetic flux is related to smaller-scale magnetic fields. These, in the solar case, are traced by plages, which are present even at cycle minimum when few or no spots are on the disk. Understanding the link between activity cycles and small-scale magnetic fields is a key issue to further progress on dynamo models, especially to improve the capability to forecast details of activity on time scales of months to years.

Location and migration patterns of active regions in the chromospheric and transition-region emissions of stars can be derived by direct images or inferred by means of indirect techniques.

5.2.1 Direct Imaging of the Chromosphere

Direct narrowband images of the chromosphere have been secured till today only for the red supergiant Betelgeuse (α Ori, M1-2Ia-Ibe), imaged by the Faint Object Camera on Hubble Space Telescope at 50 mas resolution in the mid-UV. The Betelgeuse chromosphere resulting from the *HST* images is more than twice extended with respect to the photosphere, with a single bright, unresolved area brighter than the surrounding disk with a temperature differential in excess of 200 K (e.g., Gilliland and Dupree 1996). The extended α Ori's chromosphere is apparently hard to be concealed with the view obtained by spatially resolved radio data which show a much cooler extended atmospheres (Lim et al. 1998), suggesting a copresence of hot chromospheric plasma and cooler gas. Moreover, by using UV and radio diagnostics for the plasma density, it was demonstrated that there is over an order of magnitude difference in electron density between the hot (from UV) and cool (from radio) plasma components (Harper and Brown 2006). Theoretical models attempting to account for both UV and radio data predict that the hot chromospheric plasma is confined in numerous uniformly distributed thin structures (Harper 2010).

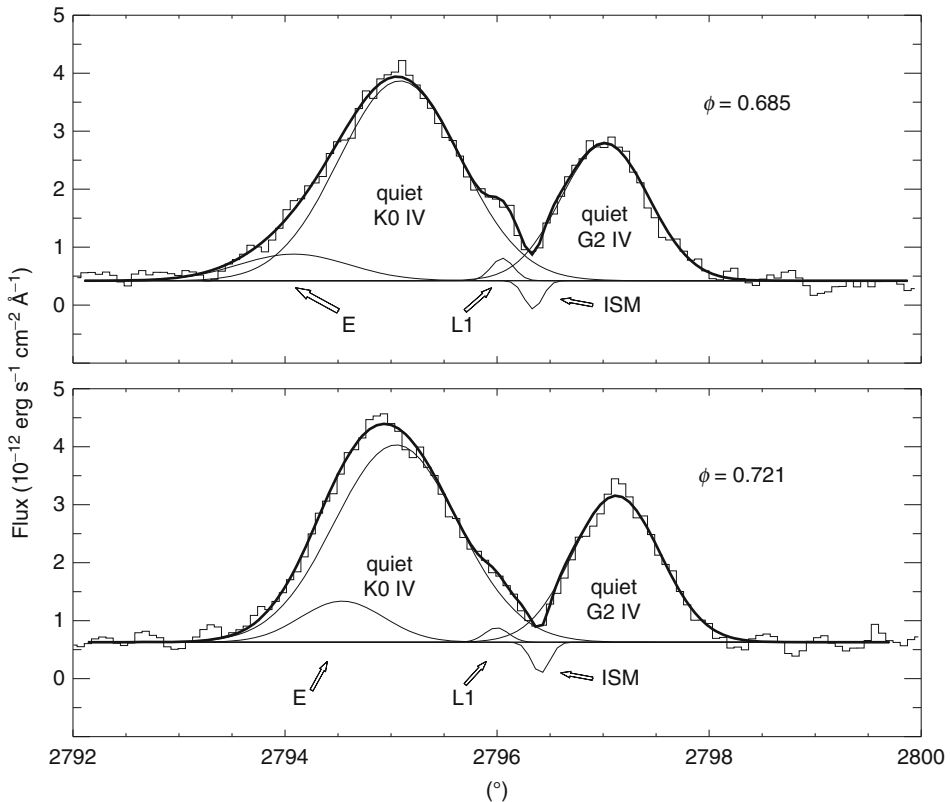
The case of Betelgeuse is unfortunately unique and will be so for the near future. In fact, obtaining a $100 \times 100 \text{ pixel}^2$ UV map sufficient to record small active regions, and crudely image the supergranulation network, requires a spatial resolution of about a tenth of a *mas* for the closest solar-like dwarf stars and at μas level to extend the sample within 100 pc. These performances can be reached by a kilometer scale interferometer like the one proposed by the *Stellar Imager* mission concept studied by NASA (Carpenter et al. 2009).

5.2.2 Spectral Imaging

An indirect technique to derive the location and migration patterns of plages in the chromospheres and transition regions is to use time series of high-resolution emission-line profiles to highlight their Doppler-shifted signatures. This method, known as “spectral imaging technique,” can be applied to rapidly rotating active stars whose rotationally broadened chromospheric (e.g., Mg II h&k 2,800 Å) or higher temperature (e.g., C IV 1,548 Å) emission lines are generally observed to be distorted and highly variable from one phase to the next. The distorted emission-line profiles are the result of several components: the uniformly distributed emission

(i.e., the “quiet” emission) due to the stellar chromosphere (or transition region), the absorption intervening in the interstellar medium (ISM) between the star and the observer, and the emission arising from plages and/or flares which are Doppler shifted according to the active region location with respect to the central meridian. These latter components are variable from one spectrum to the other because intrinsically variable and because their visibility is modulated by the stellar rotation.

A few RS CVn-type stars with rotation periods less than 3 days have been monitored in the Mg II h&k light with the *IUE* satellite to get maps of their chromosphere (e.g., Walter et al. 1987; Neff et al. 1989; Busà et al. 1999; Pagano et al. 2001). Plage signatures have been detected and followed from their appearance on the blue wing of the line profile to their crossing at the central meridian, till their disappearance when reach the red wing of the line profile (► Fig. 10-14).



■ Fig. 10-14

Mg II k line profiles of the RS CVn-type system AR Lac (G 0 IV + K2 IV), taken at two different rotational phases by *IUE*. The line profile is modeled with a multicomponent Gaussian fit. The two main Gaussians represent the emission from each stars; the absorption component is due to the interstellar medium (ISM). From the analysis of a time series of such profiles, it was possible to understand that the E component is due to an active region, which migrate across the line profile because of the stellar rotation, and the L1 component is due to emission near to the system center of mass (Pagano et al. 2001)


As an illustrative example, we report here results of the analysis of a sequence of 59 high-resolution Mg II k line profiles secured by IUE in October 1994 of AR Lac, a binary system of RS CVn-type consisting of a G2 IV primary and a K0 IV secondary stars which are spin-orbit coupled with a $\sim 1.98 d$ period:

- (a) Light curves obtained by measuring the fluxes of chromospheric and TR emission lines show eclipses that are much wider in phase than predicted by the geometrical eclipses of the stellar photospheres. This fact implies that the plasma at temperature between about 10^4 and 10^5 K extends well above the stellar photospheres.
- (b) The chromospheric emission of the G2 IV star appeared to be both compact and spatially homogeneous, while the chromosphere of the K0 IV star was spatially nonuniform with the outer hemisphere weaker than the one facing its companion. Regions of enhanced chromospheric emission corotating with the K0 IV star and localized at heights between 0.3 and $1.2 R_*$ above the photosphere have been detected. These results are consistent with the conclusions drawn by X-ray observations which suggest a compact corona surrounding the G2 IV star whereas an extended corona surrounds the K0 IV star (see Rodonò et al. 1999 and references therein). Moreover, the presence of chromospheric plasma at distances as large as $1.2 R_*$ above the K0 IV photosphere suggests that some mechanism is at work to confine the “cool” plasma in a much hotter coronal environment. The multi-temperature structure of the external atmosphere of magnetic active stars is suggested also by the observations of extremely broadened TR emission lines in some rapidly rotating giant stars; being the rotation broadened line widths up to twice that expected from the photospheric $v \sin i$, the 10^5 K plasma must be located at levels up to 10 pressure scale heights above its equilibrium altitude, raising questions as how the gas got there and how it can remain (Ayres et al. 1998).
- (c) The active regions lifetime was greater than the 4 days of coverage.
- (d) From the mean Mg II h&k surface fluxes of the active regions, the presence of heating mechanism operating in a regime close to saturation can be derived.
- (e) Absorbing material in the lines of sight toward both the K0 IV star and the G2 IV star has been detected both as sporadic episodes and in permanent corotating structures. Evidences for the existence of absorbing structures embedded in the coronal plasma were previously found by Neff et al. (1989) – who have observed the disappearing of the G2 IV chromosphere, possibly behind a large extended absorbing structure in the Mg II h&k lines – and by Walter (1996a) from an EUV light curve of AR Lac observed by the EUVE DS photometer (70–190 Å). Extended absorbing structures were also observed in H α (Collier-Cameron et al. 1990) and UV emission lines (Walter 1996b) of the active single early K dwarf AB Dor and in the UV emission lines of the eclipsing K2 V-dA binary V471 Tau by Walter and Kim (2001).
- (f) A bright feature in Mg II k light has been detected to be permanently present close to the L1 Lagrangian point. This has been interpreted as a possible signature of some hot spot due to mass exchange between the two stars. Evidence for a similar intersystem structure was found also for the RS CVn-type binary system HR 1099 by Busà et al. (1999) and, independently, by Richards and Rosolowsky (1998).

The best targets for emission-line spectral imaging are rapidly rotating stars which, by their nature, are highly active and flare frequently. Time series of UV spectra with spectral resolution $\geq 30,000$ ($< 10 \text{ km s}^{-1}$) with good enough cadence to avoid smearing the Doppler information, and long enough coverage to distinguish between persistent surface features and transient flare activity, are required to detect discrete emission components migrating through the line profile of chromospheric and transition-region lines. While it was straightforward for an

observatory/satellite like IUE performing an observing program requiring the monitoring of a same object for typically 2–6 days, as it did for a number of sources, it resulted not trivial with the Hubble Space Telescope, because of its Earth occulted orbit and the difficulties to accomplish monitoring projects given the strong pressure to get observing time. The single attempt by HST to map the outer atmosphere of a cool star, namely, HD 155555 (Dempsey et al. 2001), covered one stellar rotation and was only partially successful because flaring and non-phase-dependent variability could not be separated from rotational modulation of unchanging active regions. For this reason, up to date, only few stars have been systematically observed in the UV range for magnetic activity cycles, and the observations have been obtained only for very active stars.

5.3 Chromospheric Versus Photospheric Activity

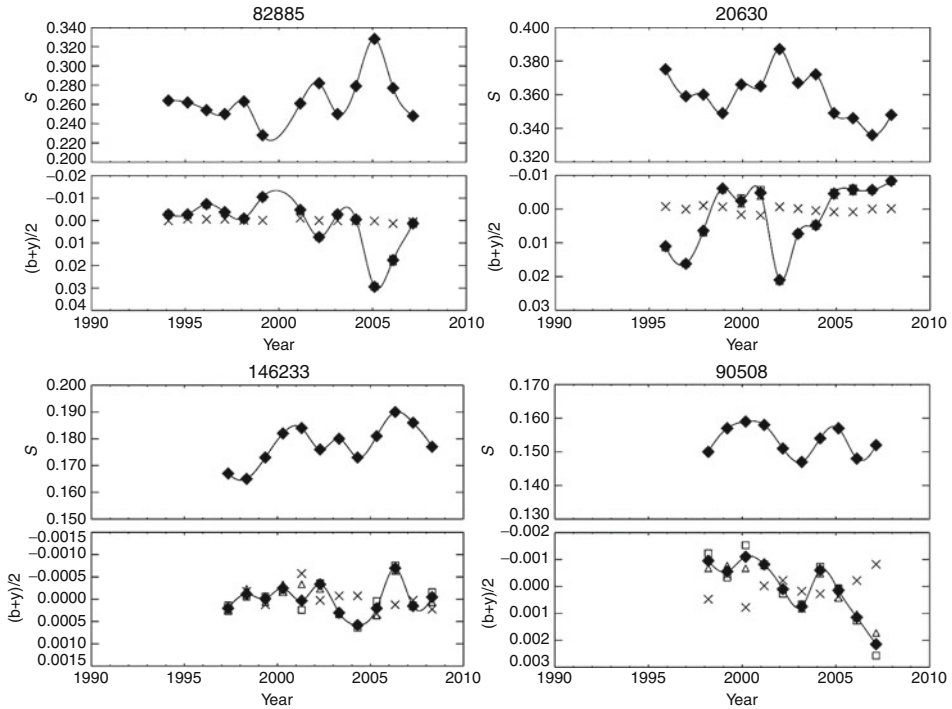
Lockwood et al. (2007) and Hall et al. (2009) used Ca II H&K data, from Mt. Wilson Observatory, and Strömgren *b*, *y* photometry, from Lowell and Fairborn Observatories, collected over a time lapse of over two decades to study the relation between photospheric and chromospheric variability for a wide sample of solar-like stars. The main conclusion of these studies can be summarized as follow: on a year-to-year time scale, the young active stars become fainter when their Ca II emission increases; on the contrary, old-less active stars, including the Sun, become brighter when their Ca II emission increases. In  Fig. 10-15, HD 82885 and HD 20630 are shown as example of young active solar-like stars; their chromospheric and photospheric light curves are anticorrelated. Instead, HD 146233 and HD 90508 are shown as example of old low-active solar-like stars; their chromospheric and photospheric light curves are correlated. As a matter of fact, young active solar-like stars are spot-dominated while older less active stars, like the Sun, are faculae dominated.

Hall et al. (2009) highlight that the Sun is not peculiar among other solar-like stars; its median photospheric variability is only slightly above the one of the solar analogs. For instance, 18 Sco (G2 V), one of the closest solar twins, with a photometric activity cycle of 7 year, exhibits a chromospheric variability very similar to the solar one, with $\sigma(\log R_{\text{HK}}) = 0.043$ from 1994 to 2008, that is comparable to 0.040 for the Sun over the same period.

The existence of two classes of objects, one spot dominated, the other faculae dominated, has been claimed also for magnetically active close-binary stars. Messina (2008) has analyzed brightness and color variations on a long time scale (tens of years) in the U, B, and V bands of a sample of these objects, finding that some of them become redder when they are fainter, as it is expected from the presence of active regions mainly consisting of cool spots. Others instead become bluer when they become fainter, i.e., when spots reach the maximum coverage. In such a case, hot photospheric faculae dominate the color variation. The reasons why two classes are present among the magnetically active close binaries is still unclear.

6 Transition-Region Lines as Probe for the Plasma Dynamics

From the analysis of UV emission lines, we can get constraints on the dynamo and atmospheric heating models. For instance, the displacement of the spectral lines with respect to the photospheric rest frame, i.e., their possible blue- or redshifts, as well as the shape of the line wings, i.e., their broadening and eventual distortion, carries information on the plasma dynamics. Understanding the plasma dynamic is key into the comprehension of quantitative details on the



■ Fig. 10-15

Ca II H&K (*top panels*) and photometric (*bottom panels*) series for spot-dominated young active solar-like stars (HD 82885 (G8 IV–V), HD 20630 (G5 V)) and faculae-dominated old and less active stars (HD 146233 (G2 V), HD 90508 (G0 V)). Diamonds show the seasonal means, while the differential series of the comparison stars are shown by “x” symbols (From Hall et al. 2009)

energy transport and dissipation processes in stellar atmospheres. Actually, plasma dynamic is a by-product of the heating mechanisms, whose precise nature is one of the major unsolved problems in solar and stellar physics. In the outer convection zone, gas kinetic energy is transformed by turbulence into sound waves and electrodynamic perturbations. The latter are thought to provide the bulk of coronal heating, because sound waves hardly propagate through the steep temperature gradient of the transition region, even though the shock dissipation of pure acoustic waves is probably an important, although not main, heating source for the chromosphere and the lower transition region (c.f., Judge et al. 2003, and references therein).

6.1 UV Line Shift Versus Temperature of Line Formation

The measured Doppler centroid of optically thin high-temperature lines reflects the emission-measure-weighted net flow speed in a highly dynamic atmosphere. Hence, the empirical dependence of their velocity shifts with temperature provides an important constraint on stellar transition region dynamics.

Observations of the Sun show that transition region emission lines are, on average, redshifted, and that redshifts increase with increasing formation temperature up to about 10^5 K. At higher temperatures, the centroid velocities decrease, crossing over to blueshifts at $T \sim 10^6$ K, reaching about -5 km s^{-1} (Peter and Judge 1999). The slope and the peak of the velocity shift versus line formation temperature are different for active regions or quiet Sun, as shown in **Fig. 10-16**. Measurements of transition region line Doppler shifts have been obtained for

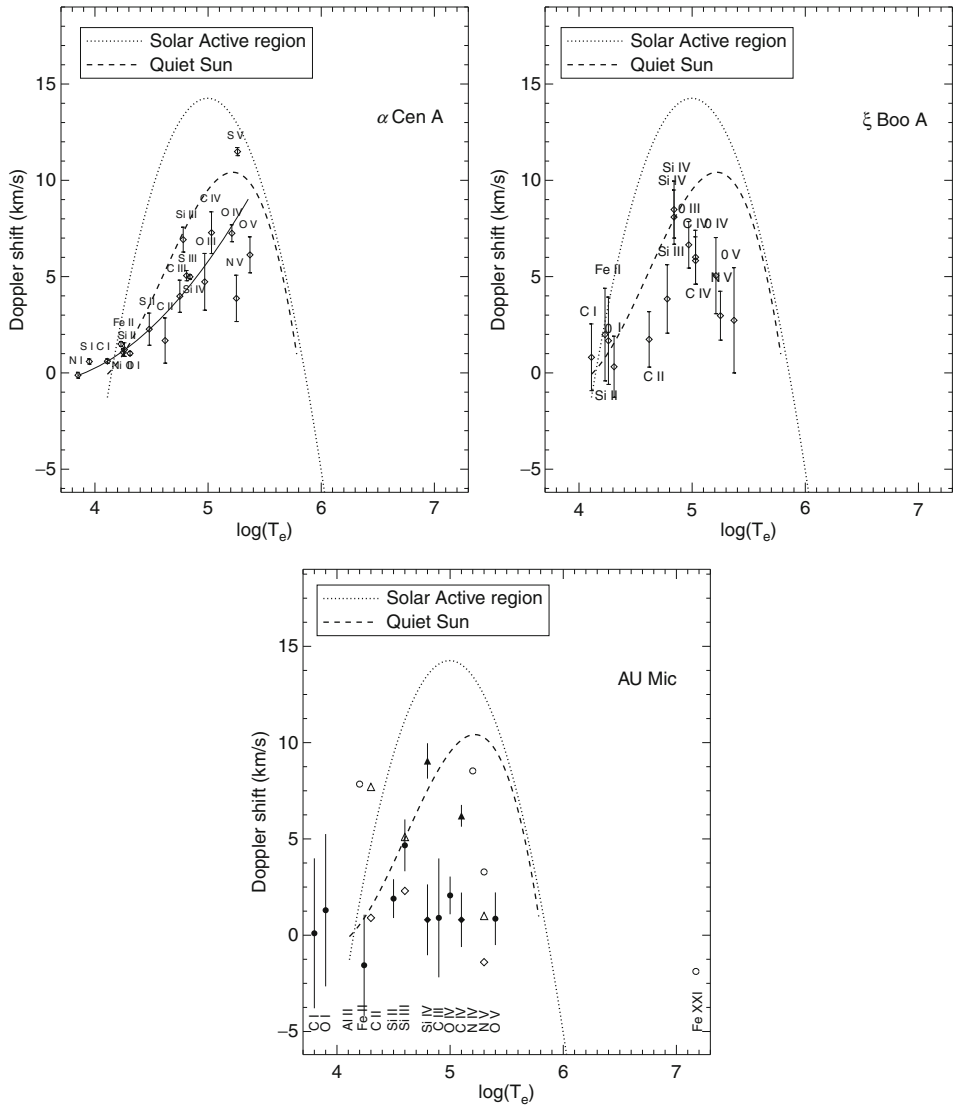




Fig. 10-16

Doppler shifts of chromospheric and transition-region lines relative to the photospheric radial velocity as a function of temperature of line formation of α Cen A, ξ Boo A, and AU Mic (from Pagano et al. 2000, 2004, 2006). The dotted and dashed lines are fits to the Doppler shifts measured by Teriaca et al. (1999) for a solar active region and the quiet Sun, respectively

a number of other stars with *IUE*, *HST*, and *FUSE*. These measurements are full disk averages, contrary to what is typically available for the Sun. When plotted against line formation temperature, a solar-like behavior is apparent for late-type stars, with the maximum redshift occurring at temperature that can be the same as for the Sun (e.g., α Cen A, see Pagano et al. 2004) or higher (e.g., Procyon, see Wood et al. 1996) or lower (ξ Boo A, see Pagano et al. 2006) as shown in  Fig. 10-16. There are, however, objects which behave differently. For instance, the very active dM1e star AU Mic ($\log L_x/L_{\text{bol}} \sim -3$) hardly shows any redshifts and certainly shows no conspicuous trend of line shift versus line formation temperature (Pagano et al. 2000; Redfield et al. 2002). This is not expected because AU Mic has an average magnetic flux density much larger than the solar average magnetic flux density,³ and redshifts of transition region lines in solar active regions, where the magnetic field is strong, are twice as large as in quiet regions (Achour et al. 1995). Also, transition region models computed for the Sun predict larger redshifts in regions permeated by strong magnetic fields than in quiet regions (Reale et al. 1996). Several other models have been proposed to account for the redshifts observed for the Sun. For instance, Peter et al. (2006) propose a 3D numerical model able to predict redshifts in transition region lines; in this model, coronal heating is produced by Joule dissipation of the currents that origin in the interaction of magnetic fields with the photospheric motion. Spadaro et al. (2006) have successfully reproduced the temperature dependence of the persistent redshifts observed on the Sun, together with the observed emission measure distribution over the entire range $\log T [K] = 4.7 - 6.1$, by using a hydrodynamic model accounting for small cool magnetic loops (length scale of order 1 Mm), transient heating (timescales of order 20 s every 100 s), and energy deposition spatially localized near the chromospheric foot points (nanoflares). Further modeling developments are proposed by Hansteen et al. (2010) and Zacharias et al. (2011). While these models are predictive for the Sun and other solar like-stars, it is clear that a general model, including the dependence for the magnetic field strength, the magnetic filling factor, the nonradiative heating rate, and the gravity, is required to shed light on the origin of stellar transition region line redshifts.

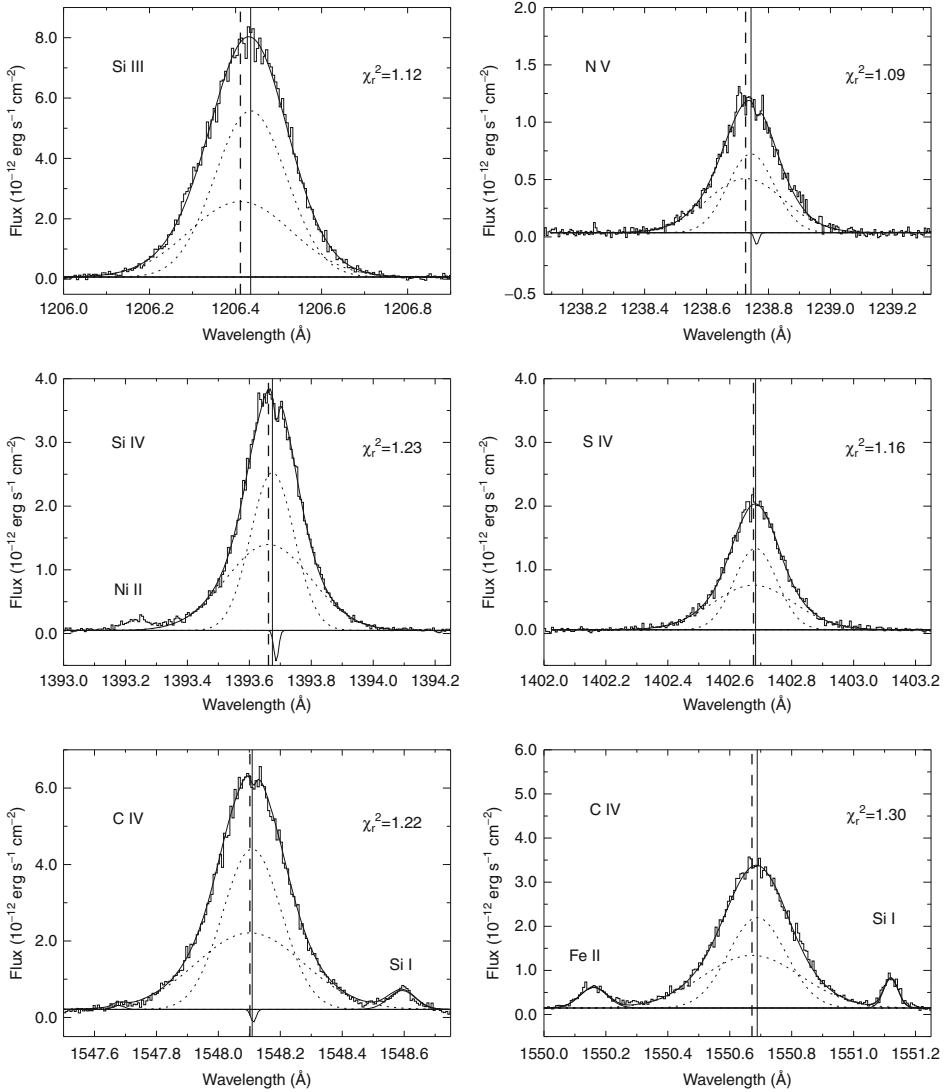
6.2 Broad Wings in TR Lines

Other important clues on plasma dynamics come from the broad wings observed in TRs and coronal emission lines (e.g., Linsky and Wood 1994; Wood et al. 1997).  Figure 10-17 shows the strongest transition region emission lines of α Cen A. Broad wings are present in the Si III λ 1,206 Å, N V λ 1,238 Å, Si IV λ 1,393 and 1,402 Å, and C IV λ 1,548 and 1,502 Å line profiles. These lines are reproduced by one narrow Gaussian component that fits the line core and one broad Gaussian component that fits the broad wings. This bimodal structure of the transition region lines is typically observed for main-sequence-type stars (e.g., AU Mic, Procyon, and α Cen A & B), giants (e.g., 31 Com, β Cet, β Dra, β Gem), rapid rotator PMS stars (e.g., AB Dor), and several RS CVn-type stars (e.g., Capella and HR 1099) (cf., Linsky and Wood 1994; Linsky et al. 1995; Pagano et al. 2000).

The broad and narrow Gaussian components have generally comparable intensity and their flux ratio is independent of the activity level of the star.

Wood et al. (1997) showed that the narrow components can be produced by turbulent wave dissipation or Alfvén wave-heating mechanisms, while the broad components – that resemble

³ $Bf = 2,300$ G and 15 G for AU Mic and the Sun, respectively, where f is the fractional area coverage of the magnetic field B



■ Fig. 10-17

Transition region lines of α Cen A (from Pagano et al. 2004), presenting a bimodal structure made by a narrow core and broad wings. This structure is typically observed for magnetic active stars. The broad wings have been interpreted as signature of highly dynamic explosive events in the lower atmosphere, but they can be due to the structure of and flows in the chromospheric network

the explosive events on the Sun and whose strength correlate with activity indicators like the X-ray surface flux – can be interpreted as a signature of “microflare” heating. Using FUSE data, Redfield et al. (2002) showed that there is a trend of decreasing relative importance of the broad components with increasing line formation temperature. Hence, assuming that the broad components are signature for “microflare” heating as inferred by Wood et al. (1997), the

spectroscopic impact of “microflare” heating might be more important in the lower TR than at $\log T \text{ (K)} = 5.5$, where the O VI lines are formed.

However, the comparative analysis of a full-Sun VUV spectrum, based on a raster scan of the whole solar disk using SUMER/SOHO, with the HST/STIS spectrum of α Cen A, led Peter (2006) to conclude that the broad components of disk-integrated stellar spectra do not contain much information on the coronal heating mechanism but rather provide valuable input on the structure of the magnetic network.

On the other hand, the very young Sun-like star EK Dra (G1.5 V), observed with HST/COS by Ayres and France (2010), shows broad Si IV 1,400 Å lines having both the narrow and the broad components highly redshifted. Such a scenario is in contrast with the association of the broad component with explosive events but is instead suggestive of a “coronal rain” made of the $T \sim 6 \times 10^4$ K gas continually accreting onto the lower atmosphere.

Bimodal Gaussian profiles have been observed also for luminous cool stars thanks to FUSE observations by Dupree et al. (2005), who argue that the broad wings are due to an opacity effect in the cool dense wind of these stars, even though Peter (2006) suggests that this interpretation is not unique.

Ayres et al. (1998) have found very broad FUV emission lines in moderate-mass ($\sim 2\text{--}3 M_{\odot}$) giants staying in the Hertzsprung gap (spectral types early F to mid-G) and in the post-helium flash “clump” ($\sim G8\text{--}K0$), suggesting that the transition region gas ($T = 10^5$ K) extends out to one R_* above the photosphere, i.e., about 50 scale heights. Fast rotator gap giants show supra-rotational broadened FUV line profiles which are generally asymmetric, with profile distortions possible signatures of Doppler-shifted surface structures. Instead, slowly rotating clump giants show transition region lines narrower and more symmetric.

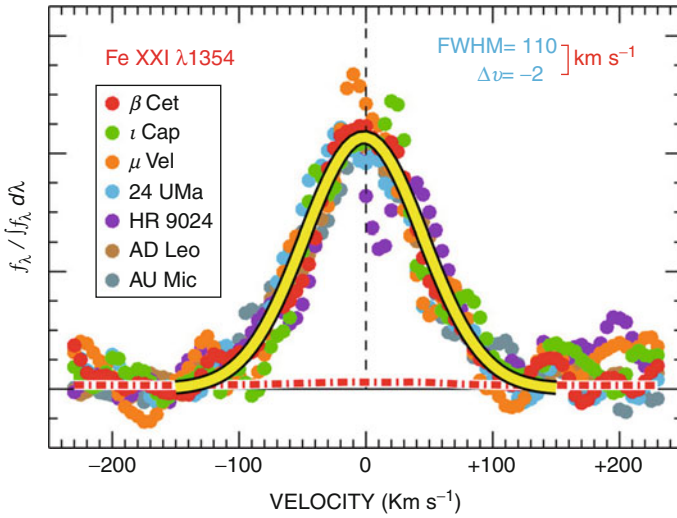
For some rapidly rotating stars, broad wings have been attributed to prominence-like, extended structures of cool ($10^4\text{--}10^5$ K) gas at 2–3 stellar radii, similar to the “slingshot prominences” seen in H α for the young, rapid rotator AB Dor (Collier-Cameron et al. 1999; Vilhu et al. 1998), and the extended loops inferred by Mg II h&k Doppler imaging for AR Lac (Pagano et al. 2001).

Extending the stellar observations to higher temperatures, as in the solar work, is not straightforward as discussed in the ► Sect. 6.3.

6.3 UV Coronal Forbidden Lines

The UV spectrum is extremely interesting also for the study of the coronal dynamics. Observing the plasma dynamics to higher coronal temperatures is relatively easy on the Sun, because many suitable strong coronal permitted lines (e.g., Mg x λ 61.025 nm) fall in the Lyman continuum region immediately below 91.2 nm, where high-resolution far-UV spectroscopy still is practical. Unfortunately, these key features are not accessible even in the nearest stars owing to interstellar extinction. High-resolution spectroscopy of permitted coronal X-ray lines, say in the important iron-L shell band at 1 keV, is not feasible at present, because contemporary high-energy missions like Chandra and XMM-Newton fall short in spectral resolution by an order of magnitude, and we must account on future missions to push that limit.

Coronal spectroscopy of active stars at high resolution has been possible up to date only thanks to the observations of coronal forbidden lines in the UV and FUV spectrum, e.g., Fe XVIII λ 974 Å, Fe XIX λ 1,118 Å, Fe XII λ 1,242 and 1,349 Å, and Fe XXI λ 1,354 Å (cf., Pagano et al. 2000; Ayres et al. 2003; Redfield et al. 2003). As shown in ► Fig. 10-18 for the Fe XXI λ



■ Fig. 10-18

Superposition of scaled Fe XXI λ 1354 profiles of seven stars, mostly G1–K0 giants but including two M dwarfs. The individual line shapes collectively define a “universal” profile that can be matched by a Gaussian with FWHM 110 km s^{-1} and an essentially negligible 2 km s^{-1} blueshift with respect to the predicted average photospheric velocities of the stars and the adopted wavelength of the Fe XXI λ 1354 (From Ayres et al. 2003)

$1,354 \text{ \AA}$, centroids of coronal lines of a number of late-type giants and M dwarfs fall at rest with respect to the photosphere; hence, the high-temperature plasma is not participating in any large systematic mass motions, but it is likely confined in stable magnetic structures. Moreover, the measured widths of the coronal forbidden lines are generally close to their thermal values (FWHM $40\text{--}90 \text{ km s}^{-1}$ at $T \sim 10^{6.2}\text{--}10^{7.0} \text{ K}$), except for some of the fastest rotating stars in the observed sample (AB Dor, 31 Com, and the G1 star of Capella) all showing broadening exceeding the convolution between the thermal and the rotational widths, consistently with excess turbulence or with emitting regions at $\sim 0.4\text{--}1.3$ stellar radii from the photosphere.

7 Stellar Coronae

Main diagnostic of the stellar corona is the X-ray emission due to the thermal plasma at temperature from 10^6 to 10^8 K and the radio emission due to the electrons moving in the magnetic fields of the loop-like structures.

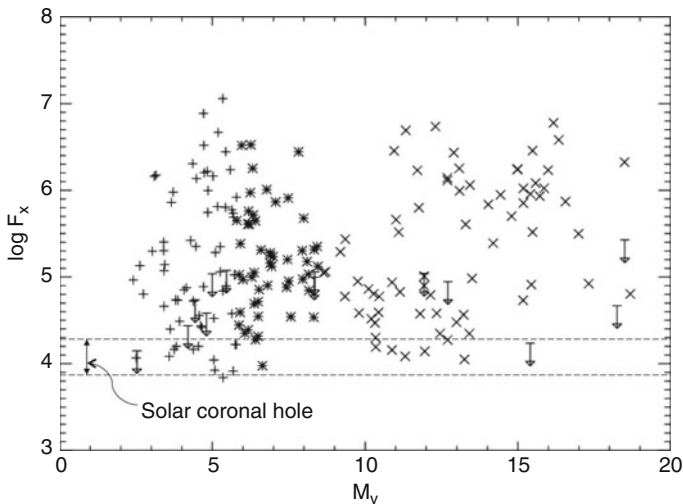
7.1 X-Ray Emission from Stellar Coronae

The first identification of coronal emission from other stars besides the Sun is that reported by Catura et al. (1975) who observed the binary system Capella with a rocket experiment in 1974. Then, Capella and Sirius have been revealed as soft X-ray sources by the ANS satellite (Mewe et al. 1975), and some RS CVn-type binaries have been detected in soft X-rays with

HEAO-1 (Walter et al. 1978; Walter et al. 1980). It was the Einstein satellite, the first X-ray observatory with imaging capability, that allowed Vaiana et al. (1981) to show that X-ray emission is present throughout the HR diagram with the exception of B-A stars and red giants. Since then, a number of X-ray observatories have contributed to the understanding of the physics of stellar coronae, including the high-sensitivity and high-spatial-resolution Chandra and XMM-Newton satellites that have been at work since 1999. The X-ray emission in stellar coronae is mainly due to bound-bound emission (lines) dominating below 10^7 K, free-free emission (bremsstrahlung) dominating above 10^7 K, free-bound emission and bound-free absorption which determine the continuum, and emission decay from a metastable state in helium-like ions. Excellent reviews on stellar coronae have been done by Favata and Micela (2003) and Güdel (2004); even not recent, these reviews cover most of the last discoveries in the field being the bulk of novelties carried by Chandra and XMM-Newton observatories very well exploited at the time of their redaction.

The first comprehensive survey of all F and G stars within 13 pc from the Sun was made by ROSAT, whose main result is that all the solar-type stars have a corona. In the ROSAT band (0.1–2.4 keV), F and G dwarfs have luminosity in the range $\log L_X \sim 26.5\text{--}29.5$ (erg s^{-1}), and their X-ray luminosity is correlated to the brightness temperature of the plasma, i.e., the star is more luminous when its coronal plasma is hotter. Moreover, the minimum flux of F to M dwarfs ($\sim 10^4 \text{ erg s}^{-1} \text{ cm}^{-2}$) corresponds to the emission of solar coronal holes (cf. [Fig. 10-19](#)).

M dwarfs have coronal emission up to a spectral type of M8 (Randich 2000 and references therein), well beyond the border with the fully convective stars. Actually, late M dwarfs are as efficient coronal emitters as other cool stars in terms of L_X/L_{bol} . Also, young ($10^7\text{--}10^8$ year) brown dwarfs have coronal emission. Preibisch and Zinnecker (2002) with Chandra/ACIS found brown dwarfs with $L_X/L_{\text{bol}} \sim 10^{-4} - 10^{-3}$ and $T = 1\text{--}2$ keV, a situation similar to that



■ Fig. 10-19

Mean X-ray surface flux versus absolute magnitude M_V for nearby main-sequence stars observed by ROSAT; plusses denote F/G-type stars, asterisks K-type stars, and crosses M-type stars (Adapted from Schmitt and Liefke 2004)

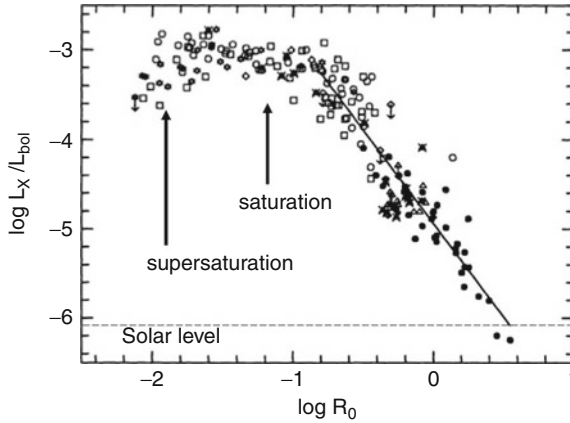
of low-mass stars, the brightest emitting in a saturation regime and the others at lower levels. Moreover, the young brown dwarfs are more active with the emission close to saturation, while the old seem to have X-ray emission mainly as flares. For instance, the sensitive X-ray data from the Chandra Orion Ultradeep Project (COUP) allowed Preibisch et al. (2005) to study 10-day-long X-ray light curves, showing that the X-ray properties (spectra, fractional X-ray luminosities, flare rates) of the young brown dwarfs are similar to those of the low-mass stars in the Orion Nebula Cluster (ONC), and thus there is no evidence for changes in the magnetic activity around the stellar/substellar boundary, which lies at M6 for ONC sources.

The active binaries (RS CVn, BY Dra, etc.) show X-ray emission with $29.0 < \log L_X < 31.5$ (erg s^{-1}). The low-resolution spectra indicate the presence of at least two components of the plasma temperature at $T \cong 3$ MK and $T \geq 10$ MK. Their coronal luminosity does not depend on the speed of rotation; they are under saturation with $L_X/L_{\text{bol}} \sim 10^{-3}$ (cf. [Sect. 7.1.1](#)).

T Tauri stars have X-ray luminosity typically between 10 and 10^4 times the solar X-ray luminosity, with spectra harder than those of dwarf stars and with a higher frequency of flares. A high fraction of PMS stars is in the saturation regime, with $L_X/L_{\text{bol}} \sim 10^{-3}$ (cf. [Sect. 7.1.1](#)). Typically, Classical T Tauri stars (CTTS) are subluminal with respect to the weak-line T Tauri stars (WTTS) (Flaccomio et al. 2003; Preibisch et al. 2005). The analysis of PMS stars surveyed with COUP (Getman et al. 2008a, b) has shown that X-ray coronal extents in fast-rotating disk-free PMS stars (WTTS) can significantly exceed the Keplerian corotation radius, whereas X-ray loop sizes in CTTS, i.e., disk and accreting systems, do not exceed the corotation radius. This is consistent with a picture of star-disk magnetic interaction in which the inner disk truncates and confines the PMS stellar magnetosphere. The current understanding is that in CTTS, X-rays are produced by two plasma components: (i) a hot low-density plasma, with frequent flaring activity, and (ii) a high-density lower temperature plasma. The former is coronal plasma related to the stellar magnetic activity. The latter component, never observed in non-accreting stars, could be plasma heated by the shock formed by the accretion process. Recently, high-resolution X-ray spectra of a few CTTS enabled measurement of individual emission lines sensitive to plasma density (i.e., He-like triplets). The majority of CTTS for which the O VII triplet lines were detected showed cool plasma with high density, $n_e > 10^{11} \text{ cm}^{-3}$ (e.g., Argiroffi et al. 2007; Robrade and Schmitt 2006). In contrast, the cool quiescent plasma of active stellar coronae is always dominated by low densities ($n_e < 10^{10} \text{ cm}^{-3}$, e.g., Testa et al. 2004; Ness et al. 2002b). This suggests that the high-density cool plasma in CTTS is not a coronal one but instead heated in accretion shocks. One complication is that mass accretion rates derived from assuming a very high efficiency of conversion of accretion energy into X-rays tend to be about one order of magnitude lower than rates derived using other methods (e.g., Günther et al. 2007; Schmitt et al. 2005).

7.1.1 Saturation and Supersaturation

[Figure 10-20](#) shows the ratio L_X/L_{bol} versus the Rossby number $R_0 = P_{\text{rot}}/\tau_c$ (in a log–log scale). For intermediate and slow rotators, $L_X/L_{\text{bol}} \propto R_0^2$ (cf., Randich 2000). For fast rotators, L_X becomes a unique function of L_{bol} , with $L_X/L_{\text{bol}} \sim 10^{-3}$ regardless of the rotation period (e.g., Pallavicini et al. 1990). Saturation occurs at Rossby numbers of about 0.1 in stars of spectral types G, K, and M. This relationship implies that the different spectral types have a different level of maximum brightness under saturation. In particular, low-mass stars in saturation regime have lower L_X than higher mass stars; thus they reach maximum levels at greater rotation period



■ Fig. 10-20

X-ray activity level, given as fraction of X-ray to the bolometric luminosity, as function of the Rossby number P_{rot}/τ_c (Adapted from [Randich 2000](#))

(i.e., lower rotation speed). [Pizzolato et al. \(2003\)](#) show that the period for which the activity is at saturation regimes ranges from 1 day for early dG stars up to ~ 4 days for early dM stars.

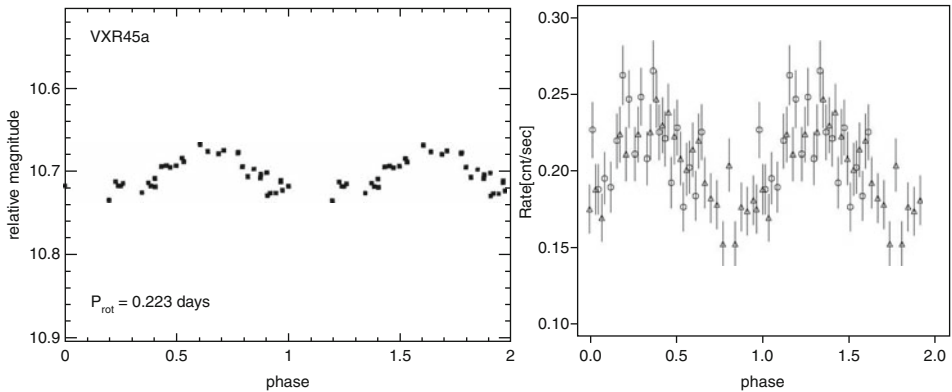
Coronal activity turns down again at rotation rates ~ 5 times faster than required for saturation; this is the so-called “supersaturation” regime. Examples of supersaturated G and K stars, with $L_X/L_{\text{bol}} \sim 10^{-3.5}$, have been found in young open clusters and among the fast-rotating components of contact W UMa binaries. Among other hypotheses, supersaturation has been explained as the result of an overall decrease of dynamo efficiency at very high rotation (e.g., [Świątek et al. 2001](#)) or as the effect of a reorganization of the coronal magnetic field (e.g., [Solanki et al. 1997](#)) or centrifugal stripping of the corona ([Jardine 2004](#)).

[Jeffries et al. \(2011\)](#) find that supersaturation is not displayed by M dwarfs. [Jeffries et al. \(2011\)](#) suggest that the rotation period instead of the Rossby number is a more accurate predictor of the onset of supersaturation, typically occurring for $P_{\text{rot}} < 0.3 d$. In the hypothesis that supersaturation is associated with topological changes in the coronal magnetic field as a consequence of the fast rotation rate, as it is for the “centrifugal stripping” model by [Jardine \(2004\)](#), then M dwarfs should supersaturate at shorter periods than K dwarfs, by factors of up to ~ 2 . Determining the X-ray emission from a small sample of rapidly rotating ($P < 0.25 d$) M dwarfs would resolve this issue.

7.1.2 Variability of Coronal Emission and Coronal Cycles

Coronal emission can vary both on short time scale – e.g., rotational modulation and flares – and on long time scale according to activity cycles. Stellar rotation periods are generally longer than typical duration of X-ray observations; for this reason, rotational modulation of X-ray emission has been detected only sporadically.

Flares will be discussed in [Sect. 10](#). As for rotational modulation, unambiguous detections have been reported by [Güdel et al. \(1995\)](#) for the young solar analogue EK Dra, by [Marino et al. \(2003\)](#) for VXR45, a G9 V star in IC 2391 (cf. [Fig. 10-21](#)), by [Hussain et al. \(2005\)](#) for the

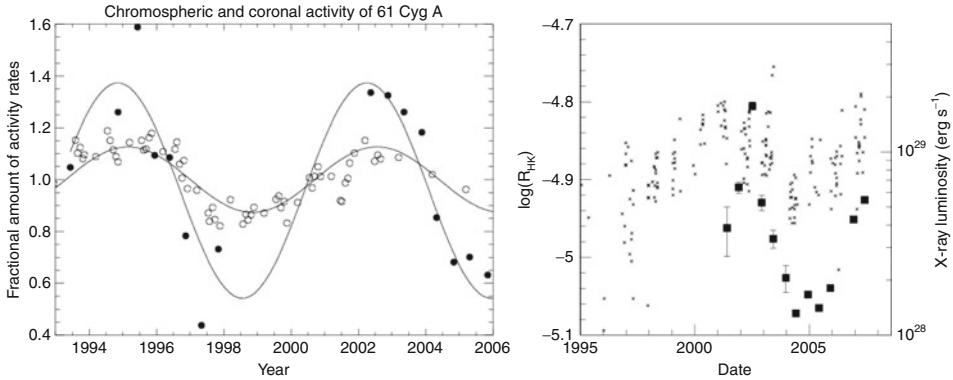


■ Fig. 10-21

Left: Optical light curve of VXR45a obtained by Patten and Simon (1996) folded with a period of 0.223 days. **Right:** X-ray light curve of VXR45a acquired with XMM-Newton in the 0.3–7.8 keV band by Marino et al. (2003), folded with the same rotation period used for the optical light curve. Circles and triangles have used for data acquired during the first and second covered rotations, respectively

K0 V AB Dor, and by Stassun et al. (2006) for a number of young stars in the ONC. AB Dor as well as many of the stars in ONC showing rotational modulation are fast rotators with an X-ray emission in the saturation regime. VXR45 is even in supersaturation regime. The presence of X-ray rotational modulation in these stars indicates a substantial inhomogeneity in the spatial distribution of active regions. Hence, saturation (and supersaturation) are not related, at least for these objects, to the filling amount of the stellar surface with active regions.

X-ray stellar cycles are instead elusive. The X-ray luminosity of the Sun follows a cycle with the same periodicity as the chromospheric cycle but with a much larger amplitude up to a factor of 100 in the Yohkoh 0.73–2.5 keV band (Acton 1996). For other stars, cycles have been detected only for α Cen B (Ayres 2009), 61 Cyg (Hempelmann et al. 2006), and HD 81809 (Favata et al. 2008), all solar-like stars with moderate activity. HD 81809 is a close visual binary of slow rotating solar-like stars (G2 V+G9 V) with a well-defined chromospheric cycle (Baliunas et al. 1995). XMM-Newton observations have clearly shown the occurrence of a coronal cycle on this star that has been modeled by Favata et al. (2008) with a simple extension of the solar case in terms of varying coverage of solar-like active regions, having large filling factor than in the solar case (cf. ► Fig. 10-22-right). The binary stars in 61 Cyg – K5 V (A)+ K7 V (B) – observed by XMM-Newton in the X-rays (0.2–2.0 keV) show both short time scales and long-term variations. In particular, 61 Cyg A has a clear coronal cycle phased with the chromospheric activity cycle (cf. ► Fig. 10-22-left). The changes of coronal properties during the cycle resemble the solar behavior. The coronal activity of 61 Cyg B is more irregular but also follows the chromospheric activity (Robrade et al. 2007). The close-by α Cen triple system, made by a G2 V (A), a K1 V (B), and the M5 V Proxima Centauri (C), has been monitored by all available X-ray observatories since the 1980. For α Cen B, Ayres (2009) confirmed the presence of a coronal cycle with period of 9 years (as the chromospheric cycle) and a depth of a factor ~ 6 , comparable to the depth of the solar coronal cycle (~ 5), while for α Cen A, he did not find any significantly variation in the 0.2–2.0 keV “ROSAT” band, despite its dramatic fainting in the high-energy



■ Fig. 10-22

Left: The coronal and chromospheric cycles of 61 Cyg A observed by XMM-Newton (0.2–2.0 keV band) and Mt. Wilson and Lowell observatory projects, respectively (Hempelmann et al. 2006 for details). **Right:** The coronal and chromospheric cycles of HD 81809 measured, thanks to XMM-Newton monitoring in the 0.2–2.5 keV band (*squares, right-hand scale*) and the monitoring of the chromospheric activity index R_{HK} from Mt. Wilson project (*crosses, left-hand scale*) (See Favata et al. 2008 for details)

tail of the coronal spectrum observed by XMM-Newton since 2003 (Robrade et al. 2005). The α Cen A behavior in the X-ray reminds the Sun's magnetic activity cycle, where the 2–3 MK active regions of sunspot maximum give way to the spatially pervasive, but cycle-independent, 1 MK quiet corona at minimum. The average coronal luminosity of α Cen B is a factor of 2 higher than that of the Sun, while α Cen A is a similar factor lower, making this star a good candidate as a star in a Mounder-like magnetic minimum.

Coronal cycles are not detected in more active stars, possibly because stochastic variability is prevalent in these objects. Moreover, any discussion of X-ray cyclic coronal variability in low-activity stars will depend crucially on the energy coverage of the measurements.

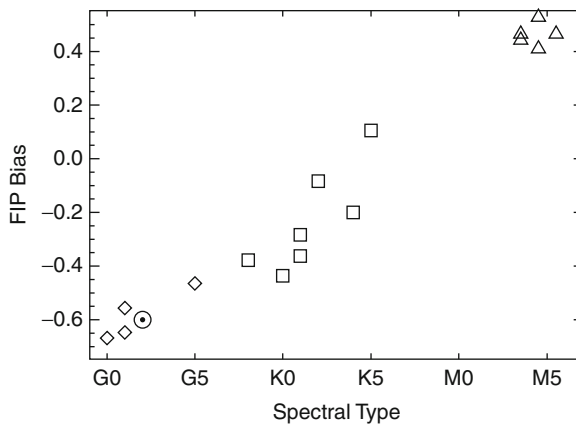
7.1.3 Coronal Abundances

High-resolution X-ray spectra provide information on coronal abundances (essentially C, N, O, Ne, Mg, Al, Si, S, Ar, Ca, Fe, Ni). Two different methods have been used to investigate stellar coronal abundance: (i) global spectral modeling that includes the determination of key elemental abundances and the emission measure distribution and (ii) ratio of emission lines that have a similar emissivity versus temperature distributions. On the Sun, coronal abundances do not reflect the photospheric composition. Although the actual abundances depend on which part of the corona is observed (with coronal holes, active regions, and flares showing quite different abundance patterns), full disk spectra show the so-called FIP (first ionization potential) effect (Feldman 1992). Essentially, elements having first ionization potential ≥ 10 eV (high FIP), like Fe or Mg, have photospheric composition, while elements having first ionization potential ≤ 10 eV (low FIP), like Ne and O, are overabundant by about a factor 4, on average, with respect to photospheric values (Feldman and Lamington 2000). An explanation for the FIP effect

is that low-FIP elements are mainly ionized in the solar chromosphere ($T \sim 5,000\text{--}10,000\text{ K}$), while high-FIP elements are mainly neutral; hence, the ionized elements are easily transported upward in the corona, where they result more abundant than the neutral elements. What about abundances in the stellar coronae? Early measurements have been hampered by the different results coming from different measurement methods. Also, the poor determination of photospheric abundances especially in fast rotators have played a role. The general scenario we have today is the following: a solar-like FIP effect is observed in cool, inactive stars. Instead, the so-called inverse FIP (I-FIP) effect, in which high-FIP elements are overabundant in corona with respect to low-FIP elements, is observed in extremely active RS CVn-type and Algol-type binaries. No FIP bias has been observed for stars with intermediate activity. There are exception of course; for instance, the cool, inactive Procyon should display a strong solar-like FIP effect. Instead, it shows no FIP bias (e.g., Sanz-Forcada et al. 2004). Wood and Linsky (2010) find a clear dependence of the FIP bias on spectral type as shown in [Fig. 10-23](#), with the earlier type stars having solar-like FIP effects and the later stars having inverse FIP effects, though the lack of photospheric abundance measurements for the M dwarfs (for which solar photospheric abundances have been assumed) means that conclusions about them are more uncertain. Anyway, no dependence of the degree of “FIP effect” on activity has been observed.

7.1.4 Density and Temperature Structure of Stellar Coronae

The first coronal densities were obtained from EUV spectra by measurements of the flux ratio of Fe lines originating from the same ionization state. As reviewed by Bowyer et al. (2000), the general scenario emerging from such measurements is that low-temperature plasma ($T \sim 1\text{ MK}$) is associated with low-density plasma, with $n_e \sim 10^9\text{--}10^{10}\text{ cm}^{-3}$, while plasma at high temperature ($T \sim 10\text{ MK}$) is typically denser, with $n_e \sim 10^{12}$. For the sake of comparison, typical solar active



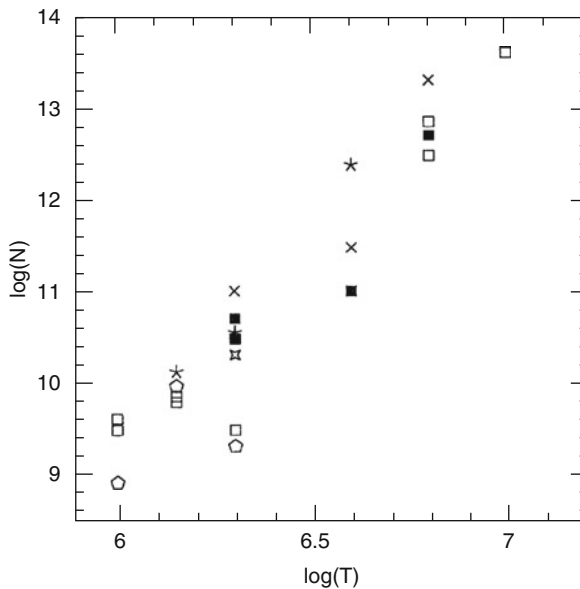
■ Fig. 10-23

High FIP bias of moderate active stars ($\log L_X < 29$) plotted versus spectral type. Different symbols mark data from different authors (see Wood and Linsky 2010, and references therein). For all the GK stars, the FIP bias calculations include corrections for measured stellar photospheric abundances, while for the M stars, solar photospheric abundances have been assumed

regions have density $n_e \sim 10^9$, with $n_e > 10^{10}$ during solar flares. The bimodal temperature/density structure emerging from EUVE data is eventually due to the lack of resolution of the temperature diagnostics.

Chandra and XMM-Newton have allowed density measurements thanks to the resolution of He-like triplets. The ratios between the *forbidden* and the *intercombination* line fluxes are sensitive to the electron density, the forbidden line being suppressed in high-density environments. The densities to which He-like transitions are sensitive increases with the formation temperature of the ion. In this way, each triplet can diagnose densities in different temperature ranges. These measurements have confirmed the trend observed by EUVE: plasma associated with warmer regions has higher density, but with the indication of a continuous transition rather than a dichotomy as shown in [Fig. 10-24](#). However, this trend may in part be affected by a bias due to the fact that triplets formed at higher temperatures are sensitive to higher values of density.

Density values obtained from the analysis of He-like triplets can be influenced by radiation pumping due to UV photons generated in chromosphere. Ness et al. (2002a) show that neglecting the influence of UV radiation field can lead to higher densities up to an order of magnitude. This effect is more pronounced for low-Z ions and hence for plasma at lower temperature. Moreover, this effect is important for stars such as Algol, for which the UV radiation field comes from the nearby hot companion. The high-density values obtained by Chandra and EUVE spectra have been questioned. For instance, for the active binary σ^2 CrB, the Fe xxI and Fe xxII He-like triplets yield a density n_e of few times 10^{12} cm^{-3} while a density $n_e < 10^{11} \text{ cm}^{-3}$ is derived from




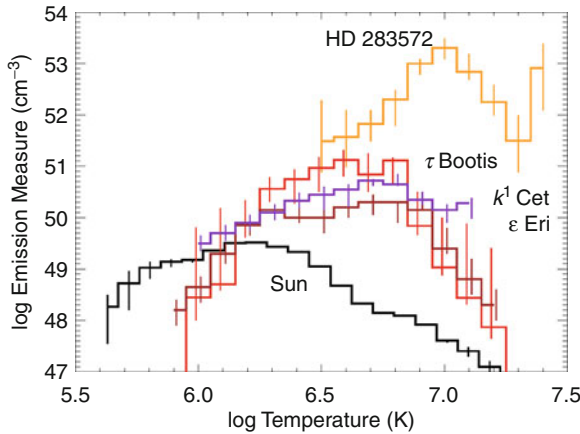
the Mg XI, despite these ions have similar formation temperatures (Osten et al. 2003). Also, line blends could systematically bias the measured line-flux ratios. In fact, many of the density measurements are near the low-density limit, so that even a small amount of blending or calibration problems can lead to inaccurate values of density, where only an upper limit would have to be estimated (e.g., Drake 2001). The resolution of this issue requires very high spectral resolution and is hence linked to the availability of next-generation X-ray telescopes.

The presence in the same corona of plasma with density n_e ranging from 10^9 to 10^{13} cm^{-3} – if it is real – is not compatible with the plasma being confined in quasi-static flux tubes. In fact, in this case, the higher density should occur in colder regions of the loop (Favata and Micela 2003).

Mewe et al. (2001) have proposed that the colder plasma ($T \sim 0.5\text{--}2 \text{ MK}$) is connected with loops that have density similar to that of solar active regions, i.e., $n_e \sim 10^9\text{--}10^{10} \text{ cm}^{-3}$, while the hotter plasma ($T \sim 5\text{--}9 \text{ MK}$) is in compact loops with density $n_e \sim 10^{12}\text{--}10^{13} \text{ cm}^{-3}$.

In the solar corona, the temperature of the plasma has a continuous distribution. Instead, coronal spectra of active stars generally require at least two temperature components to be reproduced. This result has been achieved already by the Einstein observatory despite its IPC detector had a very limited spectral resolving power (Schmitt et al. 1990), and was lately confirmed by subsequent X-ray observatories. Early investigations were aimed to understand if the 2-T (sometimes 3-T) structure is real or the result from the parameterization of a more complicated temperature structure with features having different temperature. Actually, using data from the solar corona observed by the Yohkoh satellite and reducing them to appear as if they had been observed by ROSAT, Orlando et al. (2000) found that the solar spectrum taken at the minimum of solar activity is reproduced by optically thin plasma emission model with temperature $T \sim 1 \text{ MK}$, while 2-T components are required by the solar spectrum taken at the maximum of solar activity. The same technique applied at the spectrum of a solar flare shows the presence of plasma also at $T > 10 \text{ MK}$ (Orlando et al. 2001). Hence, the presence of both cold and hot plasma seems related to the level of activity.

High-resolution spectroscopy has been very important to demonstrate that superimposed to a continuous distribution of the emission measure versus temperature, there are features that trace specific coronal components. Differential emission measure (DEM) distribution characterizes the temperature structure of the coronal plasma. Many techniques have been developed to derive DEM from X-ray or EUV spectroscopic data that resulted to be a continuous function over a wide range of temperatures. For the very active stars, the DEM shows a significant amount of plasma at high temperature (even tens of MK), with a peak around 10 MK . It is particularly interesting that the DEMs of stars very different from each other (single, binary, or evolved MS) are very similar, in particular, all the very active stars show the “bump” at high temperature, while low-activity stars do not show the high-temperature plasma (see, e.g.,  Fig. 10-25). DEMs and density measurements together suggest that the hot plasma, responsible for the “bump,” is also very dense, and is structured into spatially compact structures (loops). Ness et al. (2004) argue that, going from low to high level of activity, the hotter plasma loops fill the space between cooler loops until much of the corona is dominated by the hot plasma. As the magnetic activity level and consequently the surface magnetic filling factor increases, the coronal magnetic fields become denser, leading to increased interactions between neighboring field lines, which leads to increased heating. In their most extreme form, such interactions lead to increased levels of flaring, again resulting in increased amounts of hot, luminous plasma. This picture is essentially consistent with that described by Testa et al. (2004) and, earlier, by Güdel (1997) and Drake et al. (2000).



■ Fig. 10-25

Emission measure distributions of a PMS very active star (HD 283572), a F7 V medium activity star (τ Boo, a star hosting planet), a G5 V solar analog (κ^1 Cet), an intermediate activity level K2 V star (ϵ Eri), and the quiescent Sun (From Maggio et al. 2011, and references therein)

Unless for tidally locked rapid rotators, the youth, characterized by the high rotation regime, is the key element for a late-type star to have a high level of coronal activity. This implies that with increasing age, the X-ray spectra become softer.

7.1.5 Coronal Versus Chromospheric and TR Emission

The correlation between coronal and chromospheric emissions have been firstly observed by Ayres et al. (1981b) and Schrijver (1987). Relations among the chromospheric, transition region and coronal emissions have been derived by Ayres et al. (1995) for samples of F-K dwarfs and giants and by Fox et al. (1994) for RS CVn-type stars (cf. ▶ Table 10-1). The F9-G2 “solar-type” main-sequence stars have a significantly steeper slope in X-rays (0.1–2.4 keV) versus C iv than the cooler G8-K5 dwarfs. The populous *clump giants* follow essentially the same behavior as the cooler MS stars, showing a nearly 1:1 correlation between X-rays and C iv. However, some F dwarfs, the Hertzsprung gap giants, and virtually all of the supergiants fall systematically to lower X-ray luminosities with respect to their transition region and chromospheric luminosities.

The correlations between emission fluxes coming from different layers of the stellar external atmosphere depend very little from the source spectral and luminosity classes (cf. ▶ Fig. 10-26). This suggests that a strong coupling between different atmospheric layers does occur. The flux–flux correlations show less scatter if the “basal flux” is subtracted from the chromospheric line fluxes (e.g., Ca II and Mg II), as is expected being the “basal” chromospheric emission not due to magnetic dissipation.

Ayres et al. (1996) suggest that the nonlinearity of the corona-transition region flux–flux relation is due to the increasing coronal temperatures with increasing activity (see ▶ Sect. 7.1.4), bringing a progressively larger fraction of the emission into the X-ray band.

■ **Table 10-1**
Correlation among X-ray, UV, and EUV luminosities, normalized to the stellar bolometric fluxes
 $\log R_x = \log f_x/f_{\text{bol}}$ for contemporary observed samples of dwarfs, giants, and RS CVn-type binaries

Diagnostics	Spectral types	Slope $\pm 1\sigma$	Reference
X-rays ^a versus C IV	F9-K5 V	1.53 ± 0.06	Ayres et al. (1997)
	F9-G2V	1.72 ± 0.10	Ayres et al. (1997)
	G8-K5 V	1.32 ± 0.09	Ayres et al. (1997)
	RS CVn type	1.24 ± 0.19	Fox et al. (1994)
X-rays ^a versus Mg II h	F9-K5V	2.91 ± 0.13	Ayres et al. (1997)
	F9-G2 V	2.85 ± 0.17	Ayres et al. (1997)
	G8-K5 V	2.93 ± 0.23	Ayres et al. (1997)
	RS CVn type ^b	0.71 ± 0.91	Fox et al. (1994)
C IV versus Mg II h	F9-K5V	1.77 ± 0.12	Ayres et al. (1997)
	F9-G2V	1.64 ± 0.16	Ayres et al. (1997)
	G8-K5 V	2.03 ± 0.21	Ayres et al. (1997)
X-rays ^a versus S1 ^c	F9-K5 V, IV	1.24 ± 0.12	Ayres et al. (1997)
	RS CVn type	1.02 ± 0.17	Fox et al. (1994)
X-rays ^a versus S2 ^c	F9-K5 V, IV	1.38 ± 0.13	Ayres et al. (1997)
	RS CVn type	1.06 ± 0.18	Fox et al. (1994)

^aX-rays measured in the 0.1–2.4 keV band of *ROSAT* PSPC
^bX-rays versus Mg II h&k for RS CVn-type binaries
^cS1 (69–146 Å) and S2 (113–200 Å) bands of the *ROSAT* WFC

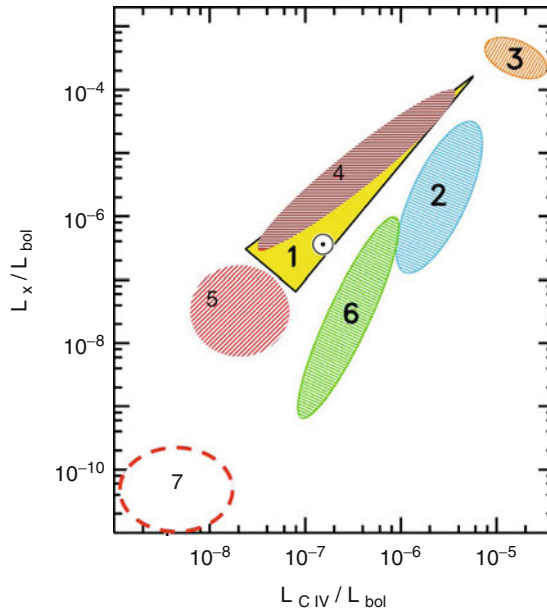
Sanz-Forcada et al. (2011a) have computed synthetic spectra for the whole EUV range (~1–912 Å) of a large sample of stars with exoplanets using as input coronal models from available X-ray observations, finding

$$\log L_{\text{EUV}} = (4.80 \pm 1.99) + (0.860 \pm 0.073) \log L_X. \tag{10.5}$$

7.2 Radio Emission from Stellar Coronae

Active stars show both quiescent and flaring radio emission. The quiescent flux, characterized by slow variations on time scales of hours and days and a low level of polarization, has brightness temperature 10^8 – 10^{10} K, i.e., about 2–3 orders of magnitude higher than the coronal temperatures measured in the soft X-rays. Actually, the radio emission is nonthermal – as suggested by the relatively flat radio spectra – typically optically thin gyrosynchrotron radiation originating from the interaction of mildly relativistic electrons and magnetic fields (e.g., Gary and Linsky 1981; Güdel 1994). A comprehensive even not recent review on “stellar radio astronomy” is given by Güdel (2002). Here we summarize the principal aspects of radio emission as observed in different categories of active stars. Radio flares will be discussed in ► Sect. 10.

RS CVn stars constitute one of the major classes of stellar system radio sources. Their radio observations allow us to directly investigate nonthermal phenomena occurring in stellar coronae. Chiuderi Drago and Franciosini (1993) reproduce their quiescent radio emission assuming flares as the origin of the accelerated electrons responsible for the gyrosincrotron



■ Fig. 10-26

X-ray/C IV flux-flux diagram. Normalization by the bolometric fluxes removes the twin biases of different distances and diameters. *Shaded zones represent* (1) GK dwarfs (the *circled dot marks cycle-average solar ratio*), (2) “X-ray deficient” Hertzsprung gap giants, (3) hyperactive RS CVn binaries, (4) active clump (G8K0) giants, (5) inactive but still coronal K0 giants, (6) GK supergiants, and (7) noncoronal (K1) red giants (Adapted from Ayres et al. 2003 and references therein)

emission. However, HR 1099, a typical RS CVn system, shows low-level radio emission lasting for several weeks with no detectable flare events (Umana et al. 1995). Apparently, the HR 1099 behavior cast doubt on the Chiuderi Drago and Franciosini (1993) model, but the possibility that accelerated electrons come from a multitude of undetected microflares cannot be ruled out.


Radio interferometry performed with VLBI, able to provide $\lesssim 1$ mas angular resolution, has revealed a typical two-components structure for very active star; a compact, and often variable core, plus an extended halo, the latter surrounding the whole system in case of active binaries like the RS CVn’s (e.g., Mutel et al. 1985; Trigilio et al. 2001). VLBA observations of UX Ari (a bright RS CVn binary system) revealed a 8.4-GHz radio source, having size comparable to the projected separation of the binary components, circularly polarized and with a polarization gradient across the resolved source, suggesting the existence of interacting magnetic fields between the components or a large magnetic loop anchored to one component (Beasley and Güdel 2000). VLBA radio maps of Algol reveal two oppositely polarized radio lobes separated 1.4 times the active K star size (Mutel et al. 1998), i.e., a very different scenario with respect to what can be expect in terms of solar analogy.



Late-type dKe-dMe stars, the so-called “flare stars,” have been widely observed in the radio domain during both flares and quiescence (see White et al. 1989 for a VLA survey of nearby flare stars). Also for these stars, gyrosynchrotron emission from mildly relativistic electrons is

considered the main origin of the observed radio luminosity. This interpretation is compatible with a spectrum that decreases toward higher frequencies. However, some spectra show a U-shaped behavior; the flux decreases with frequency until 5–8 GHz, and then increases to higher frequencies. For example, Güdel and Benz (1989) measured the spectrum of UV Cet from 0.33 to 22 GHz and found that the flux decreased from 0.33 to 8 GHz and rose from 15 to 22 GHz. They interpreted the observed spectrum as consisting of a nonthermal component (gyrosynchrotron emission) dominant at low frequencies and gyroresonance emission from the hot component of the corona plasma (that is detected in X-ray observations), which is dominant at the higher frequencies. This scenario may be appropriate only for some stars and only occasionally. In fact, White et al. (1994) showed that the observed 15 GHz fluxes of a sample of dMe stars, whose X-ray observations reveal the presence of a hot plasma component, are too low to be consistent with the Güdel and Benz (1989) interpretation. Leto et al. (2000) found that the radio fluxes due to the free–free emission of the electrons in the coronal plasma of a small sample of dMe stars lie far below the observed radio fluxes.

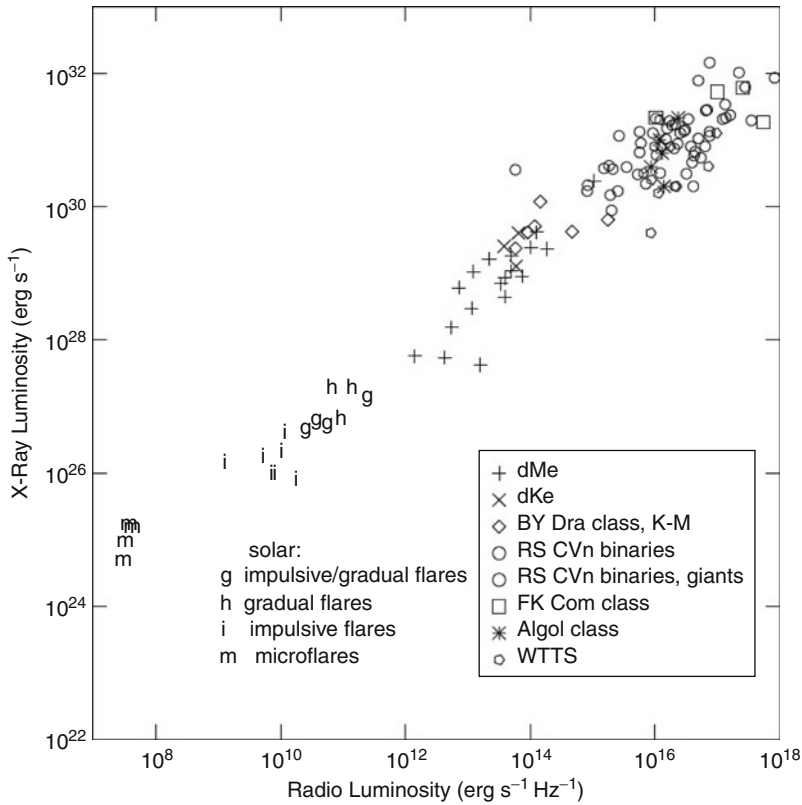
The radio corona of M active dwarfs has been resolved with VLBI/VLBA observations. For instance, YZ CMi (M4.5 Ve) has a VLBA resolved corona at 3.6 cm with size of 1.7 ± 0.3 the stellar photospheric size, not changing even during a flare, and a brightness temperature of 7.3×10^7 K, consistently with that expected from gyrosynchrotron emission (Pestalozzi et al. 2000). UV Cet B has been resolved at 3.6 cm into two components separated by $4.4 R_*$, oriented along the axis of the binary orbit, i.e., parallel and very likely close to the stellar rotation axis (Benz et al. 1998). The authors infer that large coronal loops extending to at least two stellar radii are necessary to trap the mildly relativistic electrons responsible for the gyrosynchrotron emission.

7.2.1 X-Ray Versus Radio Emission

Güdel and Benz (1993) and Benz and Güdel (1994) show that several classes of active stars and solar flares follow the relation $L_X/L_R \sim 10^{15 \pm 1}$ (Hz) (cf.  Fig. 10-27), where L_X is the soft X-ray luminosity and L_R the luminosity at 5–8 GHz, where the radio spectrum is typically flat. At the origin of such a correlation, a physical link between the mechanisms of stellar coronal heating – whose X-ray emission is a signature, and particle acceleration, which determines the radio waves, must exist; the electrons producing gyrosynchrotron radiation are generally assumed to be the tail of an energetically important nonthermal population, producing the soft X-rays.

Active binaries, like the RS CVn's, lie in the top-right corner of  Fig. 10-27. M dwarfs obey the X-ray/radio relation shown in  Fig. 10-27; however, exceptions occur. For instance, AU Mic, which is the most luminous nondegenerate X-ray source within 10 pc, has a very large radiative output from the transition region and corona, $L_{\text{TR+corona}}/L_{\text{bol}} \sim 10^{-2}$ (Pagano et al. 2000), but a radio luminosity at 8.4 GHz at least ten times less than predicted by the X-ray/radio relation (Leto et al. 2000).

Brown dwarfs have revealed very interesting behavior at radio wavelengths. In fact, they do not obey the X-ray/radio relation, being their quiescent and flaring radio emission similar to what is observed for late M dwarfs (Berger 2002; Berger et al. 2005), while their X-ray luminosity is by several orders of magnitude lower than expected according to the Güdel and Benz (1993) relation. This means that particles are accelerated and emit in the magnetic fields around BDs; however, there is not a corona in the sense of a hot thin plasma radiating X-rays.



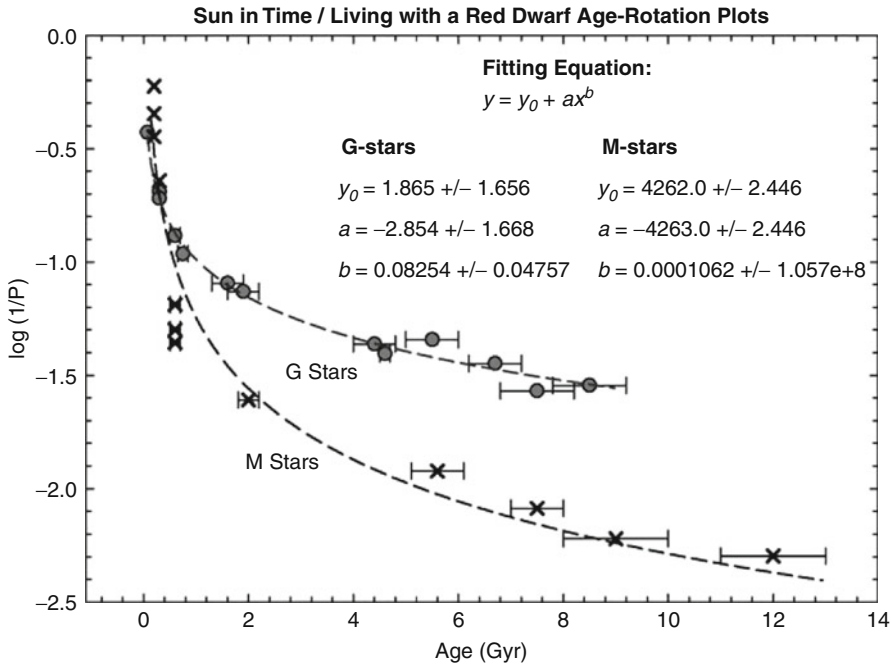
■ Fig. 10-27

Correlation between quiescent radio and X-ray luminosities of magnetically active stars (symbols) and solar flares (letters; after Benz and Güdel 1994)

8 Solar-Like Stellar Winds

The velocity structure, mass-loss rate, and ionization of the outflowing gas in the low-temperature (about 10^4 K) winds of late-type giants and supergiants can be inferred from high-resolution spectra of optically thick UV resonance lines: e.g., H I, O I, Mg II, and C II. Unfortunately, the winds of main-sequence stars like the Sun are too hot and too thin to be detected by the same diagnostics used for late-type giants and supergiants.

Spectroscopic analyses of stellar H I Ly- α lines have proven to be the best way so far to clearly detect and measure weak solar-like winds (Wood 2004). The Ly- α profile of a dwarf star having a solar-like wind changes from its initial appearance at the star because of absorption due to (a) the stellar astrosphere, (b) the local interstellar medium, and (c) the heliosphere. By modeling these absorption features, Wood et al. (2002) performed the first quantitative measurements of mass-loss rates for G and K dwarf stars. As shown in Fig. 10-28, their results suggest that mass-loss rates increase with activity (from X-ray surface flux), except for very high activity levels possibly because of fundamental changes in magnetic field topology (Wood et al. 2005). The observed trend of mass loss increase with activity for solar-like stars suggests that mass loss



■ Fig. 10-29

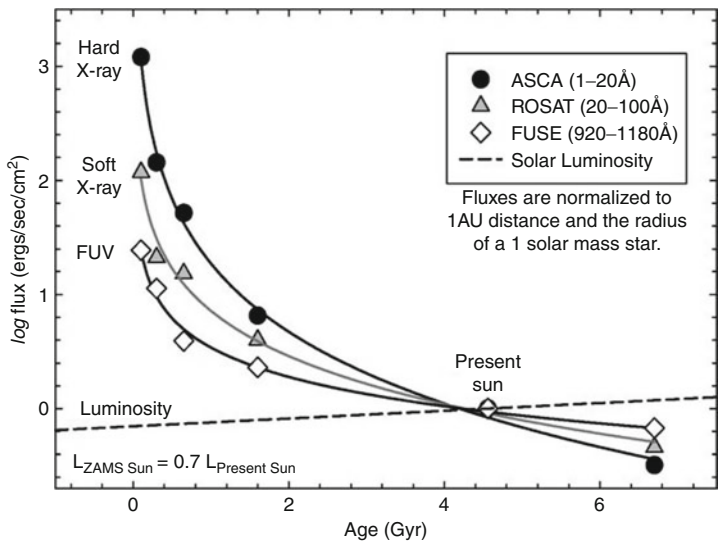
The angular velocity Ω ($\propto 1/P_{\text{rot}}$) plotted versus stellar age for solar-type stars and a representative sample of dM0–5 stars (From Guinan and Engle 2009)

was rotating more than ten times faster than today. For a given age, the dM stars have longer rotation periods indicating that they spin down faster than the more massive G stars. The rapid magnetic braking of dM stars is most likely due to more efficient dynamos (from their deeper convective zones) and their lower masses.

As a consequence of their fast rotation, young stars have vigorous magnetohydrodynamic (MHD) dynamos and correspondingly strong coronal X-ray and transition region/chromospheric FUV–UV emissions (up to several hundred times stronger than the present Sun), as shown in ► Fig. 10-30 for the solar-type stars. Specifically, the mean X-ray luminosity of the solar analogues decreases by 10^3 times from the youngest to solar-age stars in the sample.

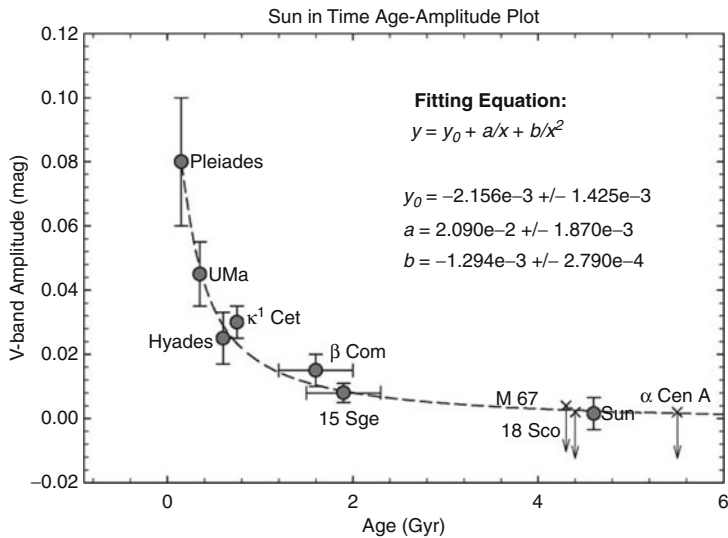
Also, the amplitudes of rotationally modulated light curves decreases with the stellar age, as shown in ► Fig. 10-31 for the G dwarfs. For instance, G dwarfs in the Pleiades (~100 Myr) have light variations in the range ~0.06–0.12 mag while the solar light modulation is typically ~1.2 mmag.

X-ray observations of open clusters have been of paramount relevance to understand (i) how coronal emission evolves with age as a consequence of the evolution of angular momentum in stars with different masses and (ii) how coronal emission depends upon stellar chemical composition. Favata and Micela (2003) show the median X-ray luminosity of G and M dwarfs as a function of stellar age from observations of star-forming regions, open clusters



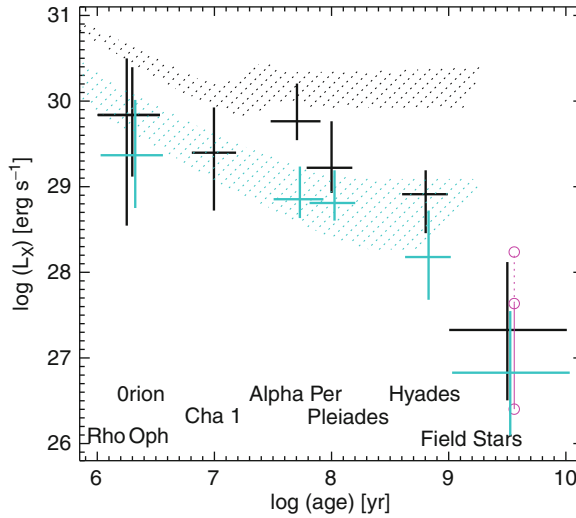
■ Fig. 10-30

Irradiances (flux densities at a reference distance of 1 AU) of the “Sun in Time” program stars are plotted for hard (1–20 Å) and soft (20–100 Å) X-rays (from ASCA and ROSAT, respectively) along with the total FUV emission fluxes (from FUSE). The fluxes have been normalized to the corresponding mean fluxes of the Sun. Also plotted is the change in the bolometric luminosity (L) of the Sun based on modern stellar evolution models. Note that the ZAMS Sun had a luminosity of 30% less than today (From Guinan and Engle 2009 and references therein)



■ Fig. 10-31

The V-band light amplitudes from starspot rotational modulations and cycles plotted against stellar age. Note that the younger stars have significant light variations due to the larger areal coverage of starspots (From Guinan and Engle 2009 and references therein)



■ Fig. 10-32

Median X-ray luminosity of G (black/heavy line) and M (cyan/light line) dwarf stars versus age. The latter have abscissa slightly shifted toward right for the sake of clarity. Vertical bars indicate the 1σ spread of the X-ray luminosity distributions. From the top to the bottom, the two dashed areas mark the range of L_X corresponding to the saturation level of $\log L_X/L_{\text{bol}} = -3$, for the mass range of pop. I dG ($0.8\text{--}1 M_{\odot}$) and dM ($0.3\text{--}0.5 M_{\odot}$) stars, respectively. The circles give the luminosity of the Sun at minimum, maximum, and flaring phases (Adapted from Favata and Micela 2003 and references therein)

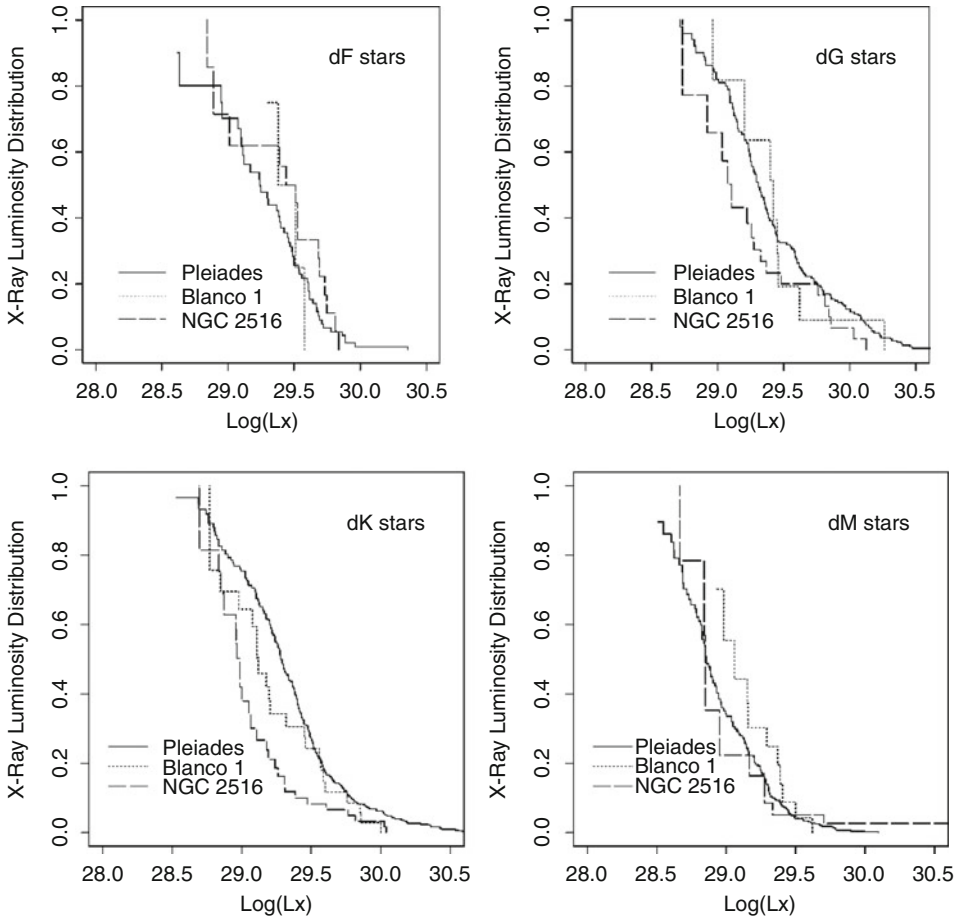
and field stars (cf. ● Fig. 10-32). The coronal emission level of solar-type stars, expressed in terms of X-ray luminosity, remains almost constant until an age of 10^8 year and then decreases steeply until the solar age. Actually, a significant fraction of G dwarfs in clusters younger than the Pleiades (~ 100 Myr) emits at the saturation level, while in older clusters, all the stars become non-saturated and the mean X-ray luminosity starts to decrease. For low-mass stars, the decrease of L_X with age starts at a later age. Most of the M dwarfs in the Pleiades and α Per and part of ones in the Hyades are saturated. The different evolution of dM with respect to dG stars can be explained as due to the dependence of rotational evolution on stellar mass; as dM stars spin down more slowly, they will emit at the saturation level for a longer time.

Also, the EUV radiation ($\sim 1\text{--}912 \text{ \AA}$) decreases with stellar age. For a large sample of stars with exoplanets, Sanz-Forcada et al. (2011a) have derived

$$\log L_{\text{EUV}} = (29.12 \pm 0.11) - (1.24 \pm 0.15) \log \tau \quad (10.6)$$

with τ the age in Gyr.

The dependence of X-ray luminosity from metallicity has been investigated, e.g., by Pillitteri et al. (2003) who computed the X-ray luminosity distributions for F, G, K, and M dwarfs for NGC 2516, Blanco 1, and the Pleiades (cf. ● Fig. 10-33). These clusters have the same age (~ 100 Myr), and metallicity values of $Z \sim 0.5, 1$, and $2 Z_{\odot}$, respectively, hence are well suited to explore the metallicity effects on coronal emission of young stars. The X-ray luminosity distributions of F dwarfs in the three clusters are indistinguishable, while metal-poor G and K dwarfs



■ Fig. 10-33

X-ray luminosity distributions, from Pillitteri et al. (2003), for dF stars (top left panel), dG stars (top right panel), dK stars (bottom left panel) and dM stars (bottom right panel) for NGC 2516 (long-dashed line), Blanco 1 (dotted line), and Pleiades (solid line)

appear to be less active than their more metal-rich counterparts. For dM stars, no significant difference is visible between the samples, indicating that the role played by metal abundance in determining their coronal emission is minor.

10 Flares

Stellar flares are unpredictable phenomena due to the sudden transformation of energy stored in the magnetic field into thermal and kinetic energy. They involve the whole stellar atmosphere from the photosphere to the corona, and their effects can be very important also for the stellar planetary environment. Excellent reviews on stellar flares have been those by Haisch et al. (1991) and Linsky (2000). For a complete description of the solar flares, see Benz (2008).

Flares on the Sun occur in association to active regions, and their frequency is in phase with the solar activity cycle. The radiation is emitted throughout the electromagnetic spectrum; radio, optical, UV, X- and gamma rays. The energy released during a medium solar flare is typically of the order of 10^{27} erg s⁻¹, but they range between 10^{22} (nanoflares) and 10^{32} erg s⁻¹. The most energetic flares are the *white light flares*, i.e., those detected in broadband optical observations. The temperature of the flaring plasma is of the order of 10–20 MK but can reach up to 100 MK.

The first study in the literature concerning stellar flares is by Joy and Humason (1949) who detected the continuum and emission lines rising in a spectrum of a dMe star, L726-8. Observations of stellar flares, as of several astrophysical phenomena, have been driven by the available instruments: blue photographic observations mainly of stars in open clusters in the 1960s–1970s; UVB photoelectric photometry of field stars in the 1970s–1980s; multiwavelength spectroscopic campaigns (UV, optical, X-ray, radio, IR) in the 1980s–1990s; X-ray imaging observations of single field stars and open clusters since the 1990s and serendipitous UV and optical observations, e.g., with GALEX, SWIFT/UVOT, and SDSS recently.

10.1 Flaring Stars

Flares are observed on all the stars that show coronal type X-ray emission: the so-called “flare stars,” brown dwarfs, active binaries (BY Dra, RS Cvn, FK Com type), T Tauri stars, hybrid stars, and clump giants, both pre-He-flash and post-He-flash. In active binaries, there are evidence of flares occurring in the circumstellar environment. In YSO, flares generally originate in the interaction between coronal magnetic field and the accreting disk. Fully convective M dwarfs and brown dwarfs show X-ray emission mostly during flares.

“Flare stars” or “UV Cet-type stars” are dwarfs of spectral type from late-Ge to M9e, with the suffix *e* indicating the presence of emission lines in the optical spectrum (cf. Gershberg et al. 1999, for a detailed catalogue UV Cet-type flare stars). Typically, flare stars are rapid rotators with respect to stars of the same spectral class without emission lines in the optical spectrum. For instance, while v_{rot} for M stars is ~ 2 km s⁻¹, for dMe’s is ~ 10 km s⁻¹ (Marcy and Chen 1992). Moreover, about 30% of dMe’s are in binary systems (Pettersen 1991). The flare star density in the solar neighborhood is 0.056 stars pc⁻³ (Shakhovskaya 1995). They have starspots in the photosphere as inferred by the observed rotational modulation of their light curves (Rodonò 1980). The flare frequency for a typical UV Cet-type star has been measured in EUV, X-ray, and B-band to be 0.16–0.37 flares h⁻¹ (Leto et al. 1997; Osten et al. 2005).

10.2 Flare Characteristics

Multiwavelength observations are essential for understanding the response of different atmospheric layers to the catastrophic release of magnetic energy we call flare: *radio waves* carry information on the role of nonthermal particles in the dynamics of flares; the *optical continuum and lines* respond to the dramatic heating of the lower chromosphere and photosphere; the *UV/EUV emission lines* provide an important source of plasma cooling, as well as lower atmospheric heating, during the gradual phase of stellar flares, and thus are valuable diagnostics of the flare evolution and the *X-ray continuum and emission lines* describe the response of the thermal coronal plasma to the flare. Hence, flares can be considered as probes for the 3D nature of the stellar external atmosphere.

Energy of stellar flares ranges from the $10^{22} \text{ erg s}^{-1}$ of nanoflares to even $10^{40} \text{ erg s}^{-1}$ of very large flares in RS CVn-type binaries, implying that a common physical origin must exist in order the manifestation of a similar phenomenology being possible despite the different energy involved.

In the optical band (e.g., U band) flares from UV Cet-type stars are characterized by impulsive rise times, lasting from few seconds to few minutes, followed by an exponential decay, often with multiple spikes (cf., [Fig. 10-34](#)), lasting typically less than 1 h (cf. [Fig. 10-35-left](#)). In YSO, instead, the flare duration goes from one hour to several days (cf. [Fig. 10-35-right](#)), as shown by soft X-rays observations (Wolk et al. 2005).

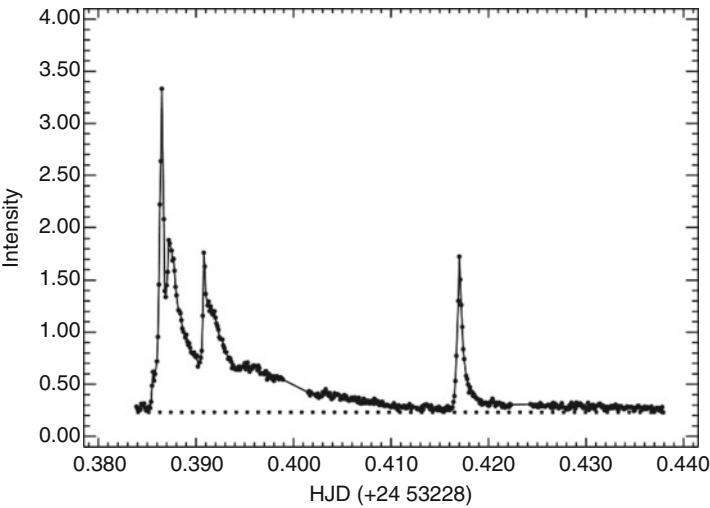


Fig. 10-34
U-band observations of a typical flare from EV Lac (M3.5e) (From Dal and Evren (2010))

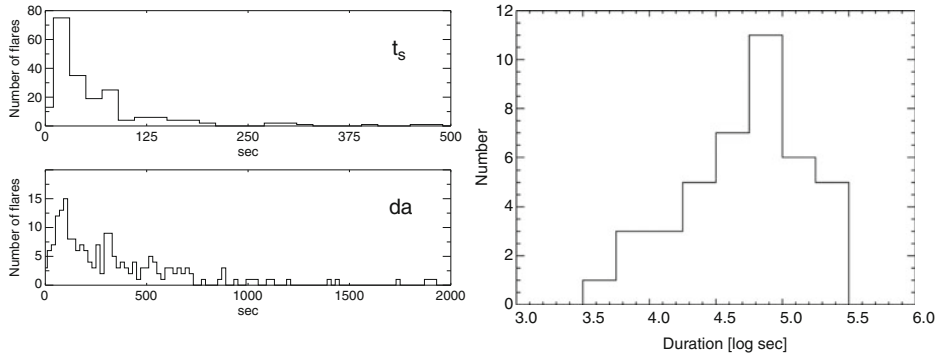


Fig. 10-35
Left: Flare rise and decay times observed on the UV Ceti-type star EV Lac (Leto et al. 1997). **Right:** Flare duration in YSOs in the Orion Nebula Cluster (Wolk et al. 2005)

Broadband optical continuum enhancements are easily recognizable only in the cooler targets, like dKe and dMe stars, because of the high contrast between flares and their low photospheric temperature. Instead, white light flares have been elusive in active binaries like RS CVn or FK Com-type binaries, and only few observations are reported in the literature. For instance, the long-lasting 9.8 days flare of YY Gem (K2 III), an FK Com-type star, with an energy output in the UBV bands of $\sim 6 \times 10^{39}$ erg corresponding to $L_{UBV}/L_{bol} \sim 3 \times 10^{-3}$ (Cutispoto et al. 1992).

A general correlation does exist between the time characteristics of a flare and its energy, in the sense that large flares tend to last longer than small flares. Actually, the time scales of flares depend from the flaring region, such as size, strength of the magnetic field where the magnetic reconnection takes place, electron density, and flaring plasma temperature, rather than the host star.

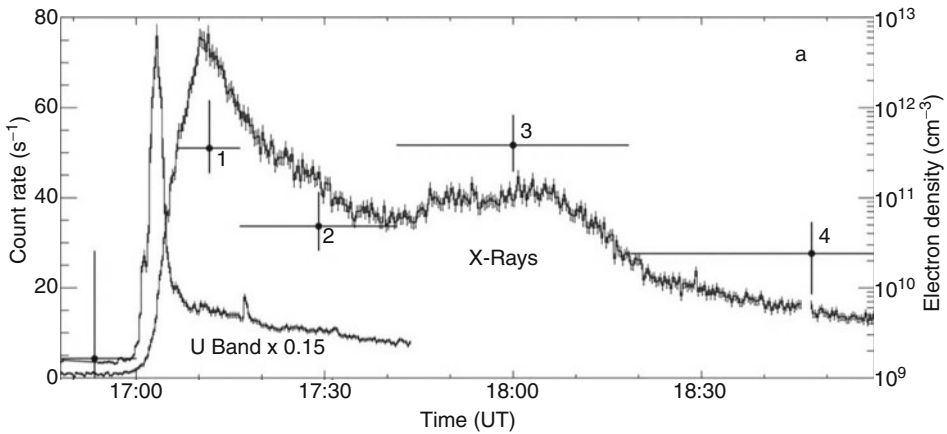
As summarized by Pettersen (1991), in some flare stars, flares appear to occur randomly (Poisson distribution). However, flares occurring in groups (e.g., Umana et al. 1995), or indication of flaring activity concentrated at preferred stellar longitude (e.g., Leto et al. 1997 and reference therein), have also been reported. A short-term cycle in flare occurrence (and also in spots) has been identified on the RS CVn-type system UX Ari (Massi et al. 2005); this is reminiscent of the short-term cycles (50–90 days) in the solar flare occurrence rate (see, e.g., Lou 2000).

10.3 Flare Classification

Two main classes of stellar flares, one with short rise times – of order of minutes – and decay times of order of a few tens of minutes and the other with longer rise times and very long decay times, exceeding several kiloseconds, are observed also in the soft X-rays. The former are similar to the solar “compact flares” which origin from the interaction of neighboring compact loops. Their rise duration occurs on time scales smaller than the dynamical time scale of filament eruption (van den Oord et al. 1988). The latter are reminiscent of solar “two-ribbon flares” which happen in complex loop arcades (10^4 – 10^5 km). “Two-ribbon flares” originate from the opening up of magnetic fields due, for instance, to a filament eruption and the return to closed field lines. However, the use of a classification scheme valid for solar flare is not straightforward in some cases. In fact, there could be physical conditions very different than on the Sun, such as significant tidal effects in close-binary stars, extremely dense packing of magnetic fields, polar magnetic fields, magnetic fields from a distributed dynamo, inter-binary magnetic fields with magnetospheres of global dimensions, and star-disk magnetic fields in pre-main-sequence stars.

10.4 The Flare “Standard Model” and the Neupert Effect

A large number of flares can be described by the so-called “standard model”: the region where the magnetic field reconnection takes place, i.e., where the flare starts, is located at the top of coronal loops. Here electrons (and ions) are accelerated to energies up to a few MeV, so that they emit gyrosynchrotron radiation and, running down along the magnetic field lines, reach the “cold” chromospheric plasma. Here they produce hard X-ray photons at the impact and heat the chromospheric plasma at temperatures typical of coronal plasma. In this way, part of the chromosphere evaporates into the corona. Radio emission, hard X-ray emission, and continuum optical emission from the heated chromosphere are related to each other within



■ Fig. 10-36

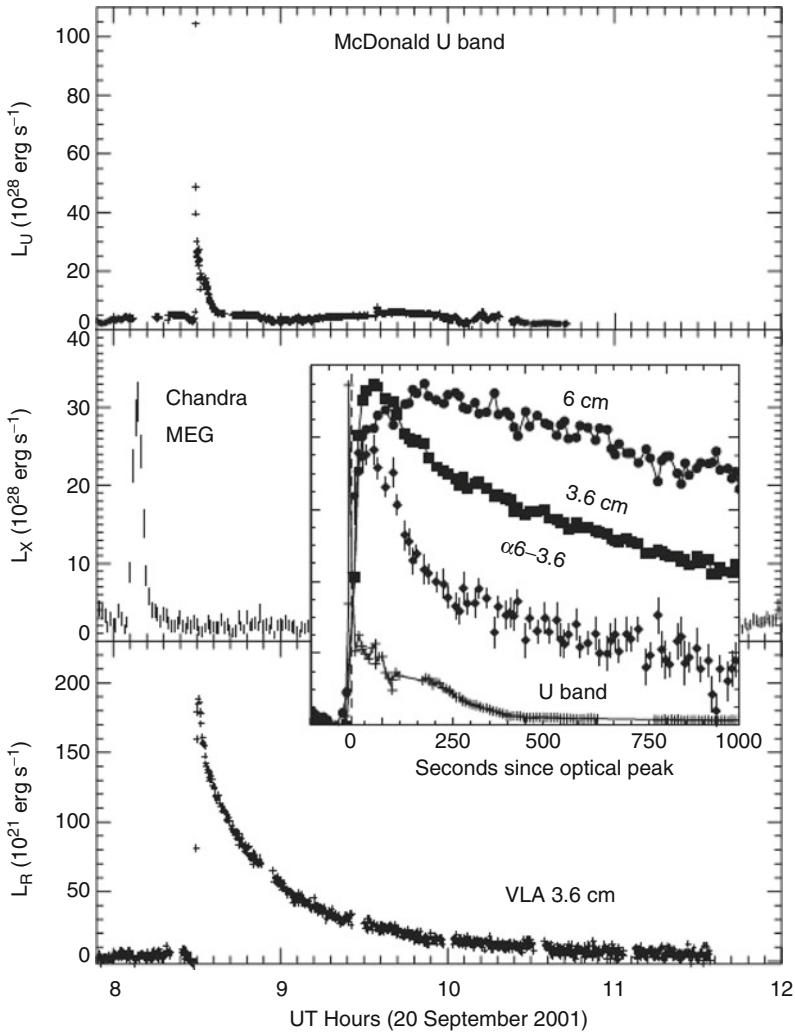
X-ray and U-band observations of a flare on Proxima Centauri (M6Ve). Crosses are the electron densities (scale on the right axis) measured, thanks to the O vii He-like triplets from XMM spectra integrated in time intervals as shown by the *horizontal bars* (After Güdel et al. 2002)

seconds and trace the impulsive phase of flares. The soft X-rays, instead, are emitted when the magnetic loop is filled with the gas that evaporates from the chromosphere; this occurs on time scales of tens of seconds to minutes. Hence, in the “standard model” scenario, that applies in the case shown in ► Fig. 10-36, there is few minute delay between the onset of the flare in the optical continuum (and radio or hard X-rays) and the onset of the soft X-rays emission.

If the “standard model” applies, the radio gyrosynchrotron emission, hard X-ray, and optical continuum emission are induced on the time scales related to the traveling path of the electrons (seconds) and then grow in proportion to the flow of high-energy particles, unless these are blocked by some trapping mechanisms. Instead, the cooling time of the thermal plasma in an extended coronal loop is governed by the mechanisms of radiation and conduction with a time scale of several minutes up to hours. Since the soft X-rays are radiated proportionally to the accumulated thermal energy, we have that $L_{O,HXR,R}(t) \propto dL_X(t)/dt$, with R, O, and XHR meaning radio, optical and hard X-rays, respectively. This relation is known as “Neupert effect.” Its observation is a good diagnostic for the chromospheric evaporation predicted by the “standard model.”

The Neupert effect is observed for ~50% of the impulsive solar flares and in some gradual solar flare (Dennis and Zarro 1993). It has been clearly detected also for many stellar flares: e.g., flares from the dMe stars AD Leo and UV Cet (Hawley et al. 1995; Güdel et al. 1996) and from the RS CVn-type binary HR 1099 (Osten et al. 2004), often with a ratio between the energy losses in the two energy bands similar to the corresponding luminosity ratio “in quiescence.”

However, from flare to flare, we can still have large departures from the “standard model” scenario. For instance, Haisch et al. (1981) describe an X-ray flare without an accompanying signal in the optical or in the UV. As opposite case, H γ and U band flares on YZ CMi have been observed by Doyle et al. (1986, 1988) without a soft X-ray counterpart. Among the several “non standard model” observations we have: an impulsive optical flare with following gradual radio



■ Fig. 10-37

Optical, radio, and soft X-ray light curves of EV Lac (M3.5e) (From Osten et al. (2005))

emission (van den Oord et al. 1996); gyrosynchrotron emission that peaks after the soft X-rays (Osten et al. 2000); the absence of any X-ray response during radio flares (e.g., Fox et al. 1994).

► Figure 10-37 shows the result of soft X-ray, radio and optical observations of the EV Lac reported by Osten et al. (2005). The optical flare peaks 20 min after the impulsive X-ray event, too far apart in time to causally connect under any realistic length scales or velocities. There is a close temporal association between the large U-band flare and the radio flare, yet no indication of an accompanying X-ray flare exists. The long-observed timescales for radiative decay of the radio emission implies low ambient electron densities ($n_e \sim 10^8 \text{ cm}^{-3}$). As argued by the

authors, the lack of a soft X-ray counterpart for the optical/radio flare could be due to different reasons; (i) the flaring structure has a too low density to produce soft-X-ray radiation⁴; (ii) the flare involves only transition region plasma, as it was observed by Ayres et al. (2001) for HR 1099; (iii) particles are trapped in the loop, so that very few of them precipitate to heat the surrounding chromosphere and produce subsequent evaporation; or (iv) most of the energy goes into accelerating particles (e.g., protons), which would allow deeper penetration of the photosphere and produce continuum emission but no line emission or significant evaporation.

An important test for the “standard model” would be the detection of non-thermal hard X-ray emission contemporary to the optical continuum impulsive phase. Despite the PDS instrument on board the BeppoSAX satellite was sensitive enough, nonthermal hard X-rays have not been detected even during large soft X-ray flares (e.g., Franciosini et al. 2001; Rodonò et al. 1999).

10.5 Signature of Coronal Mass Ejections

A solar-like CME was observed during a $\Delta U \sim 2$ mag flare from AD Leo (Rodonò et al. 1988), with the H_γ line showing a shift toward the blue of $\sim 6,000 \text{ km s}^{-1}$, that, in terms of kinetic energy, corresponds to $\sim 5 \times 10^{34} \text{ erg s}^{-1}$. The energy, mass and ejection velocity budgets are ~ 400 , 40 and 4 times higher than observed for typical solar CMEs.

A mass ejection episode has been detected also during a strong flare in the RS CVn-type binary HR 1099. Its Mg II h&k line profiles display very broad wings at flare peak, while the mid-UV continuum is enhanced. By analyzing these data, Busà et al. (1999) concluded that material was ejected by the secondary star in the direction of the primary.

A huge X-ray flare observed by Beppo-SAX on Algol (Favata and Schmitt 1999) reveals the occurrence of a CME episode: during the initial (rising) phase of the flare the best-fit absorbing column density is large, and decays on time scales of some tens of ks. The authors interpret this as possibly associated with moving, cool absorbing material in the line of sight, i.e., a major coronal mass ejection associated with the flare’s onset.

10.6 Flares and Coronal Heating

Although large flares are a conspicuous contributor to transient heating of stellar coronae, smaller-scale events – the so-called microflares and nanoflares – might play a key role in the “steady” heating of the outer layers of cool stars.

Does the energy input involve microflaring, and is therefore sporadic, or does it involve MHD waves, electric current dissipation, and/or other effects which could be quasisteady? That magnetic energy release occurring in discrete much smaller events than observed, i.e., involving events of 10^{24} – $10^{25} \text{ erg s}^{-1}$, sustains the coronal heating has been proposed and widely accepted by several authors (cf. Haisch et al. 1991; Parker 1988, and references therein). As quantified by Hudson (1991), the flare driven coronal luminosity can be expressed as

$$L_f = \int_{P_{\min}}^{P_{\max}} \frac{dN}{dP} P dP = \frac{k}{2-\alpha} [P_{\max}^{2-\alpha} - P_{\min}^{2-\alpha}] \quad (10.7)$$

⁴Quiescent X-ray structures in active M dwarfs have typically $n_e \geq 10^{10} \text{ m}^{-3}$ at $T_e \geq 3 \text{ MK}$ (cf. Sect. 7.1.4), and the thermal X-ray emission depends from n_e^2

where dN/dP is the number of flares versus power (or energy flux) per event, which empirically follows a power-law distribution, $dN/dP = kP^{-\alpha}$, with k varying from star to star. Equation 10.7 shows that if $\alpha < 2$, the coronal luminosity due to flares is dominated by the large events, that are not enough to account for coronal heating. Instead, if the statistics of micro/nano flares is such that $\alpha > 2$, then L_f is dominated by small events. Hudson (1991) argues that $\alpha \sim 1.8$ for the total power in both the solar and stellar case for flares above some threshold. Actually, our limited capability to detect low energy flares renders this one an open topic. Robinson et al. (2001) studied the statistics of transient bursts in high time resolution UV observations of AU Mic with *HST*/STIS, and concluded that the power-law slope of the occurrence rate versus time-integrated flux was considerably steeper for low-energy flares than for the rarer high-energy ones, implying that microflares potentially can account for a significant portion, if not all, of the coronal heating. However, this is such an important issue that more conclusive evidence is needed: the investigation should be extended to a larger sample of flare stars, collecting a significant number of events for each target.

Thanks to time resolution <0.01 s provided by time-tagged data, GALEX has serendipitously detected a class of flares characterized by constant temperature during the flux increase. Robinson et al. (2005) and Welsh et al. (2006) suggest these flares are due to avalanches of nanoflares in a plage. Another class of flares detected by GALEX are characterized by the increase of both flux and temperature during the flare development. These flares are probably the classical explosive events like the solar two-ribbon flares. GALEX data suggest also that in the FUV ($\lambda_{\text{eff}} \sim 1,530$ Å), the flare flux is mainly due to C IV ($\sim 50\%$), while weaker emission lines and continuum provide the remaining 50%. In the NUV ($\lambda_{\text{eff}} \sim 2,315$ Å), flux is due to Mg II (10%), Fe II (17%), Al II and C III (14%) plus continuum. Coronal lines provide 2% of the FUV and 10% of the NUV fluxes.

10.7 Superflares

Superflares occur on main-sequence stars in spectral classes F8-G8 with no unusual properties (specifically rapid rotation, high chromospheric activity, close-binary companions, or very young age). The observed superflare energy ranges from 10^{33} to 10^{38} erg s $^{-1}$. Their typical duration is about 1 h but ranges from a fraction of an hour up to days. Superflares emit radiation at least from the X-rays to the optical, with indicated temperatures (Schaefer et al. 2000) from 15,000 K (from the He I emission line) to 10 keV (from X-ray continuum). The most accepted theory for superflares is that they are caused by magnetic reconnection between fields of the primary star and a close-in Jovian planet (Rubenstein and Schaefer 2000).

11 Stellar Magnetic Activity and Extrasolar Planets

The characterization of stellar activity is fundamental also to evaluate the conditions for habitability of exoplanets. In the same way, the investigation of solar activity (Space Weather) ultimately looks for the effects of solar activity on terrestrial life; the characterization of stellar activity is required to understand the evolution of extrasolar planetary systems and of the exoplanet atmospheres. Moreover, the capability to model stellar activity is required to

unambiguously detect the weak signature of a planet often buried in the stellar intrinsic noise originating also from magnetic activity.

11.1 Stellar Magnetic Activity as Source of Noise

Magnetic activity is a source of noise in interpreting the radial velocity curves of stars hosting exoplanets. In fact, it can produce radial velocity jitter larger than the amplitude of the radial velocity variations associated to the planets, which for a $1 M_{\oplus}$ planet orbiting around a $1 M_{\odot}$ star amount to about $0.1\text{--}0.3 \text{ m s}^{-1}$ for orbital semimajor axis of 1 and 0.1 AU, respectively. The reason is that active regions induce disturbances in the shape of spectral lines. For instance, the macro- and micro-turbulence parameters change because of the action of magnetic fields on convective motions. In fact, if we consider convective cells far from surface magnetic fields, upward motions dominate with respect to downward motions, determining blueshifted line bisectors. If surface magnetic fields are present, the amount of blueshift is reduced. Meunier et al. (2010) show that convective blueshift in solar magnetized regions are reduced by $\sim 2 \text{ m s}^{-1}$ up to 8 m s^{-1} . These authors suggest that this effect is the main source of radial velocity perturbations in solar-like stars.

To extract the exoplanet signal from radial velocity curves of active stars hosting exoplanets, Lanza et al. (2010, 2011) have used spots and faculae positions and filling factors derived from spot modeling of stellar light curves to synthesize radial velocity time series that can be compared with the observed ones to cancel the effect of stellar activity.

Magnetic activity is a source of noise also for the search and characterization of exoplanet transits. The transit of a large group of spots on the solar surface causes a relative decrease in the optical flux of 3.3×10^{-3} . For comparison, the Earth and Jupiter transits across the solar disc produce a relative flux variation of the order of 10^{-4} and 10^{-2} , respectively. Therefore, to search for planetary transits in the presence of stellar activity, several techniques have been developed to filter out or fit the microvariability of the star (cf., e.g., Aigrain and Irwin 2004; Moutou et al. 2005; Bonomo and Lanza 2008; Bonomo et al. 2009).

Reducing the effect of magnetic activity is fundamental in modeling the exoplanetary transit and occultation. For instance, during the transit, the planet passes in front of the star, partially blocking the stellar light. The amount of this dimming is a direct measure of the relative size of the planet compared to the size of the star (R_p/R_*). If the planet has an atmosphere, it is not an opaque disk with a sharp edge. Instead, the opacity will gradually diminish with height as the atmosphere becomes more tenuous. The stellar light that passes through the atmosphere is absorbed by atoms and molecules or Rayleigh scattered. If at a given wavelength the atmosphere has a low opacity, the stellar light travels through the planets atmosphere unhindered, while if it has a high opacity, the atmosphere becomes transparent at a high altitude only. This results in an increase in the observed planetary radius at wavelengths having a high opacity compared to that at a low opacity. As a consequence, any observed wavelength dependence of the transit depth can be used to probe the planet atmospheric composition. As it is intuitive, any effect on the observed light curve due to stellar activity must be taken into account.

Ballerini et al. (2012) compare activity-induced flux variations in different passbands with planetary transits and quantify how they affect the determination of the planetary radius and the analysis of the transmission spectroscopy in the study of planetary atmospheres. These authors suggest that the monitoring of the systems by using broadband photometry, from visible to infrared, helps us to constrain activity effects. The ratio of the relative variations in the stellar

fluxes at short wavelength optical bands (e.g., U or B) to near-infrared ones (e.g., J or K) can be used to distinguish starspot brightness dips from planetary transits in a stellar light curve. In addition to the perturbations in the measurement of the planetary radius, starspots can affect the determinations of both the relative semimajor axis and the inclination of the planetary orbit, which have a significant impact on the derivation of the stellar density from the transit light curves.

11.2 Star Planet Interaction

Among the 755 extrasolar planets discovered up to February 2012, about 22% are massive planets ($M_p \sin i > 0.2M_J$) in tight orbits (<0.1 AU) around their parent stars. Since the pioneering paper of Cuntz et al. (2000), star–planet tidal and magnetospheric interactions have been claimed to be a possible cause of stellar activity increase. Observational support to the Star–Planet interaction (SPI) phenomenon has been provided by several authors: e.g., Shkolnik et al. (2003, 2005, 2008), who found chromospheric emission of stars hosting planets to be variable with the planet orbital period. Space-based optical photometry may have also caught this phenomenon. The variance of CoRoT-2 stellar flux is modulated in phase with the planet orbital period reaching a maximum near the time of transit (Pagano et al. 2009). Also, the MOST space telescope found a spot that persisted for hundreds of rotation cycles on τ Boo (Walker et al. 2008), while Lanza et al. (2009a, 2010) report hints for SPI for CoRoT-4 and CoRoT-6. In support to the presence of SPI, Kashyap et al. (2008) claimed that stars with hot Jupiters have a higher X-ray flux than stars with distant planets. On the contrary, Poppenhaeger et al. (2010) found no evidence of X-ray emission bias for stars hosting extrasolar planets. The nature of SPI appears to be strongly affected by both the stellar and planetary magnetic fields. Some SPI models consider that conditions similar to those observed for the Jupiter–Io system are met and consequently similar processes are expected to occur on far larger scales. However, a severe limitation of such models is that the energy released in the interaction is about three orders of magnitude smaller than needed to account for the observed chromospheric excess powers (cf. Lanza 2009 and references therein). Other models consider that the localized reconnection at the corona/planetary magnetospheric boundary can trigger a reconfiguration of the coronal magnetic field on a large scale, comparable to the star–planet separation, thus accounting for the energy budget of the star–planet magnetic interaction (SPMI), thanks to the large coronal volumes involved in the energy release (cf., e.g., Lanza 2009; Cohen et al. 2009). A review on SPI has been given by Lanza (2011).

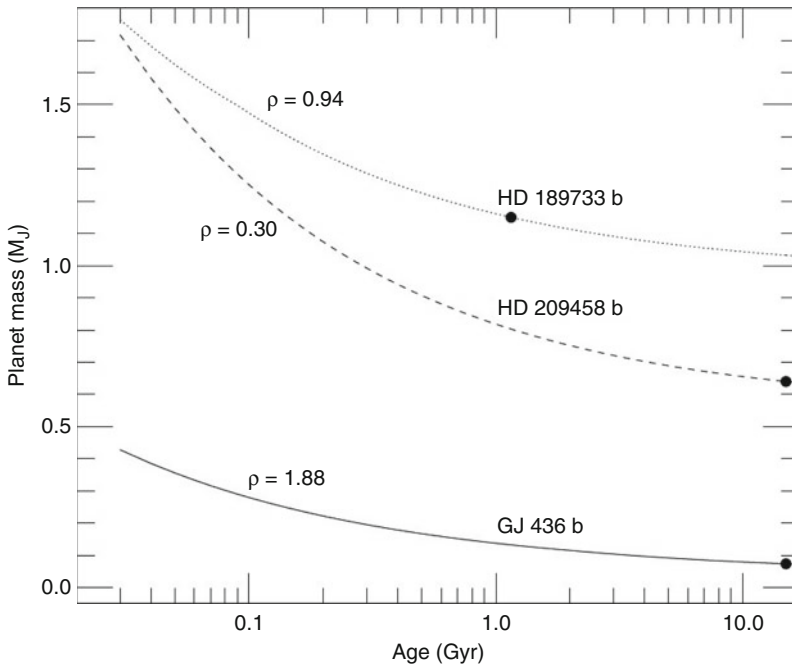
11.3 Stellar Activity Effects on Exoplanet Atmospheres

One of the most important factors affecting the planetary atmospheric evolution and water inventories is the activity associated with the chromospheres and coronae of the planets' host stars. The relevant physical phenomena include intermittent and energetic flares, coronal mass ejections (CMEs), high energetic stellar radiation, enhanced coronal X-rays, and increased chromospheric UV emission which can ionize, heat, expand, chemically modify, and erode the atmospheres of terrestrial planets during the early phase of their evolution (e.g., Lundin et al. 2007; Scalo et al. 2007; Lammer 2009). Much of the photochemistry in the planetary atmosphere is driven by the shorter wavelength photons and, therefore, our knowledge of the EUV and UV

spectra and their variability is critical. X-rays, EUV, and UV photons produce effects at different heights in the atmosphere of the planet, with X-rays penetrating deeper in the atmosphere and EUV/UV radiation mainly ionizing the atoms in the upper atmosphere.

As discussed by Lammer et al. (2011), upper atmosphere of Earth-like planets should experience non-hydrostatic conditions so that dynamically outward flowing neutral atoms can interact with the stellar plasma flow so that huge hydrogen coronae and energetic neutral atoms (ENA) can be produced via charge exchange. Valuable diagnostics of these features are the size of the extended upper atmospheres and the velocities of the surrounding hydrogen atoms.

Sanz-Forcada et al. (2010, 2011a) have studied a large sample of planet-hosting stars for which X-ray observations are available. By using the observed and synthesized X-ray and EUV SEDs, respectively, these authors conclude that massive planets ($M_p \sin i > 1.5 M_J$) may survive only if exposed to low accumulated coronal radiation. For instance, the planet HD 209458 b ($\sim 0.7 M_J$, $a \sim 0.05$ AU) might have lost more than $1 M_J$ already, and other cases, like τ Boo b, could be losing mass at a rate of $3.4 M_\oplus/\text{Gyr}$. Sanz-Forcada et al. (2011b) point out that the strongest erosive effects would take place during the first stages of the stellar life, as expected given the dependence of coronal activity from age and that planets with higher density resist better the radiation effects, as shown in ► Fig. 10-38. Given the rapid growing of the research on exoplanets, both on the experimental and theoretical sides, we expect that modeling of the



■ Fig. 10-38

Planetary mass evolution as an effect of the EUV flux from the mother star. Mass losses include evaporation by coronal radiation and losses through the Roche Lobe. Planet mean density is indicated, as well as the current stellar age of the planet, calculated using the X-ray luminosity (From Sanz-Forcada et al. 2011b)

stellar activity effects on different type of planetary atmospheres will be developed in great detail in the coming years.

Acknowledgments

I thank G. Cutispoto, A.F. Lanza, A.C. Lanzafame, G. Leto, S. Messina, and G. Scandariato for their availability, long discussions, and collaborative environment in our group at Catania Astrophysical Observatory. I am also grateful to Prof. Marcello Rodonò who introduced me to the study of magnetic activity and encouraged me to work on the implementation of new experimental facilities.

References

- Achour, H., Brekke, P., Kjeldseth-Moe, O., & Maltby, P. 1995, Observed redshifts in the solar transition region above active and quiet regions. *ApJ*, 453, 945. doi:10.1086/176454
- Acton, L. 1996, Comparison of YOHKO X-ray and other solar activity parameters for November 1991 to November 1995. *ASPC*, 109, 45
- Aigrain, S., & Irwin, M. 2004, Practical planet prospecting. *MNRAS*, 350, 331–345. doi:10.1111/j.1365-2966.2004.07657.x
- Alonso, R., Auvergne, M., Baglin, A., et al. 2008, Transiting exoplanets from the CoRoT space mission. II. CoRoT-Exo-2b: a transiting planet around an active G star. *A&A*, 482, L21–L24. doi:10.1051/0004-6361:200809431
- Antiochos, S. K., Haisch, B. M., & Stern, R. A. 1986, On the dividing line for stellar coronae. *ApJ*, 307, L55–L59. doi:10.1086/184727
- Antolin, P., & Shibata, K. 2010, The role of torsional Alfvén waves in coronal heating. *ApJ*, 712, 494–510. doi:10.1088/0004-637X/712/1/494
- Appenzeller, I., & Mundt, R. 1989, T Tauri stars. *A&AR*, 1, 291–334. doi:10.1007/BF00873081
- Argiroffi, C., Maggio, A., & Peres, G. 2007, X-ray emission from MP Muscae: an old classical T Tauri star. *A&A*, 465, L5–L8. doi:10.1051/0004-6361:20067016
- Avrett, E. H., & Loeser, R. 2008, Models of the solar chromosphere and transition region from SUMER and HRTS observations: formation of the extreme-ultraviolet spectrum of hydrogen, carbon, and oxygen. *ApJS*, 175, 229–276. doi:10.1086/523671
- Ayres, T. R. 2002, Does the Sun have a full-time comosphere? *ApJ*, 575, 1104–1115. doi:10.1086/341428
- Ayres, T. R. 2009, The cycles of α centauri. *ApJ*, 696, 1931–1949. doi:10.1088/0004-637X/696/2/1931
- Ayres, T. R. 2010, StarCAT: a catalog of space telescope imaging spectrograph ultraviolet echelle spectra of stars. *ApJS*, 187, 149–171. doi:10.1088/0067-0049/187/1/149
- Ayres, T., & France, K. 2010, Warm coronal rain on young solar analog EK draconis? *ApJ*, 723, L38–L43. doi:10.1088/2041-8205/723/1/L38
- Ayres, T. R., Linsky, J. L., Vaiana, G. S., Golub, L., & Rosner, R. 1981a, The cool half of the H-R diagram in soft X-rays. *ApJ*, 250, 293–299. doi:10.1086/159374
- Ayres, T. R., Marstad, N. C., & Linsky, J. L. 1981b, Outer atmospheres of cool stars. IX – A survey of ultraviolet emission from F-K dwarfs and giants with IUE. *ApJ*, 247, 545–559. doi:10.1086/159065
- Ayres, T. R., Fleming, T. A., Simon, T., et al. 1995, The RIASS coronathon: joint X-ray and ultraviolet observations of normal F-K stars. *ApJS*, 96, 223–259. doi:10.1086/192118
- Ayres, T. R., Simon, T., Stauffer, J. R., Stern, R. A., Pye, J. P., & Brown, A. 1996, Sleuthing the dynamo: Hubble Space Telescope faint object spectrograph observations of solar-type dwarfs in young galactic clusters. *ApJ*, 473, 279. doi:10.1086/178143
- Ayres, T. R., Brown, A., Harper, G. M., Bennett, P. D., Linsky, J. L., Carpenter, K. G., & Robinson, R. D. 1997, Digging deeper in the coronal graveyard. *ApJ*, 491, 876. doi:10.1086/304989
- Ayres, T. R., Simon, T., Stern, R. A., Drake, S. A., Wood, B. E., & Brown, A. 1998, The coronae of moderate-mass giants in the hertzsprung

- gap and the clump. *ApJ*, 496, 428. doi:10.1086/305347
- Ayres, T. R., Brown, A., Osten, R. A., Huenemoerder, D. P., Drake, J. J., Brickhouse, N. S., & Linsky, J. L. 2001, *Chandra*, EUVE, HST, and VLA multi-wavelength campaign on HR 1099: instrumental capabilities, data reduction, and initial results. *ApJ*, 549, 554–577. doi:10.1086/319051
- Ayres, T. R., Brown, A., Harper, G. M., Osten, R. A., Linsky, J. L., Wood, B. E., & Redfield, S. 2003, Space telescope imaging spectrograph survey of far-ultraviolet coronal forbidden lines in late-type stars. *ApJ*, 583, 963–984. doi:10.1086/345409
- Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. 1995, Chromospheric variations in main-sequence stars. *ApJ*, 438, 269–287. doi:10.1086/175072
- Baliunas, S. L., Donahue, R. A., Soon, W., & Henry, G. W. 1998, Activity cycles in lower main sequence and POST main sequence stars: the HK project. *ASPC*, 154, 153
- Ballerini, P., Micela, G., Lanza, A. F., & Pagano, I. 2012, Multiwavelength flux variations induced by stellar magnetic activity: effects on planetary transits. arXiv:arXiv:1201.3514
- Barnes, J. R., Lister, T. A., Hilditch, R. W., & Collier Cameron A. 2004, High-resolution Doppler images of the spotted contact binary AE Phe. *MNRAS*, 348, 1321–1331. doi:10.1111/j.1365-2966.2004.07452.x
- Beasley, A. J., & Güdel, M. 2000, VLBA Imaging of Quiescent Radio Emission from UX Arietis. *ApJ*, 529, 961. doi:10.1086/308292
- Benz, A. O. 2008, Flare observations. *LRSP*, 5, 1
- Benz, A. O., & Güdel, M. 1994, X-ray/microwave ratio of flares and coronae. *A&A*, 285, 621–630
- Benz, A. O., Conway, J., & Güdel, M. 1998, First VLBI images of a main-sequence star. *A&A*, 331, 596–600
- Berdugina, S. V. 2005, Starspots: a key to the stellar dynamo. *LRSP*, 2, 8
- Berdugina, S. V., & Usoskin, I. G. 2003, Active longitudes in sunspot activity: century scale persistence. *A&A*, 405, 1121–1128. doi:10.1051/0004-6361:20030748
- Berdugina, S. V., Berdugin, A. V., Ilyin, I., & Tuominen, I. 1999, The active RS Canum venaticorum binary II Pegasi. IV. The SPOT activity cycle. *A&A*, 350, 626–634
- Berger, E. 2002, Flaring up all over-radio activity in rapidly rotating late M and L dwarfs. *ApJ*, 572, 503–513. doi:10.1086/340301
- Berger, E., Rutledge, R. E., Reid, I. N. et al. 2005, The magnetic properties of an L dwarf derived from simultaneous Radio, X-Ray, and H α Observations. *ApJ*, 627, 960. doi:10.1086/430343
- Bertout, C. 1989, T Tauri stars – Wild as dust. *ARA&A*, 27, 351–395. doi:10.1146/annurev.aa.27.090189.002031
- Böhm-Vitense, E. 2007, Chromospheric activity in G and K main-sequence stars, and what it tells us about stellar dynamos. *ApJ*, 657, 486–493. doi:10.1086/510482
- Bonomo, A. S., & Lanza, A. F. 2008, Modelling solar-like variability for the detection of Earth-like planetary transits. I. Performance of the three-spot modelling and harmonic function fitting. *A&A*, 482, 341–347. doi:10.1051/0004-6361:20078833
- Bonomo, A. S., Aigrain, S., Bordé, P., & Lanza, A. F. 2009, Modelling solar-like variability for the detection of Earth-like planetary transits. II. Performance of the three-spot modelling, harmonic function fitting iterative nonlinear filtering, and sliding boxcar filtering. *A&A*, 495, 647–653. doi:10.1051/0004-6361:200810726
- Bowyer, S., Drake, J. J., & Vennes, S. 2000, Extreme ultraviolet astronomy. *ARA&A*, 38, 231–288. doi:10.1146/annurev.astro.38.1.231
- Brage, T., Judge, P. G., & Brekke, P. 1996, Intercombinations and allowed transitions in O IV. *ApJ*, 464, 1030. doi:10.1086/177390
- Brown, A., Linsky, J. L., & Ayres, T. R. 1994, The Corona of the K5 Giant γ Dra, and its Relation to the HybridChromosphere Stars, in *AIP Conf. Proc.* 313, *The Soft X-Ray Cosmos*, ed. E. M. Schlegel & R. Petre (New York: AIP), p. 36
- Brown, A., Korhonen, H., Berdyugina, S., et al. 2011, Starspot variability and evolution from modeling Kepler photometry of active late-type stars. *IAUS*, 273, 78–82. doi:10.1017/S1743921311015043
- Busà, I., Pagano, I., Rodonò, M., Neff, J. E., & Lanzafame, A. C. 1999, Chromospheric imaging of the active binary system V 711 Tauri = HR 1099 in December 1992. *A&A*, 350, 571–581
- Carpenter, K. G., Karovska, M., Lyon, R. G., Mozurkewich, D., & Schrijver, C. 2009, Stellar imager (SI): enhancements to the mission enabled by the constellation architecture (Ares I/Ares V). *SPIE*, 7436. doi:10.1117/12.825393
- Cassatella, A., Altamore, A., Badiali, M., & Caradini, D. 2001, On the Wilson-Bappu relationship in the Mg II k line. *A&A*, 374, 1085–1091. doi:10.1051/0004-6361:20010816
- Catala, C., Donati, J.-F., Shkolnik, E., Bohlender, D., & Alecian, E. 2007, The magnetic field of the planet-hosting star τ Bootis. *MNRAS*, 374, L42–L46. doi:10.1111/j.1745-3933.2006.00261.x

- Catalano, S., Biazzo, K., Frasca, A., & Marilli, E. 2002, Measuring starspot temperature from line depth ratios. I. The method. *A&A*, 394, 1009–1021. doi:10.1051/0004-6361:20021223
- Catura, R. C., Acton, L. W., & Johnson, H. M. 1975, Evidence for X-ray emission from Capella. *ApJ*, 196, L47–L49. doi:10.1086/181741
- Cayrel de Strobel, G., & Bentolila, C. 1989, In search of real solar twins. II. *A&A*, 211, 324–340
- Chabrier, G., & Küker, M. 2006, Large-scale α^2 -dynamo in low-mass stars and brown dwarfs. *A&A*, 446, 1027–1037. doi:10.1051/0004-6361:20042475
- Chiuderi Drago, F., & Franciosini, E. 1993, Flaring and quiescent radio emission of UX ARIETIS-A time-dependent model. *ApJ*, 410, 301–308. doi:10.1086/172746
- Cincunegui, C., Díaz, R. F., & Mauas, P. J. D. 2007, H-alpha; and the Ca II H and K lines as activity proxies for late-type stars. *A&A*, 469, 309–317. doi:10.1051/0004-6361:20066503
- Cohen, O., Drake, J. J., Kashyap, V. L., & Gombosi, T. I. 2009, The effect of magnetic spots on stellar winds and angular momentum loss. *ApJ*, 699, 1501–1510. doi:10.1088/0004-637X/699/2/1501
- Collier-Cameron, A., Duncan, D. K., Ehrenfreund, P., Foing, B. H., Kuntz, K. D., Penston, M. V., Robinson, R. D., & Soderblom, D. R. 1990, Fast spectroscopic variations on rapidly rotating cool dwarfs – Part three – Masses of circumstellar absorbing clouds on AB-Doradus. *MNRAS*, 247, 415
- Collier Cameron, A., Walter, F. M., Vilhu, O., et al. 1999, Multisite observations of surface structures on AB Doradus in 1994 November. *MNRAS*, 308, 493–509. doi:10.1046/j.1365-8711.1999.02771.x
- Cuntz, M., Saar, S. H., & Musielak, Z. E. 2000, On stellar activity enhancement due to interactions with extrasolar giant planets. *ApJ*, 533, L151–L154. doi:10.1086/312609
- Cutispoto, G., Pagano, I., & Rodonò, M. 1992, A long-duration optical flare on the FK Com-type star HD 32918 (= YY Men). *A&A*, 263, L3–L6
- Dal, H. A., & Evren, S. 2010, A new method for classifying flares of UV Ceti type stars: differences between slow and fast flares. *AJ*, 140, 483–489. doi:10.1088/0004-6256/140/2/483
- Del Zanna, G., Landini, M., & Mason, H. E. 2002, Spectroscopic diagnostics of stellar transition regions and coronae in the XUV: AU Mic in quiescence. *A&A*, 385, 968–985. doi:10.1051/0004-6361:20020164
- Dempsey, R. C., Neff, J. E., & Lim, J. 2001, Simultaneous observations of variability at all atmospheric levels of V824 ara (HD 155555). *AJ*, 122, 332–348. doi:10.1086/321102
- Dennis, B. R., & Zarro, D. M. 1993, The Neupert effect – What can it tell us about the impulsive and gradual phases of solar flares? *SoPh*, 146, 177–190. doi:10.1007/BF00662178
- Donati, J.-F., Semel, M., Carter, B. D., Rees, D. E., & Collier Cameron, A. 1997, Spectropolarimetric observations of active stars. *MNRAS*, 291, 658
- Donati, J.-F., Collier Cameron, A., & Petit, P. 2003a, Temporal fluctuations in the differential rotation of cool active stars. *MNRAS*, 345, 1187–1199. doi:10.1046/j.1365-2966.2003.07101.x
- Donati, J.-F., Collier Cameron, A., Semel, M., et al. 2003b, Dynamo processes and activity cycles of the active stars AB Doradus, LQ Hydrae and HR 1099. *MNRAS*, 345, 1145–1186. doi:10.1046/j.1365-2966.2003.07031.x
- Donati, J.-F., Jardine, M. M., Gregory, S. G., et al. 2007, Magnetic fields and accretion flows on the classical T Tauri star V2129 Oph. *MNRAS*, 380, 1297–1312. doi:10.1111/j.1365-2966.2007.12194.x
- Doyle, J. G., Butler, C. J., Haisch, B. M., & Rodonò, M. 1986, Simultaneous X-ray and optical monitoring of the flare star YZ CMi. *MNRAS*, 223, 1P–5P
- Doyle, J. G., Butler, C. J., Callanan, P. J., Tagliaferri, G., de La Reza, R., White, N. E., Torres, C. A., & Quast, G. 1988, Rotational modulation and flares on RS CVn and BY Dra systems. VIII – Simultaneous EXOSAT and H-alpha observations of a flare on the dMe star GL 644 AB (Wolf 630) on 24/25 August 1985. *A&A*, 191, 79–86
- Drake, J. J. 2001, Stellar coronae: the first 25 years. *ASPC*, 234, 53
- Drake, J. J., Peres, G., Orlando, S., Laming, J. M., & Maggio, A. 2000, On stellar coronae and solar active regions. *ApJ*, 545, 1074–1083. doi:10.1086/317820
- Dupree, A. K., Lobel, A., Young, P. R., Ake, T. B., Lin-sky, J. L., & Redfield, S. 2005, A far-ultraviolet spectroscopic survey of luminous cool stars. *ApJ*, 622, 629–652. doi:10.1086/428111
- Durney, B. R., Mihalas, D., & Robinson, R. D. 1981, A preliminary interpretation of stellar chromospheric Ca II emission variations within the framework of stellar dynamo theory. *PASP*, 93, 537–543. doi:10.1086/130878
- Eberhard, G., & Schwarzschild K. 1913, On the reversal of the calcium lines H and K in stellar spectra. *ApJ*, 38, 292–295. doi:10.1086/142037

- Engle, S. G., & Guinan, E. F. 2011, Red Dwarf Stars: Ages, Rotation, Magnetic Dynamo Activity and the Habitability of Hosted Planet. *ASPC*, 451, 285
- Favata, F., & Micela, G. 2003, Stellar coronal astronomy. *SSR*, 108, 577–708. doi:10.1023/B:SPAC.0000007491.80144.21
- Favata, F., & Schmitt, J. H. M. M. 1999, Spectroscopic analysis of a super-hot giant flare observed on Algol by BeppoSAX on 30 August 1997. *A&A*, 350, 900–916.
- Favata, F., Micela, G., Orlando, S., Schmitt, J. H. M. M., Sciortino, S., & Hall, J. 2008, The X-ray cycle in the solar-type star HD 81809. XMM-Newton observations and implications for the coronal structure. *A&A*, 490, 1121–1126. doi:10.1051/0004-6361:200809694
- Feigelson, E., Townsley, L., Güdel, M., & Stassun, K. 2007, X-ray properties of young stars and stellar clusters. Protostars and planets conference, 313–328
- Fekel, F. C., & Balachandran, S. 1993, Lithium and rapid rotation in chromospherically active single giants. *ApJ*, 403, 708–721. doi:10.1086/172242
- Feldman, U. 1992, Elemental abundances in the upper solar atmosphere. *PhysS*, 46, 202–220. doi:10.1088/0031-8949/46/3/002
- Feldman, U., & Laming, J. M. 2000, Element abundances in the upper atmospheres of the Sun and stars: update of observational results. *PhysS*, 61, 222. doi:10.1238/Physica.Regular.061a00222
- Flaccomio, E., Micela, G., & Sciortino, S. 2003, Observational clues for a role of circumstellar accretion in PMS X-ray activity. *A&A*, 397, 611–621. doi:10.1051/0004-6361:20021484
- Fontenla, J. M., Curdt, W., Haberleiter, M., Harder, J., & Tian, H. 2009, Semiempirical models of the solar atmosphere. III. Set of non-LTE models for far-ultraviolet/extreme-ultraviolet irradiance computation. *ApJ*, 707, 482–502. doi:10.1088/0004-637X/707/1/482
- Fox, D. C., Linsky, J. L., Veale, A., et al. 1994, Rotational modulation and flares on RS Canum Venaticorum and BY Draconis stars. 18: Coordinated VLA, ROSAT, and IUE observations of RS CVn binaries. *A&A*, 284, 91–104
- Franciosini, E., Pallavicini, R., & Tagliaferri, G. 2001, BeppoSAX observation of a large long-duration X-ray flare from UX Arietis. *A&A*, 375, 196–204. doi:10.1051/0004-6361:20010830
- Güdel, M. 2002, Stellar radio astronomy: probing stellar atmospheres from protostars to giants. *ARA&A*, 40, 217–261. doi:10.1146/annurev.astro.40.060401.093806
- Güdel, M. 2004, X-ray astronomy of stellar coronae. *A&AR*, 12, 71–237. doi:10.1007/s00159-004-0023-2
- Güdel, M., Audard, M., Skinner, S. L., & Horvath, M. I. 2002, X-ray evidence for flare density variations and continual chromospheric evaporation in Proxima Centauri. *ApJ*, 580, L73–L76. doi:10.1086/345404
- Günther, H. M., Schmitt, J. H. M. M., Robrade, J., & Liefke, C. 2007, X-ray emission from classical T Tauri stars: accretion shocks and coronae? *A&A*, 466, 1111–1121. doi:10.1051/0004-6361:20065669
- Gary, D. E., & Linsky, J. L. 1981, First detection of nonflare microwave emissions from the coronae of single late-type dwarf stars. *ApJ*, 250, 284–292. doi:10.1086/159373
- Gary, G. A. 2001, Plasma beta above a solar active region: rethinking the paradigm. *Sol Phys* 203, 71–86.
- Gershberg, R. E. 2005, Solar-type activity in main-sequence stars. *Stam book*. doi:10.1007/3-540-28243-2
- Gershberg, R. E., Katsova, M. M., Lovkaya, M. N., Terebizh, A. V., & Shakhovskaya, N. I. 1999, Catalogue and bibliography of the UV Cet-type flare stars and related objects in the solar vicinity. *A&AS*, 139, 555–558. doi:10.1051/aas:1999407
- Getman, K. V., Feigelson, E. D., Broos, P. S., Micela, G., & Garmire, G. P. 2008a, X-ray flares in orion young stars. I. Flare characteristics. *ApJ*, 688, 418–436. doi:10.1086/592033
- Getman, K. V., Feigelson, E. D., Micela, G., Jardine, M. M., Gregory, S. G., & Garmire, G. P. 2008b, X-ray flares in orion young stars. II. Flares, magnetospheres, and protoplanetary disks. *ApJ*, 688, 437–455. doi:10.1086/592034
- Gilliland, R. L., & Dupree, A. K. 1996, First image of the surface of a star with the Hubble Space Telescope. *ApJ*, 463, L29. doi:10.1086/310043
- Güdel, M. 1994, Quiescent microwave emission from late-type stars. *ApJS*, 90, 743–751. doi:10.1086/191899
- Güdel, M. 1997, Are coronae of magnetically active stars heated by flares? *ApJ*, 480, L121. doi:10.1086/310628
- Güdel, M., & Benz, A. O. 1989, Broad-band spectrum of dMe star radio emission. *A&A*, 211, L5–L8
- Güdel, M., & Benz, A. O. 1993, X-ray/microwave relation of different types of active stars. *ApJ*, 405, L63–L66. doi:10.1086/186766
- Güdel, M., Schmitt, J. H. M. M., Benz, A. O., & Elias, N. M., II 1995, The corona of the young solar analog EK Draconis. *A&A*, 301, 201
- Güdel, M., Benz, A. O., Schmitt, J. H. M. M., & Skinner, S. L. 1996, The Neupert effect in active

- stellar coronae: chromospheric evaporation and coronal heating in the dMe flare star binary UV Ceti. *ApJ*, 471, 1002. doi:10.1086/178027
- Guinan, E. F., & Engle, S. G. 2009, The Sun in time: age, rotation, and magnetic activity of the Sun and solar-type stars and effects on hosted planets. *IAUS*, 258, 395–408. doi:10.1017/S1743921309032050
- Haisch, B. M., & Simon, T. 1982, Further observational evidence for a coronal boundary line in the cool star region of the H-R diagram. *ApJ*, 263, 252–258. doi:10.1086/160499
- Haisch, B. M., Slee, O. B., Siegman, B. C., et al. 1981, Simultaneous X-ray, ultraviolet, optical, and radio observations of the flare star Proxima Centauri. *ApJ*, 245, 1009–1017. doi:10.1086/158878
- Haisch, B., Strong, K. T., & Rodonò, M. 1991, Flares on the Sun and other stars. *ARA&A*, 29, 275–324. doi:10.1146/annurev.aa.29.090191.001423
- Hale, G. E., & Ellerman, F. 1904, Calcium and hydrogen flocculi. *ApJ*, 19, 41. doi:10.1086/141083
- Hall, J. C., Henry, G. W., Lockwood, G. W., Skiff, B. A., & Saar, S. H. 2009, The activity and variability of the Sun and Sun-like stars. II. Contemporaneous photometry and spectroscopy of bright solar analogs. *AJ*, 138, 312–322. doi:10.1088/0004-6256/138/1/312
- Hansteen, V. H., Hara, H., De Pontieu, B., & Carlsson, M. 2010, On redshifts and blueshifts in the transition region and corona. *ApJ*, 718, 1070–1078. doi:10.1088/0004-637X/718/2/1070
- Harper, G. M. 2010, Betelgeuse: a case study of an inhomogeneous extended atmosphere. *ASPC*, 425, 152
- Harper, G. M., & Brown, A. 2006, Electron density and turbulence gradients within the extended atmosphere of the M supergiant betelgeuse (α orionis). *ApJ*, 646, 1179–1202. doi:10.1086/505073
- Hartmann, L., Soderblom, D. R., Noyes, R. W., Burnham, N., & Vaughan, A. H. 1984, An analysis of the Vaughan-Preston survey of chromospheric emission. *ApJ*, 276, 254–265. doi:10.1086/161609
- Hawley, S. L., Fisher, G. H., Simon, T., et al. 1995, Simultaneous extreme-ultraviolet explorer and optical observations of AD Leonis: evidence for large coronal loops and the Neupert effect in stellar flares. *ApJ*, 453, 464. doi:10.1086/176408
- Hempelmann, A., Robrade, J., Schmitt, J. H. M. M., Favata, F., Baliunas, S. L., & Hall, J. C. 2006, Coronal activity cycles in 61 Cygni. *A&A*, 460, 261–267. doi:10.1051/0004-6361:20065459
- Henry, T. J., Soderblom, D. R., Donahue, R. A., & Baliunas, S. L. 1996, A survey of Can H and K chromospheric emission in southern solar-type stars. *AJ*, 111, 439. doi:10.1086/117796
- Huber, K. F., Czesla, S., Wolter, U., & Schmitt, J. H. M. 2010, Planetary eclipse mapping of CoRoT-2a. Evolution, differential rotation, and spot migration. *A&A*, 514, A39. doi:10.1051/0004-6361/200913914
- Hudson, H. S. 1991, Solar flares, microflares, nanoflares, and coronal heating. *SoPh*, 133, 357–369. doi:10.1007/BF00149894
- Hussain, G. A. J., Brickhouse, N. S., Dupree, A. K., et al. 2005, Inferring coronal structure from X-ray light curves and Doppler shifts: a chandra study of AB Doradus. *ApJ*, 621, 999–1008. doi:10.1086/427647
- Järvinen, S. P., Berdyugina, S. V., Tuominen, I., Cutispoto, G., & Bos, M. 2005, Magnetic activity in the young solar analog AB Dor. Active longitudes and cycles from long-term photometry. *A&A*, 432, 657–664. doi:10.1051/0004-6361:20041998
- Jardine, M. 2004, Coronal stripping in supersaturated stars. *A&A*, 414, L5–L8. doi:10.1051/0004-6361:20031723
- Jeffries, R. D., Jackson, R. J., Briggs, K. R., Evans, P. A., & Pye, J. P. 2011, Investigating coronal saturation and supersaturation in fast-rotating M-dwarf stars. *MNRAS*, 411, 2099–2112. doi:10.1111/j.1365-2966.2010.17848.x
- Joy, A. H., & Humason, M. L. 1949, Observations of the faint dwarf star L 726–8. *PASP*, 61, 133–134. doi:10.1086/126150
- Judge, P. G., Carlsson M., & Stein R. F. 2003, On the origin of the basal emission from stellar atmospheres: analysis of solar C II lines. *ApJ*, 597, 1158–1177. doi:10.1086/381222
- Kashyap, V. L., Drake, J. J., & Saar, S. H. 2008, Extrasolar giant planets and X-ray activity. *ApJ*, 687, 1339–1354. doi:10.1086/591922
- Krivova, N. A., Solanki, S. K., Wenzlet, T., Podlipnik B. 2006, Reconstruction of solar UV irradiance since 1974, *JGRD*, 114, D00I04, doi:10.1029/2009JD012375
- Lammer, H. 2009, Ion erosion from the EUV heated and extended nitrogen atmosphere of early earth. European planetary science congress conference, 447
- Lammer, H., Eybl, V., Kislyakova, K. G., et al. 2011, UV transit observations of EUV-heated expanded thermospheres of earth-like exoplanets around M-stars: testing atmosphere evolution scenarios. *Ap&SS*, 335, 39–50. doi:10.1007/s10509-011-0646-5
- Lanza, A. F. 2009, Stellar coronal magnetic fields and star-planet interaction. *A&A*, 505, 339–350. doi:10.1051/0004-6361/200912367

- Lanza, A. F. 2010, *Stellar magnetic cycles*. IAUS, 264, 120–129. doi:10.1017/S1743921309992523
- Lanza, A. F. 2011, Searching for star-planet magnetic interaction in CoRoT observations. *Ap&SS*, 336, 303–313. doi:10.1007/s10509-011-0878-4
- Lanza, A. F., Catalano, S., Cutispoto, G., Pagano, I., & Rodonò, M. 1998, Long-term starspot evolution, activity cycle and orbital period variation of AR Lacertae. *A&A*, 332, 541–560
- Lanza, A. F., Rodonò, M., Pagano, I., Barge, P., & Llebaria, A. 2003, Modelling the rotational modulation of the Sun as a star. *A&A*, 403, 1135–1149. doi:10.1051/0004-6361:20030401
- Lanza, A. F., Rodonò, M., & Pagano, I. 2004, Multi-band modelling of the Sun as a variable star from VIRGO/SoHO data. *A&A*, 425, 707–717. doi:10.1051/0004-6361:20047028
- Lanza, A. F., Messina, S., Pagano, I., & Rodonò, M. 2006, A model to simulate synthetic light curves of solar-like stars. *AN*, 327, 21–28. doi:10.1002/asna.200510479
- Lanza, A. F., Bonomo, A. S., & Rodonò, M. 2007, Comparing different approaches to model the rotational modulation of the Sun as a star. *A&A*, 464, 741–751. doi:10.1051/0004-6361:20066575
- Lanza, A. F., Aigrain, S., Messina, S., et al. 2009a, Photospheric activity and rotation of the planet-hosting star CoRoT-4a. *A&A*, 506, 255–262. doi:10.1051/0004-6361/200811487
- Lanza, A. F., Pagano, I., Leto, G., et al. 2009b, Magnetic activity in the photosphere of CoRoT-Exo-2a. Active longitudes and short-term spot cycle in a young Sun-like star. *A&A*, 493, 193–200. doi:10.1051/0004-6361:200810591
- Lanza, A. F., Bonomo, A. S., Moutou, C., et al. 2010, Photospheric activity, rotation, and radial velocity variations of the planet-hosting star CoRoT-7. *A&A*, 520, A53. doi:10.1051/0004-6361/201014403
- Lanza, A. F., Boisse, I., Bouchy, F., Bonomo, A. S., & Moutou, C. 2011, Deriving the radial-velocity variations induced by stellar activity from high-precision photometry. Test on HD 189733 with simultaneous MOST/SOPHIE data. *A&A*, 533, A44. doi:10.1051/0004-6361/201117270
- Leto, G., Pagano, I., Buemi, C. S., & Rodonò, M. 1997, Long-term photometric flares patrol (1967–1977) on EV Lacertae: a clear evidence of a longitude concentrated flaring activity in 1970. *A&A*, 327, 1114–1122
- Leto, G., Pagano, I., Linsky, J. L., Rodonò, M., & Umana, G. 2000, VLA observation of dMe stars. *A&A*, 359, 1035–1041
- Lim, J., Carilli, C. L., White, S. M., Beasley, A. J., & Marson, R. G. 1998, Large convection cells as the source of Betelgeuse's extended atmosphere. *Nature*, 392, 575–577. doi:10.1038/33352
- Linsky, J. L. 1999, The Wilson-Bappu relation between Ca II emission and stellar luminosities. *ApJ*, 525, 776
- Linsky, J. L. 2000, Stellar flares: how common? How important? *AIPC*, 522, 389–399. doi:10.1063/1.1291741
- Linsky, J. L., & Avrett, E. H. 1970, The solar H and K lines. *PASP*, 82, 169. doi:10.1086/128904
- Linsky, J. L., & Haisch, B. M. 1979, Outer atmospheres of cool stars. I – The sharp division into solar-type and non-solar-type stars. *ApJ*, 229, L27–L32. doi:10.1086/182924
- Linsky, J. L., & Wood, B. E. 1994, High-velocity plasma in the transition region of AU microscopii: evidence for magnetic reconnection and saturated heating during quiescent and flaring conditions. *ApJ*, 430, 342–350. doi:10.1086/174409
- Linsky, J. L., Wood, B. E., Judge, P., Brown, A., Andrusis, C., & Ayres, T. R. 1995, The transition regions of Capella. *ApJ*, 442, 381–400. doi:10.1086/175447
- Linsky, J. L., Bushinsky, R., Ayres, T., Fontenla, J., & France, K. 2012, Far-ultraviolet continuum emission: applying this diagnostic to the chromospheres of solar-mass stars. *ApJ*, 745, 25. doi:10.1088/0004-637X/745/1/25
- Lockwood, G. W., Skiff, B. A., Henry, G. W., Henry, S., Radick, R. R., Baliunas, S. L., Donahue, R. A., & Soon, W. 2007, Patterns of photometric and chromospheric variation among Sun-like stars: a 20 year perspective. *ApJS*, 171, 260–303. doi:10.1086/516752
- Lou, Y.-Q. 2000, Rossby-type wave-induced periodicities in flare activities and sunspot areas or groups during solar maxima. *ApJ*, 540, 1102–1108. doi:10.1086/309387
- Lundin, R., Lammer, H., & Ribas, I. 2007, Planetary magnetic fields and solar forcing: implications for atmospheric evolution. *SSR*, 129, 245–278. doi:10.1007/s11214-007-9176-4
- Maggio, A., Sanz-Forcada, J., & Scelsi, L. 2011, Photospheric and coronal abundances in solar-type stars: the peculiar case of τ Bootis. *A&A*, 527, A144. doi:10.1051/0004-6361/201015587
- Marcy, G. W., & Chen, G. H. 1992, The rotation of M dwarfs. *ApJ*, 390, 550–559. doi:10.1086/171305
- Marino, A., Micela, G., Peres, G., & Sciortino, S. 2003, X-ray rotational modulation of a super-saturated star in IC 2391. *A&A*, 407, L63–L66. doi:10.1051/0004-6361:20031053

- Massi, M., Neidhöfer, J., Carpentier, Y., & Ros, E. 2005, Discovery of solar Rieger periodicities in another star. *A&A*, 435, L1–L4. doi:10.1051/0004-6361:200500097
- Mauas, P. J. D., Falchi, A., Pasquini, L., & Pallavicini, R. 1997, Chromospheric models of dwarf M stars. *A&A*, 326, 249–256
- Messina, S. 2008, Long-term magnetic activity in close binary systems. I. Patterns of color variations. *A&A*, 480, 495–508. doi:10.1051/0004-6361:20078932
- Messina, S., & Guinan, E. F. 2002, Magnetic activity of six young solar analogues I. Starspot cycles from long-term photometry. *A&A*, 393, 225–237. doi:10.1051/0004-6361:20021000
- Messina, S., & Guinan, E. F. 2003, Magnetic activity of six young solar analogues II. Surface differential rotation from long-term photometry. *A&A*, 409, 1017–1030. doi:10.1051/0004-6361:20031161
- Meunier, N., Lagrange, A.-M., & Desort, M. 2010, Reconstructing the solar integrated radial velocity using MDI/SOHO. *A&A*, 519, A66. doi:10.1051/0004-6361/201014199
- Mewe, R., Heise, J., Gronenschild, E. H. B. M., Brinkman, A. C., Schrijver, J., & den Boggende, A. J. F. 1975, Detection of X-ray emission from stellar coronae with ANS. *ApJ*, 202, L67–L71. doi:10.1086/181983
- Mewe, R., Raassen, A. J. J., Drake, J. J., Kaastra, J. S., van der Meer, R. L. J., & Porquet, D. 2001, CHANDRA-LETGS X-ray observations of Capella. Temperature, density and abundance diagnostics. *A&A*, 368, 888–900. doi:10.1051/0004-6361:20010026
- Moutou, C., Pont, F., Barge, P., et al. 2005, Comparative blind test of five planetary transit detection algorithms on realistic synthetic light curves. *A&A*, 437, 355–368. doi:10.1051/0004-6361:20042334
- Mutel, R. L., Molnar, L. A., Waltman, E. B., & Ghigo, F. D. 1998, Radio Emission from Algol. I. Coronal Geometry and Emission Mechanisms Determined from VLBA and Green Bank Interferometer Observations. *ApJ*, 507, 371–383. doi:10.1086/306311
- Mutel, R. L., Lestrade, J. F., Preston, R. A., & Phillips, R. B. 1985, Dual polarization VLBI observations of stellar binary systems at 5 GHz. *ApJ*, 289, 262–268. doi:10.1086/162886
- Narain, U., & Ulmschneider, P. 1990, Chromospheric and coronal heating mechanisms. *SSR*, 54, 377–445. doi:10.1007/BF00177801
- Neff, J. E., O’Neal, D., & Saar, S. H. 1995, Absolute Measurements of Starspot Area and Temperature: II Pegasi in 1989 October. *ApJ*, 452, 879. doi:10.1086/176356
- Neff, J. E., Walter, F. M., Rodonò, M., & Linsky, J. L. 1989, Rotational modulation and flares on RS Canum Venaticorum and BY Draconis stars. XI – Ultraviolet spectral images of AR Lacertae in September 1985. *A&A*, 215, 79–91
- Ness, J.-U., Mewe, R., Schmitt, J. H. M. M., & Raassen, A. J. J. 2002a, Influence of radiation fields on the density diagnostics Chandra-LETGS observations of Algol and Procyon. *ASPC*, 277, 545
- Ness, J.-U., Schmitt, J. H. M. M., Burwitz, V., Mewe, R., Raassen, A. J. J., van der Meer, R. L. J., Predehl, P., & Brinkman, A. C. 2002b, Coronal density diagnostics with helium-like triplets: CHANDRA-LETGS observations of Algol, Capella, Procyon, epsilon Eri, alpha Cen A&B, UX Ari, AD Leo, YY Gem, and HR 1099. *A&A*, 394, 911–926. doi:10.1051/0004-6361:20021146
- Ness, J.-U., Güdel, M., Schmitt, J. H. M. M., Audard, M., & Telleschi, A. 2004, On the sizes of stellar X-ray coronae. *A&A*, 427, 667–683. doi:10.1051/0004-6361:20040504
- Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984, Rotation, convection, and magnetic activity in lower main-sequence stars. *ApJ*, 279, 763–777. doi:10.1086/161945
- O’Neal, D., Saar, S. H., & Neff, J. E. 1996, Measurements of starspot area and temperature on five active, evolved stars. *ApJ*, 463, 766. doi:10.1086/177288
- O’Neal, D., Neff, J. E., & Saar, S. H. 1998, Measurements of starspot parameters on active stars using molecular bands in echelle spectra. *ApJ*, 507, 919–937. doi:10.1086/306340
- Orlando, S., Peres, G., & Reale, F. 2000, The Sun as an X-ray star. I. Deriving the emission measure distribution versus temperature of the whole solar corona from the Yohkoh/soft X-ray telescope data. *ApJ*, 528, 524–536. doi:10.1086/308137
- Orlando, S., Peres, G., & Reale, F. 2001, The Sun as an X-ray star. IV. The contribution of different regions of the corona to its X-ray spectrum. *ApJ*, 560, 499–513. doi:10.1086/322333
- Osten, R. A., Brown, A., Ayres, T. R., Linsky, J. L., Drake, S. A., Gagné, M., & Stern, R. A. 2000, Radio, X-ray, and extreme-ultraviolet coronal variability of the short-period RS Canum venaticorum σ^2 coronae Borealis. *ApJ*, 544, 953–976. doi:10.1086/317249
- Osten, R. A., Ayres, T. R., Brown, A., Linsky, J. L., & Krishnamurthi, A. 2003, Chandra, extreme ultraviolet explorer, and very large array observations of the active binary

- system σ^2 corone Borealis. *ApJ*, 582, 1073–1101. doi:10.1086/344797
- Osten, R. A., Brown, A., Ayres, T. R., et al. 2004, A multiwavelength perspective of flares on HR 1099: 4 years of coordinated campaigns. *ApJS*, 153, 317–362. doi:10.1086/420770
- Osten, R. A., Hawley, S. L., Allred, J. C., Johns-Krull, C. M., & Roark, C. 2005, From radio to X-ray: flares on the dMe flare star EV Lacertae. *ApJ*, 621, 398–416. doi:10.1086/427275
- Pace, G., Pasquini, L., & Ortolani, S. 2003, The Wilson-Bappu effect: A tool to determine stellar distances. *A&A*, 401, 997–1007. doi:10.1051/0004-6361:20030163
- Pagano, I., Linsky, J. L., Carkner, L., Robinson, R. D., Woodgate, B., & Timothy, G. 2000, HST/STIS echelle spectra of the dM1e star AU microscopii outside of flares. *ApJ*, 532, 497–513. doi:10.1086/308559
- Pagano, I., Rodonò, M., Linsky, J. L., Neff, J. E., Walter, F. M., Kovári, Z., & Matthews, L. D. 2001, Spectral imaging maps of AR Lacertae. I. Results from IUE observations in 1994 October. *A&A*, 365, 128–143. doi:10.1051/0004-6361:20000025
- Pagano, I., Linsky, J. L., Valenti, J., & Duncan, D. K. 2004, HST/STIS high resolution echelle spectra of α Centauri A (G2 V). *A&A*, 415, 331–348. doi:10.1051/0004-6361:20034002
- Pagano, I., De Martino, C., Lanza, A.F., Spadaro, D., & Linsky, J.L. 2006, Chromospheric and transition regions features in solar like stars. *ESASP*, 617, 91.
- Pagano, I., Lanza, A. F., Leto, G., Messina, S., Barge, P., & Baglin, A. 2009, CoRoT-2a magnetic activity: hints for possible star-planet interaction. *EM&P*, 105, 373–378. doi:10.1007/s11038-009-9301-3
- Pallavicini, R., Tagliaferri, G., & Stella, L. 1990, X-ray emission from solar neighbourhood flare stars – A comprehensive survey of EXOSAT results. *A&A*, 228, 403–425
- Parker, E. N. 1977, The origin of solar activity. *ARA&A*, 15, 45–68. doi:10.1146/annurev.aa.15.090177.000401
- Parker, E. N. 1988, Nanoflares and the solar X-ray corona. *ApJ*, 330, 474–479. doi:10.1086/166485
- Pasquini, L., Brocato, E., & Pallavicini, R. 1990, Chromospheric activity of evolved late-type stars – Chromospheric activity in evolved stars. *A&A*, 234, 277–283
- Patten, B. M., & Simon, T. 1996, The evolution of rotation and activity in young open clusters: IC 2391. *ApJS*, 106, 489. doi:10.1086/192346
- Pestalozzi, M. R., Benz, A. O., Conway, J. E., & Güdel, M. 2000, VLBI observations of two single dMe stars: spatial resolution and astrometry. *A&A*, 353, 569–574
- Peter, H., Gudiksen, B. V., & Nordlund, Å. 2006, Forward Modeling of the Corona of the Sun and Solar-like Stars: From a Three-dimensional Magnetohydrodynamic Model to Synthetic Extreme-Ultraviolet Spectra. *ApJ*, 638, 1086–1100. doi:10.1086/499117
- Peter, H. 2006, First high spectral resolution VUV full-Sun spectrum compared to cool stars. *A&A*, 449, 759–768. doi:10.1051/0004-6361:20054337
- Peter, H., & Judge, P. G. 1999, On the Doppler shifts of solar ultraviolet emission lines. *ApJ*, 522, 1148–1166. doi:10.1086/307672
- Petrov, P. P. 2003, T Tauri stars. *Ap*, 46, 506–529. doi:10.1023/B:ASYS.0000003267.35552.f7
- Pettersen, B. R. 1991, The nearby flare stars. *MmSAI*, 62, 217–242.
- Pillitteri, I., Micela, G., Sciortino, S., & Favata, F. 2003, The X-ray luminosity distributions of the high-metallicity open cluster Blanco 1. *A&A*, 399, 919–929. doi:10.1051/0004-6361:20021818
- Pitts, P. S., & Imhoff, C.L. 1989, A bibliography of IUE atlases and catalogues. *NASA IUE Newsletter*, 37, 1–30
- Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, The stellar activity-rotation relationship revisited: dependence of saturated and non-saturated X-ray emission regimes on stellar mass for late-type dwarfs. *A&A*, 397, 147–157. doi:10.1051/0004-6361:20021560
- Poppenhaeger, K., Robrade, J., & Schmitt, J. H. M. M. 2010, Coronal properties of planet-bearing stars. *A&A*, 515, A98. doi:10.1051/0004-6361/201014245
- Preibisch, T., & Zinnecker, H. 2002, X-ray properties of the young stellar and substellar objects in the IC 348 cluster: the Chandra view. *AJ*, 123, 1613–1628. doi:10.1086/338851
- Preibisch, T., McCaughrean, M. J., Grosso, N., et al. 2005, X-ray emission from young brown dwarfs in the Orion Nebula cluster. *ApJS*, 160, 582–593. doi:10.1086/432098
- Radick, R. R., Lockwood, G. W., Skiff, B. A., & Baliunas, S. L. 1998, Patterns of variation among Sun-like stars. *ApJS*, 118, 239–258. doi:10.1086/313135
- Randich, S. 2000, Coronal activity among open cluster stars. *ASPC*, 198, 401
- Reale, F., Peres, G., & Serio, S. 1996, Radiatively-driven downdrafts and redshifts in transition region lines. I. Reference model. *A&A*, 316, 215–228
- Redfield, S., Ayres, T. R., Linsky, J. L., Ake, T. B., Dupree, A. K., Robinson, R. D., & Young, P. R.

- 2003, A far ultraviolet spectroscopic explorer survey of coronal forbidden lines in late-type stars. *ApJ*, 585, 993–1006. doi:10.1086/346129
- Redfield, S., Linsky, J. L., Ake, T. B., Ayres, T. R., Dupree, A. K., Robinson, R. D., Wood, B. E., & Young, P. R. 2002, A far ultraviolet spectroscopic explorer survey of late-type dwarf stars. *ApJ*, 581, 626–653. doi:10.1086/344153
- Richards, M. T., & Rosolowsky, E. W. 1998, Doppler tomography of ultraviolet spectra of HR 1099. *ASPC*, 154, 2038
- Robinson, R. D., Linsky, J. L., Woodgate, B. E., & Timothy, J. G. 2001, Far-ultraviolet observations of flares on the dM0e star AU microscopii. *ApJ*, 554, 368–382. doi:10.1086/321379
- Robinson, R. D., Wheatley, J. M., Welsh, B. Y., et al. 2005, GALEX observations of an energetic ultraviolet flare on the dM4e star GJ 3685A. *ApJ*, 633, 447–451. doi:10.1086/444608
- Robrade, J., & Schmitt, J. H. M. M. 2006, XMM-Newton X-ray spectroscopy of classical T Tauri stars. *A&A*, 449, 737–747. doi:10.1051/0004-6361:20054247
- Robrade, J., Schmitt, J. H. M. M., & Favata, F. 2005, X-rays from α Centauri – The darkening of the solar twin. *A&A*, 442, 315–321. doi:10.1051/0004-6361:20053314
- Robrade, J., Schmitt, J. H. M. M., & Hempelmann, A. 2007, X-ray activity cycles in stellar coronae. *MmSAI*, 78, 311
- Rodonò, M. 1980, Stellar activity. *MmSAI*, 51, 623–658.
- Rodonò, M. 1992, Automated photometric telescopes in the study of active stars (invited). *LNP*, 397, 201. doi:10.1007/3-540-55310-X
- Rodonò, M., & Cutispoto, G. 1994, A long-term program of monitoring active stars with the Catania automated photometric telescope. *ASPC*, 64, 459
- Rodonò, M., Cutispoto, G., Pazzani, V., et al. 1986, Rotational modulation and flares on RS CVn and BY Dra-type stars. I – Photometry and SPOT models for BY Dra, AU Mic, AR Lac, II Peg and V 711 Tau (= HR 1099). *A&A*, 165, 135–156
- Rodonò, M., Houdebine, E. R., Catalano, S., et al. 1988, Simultaneous multiwavelength observations of an intense flare on Ad-Leonis. Solo and small firm conference, 53
- Rodonò, M., Pagano, I., Leto, G., Walter, F., Catalano, S., Cutispoto, G., & Umana, G. 1999, BeppoSAX observations of AR Lacertae. *A&A*, 346, 811–818
- Rodonò, M., Messina, S., Lanza, A. F., Cutispoto, G., & Teriaca, L. 2000, The magnetic activity cycle of II Pegasi: results from twenty-five years of wide-band photometry. *A&A*, 358, 624–638
- Rubenstein, E. P., & Schaefer, B. E. 2000, Are superflares on solar analogues caused by extrasolar planets? *ApJ*, 529, 1031–1033. doi:10.1086/308326
- Rosner, R., Musielak, Z. E., Cattaneo, F., Moore, R. L., & Suess, S. T. 1995, On the origin of ‘dividing lines’ for late-type giants and supergiants. *ApJ*, 442, L25–L28. doi:10.1086/187807
- Saar, S. H. 2006, Maunder minimum dwarfs: defined out of existence? *SPD*, 38, 240
- Saar, S. H., & Baliunas, S. L. 1992, Recent advances in stellar cycle research. *ASPC*, 27, 150–167
- Sanz-Forcada, J., Favata, F., & Micela, G. 2004, Coronal versus photospheric abundances of stars with different activity levels. *A&A*, 416, 281–290. doi:10.1051/0004-6361:20034466
- Sanz-Forcada, J., Ribas, I., Micela, G., Pollock, A. M. T., García-Álvarez, D., Solano, E., & Eiroa, C. 2010, A scenario of planet erosion by coronal radiation. *A&A*, 511, L8. doi:10.1051/0004-6361/200913670
- Sanz-Forcada, J., Micela, G., Ribas, I., Pollock, A. M. T., Eiroa, C., Velasco, A., Solano, E., & García-Álvarez, D. 2011a, Estimation of the XUV radiation onto close planets and their evaporation. *A&A*, 532, A6. doi:10.1051/0004-6361/201116594
- Sanz-Forcada, J., Ribas, I., Micela, G., Pollock, A., García-Álvarez, D., Solano, E., & Eiroa, C. 2011b, Planet erosion by coronal radiation. *hsa6*. conference, 607–612
- Scalo, J., Kaltenecker, L., Segura, A. G., et al. 2007, M Stars as targets for terrestrial exoplanet searches and biosignature detection. *Astrobiology*, 7, 85–166. doi:10.1089/ast.2006.0125
- Schaefer, B. E., King, J. R., & Deliyannis, C. P. 2000, Superflares on ordinary solar-type stars. *ApJ*, 529, 1026–1030. doi:10.1086/308325
- Schatzman, E. 1962, A theory of the role of magnetic activity during star formation. *AnAp*, 25, 18
- Schmitt, J. H. M. M., & Liefke, C. 2004, NEXXUS: a comprehensive ROSAT survey of coronal X-ray emission among nearby solar-like stars. *A&A*, 417, 651–665. doi:10.1051/0004-6361:20030495
- Schmitt, J. H. M. M., Collura, A., Sciortino, S., Vaiana, G. S., Harnden, F. R., Jr., & Rosner, R. 1990, Einstein observatory coronal temperatures of late-type stars. *ApJ*, 365, 704–728. doi:10.1086/169525
- Schmitt, J. H. M. M., Robrade, J., Ness, J.-U., Favata, F., & Stelzer, B. 2005, X-rays from accretion shocks in T Tauri stars: the case of BP Tau. *A&A*, 432, L35–L38. doi:10.1051/0004-6361:200500014

- Schrijver C. J. 1987, Magnetic structure in cool stars. *A&A*, 172, 111–123
- Schrijver, C. J., & Title, A. M. 2001, On the formation of polar spots in sun-like stars. *ApJ*, 551, 1099–1106. doi:10.1086/320237
- Schwarzschild, M. 1948, On noise arising from the solar granulation. *ApJ*, 107, 1. doi:10.1086/144983
- Shakhovskaya, N. I. 1995, Flare activity among nearby stars. *LNP*, 454, 61. doi:10.1007/3-540-60057-4_233
- Shkolnik, E., Walker, G. A. H., & Bohlender, D. A. 2003, Evidence for planet-induced chromospheric activity on HD 179949. *ApJ*, 597, 1092–1096. doi:10.1086/378583
- Shkolnik, E., Walker, G. A. H., Bohlender, D. A., Gu P.-G., & Kürster, M. 2005, Hot Jupiters and hot spots: the short- and long-term chromospheric activity on stars with giant planets. *ApJ*, 622, 1075–1090. doi:10.1086/428037
- Shkolnik, E., Bohlender, D. A., Walker, G. A. H., & Collier Cameron A. 2008, The on/off nature of star-planet interactions. *ApJ*, 676, 628–638. doi:10.1086/527351
- Simon, T., Ayres, T. R., Redfield, S., & Linsky, J. L. 2002, Limits on chromospheres and convection among the main-sequence stars. *ApJ*, 579, 800–809. doi:10.1086/342941
- Skumanich, A. 1972, Time scales for Ca II emission decay, rotational braking, and lithium depletion. *ApJ*, 171, 565. doi:10.1086/151310
- Solanki, S. K., Motamen, S., & Keppens, R. 1997, Polar spots and stellar spindown: is dynamo saturation needed? *A&A*, 324, 943–948
- Spadaro, D., Lanza, A. F., Karpen, J. T., & Antiochos, S. K. 2006, A transient heating model for the structure and dynamics of the solar transition region. *ApJ*, 642, 579–583. doi:10.1086/500963
- Spiegel, E. A., & Zahn J.-P. 1992, The solar tachocline. *A&A*, 265, 106–114
- Stepien, K., Schmitt, J. H. M. M., & Voges, W. 2001, ROSAT all-sky survey of W Ursae Majoris stars and the problem of supersaturation. *A&A*, 370, 157–169. doi:10.1051/0004-6361:20010197
- Stassun, K. G., van den Berg, M., Feigelson, E., & Flaccomio, E. 2006, A simultaneous optical and X-ray variability study of the Orion Nebula cluster. I. Incidence of time-correlated X-ray/optical variations. *ApJ*, 649, 914–926. doi:10.1086/506422
- Stewart, R. T., Slee, O. B., White, G. L., Budding, E., Coates, D. W., Thompson, K., & Bunton, J. D. 1989, Radio emission from EA eclipsing binaries – Evidence for kilogauss surface fields on both early-type and late-type stars. *ApJ*, 342, 463–466. doi:10.1086/167605
- Strassmeier, K. G. 2009, Starspots. *A&AR*, 17, 251–308. doi:10.1007/s00159-009-0020-6
- Teriaca, L., Banerjee, D., & Doyle, J. G. 1999, SUMER observations of Doppler shift in the quiet Sun and in an active region. *A&A*, 349, 636–648
- Testa, P., Drake, J. J., & Peres, G. 2004, The density of coronal plasma in active stellar coronae. *ApJ*, 617, 508–530. doi:10.1086/422355
- Thomas, J. H., & Weiss, N. O. 2008, Sunspots and starspots. Cambridge University Press, New York. ISBN:9780521860031
- Trigilio, C., Buemi, C. S., Umana, G., Rodonò, M., Leto, P., Beasley, A. J., & Pagano, I. 2001, The radio corona of AR Lacertae. *A&A*, 373, 181–189. doi:10.1051/0004-6361:20010566
- Umana, G., Catalano, S., Rodonò, M., & Gibson, D. M. 1989, Radio emission from selected algol systems. *SSR*, 50, 370–370. doi:10.1007/BF00215973
- Umana, G., Trigilio, C., Tumino, M., Catalano, S., & Rodonò, M. 1995, Radio flares on HR 1099 (= V 711 Tauri). *A&A*, 298, 143
- Vaiana, G. S., Cassinelli, J. P., Fabbiano, G., et al. 1981, Results from an extensive Einstein stellar survey. *ApJ*, 245, 163–182. doi:10.1086/158797
- van den Oord, G. H. J., Doyle, J. G., Rodonò, M., et al. 1996, Flare energetics: analysis of a large flare on YZ Canis Minoris observed simultaneously in the ultraviolet, optical and radio. *A&A*, 310, 908–922
- van den Oord, G. H. J., Mewe, R., & Brinkman, A. C. 1988, An EXOSAT observation of an X-ray flare and quiescent emission from the RS CVn binary sigma2 CrB. *A&A*, 205, 181–196
- Vaughan, A. H. 1983, The Mount Wilson program for stellar activity cycles. *IAUS*, 102, 113–131
- Vaughan, A. H., & Preston, G. W. 1980, A survey of chromospheric Ca II H and K emission in field stars of the solar neighborhood. *PASP*, 92, 385–391. doi:10.1086/130683
- Vernazza, J. E., Avrett, E. H., & Loeser, R. 1981, Structure of the solar chromosphere. III – Models of the EUV brightness components of the quiet-sun. *ApJS*, 45, 635–725. doi:10.1086/190731
- Vida, K., Kovári, Z., Švanda, M., Oláh, K., Strassmeier, K. G., & Bartus, J. 2007, Anti-solar differential rotation and surface flow pattern on UZ Librae. *AN*, 328, 1078. doi:10.1002/asna.200710879
- Vilhu, O., & Walter, F. M. 1987, Chromospheric-coronal activity at saturated levels. *ApJ*, 321, 958–966. doi:10.1086/165689
- Vilhu, O., Muhli, P., Huovelin, J., Hakala, P., Rucinski, S. M., & Collier Cameron, A. 1998, Ultraviolet spectroscopy of AB Doradus with the Hubble

- Space Telescope: impulsive flares and bimodal profiles of C IV λ 1549 in a young star. *AJ*, 115, 1610–1616. doi:10.1086/300268
- Vogt, S. S., & Penrod, G. D. 1983, Doppler imaging of spotted stars – Application to the RS Canum venaticorum star HR 1099. *PASP*, 95, 565–576. doi:10.1086/131208
- Vogt, S. S., Hatzes, A. P., Misch, A. A., & Kürster, M. 1999, Doppler imagery of the spotted RS Canum venaticorum star HR 1099 (V711 Tauri) from 1981 to 1992. *ApJS*, 121, 547–589. doi:10.1086/313195
- Walter, F. M. 1996a, EUVE observations of AR Lacertae: the differential emission measure and evidence for extended prominences. Australian education union conference, 129
- Walter, F. M. 1996b, Stellar plagues (review). *IAUS*, 176, 355
- Walter, F. M., & Kim, J. S. 2001, The chromosphere of V471 Tauri resolved (CD-ROM directory: contributors/walter). *ASPC*, 223, 1073
- Walter, F., Charles, P., & Bowyer, S. 1978, Discovery of quiescent X-ray emission from HR1099 and RS CVn. *Nature*, 274, 569. doi:10.1038/274569b0
- Walter, F. M., Cash, W., Charles, P. A., & Bowyer, C. S. 1980, X-rays from RS Canum venaticorum systems – A HEAO 1 survey and the development of a coronal model. *ApJ*, 236, 212–218. doi:10.1086/157735
- Walter, F. M., Neff, J. E., Gibson, D. M., Linsky, J. L., Rodonò, M., Gary, D. E., & Butler, C. J. 1987, Rotational modulation and flares on RS CVn and BY Dra stars. IV – The spatially resolved chromosphere of AR Lacertae. *A&A*, 186, 241–254
- Walker, G. A. H., Croll, B., Kuschnig, R., et al. 2007, The differential rotation of κ^1 Ceti as observed by MOST. *ApJ*, 659, 1611–1622. doi:10.1086/511851
- Walker, G. A. H., Croll, B., Matthews, J. M., et al. 2008, MOST detects variability on τ Bootis A possibly induced by its planetary companion. *A&A*, 482, 691–697. doi:10.1051/0004-6361:20078952
- Welsh, B. Y., Wheatley, J., Browne, S. E., et al. 2006, GALEX high time-resolution ultraviolet observations of dMe flare events. *A&A*, 458, 921–930. doi:10.1051/0004-6361:20065304
- White, Stephen M., Jackson, Peter D., & Kundu, Mukul R. 1989, A VLA survey of nearby flare stars. *ApJSS*, 71, 895–904. doi:10.1086/191401
- White, S. M., Lim, J., & Kundu, M. R. 1994, Radio constraints on coronal models for dMe stars. *ApJ*, 433, 293–303. doi:10.1086/173727
- White, O. R., & Livingston, W. C. 1981, Solar luminosity variation. III – Calcium K variation from solar minimum to maximum in cycle 21. *ApJ*, 249, 798–816. doi:10.1086/159338
- White, N. E., & Marshall, F. E. 1983, An X-ray survey on nine algol systems. *ApJ*, 268, L117–L120. doi:10.1086/184040
- Wilson, O. C., & Vainu Bappu, M. K. 1957, H and K emission in late-type stars: dependence of line width on luminosity and related topics. *ApJ*, 125, 661. doi:10.1086/146339
- Wolk, S. J., Harnden, F. R., Jr., Flaccomio, E., Micela, G., Favata, F., Shang, H., & Feigelson, E. D. 2005, Stellar activity on the young Suns of orion: COUP observations of K5-7 pre-main-sequence stars. *ApJS*, 160, 423–449. doi:10.1086/432099
- Wood, B. E., Harper, G. M., Linsky, J. L., & Dempsey, R. C. 1996, Goddard High-Resolution Spectrograph Observations of Procyon and HR 1099. *ApJ*, 458, 761. doi:10.1086/176857
- Wood, B. E. 2004, Astrospheres and solar-like stellar winds. *LRSF*, 1, 2
- Wood, B. E., & Linsky, J. L. 2010, Resolving the χ Boo binary with Chandra, and revealing the spectral type dependence of the coronal “FIP effect”. *ApJ*, 717, 1279–1290. doi:10.1088/0004-637X/717/2/1279
- Wood, B. E., Linsky, J. L., & Ayres, T. R. 1997, Evaluating possible heating mechanisms using the transition region line profiles of late-type stars. *ApJ*, 478, 745. doi:10.1086/303821
- Wood, B. E., Müller, H.-R., Zank, G. P., & Linsky, J. L. 2002, Measured mass-loss rates of solar-like stars as a function of age and activity. *ApJ*, 574, 412–425. doi:10.1086/340797
- Wood, B. E., Müller, H.-R., Zank, G. P., Linsky, J. L., & Redfield, S. 2005, New mass-loss measurements from astrospheric Ly-alpha; absorption. *ApJ*, 628, L143–L146. doi:10.1086/432716
- Wright, J. T. 2004, Do we know of any Maunder minimum stars? *AJ*, 128, 1273–1278. doi:10.1086/423221
- Zacharias, P., Peter, H., & Bingert, S. 2011, Investigation of mass flows in the transition region and corona in a three-dimensional numerical model approach. *A&A*, 531, A97. doi:10.1051/0004-6361/201016047

11 White Dwarf Stars

Detlev Koester

Institut für Theoretische Physik und Astrophysik, Universität Kiel,
Kiel, Germany

1	<i>Introduction</i>	561
1.1	A Few Historical Milestones	562
2	<i>Observations and Classification</i>	563
2.1	Classification and Spectral Types	564
2.2	Photometry	566
2.3	Parallaxes	566
3	<i>Stellar Parameters</i>	566
3.1	Effective Temperatures, Surface Gravities, and Abundances from Atmospheric Modeling	567
3.2	Determination of Masses and Radii	569
4	<i>Interior Structure and Evolution</i>	570
4.1	Degeneracy and the Equation of State	570
4.2	Interior Models and Mass–Radius Relation	572
4.3	Gravitational Contraction and Cooling	574
5	<i>Full Evolutionary Calculations</i>	576
5.1	Low Mass White Dwarfs	576
5.2	Origin and Evolution of Hydrogen-Deficient White Dwarfs	578
5.3	Other Evolutionary Channels	579
5.4	Evolution and the HRD	579
5.5	Mass–Radius Relation	582
5.6	Luminosity Function and Age of the Galactic Disk	584
5.7	Local Space Density and Birthrate	584
5.8	Initial–Final Mass Relation	585
6	<i>Physical Processes in the Outer Layers</i>	586
6.1	Gravitational Settling and Diffusion	587
6.2	Stellar Winds and Radiative Levitation	589
6.3	Convection and Dredge-Up	590
6.4	Accretion and Diffusion	591
6.5	Spectral Evolution of White Dwarfs: Current Status	593
6.5.1	The DB Gap	595
6.5.2	The DAO and DAB Stars	595
6.5.3	DB Stars with Hydrogen Traces: The DBA	596

6.5.4 Cool White Dwarfs with Metal Traces: DZ, DBZ, DAZ 596

6.5.5 The DQ Stars 596

6.5.6 Evidence for DA to DB Transformation by Convective Mixing? 597

6.5.7 Summary of Spectral Evolution 598

7 *Some Areas of Current Interest* 598

7.1 White Dwarfs with Debris Disks 598

7.2 Magnetic White Dwarfs 600

7.2.1 Origin of Magnetic Fields 602

7.3 Variable White Dwarfs 604

7.4 White Dwarfs in Open and Globular Clusters 605

7.4.1 White Dwarfs in Open Clusters 605

7.4.2 White Dwarfs in Globular Clusters 606

7.5 White Dwarfs in Binaries 607

7.5.1 Double Degenerates 608

8 *Suggestions for Further Reading and Useful Resources* 608

***Acknowledgments* 609**

***References* 610**

Abstract: White dwarfs are the final stage for more than 95% of all stars. Their population statistics and properties contain a wealth of information about the history of star formation in our galaxy, the ages of stellar systems, and the relation between original mass at birth and that of the final remnant. They are also interesting individually as laboratories for physical conditions not easily reached in terrestrial labs: macroscopic manifestation of the Pauli principle, high densities and pressures, and extremely high magnetic fields. After a brief introduction with some historical milestones the observational status is reviewed: spectroscopic classification, determination of stellar parameters from spectroscopic and photometric observations, effective temperatures, surface gravities, radii, and masses. The next sections deal with the physics of the interior and evolution of white dwarfs, leading to the mass–radius relation and cooling times. Going back closer to the observations again, the physical processes in the outer layers are discussed: gravitational separation, diffusion, radiative levitation, accretion, and convective mixing. This leads to a review of our current understanding of the origin of spectral types and their interrelation. A final section gives brief introductions to topics at the center of current research: white dwarfs in open and globular clusters, debris disks, the origin of accreted metals in the atmospheres, magnetic fields and their origin, variable white dwarfs, and white dwarfs in binaries. This chapter was finished in February 2010 and reflects the status of knowledge at that time.

Keywords: Accretion, Chandrasekhar mass, Convective mixing, Debris disks, Gravitational settling, Initial–final mass relation, Magnetic fields, Mass–radius relation, Stellar remnants, White dwarfs

1 Introduction

White dwarf stars are one of the known final stages of stellar evolution, besides neutron stars and black holes. With typical masses of $1/2$ the solar mass and radii of $1/80$ the solar radius they are very compact objects with interior densities of the order of 10^6 g cm^{-3} . Their existence, the mass–radius relation, and the limiting Chandrasekhar mass are macroscopic manifestations of the Pauli principle for fermions and the electron degeneracy following from it. White dwarfs (WDs) thus offer cosmic laboratories with physical conditions that cannot be achieved in terrestrial laboratories. The study of their structure and evolution tests our understanding of the equation of state under these extreme conditions. A fraction of at least 10% of all WDs show magnetic fields of up to 10^9 G in their atmospheres; this is much smaller than the fields in neutron stars, but still much larger than any fields studied in terrestrial labs. This range of field strengths is also a very difficult one to study theoretically, because Coulomb and magnetic fields in the atoms are of comparable strength and perturbation methods cannot be applied. Observations of magnetic WDs are the only possibility to test the predictions of energy levels and transition probabilities for the hydrogen and helium atoms, where such calculations are available.

Physical conditions are even more extreme in neutron stars and black holes. The advantage in the case of WDs is that their surfaces can be observed directly and studied with the classical methods of spectroscopy and stellar atmosphere modeling, giving reliable data for effective temperatures, surface gravities, chemical abundances, and the parameters derived from these data as masses, radii, and ages. Using these methods, a detailed understanding of the basic properties of WDs and their evolution has accumulated, although some even fairly fundamental questions are not yet fully understood.

There is as well a large interest in WDs beyond the community studying the physical properties of these objects, which is related to the use as a *tool* to study completely different aspects in other astrophysical areas. The spectra of the vast majority of WDs are very simple and can be modeled theoretically with high accuracy and confidence to better than 1% of the absolute flux. They are therefore primary candidates for the calibration of spectrographs and have been used for that purpose on the Hubble Space Telescope (HST) and many other spacecraft, as well as on many terrestrial telescopes.

About 95–98% of all stars end their nuclear life as a WD, and the final state of the universe is therefore one in which almost the total baryonic mass is converted into WDs, with some minor fraction in the galactic central black holes, neutron stars, and stellar black holes. The present population of WDs preserves information about the star formation rate throughout the history of our galaxy. The relation between the mass of a WD and that of its progenitor star on the main sequence (MS) indicates the total, integrated mass loss during the evolution. These are important data for an understanding of the mass budget of the galaxy: how much is locked up forever in remnants and how much is given back to the interstellar medium and can be used again for star formation? Since there is not yet a theory for the mass loss in all phases of stellar evolution, and since the observation is difficult especially in the latest phases on the Asymptotic Giant Branch (AGB), where mass loss is strongest, this is currently the only method to get reliable data.

The oldest white dwarfs in the solar neighborhood belong to the oldest objects in the galactic disk. Since it is possible to estimate their ages from theoretical evolutionary calculations, this gives a lower limit for the beginning of star formation in the disk. WDs in open and in globular clusters offer new and independent methods for age as well as distance determinations, which are often strongly coupled when using the traditional main sequence fitting method.

A rather new area of research concentrates on the study of dust and gas disks detected in the infrared around a significant fraction of WDs. The currently favored explanation is that these disks are the remnants of a smaller body like an asteroid, of a former planetary system around the progenitor star. When this body approaches the WD, it is pulled apart by the tidal forces. The study of the disk cannot give a very detailed picture of the chemical composition. However, in many cases this material is accreted onto the photosphere of the white dwarf, where it can be observed spectroscopically. If this interpretation is confirmed, it would offer a unique opportunity to study indirectly the composition of material from extrasolar planetary systems.

Many of these areas will be studied in more detail in the following sections. With more than 600 new publications every year, the WDs are a very vital area of research, which cannot be covered totally in a short chapter, aiming to provide a self-contained complete introduction to the field. At the end, therefore, the reader will find a section with suggestions for further reading.

1.1 A Few Historical Milestones

1910–1915: The first Hertzsprung–Russell diagrams constructed during this time showed three peculiar (blue–white) objects in the lower left corner, that is with hot surface temperatures, but much lower luminosity than the hot main sequence stars. These were the so-called “classical white dwarfs” 40 Eridani B, Van Maanen 2, and Sirius B. Their location could only be understood if the stars had a much smaller radius, by about a factor 100, than the main sequence stars

of corresponding spectral type. For Sirius B the mass was known from the binary orbit to be about $1 M_{\odot}$, resulting in a mean density of $\approx 10^6 \text{ g cm}^{-3}$, which was completely mysterious at that time.

1926: Fowler realized that the pressure supporting these stars is supplied by the almost completely degenerate electrons, as predicted by the just discovered Pauli principle for fermions.

1931–1939: Chandrasekhar worked out the complete theory of the equation of state at all densities, including the effects of special relativity. He discovered the mass–radius relation for white dwarfs, and the upper mass limit (the Chandrasekhar mass) of about $1.4 M_{\odot}$ for WDs with elements heavier than hydrogen in the interior.

1945: Schatzman explains the mono-elemental composition of most white dwarfs as due to gravitational settling.

1952: Mestel described the evolution of white dwarfs as a cooling process.

1958: 111 WDs known. First book on white dwarfs by Schatzman appeared.

1958–1970: Large proper motion surveys (Luyten, Bruce Proper Motion Survey; Giclas, Lowell Observatory) used the fact that WDs are close by (large proper motion) and blue, to find many candidate white dwarfs.

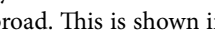
1965–1990: Extensive photometric and spectroscopic work by Greenstein and collaborators lead to major advancements in the understanding of WDs.

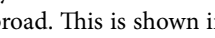
1978–1996: The International Ultraviolet Explorer (IUE), a small, but extremely successful satellite, opens up a completely new wavelength region for white dwarf research.

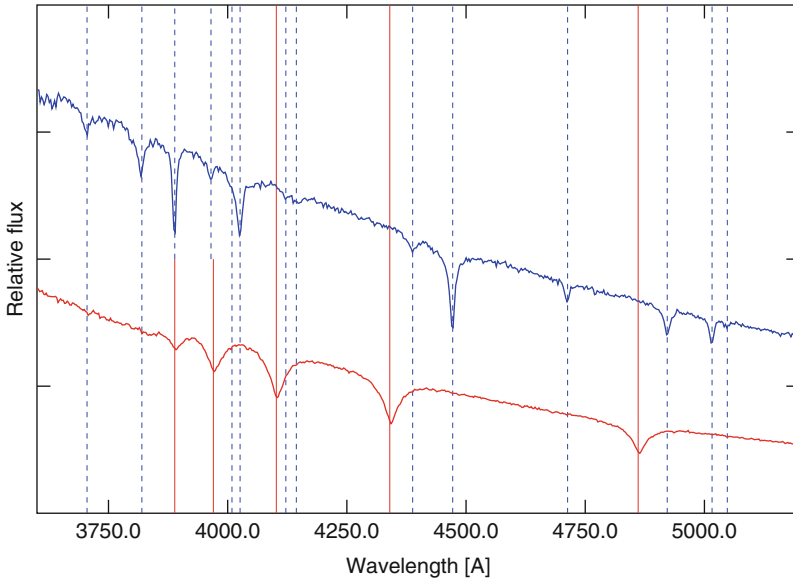
1986: The Palomar-Green survey increased the number of known white dwarfs significantly.

2004–2010: Several releases of White Dwarf Catalogs based on the Sloan Digital Sky Survey (SDSS) increased the number of spectroscopically confirmed white dwarfs to approximately 18,000.

2 Observations and Classification

Spectra of WDs differ from typical main sequence spectra in two respects: in most objects spectral lines of only one element are observed, either hydrogen, or helium, and the spectral lines are extremely broad. This is shown in  Fig. 11-1, which displays spectra of examples for the two largest groups of WDs, the so-called DAs and DBs. The spectra have a very high signal-to-noise ratio (SNR), because they are obtained from the calibration database for the HST. WDs are generally faint objects, and more typical spectra have lower SNR, even when observed with the largest telescopes.

The vertical lines in  Fig. 11-1 indicate the positions of hydrogen and helium spectral lines. It is obvious that the DA spectrum shows only hydrogen lines, and the DB spectrum only helium lines (a close inspection also shows several extremely weak carbon lines, which is not relevant for the current discussion). This indicates that the classification of white dwarf spectra was originally based on analogy with the main sequence spectra, where A indicates strong Balmer lines, and B strong neutral helium lines. The letter D stands for “degenerate,” based on the realization that electron degeneracy is the explanation for their existence. Contrary to the physical explanation of main sequence spectra as a temperature sequence, however, the main difference between spectral types of WDs is the elemental abundances. Detailed analyses of many objects have shown that there are two basic types of white dwarfs – those with hydrogen-dominated atmospheres and those with helium-dominated atmospheres. Since the latter are



■ Fig. 11-1

Spectra from the HST calibration database for a white dwarf of spectral type DA (*lower spectrum, red*), and DB (*upper spectrum, blue*). The vertical lines indicate the position of hydrogen Balmer lines (*continuous, red*) and helium lines (*dotted, blue*)

divided into several spectral types (see below), they are also often collectively called “Non-DAs,” or “hydrogen-deficient.” With very few exceptions, all other elements are at most tiny traces or completely absent from the surface layers.

The broadness of the lines indicates high gas pressure in the atmospheres. Since the pressure in the visible photosphere is caused by the weight of the higher layers, it is in a simple approximation proportional to the surface gravity, typically 10^8 cm s^{-2} , and confirming the estimates obtained from masses and radii.

2.1 Classification and Spectral Types

Over the 100 years of WD observations, extensions of the simple two classes system have been revealed and led to additions and refinements of the spectral classification system. The system in current use is defined in Sion et al. (1983); it uses a primary symbol for the elements with the strongest lines. The symbols from “P” to “V” in ► Table 11-1 are appended to the primary designation to indicate any peculiarities.

If lines of more than one element are visible, the corresponding element symbols are added. A DA with weak metal lines would be a DAZ, a star showing strong metal lines (most often the CaII H and K resonance lines), and in addition weak $H\alpha$ is a DZA, a DB with strong helium lines and weak CaII lines a DBZ, etc.

The DA group is by far the largest, with roughly 80% of all WDs. All the other types have helium-dominated atmospheres, with the exception of the recently discovered “hot DQ,” where

■ Table 11-1

Definition of primary spectral symbols

Spectral type	Characteristics
DA	Only Balmer lines; no HeI or metals present
DB	HeI lines; no H or metals present
DC	Continuous spectrum, no lines deeper than 5% in any part of the electromagnetic spectrum
DO	HeII strong; HeI or H present
DZ	Metal lines only; no H or He lines
DQ	Carbon features, either atomic or molecular in any part of the electromagnetic spectrum
P	Magnetic white dwarfs with detected polarization
H	Magnetic white dwarfs with Zeeman splitting in the spectra
X	Peculiar or unclassifiable spectrum
E	Emission lines are present
?	Uncertain assigned classification; a colon (:) may also be used
V	Optional symbol to denote variability

carbon is the most abundant element, and some of the PG1159 stars, where carbon or oxygen may dominate over helium.

The DAs are found over a very large range of effective temperatures, from about 6,000 K to over 100,000 K; for DBs the range is 10,000–30,000 K. To distinguish the objects within this range, a decimal digit was introduced as a temperature indicator, defined as $50,400/T_{eff}$, and ranging from 1 to 9. More recently this has been even more refined to half integer and double digit values. However, since the determination of effective temperatures needs a careful spectroscopic or photometric analysis with sophisticated theoretical models, such a designation seems to be in contradiction with the aim of an observational classification and has therefore only limited value.

DCs are found with effective temperatures below 10,000 K, where helium lines become invisible in the optical range. At this temperature they must be helium-rich, because otherwise strong Balmer lines would still be visible. With decreasing temperature, however, the Balmer lines also disappear and a distinction between hydrogen-rich or helium-rich atmosphere becomes extremely difficult.

DOs start around 45,000 K and extend to high temperatures above 100,000 K. They are helium-rich, but hydrogen lines are detected in some objects. The hottest group of WDs, which is sometimes included in the DO class, does not fit well into this classification system and is therefore classified as “PG1159 stars” by most authors, after the prototype star PG1159-035. These are also hydrogen-deficient stars characterized by absorption lines of HeII and highly ionized C and O.

The DZ stars show traces of heavy metals in a helium-rich atmosphere, most often Ca, followed sometimes by Mg and Fe. Because of gravitational settling (see ► Sect. 6) these elements stay in the outer layers only for times, which are always short compared to the evolutionary ages of the WDs. They cannot be primordial, but must be supplied from the outside. Because at the same metal abundance the lines are much stronger in helium-rich than in hydrogen-rich

atmospheres (because of the very low opacity of neutral helium), the hydrogen-rich equivalent of this class (DAZ) has been discovered only recently.

The “classical” DQ stars are cool, helium-rich stars below $T_{eff} \approx 11,000$ K. This class is considered separate from the DZ class, since carbon is believed to be present from the previous evolution below the thin outer helium layer, and can be dredged-up into the atmosphere by convective motions, which is indeed the accepted explanation for this group. The DQs were extended very recently to include a small group of “hot DQs” with T_{eff} between 18,000 and 24,000 K, in which carbon is the dominant element.

A spectral atlas, with an extensive coverage of most types, subtypes, and peculiarities has been published by Wesemael et al. (1993).

2.2 Photometry

White dwarfs have been observed in the Johnson UBV, Strömrgren, and other photometric systems. Before the advent of electronic detectors and large telescopes this was the only feasible way to observe larger samples for the study of population characteristics. Since about 1990 spectroscopic studies have become possible for this purpose and superseded the accuracy of the photometric studies.


The huge database of homogeneous and accurate photometry in the *ugriz* system of the Sloan Digital Sky Survey has led to a renewed interest in photometry. In particular the electronic availability of the photometry for a huge number of point sources has allowed successful targeted searches for new types of white dwarfs, or more members of the rare classes.

2.3 Parallaxes

The standard catalog for parallaxes is the “Fourth Edition of the General Catalogue of Trigonometric Stellar Parallaxes” (van Altena et al. 1995a, b). The WD parallaxes in this catalog have typical formal errors of 3.7 mas, very similar to the ≈ 20 bright WDs observed by the HIPPARCOS space project.

Parallax programs with dedicated telescopes and CCD cameras at the US Naval Observatory and similar projects at other observatories have now pushed the formal errors down to 1.0 mas and below. These accurate distances for a few hundred nearby WDs are important data for the study of the local population, space densities, and tests of the spectroscopic and photometric parameter determinations.

3 Stellar Parameters

The only method to determine masses independently of any assumption about the interior structure is the use of dynamical masses from visual binary systems containing a white dwarf. There are still only three stars with highly accurate results, which are shown in  Table 11-2.

■ Table 11-2

Parameters of visual binaries

Object	M/M_{\odot}	$R/100R_{\odot}$	Ref.
Sirius B	1.00 ± 0.02	0.8167 ± 0.038	Liebert et al. (2005b), Barstow et al. (2005)
Procyon B	0.58 ± 0.014	1.234 ± 0.032	Gatewood and Han (2006), Provencal et al. (2002)
40 Eri B	0.501 ± 0.011	1.36 ± 0.024	Shipman et al. (1997)
Stein 2051	0.50 ± 0.06	1.11 ± 0.15	Strand and Kallarkal (1989), Liebert (1976)

The fourth object, Stein 2051, is often added to this list, although only the mass ratio is known from the orbit. The mass of the non-degenerate component, a dM dwarf, is estimated from its absolute magnitude. The error of the WD mass may be larger than indicated in the table.

3.1 Effective Temperatures, Surface Gravities, and Abundances from Atmospheric Modeling

About a dozen WD-MS eclipsing binaries are known, where the fitting of the eclipse light curve together with spectroscopic and photometric constraints can be used to determine radii (Pyrzas et al. 2009; Steinfeldt et al. 2008). In general, however, radii can only be determined from an analysis of the emitted radiation with the help of model atmospheres. These models describe the physical structure (pressure, mass density, temperature) of the photospheric layers using the appropriate laws of physics, and from that calculate the emitted radiation (spectrum). In the case of WDs the chemical composition is usually simple and the following simplifying assumptions are adequate for the vast majority of WDs:

- **Homogeneous, plane parallel layers:** the depth of the atmosphere is considered to be very small compared to the radius of the star. All matter quantities (density, pressure, temperature) depend only on one geometric variable, the depth (in radial direction). The intensity of the radiation field depends on depth and the angle against the normal, but not on the azimuthal angle.
- **Hydrostatic equilibrium:** at each point within the outer layers, which have a direct influence on the emerging radiation (i.e., the atmosphere or photosphere), the gradient of the gas pressure is in equilibrium with the gravitational attraction (plus possibly the transfer of momentum by photons).
- **Radiative and convective equilibrium:** there is no energy generation or loss within the atmosphere, only transport of the energy generated in the deep interior. This transport can occur through radiation, heat conduction, or convection; the total energy flux as determined by the parameter effective temperature is constant at all depths.
- **Local Thermodynamic Equilibrium (LTE):** the matter is in thermal equilibrium corresponding to the local temperature at each layer, that is the ionization, excitation, dissociation of molecules, etc., are governed by the usual relations of thermal equilibrium (Boltzmann factors, Saha equation, Kirchhoff's law, etc.). Thermal equilibrium (i.e., the Planck function) is *not* assumed for the radiation field!

Except for the last these assumptions are well justified. Small deviations from LTE become visible around $T_{eff} = 40,000$ K, and above $\approx 60,000$ K Non-LTE atmosphere models have to be used for accurate results. If the laws of thermal equilibrium cannot be used, all important excitation and ionization reactions have to be calculated and the stationary state determined. This increases the computational effort by several orders of magnitudes.

The parameters, which describe an atmosphere model, are the effective temperature T_{eff} , the logarithm of the surface gravity, $\log g$, and the chemical abundances in the photosphere, in the following summarily denoted as XYZ . These are iterated, until a satisfactory fit between observed spectra or photometry and the model prediction is achieved. A typical example from Koester et al. (2009) is shown in ► Fig. 11-2 for the DA WD1129 + 071.

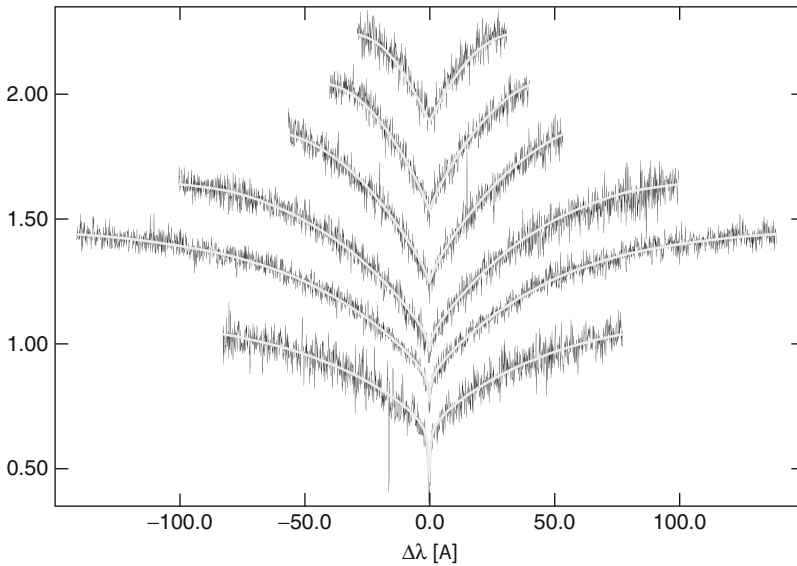
The most important quantity to be predicted by the final model is the average intensity of the stellar disk $F_\lambda(T_{eff}, \log g, XYZ)$, with wavelength λ . The relation with effective temperature is

$$\int_0^\infty F_\lambda d\lambda = \frac{\sigma}{\pi} T_{eff}^4 \quad (11.1)$$

with the radiation constant σ . Only in a few cases (e.g., study of variables) the angle-dependence of the surface intensity (i.e., the limb-darkening) is needed explicitly.

The observational quantity is the energy flux f_λ per time and area outside the terrestrial atmosphere, which is corrected for atmospheric extinction. The connection is

$$f_\lambda = \Omega F_\lambda = \frac{\pi R^2}{D^2} F_\lambda \quad (11.2)$$



■ Fig. 11-2

Spectrum of the DA white dwarf WD1129+071 (*black*) with a theoretical model (*light gray*). The relative intensity of the Balmer lines $H\alpha$ to $H8$ is shown stacked and shifted upward (except for $H\alpha$). The model parameters are $T_{eff} = 14,300$ K, $\log g = 7.84$ (cgs units)

with the solid angle of the star Ω given by the radius R and distance D . If the distance is known from a trigonometric parallax, membership in a cluster or common proper motion pair, or any other means, then the radius can be determined. The values in [Table 11-2](#) are obtained from trigonometric parallaxes.

3.2 Determination of Masses and Radii

Very few WDs have accurate dynamical mass determinations. Many more have radius determinations as described above. The surface gravity g provides a relation between mass M and radius R

$$g = \frac{GM}{R^2} \quad (11.3)$$

with gravitational constant G , from which a mass can be estimated. Another such relation is the gravitational redshift, expressed in the form of a (Doppler) velocity

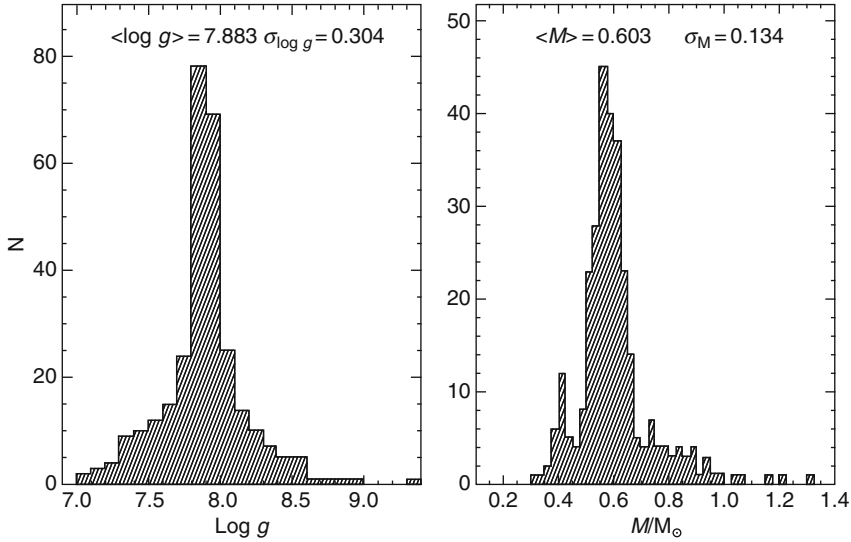
$$v_{grav} = \frac{GM}{cR} \quad (11.4)$$

with the velocity of light c . The problem is that the space motion of the white dwarf has to be known from membership in a binary or common proper motion pair, because the Doppler shift cannot be distinguished from the gravitational shift (Koester 1987; Reid 1996; Wegner and Reid 1991).

In order to determine stellar parameters for a large set of WDs, as necessary for a statistical study of properties, one has to make use of the theoretical mass–radius relation, which will be discussed in [Sect. 4](#). Originally derived by Chandrasekhar for zero temperature as a relation $R(M, XYZ)$, with XYZ denoting the interior composition, more modern versions include the small modifications for finite temperature and outer non-degenerate layers as $R(M, T_{eff}, XYZ)$. Together with a spectroscopic $\log g$ this allows the separate determination of mass and radius. [Figure 11-3](#) shows the distribution of $\log g$ and masses for 298 DA white dwarfs from the Palomar-Green (PG) Survey (Liebert et al. 2005a).

Prominent features of the mass distribution, well known since many years, are the sharp peak near $0.6 M_{\odot}$, the small secondary peak near $0.4 M_{\odot}$, and the extended tail toward higher masses. Since the higher mass WDs have smaller radius and hence smaller luminosities at the same T_{eff} , they are detected in a smaller volume only, if the survey is magnitude-limited like the PG survey. If this is corrected, the contribution of higher masses becomes significantly larger and a third peak near $0.9 M_{\odot}$ appears. The mass distribution of WDs depends on the distribution of progenitor masses, the mass loss between main sequence and WD, the star formation rate during the history of the galaxy, and the evolutionary time scales of the WD stage.

The only other large spectral type with statistically significant numbers of masses determined are the DBs. The mean mass is certainly within $\approx 0.1 M_{\odot}$ similar to that of the DAs. Studies to find smaller systematic differences are so far inconclusive. However, the secondary peaks at low as well as high masses seem to be absent in the DB mass distribution.



■ Fig. 11-3

Distribution of surface gravities and masses for 298 DA white dwarfs from Liebert et al. (2005a), Fig. 12 (Reproduced with permission from the author and the American Astronomical Society (AAS))

4 Interior Structure and Evolution

With interior densities of the order of 10^6 g cm^{-3} the mean distance r_e between electrons in completely ionized carbon is $\approx 10^{-10} \text{ cm}$. Even at interior temperatures of 10^8 K the ratio between the thermal de Broglie wavelength and the mean distance is

$$\frac{\lambda_{th}}{r_e} = \left(\frac{4\pi n_e}{3} \right)^{1/3} \frac{h}{\sqrt{2\pi m_e kT}} \approx 8 \quad (11.5)$$

with electron density n_e , mass m_e , temperature T , and the Planck and Boltzmann constants h, k . Quantum mechanical effects are therefore dominant for the electrons – they form a degenerate electron gas. Because of the Pauli principle, the electrons are forced to very high momentum states, when the spatial volume is decreased. These electrons then provide the major part of the pressure, which stabilizes the star against gravity.

4.1 Degeneracy and the Equation of State

The equation of state for the degenerate ideal gas can be derived from the distribution function in the 6-dimensional phase space of non-interacting particles $f(x, y, z, p_x, p_y, p_z, t)$. If a stationary state, spatially homogeneous distribution, and isotropic momentum distribution are assumed, the only variable remaining is the absolute value of the momentum p . Statistical

mechanics leads to the following results for particle density n , pressure P , and energy density u

$$n = \int f(p) 4\pi p^2 dp \quad (11.6)$$

$$P = \frac{1}{3} \int f(p) p v(p) 4\pi p^2 dp \quad (11.7)$$

$$u = \int f(p) \epsilon(p) 4\pi p^2 dp \quad (11.8)$$

with the velocity v and kinetic energy ϵ . Using the Boltzmann statistics for $f(p)$ leads to the classical ideal gas. For $T \rightarrow 0$, in the case of Fermions, the Boltzmann distribution has to be replaced by the Fermi–Dirac statistics

$$f(p) = \frac{2}{h^3} \frac{1}{e^{(\epsilon(p)-\mu)/kT} + 1} \quad (11.9)$$

The factor 2 comes from the internal degrees of freedom of the electron, μ is the chemical potential of the electrons, and the ratio $\eta = \mu/kT$ is used as degeneracy parameter, with large η indicating large degeneracy. To include the relativistic case that the velocity becomes comparable to the velocity of light, the general relations between energy, momentum, and velocity are used

$$\epsilon(p) = mc^2 \left[\sqrt{1 + (p/mc)^2} - 1 \right] \quad (11.10)$$

$$v(p) = \frac{\partial \epsilon}{\partial p} = \frac{p}{m} \left[1 + \left(\frac{p}{mc} \right)^2 \right]^{-1/2} \quad (11.11)$$

The equation of state for the degenerate electron gas is obtained in parametric form

$$n = \frac{2}{h^3} \int_0^\infty \frac{4\pi p^2 dp}{\exp[-\eta + \epsilon/kT] + 1} = n(T, \eta) \quad (11.12)$$

$$P = \frac{2}{3h^3} \int_0^\infty \frac{p \partial \epsilon / \partial p 4\pi p^2 dp}{\exp[-\eta + \epsilon/kT] + 1} = P(T, \eta) \quad (11.13)$$

$$u = \frac{2}{h^3} \int_0^\infty \frac{\epsilon(p) 4\pi p^2 dp}{\exp[-\eta + \epsilon/kT] + 1} = u(T, \eta) \quad (11.14)$$

For given n and T the first equation is inverted to obtain the degeneracy parameter η , which is then used to get the pressure and energy density. The necessary integrals have been redefined with non-dimensional variables. They can only be calculated numerically, but are available in tabular form as well as through numerical approximations.

An important limiting case is that of complete degeneracy at zero temperature. In this case a parameter $x = p_F/mc$ is used as a measure of the influence of relativistic effects. p_F is the Fermi momentum, the limit up to which all energy levels of the free electrons are filled. The equations (11.12) to (11.14) reduce to

$$n = \frac{8\pi m^3 c^3}{3h^3} x^3 = Bx^3 \quad (11.15)$$

$$\begin{aligned} P &= \frac{\pi m^4 c^5}{3h^3} \left[x(x^2 + 1)^{1/2} (2x^2 - 3) + \ln(x + \sqrt{1 + x^2}) \right] \\ &= Af(x) \end{aligned} \quad (11.16)$$

$$u = \frac{\pi m^4 c^5}{3h^3} \left[8x^3 (\sqrt{1 + x^2} - 1) - f(x) \right] = Ag(x) \quad (11.17)$$

In the nonrelativistic limit, $f(x) \approx 8/5 x^5$, while in the opposite limit $f(x) \approx 2 x^4$. In both cases the variable x can be eliminated from the first two equations, leading to $P \propto n^{5/3}$ or $P \propto n^{4/3}$. The transition between these two cases occurs approximately at a mass density of 10^6 g cm^{-3} in a carbon–oxygen mixture.

Although the derivation above is generally valid for fermions, in white dwarfs these results apply to the electron component only. Because of the mass appearing in (11.5), protons and neutrons become degenerate only at much higher densities, and can in a first approximation be treated as a classical ideal gas, with the total pressure being the sum of electron and ion contributions.

Deviations from the ideal gas behavior – classical or quantum mechanical – cause minor modifications (Chabrier et al. 2000; Kitsikis et al. 2005; Koester and Chanmugam 1990; Lamb and van Horn 1975; Lamb 1974; Salpeter 1961; Saumon et al. 1995; Segretain et al. 1994). The most important effect for finite-temperature WDs is the Coulomb interaction, measured with the parameter

$$\Gamma_C = \frac{Z^2 e^2}{r_0 k T} \quad (11.18)$$

with Z the charge of the ions and r_0 their mean distance. In addition to a small negative contribution to the pressure of the ions, the Coulomb interaction leads to a regular pattern of the positive ions as a body-centered cubic lattice, when $\Gamma_C \approx 175$. In analogy to the usual liquid–solid transition this is called “crystallization,” although the positive ions in WDs are completely ionized bare nuclei.

4.2 Interior Models and Mass–Radius Relation

If the electrons are highly degenerate and the thermal pressure of electrons and ions can be neglected, (11.15) and (11.16) define the complete EOS. For mono-atomic composition and complete ionization the mass density is related to the number density of electrons as

$$\rho = n_e \mu_e m_u = n_e m_u \frac{A}{Z} \quad (11.19)$$

Here m_u is the atomic mass unit, μ_e the “atomic weight per electron,” and A and Z are the mass and charge numbers of the nuclei. This EOS can be used to solve the mechanical part of the stellar structure equations (see 1 in this volume). A central density is assumed as the single parameter, and the equations of mass conservation and hydrostatic equilibrium can be integrated outward to $\rho = 0$, resulting in the density–pressure stratification, a stellar mass M , and radius R . This defines the famous mass–radius relation (MRR) for completely degenerate zero-temperature white dwarfs (Chandrasekhar 1939).

The essential features of this relation can be understood from very simple considerations. In hydrostatic equilibrium the pressure produced from the gravitational force at the center scales as

$$P_c \propto \frac{GM^2}{R^4} \quad (11.20)$$

This pressure is balanced by the degenerate electron pressure. In the non-relativistic limit

$$\frac{GM^2}{R^4} \propto \left(\frac{\rho_c}{\mu_e} \right)^{5/3} \propto \left(\frac{M}{\mu_e R^3} \right)^{5/3} \quad (11.21)$$

and therefore

$$R \propto \mu_e^{-5/3} M^{-1/3} \quad (11.22)$$

On the other hand, in the extreme-relativistic limit

$$\frac{GM^2}{R^4} \propto \left(\frac{\rho_c}{\mu_e} \right)^{4/3} \propto \left(\frac{M}{\mu_e R^3} \right)^{4/3} \quad (11.23)$$

There is only one solution $M = \text{const} = M_{Ch}$, independent of radius. As the density decreases outward, the extreme-relativistic limit is not applicable for the whole star, and this “Chandrasekhar mass” is only the asymptotic limit for infinite central density and $R = 0$. In physical units


$$M_{Ch} \mu_e^2 = 2.0182 \frac{6^{1/2}}{8\pi} \frac{(ch)^{3/2}}{G^{3/2} m_u^2} = 5.836 M_\odot \quad (11.24)$$

For interior composition of elements heavier than hydrogen, $\mu_e \approx 2$ and the limiting mass is close to $1.4 M_\odot$.

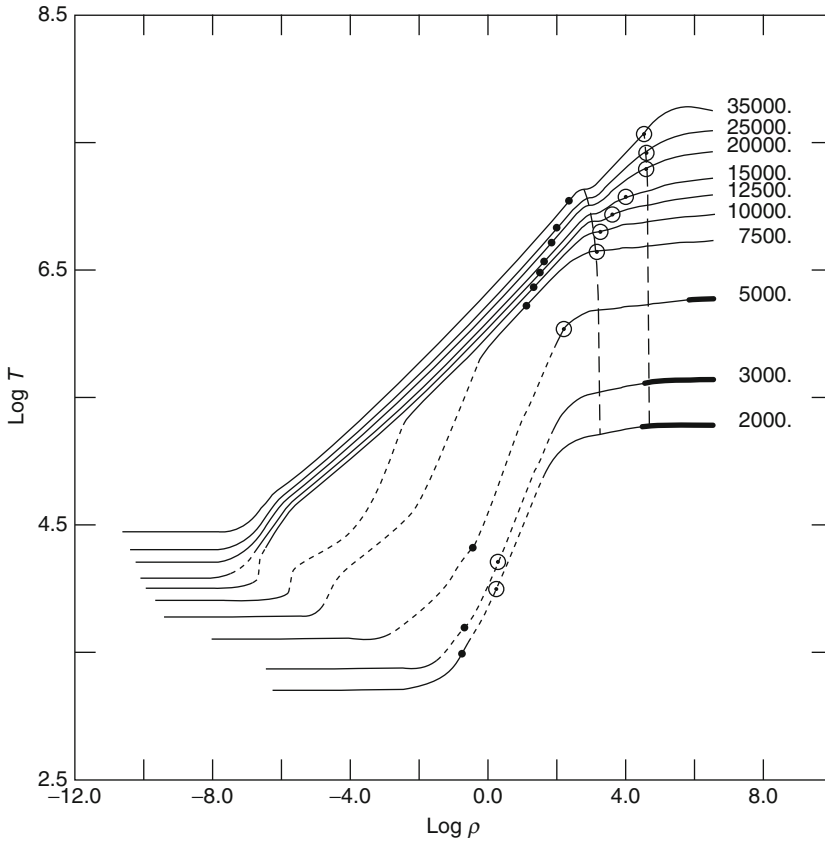
At high densities near the limiting mass the Fermi energy of the electrons can exceed the mass difference between the nuclei (A, Z) and $(A+1, Z-1)$. It then becomes energetically favorable to “capture” an electron onto the nucleus, transforming a proton into a neutron, a process called “inverse β -decay.” The neutrinos formed in this transition can escape easily from the white dwarf. For the C^{12} nucleus this instability occurs at $\rho = 3.9 \cdot 10^{10} \text{ g cm}^{-3}$, and the transformation then rapidly proceeds to Ne^{24} , because the intermediate nuclei have lower thresholds. Since these reactions increase μ_e , the mass–radius relation turns back toward smaller masses and the limiting mass is replaced by a slightly smaller maximum mass at finite density and radius.

Since the electron degeneracy is always large for the observed WDs, simple approximations have historically been used to calculate stellar models at finite temperatures and including non-ideal effects in the EOS. Such calculations make use of the fact that degenerate electrons have a very high thermal conductivity. Starting with outer boundary conditions from an atmosphere model, the equation of energy transfer can be integrated throughout the outer non-degenerate layers. When degeneracy is important, the temperature gradient becomes very small, and more than 99% of the total mass are at approximately constant temperature, identified with the central temperature of the WD. A useful numerical fit for the surface luminosity L as a function of the central temperature T_c , which can be used for a first estimate of the interior temperatures, is

$$\frac{L/L_\odot}{M/M_\odot} = 1.63 \cdot 10^{-21} T_c^{2.73} \quad (11.25)$$

For a typical observed WD with $0.6 M_\odot$ and $0.01 L_\odot$ this predicts interior temperatures of $\approx 10^7 \text{ K}$.  **Figure 11-4** shows the temperature–density structure of a $0.6 M_\odot$ model for T_{eff} from 2,000 to 35,000 K from the photosphere (left) to the center (right).

With the increasing power of modern computers, full stellar evolution calculations have been extended to cover all of the evolutionary phases. These calculations automatically provide a MRR, including the dependence on temperature, interior composition, and the previous evolution.



■ Fig. 11-4

Temperature–density structure from photosphere to center of representative DA models with $0.6 M_{\odot}$ and T_{eff} from 2,000 to 35,000 K (numbers on the right). The *solid*, *dotted*, and *thick solid* (for the three cooler epochs) portions of each curve indicate the radiative-conductive, convective, and crystallized regions. The top of the convection zone is always located in the photospheric layers. Electrons become degenerate to the right of the small filled circle on each curve. Likewise, the *bigger open circle* on each curve indicates the location to the right of which the Coulomb interaction becomes important for the ions ($\Gamma_c = 1$). The *dashed curves* define the composition transition zones, H/He at lower densities, and He/C at higher densities. Although the nearly isothermal carbon core looks small, it contains 99% of the total mass (From Fontaine et al. 2001, Fig. 1. Reproduced with permission from the author)

4.3 Gravitational Contraction and Cooling

The evolutionary calculations and their results will be discussed in ► Sect. 5. It is, however, very instructive to understand the main features of white dwarf evolution using a strongly simplified reasoning.

If a star has (momentarily) exhausted a nuclear energy source, energy can be freed by gravitational contraction. Since the star remains in hydrostatic equilibrium (as opposed to gravitational collapse), the equilibrium conditions for the central pressure can be described as

$$\frac{GM^2}{R^4} \propto GM^{2/3} \rho^{4/3} \propto \rho T \longrightarrow T \propto \rho^{1/3} \quad (11.26)$$

if the central EOS is that of an ideal classical gas. Contraction and increase of the density result in higher temperature, which can eventually start a new phase of nuclear fusion. If the density becomes high enough for electron degeneracy to be significant, the EOS changes. In the intermediate range the equations of partial degeneracy, described above, should be applied. A very simple approximation is to use the sum of the two contributions from thermal pressure of electrons and ions, and of the completely degenerate electrons. In the limits of negligible or complete degeneracy the other term is negligible, and in the intermediate region the error is at most a factor of 2. The condition of hydrostatic equilibrium is then

$$GM^{2/3} \rho^{4/3} \propto c_1 \rho T + c_2 \rho^{5/3} \quad (11.27)$$

$$T \propto \frac{GM^{2/3}}{c_1} \rho^{1/3} - \frac{c_2}{c_1} \rho^{2/3} \quad (11.28)$$

At low density the first term dominates. When the density approaches the transition region (f.e. defined by the equality of classical thermal and completely degenerate pressure), the central temperature goes through a maximum and then declines with further contraction!

A more sophisticated study (e.g., Koester and Chanmugam 1990) with the help of the virial theorem indicates that in the degenerate case, the released gravitational energy is completely used to increase the Fermi energy of the electron gas, while the decrease in thermal heat content is numerically equal to the radiated luminosity. Because of this equality the evolution of white dwarfs is called “cooling,” although it is necessarily accompanied by gravitational contraction like in ordinary non-degenerate stars, and the change of the gravitational binding energy is always of the same order of magnitude as the luminosity of the WD.

With further simplifying assumptions this can be formulated as

$$L = -\frac{dU_{th}}{dt} = -\frac{\partial U_{th}}{\partial T_c} \frac{dT_c}{dt} = -M < c_V > \frac{dT_c}{dt} \quad (11.29)$$

with U_{th} the thermal energy of ions and electrons, and $< c_V >$ an appropriately averaged value for the specific heat of the matter. Using the relation between L and T_c (☛ 11.25), the ideal gas c_V , and assuming that at $t = 0$ the WD starts at very high luminosity, this equation can be solved to give a “cooling age” for the WD with atomic mass number A

$$t \propto \frac{1}{A} \left(\frac{L}{M} \right)^\alpha \quad (11.30)$$

Using again (☛ 11.25) and adjusting the constant slightly with results from more sophisticated calculations gives the approximation

$$t[years] = \frac{1.6 \cdot 10^7}{A} \left(\frac{L/L_\odot}{M/M_\odot} \right)^{-0.634} \quad (11.31)$$


and thus ages of 10^6 to a few billion years for observed WDs.

5 Full Evolutionary Calculations


With modern stellar evolution codes the evolution can be followed from the main sequence through the phases of the asymptotic giant branch (AGB), and central star of a planetary nebula (CPN). An important ingredient is a description of the mass loss rate in all phases with semi-empirical formulae. The end of the nuclear burning phases is reached, when the pressure of the outer layers above a nuclear burning shell is too small to support further burning, and at the same time the center has entered the electron-degeneracy regime. The central temperature then does not rise anymore to ignite the next possible nuclear burning stage. After the heavy mass-loss phases at the end of the AGB and in the CPN stage the former core of the red giant constitutes the immediate progenitor of a white dwarf, a pre-white dwarf.

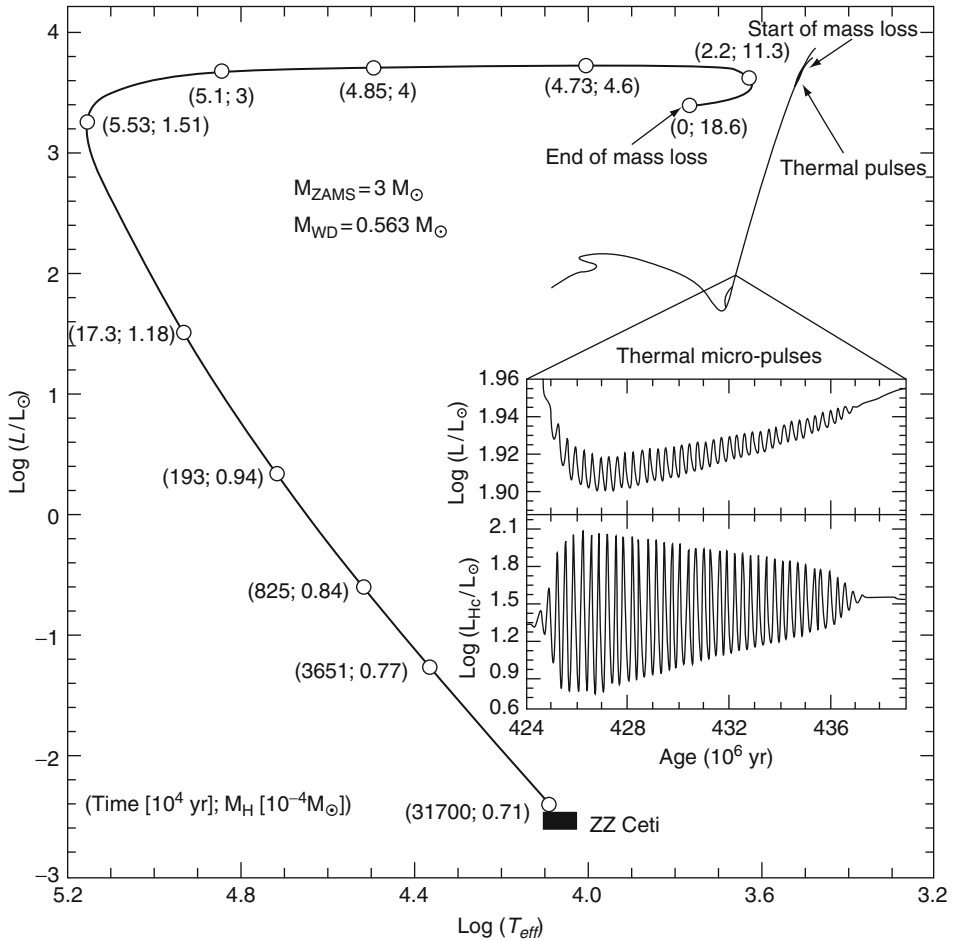
From the calculations of many groups a consensus has emerged that all single main sequence stars born with initial mass up to $\approx 8 M_{\odot}$ end their nuclear lives as WDs with masses up to the limiting mass of $1.4 M_{\odot}$, with the remainder given back to the interstellar medium. Stars with masses up to $\approx 7 M_{\odot}$ leave a carbon–oxygen (C/O) core with a mass up to $\approx 1.05 M_{\odot}$. The relative distribution of carbon and oxygen depends on the nuclear fusion rate $C^{12}(\alpha, \gamma)O^{16}$, which is still not very well determined, and the detailed treatment of convection and overshooting. This core is surrounded by a helium-rich shell (ashes of the hydrogen-burning) and an outer hydrogen-rich layer. The typical fractional masses obtained from the evolutionary calculations are 10^{-2} for helium and 10^{-4} for the hydrogen layer. There are indications that these fractions depend on the core mass (smaller for higher core masses) and that they may be modified by later evolution toward the white dwarf phase (late shell-flashes, stellar winds), but the details are not fully understood.

For stars with higher initial masses, but below the threshold for core-collapse supernovae ($\approx 8 M_{\odot}$), the carbon burning will be ignited and the core ($>1.05 M_{\odot}$) consists of an O/Ne mixture. The uncertainties of the detailed interior composition and outer shell masses remain major problems in WD evolution.

Very few calculations exist, where the evolution has been followed with the same code throughout the AGB and the WD cooling phase. Most authors use a different specialized code for the white dwarf phase, because new phenomena have to be considered, which are not as relevant in the previous phases. Examples are the crystallization of the interior, diffusion and mixing processes, and a detailed treatment of the thin outer layers, which is important for the interpretation of WD spectral types. The WD evolution calculations start with initial models of very hot pre-white dwarfs, derived by some artificial procedures (e.g., scaling a model to a different mass) from the cores obtained through pre-WD evolution. The first phases of the evolutionary tracks, in particular the derived evolutionary cooling ages, are then not very reliable. Only very recently large numbers of evolutionary tracks have been calculated all the way from main sequence to the pre-WD phase, providing very realistic starting models for the subsequent WD cooling of the whole range of masses. The masses of the outer layers are generally treated as free parameters, with the “standard” values mentioned above most often used.  [Figure 11-5](#) shows an example for such a calculation, starting with a $3 M_{\odot}$ star on the main sequence and ending with a DA white dwarf of $0.562 M_{\odot}$ and $T_{eff} \approx 12,000$ K.

5.1 Low Mass White Dwarfs

The mass distribution ( [Fig. 11-3](#)) shows a number of WDs with masses below $0.45 M_{\odot}$. This value is approximately the threshold for the mass of a helium core to ignite the helium-burning,



■ Fig. 11-5

Evolutionary track of a $3 M_{\odot}$ main sequence star through the AGB, CPN, and down to a $0.562 M_{\odot}$ DA white dwarf. Indicated in parentheses are the time since the AGB in 10^4 years, and the mass of the hydrogen layer in $10^{-4} M_{\odot}$ (From Althaus et al. 2002, Fig. 1. Reproduced with permission from the author and publisher Wiley-Blackwell)

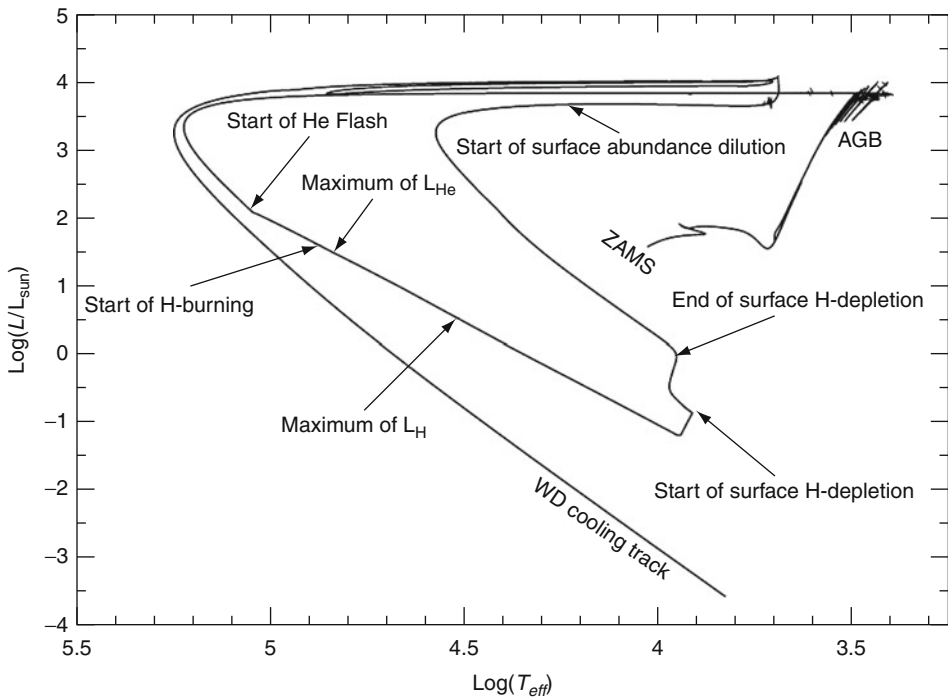
and lower mass WDs should therefore consist of He in the interior. The main sequence progenitors of such objects should have masses below $0.8 M_{\odot}$, and still be on the main sequence. The only possible explanation seems to be that these low-mass, helium-core WDs are the product of close binary evolution. The occurrence of low-mass WDs in binary systems with a pulsar companion provides support for this hypothesis. Some of these low-mass WDs probably have higher-mass (less luminous) WD companions, but a sizable fraction are apparently single stars of currently unknown origin.

The initial evolutionary phases of the binary systems are uncertain, and calculations aiming at the time scales, mass-radius relation, and observational properties in the later phases usually start by artificially removing the outer layers of a main sequence or red giant star, when the He core has reached the desired mass (Serenelli et al. 2002). Current emphasis is on more realistic calculations for the progenitor binary phase of mass transfer. This is important, because the

cooling times in the WD phase depend strongly on the thickness of the outer H layer. For thick layers residual H burning can prolong the cooling times significantly.

5.2 Origin and Evolution of Hydrogen-Deficient White Dwarfs

The above considerations are relevant for the majority of WDs with outer hydrogen envelopes. There is, however, a parallel sequence of hydrogen-deficient WDs (PG1159, DO, DB). The most widely accepted explanation is that these WDs have lost almost the complete hydrogen envelope through the so-called “born-again” scenario and evolve along a separate channel from the DAs (Iben et al. 1983). The basic idea is that the evolution depends on the timing of the last He shell-flash. If it occurs, while the star is still on the AGB, a normal DA results. If, however, there is a “very late thermal pulse” with the star already at the start of the WD cooling phase, the star moves back again to the AGB and burns most of its hydrogen, before turning back again toward the WD regime. ● Figure 11-6 shows such an evolutionary track in the HRD with different phases indicated. From the AGB the star moves through the PN stage onto the hot part of the WD cooling track. Then a final He-flash occurs, leading to a convection zone, which reaches up into the H-rich layers and leads to violent hydrogen-burning. The depletion of hydrogen



■ Fig. 11-6

Evolution of a $0.58 M_{\odot}$ remnant from the AGB through a very late thermal pulse (From Miller Bertolami et al. 2006, Fig. 2. Reproduced with permission from the author and the European Southern Observatory (ESO))

reaches the surface and any remaining hydrogen is diluted further. About 350 years after the He-flash the star reaches the red giant region for the second time, and then contracts again back toward the pre-WD stage. The time scales in this calculation are longer than observed in Sakurai's object, which is believed to be in such a born-again phase. That may be due to several uncertainties and approximations in the calculation.

On the other hand, nuclear burning, convective mixing, and element diffusion – processes which determine the chemical abundance profile above the He-shell, are treated in realistic detail. These layers become visible at the surface of the PG1159 stars and their likely precursors, the [wc] central stars of planetary nebulae with hydrogen-deficient atmospheres.

A detailed comparison of the predicted abundance profiles with the results of careful atmospheric analysis using Non-LTE models is presented in Werner and Herwig (2006). The main atmospheric constituents in the PG1159 stars are He/C/O. They show large scatter from star to star, but the class prototype, PG1159-035, has about the average abundances with 0.33/0.50/0.17 by mass. A more extreme case is HS 1517+7403, which has He/C/O = 0.85/0.13/0.02. Other elements that have been detected are N, Ne, F, Si, P, S, and in a few objects even hydrogen is seen. The comparison with model predictions confirms that the observed abundances indeed correspond to the abundances in the inter-shell region – between H- and He-shell – of the AGB progenitor. The observed scatter can be understood with differences in stellar mass, mass-loss on the AGB, or number of thermal pulses experienced.

It is generally believed that the PG1159 stars are the progenitors of the hydrogen-deficient sequence of WDs, after the heavy elements have settled down (see below). There is a caveat, however: some PG1159 stars show hydrogen, and a mass fraction of 1% H cannot be excluded for any of the group. This hydrogen would remain at the surface and necessarily turn the star into a DA at lower T_{eff} .

5.3 Other Evolutionary Channels

Subluminous O and B stars (sdO, sdB) are core helium-burning objects with hydrogen envelopes too thin to support hydrogen-burning. While there are still open questions about their origin, the future is much clearer. These objects will evolve directly into white dwarfs, without passing through the AGB. The sdB stars have masses of about $0.5 M_{\odot}$ and will thus produce low mass WDs; some sdO stars may result from binary mergers and also leave higher mass WDs. The total contribution of sdOs and sdBs to the white dwarf birthrate is estimated as $\approx 2\%$, with very large uncertainty.

5.4 Evolution and the HRD

Very widely used for the majority of WDs with C/O cores are the calculations from groups, which provide homogeneous large grids of evolutionary data for the general use. For more than a decade the calculations of Wood (1995) have set the standard. Since then many improvements have been made by various groups concerning the EOS, opacities, treatment of convection, inclusion of diffusion processes, better treatment of the crystallization, and more realistic initial models. Most of these improvements have been included in the grids published by the groups of Fontaine, Brassard, Bergeron, and collaborators in Montreal (Canada), and of Althaus, Benvenuto, and collaborators in La Plata (Argentina). Some recent papers are Brassard and Fontaine

(1998), Fontaine et al. (2001), Althaus et al. (2007, 2009), Panei et al. (2000, 2007), Benvenuto et al. (1999), and many references in these papers. Both groups make these data available from their websites. In the following some results from these evolutionary calculations and applications to observations are discussed.

Figure 11-7 shows the time evolution of DA WDs with masses $>0.45 M_{\odot}$. While the general trend is described by the simple cooling approximation of (11.31), there are significant differences mainly at high and low luminosities. These are caused by effects not taken into account in the approximation:

- At high luminosities plasmons (photons coupled to electrons in a plasma) may decay into a neutrino–antineutrino pair (Yakovlev et al. 2001), which leaves the star immediately. This energy loss may even exceed the photon luminosity and lead to an accelerated cooling.
- If the outer hydrogen layer has a mass of $10^{-4} M_{\odot}$ or larger, residual hydrogen-burning may contribute to the energy generation, slowing down the cooling.

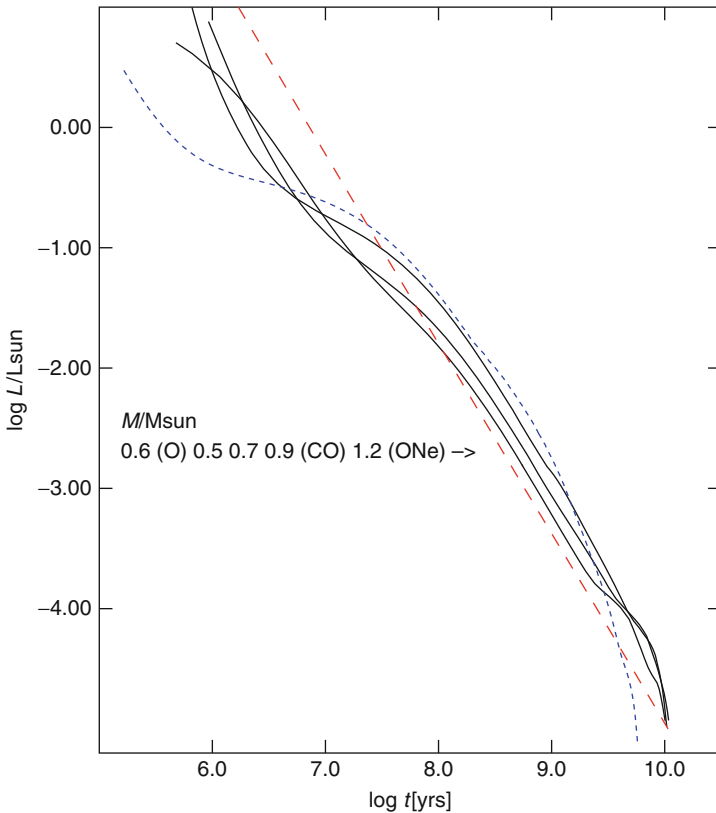


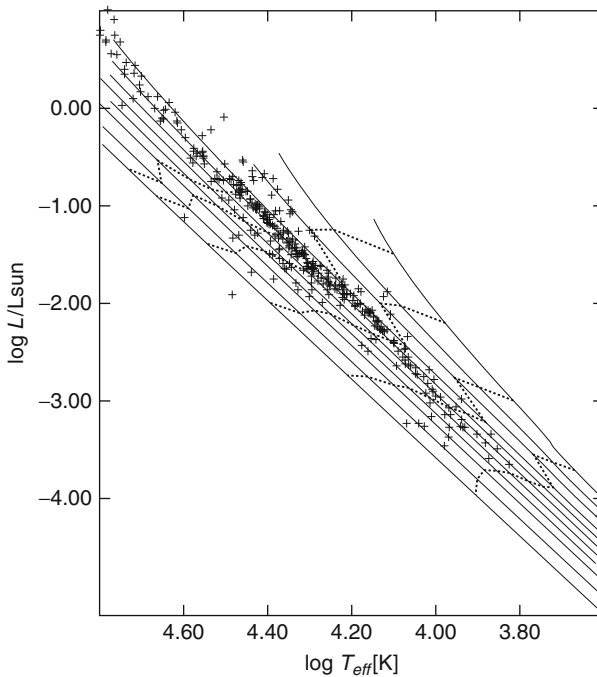
Fig. 11-7

Evolution of WDs with outer hydrogen layer and C/O ($M < 1.05 M_{\odot}$) or O/Ne cores ($M > 1.05 M_{\odot}$, blue, dotted). The dashed (red) line is the linear approximation of (11.31) for $0.6 M_{\odot}$ and $A = 16$ (Data from various papers of the La Plata group as collected on their website (see Sect. 8))

- At moderately low luminosities, with central temperatures between 10^6 and 10^7 K, depending on the mass of the WD, the crystallization starts at the center and moves outward with further cooling. This is a first order phase transition, and the latent heat released, of the order of kT per particle, slows the cooling. Moreover, the specific heat of the solid is larger than that of a gas, leading to additional slowing of the temperature decrease.
- If the interior consists of a mixture of carbon and oxygen, the crystallization will lead to enhanced oxygen content of the solid. Because oxygen is transported toward the center and the lighter carbon outward, this effect releases a small amount of gravitational energy. If this occurs at low luminosities (for low mass WDs), the ages may be increased significantly (Isern et al. 1997). A very lucid description of this effect is Montgomery et al. (1999).
- At very low luminosities quantum effects become important for the solid. Higher energy levels do not take part anymore in the heat exchange, the specific heat falls rapidly to zero as $c_V \propto T^3$, and the interior temperature and luminosity decrease rapidly. This is usually referred to as “Debye cooling.”

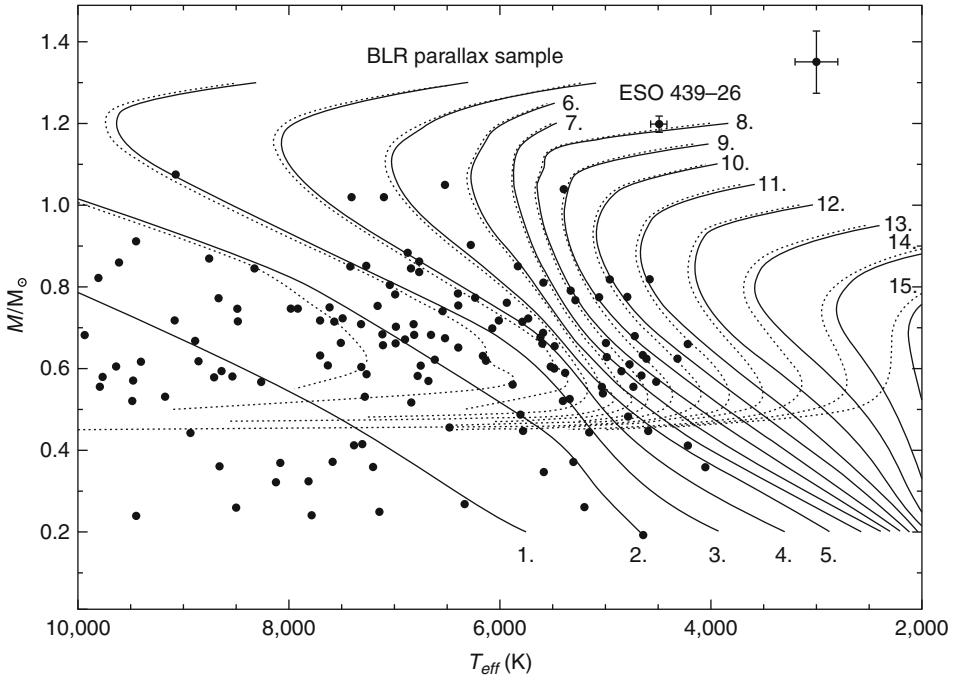
A comparison of theoretical evolution with observations is possible in the HRD.

☛ Figure 11-8 shows cooling tracks for DA white dwarfs with masses from 0.2 to $1.2 M_\odot$, and



■ Fig. 11-8

Evolution of WDs with outer hydrogen layer in the HRD. The masses are from right to left $0.2\text{--}1.2 M_\odot$, with a step of $0.1 M_\odot$. The models from 0.1 to $0.3 M_\odot$ have a helium core, those of 1.1 and $1.2 M_\odot$ a O/Ne core, and the intermediate models have C/O cores. The *dotted lines* are isochrones for (from top to bottom) $\log t_{\text{cool}}[\text{years}] = 7.0, 7.5, 8.0, 8.5, 9.0, 9.5$ (The evolutionary data are from the website of the La Plata group. The crosses are 348 observed DAs from the PG survey (Liebert et al. 2005a))



■ Fig. 11-9

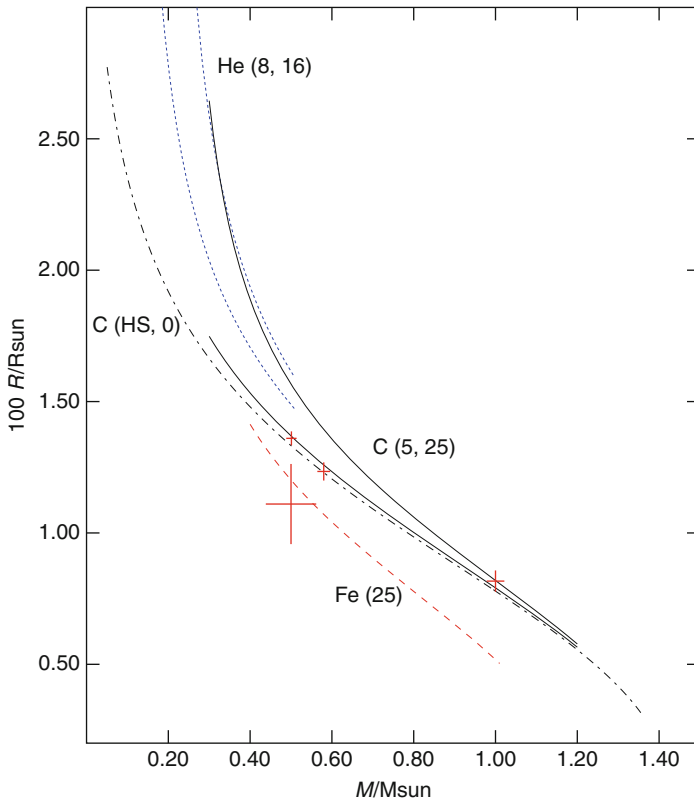
Distribution of 135 cool white dwarfs in a mass- T_{eff} diagram. These are stars with trigonometric parallaxes from Bergeron et al. (2001). The average uncertainties on the determinations of the individual values of M and T_{eff} are indicated by the cross in the upper right corner. The solid curves are isochrones for the white dwarf cooling phase only, while the dotted curves take into account the lifetimes of the progenitors on the main sequence. The ages are noted in Gyear (From Fontaine et al. 2001, Fig. 8. Reproduced with permission from the author)

observed DAs from Liebert et al. (2005a). These data are from the PG survey, which was limited to fairly blue objects, hence the cool WDs are missing in this diagram.

Another instructive way to show the connection between masses, T_{eff} , and ages is ▶ Fig. 11-9. The objects with apparent masses below $0.45 M_{\odot}$ are helium-core stars resulting from close binary evolution or unresolved double-degenerates. In the latter case the luminosity appears too large for their temperature, and therefore a large radius is derived from the known distance. For these objects a comparison with the isochrones is meaningless; for all other stars individual ages can be directly inferred. The two oldest stars appear to be close to 11 Gyear old, setting a lower limit for the star formation in the galactic disk.

5.5 Mass-Radius Relation

Another result from the cooling calculations for a range of masses is the radii as a function of mass, T_{eff} , and interior and envelope composition $R(M, T_{\text{eff}}, XYZ)$. A selection of these is



■ Fig. 11-10

Mass–Radius relations of white dwarfs. The dot-dashed (black) line is the relation for a pure carbon, zero-temperature WD (Hamada and Salpeter 1961). The two continuous (black) lines are models for carbon interior, fractional He layer mass of 10^{-2} and H layer mass of 10^{-4} , with $T_{\text{eff}} = 5,000$ K (left) and 25,000 K. The dashed (red) line is the relation for iron interior. The two dotted (blue) lines are He interior models with fractional H layer of 10^{-3} and $T_{\text{eff}} = 8,000$ K (left) and 16,000 K. The symbols with error bars are the four visual binary components from Table 11-2. The object with the large errors is Stein 2051. Only the mass ratio is known dynamically for this object; the mass of the secondary is estimated from its luminosity and the WD mass is probably more uncertain than the error bar indicates. A spectroscopic mass determination for this object gives $0.64 M_{\odot}$. The other three objects are from left to right 40 Eri B, Procyon B, and Sirius B

displayed in Fig. 11-10. Compared to the zero-temperature relation (Hamada and Salpeter 1961) the radii are always larger, but converge for large masses. For non-DA stars, without the outer H layer, the radii would be slightly smaller. The three well-determined positions of the visual binary components agree fairly well with the relations for carbon interior. Other determinations using surface gravity, gravitational redshift, and distances in various combinations have always much larger errors. They generally agree with the relations in Fig. 11-10, but are too uncertain to distinguish interior compositions or outer layer masses.

5.6 Luminosity Function and Age of the Galactic Disk

The luminosity function (LF) $n(L)$ describes the number of WDs in a luminosity interval for a stellar system, which can be the whole galactic disk, or a local sample (per pc^{-3} , or pc^{-2} if referred to a column perpendicular to the Galactic plane). Instead of the luminosity the absolute magnitude is sometimes used. For the simple WD cooling evolution the LF is proportional to the cooling time dt spent in the interval $L, L + dL$; this can be written as

$$n(L) = \int_{M_{low}}^{M_{up}} \frac{dt_{cool}(M_f, L)}{dL} \Psi(t - t_{cool}(L, M_f(M)) - t_{ms}(M)) \Phi(M) dM \quad (11.32)$$

Here M_{up} is the upper mass limit on the main sequence for stars, which become white dwarfs, currently taken as $\approx 8 M_\odot$, and M_{low} is the lowest mass for a star, which has had time to cool down to luminosity L

$$t_{ms}(M_{low}) + t_{cool}(M_f(M_{low}), L) = t \quad (11.33)$$

t_{ms} is the evolution time on the main sequence plus the following phases prior to the WD, t_{cool} the cooling time of the WD with mass M_f , which is the remnant of the original main sequence star with mass M . t is the present time, counted since star formation began in the system. $\Psi(t)$ is the time-dependent star formation rate and $\Phi(M)$ the initial mass function (IMF) on the main sequence. M_f and M are related through the initial–final mass relation (IFMR), here assumed to be a unique relation.

If the LF refers to a volume element in the solar neighborhood, a correction has to be applied which takes into account the change of the galactic scale height for WD populations of different age. [Figure 11-11](#) shows an example for the comparison of observed luminosity functions with theoretical models, where simple assumptions have been made for the used relations (Salpeter IMF, constant star formation rate).

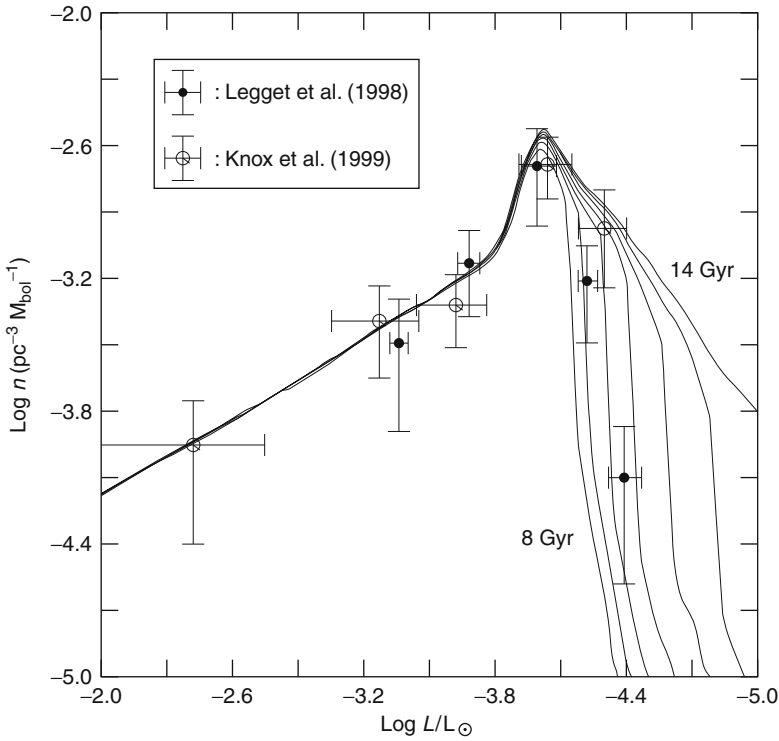
The observed LF decreases rapidly near $\log L/L_\odot = -4.5$. The widely accepted explanation is that the oldest WDs of our galactic disk are still visible today, and can be used to derive the time since the beginning of star formation in our disk (Winget et al. 1987). With the models used in [Fig. 11-11](#) this age would be between 10 and 11 Gyear, in agreement with the oldest WDs in [Fig. 11-9](#).

5.7 Local Space Density and Birthrate

Instead of ([11.32](#)) the LF can also be calculated as

$$n(L) = \int \dot{n}(M_f, t - t_{cool}(M_f, L)) \frac{dt_{cool}(M_f, L)}{dL} dM_f \quad (11.34)$$

with the WD birthrate $\dot{n}(M_f, t)$. The integral is over the range of possible WD masses. The absolute scale of the LF therefore contains implicit information on the integrated birthrate, as well as the total number density of white dwarfs, e.g., in the solar neighborhood. A careful consideration of selection effects, contributions of unseen companions in binaries, and non-DA white dwarfs lead to an estimate of $1.0 \times 10^{-12} \text{ pc}^{-3} \text{ year}^{-1}$ for the total birthrate, integrated over all masses (Liebert et al. 2005a). This is an estimate for the recent rate, because the PG survey used is a sample of hot, young WDs.



■ Fig. 11-11

Luminosity function of white dwarfs. The observed data are from Leggett et al. (1998) and Knox et al. (1999). The comparison with theoretical calculations is from Fontaine et al. (2001). Please note that the data are per interval of M_{bol} and not L (From Fontaine et al. 2001, Fig. 9. Reproduced with permission from the author)

Estimates for the birthrate of planetary nebulae, with central stars that must evolve into WDs, are about a factor 2–3 larger. Uncertainties of the distances and lifetimes, and the comparison of a local sample with PNs at much larger distances, may account for this difference, but that question is still open.

Total space densities of WDs are best determined in the immediate neighborhood. The nearly complete sample within 20 pc amounts to a total space density of $(4.8 \pm 0.5) 10^{-3} \text{ pc}^{-3}$ (Holberg et al. 2008). With ages up to 8 Gyear, the average rate over the history of the galactic disk is about a factor of 2 smaller than the current rate. This is quite plausible, since current WD formation occurs from smaller mass progenitors with higher number density in the Galaxy.

5.8 Initial–Final Mass Relation

An important relation for the theoretical predictions of the LF is the initial–final mass relation: what is the WD mass M_f for a main sequence star born with mass M_i . This relation is important

far beyond the study of WDs, since it determines the mass budget of our galaxy. Which fraction of the mass in main sequence stars is given back into the interstellar medium, to be used again in star formation; which part remains locked forever in the interior of cold WDs? And finally, closely related the question: What is the upper mass limit for stars to end as WDs versus supernovae?

In principle this would be a result of stellar evolution calculations following the star from birth to the WD stage. However, there is still no theory for the calculation of the mass loss in various phases from first principles. This mass loss is therefore described with empirical formulae fitted to observations. Nevertheless, the C/O cores predicted by these calculations in the latest phases of the AGB are used as a first theoretical estimate for the mass of the final WD.

More important has been a semi-empirical approach using white dwarfs in open clusters. In the most simple version, finding a white dwarf in an open cluster with a turnoff mass of $5 M_{\odot}$ demonstrates immediately that the progenitor star must have had a mass $>5 M_{\odot}$. It is easy to expand this argumentation to the following steps:

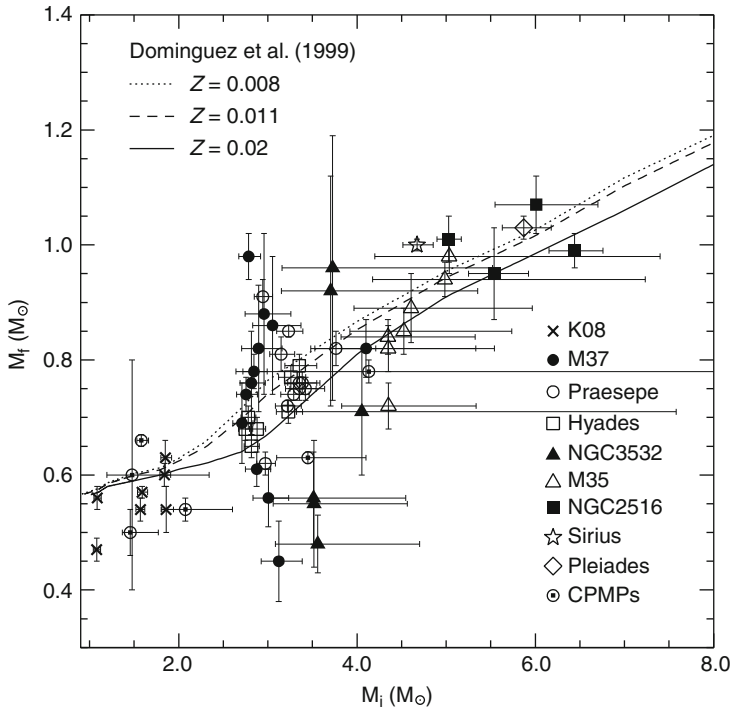
- Identify white dwarfs in an open cluster; establish likelihood of cluster membership.
- Determine spectroscopically and/or photometrically WD parameters T_{eff} , $\log g$, and the mass from the MRR.
- Get the cooling age of the WD from cooling calculations.
- Subtract cooling age from cluster age. This is the total evolution time of the progenitor from main sequence to PN.
- Compare this with stellar evolution calculations to determine the mass of the progenitor from the lifetime.

From the pioneering works of Tinsley (1974), van den Heuvel (1975), Sweeney (1976), Weidemann (1977), Romanishin and Angel (1980), Koester and Reimers (1981) based on photographic plates from small telescopes (compared to today), and early large scale computations of stellar evolution this work has been followed by numerous authors with much improved results. In addition to open clusters common proper motion pairs can be used with similar arguments, if the total age of the system can be determined from the non-degenerate companion.

❶ *Figure 11-12* shows an example for such a compilation, which includes results from a large body of literature. The open clusters are shown with different symbols; the common proper motion pairs help to define the low mass part of the relation. The theoretical curves show the core masses at the end of the AGB phase for different metallicities. The observations confirm the general trend of increasing final masses with progenitor mass. The scatter in individual clusters is still quite large, and the theoretically expected dependence of the relation on progenitor metallicity cannot yet be confirmed. The extrapolation of this relation to the highest observed WD masses indicates the upper limit of $\approx 8 M_{\odot}$ for WD formation, but the remaining uncertainty is at least $\pm 1 M_{\odot}$.

6 Physical Processes in the Outer Layers

The most obvious peculiarities observed in WDs are the two sequences of hydrogen-rich and hydrogen-deficient (or DA versus non-DA) objects and the largely mono-elemental composition of most WDs. The first is currently thought to be a result of the timing of the last thermal



■ Fig. 11-12

Initial-final mass relation from open clusters and common proper motion pairs. From Catalán et al. (2008). The theoretical core masses are from Dominguez et al. (1999). The large scatter of the NGC3532 WDs has been reduced in more recent work by Dobbie et al. (2009) (From Catalán et al. 2008, Fig. 1. Reproduced with permission from the author and publisher Wiley-Blackwell)

pulse in the He-burning shell, i.e., WDs are already born in two different kinds. The details of the chemical composition as well as their possible changes during the cooling evolution are influenced by several physical processes discussed in the following.

6.1 Gravitational Settling and Diffusion

The mono-elemental composition with the lightest element remaining floating to the top of the outer layers and being observable in the atmospheres was explained as gravitational settling by Schatzman (1945). Considering first an atmosphere of pure ionized hydrogen, the gravitational forces on protons and electrons are

$$F_p = m_p g \quad \text{and} \quad F_e = m_e g \quad (11.35)$$

with proton and electron masses m_p and m_e . This difference of the forces would lead to a change of the electron to proton concentration ratio with height, similar to the concentration gradient of O_2 and N_2 molecules in the terrestrial atmosphere. However, the particles in the stellar

atmosphere are charged, and a charge separation will produce an electric field E , until both forces are equal

$$m_p g + eE = m_e g - eE \longrightarrow E = -\frac{m_p - m_e}{2e} g \approx -\frac{m_p}{2e} g \quad (11.36)$$

Gravity and electric forces have the opposite sign for protons and the same for electrons. The charge separation is extremely small and the matter remains practically neutral, as can be seen with the help of the Poisson equations for electric and gravitational fields and their sources, starting from (11.36)

$$\operatorname{div} \mathbf{E} = -4\pi e (n_p - n_e) = -\frac{m_p}{2e} \operatorname{div} \mathbf{g} = \frac{m_p}{2e} 4\pi G n_p m_p \quad (11.37)$$

$$\left| \frac{n_p - n_e}{n_p} \right| = \frac{G m_p^2}{2e^2} \approx 4 \cdot 10^{-37} \quad (11.38)$$

which is the ratio of the gravitational to electrical force between two protons. Although the net charge is extremely small the electric field is important to secure the equilibrium between protons and electrons in the gravitational field. In writing down the hydrostatic equation for a star, the field is not considered explicitly; it is accounted for implicitly by assuming an effective molecular weight of $1/2$ for ionized hydrogen.

If a heavier element is present as a small trace, which does not change the global electric field significantly, the force on the ion with mass number A and actual charge number Z is

$$F(A, Z) = A m_p g + Z e E = m_p g \left[A - \frac{1}{2} Z \right] \quad (11.39)$$

The factor in brackets is $1/2$ for proton and electron, but 3 for doubly ionized He, or 43 for completely ionized Fe. As a consequence the heavy elements cannot be simultaneously in equilibrium with protons and electrons, they experience a net downward force. Since they are not in vacuum, but interact with the plasma, this leads to a diffusion process. Similar calculations apply to the case of hydrogen-deficient WDs, where the lightest element is He and heavier elements experience a net downward force relative to the He ions.

The collision integrals for Static Screened Coulomb Potentials, which are needed to describe this diffusion process, have been calculated and approximated with analytical fits in Paquette et al. (1986a). From these fits, diffusion coefficients and velocities v_{diff} can be calculated for the conditions in WDs. Assuming a constant reservoir of mass M_{res} , density ρ , and surface area $4\pi R^2$ at the bottom of the reservoir, the diffusion equation for the trace element abundance by mass X

$$M_{res} \frac{dX}{dt} = -4\pi R^2 X \rho v_{diff} \quad (11.40)$$

has a very simple solution as an exponential decrease of the heavy element abundance in the reservoir. The diffusion time scale is

$$\tau_{diff} = \frac{M_{res}}{4\pi R^2 \rho v_{diff}} \quad (11.41)$$

Results of such calculations are published in Paquette et al. (1986b) and Koester (2009); a few examples are given in Table 11-3. These time scales are very short for hot WDs, but even at the lower temperatures they remain always short compared to the cooling ages. If gravitational settling were the only process at work one would only observe one single element in all WDs.

■ Table 11-3

Diffusion time scales for DA models with $\log g = 8$ (left) and non-DA models (right). Columns 2–13 give the logarithm (base 10) of the time scales in years for the elements C, Na, Mg, Si, Ca, and Fe

T_{eff}	C	Na	Mg	Si	Ca	Fe	C	Na	Mg	Si	Ca	Fe
6000	4.50	4.34	4.29	4.21	4.15	4.01	6.50	6.57	6.55	6.59	6.56	6.56
10000	2.65	2.30	2.20	2.27	2.22	2.11	6.31	6.28	6.28	6.29	6.27	6.15
15000	-1.48	-2.03	-1.69	-1.70	-1.89	-1.97	5.84	5.72	5.71	5.71	5.53	5.46
20000	-1.21	-1.83	-1.55	-1.43	-1.72	-1.68	4.06	3.99	3.91	3.77	3.78	3.56
25000	-1.06	-1.67	-1.47	-1.24	-1.60	-1.57	1.90	1.66	1.57	1.42	1.57	1.37
50000	-0.41	-1.00	-1.05	-0.81	-0.98	-1.04	-0.96	-1.58	-1.59	-1.35	-1.55	-1.60

6.2 Stellar Winds and Radiative Levitation

A stellar wind, or mass loss, is necessarily connected with a velocity field in the outer layers, because the equation of continuity has to be satisfied. If this outward velocity becomes comparable to the diffusion velocities, the settling of the heavy elements will be slowed down or even prevented.

Mass loss has been observed in central stars of planetary nebulae on the horizontal track from the AGB toward high T_{eff} , including a few PG1159 stars. The mass loss rates are $10^{-6} - 10^{-9} M_{\odot} \text{ year}^{-1}$. No mass loss has been observed for white dwarfs on the cooling track, when the outer layers are compressed and $\log g$ becomes larger than 7. Mass-loss rates as low as $10^{-13} M_{\odot} \text{ year}^{-1}$, currently unobservable, would still have an effect on settling times, but theoretical predictions based on the radiation-driven wind theory do not expect such winds in the WD regime (Unglaub 2008). On the other hand, the diffusion time scales of He and heavier elements are very short and no mechanism is known at present to explain their presence together with hydrogen in the atmospheres down to 50,000 K, and cooling ages of the order of 10^6 years.

If the radiation field is not strong enough to cause mass loss, it may still transfer significant momentum on individual trace elements. This could support them in the photosphere against gravitational settling, a process called radiative levitation. The basic process is that the momentum transferred from photons absorbed by an ion is asymmetrical, with the outward momentum on average higher than the inward, because of the temperature gradient. If the re-emission of the photons is isotropic, a net momentum transfer results, reducing the effective gravitational acceleration. This process is most efficient for ions with many absorption lines in the wavelength range of the maximum of the radiation field. At very low abundances of such trace elements, every ion absorbs the maximum momentum, until the spectral lines start to saturate and an equilibrium abundance is approached. This abundance, or more correctly the equilibrium distribution in the outer layers, depends (for a given main constituent) only on the effective temperature and surface gravity (Chayer et al. 1995a, b; Fontaine and Michaud 1979; Landenberger-Schuh 2005; Michaud et al. 1976; Vauclair et al. 1979).

The importance of radiative levitation is confirmed by observations of the hottest DA white dwarfs. Whereas the optical spectra show only hydrogen, observations in the soft X-ray and EUV regions showed a clear flux deficiency compared to pure hydrogen models. When He was excluded as a source of this absorption, only heavier elements remained as explanation. However, the resolution of ROSAT (Röntgensatellit) and EUVE (Extreme Ultraviolet Explorer) was

too low to identify individual elements. This became only possible with high-resolution spectroscopy in the FUV region of the Far Ultraviolet Spectroscopic Explorer (FUSE) and in the UV with HST. Barstow et al. (2003) found that all DAs with $T_{eff} > 50,000$ K have detectable C, N, O, Si, Fe, Ni with abundances in the range $10^{-7} - 10^{-5}$ by number, relative to hydrogen. These elements are expected from radiative levitation theory to be supported; however, the detailed abundances do not agree very well with the predictions. Below 50,000 K, some DAs show metals, while others with very similar parameters apparently have pure hydrogen atmospheres. The reason for this dichotomy is not understood. Radiative levitation becomes negligible below 40,000 K.

In the parallel sequence of DOs, the hot helium-rich white dwarfs, metal traces are also present. The general trend of decreasing abundances with lower T_{eff} agrees with theoretical predictions, but individual abundances do not.

6.3 Convection and Dredge-Up

Energy transport in the non-degenerate layers can be through radiative transfer or convection, as decided by the Schwarzschild criterion for the temperature gradient. Because of the large length scales and low viscosity in stellar plasmas, the Reynolds numbers are large. The convective motion is always highly turbulent, leading to a chemically homogeneous mixed convection zone. For the determination of diffusion time scales in [Table 11-3](#) therefore the mass in the convection zone (CVZ) was used as the reservoir in the cooler models.

In a stratified structure, as expected in pre-white dwarfs from the previous evolution, convection can change the surface abundances, if during the cooling the convection zone extends downward into the transition layers. Evidence for this exists in the distribution of WD spectral types, and its interpretation requires the knowledge of the evolution of convection zones with cooling.

Convection in WDs is usually described with the mixing length theory (MLT) as applied to stars first by Böhm-Vitense (1958), although attempts have been made occasionally to use more sophisticated theories. While for MS stars like the sun the mixing-length l , or its ratio to the pressure scale height $\alpha = l/H_p$ is usually the only free parameter, it has become customary to use three more parameters for WDs. These additional dimension-less parameters (a, b, c) are introduced in the heuristic derivation of the theory and modify the average speed of convective elements, the energy flux, and the convective efficiency. A specific version of MLT is then noted f.e. as ML1/ $\alpha = 1.0$, or ML2/ $\alpha = 0.6$. ML1 and ML2 are the most often used combinations, they stand for the choices $a, b, c = 1/8, 1/2, 24$ (ML1) and $a, b, c = 1, 2, 16$ (ML2) (Fontaine et al. 1981; Jordan et al. 1998; Tassoul et al. 1990). With a given choice of the ML version, the efficiency of the convective energy transport generally increases with larger mixing length α ; the other parameters determine more subtle effects, e.g. whether the efficiency is increased more in the upper layers (low convection efficiency), or more in the deeper layers (high efficiency). The current standard for DA atmosphere models is ML2/ $\alpha = 0.6$, which has been shown by Bergeron et al. (1995) to give consistent fits to optical and UV spectra for variable DAs with $T_{eff} = 11,000 - 12,000$ K. Similar arguments favor ML2/ $\alpha = 1.25$ for DBs, although the evidence is weaker.

These “calibrations” of MLT are derived from spectroscopic observations of the atmospheres. The variable DA G29–38 has a shortest period around 200 s, which according to theoretical arguments should agree approximately with the thermal time scale of its convection

zone. However, with the standard parameters the convection zone is very thin, with a time scale of ≈ 1 s. This points to a higher convective efficiency in deeper layers. Similar conclusions are also drawn from a comparison of the mean temperature structure in 2d hydrodynamic simulations, with the MLT model atmospheres. No choice of MLT parameters can describe simultaneously the temperature structure in the upper and deeper layers; compared to a forced best fit MLT structure, the efficiency of convection needs to be lower in the upper and larger in the deeper layers. In summary, all these arguments, as well as comparisons with more sophisticated theories, demonstrate that MLT is a very poor approximation.

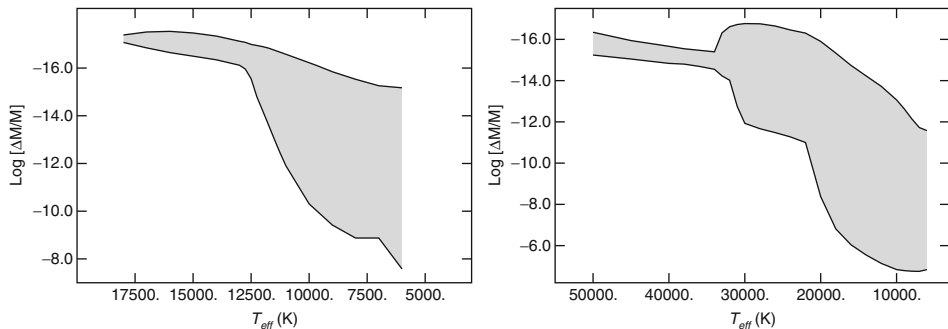
Keeping in mind these limitations, the size of convection zones and its evolution during the cooling is shown for DAs and DBs in [Fig. 11-13](#). In DAs, the convection zone develops near $T_{eff} = 18,000$ K for $0.6 M_{\odot}$ and remains shallow until about 12,500 K. Between 12,500 and 10,000 K the mass in the convection zone increases very rapidly by six orders of magnitude. Since convection is closely related to the partial ionization of hydrogen, these characteristic points in the evolution shift slightly to higher T_{eff} for larger masses (higher atmospheric pressure) and vice versa.

The picture is more complicated for the helium-rich WDs. A shallow convection zone driven by the HeII ionization is present as high as $T_{eff} = 60,000$ K. Around 33,000 K a HeI convection zone appears in higher layers, moves upward, and combines at depth with the HeII zone with further cooling, reaching down to fractional masses of $\approx 10^{-5}$. The evolution of the upper limits of convection zones for the lower T_{eff} is mostly caused by the change of atmospheric opacities; the upper limit is always in the photosphere.

The consequences of this evolution of the convection zones will be studied below.

6.4 Accretion and Diffusion

The presence of heavy elements like Ca in white dwarfs with cooling ages much larger than diffusion time scales can only be understood, if they are provided from a source outside the star. Historically, the most natural explanation has been accretion from the interstellar matter (ISM), but recently other possibilities (comets, debris disks as relics from asteroids) have come into focus.



■ Fig. 11-13

Evolution of convection zones (gray areas) during the cooling evolution in DAs (left) and DO/DBs (right) of mass $0.6 M_{\odot}$. $\Delta M/M$ is the fractional stellar mass above some depth in the outer layers

Accretion onto a spherical star is described by the theory developed by Hoyle, Lyttleton, and Bondi, and can be summarized in the equation for the accretion rate (Bondi 1952)

$$\dot{M}_{BHL} = \frac{4\pi(GM)^2\rho_\infty}{(v^2 + v_s^2)^{3/2}} \quad (11.42)$$

with ρ_∞ the density of the unperturbed ISM, v the relative velocity, and v_s the sound velocity. This is an interpolation formula between the solutions for a star moving at large velocity and a star at rest. In the words of Bondi, this equation is expected to give the order of magnitude of the accretion in a stationary state. There is some confusion in the literature concerning the factor of 4. It is an adjustment to numerical simulations and agrees with the original Hoyle-Lyttleton formula for large velocities (see Edgar 2004 for a discussion); the original Bondi formula has a factor of 2 instead. For typical WDs ($0.6 M_\odot$, 50 km s^{-1}) the rate would be of the order of 10^{-20} , 10^{-17} , $10^{-15} M_\odot \text{ year}^{-1}$ for the hot, warm, and cold phases of the ISM. All these calculations are applicable only, if the accretion is a hydrodynamical flow, i.e., the mean free path of the particles is short compared to the “gravitational radius” $\sqrt{2GM}/v$. If that is not true and the particles are freely falling onto the star, the much lower Eddington rate applies

$$\dot{M}_E = \frac{2\pi GMR}{v} \rho_\infty \quad (11.43)$$

These simple estimates do not take into account the extremely inhomogeneous nature of the ISM, the mixture of gas and dust, and a possible influence of the stellar radiation or a magnetic field on the accretion. The true rates of ISM accretion under realistic conditions are therefore only very poorly known.

The observable metal abundances at the surface are determined by the interplay of accretion and diffusion. Using now the convection zone as the reservoir, the evolution of the mass fraction X of a heavy element is

$$M_{cvz} \frac{dX}{dt} = X_{acc} \dot{M} - 4\pi r^2 X \rho w \quad (11.44)$$

$$= X_{acc} \dot{M} - \frac{X M_{cvz}}{\tau_{diff}} \quad (11.45)$$

with r , ρ , w the radius, density, and diffusion velocity at the bottom of the CVZ. Assuming the structure of the CVZ is unchanged, and $\dot{M} = \text{const}$, the solution is

$$X(t) = X(0) e^{-t/\tau_{diff}} + \frac{\tau_{diff} X_{acc} \dot{M}}{M_{cvz}} \left[1 - e^{-t/\tau_{diff}} \right] \quad (11.46)$$

where the first term on the right-hand side is the abundance at the start of the accretion phase at $t = 0$ (usually assumed to be zero). Three different phases can be distinguished with such a solution:

- For $t \ll \tau_{diff}$ the abundance increases linearly with time, with a slope only determined by the accretion rate.
- After a few times the diffusion time scale a steady state is reached with

$$X_{ss} = \frac{\tau_{diff} X_{acc} \dot{M}}{M_{cvz}} \quad (11.47)$$

Using M_{cvz} from a stellar model and a calculated diffusion time scale this equation allows an estimate of the accretion rate. Another interesting consequence of this steady state is that

the ratio of abundances for two elements is

$$\frac{X_1}{X_2} = \frac{\tau_1 X_{1,acc}}{\tau_2 X_{2,acc}} \quad (11.48)$$

In this case the abundance ratios of the accreted matter can be determined from the calculated diffusion time scales and the steady state observed abundances.

- If accretion stops, diffusion continues and the abundances decline with their diffusion time scales. Since these are usually different for different elements, the abundance ratios change exponentially with time, and conclusions about the accreted abundances become uncertain or even impossible.

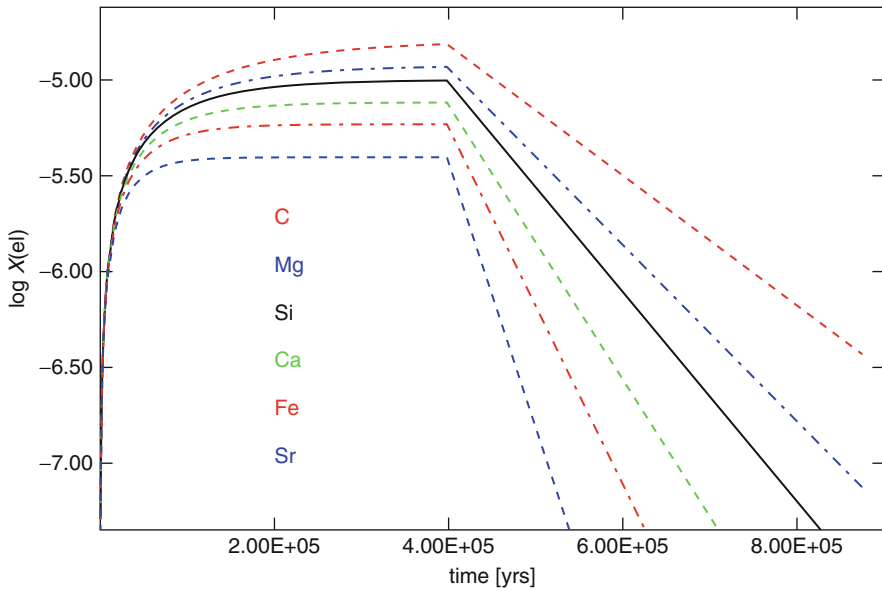
Obviously the steady state offers the best opportunity for conclusions. Looking at the numbers in [Table 11-3](#), diffusion time scales are very short in the hotter DAs, at most a few years above 12,000 K. Several DAZs in this range have been observed for a number of years without changes in the Ca abundance; it is therefore quite likely that they are in a steady state. This allowed Koester and Wilken (2006) to derive accretion rates in the range 10^{-18} to $3 \times 10^{-15} M_{\odot} \text{ year}^{-1}$ for 38 DAZs, assuming solar composition for the accreted matter. This is compatible with the predictions of the Bondi formula for the warm and cold phases of the ISM.

At least for the hotter DAZs this also allows to obtain heavy element abundance ratios for the accreted matter, which is important for the discussion of the source of this matter. The situation is much less favorable, if the diffusion time scales are long compared to observation “time scales,” which is generally the case for DBZs. A possible accretion/diffusion scenario for some of the heavy elements observed in the WD GD 368 is shown in [Fig. 11-14](#). Here it is assumed for simplicity that all elements are accreted with the same abundance, the figure thus shows only the relative changes. In this specific case, the atmospheric abundances ratios for times $t < 3 \times 10^4$ years agree with those in the accreted matter. After that time, the steady state is approached, and the ratios differ from the accreted ones by the ratios of the diffusion time scales, which are factors of 2–4 in this case. After accretion is switched off, the ratios diverge exponentially, and no conclusions about the accreted matter are possible.

The situation is more favorable, if an exponentially declining accretion rate is assumed with time scale τ_{acc} . This could be appropriate, if the source is a debris disk, which is not replenished from material further out in the system (Jura et al. 2009). In this case ([Fig. 11-15](#)), the abundances never reach a steady state, but instead begin to increase, reach a maximum, and then decline. The maxima are reached at different times for the elements, depending on their diffusion time scales. In this example, the accretion time scale is longer than all diffusion time scales. As time increases, the later decline is governed by the accretion time scale, which is the same for all elements. The abundance ratios therefore do not show the large deviations from the accreted values as in the case, where accretion is switched off. Asymptotically the element ratios approach that of the steady state, governed by the diffusion time scales.

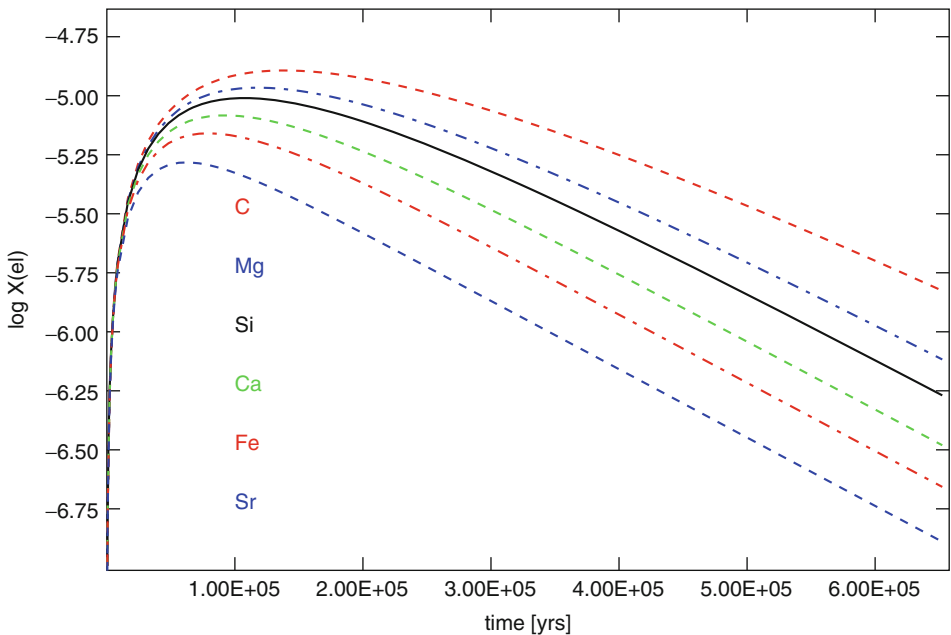
6.5 Spectral Evolution of White Dwarfs: Current Status

As discussed already, the main feature of two sequences is explained with a different origin. The hydrogen-rich objects evolve from the AGB through the hydrogen-rich central stars of planetary nebulae to the hottest DAs at $\approx 100,000$ K. They remain DAs until at least $T_{eff} \approx 12,000$ K (see below).



■ Fig. 11-14

Possible accretion/diffusion history for GD 362. The element symbols are ordered in the same sequence as the corresponding curve (From Koester 2009, Fig. 2. Reproduced with permission from the author and ESO)



■ Fig. 11-15



Alternative history for GD 362, assuming an exponentially declining accretion rate. Otherwise the meaning of the curves is the same as in Fig. 11-14 (From Koester 2009, Fig. 3. Reproduced with permission from the author and ESO)

If the last He shell-flash occurs late, at the beginning of the cooling sequence, the star moves a second time through the AGB and PN region, losing almost all of the remaining hydrogen-rich layers. The evolution goes very likely through the hydrogen-deficient central stars of Wolf-Rayet type [wc], and the PG1159 stars. Helium, carbon, and oxygen show abundances similar to that predicted from nuclear burning between the H- and He-burning shells on the AGB. Other strong arguments for this scenario are the large overabundances of neon and fluorine, which are both produced in a He-shell burning environment.

Between $\approx 120,000$ K and $75,000$ K the PG1159 stars are transformed into helium-dominated DOs; the carbon and oxygen settles down. DOs exist down to temperatures of $\approx 45,000$ K.

6.5.1 The DB Gap

For many years no helium-rich object was known in the range $30,000$ – $45,000$ K. This mystery was called the “DB gap” (Liebert et al. 1986). Within the huge new sample from the SDSS, Eisenstein et al. (2006) found several new DBs between $30,000$ and $40,000$ K and a few DOs between $40,000$ and $45,000$ K, but there is still a large deficit in this region compared to DAs of the same temperature, as well as to helium-rich objects above and below the gap. The basic explanation for the gap is in terms of diffusion and convective mixing (Fontaine and Wesemael 1987; Liebert et al. 1987). The small amount of hydrogen left mixed within the outer layers of the DOs floats to the top and becomes massive enough to be optically thick around $45,000$ K. The star now appears as a DA.

At about $30,000$ K (the exact value depending on the version of the MLT used) the He convection zone expands rapidly upward and downward. If the hydrogen layers are sufficiently thin in some stars, this hydrogen can be diluted to transform the star into a DB.  Figure 11-13 shows that the hydrogen mass needs to be of the order of 10^{-16} – $10^{-15} M_{\odot}$ (Macdonald and Vennes 1991). Again, the exact numbers depend on model details, e.g., the calculations used for the figure do not include the outer thin hydrogen layer. Among the remaining questions are the slow separation of the hydrogen in the DO phase: the settling times for He are very short, so there must be a mechanism delaying the settling. It could be a weak, as yet undetected stellar wind, or convective mixing. Even at $60,000$ K there is a small HeII convection zone remaining high in the atmosphere, which is receding to deeper layers with the cooling ( Fig. 11-13), perhaps paving the way for hydrogen to float up?

Many DOs have hydrogen mass fractions in their atmospheres of the order of 1%. If this is distributed homogeneously throughout the canonical He mass fraction of 10^{-2} , the total H mass fraction would be 10^{-4} and these stars could never become DBs. Either the He mass is much smaller in many objects, or the hydrogen is already concentrated in the upper layers in the DOs.

6.5.2 The DAO and DAB Stars

These are small groups of stars with strong Balmer lines, but HeII or HeI lines of comparable strength. Some are found close to the expected $\text{DO} \rightarrow \text{DA}$ or $\text{DA} \rightarrow \text{DB}$ transition temperatures, which suggests that these stars are just undergoing the transition and therefore show a mixed H/He atmosphere. Attempts to fit the optical spectra and overall energy distributions with

homogeneous or stratified atmosphere models, or spotted surfaces, have had only limited success. This may be expected, if the transition is a very dynamical process. Other DABs, however, have been identified as unresolved DA/DB binaries. Several DAOs have dM companions, others have low masses, and it has been suggested that they are descendants of extreme horizontal branch stars. Overall, these two classes may be rather diverse groups.

6.5.3 DB Stars with Hydrogen Traces: The DBA

In these WDs the hydrogen lines are weak, and the abundances are in the range 10^{-3} – 10^{-6} by number. Voss et al. (2007) found that more than half of all DBs show some hydrogen; this ratio has increased in the past with the quality of the observations, and it may be that at some level almost all DBs have hydrogen traces.

Taking into account the mass in the helium convection zone, which dilutes the hydrogen content, the total mass of hydrogen increases systematically from 10^{-16} to $10^{-10} M_{\odot}$ between 24,000 and 10,000 K. Whereas the smaller number is compatible with a recently formed DB from a DA with a very thin H layer, the latter number would be much too high for a DA→DB transition at 30,000 K. The only plausible explanation is that this hydrogen is accreted and accumulated at a rather low average rate of 10^{-21} – $10^{-19} M_{\odot} \text{ year}^{-1}$.

6.5.4 Cool White Dwarfs with Metal Traces: DZ, DBZ, DAZ

One of the first three “classical” white dwarfs – van Maanen 2 – showed strong Ca lines in the optical spectrum, and Ca has remained the only element found in *all* stars of these types. That is due to the fact that the CaII ion is the dominant Ca ion at the relevant temperature, and that it has the resonance lines in the optical range. With high resolution and high SNR, or using other spectral regions like the UV or FUV, several other elements can be detected, e.g., Mg, Fe, Si, and several more. It is estimated that – depending on the temperature range and parent spectral type – between 5% and 25% of all WDs show such traces.

The standard explanation has been the accretion/diffusion scenario with accretion from the ISM (Dupuis et al. 1992, 1993a,b). However, there have always been doubts, whether the necessary accretion rates are possible in the solar neighborhood, where no dense ISM clouds are known in the “local bubble.” An even more severe problem is that in most objects the abundance ratio of hydrogen (assumed to stay always on top) and the heavy elements is much less than solar (or than in the ISM). This can of course only be determined in DZs and DBZs with helium-dominated atmospheres. Somehow hydrogen is not or much less accreted. These problems, together with the recent detection of infrared excesses for many of these WDs, now favor a circumstellar dust disk as the source of the accretion. This new area of research is discussed in more detail in the next section.

6.5.5 The DQ Stars

DQ stars have a helium atmosphere and traces of carbon, in the form of CI lines or C_2 molecular bands, but generally no other metals. This case is different from the above, because carbon is expected to be the next element underlying the outer helium layer. If the helium convection zone penetrates deep enough and the helium mass is not too large, it could be expected that carbon will be dredged-up and become visible. Early explanation attempts along these lines were

dismissed, because in a simple model with a sharp transition between He and C, any mixing would immediately lead to a carbon-dominated atmosphere. Koester et al. (1982) and Pelletier et al. (1986) demonstrated that this problem could be solved by assuming a smooth transition as determined by the diffusion equilibrium between the two elements. The convection zone then only needs to reach the outer tail of the carbon distribution to provide the observed abundances. This scenario offers a possibility to estimate the helium-layer masses; recent calculations find mass fractions of 10^{-3} – 10^{-2} for the bulk of the DQ stars.

A recent addition to the WD spectral classification are the “hot DQ” stars, a small group with temperatures between 18,000 and 24,000 K. The atmospheres are dominated by carbon, with very little if any hydrogen and helium left (Dufour et al. 2008). Consulting [Fig. 11-13](#) shows that in the framework of the convective mixing/diffusion scenario DB stars with fractional He masses of the order of 10^{-8} could mix up the carbon in this temperature range. Interestingly, several hot DB stars, including the DBV prototype GD358 at 25,000 K, show traces of carbon in the UV and FUV spectra, but no other heavy elements. This may demonstrate the mixing process, where the convection zone first reaches the tail of the carbon distribution, in analogy to the cool DQ with much thicker helium envelopes, and the transformation to carbon-rich occurs perhaps, when the He convection zone combines with a cvz in the carbon layer. But where are the carbon-dominated stars below 18,000 K?

The latest twist in the DQ mysteries is the discovery of two objects with carbon lines and bands, but in addition with OI lines (Gänsicke et al. 2010). A first analysis shows these stars to have helium-rich atmospheres, polluted with carbon and oxygen, but *with an O/C ratio greater than 1*. These are the first WDs with such abundances, and it is speculated that they might be massive WDs with O/Ne cores and very thin outer He and C layers. Since these WDs probably originate from progenitors with $M > 7 M_{\odot}$, they should have masses above $1.0 M_{\odot}$. The mass determination is not possible spectroscopically and will have to wait for a determination of trigonometric parallaxes.

6.5.6 Evidence for DA to DB Transformation by Convective Mixing?

The evolution of convection zones in DAs shows a large increase in depth between 12,000 and 5,000 K. It was conjectured already by Koester (1976), Vauclair and Reisse (1977), and D’Antona and Mazzitelli (1979) that this could lead to a dredge-up and transformation of the DA into a DC. Indirect evidence for this process has accumulated and comes from different lines of argument:

- The ratio of DA to non-DA WDs is approximately 4 at $T_{eff} > 20,000$ K, but decreases to 1.6–2.0 for $T_{eff} < 12,000$ K, implying the transformation of a significant fraction to helium-dominated atmospheres, mostly of type DC below 10,000 K (Bergeron et al. 2001; Sion 1984; Sion et al. 2009; Tremblay and Bergeron 2008).
- The average surface gravity ($\log g$) of DA white dwarfs, which is ≈ 8.0 for DAs above 12,000 K, rises by about 0.2 dex at lower T_{eff} . This was first found by Bergeron et al. (1990) and confirmed in several independent studies since. It is highly unlikely that the average mass increases for the cooler white dwarfs; an alternative explanation is that helium is mixed into the hydrogen atmosphere. Because helium has a much lower opacity per mass in this temperature regime, the pressure in the visible layers is increased. Since the surface gravity is in effect measured as the pressure generating the line broadening, this cannot be distinguished from a real gravity increase. While this is an intriguing solution, there are still other

possibilities to explain the increase with deficiencies of the current atmosphere models, in particular the treatment of convection.

- Seismological studies of ZZ Ceti (variable DA) stars between 11,000 and 12,000 K show a range of fractional H masses from 10^{-4} to 10^{-10} (Castanheira and Kepler 2009). Those with the smallest values should mix with the helium layers just below 10,000 K.

6.5.7 Summary of Spectral Evolution

Obviously convective mixing and gravitational settling/diffusion are the major processes determining the spectral evolution of white dwarfs. While standard stellar evolution calculations predict “canonical” layer masses of 10^{-2} for helium and 10^{-4} for hydrogen for the typical WD with $0.6 M_{\odot}$ (with substantial modifications depending on the final mass of the WD), the consideration of observed spectral types requires a more complicated scenario. It is hard to avoid the conclusion that there is a variation of H layer masses in the DAs by several orders of magnitudes, possibly due to ongoing mass-loss toward lower luminosities than currently observed or predicted by theory. Variations of the remaining H mass could be caused by differences of the stellar masses and/or different amounts of accretion from the ISM.

Even more complicated is the non-DA sequence, but also in this case much of the evidence points to a large variation of the helium layer mass at the end of the DO phase. While the general picture seems understood, many details remain open.

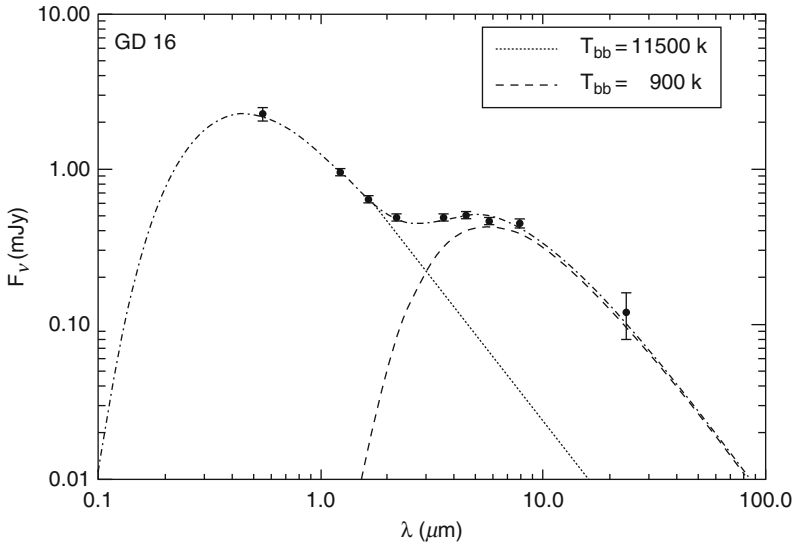
7 Some Areas of Current Interest

In this section some of the areas of WD research are reviewed, which are currently in the focus and together account for about 80% of annual publications.

7.1 White Dwarfs with Debris Disks

The first infrared excess – radiation clearly in excess of the extrapolated WD photospheric energy flux – was found for G29–38 by Zuckerman and Becklin (1987), and initially interpreted as a possible brown dwarf companion. Later studies demonstrated that the source is more likely a diffuse dust cloud in the vicinity of the star. The improvement of IR instrumentation has allowed several large scale surveys in the near and mid IR, identifying more than a dozen (14 as of Nov. 2009) WDs with similar IR excess. It is estimated that 1–3% of all white dwarfs have IR indications for circumstellar dust (Farihi et al. 2009). An example for such observations is shown in ► Fig. 11–16. The two components, WD and warm dust, are simply approximated with blackbody spectra here, but more sophisticated modeling confirms the conclusions.

While the models for the dust emission (e.g., Jura 2003) do not provide strong constraints for the morphology of the dust, the discovery of Keplerian motion in a few objects with gas emission lines (Gänsicke et al. 2006) clearly points to a disk structure, and this is now generally assumed for all objects with similar IR excess. Such a disk is believed to originate from a minor planet or some other rocky material of a former planetary system, which has been tidally disrupted in the vicinity of the WD, forming a “debris disk” (Debes and Sigurdsson 2002; Jura 2003). Simple



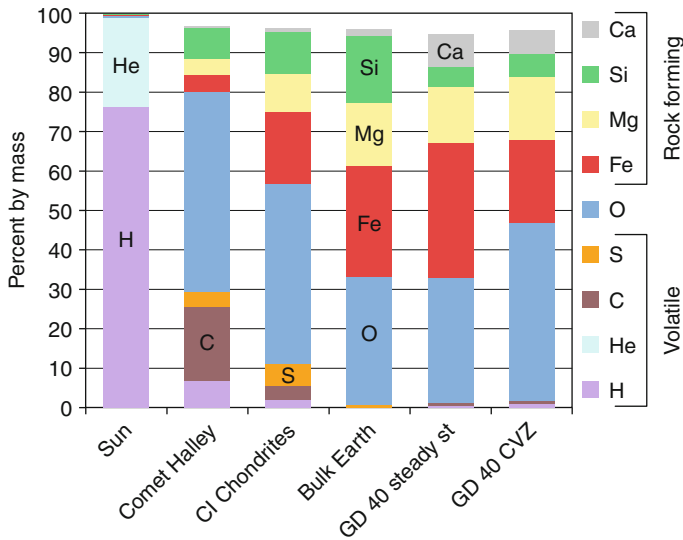
■ Fig. 11-16

Spectral energy distribution of the DAZ white dwarf GD16 and its dust disk, approximated with two blackbody spectral energy distributions (From Farihi et al. 2009, Fig. 2. Reproduced with permission from the author and Institute of Physics (IOP))

models of a geometrically thin, optically thick dust disk with temperatures of 600–1,200 K and extending between 0.1 and 1.0 R_{\odot} describe the continuum dust emission fairly well, but the observed strong silicate emission peak near 10 μ requires more sophisticated modeling.

All stars with dust disks show metal pollution in the atmospheres, i.e., they are DAZs or DBZs. Vice versa, of those with the strongest metal pollution, $\approx 50\%$ show a dust disk. This strongly suggests a connection between metal accretion and presence of a debris disk. The majority of the polluted stars, however, does not show IR excess; whether this is observational bias – the IR emission is too weak to be observed for smaller accretion rates – or whether ISM accretion has to be invoked as additional source is an open question (Kilic et al. 2006). Farihi et al. (2010) have presented strong evidence in favor of the circumstellar explanation for *all* metal-polluted WDs. In a study of 167 DZs and DAZs from the SDSS they do not find any correlation between the Ca abundance and the position above the galactic plane or the kinematic properties, as would be expected for ISM accretion. They reconsider the accretion physics for the ISM, and exclude the possibility of Bondi-Hoyle accretion within the local bubble. The total Ca masses are then much too high to be explained with the much lower Eddington rate. This together with the very large Ca/H ratio, much larger than the solar ratio, strongly argues for accretion of the metals from circumstellar material, most likely originated from rocky planetary matter with composition similar to the bulk earth composition. The total hydrogen mass (which can only increase with time in a helium WD) shows some correlation with stellar age, and interstellar accretion at a low rate is not excluded.

If the accretion source is a debris disk, it is possible to deduce the chemical composition of the debris from the photospheric abundances of the WD, using the conclusions derived in the previous section for the accretion/diffusion scenario. The situation is most obvious for the



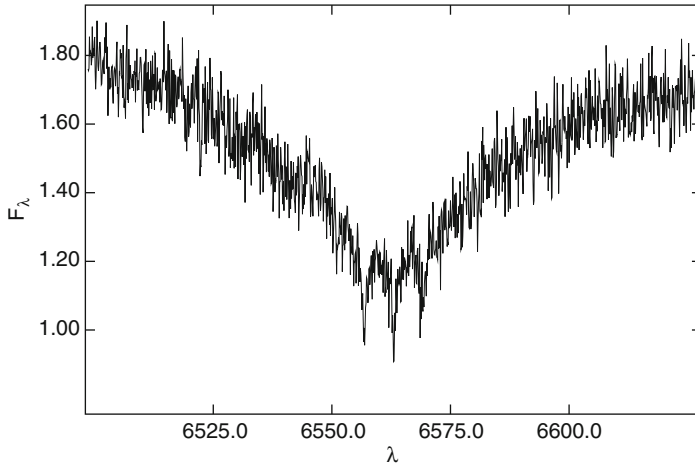
■ Fig. 11-17

GD40 volatile-deficient polluting abundances are most similar to a terrestrial planet composition. GD40 CVZ are the abundances measured directly in the photosphere, the steady state abundances have been corrected for the diffusion times assuming steady state (From Klein et al. 2010, Fig. 14. Reproduced with permission from the author and IOP)

hotter DAZs, with diffusion time scales of a few years or less. Here one can justify the assumption of a steady state and obtain heavy element abundance ratios for the debris from those in the photosphere using the calculated diffusion time scales. In helium atmospheres other arguments have to be invoked to make plausible that the diffusion process is not in a final phase of diverging abundances. If that can be achieved, the resulting abundance pattern holds important clues for the identification of the source of the accreted material (Zuckerman et al. 2007). Such an example is ► Fig. 11-17 from the analysis of the DBZA star GD40 (Klein et al. 2010). The findings are consistent with rocky planetary material in which heavy elements are largely contained within oxides, derived from a tidally disrupted minor planet at least the size of Vesta. The inferred parent body contained at most 10% of water by mass.

7.2 Magnetic White Dwarfs

More than 10% of white dwarfs show magnetic fields in the 1–100 MG range, with tails extending up to 1,000 MG and down to the observational limit of a few kG. Recent surveys with large telescopes indicate that even up to 25% may have fields in the kG range. The highest fields are very difficult or even impossible to study in terrestrial labs and the analysis of magnetic WDs (MWD) has been an important tool to verify theoretical calculations in atomic physics. Observationally these fields are detected through the Zeeman splitting of spectral lines, spectropolarimetry of lines and absorption edges, as well as broadband polarimetry of the continuum (Wickramasinghe and Ferrario 2000).



■ Fig. 11-18

Zeeman splitting of $H\alpha$ in WD0058-044 at a magnetic field of ≈ 0.32 MG

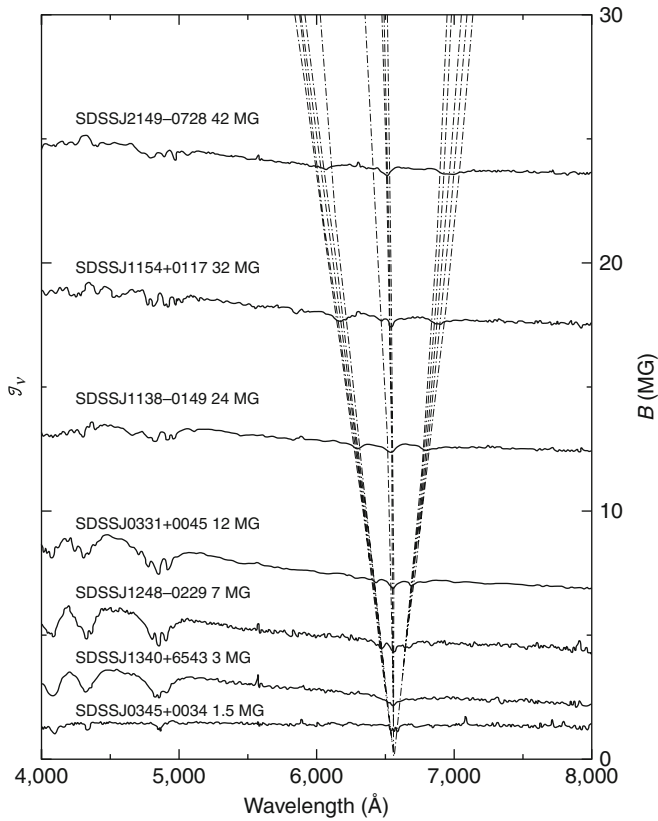
► *Figure 11-18* shows the Zeeman splitting at a low field of 0.32 MG in the white dwarf WD0058-044. ► *Figure 11-19* shows the transition to quadratic Zeeman splitting in the MG regime for several magnetic DAs detected in the SDSS.

For even higher fields the strength of the magnetic interaction becomes comparable to the Coulomb interaction in the atom and it is no longer possible to calculate the level shifts with perturbation methods. This problem could only be solved for the H and He atoms in the last 20 years with very demanding calculations on supercomputers. ► *Figure 11-20* shows the positions of the components of $H\alpha$ to $H\gamma$ for fields up to 1,000 MG. These shifts can reach several 1,000 Å, and it is not surprising that many spectral features were not identified until the advent of the theoretical calculations.

Since the magnetic fields are not homogeneous on the stellar surface, varying by a factor of two even in the most simple case of a pure dipole field, most spectral features are smeared out and become unrecognizable. The detected features are generally associated with “stationary points,” wavelengths for which the spectral shift reaches an extreme value, either maximum or minimum. Since to first order the shift is constant at these fields, relatively strong absorption features remain visible. These are used to determine the typical field strengths on the surface.

Weaker fields below the MG regime can be detected through spectropolarimetry. A spectral line such as $H\alpha$ is split into components, and on one side of the line the absorption for left-handed polarized light is stronger than for right-handed, while it is the opposite on the other side. This leads to a characteristic circular polarization pattern of a few percent at the position of the line, as shown in ► *Fig. 11-21*. The intensity spectrum shows no splitting at this resolution, but the polarization is easily recognized. A similar effect for the bound-free and free-free absorption coefficients can produce measurable polarization even in the continuum at larger field strengths.

Theoretical modeling of the atmospheres with the treatment of radiative transfer in a magnetized plasma has made large progress in recent years. The aim is to determine the field strength and the field geometry, in addition to the usual parameters T_{eff} and $\log g$ of a WD. The first assumption is typically a pure dipole geometry, but much better fits are in many cases



■ Fig. 11-19

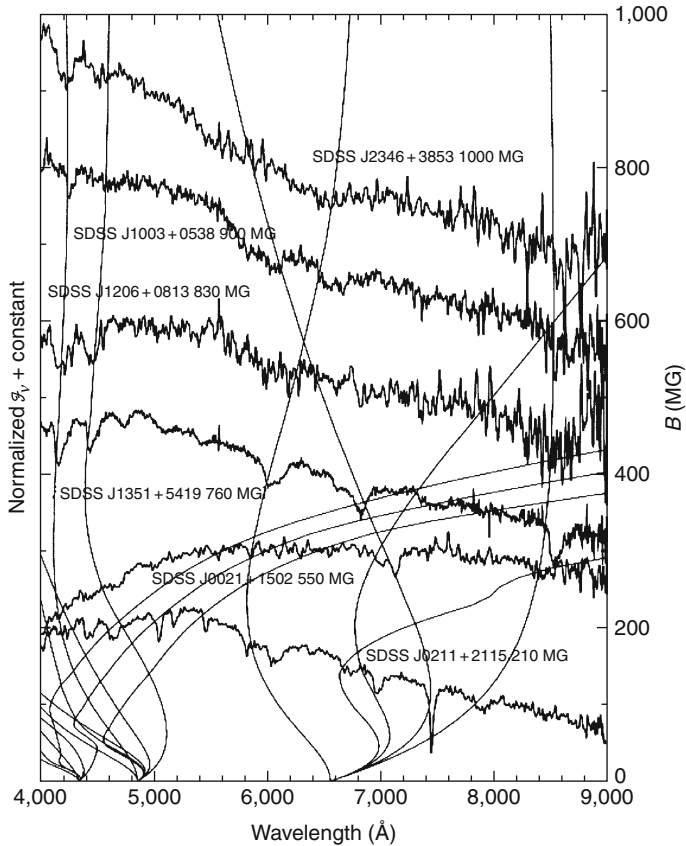
Zeeman splitting in 7 WDs from the SDSS, with fields ranging from 1.5 to 42 MG. The vertical axis (*left*) is the relative intensity, with the spectra shifted arbitrarily for clarity. The *dashed lines* show the theoretical shifts of the components of $H\alpha$ with field strength B , indicated at the right axis (From Schmidt et al. 2003, Fig. 1. Reproduced with permission from the author and IOP)

obtained by including higher order multipoles. Technically, to simplify the very demanding calculations, this is most often achieved by assuming an off-center dipole field, with the center of the dipole displaced a fraction of the radius along the magnetic axis (Kübeli et al. 2009; Wickramasinghe et al. 2002).

7.2.1 Origin of Magnetic Fields

The important facts for an understanding of the origin of these fields are:

- The incidence of at least 10% and possibly up to 25%.
- Statistically, magnetic WDs have a significantly higher mean mass of $\approx 0.9 M_{\odot}$, compared to $0.6 M_{\odot}$ for nonmagnetic WDs (Liebert 1988).

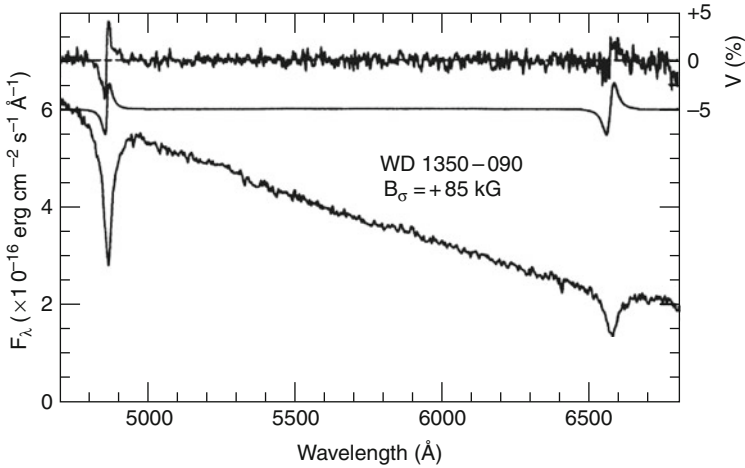


■ Fig. 11-20

Similar to ● Fig. 11-19, but for stronger fields. Many of the stronger absorption features can be identified with the position of stationary points (From Vanlandingham et al. 2005, Fig. 2. Reproduced with permission from the author and IOP)

- Contrary to nonmagnetic WDs there seem to be very few if any magnetic WDs with a detached main sequence companion (Liebert et al. 2005c).
- Rotation periods seem to have a bi-modal distribution. Most magnetic WDs are slow rotators with periods of days like nonmagnetic WDs, or even longer, up to 100 years. At the other extreme there is a small group of faster rotators, with periods of hours, down to 725 s.

Because of the high conductivity in the interior of WDs, Ohmic decay times for the fields (at least the lower multipole orders) are large ($\approx 10^{10}$ years), and the fields should be frozen in the matter. This suggests a fossil origin, and the evolution from the main sequence to WD with conservation of most of the magnetic flux. Compression of the radius by a factor of 100 thus demands fields from 0.1 to 10 kG on the main sequence. Such fields are indeed found in the group of chemically peculiar Ap and Bp stars, whereas nonmagnetic stars of the A/B spectral types have typical upper limits of 100 G. The fields in these hot stars without large convective outer layers are assumed to be fossil relics from star formation, in contrast to the lower main



■ Fig. 11-21

Circular polarization near $H\alpha$ and $H\beta$ in WD1350-090 compared with a theoretical model (*top*, labels on right axis). (*Below*) Intensity spectrum with labels on left axis (From Schmidt and Smith 1994, Fig. 1. Reproduced with permission from the author and IOP)

sequence, where a dynamo mechanism can be effective. Ap/Bp stars were long considered as the natural choice for the magnetic WD progenitors.

While the Ap stars have larger masses than the average progenitor of present day WDs, simulations using the IFMR for nonmagnetic stars do not reproduce the MWD mass distribution; one has to assume that the magnetic field changes the IFMR by decreasing the mass loss. The most significant argument against the Ap origin is, however, the recently increased fraction of MWDs. Ap/Bp stars can account for less than 7% MWDs, there must therefore be another source. The problem could be solved by assuming that 40% of main sequence stars with mass $>4.5 M_\odot$ have magnetic fields in the 10–100 G range, which are currently not detectable.

Another possible source has been identified as binary origin through the merging of two WDs. The WD EUVE J0317-853 shows the clear signature of such a merger: a very high mass of $\approx 1.35 M_\odot$ and a fast rotation with 725 s period. This may be the explanation for the group of fast rotators. A scenario proposed by Tout et al. (2008) goes even further by claiming that all MWDs with MG fields originate from binary evolution. The starting point of the argumentation is the observation that not a single WD out of 1,253 WD + dM pairs found in the SDSS shows a MG magnetic field, whereas at least 10% of single WDs do. The basic idea is that the magnetic fields are generated by dynamo action during the common envelope phase of binary evolution, with the magnetic field the stronger the closer the separation is at the end of this period. Magnetic cataclysmic variables in this view are the pairs which just avoided the merger, with fields predominantly below 10 MG; those pairs which merge form the high field MWDs.

7.3 Variable White Dwarfs

Long before the first variable white dwarf was detected the time scale for radial pulsations was estimated to be a few seconds. It was therefore surprising when Landolt (1968) found HL Tau

to vary with a period near 12.5 min. The pulsations are nonradial g-mode oscillations, with the motion at each point not confined to the radial direction. Gravity (g) mode indicates that the major restoring force is gravity and not pressure as in the radial oscillations. A single oscillation mode has an amplitude, e.g., at the surface described by spherical harmonics, with similar “quantum numbers” as in the hydrogenic wave functions. A few up to more than hundred of such modes can be excited at the same time and be observed as different frequencies in the power spectra of the light curves. The analysis of these modes – now called “asteroseismology” – can provide a wealth of information about the interior of the WD: masses, rotation rates, magnetic fields, structure of the outer H and He layers, and much more.

Variability is observed in four separate instability strips, depending on the spectral type. DAVs have T_{eff} around 11,500 K, DBVs around 25,000 K, and DOVs or variable PG1159 stars just below 20,000 K. While this newest class needs further study, for the others the position of the instability strip is related to a partial ionization zone of the most abundant species (H, He, C, O). The driving mechanism for the DOV is the classical κ mechanism, while for the DAV and DBV the interplay of convection and pulsation plays the decisive role, a mechanism termed “convective driving.”

The methods of asteroseismology, and results for white dwarfs, are discussed in the ► Chap. 4 and in the reviews by Fontaine and Brassard (2008) and Winget and Kepler (2008).

7.4 White Dwarfs in Open and Globular Clusters

The advantage of studying WDs in clusters is similar to that for other stellar classes: a sample of stars of the same age, at the same distance, the same original metallicity, and well-defined selection criteria. Additionally, the relation to the progenitor stars is well defined and the turnoff mass and metallicity of the MS stars are known.

7.4.1 White Dwarfs in Open Clusters

By far the most important aspect of studying WDs in open clusters is the determination of the initial–final mass relation described in ► Sect. 5. Young clusters with high turn-off masses are most interesting for this study and correspondingly there are only relatively few WD members, and only the brighter ones can be studied spectroscopically with sufficient accuracy to establish the IFMR relation.

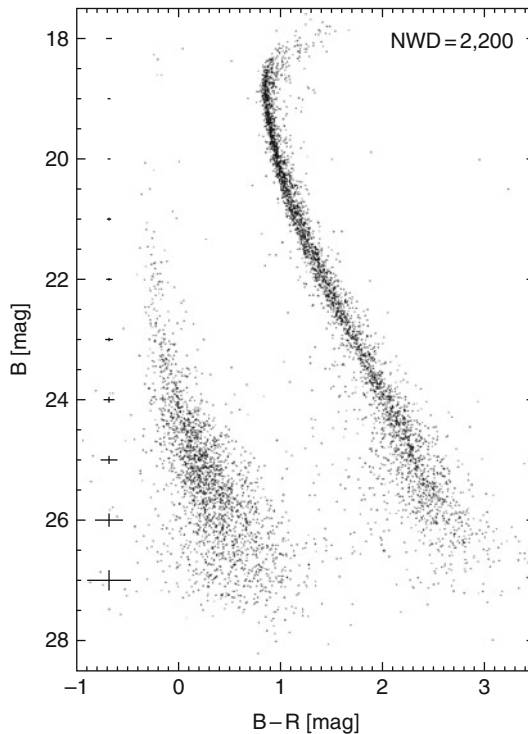
The cooling ages of WDs in the cluster provide in principle a second method for age determinations, independent of the MS fitting technique. Observational difficulties to reach the bottom of the WD cooling sequence, small samples, uncertainties about the nature of the coolest WDs (C/O cores vs. He cores, contribution of binaries, question of completeness) have hampered the comparison of the two methods and led to some conflicting results. A new method, which uses only the well studied brighter white dwarfs of the cluster seems promising (Jeffery et al. 2007).

About 100 WDs in open clusters have spectroscopically confirmed spectral types. Of these only four have a non-DA type, and none of these have the cluster membership definitely confirmed with proper motion measurements. The reason for this DA/non-DA ratio, which is very different from that of the field WDs, is unknown (Kalirai et al. 2005).

7.4.2 White Dwarfs in Globular Clusters

A few dozen WDs in the solar neighborhood have been claimed to have halo kinematics, but many of them are too young, even including the likely progenitor lifetime, to be members of the oldest stellar population. The status of individual halo white dwarfs, their space density, and possible contribution to the halo dark matter are not clear at present. Halo white dwarfs can, however, now be detected and studied in globular clusters with the HST and terrestrial large telescopes, although the WD cooling sequence is ≈ 8 magnitudes below the main sequence. Problems are obviously the faintness and the crowding of the stellar images, which make spectroscopic studies very difficult and confined to the brighter objects. The information learned about individual WDs as stars is therefore rather limited, and is mostly based on the study of color-magnitude diagrams (CMD, e.g., [Fig. 11-22](#)):

- Spectroscopic identifications have so far been obtained only for about 35 cluster WDs, and they are all of type DA. No helium-rich WDs have been identified. This is similar to the results for open clusters, and also unexplained (Davis et al. [2009](#); Kalirai et al. [2009](#)).



■ Fig. 11-22

Color-magnitude diagram for the globular cluster ω Centauri with main sequence and 2,200 white dwarfs. The data were collected with HST using the optical filters F435W and F625W, here denoted as B and R. The error bars demonstrate the intrinsic photometric errors (without calibration errors) at that magnitude (From Monelli et al. [2005](#), Fig. 1. Reproduced with permission from the author and IOP)

- The WD masses in the clusters as determined from the position in a CMD are roughly consistent with the range $0.5\text{--}0.6\,M_{\odot}$, as expected from the IFMR for low mass progenitors. Determinations independent of assuming a known distance, e.g., via a spectroscopic gravity determination, are just now becoming feasible, and indicate masses of $\approx 0.53\,M_{\odot}$ for the youngest WDs in the cluster (Kalirai et al. 2009).
- The youngest WDs in a cluster seem to have a more extended radial distribution than either their progenitors or the older WDs. The best explanation is that they have received a modest velocity kick at birth, before later relaxing to the distribution appropriate for their masses.
- The cooling sequences in color-magnitude diagram with several thousand WDs can provide well-defined luminosity functions, with information about the physics of WD cooling. Such data have been used to show that including the latent heat of the crystallization in the interior results in better fits (Winget et al. 2009).

More important is the prospect of using white dwarfs as a tool to determine ages and distances independent of the method of MS fitting. The distance determination works in a very similar way to the MS fitting technique: the position of the WD sequence in a CMD is compared to theoretical cooling tracks, or to the absolute magnitudes of WDs in the solar neighborhood with known distances. The advantage of this method is that the theoretical WD tracks are much simpler than MS evolution; they do not depend on metallicity, or assumptions about convective overshooting. One does need, however, a spectral classification to determine the presence or not of outer hydrogen layers, and, most important, an independent determination of the mass or surface gravity. This is not yet possible to the necessary extent, but the first attempts have demonstrated the potential of these methods (Kalirai et al. 2009; Layden et al. 2005).

Age determinations rely on identifying the end of the WD cooling sequence, a very challenging task at the limit of current possibilities. First results, e.g., 12 Gyear for NGC 6121, agree with main sequence determinations within the errors (Bedin et al. 2009).

7.5 White Dwarfs in Binaries

One of the most famous WDs is Sirius B, companion to the very bright star Sirius. The system is close to the sun and the binary orbit can be studied, resulting in one of the very few very accurate WD masses. Because of the 10 mag brightness difference in the optical such systems cannot be studied or even found at larger distances and are very likely underrepresented in our WD samples. If the white dwarfs in these systems are hot like Sirius B, they can be detected by the excess radiation in the UV or EUV. Subsequent attempts with HST to resolve these systems have shown that most of them have separations (and therefore periods) too large to lead to orbit determinations in the nearer future, but a few promising cases remain (Barstow et al. 2001). The contribution of these systems to the total number of WDs is uncertain, but because of the increase of the IMF toward smaller masses probably much smaller than that of WDs with low mass companions. If the definition of “Sirius-like” includes all MS companions with spectral type of K or earlier, then the contribution to the total WD number in the local sample is 8% (Holberg 2009).

Binaries are much more easily detected, if the companion is a dM star of similar brightness as the WD. Indications are colors that are much too red for a white dwarf, infrared excesses, contributions with molecular features in the red part of the optical or near IR spectrum, or emission features (from the dM) in the cores of the Balmer absorption lines. The SDSS is a rich source of such systems, with several thousands identified already.

These binaries were initially main sequence binaries in which the more massive star evolved through the giant phase and became a white dwarf. In most cases the separation of the main sequence binary was wide and the evolution of the WD is essentially the same as that of a single star. For those with smaller separation the more massive star may engulf its companion while evolving into a red giant, and the system enters the common-envelope phase, with transfer of mass and angular momentum leading to the ejection of the envelope. The study of the exposed post-common-envelope binaries is the best empirical test for the details of the common-envelope phase and the subsequent evolution to shorter periods by magnetic braking.

If the separation becomes small enough, with periods of a few hours, the binary will become a cataclysmic variable, with another mass transfer phase from the secondary to the WD. These systems are discussed in more detail in the ➤ [Chap. 13](#).

7.5.1 Double Degenerates

Binaries consisting of two WDs are interesting as a possible source of Supernova Ia explosions. If they are close, with periods below 10 h, they will lose angular momentum through the radiation of gravitational waves and merge within a Hubble time. With a combined mass larger than the Chandrasekhar limit of $1.4 M_{\odot}$, that merger would result in a supernova explosion. The recent search “Supernova Ia Progenitor Survey” (SPY, Napiwotzki et al. 2001) for radial velocity variations in apparently single WDs turned out ≈ 100 new such pairs, but only one or two fulfill the conditions as SNIa progenitor. For all others, the periods are too large to lead to a merger, or the combined masses are below the Chandrasekhar limit. These lower mass binaries might explain the high mass peak in the WD mass distribution around $0.9 M_{\odot}$ (Vennes 1999). The SNIa scenario favored by most authors now is the single degenerate channel, with a white dwarf accreting mass from a non-degenerate companion.

Double-degenerates can also be detected in a color-magnitude diagram, because they seem too bright for their color. Assuming a single star, the radius determined is large, indicating an apparent very low mass.

8 Suggestions for Further Reading and Useful Resources

More than 600 research papers on white dwarfs are published every year, and not all of their areas of research are covered in this chapter. Moreover, because of space limitations, only very few references are given in the text, and the selection is highly biased. Sometimes historically important papers are cited, sometimes more recent papers, because they give the latest results, or a good summary of the previous literature. When statements are made without reference, it should not be assumed that they are ideas of this author. To correct for this bias the reader is encouraged to consult reviews or other important papers, a selection of which is given here.

General aspects of physics and astrophysics of WDs are discussed in Hansen (2004) and Koester and Chanmugam (1990). The latter covers all aspects of the interior physics, evolution, and atmospheres, and a brief historical review. The former concentrates on “cool white dwarfs,” but covers much of the physics relevant for all WDs. Observational data and analysis of cool white dwarfs is also the topic in Hansen and Liebert (2003) and Bergeron et al. (2001).

An overview of white dwarf evolution with special emphasis on the applications of white dwarfs to determine ages of stellar systems (“white dwarf cosmochronology”) is Fontaine et al. (2001). Among papers to read for the topic of variable WDs are the reviews by Winget and Kepler (2008), Fontaine et al. (2003), and Fontaine and Brassard (2008).

Kinematics and galactic distribution are discussed in Reid (2005), the local sample of white dwarfs in Sion et al. (2009) and Holberg et al. (2008). Magnetic fields are surveyed in Wickramasinghe and Ferrario (2000), Jordan (2009), Tout et al. (2008), Wickramasinghe and Ferrario (2005), Liebert et al. (2005c), Ferrario and Wickramasinghe (2005).

Further reading for WD-MS binaries are Silvestri et al. (2007), Rebassa-Mansergas et al. (2009), Augusteijn et al. (2008), for double degenerates Napiwotzki et al. (2007) and Parthasarathy et al. (2007); and for rotation and angular momentum in white dwarfs Berger et al. (2005) and Karl et al. (2005).

The field of WDs with debris disks is so new and fast changing that any reference list will shortly be outdated. Good starting points are currently Farihi et al. (2009), Kilic et al. (2006), Gänsicke et al. (2006), Farihi et al. (2008), Mullally et al. (2007), von Hippel et al. (2007), Debes and Sigurdsson (2002), Jura (2003), Zuckerman (2001).

As a general source of information about the current status of white dwarf research the “European Workshops on White Dwarfs” are highly recommended. This series of meetings was founded in Kiel (Germany) in 1974 and has been held bi-annually with few exceptions since. Proceedings have been published since the meeting of 1990 in Toulouse.

Valuable resources are of course also available in the WEB. The SAO/NASA Astrophysics Data System (ADS) and the SIMBAD database at Centre de Données astronomiques de Strasbourg hardly need mentioning – no astronomer can work without those. For white dwarfs very useful are the collections of evolutionary tracks, mass–radius relations, colors, and other information of the Argentinian group at La Plata

<http://fcaglp.fcaglp.unlp.edu.ar/evolgroup/tracks.html>

and the Canadian group at Montreal

<http://www.astro.umontreal.ca/~bergeron/CoolingModels>

Matt Woods evolutionary tracks for cooling white dwarfs are older, but still widely used. They are available at

<http://www.astro.fit.edu/wood/wd.html>

Many PhD theses with white dwarf or related topics are collected at

<http://whitedwarf.org>

This includes a number of authors, who are well-known white dwarf experts today! The famous McCook&Sion catalogue of spectroscopically confirmed white dwarfs is online at

<http://www.astronomy.villanova.edu/WDCatalog/index.html>

Acknowledgment

I am deeply grateful to my wife Brigitte for her continuous and essential support throughout my career. Even now, years after official retirement, she tells me to do some useful work at the department, instead of staying home all day.

References

- Althaus, L. G., García-Berro, E., Isern, J., Córscico, A. H., & Rohrmann, R. D. 2007, *A&A*, 465, 249
- Althaus, L. G., Panei, J. A., Miller Bertolami, M. M., García-Berro, E., Córscico, A. H., Romero, A. D., Kepler, S. O., & Rohrmann, R. D. 2009, *ApJ*, 704, 1605
- Althaus, L. G., Serenelli, A. M., Córscico, A. H., & Benvenuto, O. G. 2002, *MNRAS*, 330, 685
- Augusteijn, T., Greimel, R., van den Besselaar, E. J. M., Groot, P. J., & Morales-Rueda, L. 2008, *A&A*, 486, 843
- Barstow, M. A., Bond, H. E., Burleigh, M. R., & Holberg, J. B. 2001, *MNRAS*, 322, 891
- Barstow, M. A., Bond, H. E., Holberg, J. B., Burleigh, M. R., Hubeny, I., & Koester, D. 2005, *MNRAS*, 362, 1134
- Barstow, M. A., Good, S. A., Holberg, J. B., Hubeny, I., Bannister, N. P., Bruhweiler, F. C., Burleigh, M. R., & Napiwotzki, R. 2003, *MNRAS*, 341, 870
- Bedin, L. R., Salaris, M., Piotto, G., Anderson, J., King, I. R., & Cassisi, S. 2009, *ApJ*, 697, 965
- Benvenuto, O. G., Althaus, L. G., & Torres, D. F. 1999, *MNRAS*, 305, 905
- Berger, L., Koester, D., Napiwotzki, R., Reid, I. N., & Zuckerman, B. 2005, *A&A*, 444, 565
- Bergeron, P., Leggett, S. K., & Ruiz, M. T. 2001, *ApJS*, 133, 413
- Bergeron, P., Wesemael, F., Fontaine, G., & Liebert, J. 1990, *ApJ*, 351, L21
- Bergeron, P., Wesemael, F., Lamontagne, R., Fontaine, G., Saffer, R. A., & Allard, N. F. 1995, *ApJ*, 449, 258
- Böhm-Vitense, E. 1958, *Zeitschrift für Astrophysik*, 46, 108
- Bondi, H. 1952, *MNRAS*, 112, 195
- Brassard, P., & Fontaine, G. 1998, *J R Astron Soc Can*, 92, 315
- Castanheira, B. G., & Kepler, S. O. 2009, *MNRAS*, 396, 1709
- Catalán, S., Isern, J., García-Berro, E., & Ribas, I. 2008, *MNRAS*, 387, 1693
- Chabrier, G., Brassard, P., Fontaine, G., & Saumon, D. 2000, *ApJ*, 543, 216
- Chandrasekhar, S. 1939, *An Introduction to the Study of Stellar Structure* (Chicago, IL: The University of Chicago Press)
- Chayer, P., Fontaine, G., & Wesemael, F. 1995a, *ApJS*, 99, 189
- Chayer, P., Vennes, S., Pradhan, A. K., Thejll, P., Beauchamp, A., Fontaine, G., & Wesemael, F. 1995b, *ApJ*, 454, 429
- D'Antona, F., & Mazzitelli, I. 1979, *A&A*, 74, 161
- Davis, D. S., Richer, H. B., Rich, R. M., Reitzel, D. R., & Kalirai, J. S. 2009, *ApJ*, 705, 398
- Debes, J. H., & Sigurdsson, S. 2002, *ApJ*, 572, 556
- Dobbie, P. D., Napiwotzki, R., Burleigh, M. R., Williams, K. A., Sharp, R., Barstow, M. A., Casewell, S. L., & Hubeny, I. 2009, *MNRAS*, 395, 2248
- Dominguez, I., Chieffi, A., Limongi, M., & Straniero, O. 1999, *ApJ*, 524, 226
- Dufour, P., Fontaine, G., Liebert, J., Schmidt, G. D., & Behara, N. 2008, *ApJ*, 683, 978
- Dupuis, J., Fontaine, G., Pelletier, C., & Wesemael, F. 1992, *ApJS*, 82, 505
- Dupuis, J., Fontaine, G., Pelletier, C., & Wesemael, F. 1993a, *ApJS*, 84, 73
- Dupuis, J., Fontaine, G., & Wesemael, F. 1993b, *ApJS*, 87, 345
- Edgar, R. 2004, *New Astro Rev*, 48, 843
- Eisenstein, D. J., et al. 2006, *AJ*, 132, 676
- Farihi, J., Barstow, M. A., Redfield, S., Dufour, P., & Hambly, N. C. 2010, *ArXiv e-print 1001.5025*
- Farihi, J., Jura, M., & Zuckerman, B. 2009, *ApJ*, 694, 805
- Farihi, J., Zuckerman, B., & Becklin, E. E. 2008, *ApJ*, 674, 431
- Ferrario, L., & Wickramasinghe, D. T. 2005, *MNRAS*, 356, 615
- Fontaine, G., & Brassard, P. 2008, *PASP*, 120, 1043
- Fontaine, G., Brassard, P., & Bergeron, P. 2001, *PASP*, 113, 409
- Fontaine, G., Brassard, P., & Charpinet, S. 2003, *Ap&SS*, 284, 257
- Fontaine, G., & Michaud, G. 1979, *ApJ*, 231, 826
- Fontaine, G., Villeneuve, B., & Wilson, J. 1981, *ApJ*, 243, 550
- Fontaine, G., & Wesemael, F. 1987, in *Conference on Faint Blue Stars*, 2nd, Tucson, AZ, June 1–5, 1987, *Proceedings Schenectady* (New York: L. Davis Press, Inc.), pp 319–326. Discussion, pp 327, 328
- Gänsicke, B. T., Koester, D., Girven, J., Marsh, T. R., & Steeghs, D. 2010, *Science*, 327, 188
- Gänsicke, B. T., Marsh, T. R., Southworth, J., & Rebassa-Mansergas, A. 2006, *Science*, 314, 1908
- Gatewood, G., & Han, I. 2006, *AJ*, 131, 1015
- Hamada, T., & Salpeter, E. E. 1961, *ApJ*, 134, 683
- Hansen, B. 2004, *Phys Rep*, 399, 1
- Hansen, B. M. S., & Liebert, J. 2003, *ARA&A*, 41, 465
- Holberg, J. B. 2009, *J Phys Conf Ser*, 172, 012022
- Holberg, J. B., Sion, E. M., Oswalt, T., McCook, G. P., Foran, S., & Subasavage, J. P. 2008, *AJ*, 135, 1225
- Iben, I., Jr., Kaler, J. B., Truran, J. W., & Renzini, A. 1983, *ApJ*, 264, 605

- Isern, J., Mochkovitch, R., Garcia-Berro, E., & Hernanz, M. 1997, *ApJ*, 485, 308
- Jeffery, E. J., von Hippel, T., Jefferys, W. H., Winget, D. E., Stein, N., & De Gennaro, S. 2007, *ApJ*, 658, 391
- Jordan, S. 2009, in *Cosmic Magnetic Fields: From Planets, to Stars and Galaxies*, Proceedings IAU Symposium No. 259, eds. K.G. Strassmeier, A.G. Kosovichev & J. Beckman, 369
- Jordan, S., et al. 1998, *A&A*, 330, 277
- Jura, M. 2003, *ApJ*, 584, L91
- Jura, M., Muno, M. P., Farihi, J., & Zuckerman, B. 2009, *ApJ*, 699, 1473
- Kalirai, J. S., Richer, H. B., Reitzel, D., Hansen, B. M. S., Rich, R. M., Fahlman, G. G., Gibson, B. K., & von Hippel, T. 2005, *ApJ*, 618, L123
- Kalirai, J. S., Saul Davis, D., Richer, H. B., Bergeron, P., Catelan, M., Hansen, B. M. S., & Rich, R. M. 2009, *ApJ*, 705, 408
- Karl, C. A., Napiwotzki, R., Heber, U., Dreizler, S., Koester, D., & Reid, I. N. 2005, *A&A*, 434, 637
- Kilic, M., von Hippel, T., Leggett, S. K., & Winget, D. E. 2006, *ApJ*, 646, 474
- Kitsikis, A., Fontaine, G., & Brassard, P. 2005, in *ASP Conf. Ser. 334*, 14th European workshop on white dwarfs, eds. D. Koester & S. Moehler (San Francisco: ASP), 65
- Klein, B., Jura, M., Koester, D., Zuckerman, B., & Melis, C. 2010, *ApJ*, 709, 950
- Knox, R. A., Hawkins, M. R. S., & Hambly, N. C. 1999, *MNRAS*, 306, 736
- Koester, D. 1976, *A&A*, 52, 415
- Koester, D. 1987, *ApJ*, 322, 852
- Koester, D. 2009, *A&A*, 498, 517
- Koester, D., & Chanmugam, G. 1990, *Rep Prog Phys*, 53, 837–915
- Koester, D., & Reimers, D. 1981, *A&A*, 99, L8
- Koester, D., Voss, B., Napiwotzki, R., Christlieb, N., Homeier, D., Lisker, T., Reimers, D., & Heber, U. 2009, *A&A*, 505, 441
- Koester, D., Weidemann, V., & Zeidler, E. 1982, *A&A*, 116, 147
- Koester, D., & Wilken, D. 2006, *A&A*, 453, 1051
- Külebi, B., Jordan, S., Euchner, F., Gänsicke, B. T., & Hirsch, H. 2009, *A&A*, 506, 1341
- Lamb, D. Q., & van Horn, H. M. 1975, *ApJ*, 200, 306
- Lamb, D. Q. J. 1974, PhD thesis, University of Rochester
- Landenberger-Schuh, S. 2005, PhD thesis, University of Tübingen
- Landolt, A. U. 1968, *ApJ*, 153, 151
- Layden, A. C., Sarajedini, A., von Hippel, T., & Cool, A. M. 2005, *ApJ*, 632, 266
- Leggett, S. K., Ruiz, M. T., & Bergeron, P. 1998, *ApJ*, 497, 294
- Liebert, J. 1976, *ApJ*, 210, 715
- Liebert, J. 1988, *PASP*, 100, 1302
- Liebert, J., Bergeron, P., & Holberg, J. B. 2005a, *ApJS*, 156, 47
- Liebert, J., Young, P. A., Arnett, D., Holberg, J. B., & Williams, K. A. 2005b, *ApJ*, 630, L69
- Liebert, J., et al. 2005c, *AJ*, 129, 2376
- Liebert, J., Fontaine, G., & Wesemael, F. 1987, *Mem-SAIt*, 58, 17
- Liebert, J., Wesemael, F., Hansen, C. J., Fontaine, G., Shipman, H. L., Sion, E. M., Winget, D. E., & Green, R. F. 1986, *ApJ*, 309, 241
- Macdonald, J., & Vennes, S. 1991, *ApJ*, 371, 719
- Michaud, G., Charland, Y., Vauclair, S., & Vauclair, G. 1976, *ApJ*, 210, 447
- Miller Bertolami, M. M., Althaus, L. G., Serenelli, A. M., & Panei, J. A. 2006, *A&A*, 449, 313
- Monelli, M., et al. 2005, *ApJ*, 621, L117
- Montgomery, M. H., Klumpe, E. W., Winget, D. E., & Wood, M. A. 1999, *ApJ*, 525, 482
- Mullally, F., Kilic, M., Reach, W. T., Kuchner, M. J., von Hippel, T., Burrows, A., & Winget, D. E. 2007, *ApJS*, 171, 206
- Napiwotzki, R., et al. 2001, *Astronomische Nachrichten*, 322, 411
- Napiwotzki, R., et al. 2007, in *ASP Conf. Ser. 372*, 15th European workshop on white dwarfs, eds. R. Napiwotzki & M. R. Burleigh (San Francisco: ASP), 387
- Panei, J. A., Althaus, L. G., & Benvenuto, O. G. 2000, *A&A*, 353, 970
- Panei, J. A., Althaus, L. G., Chen, X., & Han, Z. 2007, *MNRAS*, 382, 779
- Paquette, C., Pelletier, C., Fontaine, G., & Michaud, G. 1986a, *ApJS*, 61, 177
- Paquette, C., Pelletier, C., Fontaine, G., & Michaud, G. 1986b, *ApJS*, 61, 197
- Parthasarathy, M., Branch, D., Jeffery, D. J., & Baron, E. 2007, *New Astron Rev*, 51, 524
- Pelletier, C., Fontaine, G., Wesemael, F., Michaud, G., & Wegner, G. 1986, *ApJ*, 307, 242
- Provencal, J. L., Shipman, H. L., Koester, D., Wesemael, F., & Bergeron, P. 2002, *ApJ*, 568, 324
- Pyrzas, S., et al. 2009, *MNRAS*, 394, 978
- Rebassa-Mansergas, A., Gänsicke, B. T., & Koester, D. 2009, *J Phys Conf Ser*, 172, 012025
- Reid, I. N. 1996, *AJ*, 111, 2000
- Reid, I. N. 2005, *ARA&A*, 43, 247
- Romanishin, W., & Angel, J. R. P. 1980, *ApJ*, 235, 992
- Salpeter, E. E. 1961, *ApJ*, 134, 669
- Saumon, D., Chabrier, G., & van Horn, H. M. 1995, *ApJS*, 99, 713
- Schatzman, E. 1945, *AnAp*, 8, 143
- Schmidt, G. D., & Smith, P. S. 1994, *ApJ*, 423, L63
- Schmidt, G. D., et al. 2003, *ApJ*, 595, 1101

- Segretain, L., Chabrier, G., Hernanz, M., Garcia-Berro, E., Isern, J., & Mochkovitch, R. 1994, *ApJ*, 434, 641
- Serenelli, A. M., Althaus, L. G., Rohrmann, R. D., & Benvenuto, O. G. 2002, *MNRAS*, 337, 1091
- Shipman, H. L., Provencal, J. L., Hog, E., & Thejll, P. 1997, *ApJ*, 488, L43
- Silvestri, N. M., et al. 2007, *AJ*, 134, 741
- Sion, E. M. 1984, *ApJ*, 282, 612
- Sion, E. M., Greenstein, J. L., Landstreet, J. D., Liebert, J., Shipman, H. L., & Wegner, G. A. 1983, *ApJ*, 269, 253
- Sion, E. M., Holberg, J. B., Oswalt, T. D., McCook, G. P., & Wasatonic, R. 2009, *AJ*, 138, 1681
- Steinfadt, J. D. R., Bildsten, L., & Howell, S. B. 2008, *ApJ*, 677, L113
- Strand, K. A., & Kallarkal, V. V. 1989, in *Lecture Notes in Physics*, Vol. 328, ed. G. Wegner (Berlin: Springer Verlag), pp 413–415. *IAU Colloq.* 114: White Dwarfs
- Sweeney, M. A. 1976, *A&A*, 49, 375
- Tassoul, M., Fontaine, G., & Winget, D. E. 1990, *ApJS*, 72, 335
- Tinsley, B. M. 1974, *PASP*, 86, 554
- Tout, C. A., Wickramasinghe, D. T., Liebert, J., Ferrario, L., & Pringle, J. E. 2008, *MNRAS*, 387, 897
- Tremblay, P., & Bergeron, P. 2008, *ApJ*, 672, 1144
- Unglaub, K. 2008, *A&A*, 486, 923
- van Altena, W. F., Lee, J. T., & Hoffleit, D. 1995a, *VizieR Online Data Catalog*, 1174, 0
- van Altena, W. F., Lee, J. T., & Hoffleit, E. D. 1995b, *The General Catalogue of Trigonometric [Stellar] Parallaxes* (New Haven, CT: Yale University Observatory). [c1995, 4th ed., completely revised and enlarged
- van den Heuvel, E. P. J. 1975, *ApJ*, 196, L121
- Vanlandingham, K. M., et al. 2005, *AJ*, 130, 734
- Vauclair, G., & Reisse, C. 1977, *A&A*, 61, 415
- Vauclair, G., Vauclair, S., & Greenstein, J. L. 1979, *A&A*, 80, 79
- Vennes, S. 1999, *ApJ*, 525, 995
- von Hippel, T., Kuchner, M. J., Kilic, M., Mullally, F., & Reach, W. T. 2007, *ApJ*, 662, 544
- Voss, B., Koester, D., Napiwotzki, R., Christlieb, N., & Reimers, D. 2007, *A&A*, 470, 1079
- Wegner, G., & Reid, I. N. 1991, *ApJ*, 375, 674
- Weidemann, V. 1977, *A&A*, 59, 411
- Werner, K., & Herwig, F. 2006, *PASP*, 118, 183
- Wesemael, F., Greenstein, J. L., Liebert, J., Lamon-tagne, R., Fontaine, G., Bergeron, P., & Glaspey, J. W. 1993, *PASP*, 105, 761
- Wickramasinghe, D. T., & Ferrario, L. 2000, *PASP*, 112, 873
- Wickramasinghe, D. T. 2005, *MNRAS*, 356, 1576
- Wickramasinghe, D. T., Schmidt, G., Ferrario, L., & Vennes, S. 2002, *MNRAS*, 332, 29
- Winget, D. E., Hansen, C. J., Liebert, J., van Horn, H. M., Fontaine, G., Nather, R. E., Kepler, S. O., & Lamb, D. Q. 1987, *ApJ*, 315, L77
- Winget, D. E., & Kepler, S. O. 2008, *ARA&A*, 46, 157
- Winget, D. E., Kepler, S. O., Campos, F., Montgomery, M. H., Girardi, L., Bergeron, P., & Williams, K. 2009, *ApJ*, 693, L6
- Wood, M. A. 1995, in *White Dwarfs, Proceedings of the 9th European Workshop on White Dwarfs* (Berlin/Heidelberg/New York: Springer-Verlag) XII, 348. Also *Lecture Notes in Physics*, Vol. 443, eds. D. Koester & K. Werner, 51
- Yakovlev, D. G., Kaminker, A. D., Gnedin, O. Y., & Haensel, P. 2001, *Phys Rep*, 354, 1
- Zuckerman, B. 2001, *ARA&A*, 39, 549
- Zuckerman, B., & Becklin, E. E. 1987, *Nature*, 330, 138
- Zuckerman, B., Koester, D., Melis, C., Hansen, B. M., & Jura, M. 2007, *ApJ*, 671, 872

12 Black Holes and Neutron Stars

Thomas J. Maccarone

School of Physics and Astronomy, University of Southampton,
Southampton, Hampshire, United Kingdom

1	<i>Introduction</i>	615
2	<i>Neutron Stars</i>	616
2.1	Structure of Neutron Stars	617
3	<i>Black Holes</i>	617
4	<i>X-Ray Binaries</i>	619
4.1	Classes of X-Ray Binaries	619
5	<i>The Basics of Accretion</i>	620
5.1	Geometrically Thin Accretion Disks	622
5.2	Comptonization Regions	622
5.3	Reprocessing Components	623
5.4	Emission from Relativistic Jets	624
5.4.1	Emission from the Neutron Star Surface and Boundary Layer	624
5.5	Variability Measurement Techniques	625
5.6	Disks: Formation and Stability	626
6	<i>High Magnetic Field Neutron Star Accretors</i>	627
6.1	Low Magnetic Field Neutron Star Accretors	628
6.2	Black Holes in X-Ray Binaries	629
6.3	Spectral States	629
6.4	Jets and Spectral States	631
6.4.1	Outbursts and Spectral State Transitions	633
6.5	Mass Estimates	634
6.6	Constraints on the Neutron Star Equation of State	635
6.7	Distance Estimates	636
6.8	Key Phenomenological Differences Between Black Holes and Neutron Stars	637
6.9	Evidence for Black Hole Spin	638
6.10	Isolated Black Holes	640
6.11	Neutron Stars not in Accreting Binaries	640
6.12	Pulsars and Other Rotation-Powered Emitters	641
6.12.1	Pulsar Emission Mechanisms	641
6.13	Millisecond Pulsars	642
6.13.1	Anomalous X-Ray Pulsars and Soft Gamma Repeaters	644
6.13.2	Old Isolated Neutron Stars	644

7 *Populations of Compact Objects* 644

7.1 Observations 644

7.2 Extragalactic X-Ray Binaries, Intermediate Mass Black Holes, and Other
 Unusual Objects 645

8 *Formation of Compact Objects and Compact Object Binaries* 647

8.1 Black Hole and Neutron Star Formation in Supernovae and Related Events 647

8.2 Natal Kicks 649

9 *Conclusions* 649

References 649

Abstract: Black holes and neutron stars are the compact remnants of massive stars. They represent one of the key intersections between astronomy and fundamental physics – both are classes of objects which are sufficiently compact that Newtonian gravity cannot be used to describe the forces near their surfaces. The structure of neutron stars furthermore presents one of the few key tests of the equation of state of nuclear matter. This chapter will review some of the key theoretical results underpinning the current understanding of neutron stars and black holes. It will also describe the observations of neutron stars and black holes both in isolation (as thermal emitters and as radio pulsars for the case of neutron stars) and in close binaries, where accretion processes can make these objects bright X-ray sources. Additionally, this chapter will detail the formation processes of both neutron stars and black holes in general, and also the formation of close binaries containing such objects.

Keywords: (Including multiple): close, Pulsars: general, Radio continuum: stars, (Stars:) binaries, Stars: neutron, X-rays: binaries, X-rays: bursts

1 Introduction

The observational study of neutron stars and black holes has developed almost entirely in the past 50 years; the articles for the first edition of “Stars and Stellar Systems” were set to press just as the first cosmic X-ray sources were being discovered. These soon came to be understood to provide the first observational evidence for neutron stars and black holes in the Galaxy. More than 5 years later came the first publication of the discovery of radio pulsars. Some observational evidence for the existence of supermassive black holes had been found already by that time [for example, the discovery of strong, extremely broad optical emission lines from some galactic nuclei, seen by Seyfert in 1943, and an optical jet being ejected by the supermassive black hole in M87 has been known even further back (Curtis 1918)]. However, the interpretation of the active galactic nuclei as accretion onto black holes did not gain traction until well into the 1960s (Salpeter 1964) nor did it achieve a consensus until the 1970s. Over the past several decades, the wealth of data from across the wavelength spectrum, from radio through gamma-rays, has greatly increased the level of understanding of how matter behaves near compact objects.

Black holes and neutron stars are astrophysically important objects for a variety of reasons. Both classes of objects represent the end stages of stellar evolution of massive stars – the stars that form the bulk of the heavy elements in the Universe. Understanding the formation of black holes and neutron stars is thus an essential part of understanding how the Universe is chemically enriched to the point in can produce life. After all, matter locked up in a black hole or a neutron star represents core material which has not been injected into the interstellar medium.

Both classes of compact objects also represent objects which can be used to test ideas about fundamental physics. Black holes can be used to determine whether general relativity is, indeed, the correct formulation of gravity. General relativity is also important for understanding the structure of neutron stars, and neutron stars also provide one of Nature’s opportunities to study the properties of matter at nuclear density. Both pulsars and black hole X-ray binaries have been associated with high-energy gamma-ray emission, and, more tenuously, with cosmic rays – they may thus also serve as particle accelerators more powerful than any yet built on Earth. These objects are also strong candidates for producing the low-energy positron background seen from the Galactic Center region.

Comparisons between the accretion process onto stellar-mass black holes and that seen onto the supermassive black holes in active galactic nuclei can yield important information about which parts of the accretion process are “scale-free,” and which show significant differences as a function of the accretor mass. There are also facets of the accretion process which can be studied only with stellar-mass black holes because of the higher signal-to-noise ratio one can obtain from them and because the characteristic timescales of accretion are generally proportional to black hole mass, making some variability features inaccessible in supermassive black holes on timescales shorter than the lifetimes of typical astronomers. Accretion onto neutron stars can be used to compare with accretion onto black holes. Since the mass differences between neutron stars and stellar-mass black holes are relatively small and the gravitational potential well depths are nearly the same for both cases, the differences can often be ascribed to the presence of a surface and/or a magnetic field in the neutron star accretors.

More speculatively, both neutron stars and black holes may be important sources of gravitational radiation. Mergers involving two black holes, two neutron stars, or a black hole and a neutron star are among the few candidate classes of event that should be able to produce gravitational radiation both strong enough and in the right frequency range for detection by ground-based gravitational wave detectors. Pulsar binaries already represent the best evidence to date for indirect gravitational radiation as their excellent clocks can be used to measure their orbital period derivatives to a precision greater than other techniques.

This article will start with an introduction of the general properties of neutron stars and black holes, followed by a discussion of the mechanisms which allow them to cause the emission of radiation (either directly from neutron stars themselves, or from accretion disks surrounding compact objects from either class), and will close with a discussion of their formation and evolution. While this approach may seem counterintuitive, given that, obviously, a compact object must form before it can emit, an understanding of the emission processes from compact objects is necessary for understanding the selection effects that come into population studies of these classes of objects.

2 Neutron Stars

The first suggestion that neutron stars might exist came from the work of Baade and Zwicky (1934), who showed that the total energy radiated in a supernova was about 10^{51} ergs and argued that one of the few plausible sources of such an amount of energy would be from collapsing a star down to a very small radius, at which point it would have nuclear density, and consist primarily of neutrons. It has since been realized that the total energy release from supernovae is dominated by neutrinos, rather than radiation or kinetic energy, and hence that the total energy release from a supernova is about an order of magnitude larger than Baade and Zwicky had estimated. On the other hand, it has become more widely accepted that the means by which this energy is generated is the gravitational collapse of the core of a massive star into a neutron star or a black hole.

Neutron stars are upheld by degeneracy pressure between nucleons (and, perhaps, free quarks) in their cores. The term “neutron star” implies that these objects are composed almost entirely of neutrons. However, there is strong evidence that these stars contain thin envelopes of normal matter, and there are speculative suggestions that these stars may contain substantial components of other more exotic forms of matter, such as free quarks. The density inside a neutron star is large enough that protons and electrons can combine to form neutrons.

2.1 Structure of Neutron Stars

A full understanding of the structure of neutron stars requires an in-depth knowledge of nearly all areas of modern physics. Neutron stars have combinations of masses and radii that leave them within a factor of a few of becoming black holes. Their structure depends heavily on the interactions between nucleons at densities slightly higher than those in atomic nuclei. Their inner cores are often thought to be superconducting, and in some cases, they may contain substantial numbers of “free” quarks.

One of the key reasons that studies of neutron stars are important is that the neutron star equation of state places important constraints on how matter behaves at nuclear densities. Equations of state are referred to as “hard” if they have a weak dependence of radius on mass, and “soft” if they have a stronger dependence of radius on mass. Both the maximum and minimum masses of neutron stars are of considerable interest. The maximum mass of a neutron star is determined by the stiffness of the equation of state of the neutron star. An equation of state is referred to as soft if density depends strongly on pressure, while it is referred to as stiff if density depends weakly on pressure. Soft equations of state therefore can be ruled out if high-mass neutron stars can be observed.

The minimum mass of a neutron star comes from the requirement that β -decay of neutrons in the star is not effective. Under normal densities, free neutrons *will* decay via β emission. However, at very high matter densities, all electron energies levels up to the energy of the electron to be emitted can be filled (see, e.g., Shapiro and Teukolsky 1983 reference work for a more detailed discussion). If the density of a star is below this critical density, it cannot contain the number of free neutrons seen in neutron stars; the neutrons will decay via β emission into electrons and protons. The star will then be supported by electron degeneracy pressure rather than nucleon degeneracy pressure. Under normal conditions of stellar evolution, this minimum mass for a neutron star is not physically relevant – the minimum mass for a neutron star is smaller than the maximum mass for a white dwarf, and normal stellar evolutionary processes will not pass through the region of parameter space with very small radius and medium core mass, where a low mass neutron star might exist. However, under some circumstances, binary evolution might produce an object which starts out as a neutron star but eventually falls below the minimum neutron star mass. This could happen, for example, if the neutron star is in a close binary with either a black hole or a heavier neutron star, and is transferring matter onto that object by Roche lobe overflow (see below). This transformation would be likely to result in a neutrino-rich explosion.

3 Black Holes

Black holes are in many ways Nature’s simplest objects. Like fundamental particles, they can be described by only three numbers – charge, mass, and spin. Unlike fundamental particles, only their charges are likely to be effectively quantized – the spins and masses of black holes should reside in the fully “classical” limit. Charge of black holes is generally neglected for astrophysical purposes as a charged black hole should quickly acquire a particle of the opposite charge and neutralize itself.

A few key results from general relativity are important for understanding observable effects seen from the region near black holes. In general relativity, gravitational redshifts exist – changes

in the wavelength of light for observers located at different depths in gravitational potential wells. A black hole is an object which is sufficiently dense that it has a surface where the gravitational redshift becomes infinite. This surface is called the event horizon.

The equations for a nonrotating black hole were solved by Karl Schwarzschild in 1916, relatively soon after Einstein published his theory of general relativity. As a result, nonrotating black holes are often referred to as “Schwarzschild black holes.” The radius of the event horizon is called the Schwarzschild radius and is $R_{SCH} = 2GM/c^2$. A related scale parameter called the gravitational radius, defined to be GM/c^2 , is often used in relativity. It is an interesting curiosity that this expression can be obtained correctly, albeit through incorrect reasoning, by setting the Newtonian escape velocity equal to the speed of light and solving for radius.

More than about 20 Schwarzschild radii from a black hole, the general relativistic corrections to Newtonian gravity are quite small. However, close to a black hole, there are some fundamental differences between general relativity and Newtonian gravity, apart from just the gravitational redshift. One important such effect with observable consequences is that there exists an innermost radius, inside which no stable circular orbits can exist. It is generally presumed that emission from an accretion disk will also be truncated at this radius as the orbits will be “plunging orbits” over which viscosity is unlikely to be important. Some debate has existed in recent years about whether magnetic coupling across this boundary could lead to significant emission from inside the innermost stable circular orbit, but the most recent work on the topic indicates that such effects are likely to be quite small (Shafee et al. 2006).

Rotating black holes have presented a greater challenge to theorists. The solution to Einstein’s equations for rotating black holes was first presented by Kerr (1963) – it is called the Kerr metric, and rotating black holes are often referred to as “Kerr black holes.” In fact, given the difficulty in obtaining the solution, it is surprising how simply one can express the key observable results of the Kerr metric. The dimensionless angular momentum a is defined to be the angular momentum J divided by Mc . If one expresses the Schwarzschild radius as r_s in gravitational radii, then the event horizon of a Kerr black hole can be given as $\frac{r_s + \sqrt{r_s^2 - 4a^2}}{2}$. Therefore, as a approaches unity, the event horizon radius approaches the gravitational radius.

A second important characteristic radius for Kerr black holes is the radius of the “ergosphere.” The ergosphere is defined as the region within which a stationary particle would not experience a proper time. Within this region, particles are “dragged” by the inertia of the black hole to rotate in the same direction as the black hole. The radius of the ergosphere is given by $\frac{r_s + \sqrt{r_s^2 - 4a^2 \cos^2 \theta}}{2}$.

Finally, we can consider the innermost stable circular orbits around Kerr black holes. As a result of the inertia effects of the black hole on the spacetime around it, the innermost stable circular orbit differs for prograde and retrograde rotation. For prograde rotation about a maximally rotating black hole, the innermost stable circular orbit is at the event horizon, while for counter-rotating orbits, the innermost radius is at nine gravitational radii.

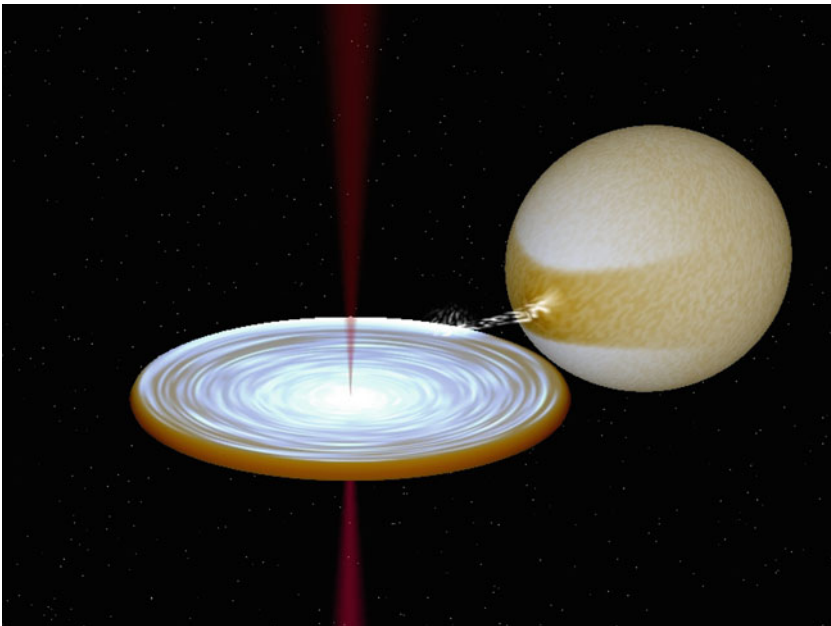
The case of rotation with an axis which is misaligned from the spin axis of the black hole is an interesting one. Exactly what happens in such circumstances is rather difficult to calculate. For a viscous disk of gas, there should be some radius within which the inertial frame dragging due to the black hole’s spin forces the accretion disk to become warped, such that it is forced to rotate with an axis closer to that of the black hole’s at smaller radii (Bardeen and Petterson 1975). The details of how that happens depend not just on general relativistic effects, but also on more complicated disk physics.

4 X-Ray Binaries

Much of what is currently known about neutron stars, and nearly all that is currently known about stellar-mass black holes comes from observations of X-ray binaries (additional knowledge about black holes comes from studies of active galactic nuclei, covered in another volume of this series). An X-ray binary is a system containing either a neutron star or a black hole in a close binary system with a “normal star” (including, in some cases, white dwarf donor stars and objects of insufficient mass to support hydrogen fusion). The system must be close enough for mass to be transferred from the “normal” star onto the black hole or neutron star. • [Fig. 12-1](#) shows a computer-generated visualization of an X-ray binary.

4.1 Classes of X-Ray Binaries

X-ray binaries can be classified both by the nature of the compact accretor (black hole or neutron star) and by the nature of the donor star. There exists a broader variety of classes of donor stars than accretors. Two broad classes, high-mass X-ray binaries (with mass donors of $> 8M_{\odot}$) and low-mass X-ray binaries (with mass donors less than about $1 M_{\odot}$) have been used broadly in the literature. This classification obviously ignores the existence of X-ray binaries whose donor



■ Fig. 12-1

A computer-generated visualization of an X-ray binary, made using Rob Hynes' BinSim tool (Hynes 2002). The figure illustrates the key features of accreting binaries – a deformed donor star, overflowing its Roche lobe and transferring material through an accretion stream into an accretion disk, with some matter and power expelled from the system in the form of a jet

stars are in the intermediate range of $1\text{--}8 M_{\odot}$, a class whose membership in catalogs of known objects has expanded significantly in the past decade or so. These systems are sometimes classified as low-mass X-ray binaries, but in other cases, they are classified, more appropriately, as intermediate-mass X-ray binaries.

As is traditional when astronomers break objects into two classes, the need for subclasses of X-ray binaries has become apparent. Both the high-mass and low-mass X-ray binaries can contain either black holes or neutron stars as accretors. Additionally, there are subclasses based on the specifics of the type of donor star, and properties of the accretion.

5 The Basics of Accretion

It is first noted that much of the material in this section is drawn from the excellent reference work by Frank et al. (1995), which covers the basics of accretion in considerable detail. There are two major processes by which accretion can take place. The first is Roche lobe overflow, where matter spills from a donor star onto an accretor, because some of the matter in the outer regions of the donor star is more strongly gravitationally attracted to the accretor than to the donor. The second is wind accretion, where a star captures material being ejected from another star.

Let us first consider the case of Roche lobe overflow. The Roche lobe of a star is a roughly teardrop-shaped surface of equipotential in a binary star system, calculated by taking into account both the gravitational and rotational effects on the potential. The inner Lagrangian point, L_1 , is the location where the two Roche lobes meet. It is a saddle point for the Roche potential. When a star is larger than its Roche lobe matter will be transferred from it onto its binary companion through the L_1 point.

A few key rules of thumb regarding Roche lobe overflowing systems are important. First, there are a few well-established analytic approximations for the sizes of Roche lobes:

$$\frac{R_2}{a} = \frac{0.49q^{2/3}}{0.6q^{2/3} + \ln(1 + q^{1/3})}, \quad (12.1)$$

which comes from Eggleton (1983), and $R_2/a = \text{Max}(0.462[1/(1 + q)]^{1/3}, 0.38 + 0.2\log_{10} q)$, which comes from Paczynski (1971). In both formulae, $q = M_2/M_1$, where M_2 is the mass of the donor star, and M_1 is the mass of the accretor. Eggleton's formula is accurate to higher precision over a wider range of mass ratios.

Paczynski's formula, at least for low-mass ratios, where the first term will dominate, is more convenient for use in analytical approximations. One of the more well-known and useful applications of Paczynski's formula is the derivation of the mass-period relation for X-ray binaries. For low-mass main-sequence stars, it can be shown that the mass of the donor star in a Roche lobe overflowing system will be 0.11 times the orbital period in hours. The orbital periods of binary systems are generally much easier to measure than the masses of their secondary stars. It should be noted that evolved stars will be considerably less dense than main-sequence stars of the same mass, so the period-mass relation should normally be taken as a useful means of estimating an upper limit to the donor star mass.

Accretion will not take place merely because a star fills its Roche lobe; it must overflow the Roche lobe. The accretion process is thus continually driven by three processes in Roche lobe overflowing binaries, with the dominant process depending on the mass and evolutionary state of the donor star. For relatively high-mass stars, the nuclear evolution of the donor star leads to its expansion and, hence, to continued accretion.

In the absence of substantial evolution in the radius of the donor star, angular momentum loss is required to keep the two stars in contact. For stars in very short-period orbits, gravitational radiation will provide a substantial loss of angular momentum. This can be especially important in systems where the mass donor is a white dwarf star, since white dwarfs can be quite massive even for very short orbital periods. For stars in intermediate period orbits, magnetic braking will be the dominant source of angular momentum loss. Magnetic braking occurs when the wind of the mass donor is forced to move through its magnetic field. The details of how the magnetic braking process works and what is an appropriate prescription for the magnetic braking angular momentum loss are still areas of active research.

The other major process by which mass transfer can take place is through wind accretion. Typically, for X-ray binaries, wind accretion comes from the winds of massive stars. Among these systems, the overall number of X-ray binaries where the donor star is a Be star, rather than a more “normal” O/B star dominates. In any given epoch, however, the majority of the wind-fed X-ray binaries will be those fed by the more spherically symmetric winds of massive stars, rather than from the excretion disks of Be stars. There also exists a small class of objects called symbiotic stars, where accretion takes place from a red giant donor’s wind. A much larger population of symbiotic stars with white-dwarf accretors is known to exist as well.

The problem of wind mass transfer can be modelled in a manner very similar to the Bondi–Hoyle accretion problem. The compact object in a binary system is assumed to accrete any gas that becomes gravitationally bound to it. This leads to a rather strong dependence of the mass accretion rate on the wind speed of the donor star and, hence, to large uncertainties in the predicted accretion rate.

Once mass transfer has started, an important question is why the gas falls in toward the compact object. Early models of accretion disks posited the existence of a form of viscosity that could lead gas to fall inward. They parameterized the viscosity in terms of a parameter α , such that:

$$\nu = \alpha c_s H, \quad (12.2)$$

where ν is the viscosity, c_s is the sound speed in the gas, and H is the scale height of the accretion disk.

The magnetorotational instability (Balbus and Hawley 1991) is the most commonly invoked mechanism for producing this viscosity. This instability requires only a small vertical magnetic field to work. The magnetic forces act toward forcing rigid body rotation in the disk. The effect, then on a fluid element which is displaced outward is that the magnetic field forces it to rotate faster than the Keplerian speed at its new radius, pushing it still further outward. This hence enables a small amount of gas to transport a large amount of angular momentum outward through the disk, allowing the rest of the gas to flow inward.

The efficiency of accretion is an important parameter. Apart from matter-anti-matter annihilation, accretion onto black holes and neutron stars is the most efficient process in the universe in terms of the amount of energy that can be produced per unit matter. Nuclear fusion can produce about 9 MeV per nucleon, if one considers the fusion of hydrogen all the way into iron. If one defines the efficiency of energy production of a process, η to be $E/(Mc^2)$, this corresponds to an efficiency of about 10^{-2} for nuclear fusion. For fusion of hydrogen into helium, the efficiency is a factor of about 10 lower than this value. In the Newtonian limit, the efficiency of accretion is $GM/(R_*)$. For a white dwarf, this yields a value of about 10^{-4} – smaller than the efficiency of nuclear fusion – while for accretion onto a neutron star, this yields about 0.2. However, only half the gravitational energy of a Keplerian disk is radiated, with the other half going into the increase in the rotational energy of the material as it falls inward.

At the surface of the neutron star, a boundary layer is formed. Some of the rotational energy in the boundary layer will be radiated away, and some will go into increasing the rotational velocity of the neutron star. Therefore, typically, $GM/(2R_*)$ is often taken as the radiative efficiency of a neutron star.

For a black hole, the situation is much more complicated. Newtonian gravity clearly does not apply, and a fully relativistic calculation must be done. It can be shown that the radiative efficiency of a nonrotating black hole's accretion disk will be about 0.06, while that for a prograde disk around a maximally rotating black hole will be about 0.42 (Shapiro and Teukolsky 1983). These values are derived by taking the inner radius of the disk to be the radius at which the innermost stable circular orbit of the disk is found and applying the relativistic corrections due to gravitational redshifts from that radius. However, it bears mention that black holes, due to the presence of the event horizon, need not radiate away all the power released by the gravitational fall of material into them. This is especially true if the gravitational energy is locked up in protons, rather than in electrons. Furthermore, in nearly all classes of accreting objects, there exists evidence for jets launched from somewhere in the inner accretion flows. There are some indications that the jets may, in some cases, carry away more kinetic power than the radiated luminosity of the accretion flow.

Spectral modelling of X-ray binaries has typically focused (not surprisingly) on their X-ray and γ -ray emission, although in recent years, a growing number of simultaneous or nearly simultaneous monitoring campaigns have been undertaken including radio, infrared, and optical data, in addition to the high-energy data. Several components are frequently required to fit the spectra of X-ray binaries. The physics behind these components will be described in this section, with the details of the spectra of real accreting objects described in [Sect. 6.3](#).

5.1 Geometrically Thin Accretion Disks

The “standard” model of an accretion disk around a relativistic star comes from Shakura and Sunyaev (1973). In the context of the Shakura–Sunyaev model, the X-ray spectrum of a geometrically thin, optically thick accretion disk can be described very well as the sum of a series of annuli, each of which emits as a black body with a temperature scaling as $R^{-3/4}$. Some corrections are made for relativistic effects and for the effects of scattering of the photons in the atmosphere of the accretion disk so that estimating the best-fitting inner disk radius, as naively estimated from model fits underestimates the real inner disk radius, typically by a factor of about 3 (see for example, Merloni et al. 2000).

5.2 Comptonization Regions

It was realized early on in the history of high-energy astrophysics that some X-ray binaries showed power in the soft γ -rays at luminosities far in excess of what could be explained from purely thermal models for energy production, and that some sources would switch between being dominated by soft X-rays and dominated by soft gamma-rays (e.g., Tananbaum et al. 1972). A natural explanation of such emission is Compton scattering in a geometrically thick, optically thin cloud of hot gas (Thorne and Price 1975). Such clouds can be maintained as hot if there is weak Coloumb coupling between the protons and electrons in the plasma so that most of the energy is advected by the black hole, rather than radiated away (Ichimaru 1977; Narayan and Yi 1994).

Compton scattering can be expected to produce power-law spectra, even if the energies of the electrons off which the scattering takes place do not follow a power-law distribution. Each Compton scattering event will, statistically speaking, result in the multiplication of the photon energy by a constant (presuming that the initial photon energy is small relative to the typical electron energy). This will be $\frac{4kT}{m_e c^2}$ in the nonrelativistic regime, and $(\frac{4kT}{m_e c^2})^2$ in the highly relativistic regime. The number of scatterings will follow $\text{Max}(\tau, \tau^2)$ to good approximation for scattering in a homogeneous, spherical medium, where τ is the optical depth of the medium (there will be some deviations from the latter law due to the fact that Compton scattering has some directional dependence and due to the fact that the scattering medium cannot be infinite in size). This leads to a power-law distribution of photons since the number of photons at high energies will be the fraction of photons at low energies multiplied by the probability that enough amplifications take place to reach the high energy. As the photon energies approach the kT of the electrons, the amplification will become weaker, leading to a cutoff in the energy spectrum.

Additionally, the results above can be used, in some cases, to rule out thermal Comptonization as an emission model. For sufficiently high electron temperatures, $4kT \gg m_e c^2$. This can then lead to substantial amplification of the photon energy per scattering. A result of Compton scattering in such a high-temperature medium would then be to produce “bumps” in the spectrum, rather than a smooth power law. Therefore, a smooth power-law spectrum with no cutoff in the power law out to energies of ~ 1 MeV is a strong indicator that thermal Comptonization is not the dominant energy production mechanism for the power law (see, e.g., Coppi 1999). It can be, in such cases, that the electrons have a nonthermal energy distribution, as might be expected if they are accelerated in magnetic reconnection events, for example. This may represent a good explanation of what is happening in the handful of systems with power laws out to several MeV, with no discontinuities in the soft X-rays, as would be expected for thermal Comptonization from a $kT \sim 300$ keV plasma.

5.3 Reprocessing Components

Some additional components are often seen in the spectra of X-ray binaries which are best explained as reprocessing of hard X-rays producing in the Comptonizing region by the optically thick, geometrically thin accretion disk. Notably, these include emission lines (the strongest being from iron $K\alpha$ at 6.4–6.7 keV, depending on the ionization state of the disk), and a “reflection bump” which is a continuum component peaking around 30 keV, as well as some edges, primarily due to partially ionized states of oxygen and iron. The first discussions of the effects of photoionization of cold gas by X-ray sources can be found in Basko et al. (1974), while the most commonly used spectral models for this effect are presented in Magdziarz and Zdziarski (1995).

Additionally, in many low-mass X-ray binaries, much of the emission from the geometrically thin, optically thick accretion disk is thought to be reprocessed X-ray emission, rather than energy produced from viscous heating of the disk (in most high-mass X-ray binaries, the optical luminosity of the donor star will dominate over that of the accretion flow components). Support for this idea comes largely from the evidence that the optical luminosities of X-ray binaries increase both with X-ray luminosity and with orbital period, following well what is expected for a model in which the optical emission is reprocessed X-ray emission (van Paradijs and McClintock 1994), but there exists also additional timing information that indicates this in many cases.

5.4 Emission from Relativistic Jets

Some X-ray binaries show strong radio emission. An excellent review of jets from X-ray binaries is that of Fender (2006). It can be shown easily that the radio emission from X-ray binaries must come from an outflow, rather than from the accretion inflow. The brightness temperature of a radio source can be defined as:

$$T_B = \frac{c^2}{k_b \nu^2} I_\nu \quad (12.3)$$

Note that this is referred to as a brightness temperature because it is the temperature required for the flux to be produced in the Rayleigh–Jeans part of blackbody emitter’s spectrum. Where the brightness temperature is above 10^{13} K, the emission mechanism cannot be unbeamed incoherent synchrotron radiation. One can thus use this argument to infer that the radio emission region in many systems is likely to be considerably larger than the binary separation, thus requiring the emission come from an outflow.

In many systems, the jet emission shows a flat spectrum, all the way from the radio through the infrared. This flat spectrum radio emission is most commonly explained as the emission from a compact, conical jet (e.g., Blandford and Konigl 1979; Hjellming and Johnston 1988). In the context of these models, the flat spectrum occurs due to the superposition of a series of synchrotron self-absorbed peaks, with the self-absorption frequency varying as a function of radius.

In other systems, where the jet power is emitted over a more spatially extended region, the jet spectrum is a $f_\nu \propto \nu^{-0.7}$ power law. This power law is what is expected from synchrotron emission from a distribution of relativistic electrons with a power-law index of 2.4 (i.e., $\frac{dN}{dE} \propto E^{-2.4}$). In a few cases, emission has also been seen from an interaction site far from the X-ray binary, where the jet presumably collided with an overdensity in the interstellar medium.

The historically best well-known radio emitting X-ray binaries are all high-mass X-ray binaries – Cygnus X-1, Cygnus X-3, and SS 433. The latter two, in particular, have very dense stellar winds in the vicinity of where the jet is launched, and it is likely in both cases that the radiative power of the jet is enhanced by interactions with the stellar wind. In SS 433, emission lines are seen in both the optical and X-rays (Margon et al. 1979; Kotani et al. 1994), indicating that the jet is likely to be entraining material from a thick accretion disk.

5.4.1 Emission from the Neutron Star Surface and Boundary Layer

Finally, in the neutron star systems, there can also be emission from either the surface of the neutron star or from a boundary layer very close to it. In some systems, where the accretion rate is high and the boundary layer is optically thick, this emission is rather straightforward and is well fitted by a black-body component in the X-ray spectrum. In some other cases, there exists little evidence for boundary layer emission in the X-rays – although this can indicate either a low-temperature boundary layer, emitting at sub-X-ray energies, or a higher temperature, optically thin boundary layer, merged in with the Comptonizing component.

Emission directly from the surface of a neutron star in an X-ray binary is especially important. It is believed that during outbursts of X-ray transient systems, the neutron star itself can be heated by the accretion energy. Understanding the rate of cooling of the neutron star can be important for understanding the internal structure of the neutron star – if the neutron star’s

internal energy, as estimated from its surface temperature, varies more quickly than the integration of its thermal emission over time, then one needs to invoke other cooling mechanisms, such as, for example, neutrino cooling (see, e.g., Rutledge et al. 2002; Wijnands et al. 2004).

5.5 Variability Measurement Techniques

For a comprehensive review of variability of X-ray binaries, see van der Klis (2006). Variability studies hold an important place in the current state of understanding of black hole and neutron star accretion disks. X-ray binaries frequently show variability with $\sim 30\%$ root-mean squared amplitudes on timescales of seconds and can show variability of factors of about 10^7 on timescales of years. Results from variability studies are, in general, much more constraining than spectral observations. Most spectroscopic results, especially for continuum components, can be explained by a wide range of source geometries. In contrast, many variability phenomena have been well known for decades without satisfactory theoretical explanations. At the same time, calibration of variability data is far easier than calibration of spectroscopic data, meaning that any well-understood results from X-ray timing is likely to be far more robust than an equally well-understood results from X-ray spectroscopy.

A wide range of timescales of variability can be seen from X-ray binaries. A major part of the reason for this is that the range of radii which can play a relevant role in an X-ray binary is larger than in most other classes of systems. Indeed, the radius of a neutron star is about 10 km, and the Schwarzschild radius of a $10 M_{\odot}$ black hole is about 30 km. The light crossing times at these radii represent the shortest timescales on which one would expect to see variability from an X-ray binary. At the same time, the radius of the accretion disk can often be as large as 10^{12} cm, a factor of about 10^6 larger than the smallest spatial scale. The viscous timescale – the timescale on which matter will flow from the outer disk to the black hole’s event horizon or the neutron star’s surface – at this radius can be weeks.

The power spectrum (the squared magnitude of the Fourier spectrum) represents the major tool used in rapid variability studies of X-ray binaries. The power spectra of accretion disks are rather unlike the power spectra of most other astronomical objects and, hence, bear some specific discussion. In most areas of astronomy, time series are searched for periodicities – for example, those of stellar pulsations and of orbits. Searches for periodicities are important in understanding accretion flows, as well, but the power spectra made in the X-rays and in the few cases studied to date of the optical and infrared, time series of X-ray binaries often show broader features. These features can usually be well-fitted by Lorentzian components. If the Q value of the Lorentzian is very small, the feature is generally referred to as a “noise component.” Noise components are also often well fitted with power laws. Higher Q Lorentzians are usually referred to as “quasi-periodic” oscillations. Coherent oscillations have also been found and in all cases to date, these oscillations have been associated with the rotational frequency of an accreting neutron star or the orbital period of the binary system. For longer time-scale variability (i.e., the variability on timescales much longer than seconds), there has generally not been good sampling of most sources, and less rigorous analysis techniques are generally used.

Time lags are another tool used to try to understand X-ray variability in X-ray binaries. One can make light curves in different X-ray energies and then look either with a cross-correlation function or a cross-spectrum for time lags between the different bands. These can, for example, be used to constrain the size of a region in which Compton scattering is taking place as the Compton scatterings will increase the path lengths which must be followed by higher-energy

photons relative to the path lengths followed by the lower-energy “seed” photons. In fact, models which explain the time lags solely through light travel times often fail to reproduce the observed light curves of X-ray binaries in detail (see, e.g., Maccarone et al. 2000), so these lags represent only upper limits on the size of the emission regions in most cases.

5.6 Disks: Formation and Stability

Accretion frequently occurs in accretion disks. However, the fact that accretion takes place in a system does not immediately imply that an accretion disk will form. In some cases, the accretion stream will impact the accretor directly or will be channeled by the accretor’s magnetic field before a disk forms. Disks form only when the angular momentum of the accreted material is larger than that of a circular orbit either at the radius of either the innermost stable circular orbit around a relativistic object or the radius around a less relativistic object. The radius for which the angular momentum of the accreted gas equals the angular momentum of a Keplerian orbit is called the circularization radius. This will typically be quite similar to the outer radius of the disk, but in some cases, the angular momentum transported outward by the accretion process will lead to a somewhat larger outer disk radius. In some cases, there will also be large magnetic fields with which to contend; the high magnetic field neutron star accretors are mostly fed by the fast winds of massive stars, and often the magnetic pressure of the neutron star is larger than the ram pressure of the gas at the circularization radius so that a disk does not form, and instead, what material is accreted by the neutron star is channelled along its magnetic field lines to the magnetic poles. See again, Frank et al. (1995) and references within for more details of disk formation.

Accretion disks around all sorts of compact objects – white dwarfs, neutron stars, and black holes – are subject to instabilities. The most prominent of these instabilities is the ionization instability. Many aspects of disk stability were first studied in cataclysmic variable stars (e.g., Smak 1984). Much can be learned from such studies, but there are some fundamental differences between the properties of CVs and of X-ray binaries. In particular, the irradiation of the outer accretion disk by the inner accretion disk is far more important in X-ray binaries, where the overall efficiency of accretion is orders of magnitude larger (e.g., Dubus et al. 1999). The essence of the irradiation instability model is that the effective viscosity of gas is larger when the gas is highly ionized than it is when gas is largely neutral. Other sorts of disk instabilities may be relevant as well. Several X-ray binaries show precession, also thought to be due to radiation driven warping of the accretion disk.

Most of the known X-ray binaries are X-ray transients, in the sense that they have been seen to vary in X-ray luminosity by several orders of magnitude over the history of X-ray astronomy. They belong, broadly speaking, to two classes, known as soft X-ray transients (in which the accretors are either black holes or low magnetic field neutron stars, and the spectra near the peak of the outbursts are hence usually dominated by emission at about 1–3 keV) and hard X-ray transients (in which the accretors are high magnetic field neutron stars, and the spectra typically peak at about 15 keV). It is also likely that among the low-mass X-ray binary systems thought to be persistent, there are probably some sources in transient systems with long outbursts, rather than being bona fide persistent systems, given that some systems like GRS 1915+105 have been known to be bright and apparently persistent for about half the history of X-ray astronomy.

Hard X-ray transient behavior is seen only from neutron stars in Be X-ray binaries. In these systems, the accretion onto the neutron star is fed by the decretion disk of the Be star. The orbits

of the neutron stars are usually highly eccentric and do not appear to be coplanar with the decretion disk (see the section on neutron star natal kicks below for a discussion of why), and strong X-ray emission is observed primarily during passages through the plane of the excretion disk. The outbursts thus typically occur periodically, although additional outbursting is sometimes seen which is generally attributed to variations in the excretion disk itself.

Soft X-ray transient behavior is seen from a subset of low-mass X-ray binaries, in which the ionization instability acts. Because the radii of black hole accretion disks will be larger than those of neutron star accretion disks at the same orbital period, and because the ionization instability depends on the temperature of the outer accretion disk, it is expected that a larger fraction of black hole X-ray binaries should be transients than neutron star X-ray binaries. It should also be noted that the lowest-mass black holes would be most likely to end up as persistent sources.

Testing this idea is difficult due to selection effects. Neutron star X-ray binaries can clearly be identified in many cases when Type I X-ray bursts (large increases in luminosity for about 100 s, which are thought to be due to thermonuclear runaway on the surface of a neutron star) are seen. Black hole X-ray binaries have no such clear signature in the X-rays. Measurement of a mass is difficult when a source is in outburst since doing so requires spectroscopic observations of the companion star, which is typically swamped by the optical emission from the outer disk during outbursts. Despite these selection effects, there is still reasonably good evidence that all or nearly all of the persistent low mass X-ray binaries have neutron star primaries – most have shown Type I X-ray bursts or some other variability phenomena (e.g., kilohertz quasi-periodic oscillations, discussed below) seen in bursters, but not seen from any dynamically confirmed black hole candidate. There exists one persistent system, 4U 1957+11, which has never shown Type-I X-ray bursts, and has never shown kilohertz quasi-periodic oscillations. It is considered a strong candidate for being a black hole X-ray binary, but is far from confirmed as such.

The behavior of X-ray transients in outburst is complex. A relatively large fraction of the transients follow a “fast rise, exponential decay” profile in time, but a significant fraction show quite different behavior (Chen et al. 1997). A few qualitative statements can be made about the properties of the outbursts as a function of orbital period. Longer period systems have, as a class, higher peak luminosities and longer outbursts than shorter period systems. These can both be explained as a consequence of their having physically larger accretion disks, which have stored up a larger mass reservoir at the time they reach a critical ionization level to start a new outburst.

More recently, using monitoring campaigns made by the pointed instruments on Chandra and XMM, it has been shown that there exists a population of transient X-ray sources whose outburst are quite faint (e.g., Wijnands et al. 2006). The campaigns which have found these objects have been focused either on the Galactic Center region or on globular clusters – regions expected to contain relatively large numbers of mass-transferring binary systems within a field of view of a pointed X-ray instrument. In globular clusters, the evolutionary processes by which X-ray binaries form are very different than elsewhere in the Galaxy, while the Galactic Center region sources are difficult to followup in other wavelengths. There thus remains considerable uncertainty about the properties of X-ray transients which peak at fluxes below the sensitivity of current and past all-sky monitors.

6 High Magnetic Field Neutron Star Accretors

One class of X-ray binaries contains neutron stars whose magnetic fields are sufficiently large as to channel the accretion flows down the magnetic poles. This class consists predominantly

but not entirely of systems with high-mass donor stars. It is generally believed that the accretion process can bury the magnetic field of the neutron star and, perhaps, that it can cause the neutron star's magnetic field to decay (e.g., Cumming et al. 2001). As a result, high-mass X-ray binaries, over whose lifetimes relatively little total accretion occurs, should have predominantly high magnetic field neutron star companions, while only the fraction of low-mass X-ray binaries in which accretion has started recently should have very high magnetic fields.

While the primary distinguishing characteristic of high magnetic field neutron stars is the presence of pulsations, there are properties of the X-ray spectra of this class of objects which also clearly distinguish these objects from low magnetic field neutron star accretors or from black hole accretors. The typical X-ray spectral shape of a high magnetic field neutron star is reasonably well modelled by a broken power law with a very flat spectrum below about 15 keV and a very steep spectrum above about 15 keV.

Additionally, the high magnetic field systems often show cyclotron resonance lines in their spectra. These lines are seen at:

$$E = 11.6 \frac{B}{10^{12} \text{G}} (1 + z_g)^{-1} \text{keV}, \quad (12.4)$$

where E is the energy of the cyclotron line, B is the neutron star's magnetic field, and z_g is the gravitational redshift of the neutron star. As a result, these lines are useful for making estimates of neutron star magnetic fields. The lines tend to be broad because they are usually formed over a finite sized region near the surface of the neutron star, rather than at the surface of the neutron star. See Wilms et al. (2009) for a detailed review of the detections of cyclotron lines from neutron stars.

6.1 Low Magnetic Field Neutron Star Accretors

Most neutron stars in low-mass X-ray binaries have low magnetic fields. There is no definitive evidence for the magnetic field being dynamically important in these systems, but there are suggestions that some of the quasi-periodic oscillations seen in neutron star X-ray binaries may require magnetic effects from the star itself. There is nearly a one-to-one correspondence between a neutron star in X-ray binary having a low magnetic field and its being in a low-mass X-ray binary; only a handful of low-mass X-ray binaries contain X-ray pulsars.

An intermediate class of neutron star accretors is the class of millisecond X-ray pulsars (see Wijnands 2004 for a review). These systems show pulsations with periods in a range similar to that seen from millisecond radio pulsars (i.e., a few hundred Hz) but otherwise show X-ray properties that are similar to those seen in low-luminosity, low magnetic field neutron stars. The known millisecond X-ray pulsars are all short period systems, with very low-mass donor stars. They all have low mean accretion rates. A few recent observations have found intermittent coherent pulsations from neutron star X-ray binaries, clouding the distinction between millisecond X-ray pulsars and other classes of low magnetic field neutron stars (e.g., Casella et al. 2008).

The low magnetic field neutron stars exhibit a phenomenon known as Type-I X-ray bursts, first discovered by Grindlay et al. (1976). These bursts are increases in X-ray luminosity typically lasting a few seconds. They are often, but not always Eddington limited. Their X-ray spectra are typically well fitted by modified blackbody models. These bursts are thought to be caused by runaway thermonuclear burning in a layer on the surface of the neutron star (e.g., Woosley and

Taam 1976; Swank et al. 1977). During the bursts, sometimes coherent oscillations can be seen in the X-ray flux (e.g., Strohmayer et al. 1996). A subclass of thermonuclear bursts, called superbursts, has also been seen. These bursts typically last 3–5 h. They are most frequently associated with runaway carbon burning, but alternatively, it has been suggested that they may be associated with electron capture by protons in the neutron star's atmosphere. At present, superbursts have been seen only from sources with accretion rates from 10–30% of the Eddington rate. See Kuulkers (2004) for a review of the superbursts.

6.2 Black Holes in X-Ray Binaries

Relatively early in the history of X-ray astronomy, some good candidates were found for containing black holes. Traditionally, the level of proof required for astronomers to call an object a black hole is to demonstrate that it produces no fusion-based optical light (i.e., that all its optical light can be well explained as coming from some combination of the accretion disk and the companion star) and to demonstrate that its mass is significantly larger than the maximum mass for a neutron star under any reasonable equation of state. In other cases, a much higher burden of proof is invoked, and it is required that an event horizon be shown to exist. This requirement is nearly impossible to satisfy through electromagnetic observations as an object with a sufficiently large gravitational redshift will be indistinguishable from a bona fide black hole (Abramowicz et al. 2002).

6.3 Spectral States

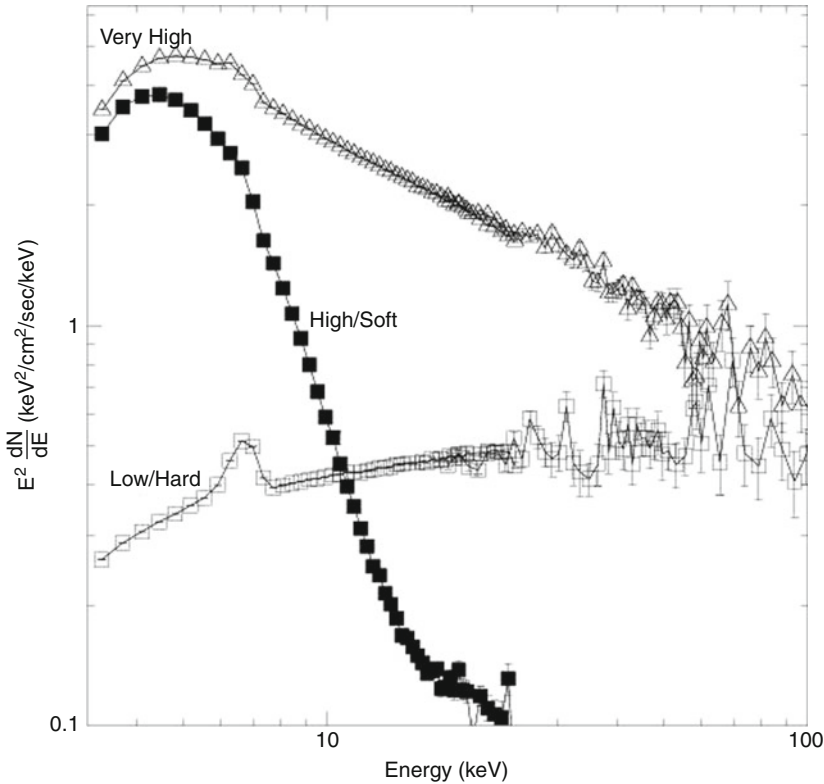
Accreting black holes and accreting low magnetic field neutron stars both exhibit what is known as spectral state phenomenology. Across a wide range of objects, certain patterns in behavior of the spectral energy distribution and the power spectral densities of sources are found to be well correlated with one another. Historically, different nomenclature has been used to describe the spectral states of black holes and the spectral states of neutron stars, but the basic phenomenology of the spectral state behavior seems to be nearly independent of the nature of the compact object.

The state in which an X-ray binary will find itself is mostly a function of the mass accretion rate but also must depend on other parameters. It is customary to discuss luminosities of X-ray sources in terms of the fraction of the Eddington luminosity at which they are emitting. The Eddington luminosity is defined as the luminosity at which the gravitational force inward on a spherically symmetric, fully ionized accretion flow is equal to the radiation pressure outward. The value of the Eddington luminosity is $4\pi GMm_p c/\sigma_t$, where G is the gravitational constant, M is the mass of the accretor, m_p is the proton mass, c is the speed of light, and σ_t is the Thompson cross section. Numerically, the Eddington luminosity is $1.38 \times 10^{38} (M/M_\odot)$ ergs/s. The Eddington limit may be violated if, for example, an accretion flow is not spherically symmetric or if the emission is coming from a region that is moving relativistically toward the observer so that doppler boosting is important.

The current state of understanding of how X-ray binaries behave is best at intermediate fractions of the Eddington luminosity. In particular, at a few percent of the Eddington luminosity, X-ray binaries are frequently found in what is known as a high/soft state. In this state,

the accretion flow is well described by classical models of geometrically thin, optically thick accretion disks. The high soft state generally has a weak but existent power-law tail to the spectrum. In the soft states, typical root mean squared variability amplitudes, with only the two persistent sources, Cygnus X-1 and LMC X-1, showing stronger variability (e.g., Nowak et al. 2001) – these two sources also show the most luminous power laws in the soft state. It has alternatively been proposed to call the soft state the “thermal dominant state” (McClintock and Remillard 2006) and also to refer to the cases with the very weakest power-law tails as “ultrasoft states” (Done et al. 2007). In [Fig. 12-2](#), we show examples of the spectra of an X-ray transient in the canonical spectral states.

At lower luminosities, X-ray binaries are typically found in what is known in the low/hard state. A well-developed phenomenological picture exists describing the behavior of systems in this state, although the nature of the emission geometry remains a topic of ongoing research. For black holes in the low hard state, the X-ray spectrum is typically well modelled by a cutoff



■ Fig. 12-2

The X-ray spectra of XTE J1744-288 in the three canonical states, made using data from the Rossi X-ray Timing Explorer. In the low hard, state, one can see power law emission extending over nearly the full energy range, with possible evidence for a cutoff at the highest energies. In the high soft state, it is clear that nearly all the emission is coming from a quasi-thermal component. In the very high state, both a quasi thermal component (below about 7 keV), and a power law component (at the higher energies) can be seen in the spectrum


power law, plus a contribution from a reflection component, with a contribution from a disk at somewhat lower energies than where it is seen in the high/soft state. The root-mean squared amplitude of variability in the hard state is typically about 30%.

A different phenomenology is seen both during state transitions and during some extended episodes of very high luminosity. In these spectral states, sources are seen to have strong emission from both a quasi-thermal component and a power-law component, typically with a spectral index of about 2.5. Strong variability is seen and, frequently, strong quasi-periodic oscillations are seen. Again, varying terminology is used to describe this state. Traditionally, the term very high state has been used to describe these states at high luminosities, and intermediate state to describe them at lower luminosities (Nowak 1995), while the more recent review article of McClintock and Remillard (2006) proposes calling such states “steep power law” states.

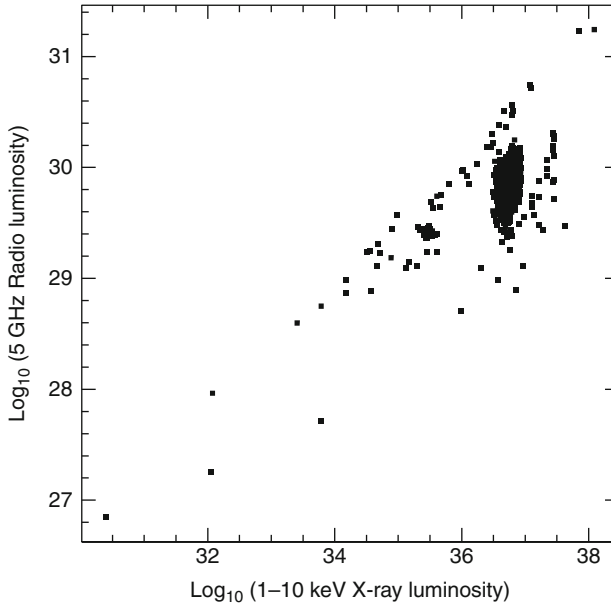
We note that there are different sets of terminology sometimes used to describe the spectral states of neutron stars in the literature. In particular, the neutron stars in low/hard type states are often called “island state” systems and, in high/soft states, are often called “banana” state systems. In their very high states, they are often called “Z-sources.” This terminology comes from color-color plots of neutron star X-ray binaries made with the EXOSAT observatory. The brightest systems traced out a Z-shaped track in the color-color space, while the fainter systems traced out a track that looked like an atoll, with an island-shaped region and a banana-shaped region. At the time, all known non-pulsing neutron star X-ray binaries fell into one of the two categories, so the sources were broken up into the Z-source and atoll source categories. Since that time, a transient neutron star X-ray binary went through both the atoll states and the Z-source states and returned to quiescence, indicating that the atoll and Z classification is primarily a luminosity classification (Homan et al. 2007).

6.4 Jets and Spectral States

The properties of relativistic jets in X-ray binaries are well correlated with the X-ray properties of the systems. For the most part, the strength of jet emission is correlated with the strength of hard X-ray emission. Additionally, the properties of the jets are different for systems with neutron star accretors than they are for systems with black hole accretors.

In hard-state black hole X-ray binaries, the radio emission from X-ray binaries has a flat spectrum due to synchrotron self-absorption, which typically breaks in the range between the near-infrared and near-ultraviolet. It is usually difficult to measure the spectral index of the jet in the non-absorbed part of the spectrum because there is usually only a very narrow range of wavelengths where the jet is both optically thin and significantly brighter than the accretion disk. Most of the black hole X-ray binaries fit on a relation where $L_R \propto L_X^{0.7}$ (Gallo et al. 2003). A few recently discovered sources seem to have radio fluxes significantly below the predictions of this relation, and there are hints at possible parallel tracks in this relation, with similar power-law indices, but different normalization (see, e.g., Fender et al. 2010, whose compilation of data points has been plotted in  Fig. 12-3).

Less information is available about neutron star X-ray binaries in low/hard states. The number of radio detections of low/hard state neutron star X-ray binaries is much smaller than the corresponding number of black hole detections. No radio spectra have been measured for these sources which are of sufficient quality to distinguish between a spectral index of 0.0 and one of -0.7 (the two extremes typically seen from black hole X-ray binaries), but the detection



■ Fig. 12-3

The radio versus X-ray luminosity for a compilation of X-ray binaries, mostly in hard spectral states. The overall trend of $L_X \propto L_R^{0.7}$ can be seen from the data, as well as some considerable scatter. The turndown from this relation at high X-ray luminosities is mostly from data points on Cygnus X-1 as it approaches the high/soft spectral state. This figure uses the same data as that of Fender et al. (2010)

of jet emission in the mid-infrared from the neutron star 4U 0614+091 requires that the spectral index be relatively close to a flat spectrum, given that it has never been seen to be an extremely bright radio source (Migliari et al. 2006). Therefore, it seems likely that the analogy between neutron stars and black holes holds for the spectral shapes of the neutron star jets at low luminosities.

The relationship between X-ray and radio luminosities of neutron star X-ray binaries that best fits the existing data is steeper than the relationship for black hole X-ray binaries – for the neutron star systems, $L_R \propto L_X^{1.4}$ (Migliari and Fender 2006). A consequence of this, plus the generally lower X-ray luminosities of neutron star X-ray binaries is that the neutron star systems are much fainter in the radio and, hence, much harder to detect than are the black hole X-ray binaries. The most likely cause of this different relationship is that the black hole systems advect energy across their event horizons, while the neutron stars have large, optically thin, geometrically thick boundary layers. As a result, in the neutron star X-ray binaries, $L_X \propto \dot{m}$, while in the black hole X-ray binaries, $L_X \propto \dot{m}^2$. If the jet production and radiative efficiency of the jet then scale with \dot{m} , the observed correlations can be produced (Körding et al. 2006).

Jets are often seen in very high states and in state transitions. In these states, the jets are often seen to be very bright, to have optically thin radio spectra, and to have measurable proper motions. A rapid growth in studies of radio properties of X-ray binaries began in the 1990s when the first “microquasars” were seen – that is, X-ray binaries with apparently superluminal

jet proper motions (e.g., Mirabel and Rodriguez 1994). The neutron star systems have also shown similar jets. Circinus X-1, which is a peculiar system likely to have a highly eccentric orbit, in particular, has shown evidence for outflows at close to the speed of light (with apparent superluminal motions) during its periastron passages (Fender et al. 2004).

There are many radio observations made during high/soft states, as well, both for neutron stars and for black holes. Most of these observations are non-detections. In a few cases, black holes have been observed to show radio emission in high/soft states, but these have all been cases where it was possible (and, indeed, likely) that the radio emitting plasma was ejected from the accretion flow during a previous state transition, and its emission was still decaying at the time of the high/soft state observation. The upper limits on other radio emission from high/soft state black holes are as stringent as 50 times the expected flux from extrapolating the hard state radio/X-ray correlation (e.g., Corbel et al. 2004).

There are detections of radio emission from at least two soft-state neutron star X-ray binaries, 4U 1820-30 and Ser X-1 (Migliari et al. 2004). In these systems, the radio flux is the same in the high soft state as it was in the peak of the low hard state. Thus, there is still reduced radio emission from what was expected from the extrapolation of the hard-state correlation, but not reduced so much as to make the source undetectable in the radio. A possible explanation is that the radio emission from neutron star X-ray binaries is related to the boundary layer (Maccarone 2008). A boundary layer provides large-scale height differentially rotating plasma. On the other hand, an example of a neutron star with a stringent upper limits on its soft state radio flux also exists (Tudose et al. 2009), so the presence of radio emission in the high soft states of neutron stars is not a universal difference between neutron stars and black holes.

The high-mass X-ray binaries with accretion powered pulsars have not been detected as radio sources. Because of the truncation of their disks at large radii by the magnetic fields of the neutron stars, it is presumed that these systems are poor locations for jet launching. On the other hand, the searches of these systems for radio emission are, to date, quite shallow. One class of binaries which are likely to consist of neutron stars and high-mass stars has shown strong radio emission. These are “gamma-ray binaries,” LS 5039, LS I +61 303, and PSR B1259-63. It is generally thought that these systems are composed of a rotation-powered pulsar whose pulsar wind interacts with the stellar wind of the high mass star (e.g., Dubus 2006; Dhawan et al. 2006), although models do exist for explaining the emission from this source through interactions between a black hole-powered jet and a stellar wind (Massi and Kaufman Bernado 2009).

6.4.1 Outbursts and Spectral State Transitions

Most known outbursts of transient low-mass X-ray binaries exhibit multiple spectral states. Outbursting sources typically rise in luminosity in the hard state, and somewhere rather close to their maximum fluxes make a transition to the soft state. They then fade within the soft state and make a transition back to the hard state at a lower luminosity than where they transited from the hard state to the soft. This hysteresis effect in spectral state transitions occurs in both neutron star and black hole systems (Maccarone and Coppi 2003). Typically, the transition from the hard state into the soft state takes place near the peak of the outburst, while the transition from the soft state to the hard state takes place at about 2% of the Eddington limit (Maccarone 2003). A subset of low-mass X-ray binary transient events seems to take place in which the

source never enters a soft state. These are all outbursts whose peak luminosities are low and are predominantly from short period systems (Brocksopp et al. 2004).

6.5 Mass Estimates

It is obviously of great importance to have a good understanding of the masses of compact objects in close binaries. Breaking up classes of compact binaries into neutron stars and black holes requires accurate mass estimates so that one can determine which objects have masses larger than the maximum mass for a neutron star, under even the inclusive equations of state.

The first step in determining a compact object's mass is to measure its mass function:

$$f(M_1, M_2) = \frac{(M_2 \sin i)^3}{(M_1 + M_2)^2} = \frac{4\pi^2}{G} \frac{(a \sin i)^3}{P_b^2}, \quad (12.5)$$

where M_1 and M_2 are the masses of the two stars in the system, P_b is the binary orbital period, G is the gravitational constant, i is the inclination angle of the binary, with $i = 90^\circ$ for an edge on binary, and a is the semimajor axis of the binary. The mass function then sets a lower limit on m_2 .

The mass function thus represents a robust lower limit on the mass of an object. For making strong claims that an object is heavier than the maximum mass of a neutron star, the mass function can often be sufficient. Care must be taken to determine where the observed spectral lines are generated, and to ensure that the velocities measured are representative of the velocity of the center of mass of the donor star. Ordinarily, this does not represent a terribly serious issue, but in the case of radial velocity curves measured from emission lines generated in the stellar wind of a high mass donor star, it can be very serious; the X-rays from the X-ray binary may be capable of ionizing the stellar wind in the inner regions of the binary system but not in the regions shielded from the X-rays by the donor star. This can result in a significant overestimate of the radial velocity of the donor star if one does not correct for it. This issue has been discussed in detail for the case of the Galactic wind-fed system Cygnus X-3 (van Kerkwijk 1993), but the basic arguments apply more generally.

For making a reliable census of masses of black holes or of neutron stars, the inclination angle must be estimated in some manner. Eclipses can provide excellent constraints on the inclination angles of systems, and the use of pulsar timing, in the systems where the accretor is a high magnetic field neutron star, can also provide additional constraints on the orbital parameters of the system. For this reason, there exist many HMXB neutron star systems with mass measurements accurate to about 10%, whereas only one black hole, which is in an eclipsing binary, has a mass measurement so precise (Orosz et al. 2007).

In many ways, though, accurate measurements of black hole masses and of the masses of neutron stars in LMXBs are more interesting than accurate mass estimates of the neutron stars in high-mass X-ray binaries. The neutron stars in high-mass X-ray binaries are generally unlikely to accrete much material before the donor star's lifetime ends. If normal stellar evolution leads to the production of neutron stars within a narrow range of masses (i.e., typically just larger than the Chandrasekhar mass), only in low-mass X-ray binaries are the most massive neutron stars likely to be formed. There is even more uncertainty about the masses of black holes. Theoretical work on black hole formation predicts a range of masses (e.g., Fryer and Kalogera 2001).

The inclination angles in low-mass X-ray binary systems are generally estimated using ellipsoidal modulations of the donor stars – the variations in the flux of the donor star as seen from Earth due to the nonspherical shape of the donor star. The sizes of the ellipsoidal modulations seen can then be compared with models based on a Roche geometry of the donor star and converted into an inclination angle. Measurements of the amplitude of the ellipsoidal variations can generally be made quite precisely. A major source of uncertainty can often be the ratio of the flux of the accretion disk and jet in the optical or infrared band used to estimate the size of the modulations. While the measurements are generally made while the black hole or neutron star is “quiescent,” the bolometric luminosity of the accretion flow can often still be a substantial fraction of the luminosity of the donor star.

Eccentric radio pulsar binaries can be used to make mass estimates purely through pulse timing behavior. These systems are typically found only in globular clusters, where stellar encounters can induce eccentricities in binaries faster than the binaries will re-circularize themselves. One needs to measure both the relativistic effect of the advance of the periastron, plus the Keplerian motions in order to measure the masses of both objects in a binary system purely from pulsar timing. The details and results of this technique can be found in the review article of Lorimer (2008).

6.6 Constraints on the Neutron Star Equation of State

Understanding the equation of state of nuclear matter is one of the key goals of neutron star research. There are several means to do this. The simplest is simply to find the most massive neutron stars. The softest equations of state can be ruled out by finding sufficiently massive neutron stars, as in the softest equations of state, the threshold mass at which an object must become a black hole, rather than a neutron star, is lowest. The lack of clear evidence for neutron stars more massive than about $1.8 M_{\odot}$ favors soft equations of state, but as noted above, the low-mass X-ray binaries’ neutron stars are most likely to be the heaviest neutron stars, and the few mass estimates for those objects are generally poorly constrained.

For limiting the range of possible harder equations of state, however, it becomes necessary to limit the available parameter space in the mass-radius relation. This can be done by measuring the gravitational redshift at the surface of the neutron star. At the present time, there exist some controversial measurements of the gravitational redshift of a single neutron star from absorption lines in the X-ray spectrum of a Type I burst (Cottam et al. 2002; Galloway et al. 2009). It can also be done by detailed spectral modelling of the surface emission from Type I X-ray bursts (Özel et al. 2009).

Neutron star X-ray binaries often show “kilohertz quasi-periodic oscillations” – quasi-periodic oscillations at frequencies of about 600–1200 Hz. Combined mass-radius constraints may also be possible if a clear model for kilohertz quasi-periodic oscillations can be found, and the results could be particularly effective for determining the actual masses and radii of specific neutron stars. While considerable debate still exists about what is the correct model for the QPOs (or indeed, whether any currently proposed model is correct), in the long term, the quasi-periodic oscillations are quite likely to provide the strongest constraints on the neutron star equation of state, since the measurements of the QPO frequencies are highly precise (typically accurate to about 1%), and far less susceptible to calibration uncertainties or radiative transfer effects than are spectral signatures.

6.7 Distance Estimates

A variety of techniques are used for measuring distances to black holes and neutron stars in the Galaxy. The most robust of these, the geometric parallax, has been applied for about 20 neutron star sources [one X-ray binaries – Sco X-1 (Bradshaw et al. 1999), about a dozen pulsars (Chatterjee et al. 2009), and one isolated non-pulsing neutron star (Kaplan et al. 2002)]. One black hole's parallax distance has recently been measured as well (Miller-Jones et al. 2009). Apart from parallax distances, there are also cases of firm associations with star clusters and supernova remnants which can be used to make reliable distance estimates. The comparisons between parallax distance estimates and other distance measurement techniques are important for understanding the errors in the other techniques.

Nearly all parallax distance measurements of neutron stars to date have made use of very long baseline radio interferometry, with a much smaller number measured in the optical with the Hubble Space Telescope (Chatterjee et al. 2009). Present instrumentation at other wavelengths lacks the angular resolution needed to make parallax measurements. The ease of measuring the parallaxes of radio pulsars can be improved by “pulsar gating” – taking the known timing solution of the pulsar and making an image only when the pulsar is bright, resulting in a significant reduction of noise with relatively little loss of signal. The limits to radio parallaxes are typically severe at distances of a few kpc, although one parallax measurement has been made of a pulsar at 7.2 kpc. At larger distances, the uncertainties with which one can do phase referencing in radio astronomy – the measurement of positions based on the phase differences seen from different sources – is limited by ionospheric effects. More accurate radio astrometry will be possible with an upgraded bandwidth Very Long Baseline Array because it will allow the use of phase calibrators which are fainter and, hence, closer to the source of interest. Timing parallax measurements can be made for some of the pulsars which act as the most accurate clocks – residuals in the timing solution which correlate with time of year can be used to estimate the position

Where parallax measurements have not been made, the distance measurements techniques generally differ quite strongly between radio pulsars and X-ray binaries. When the distance to a radio pulsar cannot be measured either through a parallax or by an association with a star cluster or other object whose distance is known, dispersion measure distances are often used. The dispersion measure of a pulsar is the number of free electrons between the pulsar and the observer. Dispersion of radio waves as they pass through ionized gas leads to a time delay which is function of photon frequency. By comparing the pulsar's pulse arrival times at different radio frequencies, one can thus measure the column density of free electrons to the pulsar. There exist models of the free electron distribution in the Galaxy which can be used to estimate the distances of these pulsars (e.g., Cordes and Lazio 2003). Distances estimated from dispersion measures are usually accurate to within a factor of about two within the Galactic Plane and fail quite badly for most globular cluster pulsars.

The distances to neutron star X-ray binaries can often be estimated from Type I X-ray bursts. When these bursts show radius expansions over the course of the burst duration, this generally gives a good indication that the burst is Eddington limited. The Eddington limited bursts have often been suggested to be good standard candles, and repeated observations of a large number of Type I bursts in globular clusters have helped to verify this (Kuulkers et al. 2003).

However, not all neutron star X-ray binaries have shown radius expansion bursts, and, of course, no black hole X-ray binaries show Type I bursts. For these objects, usually the best distance estimates come from using the donor star, using the photometric parallax technique.

A crude estimate of the distance to a donor star can be made by measuring the flux of the donor star and comparing it with the luminosity of a star of its spectral type. However, the evolutionary paths followed by the donor stars in low-mass X-ray binaries are often nonstandard, and the stars themselves are always nonspherical and rapidly rotating. More reliable distance estimates come from making an estimate of the radius of the donor star directly from the data, rather than under the assumption that the donor star has the mass and radius typical for its spectral type. The estimation of the stellar radius is most reliably done by measuring the rotational broadening of the lines in the donor star's spectrum, combined with the fact that the star must rotate synchronously with the binary orbit. Absent a reliable estimation of the star's rotational velocity, the relation between the orbital period and the density of the donor star can be used to estimate the donor star's radius – this technique requires an estimate of the donor star's mass, but this can come from the mass ratio between the donor star and the accretor if a rotation curve for the accretion disk can be measured from disk emission lines. The status of distance estimation techniques for X-ray binaries is well discussed in Jonker and Nelemans (2004) and references within.

A few other techniques are sometimes used, but typically place only bounds on the distances to low mass X-ray binaries, rather than giving actual distance estimates. One is the use of two-sided jet proper motions. This technique gives a robust upper limit on the distance to an object based on requiring no motions faster than the speed of light. It has also been shown that sources will typically be at distances very close to the upper limits derived from this technique (Fender 2003).

The other prominently used technique is to look for evidence of absorption, either in the form of interstellar lines such as the Na D line in an optical spectrum or HI absorption in a radio spectrum. One can then use the rotation curve of the Galaxy to estimate the distances of the absorbing material to place a lower limit on the distance to the X-ray source, and somewhat more speculatively, can use the absence of absorption from more distant gas to place an upper limit on the distance (e.g., Hynes et al. 2004). The state transition flux from the high/soft state to the low/hard state can also provide an approximate distance estimator (Maccarone 2003), with an accuracy similar to that from dispersion measure distance estimates of pulsars. While these techniques are less reliable, some of them can provide distance estimates for X-ray binaries with large foreground column densities which make it impossible to make measurements of the donor star.

6.8 Key Phenomenological Differences Between Black Holes and Neutron Stars

The existence of black holes is one of the key predictions of general relativity. A key goal of observational astronomy is then to attempt to prove that black holes do, in fact, exist in Nature. The first step in this process is to find objects heavier than the theoretical maximum mass of a neutron star under equations of state which are not in conflict with nuclear physics experiments. Astronomers have, in fact, found about two dozen such objects, heavier than $3 M_{\odot}$.

The next step in this process is to demonstrate that the objects heavier than $3 M_{\odot}$ are in some ways qualitatively different from those lighter objects, which are likely to be neutron stars, and to demonstrate that the observed differences are related to the existence of a solid surface in the neutron stars, which is absent in the black holes. Considerable progress has been made in recent years on this topic.

The most prominent surface effects of neutron stars are Type I X-ray bursts, pulsations, and surface emission in quiescence. All are evidence of a surface, but in no case is the absence of these effects clear evidence of the absence of a surface. Pulsations are generally absent in low magnetic field neutron stars. There are objects, known to be neutron stars due to observations of Type I X-ray bursts during their outbursts, but which have shown no surface emission in quiescence. This can be explained by enhanced cooling processes, perhaps related to a relatively high neutron star mass (Jonker et al. 2007). Prior to those observations, there had been a relatively clear picture that the quiescent luminosities of low-mass X-ray binaries with neutron stars were systematically higher than those of black holes with the same orbital periods (e.g., Garcia et al. 2001). Type I X-ray bursts are only expected from low magnetic field neutron stars, and only in a certain range of accretion rates. At low accretion rates, it can take a very long time before enough material piles up to allow runaway nuclear burning to take place, leading to a sufficiently low rate of bursts that one cannot expect to detect bursts. At higher accretion rates, steady nuclear burning is expected to take place. On the other hand, in the intermediate accretion rate range where frequent bursts are expected, they are always seen (Remillard et al. 2006). Therefore, the absence of both bursts and pulsations from a well-observed source may be taken as reasonably good evidence for an event horizon. The above-mentioned point, though, that one cannot rule out observationally the possibility that a mass configuration could exist in which a star could have a surface gravitational redshift of $\sim 1,000$, remains (Abramowicz et al. 2002).

6.9 Evidence for Black Hole Spin

In the past few years, a variety of lines of evidence have all started to indicate that most stellar-mass black holes have substantial angular momentum. This is perhaps not surprising since massive stars routinely have angular momenta in excess of the maximum angular momentum for black holes of the same mass. On the other hand, it is difficult to measure the black hole spin with high precision because its effects are generally relatively minor and are often degenerate with other effects.

The observations of relativistic jets have been suggested in some cases to be strong evidence of black hole spin. This is partly motivated by the work of Blandford and Znajek (1977), who argued that jets could be powered in part by tapping the spin energy of the black hole. However, the observations of jets from other classes of objects which have no event horizons, and from neutron stars, which cannot be rotating quickly in terms of specific angular momentum (their break-up speeds are at about 0.1 times the speed they would need to be at to have angular momentum per unit mass similar to maximally rotating black holes).

Several lines of evidence have been proposed in recent years to show that black holes are spinning. These fall, broadly speaking, into two classes: attempts to make precision measurements of the black hole spins, and attempts to show effects which qualitatively demonstrate that the black holes must be rotating, at least at some speed. In the former category are iron-line measurements, disk spectral fitting measurements, and specific modelling of certain quasi-periodic oscillations, while in the latter are other work on quasi-periodic oscillations, and attempts to demonstrate that frame dragging is taking place in accretion disks.

There are two classes of X-ray spectral evidence for spinning black holes. The first is from the use of broad iron lines. Iron fluorescence lines can be generated as a hard X-ray producing component illuminates the cold neutral material in the inner accretion disk. The profile of the iron line will not be a delta function in energy, but rather will be severely affected by the special

relativistic effects due to Doppler motions of the material around the black hole in a rotating accretion disk, as well as by the general relativistic effects of the gravitational redshift due to the fact that most of the iron line flux will be generated very deep in the gravitational potential well of the black hole. Because the orbital motions approach rather close to the speed of light in the inner accretion disk, the Doppler effect not only provide a change in the photon energies, but also provides a strong boosting of the luminosity of the blue wing of the iron line.

The key advantages of using the iron line to fit black hole spin relative to using disk continuum fitting are that the spin measurement is independent of the distance to the black hole and of the mass of the black hole. These advantages remove two of the biggest uncertainties in nearly all measurements of anything for stellar-mass black holes in the Galaxy, which is why proper calibration of this technique is so attractive for understanding black holes.

In detail, the fit results are susceptible to a number of uncertainties, only some of which have been studied in detail in the literature to date. As the iron lines need to be fitted over a very broad range of energies – typically from about 4–8 keV, the continuum must be known to high precision over this energy range, and the instrument being used must be well calibrated within this energy range. The requirements of an accurate continuum model and of an excellent flux calibration in order to study an emission line are rather different than the usual requirements for making reliable, detailed measurements of stellar lines in the optical, but are a consequence of having lines with significant flux over a range of a factor of two in photon energy.

In X-ray binaries in their bright states, the continuum over this range of energies is often affected by the fact that the disk blackbody component and the Comptonized component meet near 4 keV, and the high spectral resolution instruments which can be used to estimate the shape of the iron line usually cut off at about 10 keV, leaving relatively little lever arm for measuring the shape of the power-law component. Additionally, the iron line is accompanied by an iron edge which is also relativistically smeared. Without good coverage of the spectrum up to energies well beyond the edge energy, there can be degeneracies between the shape of the line and the energy of the edge.

There are a few additional free parameters and assumptions in iron-line spin estimates. One parameter is the inclination angle of the inner accretion disk. Another parameter is the emissivity profile of the iron line (i.e., the amount of line emission that comes from the disk as a function of radius). The emissivity profile is usually assumed to be a power law, and its index is fit. The fitting also requires as parameters the inner and outer radius contributing to this emissivity. The ionization parameter of the emission region is usually taken to be a constant, although it is not likely actually to be a constant. Finally, there are assumptions that the disk is planar and tied to the plane of the black hole spin in the inner region. Furthermore, it is also assumed that there are no other components of iron emission, and that the orbits in the accretion disks are circular, rather than elliptical.

Measurements of spins of black holes from continuum spectra have a different set of advantages and disadvantages. From the X-ray continuum, one can attempt to estimate the inner disk radius. However, this method is strongly susceptible to uncertainties in the black hole mass, the black hole distance, and the inclination angle of the inner accretion disk and relies on the assumption that the accretion disk extends in to the innermost stable circular orbit around the black hole. Continuum fitting is susceptible to the same problems with uncertainties in the instrument response matrix that affect iron-line fitting. It is very susceptible to errors in estimating the effects of Compton upscattering of disk photons and, to date, has been attempted only with sources in the ultrasoft state. On the other hand, it can be done with sources which are much fainter than those for which iron-line measurements can be made.

At the present time, the spin estimates from reflection features span the full range of prograde black hole spins possible (Miller et al. 2009). The disk continuum fits also span a large range in a , from about 0.1–0.8 (Shafee et al. 2006; Davis et al. 2006). Unfortunately, at the present time, for the two sources for which both methods have been tried, GRO J1655-40 and 4U 1543-47, only marginal agreement is seen between the two methods, with the iron-line method giving larger spin values.

Using quasi-periodic oscillations to measure black hole spins ultimately has the greatest potential for being used to give reliable and precise spin measurements. Narrow features in power spectra are far easier to calibrate properly than spectral measurements and are not subject to any reprocessing effects (except in that the timing signature can be weakened by photon scattering).

There are two classes of QPO that have been used as evidence for black hole spin. The first is the presence of oscillations at about 1–10 Hz, seen in many black hole X-ray binaries. These have been suggested to be evidence of a Lense–Thirring precession, an effect due to relativistic frame-dragging of tilted accretion disk (e.g., Fragile et al. 2001). Precision spin measurements are not possible from the frequency of the L-T oscillation because the relation between the black hole spin and the QPO frequency will depend rather strongly on details on angular momentum transport in accretion disks.

The other QPOs, which may have potential for providing precision spin measurements, are the high-frequency QPOs seen from several black holes in a 2:3 frequency ratio (e.g., Strohmayer 2001; Abramowicz and Kluzniak 2001). Several mechanisms exist for explaining these QPOs – for example, resonances between relativistic coordinate frequencies (Abramowicz and Kluzniak 2001) and fundamental modes of oscillation of a torus around the black hole (Rezzolla et al. 2003). All QPO-based methods to date can give precision measurements of the black hole spin, but all depend on knowing also the mass of the black hole to high accuracy, and the different models give substantially different spin values. The advantages of using QPOs to measure spin of black holes are then that the method is distance independent, and that the measurements are far less susceptible to non-statistical errors than are the spectral measurements. The key disadvantages are that the measured spins are strongly dependent on both what is the actual correct model and on the black hole mass.

6.10 Isolated Black Holes

Black holes which are not accreting are nearly impossible to detect astrophysically. Searches have been made for evidence of heavy, dark stars in binary systems with no evidence for accretion taking place and have not yielded any strong candidate black holes. A few stellar-mass black hole candidates have been found in microlensing searches (Bennett et al. 2002). It has also been suggested that isolated black holes in very dense regions of the interstellar medium, such as molecular clouds, might be detectable through accretion signatures, either in high energies or in radio (e.g., Armitage and Natarajan 1999; Maccarone 2005).

6.11 Neutron Stars not in Accreting Binaries

Neutron stars, with solid surfaces and magnetic fields, can be detected even when not accreting. There are several classes of such neutron stars known. Two of them, the rotation-powered

pulsars and the magnetars, have well-defined characteristics clearly suggested by their names. Additionally, there exist a few neutron stars whose emission is dominated by thermal cooling of the initially hot star. More recently, two new classes of objects have been proposed, which may, in some cases, be objects which can also fit into the already known classes. The term “isolated neutron stars” is often used to refer to a small subset of the known neutron stars not in binary systems, and the “central compact objects” which are distinguished by their locations near the centers of young supernova remnants, rather than any particular physical properties they possess. A review of the state of the art of all classes of isolated neutron stars can be found in Kaspi et al. (2006). More detailed discussion of the rotation-powered pulsars can be found, for example, in Lyne and Graham-Smith 2006 and of binary and millisecond pulsars in Lorimer (2008).

6.12 Pulsars and Other Rotation-Powered Emitters

While the first neutron star to be observed and identified as an unusual object was the X-ray binary Sco X-1, the early observations of radio pulsars half a decade later represented the first truly strong evidence for the existence of neutron stars. The accurate clocks provided by pulsars have allowed measurements to precision higher than any other class of astronomical measurement – in a few cases, the masses of pulsars are known to such high precision that the largest source of error is in the measurement of GM_{\odot} , the product of the gravitational constant times the solar mass, rather than in any of the measurements specific to the pulsar itself. This timing precision provides for an inexpensive technique for searching for gravitational radiation – looking for correlated residuals in the timing behavior of several pulsars. Pulsars are additionally an important source of radiation across all wavelengths and may themselves be sources of gravitational radiation.

6.12.1 Pulsar Emission Mechanisms

Understanding where the power for pulsars comes from is relatively straightforward. Assuming that the rotation axis of a neutron star is misaligned with the magnetic field of the neutron star, one can calculate the expected power due to dipole radiation. More difficult is to understand how that magnetic dipole power is converted into the radiation observed from pulsars. The rotational frequencies of pulsars range from about 0.2–700 Hz. Even at 700 Hz, the direct dipole emission frequency is well below the plasma frequency of the solar system. Therefore, it cannot be the dipole power itself which is observed from the Earth.

On the other hand, the dipole spin down is thought to dominate total energy loss from rotation-powered pulsars. Since the spin-down power is:

$$\dot{E} = I\dot{\Omega}\Omega = 4\pi^2 I\dot{P}P^{-3} = \frac{\Omega^4 R^6 B_0^2 \sin^2 \alpha}{6c^3}, \quad (12.6)$$

where Ω is the rotation rate of the neutron star in angular velocity, I is the moment of inertia of the neutron star, usually assumed to be 10^{45} g cm^2 , P is the spin period of the neutron star, \dot{P} is the spindown rate of the neutron star, R is the neutron star radius, B_0 is its surface magnetic field, and α is the angle between the neutron star's magnetic field axis and its rotation axis. Under these assumptions, plus the additional assumption that α is 90° , one can estimate the

magnetic field of a neutron star as:

$$B_0 = 3.2 \times 10^{19} (P\dot{P})^{1/2} \text{G}. \quad (12.7)$$


Instead, some mechanism must exist for accelerating particles in the pulsar's magnetosphere. It was originally believed that because the gravitational binding energy per particle at the surface of a neutron star is much larger than the thermal energy per particle, that neutron stars should not have substantial atmospheres (e.g., Hoyle et al. 1964; Pacini 1967, 1968). However, it was shown shortly thereafter, by Goldreich and Julian (1969) that a rotating magnetic field leads to a surface layer with charge separation such that the electric force near the surface of the neutron star is larger than the gravitational force at the surface, meaning that a neutron star surrounded by a vacuum would have a surface layer not in dynamical equilibrium.

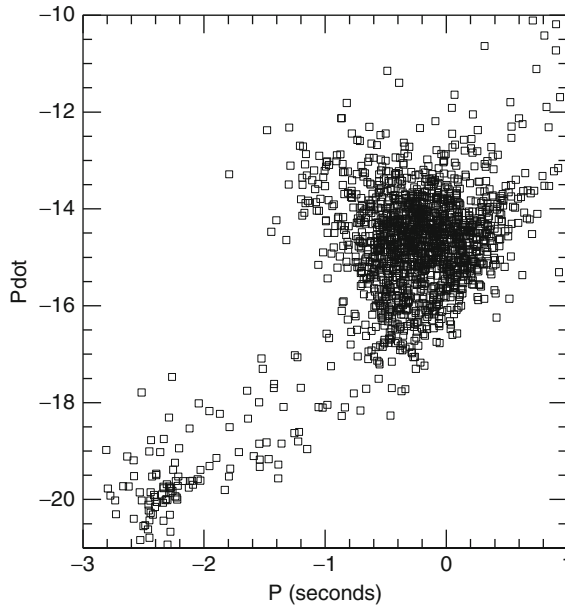
The minimum number of charged particles in a neutron star atmosphere to satisfy the Goldreich and Julian criterion is insufficient for producing strong radio emission. Instead, it is generally believed that some pair production process must be taking place within the neutron star atmosphere. In agreement with this idea is the observation of the “pulsar death line” – that there exist no pulsars which are sufficiently slowly spinning and sufficiently low in magnetic field (see, e.g., Chen and Ruderman 1993).

The radio emission from rotation-powered pulsars cannot be incoherent synchrotron radiation. There exists a limit on the brightness temperature which can be produced by incoherent synchrotron emission. Above this limit, the “Compton catastrophe” sets in – that is, Compton scattering works to cool electrons more efficiently than does synchrotron emission, and any additional power added to the system serves only to increase the luminosity of the Compton scattering component. A possible solution to this problem is that the radio emission from pulsars is, indeed, synchrotron emission but is coherent emission – electrons move in “bunches.” Since the radiated power from the synchrotron process scales as the square of the charge of the radiating particle, this bunching of electrons can lead to significantly higher ratios of synchrotron to Compton emission than can be obtained from single, incoherent electrons. More recently, plasma maser models, and plasma turbulence models have instead been more commonly proposed to explain the high brightness temperatures of pulsars (e.g. Hankins et al. 2003).

Rotation-powered pulsars usually emit most of their bolometric luminosity outside the radio band. The radio luminosities are typically less than about 10^{-6} of the spin-down luminosities, while the X-ray luminosities are typically about 0.1% of the spindown luminosities, although pulsars with more rapid spindown may have higher ratios of X-ray emission to spin-down luminosity, reaching 80% for some of the pulsars with the highest spin-down luminosities (Becker and Trümper 1997; Possenti et al. 2002). The Fermi Gamma-ray Large Area Space Telescope has recently succeeded in discovering several of pulsars as γ -ray sources before they were discovered in radio surveys. A few were quickly discovered as radio pulsars as well (Camilo et al. 2009).

6.13 Millisecond Pulsars

It is clear from examining  Fig. 12-4 that there are two distinct populations of pulsars. Most of the pulsars follow a track with inferred magnetic field values of about 10^{12} G and with spin rates of about 1 s. A significant fraction of the pulsars have inferred magnetic fields of about 10^8 G and spin periods of about 10 ms. This faster subclass of neutron stars is known as the millisecond



■ Fig. 12-4

The plot of period derivative versus period for all the pulsars in the Australia Telescope National Facility's pulsar database with period derivative measurements (Manchester et al. 2005). The region in the lower left of the diagram where no pulsars are seen is the “pulsar graveyard”, where it is believed that the pulsars are rotating too slowly to produce pair cascades. The millisecond pulsar region includes some globular clusters whose period derivatives may include contributions from acceleration in the gravitational potentials of the clusters

pulsars, the first of which was discovered in 1982 by Backer et al. and collaborators. They are sometimes referred to as “recycled” pulsars because they are thought to be recycled from dead pulsars in low-mass X-ray binaries (Alpar et al. 1982).

The evolutionary scenarios for producing millisecond pulsars usually invoke a normal process of creating a pulsar in a binary system, followed by a phase as a “normal” pulsar. At some point, the neutron star's companion begins to overflow its Roche lobe, and accretion onto the neutron star begins. The accretion process serves to provide a torque to the neutron star, which can increase its spin rate to millisecond periods with only about $0.1 M_{\odot}$ of material being added to the neutron star. Recently, a source has been discovered which appears to have spent time as both an X-ray binary and a radio pulsar within the past decade (Archibald et al. 2009).

Accretion is generally thought to reduce the magnetic fields of neutron stars. Most of the observational evidence of isolated neutron stars indicates that they have little or no magnetic field decay. Several theoretical mechanisms exist for explaining why accretion might cause field decay in neutron stars. One possibility is ohmic decay – the magnetic fields in neutron stars are driven by electron motions, and electrical resistivity can lead to a loss of these magnetic fields (see, e.g., Bhattacharya and Datta 1996). The chief alternative possibility is that the magnetic fields do not decay but are merely buried by accreted matter (e.g., Cumming et al. 2001).

Recently, an additional class of radio sources has been detected through the detection of individual pulsars of emission, rather than by Fourier transforming the power received by large radio telescopes. These sources, called rotating radio transients (or RRATs) repeat their pulsations periodically but with a very low duty cycle for detecting the pulses (McLaughlin et al. 2006). It has been suggested that at least some of the RRATs may be relatively normal radio pulsars at large distances but with a larger than typical distribution of pulse heights (Weltevrede et al. 2006).

6.13.1 Anomalous X-Ray Pulsars and Soft Gamma Repeaters

Two classes of isolated neutron stars show properties which share some broad similarities with one another but which are rather different from those seen from other classes of neutron stars. These are the anomalous X-ray pulsars, which are typically detected by showing pulsations in the X-rays and the soft gamma repeaters, which show bursts of gamma-rays which are not too different from those seen in cosmic gamma-ray bursts, except with relatively softer spectra, and with repeated activity coming from single positions on the sky. Both these classes are now thought to have much or all of their emission powered by the magnetic fields of the neutron stars, which are thought to be typically greater than 10^{14} G (Thompson and Duncan 1995).

6.13.2 Old Isolated Neutron Stars

The bulk of the neutron stars in the Galaxy rotate slowly enough to be past the pulsar death line and are not in close binaries. These objects emit thermal radiation from their still-cooling crusts with their emission typically peaking in the far-ultraviolet through soft X-ray band. Because they are both extremely faint, and their emission peaks in the bands most susceptible to interstellar absorption, these neutron stars can generally be observed only if they are very nearby. ROSAT discovered seven such object (see Kaplan 2008 for a review). One particular area of interest with these systems is fitting the spectra produced in their magnetized atmospheres, which can be used both to probe both the properties of highly magnetized matter and to make estimates of the surface gravitational redshifts of these neutron stars (see, e.g., Ho 2007).

7 Populations of Compact Objects

In addition to developing an understanding of how compact objects can produce radiation, it is important to develop an understanding of how they form, and of how often and why they end up in binary systems.

7.1 Observations

Relatively good observational constraints exist for populations of X-ray binaries and of pulsars (although neither is immune from selection effects). Even for these two classes of sources, there is strong evidence that the Galactic population is not well sampled at large distances, especially

in the Galactic Plane. For the other classes of compact objects – magnetars, old neutron stars, and isolated black holes – there exist such small numbers of objects that it is difficult to make and global statements about the source populations.


Even the observations of pulsars and of X-ray binaries are limited by selection effects, some of which are not yet well understood. A large majority of the pulsars in the Australia Telescope National Facility's pulsar database are within about 3 kpc. Some have been detected only through scintillation amplification of their brightnesses. The lack of detections of more distant pulsars is likely an effect both of lack of sensitivity of existing radio observatories and the difficulty in finding pulsations of pulsars with large dispersion measures. Most of the known X-ray binaries are on the near side of the Galactic Center, as well. The lack of more distant X-ray binaries is due to the combined effects of lack of sensitivity to large distance objects, especially in the all-sky monitors produced to date and the increasing effects of foreground absorption of soft X-rays in the Galactic Plane.

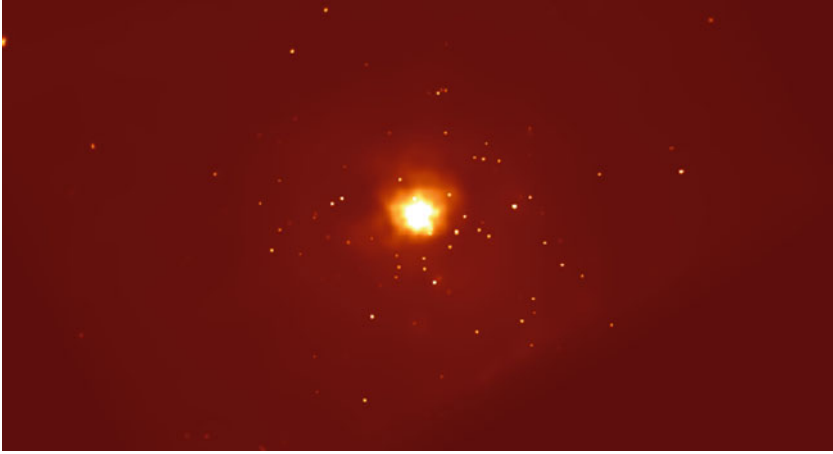
The interesting point for which there is the strongest observational evidence is that neutron stars have space velocities considerably higher than those of the massive stars thought to form them. The proper motions of many pulsars have been measured, and the best-fitting velocity distributions indicate that many pulsars are likely to have escaped the Galaxy (see, e.g., Hobbs et al. 2005). The scale height of low-mass X-ray binaries is about 1 kpc, larger than for any “normal” stellar component. Some black holes show evidence for large space velocities, as well, and the black holes also show a large-scale height (Jonker and Nelemans 2004).

7.2 Extragalactic X-Ray Binaries, Intermediate Mass Black Holes, and Other Unusual Objects

In recent years with the launch of the Chandra and XMM-Newton observatories, it has been possible to make detailed studies of the X-ray point source populations of a large number of nearby galaxies. Two major motivations exist for this work. One is to understand X-ray source populations in order to gather new data to test binary stellar evolution models, and, in the case of globular cluster X-ray sources, new data to test dynamical formation mechanisms for close binaries. The other motivation is to look for sources whose parameters fall outside the range seen in the Milky Way.

There are several problems which plague studies of X-ray source populations within the Galaxy. The most important is the difficulty in measuring source distances in the Galaxy (see above). Without reliable distance estimates, both the luminosity distributions of sources and the scale heights of source classes are difficult to determine. Another key problem is that the Galactic foreground absorption means that samples of X-ray binaries are incomplete in the Galaxy in ways that are difficult to quantify. A third problem is simply the small size of the Galactic X-ray binary population. On the other hand, for faint sources, only the Milky Way galaxy, and, to some extent the nearby dwarf spheroidal galaxies, can be studied in the X-rays with current instrumentation.

Globular cluster X-ray sources are, in particular, something which can be studied in a much different manner in nearby galaxies than in the Milky Way (see  Fig. 12-5 for an X-ray image of a nearby elliptical galaxy). Only 12 of the Milky Way's 150 or so globular clusters has been found to contain a bright ($L_X > 10^{35}$ ergs/s) X-ray source, although two of them have now been shown to contain at least two such sources at various times. One key result of extragalactic globular cluster X-ray source studies include the discovery that metal rich clusters are several times more likely to contain X-ray sources than metal poor clusters (Kundu et al. 2002), something which



■ Fig. 12-5

An X-ray image from the Chandra X-ray observatory of the galaxy NGC 4472, the optically brightest galaxy in the Virgo Cluster. This image contains about 140 X-ray point sources. About 40% of the X-ray sources in regions with good optical coverage from the Hubble Space Telescope are associated with globular clusters. The diffuse emission in the center of the galaxy is truly diffuse hot gas emission, rather than a superposition of X-ray point sources

had been suggested in the Milky Way (e.g., Silk and Arons 1975) but could be proven with the small number of Milky Way cluster sources, and the correlations between metallicity and other parameters for Milky Way clusters. Extragalactic clusters have also shown the first clear evidence for a stellar-mass black hole in a globular cluster (Maccarone et al. 2007) and the best evidence to date for an intermediate mass black hole (i.e., a black hole of more than about $1,000 M_{\odot}$) in a globular cluster (Gebhardt et al. 2005; Ulvestad et al. 2007).

Additionally, searches of nearby galaxies have turned up the class of ultraluminous X-ray sources. The standard definition of ultraluminous is having measured luminosity greater than 10^{39} ergs/s, with the measurements of nearby galaxies nearly all being made in the 0.5–10 keV energy range where imaging instruments are effective. Many of the Galactic black hole X-ray binaries have, at least briefly, reached this luminosity threshold, but only GRS 1915+105 has maintained a luminosity at or near this level for an extended period of time. Many of the ultraluminous X-ray sources have remained steadily above 10^{39} ergs/s over time scales of years or longer, and some have steadily emitted at luminosities over 10^{40} ergs/s.

These ultraluminous X-ray sources are of great interest because either they are manifesting a mode of accretion not observed in the Milky Way or they harbor intermediate mass black holes as their accretors. A few of the brightest ultraluminous X-ray sources show X-ray spectra which are best fitted with disk components of lower temperature than those seen from high luminosity black hole X-ray binaries in the Galaxy, and where the normalizations of the disk spectral components imply a large inner radius for the accretion disk. If one assumes that the disk black-body model is the physically correct description of the data, rather than merely a model which is statistically consistent with the data, and that the disk extends in to the innermost stable circular orbit around the black hole, then the implied black hole masses of this subclass of ultraluminous X-ray sources are typically $\sim 1,000 M_{\odot}$ – much larger than what can be produced through normal stellar evolutionary channels, but much smaller than the black holes seen in galactic nuclei.

Several caveats apply here, and several alternative models exist. First, the X-ray spectra of the ultraluminous X-ray sources with cool disks look qualitatively different from the X-ray spectra of stellar-mass black holes in all the sub-Eddington spectral states – they typically have X-ray spectra with bright cool disks but dominant power-law components (Soria et al. 2007). There exist mechanisms to explain their high luminosities by relativistic beaming (Körding et al. 2006), or by some combination of geometric beaming and the capability of nonspherical flows actually to exceed the Eddington limit (e.g., King 2002).

Other evidence, apart from ultraluminous X-ray sources has been suggested for the existence of intermediate-mass black holes. In globular clusters, one can attempt to measure the distribution of stellar velocities around the cluster centers. If a sufficiently massive black hole is present, it should dominate the gravitational potential for a large enough number of stars that their motions should indicate its presence. To date, there are some suggestive hints for intermediate-mass black holes from these observations, but the dynamical data taken alone have always been consistent with alternative explanations, such as anisotropic velocity distributions for the stars in the cluster (see, e.g., Noyola et al. 2008). In one case, the extragalactic globular cluster G1 in M 31, there exists both dynamical evidence (Gebhardt et al. 2005) and accretion evidence, in the form of both radio and X-ray emission (Trudolyubov and Priedhorsky 2004; Ulvestad et al. 2007), for the existence of an intermediate mass black hole.

8 Formation of Compact Objects and Compact Object Binaries

The formation of black holes and neutron stars, and the understanding of how these systems end up in binaries is an important process to understand in astrophysics. The supernova explosions that produce neutron stars and black holes also contribute substantial fractions of many of the heavy elements in the Universe. They represent one of the few known classes of events capable of producing detectable neutrino emission. They may be responsible for a substantial fraction of the kinetic energy input into the Galaxy and other galaxies.

Additionally, the formation of X-ray binaries represents an opportunity to test theories of the evolution of close binary stars. Understanding binary evolution is a key certainly to understanding how gravitational wave emitting binaries form and how Type Ia supernovae occur. It may also be important for understanding cosmic γ -ray bursts, and the formation and evolution of exotic classes of optically emitting stars, such as blue stragglers, extreme horizontal branch stars, and subdwarf B stars. At the same time, X-ray binaries are one of the very few classes of individual stars that can be seen out to distances of 20 Mpc, in which is enclosed a wide range of types of galaxies, so that the effects of stellar population on source class production can be examined.

8.1 Black Hole and Neutron Star Formation in Supernovae and Related Events

Black holes and neutron stars form in core collapse supernovae. When the core of the star has exhausted its nuclear fuel by fusing all its material into iron-peak elements, its only available source of energy is gravitational collapse. At the low end of the range of stellar masses sufficient to undergo core collapse, nucleonic degeneracy pressure will be sufficient to support the core

after this happens, leaving a neutron star. At the upper end of this mass range, a full collapse will take place, turning nearly the entire star into a black hole (see, e.g., Fryer and Kalogera 2001 for a discussion).

In the case where the core of the massive star turns into a neutron star, there remain two possibilities. In some cases, the gravitational binding energy released by the collapse will be sufficient to blow away the entire remaining mass in the star in the ensuing supernova. For somewhat more massive stars, the supernova explosion unbinds only a fraction of the material from the outer star, and a fallback of the remaining material onto the young neutron star takes place, leading to the formation of a black hole through the accretion of this material, rather than through the prompt collapse scenario. As a result, these “fallback” black holes might be expected to have natal kicks. The small amount of observational data that exists to test this idea supports it – GRO J1655-40 is a relatively low mass ($\sim 6 M_{\odot}$) black hole, and has a space velocity of about 100 km/s relative to its local standard of rest, a value which is difficult to accommodate with just a Blaauw kick (e.g., Nelemans et al. 1999). Additionally, a recent reexamination of the distances to low mass X-ray binaries has indicated that the scale height of the black hole LMXB population is about 1 kpc, similar to that of the neutron star LMXB population, and much larger than the scale height of the donor star population (Jonker and Nelemans 2004).

It is important to note that the mass estimates for black holes and neutron stars come predominantly from compact objects in close binary systems. Because of this, the measured mass functions of black holes, especially may be severely biased. Theoretical work indicates that the binary evolutionary processes that lead to black hole X-ray binary formation can have significant effects on the masses of the black holes that are produced.

The formation channels for compact objects in binary systems are necessarily rather complicated in most cases. Simply evoking independent evolution of two stars in a binary system would yield the result, for example, that it is impossible to form most neutron star X-ray binaries. The mass loss of the supernova explosion will take away more than half the mass of the binary system in most such cases. The most prominent solution to this problem is to invoke common envelope evolution early in the lifetime of the binary system. What this means is that when the heavier of the two stars in a binary system evolves off the main sequence, it expands to the point where it envelopes the other star. During this process, a considerable amount of mass can be ejected, making possible the formation of low-mass X-ray binaries with neutron stars. See, e.g., Taam and Sandquist (2000) for a review of common envelope evolution.

It is also possible, but still somewhat controversial, that some neutron stars through accretion induced collapse events in systems with accreting white dwarfs (or for two white dwarfs to merge). In particular, it has been suggested that a large fraction of the neutron stars in globular clusters should form in this manner (e.g., Ivanova et al. 2008) in order to solve the problem that very few neutron stars would be retained in globular clusters given a velocity distribution similar to that seen for isolated pulsars.

The formation scenarios for intermediate-mass black holes are more numerous and more controversial. Stars formed from the metal free gas in the early universe should have masses of a ~ 100 – $1,000 M_{\odot}$. Most evolutionary calculations indicate that these stars should have very little mass loss during their lifetimes, and that in some mass ranges, they should undergo prompt collapses, in which most or all of the stellar mass ends up in a black hole (e.g., Heger and Woosley 2002). In other mass ranges, they are thought to undergo pair instability supernovae, and to be destroyed entirely. Lower redshift star formation may lead to the formation of intermediate-mass black holes as well, especially in dense star clusters, where runaway mergers of massive stars may lead to the creation of a super-massive star (Portegies Zwart and

McMillan 2002). This supermassive star would have stronger stellar winds than a Population III star of the same mass but might be able to reach an equilibrium at a few hundred solar masses.

8.2 Natal Kicks

The large space velocities of pulsars and the large-scale heights of low-mass X-ray binaries containing neutron stars indicate that it is likely neutron stars are born with a velocity kick. In order to establish clearly that an actual velocity kick is imparted to the neutron star at birth, it is necessary to demonstrate that other mechanisms of increasing the neutron star's velocity cannot be responsible for the observed space velocities and spatial distributions.

Rather large space velocities for supernova products can be achieved without a velocity impulse being applied during (or shortly after) the supernova, if the supernova takes place in a binary system (Blaauw 1961). Rapid mass loss from a star, taking place over a timescale much shorter than the orbital period of the binary, should lead to a recoil velocity being applied to the binary system such to make up for the momentum lost to the ejected material. Even a spherically symmetric explosion will carry away linear momentum, given that the spherically symmetric explosion will come from a star which has a motion about the center of mass of the binary.

Several mechanisms exist for explaining the natal kicks. One possibility is that the supernova explosion itself may be asymmetric. Another is that the proto-neutron star may be emitting radiation or neutrinos asymmetrically.

9 Conclusions

A wealth of knowledge has been collected about neutron stars and black holes in the roughly five decades in which astronomers have been studying them. These objects have already placed important limits on fundamental physics, and three Nobel Prizes have been awarded for work on them. At the same time, many of the key questions remain open, or, at least, have answers which have not yet reached consensus level. The coming years should yield more clear answers on issues such as the mass distribution and formation mechanisms for black holes and neutron stars, the neutron star equation of state, the spin distributions of black holes, the mechanisms by which jets are launched, the reasons for variability in accretion disks in general, and quasi-periodic oscillations in particular and the mechanisms by which pulsars emit, among other questions.

References

- | | |
|---|--|
| <p>Abramowicz, M. A., & Kluzniak, W. 2001, <i>A&A</i>, 374, L16</p> <p>Abramowicz, M. A., Kluzniak, W., & Lasota, J.-P. 2002, <i>A&A</i>, 396, L31</p> <p>Alpar, M. A., Cheng, A. F., Ruderman, M. A., & Shaham, J. 1982, <i>Nature</i>, 300, 728</p> <p>Archibald, A. M., et al. 2009, <i>Science</i>, 324, 1411</p> <p>Armitage, P. J., & Natarajan, P. 1999, <i>ApJ</i>, 523, L7</p> <p>Baade, W., & Zwicky, F. 1934, <i>PNAS</i>, 20, 254</p> | <p>Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M., & Goss, W. M. 1982, <i>Nature</i>, 200, 615</p> <p>Balbus, S. A., & Hawley J. F. 1991, <i>ApJ</i>, 376, 214</p> <p>Bardeen, J. M., & Petterson, J. A. 1975, <i>ApJ</i>, 195, L65</p> <p>Basko, M. M., Sunayev, R. A. & Titarchuk, L. G. 1974, <i>A&A</i>, 31, 249</p> <p>Becker, W., & Trümper, J. 1997, <i>A&A</i>, 326, 682</p> <p>Bennett, D. P., et al. 2002, <i>ApJ</i>, 579, 639</p> <p>Bhattacharya, D., & Datta, B. 1996, <i>MNRAS</i>, 282, 1059</p> |
|---|--|

- Blaauw, A. 1961, *BAN*, 15, 265
- Blandford, R. D., & Znajek, R. L. 1977, *MNRAS*, 179, 433
- Blandford, R. D., & Konigl, A. 1979, *ApJ*, 232, 34
- Bradshaw, C. F., Fomalont, E. B., & Geldzahler, B. J. 1999, *ApJ*, 512, L121
- Brocksopp, C., Bandyopadhyay, R. M., & Fender, R. P. 2004, *NewA*, 9, 249
- Camilo, F., et al. 2009, *ApJ*, 705, 1
- Casella, P., Altamirano, D., Patruno, A., Wijnands, R., & van der Klis, M. 2008, *ApJ*, 674, L41
- Chatterjee, S., et al. 2009, *ApJ*, 698, 250
- Chen, K., & Ruderman, M. 1993, 402, 264
- Chen, W., Shrader, C. R., & Livio, M. 1997, *ApJ*, 491, 312
- Coppi, P. S. 1999, *ASPC*, 161, 375
- Corbel, S., Fender, R. P., Tomsick, J. A., Tzioumis, A. K., & Tingay, S. 2004, *ApJ*, 617, 1272
- Cordes, J. M., & Lazio, T. J. W. 2003, *astro-ph/031598*
- Cottam, J., Paerels, F., & Mendez, M. 2002, *Nature*, 420, 51
- Cumming, A., Zweibel, E., & Bildsten, L. 2001, *ApJ*, 557, 958
- Curtis, H. D. 1918, *PLiCO*, 13, 55
- Davis, S. W., Done, C., & Blaes, O. M. 2006, *ApJ*, 647, 525
- Dhawan, V., Mioduszewski, A., & Rupen, M. 2006, VI Microquasar Workshop: Microquasars and Beyond. *Proceedings of Science*, MQW6, 52
- Done, C., Gierlinski, M., & Kubota, A. 2007, *A&ARv*, 15, 1
- Dubus, G. 2006, *A&A*, 456, 801
- Dubus, G., Lasota, J.-P., Hameury, J.-M., & Charles, P. 1999, *MNRAS*, 303, 139
- Eggleton, P. P. 1983, *ApJ*, 268, 368
- Fender, R. P. 2003, *MNRAS*, 340, 1353
- Fender, R. P. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin, & M. van der Klis (Cambridge: Cambridge University Press)
- Fender, R. P., Gallo, E., & Russell, D. M. 2010, *MNRAS*, 406, 1425
- Fender, R., Wu, K., Johnston, H., Tzioumis, T., Jonker, P., Spencer, R., & van der Klis, M. 2004, *Nature*, 427, 222
- Fragile, P. C., Mathews, G., & Wilson, J. 2001 *ApJ*, 553, 955
- Frank, J., King, A., & Raine, D. 1995, *Accretion Power in Astrophysics* (Cambridge: Cambridge University Press)
- Fryer, C. L., & Kalogera, V. 2001, *ApJ*, 554, 548
- Gallo, E., Fender, R. P., & Pooley, G. G. 2003, *MNRAS*, 344, 60
- Galloway, D., Lin, J., Chakrabarty, D., & Hartman, J. M. 2009, *astro-ph/0910.5546*
- Garcia, M. R., McClintock, J. E., Narayan, R., Callanan, P., Barret, D., & Murray, S. S. 2001, *ApJ*, 553, L47
- Gebhardt, K., Rich, R. M., & Ho, L. C. 2005, *ApJ*, 634, 1093
- Goldreich, P., & Julian, W. H. 1969, *ApJ*, 157, 869
- Grindlay, J., Gursky, H., Schnopper, H., Parsignault, D. R., Heise, J., Brinkman, A. C., & Schrijver, J. 1976, *ApJ*, 205, L127
- Hankins, T. H., Kern, J. S., Weatherall, J. C., Eilek, J. A., 2003, *Nature*, 422, 141
- Heger, A., & Woosley, S. E. 2002, *ApJ*, 567, 532
- Hjellming, R. M., & Johnston, K. J. 1988, *ApJ*, 328, 600
- Ho, W. C. G. 2007, *MNRAS*, 380, L71
- Hobbs, G., Lorimer, D. R., Lyne, A. G., & Kramer, M. 2005, *MNRAS*, 360, 974
- Homan, J., van der Klis, M., Wijnands, R., Bel-loni, T., Fender, R., Klein-Wolt, M., Casella, P., Mendez, M., Gallo, E., Lewin, W. H. G., & Gehrels, N. 2007, *ApJ*, 656, 420
- Hoyle, F., Fowler, W. A., Burbidge, G. R., & Burbidge, E. M. 1964, *ApJ*, 139, 909
- Hynes, R. I. 2002, in "The Physics of Cataclysmic Variables and Related Objects," *Astronomical Society of the Pacific Conference Proceedings*, Vol. 261, ed. B. T. Gänsicke, K. Beuermann, K. Reisch (San Francisco: Astronomical Society of the Pacific), 676
- Hynes, R. I., Steeghs, D., Casares, J., Charles, P. A., & O'Brien, K. 2004, *ApJ*, 609, 317
- Ichimaru, S. 1977, *ApJ*, 214, 840
- Ivanova, N., Heinke, C. O., Rasio, F. A., Belczynski, K., & Fregeau, J. M. 2008, *MNRAS*, 386, 553
- Jonker, P. G., & Nelemans, G. 2004, *MNRAS*, 354, 355
- Jonker, P. G., Steeghs, D., Chakrabarty, D., & Juett, A. M. 2007, *ApJ*, 665, L147
- Körding, E., Fender, R. P., & Migliari, S. 2006, *MNRAS*, 369, 1451
- Kaplan, D. L. 2008, *AIPC*, 983, 331
- Kaplan, D., van Kerkwijk, M. H., & Anderson, J. 2002, *ApJ*, 571, 447
- Kaspi, V., Roberts, M. S. E., & Harding, A. K. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin, & M. van der Klis (Cambridge: Cambridge University Press)
- Kerr, R. P. 1963, *PhRvL*, 11, 237
- King, A. R. 2002, *MNRAS*, 335, L13
- Koerding, E., Falcke, H., & Markoff, S. 2002, *A&A*, 382, L13
- Kotani, T., Kawai, N., Aoki, T., Doty, J., Matsuoka, M., Mitsuda, K., Nagase, F., Ricker, G., & White, N. E. 1994, *PASJ*, 46, L147
- Kundu, A., Maccarone, T. J., & Zepf, S. E. 2002, *ApJ*, 574, L5
- Kuulkers, E. 2004, *NuPhS*, 132, 466

- Kuulkers, E., den Hartog, P. R., in't Zand, J. J. M., Verbunt, F. W. M., Harris, W. E., & Cocchi, M. 2003, *A&A*, 399, 663
- Lorimer, D. R. 2008, *LRR*, 11, 8
- Lyne, A. G., & Graham-Smith, F. 2006, *Pulsar Astronomy* (Cambridge: Cambridge University Press)
- Maccarone, T. J. 2003, *A&A*, 409, 697
- Maccarone, T. J. 2005, *MNRAS*, 360, L30
- Maccarone, T. J. 2008, *ASPC*, 401, 191
- Maccarone, T. J., & Coppi, P. S. 2003, *MNRAS*, 338, 189
- Maccarone, T. J., Coppi, P. S., & Poutanen, J. 2000, *ApJ*, 537, L107
- Maccarone, T. J., Kundu, A., Zepf, S. E., & Rhode K. L. 2007, *Nature*, 445, 183
- Magdziarz, P., Zdziarski, A. A. 1995, *MNRAS*, 273, 837
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, *AJ*, 129, 1993
- Margon, B., Grandi, S., Stone, R. P. S., & Ford, H. C. 1979, *ApJ*, 233, L63
- Massi, M., & Kaufman Bernado, M. 2009, *ApJ*, 702, 1179
- McClintock, J. E., & Remillard, R. A. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin, & M. van der Klis (Cambridge: Cambridge University Press)
- McLaughlin, M., et al. 2006, *Nature*, 439, 817
- Merloni, A., Fabian, A. C., & Ross, R. R. 2000, *MNRAS*, 313, 193
- Migliari, S., & Fender, R. P. 2006, *MNRAS*, 366, 79
- Migliari, S., Fender, R. P., Rupen, M., Wachter, S., Jonker, P. G., Homan, J., & van der Klis, M. 2004, *MNRAS*, 351, 186
- Migliari, S., Tomsick, J. A., Maccarone, T. J., Gallo, E., Fender, R. P., Nelemans, G. P., & Russell, D. M. 2006, *ApJ*, 643, L41
- Miller, J. M., Reynolds, C. S., Fabian, A. C., Miniutti, G., & Gallo, L. C. 2009, *ApJ*, 697, 900
- Miller-Jones, J. C. A., Jonker, P. G., Dhawan, V., Briskin, W., Rupen, M. P., Nelemans, G., & Gallo, E. 2009, *astro-ph/0910.5253*
- Mirabel, I. F., & Rodriguez, L. F. 1994, *Nature*, 371, 46
- Narayan, R., & Yi, I. 1994, *ApJ*, 428, L13
- Nelemans, G., Tauris, T. M., & van den Heuvel, E. P. J. 1999, *A&A*, 352, L87
- Nowak, M. A. 1995, *PASP*, 107, 1207
- Nowak, M. A., Wilms, J., Heindl, W. A., Pottschmidt, K., Dove, J. B., & Begelman, M. C. 2001, *MNRAS*, 320, 316
- Noyola, E., Gebhardt, K., & Bergmann, M. 2008, *ApJ*, 676, 1008
- Orosz, J. A., et al. 2007, *Nature*, 449, 872
- Özel, F., Gülver, T., & Psaltis, D. 2009, *ApJ*, 693, 1775
- Pacini, F. 1967, *Nature*, 216, 567
- Pacini, F. 1968, *Nature*, 219, 145
- Paczynski, B. 1971, *ARA&A*, 9, 183
- Portegies Zwart, S. F., & McMillan, S. L. W. 2002, *ApJ*, 576, 899
- Possenti, A., Cerutti, R., Colpi, M., & Mereghetti, S. 2002, *A&A*, 387, 993
- Remillard, R. A., Lin, D., Cooper, R. L., & Narayan, R. 2006, *ApJ*, 646, 407
- Rezzolla, L., Yoshida, S., Maccarone, T. J., & Zanotti, O. 2003, *MNRAS*, 344, L37
- Rutledge, R. E., Bildsten, L., Brown, E. F., Pavlov, G., Zavlin, V. E., & Ushomirsky, G. 2002, *ApJ*, 580, 413
- Salpeter, E. E. 1964, *ApJ*, 140, 796
- Schwarzschild, K. 1916, *AbhKP*, 189
- Seyfert, C. 1943, *ApJ*, 97, 28
- Shafee, R., McClintock, J. E., Narayan, R., Davis, S. W., Li, L.-X., & Remillard, R. A. 2006, *ApJ*, 636, L113
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shapiro, S. L., & Teukolsky, S. A. (1983), in *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects* (New York: Wiley-Interscience)
- Silk, J., Arons, J. 1975, *ApJ*, 200, L131
- Smak, J. 1984, *AcA*, 34, 161
- Soria, R., Goncalves, A. C., & Kuncic, Z. 2007, *AIPC*, 924, 919
- Strohmayer, T. E. 2001, *ApJ*, 552, L49
- Strohmayer, T. E., Zhang, W., Swank, J. H., Smale, A., Titarchuk, L., Day, C., & Lee, U. 1996, *ApJ*, 469, L9
- Swank, J. H., Becker, R. H., Boldt, E. A., Holt, S. S., Pravdo, S. H., & Serlemitsos, P. J. 1977, *ApJ*, 212, L73
- Taam, R. E., & Sandquist, E. L. 2000, *ARA&A*, 38, 113
- Tananbaum, H., Gursky, H., Kellogg, E., Giacconi, R., & Jones, C. 1972, *ApJ*, 177, L5
- Thompson, C., & Duncan, R. C. 1995, *MNRAS*, 275, 255
- Thorne, K. S., & Price, R. H. 1975, *ApJ*, 195, L101
- Trudolyubov, S. P., & Priedhorsky, W. C. 2004, *ApJ*, 616, 821
- Tudose, V., Fender, R. P., Linares, M., Maitra, D., & van der Klis, M., 2009, *MNRAS*, 400, 2111
- Ulvstad, J. S., Greene, J. E., & Ho, L. C. 2007, *ApJ*, 661, L151
- van der Klis, M. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin, & M. van der Klis (Cambridge: Cambridge University Press)
- van Kerkwijk, M. 1993, *A&A*, 276, L9
- van Paradijs, J., & McClintock, J. E. 1994, *A&A*, 290, 133
- Weltevredre, P., Stappers, B. W., Rankin, J. M., & Wright, G. A. E. 2006, *ApJ*, 645, L149
- Wijnands, R. 2004, *NuPhS*, 132, 496

Wijnands, R., et al. 2006, A&A, 449, 1117

Wijnands, R., Homan, J., Miller, J. M., & Lewin,
W. H. G. 2004, ApJ, 606, L61

Wilms, J., Schönherr, G., Schmid, J., Dauser, T., &
Kreykenbohm, I. 2009, AIPC, 1126, 295

Woosley, S. E., & Taam R. E. 1976, Nature, 263, 101

13 Binaries and Multiple Stellar Systems

Elliott Horch

Department of Physics, Southern Connecticut State University,
New Haven, CT, USA

1	<i>Introduction</i>	655
2	<i>Observational Techniques</i>	658
2.1	The Dynamical Problem	658
2.2	Visual Binaries	659
2.2.1	The Visual Orbit	659
2.2.2	Overview of the Observational Data	662
2.2.3	Speckle Interferometry	663
2.2.4	Long-Baseline Optical Interferometry	665
2.2.5	Adaptive Optics	666
2.2.6	Space-Based Observations	666
2.2.7	Lunar Occultation Measurements of Binaries	668
2.3	Spectroscopic Binaries	669
2.3.1	The Spectroscopic Orbit	669
2.3.2	Overview of Spectroscopic Data	672
2.3.3	Observational Techniques	672
2.4	Eclipsing Binaries	673
2.4.1	Orbital Information Determined from Observation	673
2.4.2	Observational Data	675
2.5	Complementary Observations	675
3	<i>Basic Stellar Astrophysics with Binaries</i>	676
3.1	The Mass–Luminosity Relation and Mass–Radius Relation	676
3.2	Formation of Binary and Multiple Systems	677
3.3	The Algol Paradox	679
3.4	Other Tests of Stellar Evolution	679
4	<i>Statistics of Binaries and Implications</i>	680
4.1	The Field Population	680
4.2	Pre-main Sequence Binaries	682
4.3	Thin- and Thick-Disk Binaries	684
4.4	Binary Stars in Clusters	685

5	<i>Interacting Binaries</i>	686
5.1	General Comments on Interacting Binaries	686
5.2	Important Types of Interacting Binaries	688
5.2.1	Type Ia Supernovae	688
5.2.2	X-Ray Binaries	688
5.2.3	Blue Stragglers	689
	<i>References</i>	689

Abstract: Binary and multiple stellar systems have importance in three main areas of astronomy and astrophysics. First, because of the relatively simple gravitational interaction at work in the case of binary stars, these systems provide a basic check on stellar structure and evolution theory since the masses may be determined through observation. When these masses can be linked to other properties of the two stars, such as luminosity, color, and radius, they can provide very stringent constraints on stellar models. Second, the statistics of binary and multiple star systems provide clues to star formation mechanisms and environmental effects in the galactic gravitational potential and in clusters. Although a number of good results have been obtained in nearby star clusters and associations, knowledge of the field population has been somewhat limited until recently by a lack of large, complete samples of binaries. However, there appears to be a great deal of promise in this area for the coming decade in part due to astrometric satellites such as *Hipparcos* and *Gaia*. Third, the binary scenario is invoked to explain several important types of astrophysical phenomena such as Type Ia supernovae, cataclysmic variables, and stellar x-ray sources. Since the first of these mentioned is a standard candle for the extragalactic distance scale, it may even be said binary stars play a minor role in field of cosmology. However, in this chapter, the focus will mainly be on normal stars in binary and multiple-stellar systems. The basic physics of binaries will be reviewed, and the observational methods in use today will be discussed together with their limitations and prospects for the future. Finally, an overview of the current science in the three main areas mentioned where binaries have a significant impact will be given.

Keywords: Astrometric, Binary star orbits, Detached binaries, Interacting binaries, Interferometric, Spectroscopic, Star clusters, Star formation, Stellar astrophysics, Stellar masses

1 Introduction

When two stars appear close together on the sky, the pair is known as a double star. However, such a case could represent just a chance alignment along the line of sight, with one star actually much farther away from the observer than the other. This is the rather uninteresting case of an optical double star. However, it is possible that the two stars are not only close together on the sky but also near one another in space, in which case they will orbit one another due to their mutual gravitational interaction. This was first suggested by John Michell (1767), but it was Sir William Herschel who first referred to this kind of object as a “binary sidereal system,” what is now called a binary star (Argyle 2004). Determining whether a given double star is actually a binary requires patient observation. If only an optical double, the relative position of the pair should change in a linear fashion over time, indicating no acceleration of one body relative to the other. If it is a binary, a curved path should be noted over time, and indeed, if the pair is followed long enough, one may see the two stars complete an orbital ellipse around one another. Herschel was the first astronomer to put forth serious efforts to observe the relative motions of many double stars in order to conclusively determine if some of them were binaries. After some 25 years of work, he published the result that several of the systems he had been observing over that time were in fact binaries, and not merely optical doubles (Herschel 1803). Thus the field of visual binary star astronomy began.

Through the nineteenth and well into the twentieth centuries, the search for binary stars and the determination of their orbits was a principal focus of the observational efforts of

astronomers such as William Herschel and his son John Herschel, F. G. W. (Wilhelm) Struve and his son Otto, W. R. Dawes, S. W. Burnham, and R. G. Aitken. Largely through the efforts of these and other observers, binary star astronomy began to make its first contribution to astrophysics: the empirical determination of mass information about stars other than the Sun by application of Newton's law of gravity to the observational data. The first orbit calculation, circa 1830, was of ξ Ursa Majoris, which has a period of approximately 60 years. At least three astronomers developed methods of calculating the orbit from the observational data in that timeframe: Savary, Encke, and John Herschel. By 1935, over 100 visual orbits had been calculated and tabulated in Aitken's classic work, *The Binary Stars* (Aitken 1935).

Once photography allowed for the collection of the spectrum of a star, examples of cases where the absorption lines appeared to be double and regularly changing in time led to the realization that such stars were also binary, although in almost all cases, the two stars were very close together on the sky so that the light received could not be resolved into two stars in an image. The first example of the detection of this phenomenon was with the star Mizar (ζ Ursa Majoris) by E. C. Pickering in 1889 (Aitken 1935). This gave rise to a new field of study: that of the so-called spectroscopic binary stars. The changing nature of the spectral lines could be interpreted as the Doppler shift of the wavelength of the lines as the two stars moved toward and away from the observer, with the complete cycle determining orbital period of the binary itself as well as other properties of the orbit. In addition, the orbital periods of these systems were in general much shorter than those of the visual binaries, meaning that the orbital information for many systems could be determined in a much shorter period of time, with more complete orbital coverage.

Pickering was also one of the first astronomers to work on a third type of binary stars in the late nineteenth and early twentieth century (Aitken 1935). At that time, it was known that the star Algol (β Persei) was a periodic variable star, where the variation in light intensity was constant for some time, then dimmed for a short time, then returned to the original brightness. Such Algol-type variables can be explained by suggesting that such systems are actually binary stars where the plane of the orbit is perpendicular to the plane of the sky. In this case, each star will pass in front of the other star once per orbital period, resulting in two dips in light intensity at regular intervals. The dips would not be expected to be the same depth, since in general the two stars will not be the same size or brightness. Thus, Algol and other variable stars of this type can be explained as binary systems.

Two other types of binaries that are sometimes discussed are the astrometric binaries and spectrum binaries. An astrometric binary is when one observes extremely small, periodic shifts in the position of a star relative to fixed background stars. From this, one may deduce that a gravitational acceleration is being exerted on the star, albeit by an unseen companion. Usually, the second star will be of far lower luminosity than the first so that the light detected is dominated by the brighter star, yet the gravitational influence is sufficient to see the reflex motion as the brighter object orbits the common center of mass. A classic discovery of this type is the motion of the main-sequence star Sirius A (α Canis Majoris), though, subsequently, observations were able to detect the faint companion star, which is a white dwarf. In a similar way, a spectrum binary is a star in which variations are observed in the spectrum and the object is therefore presumed to be binary, but no firm conclusion can be made about the nature of the orbit of the system. This might occur for example if the period is very long so that the shift of spectral lines is only marginally detectable with a given spectrograph.

It is interesting to note that the three main types of binaries are divided by observational technique, and not by object morphology, as the names indicate: astrometric observations are

used to obtain orbital information of visual (and astrometric) systems, spectroscopy is needed for spectroscopic (and spectrum) binaries, and photometry is the main tool when studying eclipsing binary stars. It is obviously possible to have a given binary system be both an eclipsing and spectroscopic system for example (meaning that it can be successfully observed with both techniques), or to be both a visual binary and a spectroscopic system, but through most of the history of binary star astronomy, there have been few examples of such good fortune. The implications of this fact have made it difficult from an observational point of view to develop complete, statistically significant samples of binaries from which to study their properties. It is easier to measure a large orbital velocity using spectroscopy since that will produce a larger Doppler shift and a greater change in the position of the lines being observed. However, this technique therefore favors small separations and/or large masses, since Kepler's laws tell us that the larger the separation and/or the smaller the masses, the smaller the orbital speed. On the other hand, observations of visual binaries favor larger separations so that the two stars can be clearly resolved. Thus, it has often been said that spectroscopic and astrometric techniques are complementary methods of studying binaries. However, modern high-resolution imaging and high-precision spectroscopy have finally bridged this gap: although an old field in comparison to many other areas of active research in astrophysics, binary star astronomy has greatly benefited from the application of modern observational techniques to obtain new and significant results.

Multiple stellar systems, that is, cases of more than two stars that are gravitationally bound, also exist. In fact, the work of Abt and Levy (1976) and Duquennoy and Mayor (1991) both show that multiple stellar systems of all types (including binaries) are roughly as common as single stars, at least for solar-type stars. Duquennoy and Mayor, for example, found that the relative frequency of single stars, binaries, trinary stars, and quadruple stars is approximately 57%:38%:4%:1%, respectively. In multiple stellar systems, one refers to the components of such systems, starting with the brightest member (the primary star) labeled as the A component, and then companions at successively larger separations from the primary labeled as the B, C, D components, and so on. Sometimes, a close component is discovered after a wider component; in these cases, a lower-case letter is used to indicate this. Each pair of components can be studied with one or more of the techniques already discussed. Studies of orbital stability suggest that long-lived multiple star systems will form in a hierarchical way, with each successively wider component at a separation much larger than that of the previous component relative to the primary star. In more general terms, both the process of formation and the environment can influence the statistics of orbital parameters, such as period, eccentricity, and so on. In dense stellar environments, such as star clusters, systems of wide separation are thought to be dispersed over time, whereas smaller separation systems evolve toward smaller separations through stellar encounters. Hierarchical triple stars may be preferentially formed in encounters of two binary systems, where one component is cast out of the system. When a component of a multiple system is ejected, it is often with high velocity and is known as a runaway. Such runaway stars can sometimes be identified as high proper-motion objects that are heading away from a known multiple system at approximately the same distance.

It can also be convenient to categorize binary stars in terms of physical processes. Most "normal" binaries observed with the techniques described are physically separated stars whose main interaction is simply to orbit one another; that is, they are detached systems. However, in the case of very close binaries, there can be effects such as rotational synchronization and mass transfer between the two stars that significantly change the evolution of the system and its astrophysical properties. In the most extreme case, the surfaces of the stars can have severe distortion

to the star and a common surface, leading to the notion of a contact binary. In less extreme cases, mass infall to the other star can result in many interesting processes that are observed in X-ray binaries, cataclysmic variable stars, and Type I supernovae. These are phenomena that can be explained by the model of a “semidetached” binary system.

2 Observational Techniques

2.1 The Dynamical Problem

It is useful to begin with a brief synopsis of the two-body problem from classical mechanics. When one star orbits another, the force of gravity provides an acceleration that leads to the deviation from rectilinear motion. The force on one star due to the other follows Newton’s famous equation:

$$\mathbf{F}_{12} = -G \frac{m_1 m_2}{|\mathbf{r}|^2} \hat{\mathbf{r}}, \quad (13.1)$$

where \mathbf{r} is the relative coordinate vector, $\mathbf{r}_1 - \mathbf{r}_2$, and $\hat{\mathbf{r}}$ is the unit vector in the \mathbf{r} direction. The force is directed along a line joining the two masses, and is attractive, as indicated by the sign of the right hand side of (13.1). Since the forces acting on these masses are radial in the relative coordinate and the center of mass will be along the line joining the two masses, there is no net torque in the absence of external forces on the system. Therefore, the angular momentum vector,

$$\mathbf{L} = \mathbf{r} \times (m_1 + m_2) \mathbf{v}, \quad (13.2)$$

is conserved. The total mechanical energy, in other words the sum of the kinetic and potential energies (K and U respectively), is by definition

$$E_{\text{TOT}} = K + U = \left(\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 \right) - G \frac{m_1 m_2}{|\mathbf{r}|}, \quad (13.3)$$

which is also conserved in the absence of external forces. Defining the center of mass of the system as

$$\mathbf{r}_{\text{CM}} = \frac{m_1}{m_1 + m_2} \mathbf{r}_1 + \frac{m_2}{m_1 + m_2} \mathbf{r}_2, \quad (13.4)$$

it can be shown that the velocity of the center of mass remains constant (and is usually assumed to be zero without loss of generality). Because the only two forces, \mathbf{F}_{12} and \mathbf{F}_{21} , are equal and opposite according to the form of (13.1) (and as required by Newton’s third law of motion), the above equation can be used to show that the velocities and accelerations of the two bodies are related as follows:

$$\mathbf{v}_2 = -\frac{m_1}{m_2} \mathbf{v}_1 \text{ and } \mathbf{a}_2 = -\frac{m_1}{m_2} \mathbf{a}_1. \quad (13.5)$$

By using the above and doing some algebraic manipulation on (13.3), the expression for the kinetic energy may be rewritten as

$$K = \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} v^2 \quad (13.6)$$

where

$$\mathbf{v} = \frac{d\mathbf{r}_1}{dt} - \frac{d\mathbf{r}_2}{dt} = \frac{d\mathbf{r}}{dt}, \quad (13.7)$$

and \mathbf{r}_1 and \mathbf{r}_2 are the distance vectors from the center of mass to m_1 and m_2 , respectively, and \mathbf{r} is the relative coordinate vector. (The vector directions of \mathbf{r}_1 and \mathbf{r}_2 are opposite, leading to the minus sign in the above.) Finally, rewriting (13.3) in total,

$$E_{\text{TOT}} = \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} \left(\frac{d\mathbf{r}}{dt} \right)^2 - G \frac{\frac{m_1 m_2}{m_1 + m_2} (m_1 + m_2)}{r}. \quad (13.8)$$

The quantity $\mu = \frac{m_1 m_2}{m_1 + m_2}$ is referred to as the reduced mass, and recognizing $m_1 + m_2$ as the total mass of the system, (13.8) can be written in final form as

$$E_{\text{TOT}} = \frac{1}{2} \mu \left(\frac{d\mathbf{r}}{dt} \right)^2 - G \frac{\mu m_{\text{TOT}}}{r}. \quad (13.9)$$

Thus, the total energy of the system is identical to that of a particle of mass μ moving in a potential well of strength determined by the total mass. It is well known that the solution to this differential equation is most easily expressed in plane polar coordinates, that is, $\mathbf{r} = (r, \theta)$, and will be one of three conic sections, depending on the total energy. If the total energy is greater than zero, a hyperbola is obtained; if $E_{\text{TOT}} = 0$ then the solution is a parabola; and if $E_{\text{TOT}} < 0$, an ellipse. The first two cases represent unbound orbits with no defined orbital period, whereas the ellipse of course does yield periodic motion. In that case, the period P is related to the semimajor axis a of the ellipse and the total mass using Newton's formulation of Kepler's harmonic law:

$$P^2 = \frac{4\pi^2}{G \cdot m_{\text{TOT}}} a^3, \quad (13.10)$$


where G is the universal gravitational constant. This shows for example that, if the period and semimajor axis can be determined, a total mass can be calculated.

2.2 Visual Binaries

2.2.1 The Visual Orbit

In the case of visual binaries, the angular separation, orientation, and brightness of the two stars relative to one another can be observed over time. If we define the brighter star in the system as the primary star and the dimmer as the secondary star, then the typical definition of the orientation is the angle between a line drawn from the primary to the celestial north pole and a line drawn from the primary through the secondary. North through east usually defines the positive sense of this angle, which is known as the position angle and is usually written as θ .

Over time, the astrometric data (separation and position angle) will trace out the relative orbit of the system projected onto the plane of the sky. In the plane of the true orbit, the motion will trace out an ellipse as discussed in the previous section, and when this shape is projected onto the plane of the sky, the result is also an ellipse. (That is, an ellipse viewed at an angle is still an ellipse, albeit one with different characteristics in general.) Two complications arise when attempting to determine the total mass of the system from the observational data, however. First, since the true orbital ellipse will not be in the plane of the sky in general, the angles that allow for the transformation between the apparent ellipse and the true ellipse must be known. Second, even after the geometry of the true orbit is known, the semimajor axis will still be an angular quantity, and therefore the distance to the system is needed to convert this value into an actual separation distance in linear units, as required for entry into (13.10).

The first problem has been elegantly solved by several astronomers dating back to the nineteenth century; a good example of a modern approach is that of Docobo (1985). These solutions show that seven parameters characterize the motion of the system on the sky. These seven parameters can be stated in various ways but encapsulate the transformation of coordinates between the plane of the true orbit and the plane of the sky, as well as the properties of the true ellipse (size, eccentricity, and orientation) in the plane of the true orbit. The most common formalism, which makes it very easy to compare the observed quantities to the seven orbital parameters, requires the definition of three angular quantities. The first of these is the inclination angle, i , which is the angle between the plane of the true orbit and the plane of the sky. Considering then the intersection of these two planes, a line is formed which cuts through the orbital ellipse and contains the location of the primary, which is a focus of the true relative orbital ellipse. There will be two points on the orbit that lie along that line: these are points of contact between the apparent and true ellipses, and are known as the nodes. (The line connecting the two points and containing the primary focus of the ellipse is known as the line of nodes.) Although some texts differentiate between the ascending node and descending node, the node that has a position angle between 0° and 180° is often referred to as the “nodal point,” or in some references, simply the “node,” written Ω . Finally, there will be some angle in the plane of the true orbit between the nodal point and the location of periastron (closest approach of the two stars), which lies along the semimajor axis of the ellipse; this is usually written ω .  Figure 13-1 illustrates the geometry of the situation. The three angular quantities – i , Ω , and ω – are generally reported in degrees. Four other parameters relate to the physical properties of the ellipse: (1) P , the period of the system (in days or years); (2) a , the semimajor axis of the system, in arc seconds or milli-arc seconds (mas); (3) T , the time of periastron passage (in units matching the period, either Julian date or Besselian year); and (4) e , the eccentricity of the orbital ellipse.

Adding the three geometrical parameters to the four physical parameters, the seven orbital parameters are then related to the observable quantities (ρ , θ , and t) according to the following equations:

$$\mu = \frac{360^\circ}{P} \quad (13.11)$$

$$M = \mu (t - T) = E - e \sin E \quad (13.12)$$

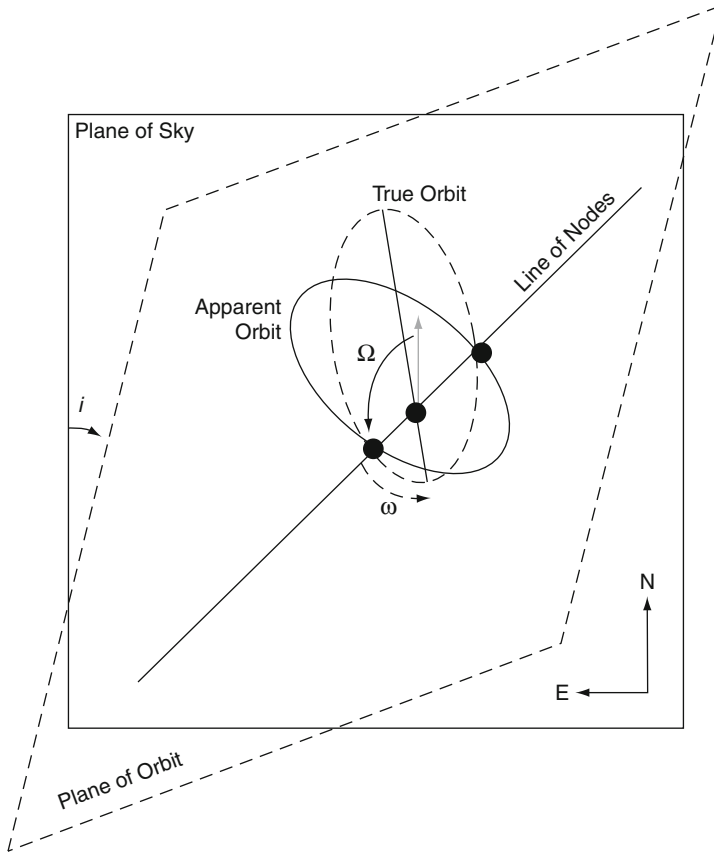
$$r = a (1 - e \cos E) \quad (13.13)$$

$$\tan\left(\frac{1}{2}\nu\right) = \frac{\sqrt{1+e}}{\sqrt{1-e}} \tan\left(\frac{1}{2}E\right) \quad (13.14)$$

$$\tan(\theta - \Omega) = \pm \tan(\nu + \omega) \cos i \quad (13.15)$$

$$\rho = r \frac{\cos(\nu + \omega)}{\cos(\theta - \Omega)} \quad (13.16)$$

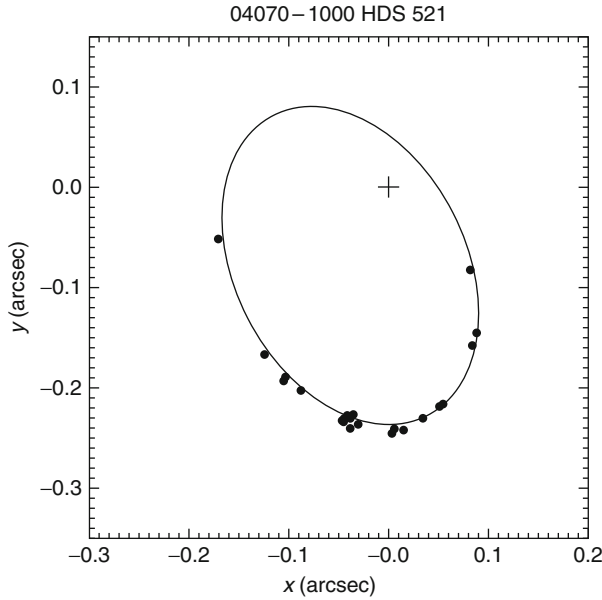
In the above, five intermediate quantities are mathematically defined. These are known as μ , the average number of degrees through which the companion travels per year, M , the mean anomaly, E , the eccentric anomaly, ν , the true anomaly, and r , the radius vector. For further information on these quantities, a good reference is Alzner (2004).



■ Fig. 13-1

Geometry for the calculation of a visual binary star orbit. In the plane of the true orbit, quantities are drawn with *dashed lines*, whereas in the plane of the sky, *solid lines* are used. The line of nodes, which marks the intersection of the two planes, is shown as a *solid line*

From the above relations, if one chooses a set of orbital parameters, then it is possible to compute the separation and position angle as a function of time. (That is, one may compute an ephemeris for the system). If a sequence of observations (ρ_i , θ_i , t_i) is available, then the ephemeris predictions may be compared with the observational results. Using various minimization algorithms, one can then choose new orbital elements in an iterative way until such point as the ephemeris predictions and the observational data are consistent, usually judged by computing a reduced chi-squared statistic. One typical difficulty is that the chi-squared parameter space is usually quite nonlinear, and if the initial parameters are not close enough to the final orbit, the procedure may not converge to the global minimum of the reduced chi-squared function. For this reason, techniques have been devised to search larger sections of the parameter space prior to minimization in order to insure proper convergence. Examples include simulated annealing (Pourbaix 2002) and grid searching (Hartkopf et al. 1996). ▶ Figure 13-2 shows an example of typical quality for speckle data; the orbit is that of Balega et al. (2007).



■ Fig. 13-2

The visual orbit of HDS 521, a binary discovered by the Hipparcos satellite and subsequently shown to have orbital motion through speckle observations over the last 15 years

In order to obtain the size of the orbit, the distance to the system must be determined in some way. The typical method is trigonometric parallax π , where the quantity (a/π) then gives the true system separation in astronomical units. Rearranging (► 13.10), it can be seen that

$$m_{\text{TOT}} = \frac{a^3}{\pi^3 P^2}, \quad (13.17)$$

with the mass sum then in solar units if the period is entered in years. Using standard error propagation formulas, the standard fractional uncertainty in the mass is then given by

$$\frac{\delta m_{\text{TOT}}}{m_{\text{TOT}}} = \sqrt{3 \left(\frac{\delta a}{a} \right)^2 + 3 \left(\frac{\delta \pi}{\pi} \right)^2 + 2 \left(\frac{\delta P}{P} \right)^2}. \quad (13.18)$$

This formula shows the fundamental difficulty of the problem of obtaining stellar mass information from visual binary stars: one must know both the semimajor axis and the parallax to very high precision if a small uncertainty in mass is desired. (There is generally no issue in determining the period to as high a precision as needed, at least compared to the other two quantities involved.)

2.2.2 Overview of the Observational Data

The most complete catalog of observations of visual binary stars is the Washington Double Star Catalog (WDS), maintained by the U.S. Naval Observatory, <http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/wds/WDS>. This resource includes over 700,000 measures

of approximately 100,000 systems and shows quite a range of techniques for obtaining the basic observed quantities over the long history of binary star observations. The great majority of observations taken before 1970 were using the technique of visual micrometry. The visual micrometer is essentially a customized eyepiece with a movable crosshair system that allows the observer to mark and measure the position angle and separation of the two stars. Past generations of binary star observers took great care in extracting the highest possible precision from this method and were able to measure systems far below the seeing-limited resolution in many cases due to their skill and visual acuity, in essence overcoming the blurring effects of the atmosphere in the images seen to deduce small separations. Aitken (1935) has a complete discussion of this instrument and its use. The precision of the best observers using this technique approached 10 mas in separation.

Nonetheless, the precision of the astrometry obtainable with visual micrometry has been superseded by the modern high-resolution techniques. These include interferometric techniques such as speckle interferometry, long baseline optical interferometry, Fine Guidance Sensor observations using the Hubble Space Telescope and other space-based observations, and adaptive optics. These techniques, particularly speckle interferometry, make up the bulk of the observations in the WDS since 1970, and a separate catalog of these observations is also maintained at the U.S. Naval Observatory, the Fourth Catalog of Interferometric Measurements of Binary Stars, <http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/wds/int4>. This catalog currently has over 100,000 observations of resolved double stars, with precision in separation from 1 to a few mas, depending on technique and telescope aperture.

A third resource again due to work at the U.S. Naval Observatory, is the Sixth Catalogue of Orbits of Visual Binary Stars, <http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/wds/orb6>. As of this writing, the catalog contains orbital elements for over 2,000 visual binaries. Most systems do have parallax determinations, but the quality of the orbital elements varies considerably, as do the distances, and estimates of the mass sums come with considerable uncertainty in the majority of cases. The catalog has a rating system of 1 (best) to 10 (worst) so that the user has some indication of the quality of the particular orbital calculation.

2.2.3 Speckle Interferometry

In speckle interferometry, the observation consists of several hundred to a few thousand short exposure images of the target. In each exposure, the atmospheric fluctuations are effectively frozen and the frame represents a distorted image of the target. The distortions are linked to the fact that the wavefronts of light passing down through the atmosphere from the target encounter variations in the refractive index that then lead to a flat wavefront becoming rough or corrugated. When the light enters the telescope aperture, the phase of the wavefront is roughly uniform across a distance of roughly 10–20 cm, known as the Fried parameter, but varies greatly from one side of a large telescope aperture to the other. This phenomenon is an example of what is known as partial coherence in statistical optics. One may then think of the large telescope aperture consisting of many sub-apertures, each of which has a phase uncorrelated with that of the other sub-apertures.

Since there are two relevant length scales on the plane of the aperture, namely the Fried parameter and the size of the telescope aperture itself, one may expect the same basic structure on the image plane. The result is an image that contains many small bright points, or speckles, that are confined within a much larger envelope of light, the seeing disk. The size of the speckles

is that of the diffraction limited spot size of the telescope aperture as a whole, whereas the size of the seeing disk is set by the size of the Fried parameter. Since optics tells us that the image is the modulus-square of the Fourier transform of the value of the electromagnetic field inside the aperture, there is an inverse relationship in the size scales: the larger the telescope aperture, the smaller the speckles, and the smaller the Fried parameter, the larger the seeing disk. As the atmospheric fluctuations change above the telescope, then the position and intensity of individual speckles changes within the seeing disk, though the overall size of the seeing disk itself remains relatively constant over typical observation times. In a long-exposure image, the speckle nature of the image washes out, even though in a short exposure image there is high-resolution information, as evidenced by the presence of the speckles themselves. On the other hand, if the exposure is short enough, then each star in the object will create a nearly-identical speckle pattern on the image plane (since the light from both stars travels down to the telescope through nearly the same column of air), and therefore for a double star, a double-image effect is present in individual frames.

In order to retrieve the high-resolution information, one must first have a camera capable of keeping up with the timescale of the atmospheric turbulence, which is on the order of 10–30 ms under typical observing conditions. The sequence of short exposure images (frames) taken during a typical speckle observation is needed to keep the contrast of speckles high – that is, to prevent significant blurring of speckles that would lead to reduced resolution overall. But, if this is achieved, then the path back to a high-resolution image is essentially a deconvolution problem, although not a trivial one by any means. The first step in the process is to form the average autocorrelation function of the frame sequence, given by:

$$\langle A_n(x, y) \rangle = \left\langle \iint dx' dy' I_n(x', y') I_n(x' + x, y' + y) \right\rangle, \quad (13.19)$$

where the brackets indicate an average over all image frames. It can be shown that this function retains diffraction-limited information. This can be most clearly seen by thinking of double stars of decreasing separation on the image plane. In wide doubles, where the separation is larger than the seeing disk, the system is classically resolved, though each speckle pattern will be nearly a copy of the other. As the separation becomes smaller than the seeing disk, then the speckle patterns begin to significantly overlap, but the presence of two stars can still be noted by the occurrence of speckle pairs at a given separation. That is what the autocorrelation function calculates. If a system is a double star, then peaks will form at the positive and negative vector separation of the secondary star (as well as at the origin), even if the two stars are only separated by a speckle width, i.e., at the diffraction limit.

Since the autocorrelation function is symmetric, it leaves an ambiguity in the position angle of the secondary star between the true position angle and the location 180° away from that value. This has been solved in various ways, including directed vector correlation (Bagnuolo 1988), shift-and-add techniques (e.g., Bates and Cady 1980), and bispectral analysis (Lohmann et al. 1983). In directed vector autocorrelation, when the autocorrelation function of a frame is computed, the number of counts in the two pixels (x, y) and (x', y') are compared, and the autocorrelation function is only incremented in the location corresponding to the direction of the brighter speckle to the fainter. In some cases, due to small sample statistics, one might obtain a brighter speckle in the location of the dimmer star, but when integrating over many speckles and averaging over many frames, an asymmetry will develop between the two peaks due the secondary in the autocorrelation function, with the higher of the peaks presumably corresponding to the true position angle. Shift-and-add techniques involve focusing on a

particular image feature such as the brightest speckle and lining up this feature in all frames. A final result is obtained by co-adding all shifted frames. In the bispectral approach, a second correlation function is computed called the triple correlation (or equivalently, one may compute its Fourier transform, known as the bispectrum). It can be shown that this data product preserves a diffraction-limited estimate of the derivative of the phase of the object's Fourier transform. In combining this information with that from the autocorrelation function (which when Fourier transformed is the modulus-square of the magnitude of the object's Fourier transform), one can reconstruct a diffraction-limited image. In most cases, this leaves no ambiguity in the position of the second star.

Speckle interferometry has been the workhorse technique in the observation of visual binaries since the 1970s because of its efficiency and robustness. Seasoned speckle observers routinely observe more than 200 objects per night, and relatively inexpensive, easily transportable cameras can be built to take the observations. Thus, the major speckle programs have followed thousands of objects over the last 30 years, yielding over 100 high-precision visual orbits, many from the long series of papers from the Center for High Angular Resolution Astronomy (CHARA) speckle program (see, e.g., Mason et al. 1997), and, more recently, including work from the Balega et al. (2006) and others.

2.2.4 Long-Baseline Optical Interferometry

Long-baseline optical interferometry has a long history in astronomy, beginning with the observations of Michelson in the 1920s of the diameters of several stars from the Mount Wilson 100-in. telescope. Michelson used an aperture mask that only accepted the light from two small but widely separated sub-apertures, bringing the light together on a common focal plane. This produces an interference pattern, but the fringe minima can be related to the diameter of the star, or alternatively, the separation of a binary system. A second important early contribution to high-resolution astronomy was the stellar intensity interferometer built by Hanbury Brown et al. (1967, 1970) and operated in the late 1960s and early 1970s at Narrabri in Australia. This system had two extremely large (6.5-m full diameter) mosaic telescopes that could be moved on tracks to different separations and orientations. At the focal plane of each telescope was a photomultiplier tube, which registered the detected photons per time interval. By correlating these intensity signals, it is possible to relate the result to the spatial structure of the source.

Observations of binary stars were greatly aided by the development and use of the Mark III Interferometer in the 1980s and early 1990s. This instrument was essentially of the Michelson type but could compensate for atmospheric fluctuations before bringing the light from two different apertures together to interfere. A number of spectroscopic binary stars were observed and resolved for the first time using the system; as discussed later in this chapter, the combined visual/spectroscopic orbit for a double-lined spectroscopic binary star gives a complete solution to the astrophysical quantities of interest (individual masses, luminosities, and effective temperatures) without recourse to an external distance measure, and indeed the distance also falls out of the solution. This was one of the first instruments to truly cross the divide between the visual and spectroscopic binary stars, and also pointed the way for future progress in the field of interferometry.

Current generation long-baseline optical interferometers include the Navy Prototype Optical Interferometer (NPOI) located at Anderson Mesa, near Flagstaff, Arizona, the CHARA Array at Mount Wilson, the Keck Interferometer, where the light from the two Keck 10-m

Telescopes on Mauna Kea is combined, and the Very Large Telescope Interferometer (VLTI) at the European Southern Observatory. Each of these systems has separation between the most distant stations (baselines) of 100 m or more, and therefore can achieve ultrahigh-resolution information of sources observed. While some of the work done with these systems has been on shorter period binary stars, these instruments have mainly focused of late on the problem of determining stellar radii and imaging stellar surfaces. Nonetheless, when used for binary work, they can produce visual orbital elements of exquisite quality, as in Muterspaugh et al. (2008).

2.2.5 Adaptive Optics

Adaptive optics is another technique that allows for high-resolution images (i.e., significantly higher than that of the seeing disk) to be collected of a target. The basic idea of this method is to correct for the corrugation of the wave front imposed by the atmosphere (and resulting in a seeing-limited image) in real time with a deformable mirror system. A common method is to split the beam and send part of a collimated beam from the target through a lenslet array. When the images from the lenslet array are recorded on a fast-readout detector, image shifts will be present due to the slope of the portion of the wavefront that is collected by each lenslet. From this set of image shifts, one may calculate the amount of deformation needed for the moveable mirror to re-flatten the wavefront from the target.

Adaptive optics has been highly successful in many areas of astronomy, but for precise binary star measurements, it has not been widely adopted. Initially, the technique was thought to avoid the so-called “magnitude difference” problem that plagued speckle interferometry through the mid-1990s. This effect was to render unreliable the magnitude difference between the primary and secondary star of a binary system obtained from the technique due to detector nonlinearities and observing constraints. However, it was found that reliable photometry from adaptive optics images is also not as easily obtained as first thought: it was still necessary to take 0.1 s exposure images to avoid systematic errors from the instrumentation itself (ten Brummelaar et al. 2000). Astrometrically, adaptive optics has not been shown to offer any advantage over speckle interferometry. However, the main advantage is that fainter targets may certainly be observed with adaptive optics, due to the fact that the correction to the wavefront is made in real time and a true integrating detector (such as a low noise CCD camera) may be used to collect the image. Also, many adaptive optics systems are optimized for use in the infrared so that the resolution is lower than typical results with speckle interferometry. Thus, the choice of instrument is influenced by the characteristics of the objects to be observed.

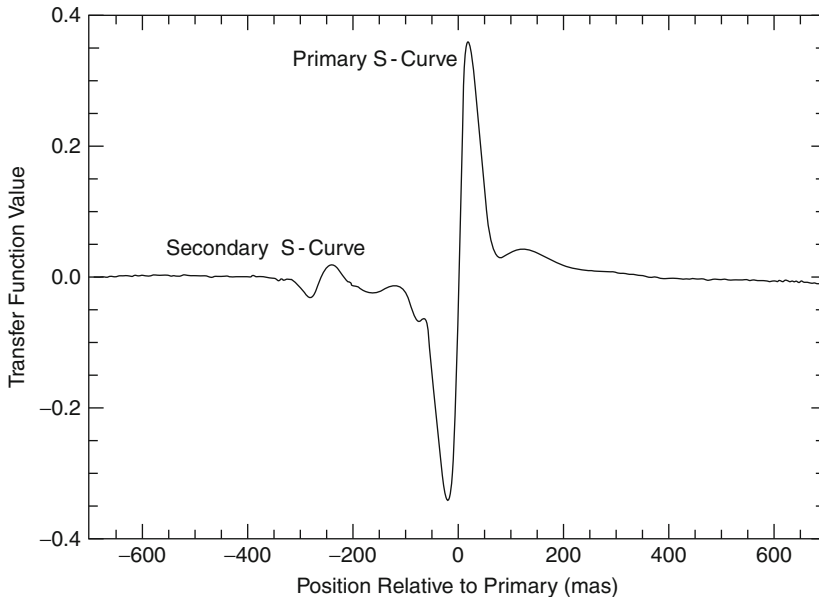
2.2.6 Space-Based Observations

Two important sources of space-based measures of binary stars are Fine Guidance Sensor (FGS) observations using the Hubble Space Telescope, and binary star measures by *Hipparcos*, the European astrometric satellite. The main function of the Fine Guidance Sensors aboard the Hubble Space Telescope is stable pointing of the telescope while other science cameras are in use. However, one of the three sensors can also be scheduled for science observations and has excellent characteristic for determining both astrometric and photometric properties of binary star

systems. The heart of the Fine Guidance Sensor system is a Koesters prism. It is an amplitude-splitting interferometer with two phototubes that collect the signal from the two exit faces of the prism. For small-separation binary star work, the instrument is scanned back and forth across an object, collecting signals at regular time intervals (nominally 1 ms). These signals are then combined into a transfer function

$$T(x) = \frac{A(x) - B(x)}{A(x) + B(x)}, \quad (13.20)$$

where $A(x)$ is the number of photons detected in phototube A as a function of scan position and $B(x)$ is the number of photons detected in phototube B as a function of scan position. When the signal is the same in both channels, as when the scan position is centered on the star, then the numerator of the transfer function will be zero, but as the scan position changes even slightly from that, then one can achieve the condition of constructive interference in one arms of the interferometer with destructive interference in the other arm. In those cases, the transfer function has value ± 1 , depending on which side of the zero point the telescope is pointing. If the light coming through the system were monochromatic, then a series of maxima and minima would be detected as a function of scan position, but the system is essentially panchromatic through the visible, leading to rapid loss of contrast in the fringe maxima and minima. In the end, what is recorded for typical stellar sources is a single maximum on one side of the center position and a single minimum on the other side. An example is shown in [Fig. 13-3](#). Because



■ Fig. 13-3

An S-curve of HD 157948 obtained using the Fine Guidance Sensors of the Hubble Space Telescope. Two components can be seen, illustrating the signature of a binary star. The primary of this system is actually a spectroscopic binary star, but the components are highly blended in the FGS scan

of the appearance, the transfer function is sometimes called the “S-curve.” Each point source in the field will generate its own S-curve offset from the zero position by some amount. Blended S-curves are separated through fitting a single star S-curve convolved with a binary star model.

The FGS system has two independent interferometers operating simultaneously in orthogonal directions. To determine the position angle and separation, information from both must be combined. Because of the simplicity of the image in the case of a binary star, complete position angle and separation information can be obtained, as well as two estimates of the brightness ratio of the system. Due to the sensitivity of the system, separations as small as 10 mas can be measured reliably, and stars as faint as magnitude 17 can be observed successfully. For this reason, FGS has been used to search for binary stars in nearby open clusters (Franz et al. 1992). FGS has also helped to define the lower end of the Mass-Luminosity Relation (Henry et al. 1999).

The *Hipparcos* Satellite was launched in 1989 and took both astrometric and photometric data of about 118,000 nearby stars until 1993 (ESA 1997). The satellite was designed to record data with a photomultiplier tube as stars passed over a grid on the image plane so that the signal obtained had a regular modulation in time for a single star. In the case of a close double star, the details of the modulation pattern would change, from which the separation and position angle could eventually be determined. The main result of the publication of the *Hipparcos* Catalogue was to give much better distances to stars than had been possible from the ground, but the satellite also serendipitously discovered about 3,400 previously unknown binary systems. Since the publication of the catalogue, binary star observers have made many follow-up observations of these stars to determine which ones are in fact binary stars and to determine orbital parameters for objects in this subset.

2.2.7 Lunar Occultation Measurements of Binaries

When the dark limb of the moon passes over a star, one may observe the change in brightness in time to reveal the classic diffraction pattern of a “knife edge,” where the exact moment of occultation by the Moon can be easily measured to high precision, if the detector is sufficiently fast. If the system is a close binary star, then the stars pass behind the Moon in rapid succession, leading to the superposition of two such interference patterns. The relative displacement of these patterns yields the projected separation along the direction of the Moon’s motion as well as the brightness ratio of the components. The *Catalogue of High Angular Resolution Measurements* (Richichi and Percheron 2002) contains hundreds of projected separation measures of binaries, some as small as 1–2 mas. In addition, the diameters of the stars can sometimes be estimated. The *Fourth Catalogue of Interferometric Measures of Binary Stars* discussed in ► Sect. 2.2.2 also contains occultation measures.

The value of such observations over time is place constraints on the masses, luminosities, and temperatures of the stars for stellar astrophysical studies. The main limitations are the lack of full-sky coverage, necessity to observe at a given moment, and the fact that only one-dimensional positional information is obtained, but on the other hand it is a very inexpensive technique and can be used to V magnitudes of as faint as 18. However, there has been a decline in the use of the technique in favor of long-baseline optical interferometry, which offers comparable resolution and no restrictions on sky position (although it is limited to much brighter stars).

2.3 Spectroscopic Binaries

2.3.1 The Spectroscopic Orbit

For a spectroscopic binary star, the observed spectrum will exhibit line shifts due to the Doppler effect. Consider the double-lined case. If the two stars are comparable in brightness and have similar line features, double lines will be observed in the spectrum when one of the two stars is moving toward the observer and the other star is moving away. Recalling the (nonrelativistic) Doppler formula, one has that

$$\frac{\Delta\lambda}{\lambda_0} = \frac{v}{c}, \quad (13.21)$$

where $\Delta\lambda$ is the shift of the wavelength relative to the rest wavelength for the line, λ_0 , v is the radial velocity of the component, and c is the speed of light. The shift $\Delta\lambda$ is taken to be positive when the object is receding, and negative when the object is approaching. Thus, the observational data yield velocities along the line of sight to the observer, that is, only one component of the velocity vector as a function of time. The orbital period of the system is easily determined from a sequence of velocities as the periodicity of the velocity variations is the same as that of the orbit.

The simplest case one can consider is that of two stars moving in circular orbits about the center of mass in a plane perpendicular to the sky, that is, viewed edge-on so that $i = 90^\circ$ (Zeilik and Gregory 1998). The stars in such a system will have uniform orbital speeds given by

$$v_1 = \frac{2\pi r_1}{P} \text{ and } v_2 = \frac{2\pi r_2}{P}, \quad (13.22)$$

where r_1 and r_2 are the radii of each orbit relative to the center of mass and P is the orbital period. The observed radial velocities will be the projection of the velocity vector as each star completes its orbit; suppose that at time $t = 0$, v_1 is directed toward the observer and v_2 is directed away, then since the velocity vector uniformly changes direction, the component along the line of sight in each case will be

$$v_{1,\text{obs}} = -v_1 \cos\left(\frac{2\pi}{P}t\right) \text{ and } v_{2,\text{obs}} = v_2 \cos\left(\frac{2\pi}{P}t\right). \quad (13.23)$$

On the other hand, the definition of the center of mass allows one to deduce that

$$m_1 r_1 = m_2 r_2. \quad (13.24)$$

Combining and rearranging these two equations, it is easy to show that

$$\frac{m_1}{m_2} = \frac{r_2}{r_1} = \frac{v_1}{v_2}. \quad (13.25)$$

The relative orbit in this case will have semimajor axis given by $a = r_1 + r_2$, and so rewriting Kepler's harmonic law (☛ 13.10) with the help of (☛ 13.21), we have

$$m_1 + m_2 = \frac{4\pi^2 \left[\frac{P}{2\pi} (v_1 + v_2) \right]^3}{GP^2} = \frac{P (v_1 + v_2)^3}{2\pi G}. \quad (13.26)$$

☛ Equations 13.24 and ☛ 13.25 represent two equations in two unknowns, namely m_1 and m_2 , and so individual masses can be obtained. The orbital radii can be obtained in a similar way, if the semimajor axis in (☛ 13.25) is expressed in terms of r_1 and r_2 , instead of velocities.

Most systems will neither have an orbital inclination of 90° nor a circular relative orbit, however. If the observed velocity variation is sinusoidal, then the orbits will still be circular and the orbital inclination merely adds a projection term into the observed velocities as follows:

$$v_{1,\text{obs}} = -v_1 \sin i \cos\left(\frac{2\pi}{P}t\right) \text{ and } v_{2,\text{obs}} = v_2 \sin i \cos\left(\frac{2\pi}{P}t\right). \quad (13.27)$$

Comparing the above with (◆ 13.24), one can see that it is still possible to derive the mass ratio of the system, though without the inclination angle, neither the total mass of the system nor the semimajor axis a can be determined. However, the quantity $a \sin i$ can be calculated, and this serves as a lower limit to the value of a itself. If the orbit has a nonzero eccentricity, then the shape of the velocity curves will be affected, but one can determine from the shape both the eccentricity and the angle between the major axis and the line of nodes in the true orbit plane (ω).

By plotting the velocities of the primary and secondary as a function of time, i.e., the velocity curve, it can be seen that the curves cross twice per orbital period. These points occur when the motion of both stars is tangential to the plane of the sky and hence neither has a component along the line of sight. Two examples are shown in ◆ Fig. 13-4. The motion of the system as a whole may still have a component along the line of sight so that these crossing points will not occur at zero-velocity in general. The amplitude of the variations gives the velocity of the component relative to the center of mass motion. The geometry of the orbit and its orientation relative to the plane of the sky determine the exact shape of the velocity curve; similar to the case of visual orbits, this leads to a convenient parameterization that includes (1) the orbital period P , (2) the angle in the plane of the true orbit between the line of nodes and the semimajor axis, ω , (3) the time of periastron passage T , (4) the eccentricity e , (5) the system velocity γ , (6) the velocity amplitudes K_1 and K_2 , and (7) the mass fraction $q = m_2/m_1$. It is also possible to derive from these data $a \sin i$, which is also usually stated with the above quantities. Note that four of these parameters are identical to those solved for in the case of a visual orbit.

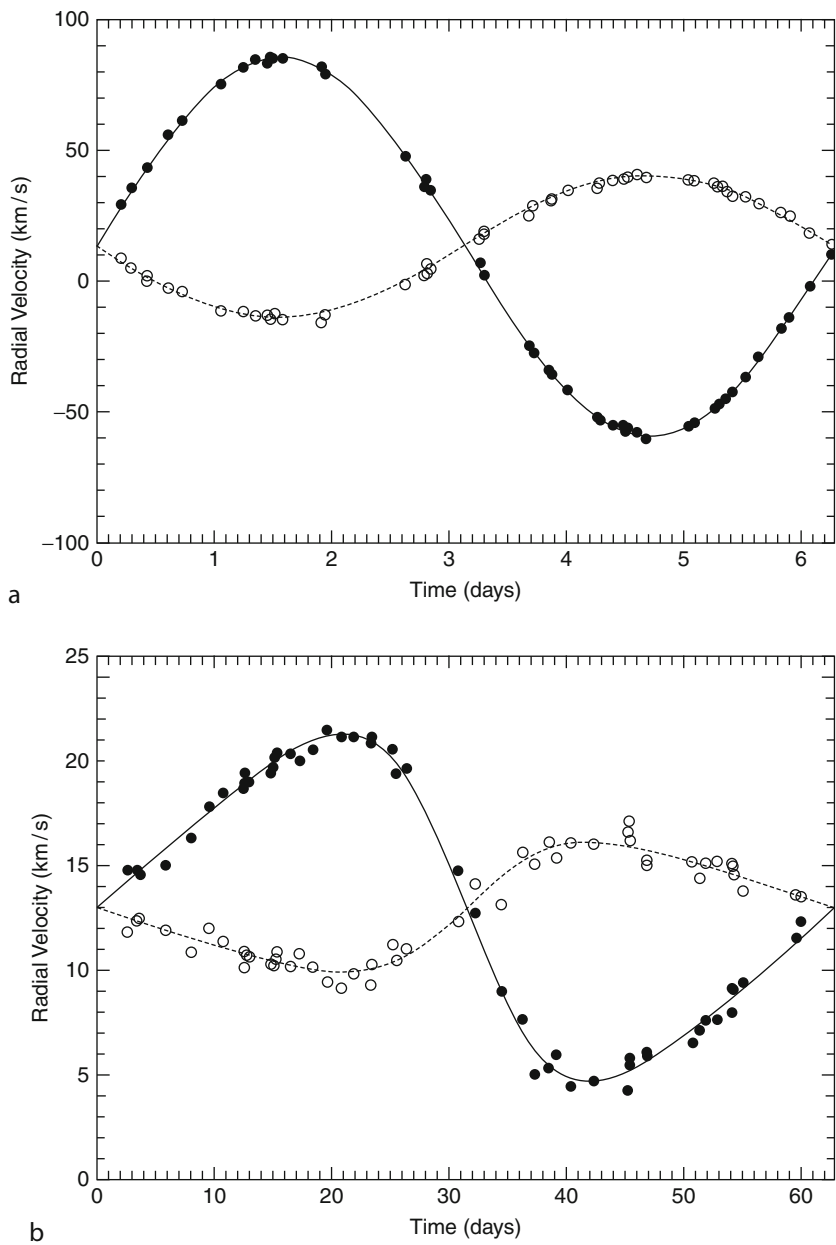
In the case of a single-lined spectroscopic binary, only the velocity curve of the primary is obtained. This makes it impossible to obtain the mass fraction of the system. Nonetheless, Kepler's harmonic law is still valid, as is the center of mass equation ((◆ 13.23) above). Therefore, one may write that

$$(m_1 + m_2) P^2 = (r_1 + r_2)^3 = r_1^3 \left(1 + \frac{r_2}{r_1}\right)^3 = r_1^3 \left(1 + \frac{m_1}{m_2}\right)^3 = \frac{r_{1,\text{obs}}^3}{m_2^3 \sin^3 i} (m_1 + m_2)^3. \quad (13.28)$$

This leads to the definition of the mass function, $f(m_1, m_2)$ as:

$$f(m_1, m_2) = \frac{r_{1,\text{obs}}^3}{P^2} = \frac{m_2^3 \sin^3 i}{(m_1 + m_2)^2}. \quad (13.29)$$

For a circular orbit, $r_{1,\text{obs}}$ can be obtained from the velocity amplitude and the period using (◆ 13.22). Although the mass function falls short of giving individual masses of the two stars, it can still be useful statistically. For a large sample of binaries, if one assumes that the inclination angle has a uniform random distribution from -90° to $+90^\circ$, then the expectation value of $\sin^3 i$ is 0.59. However, it is also easier to detect a system at high inclination since the projected velocity amplitude is greater. Therefore, the observed number will be biased higher. This can be modeled based on the velocity precision of a given spectrograph and other factors. Once estimated, then the remaining factors only depend on mass, and so a given distribution of masses can be used to generate a mass function and compared with the observational data.



■ Fig. 13-4
Schematic velocity curves of *double-lined* spectroscopic binary stars. (a) Curves obtained with a circular orbit. (b) Curves obtained with eccentricity 0.5

2.3.2 Overview of Spectroscopic Data

The most complete catalog of spectroscopic orbital elements is the Ninth Catalogue of Spectroscopic Binary Orbits, <http://sb9.astro.ulb.ac.be/>. The database contains orbits for some 2,400 systems. Quality is currently graded on a scale of 1 (worst) to 5 (best), as determined by those who have maintained the catalog, though the scale is soon to be changed in favor of a more objective scoring system. The data as a whole show several interesting characteristics. First, the orbital periods represented in the catalog span a very large range, from 10^{-1} to 10^5 days. As expected from Kepler's harmonic law, the longer periods correlate to the smaller velocity amplitudes. The data also show a relationship between period and eccentricity in the sense that longer-period systems span the complete range of eccentricity from 0 to 1, whereas short period systems (those with periods less than approximately 10 days) tend to have lower eccentricities. This is expected due to the tidal interactions that come into play when the two stars are in close proximity so that short-period systems become circularized on a time scale of one to a few billion years. Thus, it is worth noting that the simplification of a circular orbit used above is more useful than it might at first appear.

2.3.3 Observational Techniques

The range of techniques that have been used in the observation of spectroscopic binary star systems is much less varied than in the case of visual binary stars. Starting with the inception of the technique through the 1970s, spectra were collected on photographic plates. A major advance occurred when Griffin (1970) developed cross-correlation spectroscopy. In this method, the spectrum of the star is passed through a mask which transmits light only at wavelengths where prominent absorption features exist. The transmitted light is then fed into a photomultiplier tube. The mask can be moved back and forth very precisely until the number of counts registered by the photomultiplier tube is a minimum. This marks the point where the stellar absorption lines lie directly on top of the transmission windows on the mask. Precisions of less than 1 km/s can be determined in this way, and the successful development of the technique led later to the CORAVEL spectrometer of Mayor et al. (1983), which has been used for a number of important projects in binary star astronomy.

In many other spectrographs, CCDs have been used in the visible region of the spectrum and infrared arrays for wavelengths longer than the visible cutoff. The optical design of the spectrograph has remained essentially the same. A typical modern example is the Keck High Resolution Echelle Spectrometer (HIRES), which can achieve radial velocity precision of 20–30 m/s for spectroscopic binary star work. It is primarily used in the visible range of the spectrum. As its name indicates, it is a cross-dispersed instrument, meaning that different pieces of the spectrum are focused in strips onto a single large format detector.

A variation on this theme important for binary star work is the multi-object spectrograph, for example, the Hydra instrument at the WIYN 3.5-m telescope, located at Kitt Peak National Observatory in the United States. This instrument has the ability to configure the locations of approximately 100 optical fibers on the image plane so that each can collect the light of one star. This is particularly useful in studying open star clusters (Mathieu 2000). Radial velocities obtained in this way are not as precise as with the Keck instrument described above; they are in the range of a few tenths of a km/s. However, this is still quite sufficient for the detection of spectroscopic binaries and determination of their orbital elements.


Infrared spectrometers have become extremely important in binary star astronomy in the last two decades for two reasons. First, it allows for the observation pre-main sequence binaries, which are often very red and therefore bright in the infrared. Thus, the determination of spectroscopic orbital elements in addition to a visual orbit can give vital mass and dynamical information about this critical phase of stellar evolution. Second, in a single-lined spectroscopic binary, the secondary is too faint to reliably measure the location of the absorption features in the spectrum. If the binary consists of two main sequence stars, this implies that the secondary is also much redder than the primary. If the system were observed in the infrared, the contrast between the secondary and primary would be reduced. This leads to the possibility of detecting the secondary in the infrared, that is, effectively turning the system into a double-lined system by observing it with an infrared-sensitive spectrometer (Mazeh et al. 2002). A good example of a spectrometer that has been used for binary work of both kinds just described is the Keck Near Infrared Spectrometer (NIRSPEC) (e.g., Prato et al. 2002). This again is an Echelle system, with nine orders of the spectrum focused onto a $1,024 \times 1,024$ infrared array centered on wavelength $1.55 \mu\text{m}$, with typical precision of better than 1 km/s .

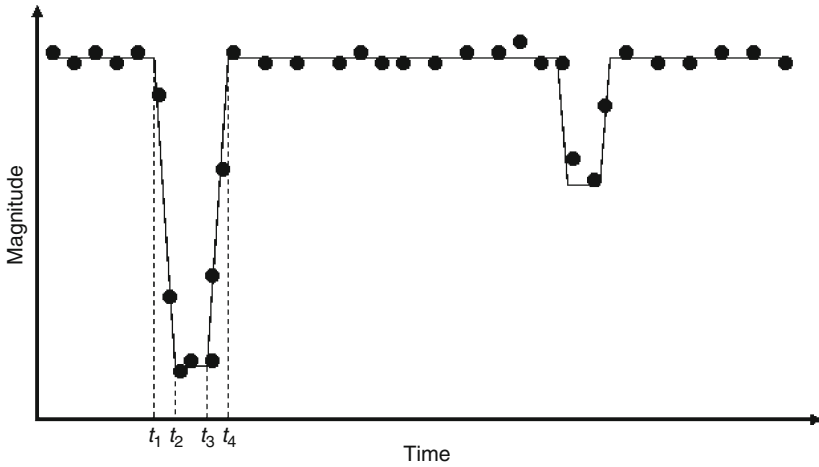
In the reduction of spectra, the most commonly used method is to establish the degree of correlation between the observed spectrum and template spectra. For example, the TOD-COR algorithm (Mazeh and Zucker 1994) works by starting with a library of single-star spectra and assumes that the observed spectrum is a combination of two spectra from the library with different fluxes and velocities. A cross-correlation is performed for various pairs of template spectra, which are combined to approximate the observed spectrum. When the best fit is found, the velocities of primary and secondary are inferred. This is in some sense a variation on the mask idea of Griffen discussed above, but carried out when the data are reduced, not during the observation.

Finally, it is important to note the contribution of recent and ongoing spectroscopic surveys done for other purposes on the knowledge base of binary stars. Two examples are the Geneva–Copenhagen survey (Nordström et al. 2004) and the Radial Velocity Experiment (RAVE) (Steinmetz et al. 2006). These are studies primarily aimed at the construction of large databases of kinematics, metallicity, and relative abundance information for galactic structure studies, but they have also discovered a number of spectroscopic binary stars. In some cases, mass fractions have also been obtained, though without complete orbital information. Searches for extrasolar planets have likewise identified new binaries via spectroscopic and photometric studies. These samples provide new binaries that could be used in the future for comparisons with stellar structure and evolution theory, for example.

2.4 Eclipsing Binaries

2.4.1 Orbital Information Determined from Observation

Eclipsing binaries are detected through photometric means; several thousand such systems are known. If one takes a sequence of observations of the magnitude of the object in time, dips in brightness can be observed at regular intervals. This is shown in  Fig. 13-5. There will be two dips per orbital period. When the cooler star passes in front of the hotter star, this is known as the primary eclipse and will be the deeper of the two minima in light. The secondary eclipse occurs when the cooler star passes behind the hotter. By definition, in order to observe



■ Fig. 13-5

Schematic of the light curve of an eclipsing binary star. The points t_1 through t_4 , described in the text, are shown for the deeper eclipse

an eclipse, the orbital plane must be perpendicular to the plane of the sky, or very nearly so. Primary and secondary eclipses can be either total (if the star in front is larger), annular (if the star in front is smaller), or partial if the orbital inclination angle differs slightly from 90° . Total and annular eclipses result in flat-bottomed minima in the light curve, while partial eclipses have rounded bottoms. This same method is now used also in the detection of extrasolar planets.

The orbital geometry will in general determine the spacing and length of the eclipses. If we take the case of a circular orbit viewed edge-on, then the spacing between eclipses will be half the orbital period and the length of both primary and secondary eclipse will be equal. Information about the diameters of the two stars can then be read off of the light curve. The orbital velocity of the circular orbit will be constant so that

$$a = \frac{vP}{2\pi} \quad (13.30)$$

If one defines four points on the light curve as the onset of the eclipse (first contact, t_1), the onset of the eclipse minimum (second contact, t_2), the end of the eclipse minimum (third contact, t_3), and the end of the eclipse (fourth contact, t_4), then the radius of the smaller star will be

$$R_s = \frac{1}{2}v(t_2 - t_1) = \frac{1}{2}v(t_4 - t_3), \quad (13.31)$$

where v is the orbital velocity and R_s is the radius of the smaller star. Also, one may write that the sum of the two stellar radii is given by:

$$R_l + R_s = \frac{1}{2}v(t_4 - t_1), \quad (13.32)$$

where R_l is the radius of the larger star. On the other hand, the orbital velocity is related to the semimajor axis of the relative orbit as

$$v = \frac{2\pi a}{P} \quad (13.33)$$

And so, the two radii can be expressed finally as

$$\begin{aligned}\frac{R_1}{a} &= \frac{\pi(t_4 - t_2)}{P} \\ \frac{R_2}{a} &= \frac{\pi(t_2 - t_1)}{P}\end{aligned}\tag{13.34}$$

An estimate of the size of the orbit is needed to determine the radii, but if the system is also a double-lined spectroscopic binary, this would be available from the spectroscopic orbit solution, or alternatively, if the system has a visual orbit and the distance is known.

2.4.2 Observational Data

There are three main catalogues of information for eclipsing binary stars. Both the General Catalog of Variable Stars (GCVS, <http://www.sai.msu.su/groups/cluster/gcvs/>), maintained by the Sternberg Astronomical Institute, Moscow, Russia, and the American Association of Variable Star Observers webpage (<http://www.aavso.org/>) list a large number of variable stars, including eclipsing binaries. Recent observations may be plotted with the software tools available on these websites. On the other hand, ephemerides of eclipsing binaries may be found at <http://www.as.up.krakow.pl/o-c/cont.html>, a website maintained by the Mt. Suhora Astronomical Observation in Poland.

2.5 Complementary Observations

There are two important cases of complementary observations of binaries, that is, occasions when a system is observable using more than one of the techniques discussed above. The first case is that of a spectroscopic and eclipsing system. The fact that the system is eclipsing indicates that the inclination angle, i , is 90° , or very nearly so. If the spectroscopic orbit is double lined, then the true orbital velocities can be determined from (● 13.25) above, and therefore the total mass follows from (● 13.26). The individual masses can then be obtained. Also, the absolute dimensions of the system are completely determined: the eccentricity, the true semimajor axis, and, from the center of mass equation, the individual distances from the center of mass. Furthermore, a distance to the system can at least be estimated by using the method of spectroscopic parallax, that is, measuring the spectral type of the larger star (e.g., during eclipse when the light from the smaller star is blocked), inferring the absolute magnitude, and the measuring apparent magnitude of the system during eclipse of the smaller star. Surface temperatures may also be inferred from the spectrum: the smaller star's spectrum may be obtained by subtracting that of the larger star from the composite spectrum, taken when neither star is in eclipse. Since the light curves of the eclipses also give the relative sizes of the two stars, the relative densities follow from the masses. The case of a single-lined spectroscopic binary with eclipses, the mass function may be rewritten without the $\sin^3 i$ term (since the inclination is 90°), but individual masses are still not directly determined. If one is prepared to make an assumption about the mass of the primary based on its spectral type, then one can at least estimate the secondary mass.

The second complementary case is that of a spectroscopic and visual binary star. Here, the visual orbit gives the inclination angle, so again, if the system is double-lined, then the individual masses are again determined: the mass sum from the visual orbit and the mass ratio from the spectroscopic orbit. A direct measure of the distance to the system is also obtained without

recourse to parallax or spectrum analysis since the semimajor axis is determined from both orbits. The quantity $a \sin i$ is determined from the spectroscopic orbit, but with the inclination provided from the visual orbit, a in absolute distance units is known, while a in angular units comes from the visual orbit, thus the distance follows from geometry. In the case of a single-lined system, individual masses can still be determined since the inclination angle and mass sum can be used from the visual orbit to make the only unknown in the mass function the secondary mass. However, the distance to the system must be determined from some other means if the true dimensions of the orbit are desired, and the parallax uncertainty is likely to dominate the precision possible in this case.

3 Basic Stellar Astrophysics with Binaries

3.1 The Mass–Luminosity Relation and Mass–Radius Relation

Since binary stars give information about the masses and radii of stars, they are an excellent laboratory with which to test stellar structure and evolution theory. However, to be meaningful for the understanding of stellar astrophysics, masses must be combined with other observables such as luminosity, effective temperature, metallicity, and helium abundance. The most basic relationship that is provided by the theory and verified with binary star observations is the mass–luminosity relation (MLR). The current understanding of the empirical mass–luminosity relation (MLR) begins with the study of eclipsing and spectroscopic binaries, which has yielded excellent results for stars more massive than the Sun. However, the relation is on much less firm ground for lower mass stars (see e.g., Andersen 1991; Henry and McCarthy 1990, 1993; Popper 1980). The classic work of Henry and McCarthy resulted in some of the best infrared speckle data of low-mass stars on the systems that are closest to the solar system. These data are still comprised of observations of surprisingly few systems; when limited by ground-based observations as they were, there were few systems that were both bright enough and close enough to observe with their instrumentation. Recent improvements in the MLR have been made with the incorporation of Hipparcos parallaxes (Martin et al. 1998; Söderhjelm 1999), and Space Telescope results (Henry et al. 1999, Torres et al. 2010), but again these are primarily mass redeterminations of well-known systems. Henry et al. gave us a significantly clearer picture of the lowest mass systems, but the range 1.0–0.2 solar masses still remains extremely uncertain. More data are still needed to improve this portion of the diagram, especially to investigate so-called “second parameter” effects in the MLR such as metallicity and age.

The recently published Geneva–Copenhagen spectroscopic catalog is a substantial resource in this area, providing metallicities and some mass fractions for many F and G stars of interest. For example, HDS 521 (shown in ► Fig. 13-2) has $[\text{Fe}/\text{H}] = -0.27$ and mass fraction 0.773 ± 0.110 . From these data, as well as the orbital elements and photometry from speckle observations, individual masses, luminosities, and colors are obtained (e.g., mass values of 0.83 and 0.76 solar masses can be obtained from the orbit in ► Fig. 13-2, with uncertainties dominated by parallax). Metallicity information exists in other catalogs as well including for some giants and red dwarfs, for example, the $[\text{Fe}/\text{H}]$ catalog of Cayrel de Strobel et al. (2001) and the ELODIE catalog (Prugniel and Soubiran 2001).

Andersen (1991) discussed what is needed observationally before meaningful comparisons with stellar theory can be made. Measures of mass should have uncertainties of less than 2%,

radius uncertainties, less than 1% (note that the radius can be obtained from the luminosity via $L = 4\pi R^2 \sigma T_{\text{eff}}^4$); effective temperature uncertainties, less than 2%, and metallicity uncertainty, less than 25%. With reliable speckle photometry and multiple observations, combined with recent metallicity determinations, only the masses fall short of these criteria in many cases. However, if the components of a binary can be placed on the H-R diagram, there can be a meaningful comparison made from the photometry and metal abundance alone especially in the case of at least one evolved component. Turning the problem around, one may establish preliminary masses for a system if one has good confidence in a particular set of theoretical isochrones, that is, the locus of points that in theoretical calculations have the same age for a range of mass, and photometry. These could then be used to form a statistical study that correlates the masses with other parameters.

In eclipsing-spectroscopic systems, the length of time in total or annular eclipse can be used to obtain the radius of each star, as discussed above. This gives another opportunity to compare directly with theoretical predictions of the size of stars as a function of mass. The sample of systems that can be used in this way is relatively small; however, data on non-eclipsing binaries is now possible with the current generation of long-baseline optical interferometers discussed in ▶ Sect. 2.2.4, which has increased the sample of M dwarfs by 50%, for example. Observed values of the radii suggest that the current theoretical models underestimate the radii of low-mass stars, particularly those of higher metallicity (Berger et al. 2006).

3.2 Formation of Binary and Multiple Systems

There have been three main theories of binary formation that have been studied over the last two decades: (1) capture, (2) prompt fragmentation, and (3) delayed breakup (Tohline 2002). In capture, stars form as initially single, and through gravitational interactions result in binaries after the fact through stellar encounters. However, if the two stars are not gravitationally bound to begin with, then the total energy of the pair is greater than or equal to zero. Therefore, there must be some loss of energy to create the bound orbit. For example, the encounter might involve a third star, and through the dynamics of the situation, this third companion might receive enough kinetic energy to reduce the total energy of the other two stars below zero, binding them together. With the extra velocity imparted to the third star, it would soon escape the vicinity of the new binary. Another mechanism that has been suggested is the dissipation of energy through the tidal interaction between one star with the protostellar disk of the other star (Larson 2002). However, if the relative velocity between the two approaching stars is too high, then the outcome is simply the disruption of the disk as the stars pass each other and not enough energy is removed to result in a bound orbit (Clarke and Pringle 1993).

In the prompt fragmentation picture, during or soon after the free-fall collapse of the rotating protostellar gas cloud, the cloud breaks into two pieces, transferring much of its spin angular momentum to the orbital angular momentum of the two protostars. A free-fall collapse that proceeds from a nearly spherical spinning gas cloud that is uniform in density is known as a homologous collapse, while if the density is concentrated toward the center of the cloud, the collapse is referred to as nonhomologous. Through numerical studies, it is found that homologous collapses can lead to configurations at the end of free fall collapse that are susceptible to fragmentation, though pinning down the general conditions under which such fragmentations to occur has been difficult to accomplish. Since many of these calculations have not continued

to the point of actual star formation, conclusions about binary formation from this method have been somewhat indirect.

In delayed breakup or disk fragmentation models, the disk around a central object can become unstable to non-axisymmetric perturbations if the accretion rate onto the disk from the surrounding gas cloud is high enough (Bonnell 1994). Also, the equation of state of the gas is an important consideration, with stiffer equations of state less likely to lead to fragmentation. Work on both prompt fragmentation and delayed breakup has been able to produce binary and multiple systems, which leads to current thinking of fragmentation as the dominant formation mechanism. A current incompleteness of these models however is the inability to produce close binary systems (those with separations of less than 10 AU). This limit is imposed because the pressure-supported core that is formed at the end of free fall collapse is stable against fragmentation, and such an object has a typical size of 4 AU and mass of a few Jupiter masses. Therefore, the formation of small separation binaries is almost certainly the result of interactions with other stars that lead to a decrease in orbital distance.

In the last 15 years, some specific predictions of binary formation theory have emerged that have been tested observationally. For example, some fragmentation models (e.g., Clarke 1999) predict that the properties of the secondary in a binary system should be independent of the primary mass, while in capture models (McDonald and Clarke 1993), the higher-mass stars have companions with greater frequency. These different predictions can be evaluated by observing the companion star fraction (CSF) for a sample of appropriate stars. This is defined as follows:

$$\text{CSF} = \frac{B + 2T + 3Q + \dots}{S + B + T + Q + \dots}, \quad (13.35)$$

where S is the number of single stars, B is the number of binaries, T the number of triples, Q the number of quadruples, and so on. In general, imaging methods (including speckle interferometry) can be used to determine the CSF of a sample, if orbital motion is confirmed or may reasonably be assumed. If orbit determinations are possible, then further statistics such as the mass ratio distribution, the period distribution, and correlations between the orbital parameters can be observationally determined. Spectroscopic binaries have provided much of this information to date due to their generally shorter periods. If orbits are not determined, the mass-ratio distribution can still be obtained indirectly, if the magnitude differences of the systems and the evolutionary status of the companions are known.

Bate et al. (2003) has completed very detailed calculations of star formation in a cluster environment. Starting with a spherical gas cloud with density fluctuations built in, the simulations allow for the concurrent collapse of many different subregions. (These simulations can be viewed from <http://www.astro.ex.ac.uk/people/mbate/>.) As the calculation proceeds, one finds that the density fluctuations evolve into high-density filaments along which stars begin to form. As they form, the stars interact gravitationally, leading to phenomena such as the disruption of circumstellar disks, the formation of binary systems with a ring of circumbinary material, and so on. Bate et al. finds that the CSF produced by his simulation is in agreement with that determined observationally for young clusters, though the statistics are quite low. This work firmly establishes the fact that binary formation is a natural channel of star formation, perhaps even the dominant channel.

Speckle interferometry has been a useful method for examining the CSF in young clusters and star-forming regions. In this case, all stars in the sample may be thought of as coeval, having the same metallicity, and equidistant from the solar system. The same studies that gave information about H-R diagram position of pre-main sequence components in the early 1990s

(Ghez et al. 1993, 1997; Leinert et al. 1993), as well as optical speckle interferometry was also used to determine binary statistics. The Space Telescope fine guidance sensors have also been used in the study of cluster binaries (e.g., Franz et al. 1992), where the ages are well known. These studies have generally shown that young stars have a very high CSF, statistically consistent with the proposition that the majority of stars form in multiple systems.

More recent work by Patience et al. (1998) has given a significantly clearer picture of the CSF and mass-ratio distribution in the Hyades cluster, allowing for a meaningful comparison of theory and observation. In particular, they ruled out capture and disk-assisted capture as primary mechanisms of binary formation on the basis of mass-ratio statistics of the cluster binaries. This and other studies point to fragmentation as the primary mechanism of binary formation. It is not fair to say that there is universal agreement on this issue, since, e.g., Abt and Willmarth (1999) have published spectroscopic results, indicating evidence for capture as the primary binary formation mechanism in some open clusters.

3.3 The Algol Paradox

Although capture of a companion is one way in which a binary system can be formed in theory, current work suggests that in most cases the two stars in the system were formed at the same time and have been gravitationally bound since that time. Therefore, at least in most binary systems, the components should have the same age. This hypothesis can be checked if enough information is known to place the two stars on the H-R diagram. When this was first done for the Algol system (β Persei), it was found that the less massive star had evolved off of the main sequence, while the more massive star was still unevolved. Since stellar structure theory predicts that the time spent on the main sequence is proportional to $m^{-3.5}$ where m is the star's mass, this situation seems paradoxical at first.

The generally accepted solution to this dilemma is that the two components form an interactive system where the star that was originally more massive has donated mass to the less massive component so that over time the masses have reversed. In this picture, the star that was more massive evolved off the main sequence, but in doing so became much larger in size. Because of the presence of a relatively close companion, the surface of the evolved star can become distorted, and some material can be pulled toward the companion, adding to its mass. Over a long time scale ($\sim 10^9$ years), enough material can be exchanged to result in a lower mass for the evolved star than the companion.

3.4 Other Tests of Stellar Evolution

A small number of very detailed comparisons between stellar models and observational data exist; often these studies focus on one or a few systems which have been observed with multiple technique. A recent example is the work of Torres and his collaborators on Capella (α Aurigae) (Torres et al. 2009). Capella is an example of a spectroscopic and visual binary, although the separation of the system can only be precisely measured with interferometric techniques. By virtue of both a spectroscopic and visual orbit, it is possible to obtain individual masses and a distance to the system. Both stars in the system have evolved off of the main sequence, making the system a particularly interesting test of stellar structure and evolution models, since the evolution is presumably proceeding quite rapidly for both stars. Torres et al. present an exhaustive

analysis of all of the data taken of the system over a 100-year time frame, leading to fractional uncertainties in the masses of 0.7% and 0.5% for the two stars. They find that it is not possible to match the H-R diagram positions with the derived properties of the two stars with any of three different stellar evolution calculations of leading theoretical groups. This discrepancy is both exciting and perplexing: there are few systems that allow such high precision in the masses and other physical parameters to be determined, but no solution so far exists to explain the discrepancies between theory and observation in this case.

While high-precision studies are extremely valuable for the comparison with stellar evolution calculations, there is also a need for studies that span a meaningful range in parameters such as metallicity, helium abundance, and so on. Because of the distribution of Population I and Population II stars in the Galaxy, obtaining a meaningful sample of lower-metallicity systems necessitates looking at targets that are more distant and therefore necessarily fainter on average. Often, these systems are beyond the capabilities of the long-baseline optical interferometers. Thus, one must ask what progress can be made with lower-resolution techniques that can successfully be used to observe these fainter targets.

As discussed earlier, the most widely used technique for visual binary star observations over the last 30–40 years has been speckle interferometry due to its relative ease of use and efficiency (i.e., the ability to observe 200 or more objects per night). However, one of the main limitations of the method through most of its use has been the difficulty in obtaining good quality magnitude differences from the data. This is due to a combination of factors, including nonlinearities associated with the detectors that have commonly been used as well as atmospheric effects. In the last decade, these effects have been studied and overcome by some observers, leading to magnitude and color information for the components of many binary systems. Combining this information with total magnitudes and colors available from normal photometric surveys, the components can then be placed on the H-R diagram. An example of this type of work can be found in Davidson et al. (2009). Masses of the components can be derived by fitting a theoretical isochrone to the two points. In systems with known masses from orbit calculations, this method can be used to compare the photometrically determined values with the dynamical ones, which in principle places constraints on the theoretical models. (It would be assumed in this case that the system is coeval.) In systems with no previous mass determinations, the method could be used to predict mass values; it would not be necessary to wait for orbital determinations, although here one is relying on the correctness of the stellar models that generate the isochrones. Davidson et al. found that the photometric masses were consistent with the dynamical masses and generally had much lower uncertainties. As a way to obtain much more complete mass information for a large sample of close visual binaries, speckle photometry would appear to show significant promise.

4 Statistics of Binaries and Implications

4.1 The Field Population

Current knowledge of the field population in the solar neighborhood has been primarily determined through spectroscopic means. Abt and Levy (1976) pointed out that binary and multiple stars systems are more common than single stars, a conclusion which has been confirmed by Duquenooy and Mayor (1991) in their distance-limited sample of solar-type stars. A subsequent


update (Udry et al. 1998) more than doubles the number of orbits in the 1991 paper. However, the field CSF seems to be below that measured in the star-forming regions, leading to the notion that many stars in the field population have been stripped of their companions. Mason et al. (1998) compared chromospherically active and non-active stars in the field via speckle interferometry to gain further insight on the CSF as a function of age. They found that the active (and presumably younger) stars have a higher CSF than the non-active stars. On the other hand, Duchêne (1999) reanalyzed some of the previous studies of star-forming regions and concluded that some regions do not have binary excesses. Thus, while there has been an evolving discussion in the field in recent years, there is no firm consensus yet on what the basic properties of the binary population tell us about formation scenarios.

Duquennoy and Mayor (1991) found that solar type stars have a period distribution that is given by a log-normal function with average period of approximately 180 years. However, others have argued that, at least for wider separations, that is, periods much larger than 180 years, the observed distribution follows the Öpik relation (1927), a scale-independent law with decreasing frequency at larger separations. However, observational data on wide binaries is far from complete. Often, they are identified only as pairs with common proper motion, as suggested by Luyten (1927), but due to the timescale of the orbit implied, one does not have orbital elements to confirm binarity or have a firm idea concerning the orbital period. Nonetheless, several thousand binaries identified in this way have been catalogued (Luyten 1997). Another important feature of the spectroscopic results of Duquennoy and Mayor and others is the period–eccentricity relation. For short-period systems (less than about 10 days), virtually all systems observed have near-zero eccentricity. However, for larger periods, the eccentricity distribution is essentially random, ranging from 0 to 1. This is explained by tidal interactions between the stars since, in a short-period eccentric system, the two stars could come quite close together at the point of closest approach. The creation of the tidal bulges dissipates energy in the system, leading to a more and more circular orbit over time.

The Hipparcos mission and subsequent release of the Hipparcos Catalogue (ESA 1997), substantially improved parallaxes of stars within a couple of 100 pc of the Sun. Hipparcos parallaxes have typical uncertainties in the range of two to five times smaller than ground-based parallax results (i.e., from 1 to 2 milliarcseconds [mas]). This was an extremely important development for visual binary star astronomy, but as impressive as this achievement is, the dominant source of mass uncertainty remains the parallax for most speckle binaries. Take, for example, the well-known binary BU 151 (HIP 101769) has Hipparcos parallax of 33.49 ± 0.88 mas, and the most precise orbital parameters are those of Hartkopf (1998), where $a = 439.4 \pm 0.6$ mas and $P = 26.6327 \pm 0.0110$ years. The fractional errors in the observed quantities are therefore $\delta a/a = 0.00137$, $\delta P/P = 0.00041$, and $\delta \pi/\pi = 0.02628$. The parallax error is nearly a factor of 20 larger than the other two contributors, even for this system which is only 30 pc from the sun, and even post-Hipparcos.

The real solution to this situation in this and many other cases is the Gaia satellite mission to be launched in the coming years. This mission plan includes determination of parallaxes to ~ 20 microarcseconds (μ as) for stars brighter than 15th magnitude, or in other words, roughly 50–100 times smaller than Hipparcos. In the case of many resolvable binaries, the parallax uncertainty will then be comparable to or less than that of the semimajor axis. For BU 151AB, the fractional error in parallax using an estimate of $\delta \pi/\pi = 20 \mu$ as becomes 0.00060, and the overall error budget from the three observables would be less than 0.5%, approximately ten times better than the current value (using the Hipparcos parallax) of 4.6%. This exercise demonstrates the importance of space-based astrometry in unlocking the power of the speckle

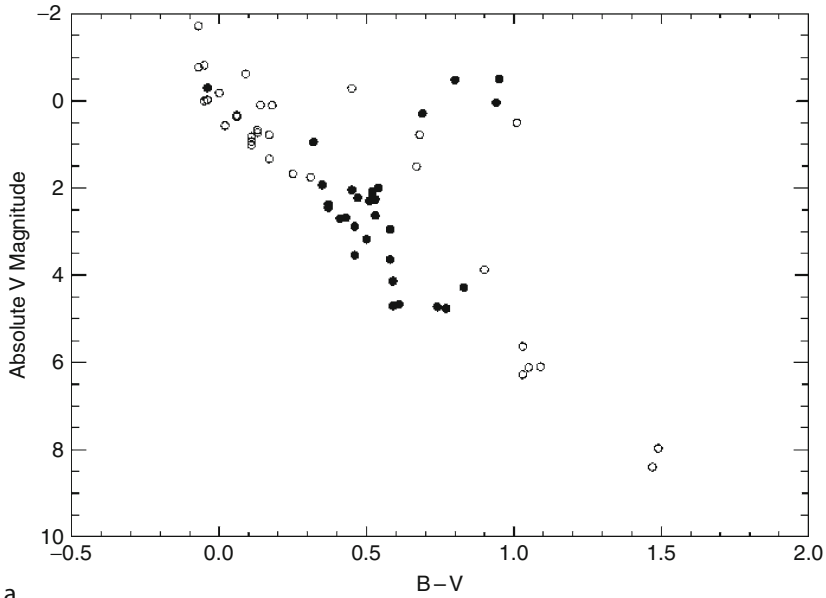
binaries for astrophysics, and there are dozens of other bright, well-known systems awaiting this kind of revision. In addition to the speckle binaries, Hipparcos discovered some 3,400 new double-star systems (the Hipparcos Double Stars, or HDSs), and there has been a push by the major speckle groups to observe and characterize these and other recently discovered systems. Orbits derived from speckle astrometry have already been determined in ~ 20 cases (Balega et al. 2002, 2005, 2006; Cvetkovi 2008; Horch et al. 2010; Söderhjelm 1999). This already represents ~ 40 new points on the MLR, and there is at least some range in metallicity represented. With a few more years of observing, this number will substantially increase. This work represents a large addition to the speckle binary sample, and binary star observers continue to take data on these systems in order to have the best possible orbital elements ready when Gaia parallaxes become available.

However, the sample of speckle binaries with the best quality orbital elements (say, those with uncertainty in semimajor axis of less than 1%) consists mainly of systems within about 100 pc and is limited to mainly main sequence systems with composite spectral types earlier than K0 and a narrow range of $[\text{Fe}/\text{H}]$ near the solar value. Over 1,100 of the Hipparcos discoveries are within 250 pc, dozens of which are now known to be gravitationally bound with reasonably short periods (a few decades or less) and again consist of mainly main sequence systems (though many more are lower down the main sequence). An H-R diagram of these two samples is shown in  Fig. 13-6. In cases where the metallicity is known, evolutionary studies such as discussed earlier where the components of the system on the H-R diagram can be done with very good speckle photometry, though precise ages will not be possible in cases where neither component is evolved.

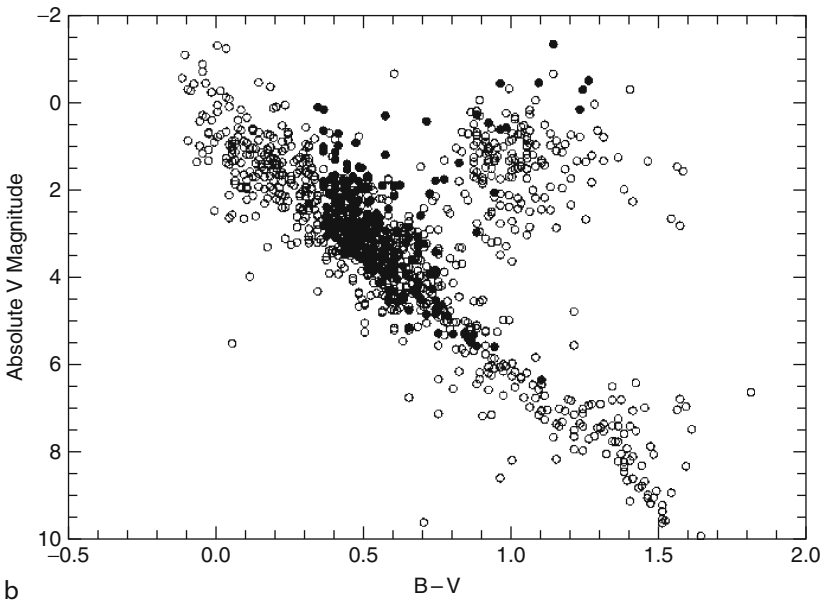
4.2 Pre-main Sequence Binaries

During the 1990s, many binaries were discovered in nearby star-forming regions, with both high-resolution imaging and spectroscopic methods, and statistical studies of these systems were completed (Ghez et al. 1993; Leinert et al. 1993; Mathieu 1989). The high-resolution imaging studies mentioned were completed using speckle interferometry with infrared arrays, which had just become available in sufficiently large formats for such work. Taken together, these studies indicate that the frequency of pre-main sequence binaries is at least as high, and perhaps significantly higher, than the field population of binaries. When sufficient photometric information exists to place the components of pre-main sequence binaries discovered on the H-R diagram, it is found that components of these systems populate the entire region of the diagram where single pre-main sequence stars are located, and this is read as an indication that the formation of the system as a binary occurs well before the final descent to the main sequence from higher luminosities.

Several other features of these systems are worthy of note. Pre-main sequence binaries appear similar to the field population in terms of the orbital characteristics. For example, the shortest period systems generally have only low eccentricities, whereas longer-period systems span the range of eccentricity from 0 to near 1. These systems also appear to follow the same log-normal distribution of periods as the field population, with a comparable median period. (Duquennoy and Mayor (1991) state 180 years as the median period for solar-type field binaries.) For large separation systems, there is a wide range of secondary masses, with a distribution that appears similar to the initial mass function. Finally, multiple systems are found with approximately the same frequency (or greater) than the field, and the hierarchical nature of most of these systems is also already established.



a



b

■ Fig. 13-6

(a) The H-R diagram of visual binaries with uncertainties in semimajor axis of less than 1% from the USNO 6th Orbit Catalog. (b) The H-R diagram of binary stars discovered by the Hipparcos satellite that are within 250 pc of the Sun. This sample shows a range of evolutionary states and metallicity. In both plots, *filled circles* indicate stars with a known value for $[\text{Fe}/\text{H}]$ in the literature

Pre-main sequence systems however also have some significant differences with the field population. The most important example of this is that a large fraction of pre-main sequence systems appear to have either circumstellar or circumbinary disks, as judged from 2 to 10 μm flux excess, and there is evidence for mass accretion in some pre-main sequence systems from X-ray observations. If the pre-main sequence frequency of binaries and multiple stars is higher than the field, then this indicates that there is a mechanism for loss of companions, probably through dynamical encounters with other stars. However, such encounters would likely change the average of the period distribution, hardening the remaining systems, which is not clearly observed (Mathieu 1994).

4.3 Thin- and Thick-Disk Binaries

Allen et al. (2000) have compiled a catalog of 122 common proper motion companions of high-velocity stars, i.e., the systems are presumably wide but physically bound binaries. While the distribution of separations appears to generally follow the so-called Öpik relation at the large separation end in the sense that for wide binaries, interactions with other stars tend to enlarge the separation over time, there appeared to be more small-separation binaries in a sample they judged to belong to a more disk-like population, though the statistics in this region of their plots was low. As the same authors previously noted (Allen et al. 1997), there may be an important transition in binary formation mechanisms at separations of ~ 25 AU. Thus, in the small separation end of their later work, a difference in terms of the disk-like and halo-like populations remains interesting and worthy of further study.

Zinnecker and his collaborators have investigated a possible difference between the average separations of Population I and Population II binaries based on adaptive optics observations (Zinnecker et al. 2004). More recently, Abt (2008) has published an analysis of spectroscopic data, indicating that there are not as many short-period spectroscopic binaries of low metallicity as there are in metal-rich systems. This may be the result of differences in formation (e.g., more turbulent clouds in the early galaxy), while on the other hand the star formation process could be essentially static over the history of the galaxy (consistent with the work of Larson, e.g., 2001, 2002), and any differences between the two populations of binaries would be more likely be due to post-formation environment and age.

Differentiating the thin- and thick-disk samples is important not only for resolving that question but also for establishing the correlation in the MLR in the metallicity range -1.0 to $+0.3$ and mass range 0.08 – 2.0 solar masses. Because of previous difficulties with speckle photometry, the component magnitudes are generally not well known, but metallicities are now in the literature for many speckle binaries with existing orbits. Although in most cases these stars have composite spectral types in the F and G range, there will be many examples where the secondary is in the subsolar mass range, even if the primary is not. Modern speckle cameras can provide the effective temperatures and luminosities that will finally allow the use of these systems in astrophysical studies.

The nature of the “transition region” at 25 AU observed at sub-arcsecond separations by Allen et al. is not clear at present. It could indicate a difference in star formation mechanisms, but it is not known if it is seen for different age populations. However, this separation compares extremely well to the typical separations of the HDSSs, where the sample of interest has apparent separations in the range of 0.1 – 1.0 arcsec and distances of typically 25 – 250 pc. They therefore span a range of 2.5 – 250 AU in their physical separations, and are ideal to follow up on

the previous work regarding binary formation mechanisms. With the next few years, modern instrumentation can be expected to push well through this transition region, down to a few AU at a distance of 250 pc.

Environment may be a sizeable effect in explaining the result of Zinnecker et al. It is well known that wide binaries become wider through stellar encounters. The most detailed study to date is that of Weinberg et al. (1987), where calculations were performed assuming a radially symmetric distribution of perturbers. It is not known in detail how this might be affected in a nonradially symmetric distribution (such as the galactic disk) over time, but it is possible that tidal effects in the galactic potential can become large enough in some cases to dramatically increase binary separations over time. This is not likely to be strong enough in the thick disk to explain a large difference between high- and low-metallicity samples. More work is warranted, but formation would seem to be more likely to be the dominant effect in the distribution of orbital parameters at this point.

Nonetheless, binaries may also eventually provide clues to understanding the structure and mass distribution within the Galactic disk. Knowledge of the mass distribution and the volume density of gravitating matter would yield information about the nature of the dark matter component in the solar neighborhood, which is of principal importance in understanding the structure of the Milky Way on larger scales, e.g., flattening of the Milky Way halo. The separation histogram could give an independent estimate of the dark matter mass density if it can be accurately measured and disentangled in terms of formation and galactic structure effects.

There are two different approaches in determining the mass density of the solar neighborhood. The first approach determines the local volume density or Oort's limit, while the other approach measures the integral surface density of the disk (Kuijken and Gilmore 1989a, b, c). Both types of experiments are based on the kinematic properties together with the determination of the distribution of a "test" population of stars perpendicular to the Galactic plane. Siegel et al. (2002) have already pointed out the importance of determining the effect of binaries in these kinds of studies. The unresolved binaries manifest themselves in two observational effects: (1) They increase the apparent luminosity of a "star" and (2) They affect the apparent color. The binary correction is therefore extremely important in estimating the photometric distances and in determining the correct density law. Usually, the field binary fraction is assumed to be near 50%, as determined in G dwarf spectroscopic binaries discussed in [Sect. 2.3](#), but there is a controversy about the binary fraction for halo and thick-disk stars and whether the binary fraction changes with spectral type (Lada 2006). Clarification of these issues is important in the dynamical studies of the density distribution in the galactic disk.

Dynamical studies of the galactic disk based on stellar parallax measurements usually suffer from the lack of knowledge about binarity. Unresolved binaries affect the measurements of radial velocities of these objects and thus lead to an overestimate of the velocity dispersion of the samples. This effect has not been studied in the literature; one must be aware of the potential for observational biases in such statistical studies.

4.4 Binary Stars in Clusters

Binary stars can be used as probes of the cluster environment in several ways. First, if orbital information can be determined for spectroscopic/visual or eclipsing/spectroscopic binaries, then an independent measure of the distance to the system can be obtained. Binaries have been studied in several open clusters where it is found that they concentrate toward the center of the

cluster. This can be explained by the higher mass of these objects compared with single stars. The orbital parameters of binaries are also thought to be affected by the gravitational potential in clusters. In general, if a system is below a certain separation (~ 30 AU), then the encounters with other stars will lead to a decrease in the separation with time. This is said to be a “hardening” of the binary system. Above the same limit, the separation will increase with time or “soften,” until such time as the system has dispersed.

The basic mechanism for binary star formation is fragmentation with subsequent gravitational interactions in the cluster environment, as discussed earlier. However, the evidence suggests that the formation of stable multiple star systems is the result of dynamical interactions (Aarseth 2004). Mathematical models of binary–binary interactions show that a typical formation process for a triple star system is process where, in the wider binary, one component is ejected and the other component becomes bound to the harder binary. A variation on this process is that two binaries interact after which two single stars are ejected leaving a binary with altered orbital parameters. In both cases, the original hard binary will have its orbital parameters affected, and velocities of the final products of the interaction can be quite high. In some cases, the hard binary or triple can leave the cluster environment in which it is formed. Sterzik and Tokovinin (2002) have compared observed triple star statistics with the results of model calculations and found reasonable agreement in physical parameters of such systems. In particular, a weak co-alignment of orbital planes was predicted from the cluster simulations, which is in fact observed in the field population of triple stars.

In globular clusters, it appears from recent research that rich globular clusters retain a much larger fraction of neutron stars than one would expect from dynamical calculations (Pfahl et al. 2002a). When they form, neutron stars are thought to have very high speeds, in excess of 200 km/s, which is far in excess of the escape speed of a typical globular cluster. Such high speeds are known as neutron star kicks. Such kicks are observed in the field population of neutron stars in the form of high proper motions. A possible explanation of the high retention rate is that neutron stars with a massive companion are preferentially retained. In this case, the companion is large enough to survive the formation of the neutron star and can in theory reduce the kick enough so that the system is retained by the cluster.

5 Interacting Binaries

5.1 General Comments on Interacting Binaries

Interacting binaries are defined by the proximity of the two stars. If they are close enough to affect the physical properties of the two stars, then they are considered interacting. As the separation of a particular pair of stars is decreased, there will be a point at which the surfaces of the two stars start to deform. The tidal bulges produced will first start to dissipate the rotational rates of the stars, and eventually they will rotate with the same period as that of the orbit of the system. If the eccentricity of the orbit is high enough, then the tidal forces will also be periodic, leading to pulsation of the stars. These in turn can dissipate orbital energy from the system, but without significant loss of angular momentum (Swihart 1992). It can easily be shown that at constant angular momentum, the lowest energy configuration is that of a circular orbit. Thus,

it is expected (and the observations also confirm) that close binaries have orbital eccentricity zero and rotate synchronously with the orbit.

If one considers a small amount of material near the surface of the more massive star in a binary system, then it will not be found far from the star despite any orbital influence on it by the second star in the system, provided that the velocity of this material is small. In general, there will be a surface enclosing the massive star inside of which the material must be contained. This is the so-called zero-velocity surface: in order to reach a point on this surface, the material must lose all of its velocity. A similar surface would exist surrounding the second star, and the average distance from the star to the zero-velocity surface will of course depend on the velocity of the material. If the velocity is high enough and/or the stars are close enough, the two zero-velocity surfaces begin to overlap, meaning that there is a pathway where material can pass from within the zero-velocity surface of one star into the zero-velocity surface of the other, even though it is still bound to the binary system as a whole. If the two surfaces just touch each other in a single point, this point is the Lagrangian point L_1 , the point at which the material would feel equal and opposite gravitational pull from both stars. The zero-velocity surfaces of the two stars in this special case are known as the Roche lobes of each star.

For larger separations, the Roche lobes are nearly spherical, but when the separation becomes small, then the surfaces become distorted. By considering the balance of forces in the rotating frame of the system, it can be shown that the surface of the star itself must take on the same shape as the Roche lobe, albeit smaller in size in general. If the stars are on the main sequence and do not have substantial stellar winds, then there is no significant transfer of material in the system, although the evolution can be affected by the distortion in shape. However, once the more massive star leaves the main sequence, it expands significantly on its way to becoming a red giant. As the star fills its Roche lobe, material near the Lagrangian L_1 will pass into the Roche lobe of the secondary star, and subsequently be attracted to the secondary star over time. Since the rate of expansion of the primary star is slow compared to the free-fall timescale of the material, once it is within the Roche lobe of the secondary it is not expected that the primary star would overfill its Roche lobe significantly. There is also a natural equilibrium that sets in because the more the star expands beyond its Roche lobe, the greater the overlap of the two lobes, and the faster the rate of mass infall to the secondary star.

Later on in the evolution of the system, the evolved star may become a white dwarf or neutron star, while the star that was on the main sequence will evolve to the giant branch and may fill its Roche lobe, thereby donating material back to the other star. If this is the configuration, then there is a key difference from the case considered just above: the star receiving the matter is now a very compact object so that much of the incoming material will not immediately fall onto the surface, but rather will begin to circle the compact component. Over time, this will create a disk. As new material drops onto the disk, it creates a hotspot through the collisions, which is over time transferred throughout the disk. Therefore, the disk can have a very high temperature and luminosity, contributing significantly to the energy output of the binary system.

If the compact companion is a neutron star instead of a white dwarf, then the higher mass means that the infalling material will be traveling with much higher velocities when it hits the accretion disk. In addition, many neutron stars have strong magnetic fields. The result is that charged particles will be funneled onto the magnetic poles of the star at high energy, creating X-rays. If the spin axis and magnetic axis have different orientations, then what will be observed is a pulsed X-ray signature as the neutron star rotates.

5.2 Important Types of Interacting Binaries

5.2.1 Type Ia Supernovae

The main characteristic that places a supernova in the Type I category is the lack of hydrogen lines in the spectrum of the object. Other spectral features change with time, but the most prominent features include absorption lines of singly-ionized metals that are heavily Doppler broadened and blueshifted. For example, the CaII doublet near 393.4–396.8 nm can be so broadened that it appears as a single feature that is blueshifted by 20.0 nm. This is evidence that the material is expanding out from the explosion and is traveling at 10,000–15,000 km/s.

In general, the light curve and spectral characteristics of Type Ia supernovae are quite uniform. The peak light is reached 10–20 days after the initial event, and after peak light, the curve decays rapidly for the first 100 days, then more gradually. Peak luminosity of these events is on the order of 10^{43} ergs, which corresponds to an absolute visual magnitude of approximately –18.2. The total energy release over time is estimated to be 10^{51} ergs. The rate of occurrence of Type Ia supernovae in spiral galaxies is approximately 1 per 36 years, and is thought to be less concentrated in the spiral arms of such galaxies than Type II supernovae.

The standard model of Type Ia supernovae is the carbon deflagration of a white dwarf. Although there are several variations along this line, the basic picture is that of a white dwarf star that exists in a semidetached binary system. The white dwarf has a degenerate carbon-oxygen core and accretes matter from its companion. It grows in mass to the Chandrasekhar limit, and carbon burning is initiated in the core. Since the mass represented by the Chandrasekhar limit is well-known, approximately 1.3 solar masses, the energy released in the explosion, should be essentially the same for every such event. They may therefore be thought of as “standard candles” and used as distance indicators to external galaxies.

White dwarfs in binary systems also can explain cataclysmic variables, that is, stars with irregular, large increases in brightness that occur without regular pattern in time. As more hydrogen accumulates near the surface of the white dwarf, temperatures and densities may be high enough to briefly initiate hydrogen fusion in episodic fashion, leading to a dramatic increase in light from the system.

5.2.2 X-Ray Binaries

If the companion star in a semidetached binary system is a neutron star or black hole, then the system forms an X-ray binary. Observationally, these systems will appear as unresolved stars or single-lined spectroscopic binaries and usually are described by the mass of the system: high-, intermediate-, and low-mass X-ray binaries (HMXB, IMXB, and LMXB, respectively). In all three cases, the X-rays are produced as matter from the luminous companion falls onto the disk of the compact companion, but the evolutionary phase of the companion star determines the mechanism of the infall. When the companion is on or near the main sequence, it will not fill its Roche lobe, and material from the stellar surface cannot be directly accreted onto the compact companion. Instead, it is provided by a stellar wind. This is a relatively long-lived and low-mass accretion, where the system is observed as a low-luminosity, hard (1–10 keV) X-ray source (Pfahl et al. 2002b).

Later on, in the evolution of the companion star, it will become a red giant, and eventually fill its Roche lobe. This initiates mass transfer through the Lagrange point L_1 , which is a much faster

and short-lived process. In this case, an accretion disk will form around the compact star, and the X-ray spectrum will be more luminous but softer. Since high-mass stars evolve much more quickly than low mass stars, it is thought that many sources of this type are LMXBs. In contrast, not only are evolved HMXBs short-lived, but a HMXB where the companion is a giant can accrete matter at a rate that generates a luminosity close to the Eddington limit, and so material can actually be blown away from the binary system in this way, disrupting the system over time. In general, the details of the evolution can be complicated by the mass transfer so that current theories predict a wide range of X-ray spectra, depending on the mass of the companion star.

5.2.3 Blue Stragglers

In many star clusters, the H-R diagram, shows a well-defined population of giants and other evolved stars that joins the main sequence at the point known as the turnoff. The pattern is explained very easily since the cluster formed at one time and all the stars in it have the same age, regardless of mass. Therefore, the cluster H-R diagram follows an isochrone. The turnoff point effectively determines the age of the cluster. In the observed H-R diagrams of dense clusters, particularly globular clusters, there are stars that appear to be on the main sequence but above the turnoff point. These are known as blue stragglers. In some cases, the masses of these stars are thought to be nearly twice the turnoff mass, and there are even examples where the object is apparently above twice the turnoff mass (Shara et al. 1997).

Since these stars are bluer and hotter than other stars on the main sequence, they should have already evolved off of the main sequence if their evolution proceeded in the normal way. The idea that these stars are formed from merger events between stars has become the leading theory to explain this phenomenon. These mergers may be formed from direct stellar collisions or from orbital decay of a close binary system. The rate of the former has been calculated to be low except in the cores of the densest clusters, but the latter may happen when a small-separation binary is involved with an encounter with a single perturbing star or as the result of multiple encounters in the cluster environment. Factors such as the density of stars, velocity dispersion, mass function of the cluster, and the binary fraction all affect the details of the theoretical studies that have been done to date (Lombardi et al. 2002).

References

-
- | | |
|---|--|
| <p>Aarseth, S. J. 2004, Formation and evolution of hierarchical systems, <i>Revista Mexicana de Astronomía y Astrofísica (Serie de Conferencias)</i>, 21, 156–162</p> <p>Abt, H. A. 2008, The difference between metal-poor and metal-rich binaries, <i>AJ</i>, 135, 722–725</p> <p>Abt, H. A., & Levy, S. G. 1976, Multiplicity among solar-type stars, <i>Astrophys J Suppl Series</i>, 30, 273–306</p> <p>Abt, H. A., & Willmarth, D. W. 1999, Binaries in the Praesepe and Coma Star Clusters and their implications for binary evolution, <i>ApJ</i>, 521, 682–690</p> | <p>Aitken, R. G. 1935, <i>The Binary Stars</i> (New York: McGraw-Hill)</p> <p>Allen, C., Poveda, A., & Herrera, M. A. 1997, The distribution of separations of wide binaries, in <i>Visual Double Stars: Formation, Dynamics, and Evolutionary Tracks</i>, eds. J. A. Docobo et al. (Dordrecht: Kluwer), 133–143</p> <p>Allen, C., Poveda, A., & Herrera, M. A. 2000, Wide binaries among high-velocity and metal poor stars, <i>A&A</i>, 356, 529–540</p> <p>Alzner, A. 2004, The orbital elements of a visual binary star, in <i>Observing and Measuring Double Stars</i>, ed. R. Argyle (Berlin, Heidelberg, New York: Springer), 53–62</p> |
|---|--|

- Andersen, J. 1991, Accurate masses and radii of normal stars, *Astron Astrophys Rev*, 3, 91–126
- Argyle, R. A. 2004, *Observing and Measuring Double Stars* (Berlin, Heidelberg, New York: Springer)
- Bagnuolo, W. G. 1988, Binary-intensity ratios by the fork algorithm, *Opt Lett*, 13, 907–909
- Balega, I. I., Balega, Y. Y., Hofmann, K.-H., Maksimov, A. F., Pluzhnik, E. A., Schertl, D., Shkhagosheva, Z. U., & Weigelt, G. 2002, Speckle interferometry of nearby multiple stars, *A&A*, 385, 87–93
- Balega, I. I., Balega, Y. Y., Hofmann, K.-H., Pluzhnik, E. A., Schertl, D., Shkhagosheva, Z. U., & Weigelt, G. 2005, Orbits of new Hipparcos binaries. I, *A&A*, 433, 591–596
- Balega, I. I., Balega, Y. Y., Hofmann, K.-H., Malogolovets, E. V., Shkhagosheva, Z. U., & Weigelt, G. 2006, Orbits of new Hipparcos binaries. II, *A&A*, 448, 703–707
- Balega, Y. Y., Beuzit, J.-L., Delfosse, X., Forveille, T., Perrier, C., Mayor, M., Ségransan, D., Udry, S., Tokovinin, A. A., Schertl, D., Weigelt, G., Balega, I. I., & Malogolovets, E. V. 2007, Accurate masses of low mass stars GJ 765.2AB ($0.83 M_{\odot} + 0.76 M_{\odot}$), *A&A*, 464, 635–640
- Bate, M. R., Bonnell, I. A., & Bromm, V. 2003, The formation of a star cluster: predicting the properties of stars and brown dwarfs, *MNRAS*, 339, 577–599
- Bates, R. H. T., & Cady, F. M. 1980, Towards true imaging by wideband speckle interferometry, *Opt Commun*, 32, 365–369
- Berger, D. H., Gies, D. R., McAlister, H. A., ten Brummelaar, T. A., Henry, T. J., Sturmann, J., Sturmann, L., Turner, N. H., Ridgway, S. T., & Aufdenberg, J. P. 2006, First results from the CHARA Array. IV. The interferometric radii of low-mass stars, *ApJ*, 644, 475–483
- Bonnell, I. A. 1994, A new binary formation mechanism, *MNRAS*, 269, 837–848
- Cayrel de Strobel, G., Soubiran, C., & Ralite, N. 2001, A catalogue of [Fe/H] determinations of F, G, K stars: 2001 Edition, *A&A*, 373, 159–163
- Clarke, C. J. 1999, Fragmentation of cold slabs: application to the formation of clusters, *MNRAS*, 307, 328–336
- Clarke, C. J., & Pringle, J. E. 1993, Accretion disc response to a stellar fly-by, *MNRAS*, 261, 190–202
- Cvetkovi, Z. 2008, First orbits for five binaries, *AJ*, 136, 1746–1752
- Davidson, J. W., Jr., Baptista, B. J., Horch, E. P., Franz, O. G., & van Altena, W. F. 2009, A photometric analysis of seventeen binary stars using speckle imaging, *AJ*, 138, 1354–1364
- Docobo, J. A. 1985, On the analytic calculation of visual double star orbits, *Celest Mech*, 36, 143–153
- Duchêne, G. 1999, Binary fraction in low-mass star forming regions: a reexamination of the possible excess and implications, *A&A*, 341, 547–552
- Duquennoy, A., & Mayor, M. 1991, Multiplicity among solar-type stars in the solar neighborhood, *A&A*, 248, 485–524
- ESA, 1997, The Hipparcos and Tycho catalogues, European Space Agency SP 1200
- Franz, O. G., Wasserman, L. H., Nelan, E., Lattanzi, M. G., Bucciarelli, B., & Taff, L. G. 1992, Binary star observations with the Hubble Space Telescope fine guidance sensors. II – Bright Hyades, *AJ*, 103, 190–196
- Ghez, A. M., Neugebauer, G., & Matthews, K. 1993, The multiplicity of T Tauri stars in the star forming regions Taurus-Auriga and Ophiuchus-Scorpius: a 2.2 micron speckle imaging survey, *AJ*, 106, 2005–2023
- Ghez, A. M., White, R. J., & Simon, M. 1997, High spatial resolution imaging of pre-mainsequence binary stars: resolving the relationship between disks and close companions, *ApJ*, 490, 353–367
- Griffin, R. F. 1970, Photoelectric radial velocities of 87 seventh-magnitude K stars previously observed by Redman, *MNRAS*, 148, 211–225
- Hanbury Brown, R., Davis, J., & Allen, L. R. 1967, The stellar interferometer at Narrabri Observatory I. A. Description of the instrument and the observational procedure, *MNRAS*, 137, 375–392
- Hanbury Brown, R., Davis, J., Herbison-Evans, D., & Allen, L. R. 1970, A study of γ^2 Velorum with a stellar intensity interferometer, *MNRAS*, 148, 103–117
- Hartkopf, W. I. 1998, Private communication
- Hartkopf, W. I., Mason, B. D., & McAlister, H. A. 1996, Binary star orbits from speckle interferometry. VIII. Orbits of 37 close visual systems, *AJ*, 111, 370–392. doi:10.1086/117790
- Henry, T. J., & McCarthy, D. W., Jr. 1990, A systematic search for brown dwarfs orbiting nearby stars, *ApJ*, 350, 334–347
- Henry, T. J., & McCarthy, D. W., Jr. 1993, The mass-luminosity relation for stars of mass 1.0 to $0.08 M_{\odot}$, *AJ*, 106, 773–789
- Henry, T. J., Franz, O. G., Wasserman, L. H., Benedict, G. F., Shelus, P. J., Ianna, P. A., Kirkpatrick, J. D., & McCarthy, D. W., Jr. 1999, The optical mass-luminosity relation at the end of the main sequence ($0.08\text{--}0.20 M_{\odot}$), *AJ*, 512, 864–873
- Herschel, W. 1803, Account of the changes that have happened, during the last twenty-five years, in

- the relative situation of double-stars; with an investigation of the cause to which they are owing, *Philos Trans R Soc Lond*, 93, 339–382
- Horch, E. P., Falta, D., Anderson, L. M., DeSousa, M. D., Minitier, C. M., Ahmed, T., & van Altena, W. F. 2010, CCD speckle observations of binary stars with the WIYN Telescope. VI. Measures during 2007–2008, *AJ*, 139, 205–215
- Kuijken, K., & Gilmore, G. 1989a, The mass distribution in the galactic disc. I. Technique to determine the surface mass density of the disc near the sun, *MNRAS*, 239, 571–603
- Kuijken, K., & Gilmore, G. 1989b, The mass distribution in the galactic disc. II. Determination of the surface mass density of the galactic disk near the sun, *MNRAS*, 239, 605–649
- Kuijken, K., & Gilmore, G. 1989c, The mass distribution in the galactic disc. III. The local volume mass density, *MNRAS*, 239, 651–664
- Lada, C. J. 2006, Stellar multiplicity and the initial mass function: most stars are single, *ApJ*, 640, L63–L66
- Larson, R. B. 2001, Implications of Binary Properties for Theories of Star Formation, in *The Formation of Binary Stars*, Proc. IAU Symp. 200, eds. H. Zinnecker, & R. D. Mathieu (Dordrecht: Reidel), 93
- Larson, R. B. 2002, The role of tidal interactions in star formation, *MNRAS*, 332, 155–164
- Leinert, Ch., Zinnecker, H., Weitzel, N., Christou, J., Ridgway, S. T., Jameson, R., & Haas, M. 1993, A systematic search for young binaries in Taurus, *A&A*, 278, 129–149
- Luyten, W. J. 1927, Sixteen double stars with large proper motion, *Harv Coll Observatory Bull*, 852, 14–15
- Luyten, W. J. 1997, LDS catalogue: doubles with common proper motion (Luyten 1940–87) (Minnesota III: Publ. Astr. Obs. Univ.), part 3, 35
- Lohmann, A., Weigelt, G., & Wirtitzer, B. 1983, Speckle masking in astronomy – triple correlation theory and applications, *Appl Opt*, 22, 4028–4037
- Lombardi, J. C., Warren, J. S., Rasio, F. A., Sills, A., & Warren, A. R. 2002, Stellar collisions and the interior structure of blue stragglers, *ApJ*, 568, 939–953.
- Martin, C., Mignard, F., Hartkopf, W. I., & McAlister, H. A. 1998, Mass determinations of astrometric binaries with Hipparcos. III. New results for 28 systems, *A&AS*, 133, 149–162
- McDonald, J. M., & Clarke, C. J. 1993, Dynamical biasing in binary star formation – implications for brown dwarfs in binaries, *MNRAS*, 262, 800–804
- Mason, B. D., McAlister, H. A., Hartkopf, W. I., Griffen, R. F., & Griffen, R. E. M. 1997, Binary star orbits from speckle interferometry. X. Speckle-spectroscopic orbits of HR 233, 36 Tau and 73 Leo, *AJ*, 114, 1607–1622
- Mason, B. D., Henry, T. J., Hartkopf, W. I., ten Brummelaar, T., & Soderblom, D. R. 1998, A multiplicity survey of chromospherically active and inactive stars, *AJ*, 116, 2975–2983
- Mathieu, R. D. 1989, Spectroscopic binaries among low-mass pre-main sequence stars, *Highlights Astron*, 8, 111
- Mathieu, R. D. 1994, Pre-main sequence binary stars, *ARAA*, 32, 465–530
- Mathieu, R. D. 2000, The WIYN open cluster study, in *Stars Clusters and Associations*, ASP Conference Series, Vol. 198, eds. R. Pallavicini et al. (San Francisco: Astronomical Society of the Pacific), 517–532
- Mayor, M., Imbert, M., Andersen, J., Ardeberg, A., Baranne, A., Benz, W., Ischi, E., Lindgren, H., Martin, N., Maurice, E., Nordstrom, B., & Prevot, L. 1983, Radial velocities of southern stars obtained with the photoelectric scanner CORAVEL. I – observations of 169 giant stars in 47 Tucanae, *A&AS*, 54, 495–503
- Mazeh, T., & Zucker, S. 1994, TODCOR: a two-dimensional correlation technique to analyze stellar spectra in search of faint companions, *Astrophys Space Sci*, 212, 349–356
- Mazeh, T., Prato, L., Simon, M., Goldberg, E., Norman, D., & Zucker, Z. 2002, Infrared detection of low-mass secondaries in spectroscopic binaries, *ApJ*, 564, 1007–1014
- Michell, J. 1767, An inquiry into the probable parallax, and magnitude of the fixed stars, from the quantity of light which they afford us, and the particular circumstances of their situation, *Philos Trans* (1683–1775) 57, 234–264
- Muterspaugh, M. W., Lane, B. F., Fekel, F. C., Konacki, M., Burke, B. F., Kulkarni, S. R., Colavita, M. M., Shao, M., & Wiktorowicz, S. J. 2008, Masses, luminosities, and orbital coplanarities of the μ Orionis quadruple-star system from phases differential astrometry, *AJ*, 135, 766–776. doi: 10.1088/0004-6256/135/3/766
- Nordström, B., Mayor, M., Andersen, J., Holmberg, J., Pont, F., Jorgensen, B. R., Olsen, E. H., Udry, S., & Mowlavi, N. 2004, The Geneva-Copenhagen survey of the Solar neighbourhood. Ages, metallicities, and kinematic properties of 14 000 F and G dwarfs, *A&A*, 418, 989–1019
- Öpik, E. 1927, Measures of Double Stars (1924–1926), *Publications de L'Observatoire Astronomique de l'Universite de Tartu*, 26, 1–2

- Patience, J., Ghez, A. M., Reid, I. N., Weinberger, A. J., & Matthews, K. 1998, The multiplicity of the Hyades and its implications for binary star formation and evolution, *AJ*, 115, 1972–1988
- Pfahl, E., Rappaport, S., & Podsiadlowski, P. 2002a, A comprehensive study of neutron star retention in globular clusters, *ApJ*, 573, 283–305
- Pfahl, E., Rappaport, S., & Podsiadlowski, P. 2002b, On the population of wind-accreting neutron stars in the galaxy, *ApJ*, 571, L37–L40
- Popper, D. 1980, Stellar masses, *ARAA*, 18, 115–164
- Pourbaix, D. 2002, Simultaneous least-squares adjustment of visual and spectroscopic observations of binary stars, *A&AS*, 131, 377–382. doi:10.1051/aas:1998275
- Prato, L., Simon, M., Mazeh, T., McLean, I. S., Norman, D., & Zucker, S. 2002, The smallest mass ratio young star spectroscopic binaries, *ApJ*, 569, 863–871
- Prugniel, P., & Soubiran, C. 2001, A database of high and medium-resolution stellar spectra, *A&A*, 369, 1048–1057
- Richichi, A., & Percheron, P. 2002, CHARM: a catalog of high angular resolution measurements. *A&A*, 386, 492–503. doi: 10.1051/0004-6361:20020236
- Shara, M. M., Saffer, R. A., & Livio, M. 1997, The first direct measurement of the mass of a blue straggler in the core of a globular cluster: BSS 19 in 47 Tucanae, *ApJ*, 489, L59–L62
- Siegel, M. H., Majewski, S. R., Reid, I. N., & Thompson, I. B. 2002, Star Counts Dedivivus. IV. Density laws through photometric parallaxes, *ApJ*, 578, 151–175
- Söderhjelm, S. 1999, Visual binary orbits and masses post Hipparcos, *A&A*, 341, 121–140
- Steinmetz, M., Zwitter, T., Siebert, A., Watson, F. G., Freeman, K. C., Munari, U., Campbell, R., Williams, M., Seabroke, G. M., Wyse, R. F. G., Parker, Q. A., Bienaymé, O., Roeser, S., Gibson, B. K., Gilmore, G., Grebel, E. K., Helmi, A., Navarro, J. F., Burton, D., Cass, C. J. P., Dawe, J. A., Fiegert, K., Hartley, M., Russell, K. S., Saunders, W., Enke, H., Bailin, J., Binney, J., Bland-Hawthorn, J., Boeche, C., Dehnen, W., Eisenstein, D. J., Evans, N. W., Fiorucci, M., Fulbright, J. P., Gerhard, O., Jauregi, U., Kelz, A., Mijovi, L., Minchev, I., Parmentier, G., Peñarrubia, J., Quillen, A. C., Read, M. A., Ruchti, G., Scholz, R.-D., Siviero, A., Smith, M. C., Sordo, R., Veltz, L., Vidrih, S., von Berlepsch, R., Boyle, B. J., & Schilbach, E. 2006, The Radial Velocity Experiment (RAVE): first data release, *AJ*, 132, 1645–1668
- Sterzik, M. F., & Tokovinin, A. A. 2002, Relative orientation of orbits in triple stars, *A&A*, 384, 1030–1037
- Swihart, T. L. 1992, *Quantitative Astronomy* (Upper Saddle River: Prentice Hall)
- ten Brummelaar, T., Mason, B. D., McAlister, H. A., Roberts, L. C., Jr., Turner, N. H., Hartkopf, W. I., & Bagnuolo, W. G., Jr. 2000, Binary star differential photometry using the adaptive optics system at Mount Wilson Observatory, *AJ*, 119, 2403–2414
- Tohline, J. E. 2002, The origin of binary stars, *ARAA*, 40, 349–385. doi:10.1146/annurev.astro.40.060401.093810
- Torres, G., Claret, A., & Young, P. A. 2009, Binary orbit, physical properties, and evolutionary state of Capella (α Aurigae), *ApJ*, 700, 1349–1381
- Torres, G., Andersen, J., & Giménez, A. 2010, Accurate masses and radii of normal stars: modern results and applications, *Astron Astrophys Rev*. doi:10.1007/s00159-009-0025-1
- Udry, S., Mayor, M., Latham, D. W., Stefanik, R. P., Torres, G., Mazeh, T., Goldberg, D., Andersen, J., & Nordstrom, B. 1998, A Survey for Spectroscopic Binaries in a Large Sample of G Dwarfs, in *ASP Conf. Ser. 154, The Tenth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, eds. R. A. Donahue, & J. A. Bookbinder (San Francisco: ASP), 2148–2155
- Weinberg, M. D., Shapiro, S. L., & Wasserman, I. 1987, The dynamical fate of wide binaries in the solar neighborhood, *ApJ*, 312, 367–389
- Zinnecker, H., Köhler, R., & Jahreiß, H. 2004, Binary statistics among population II stars, *Revista Mexicana de Astronomía y Astrofísica (Serie de Conferencias)*, 21, 33–36
- Zeilik, M., & Gregory, S. A. 1998, *Introductory Astronomy and Astrophysics* (New York: Saunders College Publishing)

14 **Supernovae and Gamma-Ray Bursts**

Philipp Podsiadlowski

Sub-department of Astrophysics, University of Oxford, Oxford, UK

1	<i>Introduction</i>	695
2	<i>Major Explosion Mechanisms</i>	696
2.1	Core-Collapse Supernovae	696
2.2	Thermonuclear Explosions	698
3	<i>Supernova Classification</i>	699
3.1	Main Classification Scheme	700
3.2	Complications	700
3.3	The Diversity of Core-Collapse Supernova Lightcurves	702
3.3.1	Understanding SN II-P Lightcurves	703
4	<i>Supernova 1987A in the Large Magellanic Cloud</i>	705
4.1	The Mystery of the Progenitor Star	705
4.2	A Binary Merger Model for the Progenitor of SN 1987A	707
5	<i>Type Ia Supernovae</i>	708
5.1	Type Ia Supernovae as Cosmological Distance Candles	708
5.2	The Progenitors of SNe Ia	709
5.2.1	The Single-Degenerate Model	709
5.2.2	The Double-Degenerate Model	710
6	<i>Gamma-Ray Bursts, Collapsars, and Hypernovae</i>	710
6.1	The History of Gamma-Ray Bursts	710
6.2	The Main Properties of GRBs	711
6.2.1	Relativistic Beaming of GRBs	712
6.3	The Relativistic Fireball Model	713
6.3.1	The Compactness and the Baryon Loading Problem	713
6.3.2	The Role of Shocks	714
6.4	The Collapsar Model and the Central Engine	714
6.4.1	The GRB Energy Source	715
6.4.2	The Production of a Relativistic Jet	716
6.5	Hypernovae: The LGRB – Supernova Connection	716
6.6	The Progenitors of GRBs	717
6.6.1	Single-Star Progenitor Models for LGRBs	717
6.6.2	Binary Progenitor Models	717

7 *The Diversity of Supernova Explosions* 718

7.1 The Role of Binarity 718

7.1.1 Types of Binary Interactions 718

7.1.2 Black Hole or Neutron Star? 721

7.1.3 Electron-Capture Supernovae in Close Binaries 722

7.2 The Diversity of Supernova Explosions 723

7.2.1 Neutron Stars 723

7.2.2 Black Holes 725

7.2.3 Thermonuclear Explosions 726

7.3 Rotation and Magnetic Fields 727

7.4 Metallicity 727

7.5 Dynamical Interactions 728

8 *Current Topics* 729

8.1 The Diversity of SN Ia Progenitors 729

8.2 Superluminous Supernovae 729

8.3 Supernovae with a Circumstellar Medium 730

8.4 Supernovae in the Middle of Nowhere 730

8.5 Faint and Failed Supernovae 731

8.6 Supernova Kicks 731

References 732

Abstract: Supernovae are explosions of stars which are triggered either by the implosion of the core of a star or a thermonuclear runaway, causing a bright optical display lasting for weeks to years. This chapter first explains the main explosion types, how they are classified, and the principles that determine their lightcurves. It then discusses in more detail some of the most important supernova types, specifically SN 1987A, the last naked-eye supernova near our own Galaxy, Type Ia supernovae that have been used as standardizable cosmological distance candles, and gamma-ray bursts and their related supernovae. Special emphasis is given to the link of the various supernova types to their progenitor systems and a discussion of any outstanding issues. Causes for the large diversity of supernova types and subtypes are then systematically explored: these include binarity, the explosion mechanisms, rotation, metallicity, and dynamical effects. Finally, some of the major topics of current interest are briefly discussed.

1 Introduction

Broadly speaking, a *supernova* is the explosion of a star or stellar system. It derives its name from a combination of the words *nova*, referring to a “new” star suddenly appearing in the sky, and *super* denoting a particularly bright “new” star. Novae and supernovae have been observed and described as important cosmic events for at least a few thousand years; the first recorded by the Chinese was SN 189 which occurred in 189 AD. The fact that novae and supernovae represent different physical phenomena was realized only in the 1930s by Baade and Zwicky (1934).¹ That these might involve the final collapse of the core of a massive star to a neutron star was proposed soon thereafter by Gamow and Schoenberg (1941), but it took until 1987 and the occurrence of supernova 1987A (SN 1987A) that this was finally confirmed observationally for at least one of the main supernova types.² We now know that there are at least two major mechanisms producing a supernova, although in recent years with the discoveries of hundreds of new supernovae, it has become clear that there is an enormous diversity of supernova types and subtypes; there are almost certainly more than just two explosion mechanisms. In this chapter, we will first discuss the two major explosion mechanisms, the classical classification scheme and how a lot of the observed diversity can be understood relatively simply by variations of the envelope properties of the exploding stars. We will then discuss SN 1987A in some detail, the last naked-eye supernova which has been one of the major astronomical events of the 1980s and which has provided a major impetus in the supernova field, and the class of Type Ia supernovae, which 10 years later provided the first evidence for an accelerating universe. Gamma-ray bursts and related hypernovae, some of the largest explosions in the universe, are the topic of the next section. This is followed by an overview over the various physical reasons that cause the observed diversity, and the chapter ends with a number of selected topics of particular current interest.

¹Unlike supernovae that generally involve the whole star in the explosion, novae are now understood to be thermonuclear explosions in the envelopes of white dwarfs.

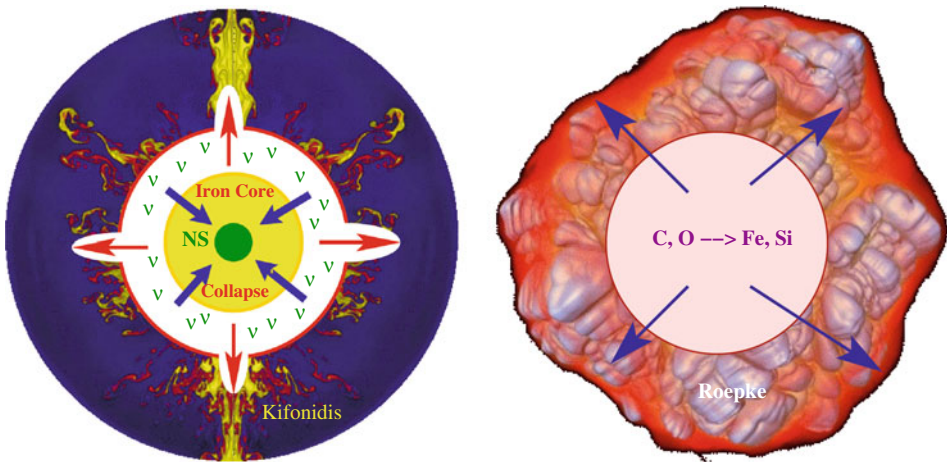
²Traditionally, supernovae are named after the year and the order in the year in which the supernova was reported; therefore, SN 1987A was the first supernova that was reported in 1987. Today, with the discovery of hundreds of supernovae per year, not all supernovae are named based on this convention.

2 Major Explosion Mechanisms

For at least a few decades, it has been realized that there are (at least) two main supernova explosion mechanisms: *core-collapse supernovae* involving the final phase in the evolution of a massive star and *thermonuclear explosions*, most likely related to white dwarfs approaching the Chandrasekhar limit (● Fig. 14-1).³

2.1 Core-Collapse Supernovae

The evolution of stars and, in particular, massive stars is characterized by an alternation of nuclear burning phases and contraction phases. For example, the Sun is presently burning hydrogen in its core at a temperature of $\sim 10^7$ K. After it has consumed all of its hydrogen in the core, the core will start to contract and heat up until ultimately helium will ignite (at a characteristic temperature of $\sim 10^8$ K). After helium burning, this cycle terminates for the Sun, as it develops a degenerate carbon/oxygen (CO) core and ends its evolution as a CO white dwarf. For significantly more massive stars ($M \gtrsim 11 M_\odot$), the alternation of contraction and burning phases continues until ultimately the star has developed an iron core, surrounded by an onion-like structure consisting of shells of increasingly lower mean-atomic mass. Since iron is the most stable nucleus (i.e., has the highest nuclear binding energy per baryon), no more energy can be generated by fusing iron with other nuclei. Therefore, if the core exceeds the Chandrasekhar mass for iron, there is no longer a cold hydrostatic equilibrium configuration, and the core has to contract/collapse as it cools and loses its thermal-pressure support. While this contraction

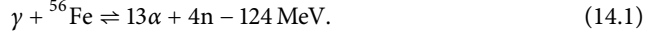


■ Fig. 14-1

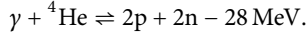
Schematic figures contrasting a core-collapse supernova in the core of a massive star (left) to a thermonuclear explosion in a CO white dwarf (right)

³The *Chandrasekhar limit* defines the maximum mass at which a zero-temperature, self-gravitating object can be supported by electron degeneracy pressure. For typical white dwarf compositions, this mass is close to $1.4 M_\odot$.

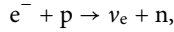
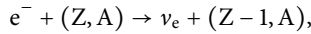
may start slowly, it soon accelerate because of a number of instabilities. At the end of silicon burning, the final nuclear burning phase producing the iron core, the central temperature and density of the core, T_c and ρ_c , are $\sim 8 \times 10^9$ K and $\sim 4 \times 10^{12}$ kg m $^{-3}$, respectively. Once the core has contracted enough and reaches a temperature $T_c \sim 10^{11}$ K, the iron and other heavy nuclei start to be photodissociated into α particles (^4He) and neutrons:



This photodissociation involves *endothermic reactions* that require energy (124 MeV); as a consequence, the temperature increases less rapidly than pressure, accelerating the contraction. At a slightly higher temperature ($T_c \sim 2 \times 10^{11}$ K), the α particles themselves are photodisintegrated into protons and neutrons:



The maximization of entropy favors the right-hand sides (because of the larger number of particles); in the final state of statistical equilibrium, the core will consist mainly of protons and neutrons. Note also that these reactions undo all of the nuclear fusion reactions of the previous nuclear burning phases. This is possible since the gravitational energy that is released in this phase corresponds to roughly 10 % of the rest-mass energy of the core ($\sim 3 \times 10^{46}$ J), which exceeds the total efficiency of nuclear burning from H to Fe by more than a factor of 10. In addition to these reactions, the material becomes increasingly neutron rich (core *neutronization*) due to electron captures such as



where Z and A refer to the charge and atomic number of a nucleus. Since these reactions take away electrons that provide an important pressure support, this further accelerates the contraction, helping to turn the initial contraction rapidly into a free-fall collapse. Most of the energy that is released in the collapse is ultimately converted into neutrinos (by the electron-capture processes listed above and other thermal neutrino processes), which freely escape from the core, at least initially.

This collapse is only stopped once matter reaches nuclear densities (ρ_{nuc}) and the strong force becomes important, providing a sudden repulsive force. Because of the initial overcompression of the matter, now mainly composed of neutrons, the core bounces and drives an outward moving shock into the still infalling outer core. It was once hoped that this shock, which initially carries an energy of $\sim 10^{44}$ J could reverse the infall of the outer core and cause an outflow, i.e., drive a *prompt explosion*. But because of the continued photodisintegration of the infalling material, which requires $\sim 10^{44}$ J for $0.1 M_\odot$ of Fe, this energy is quickly consumed; a prompt shock is always found to stall and is unable to drive an explosion.

The total energy that is released in the collapse is of order the binding energy of the neutron star forming at the center ($GM_{\text{NS}}^2/R_{\text{NS}} \sim 3 \times 10^{46}$ J $\simeq 0.1 M_{\text{NS}} c^2$ for $M_{\text{NS}} \simeq 1.4 M_\odot$ and $R_{\text{NS}} \simeq 10$ km). This is several orders of magnitude more than the binding energy of the outer core ($E_{\text{core}} \simeq 10^{44}$ J). However, most of this energy escapes freely in the form of neutrinos that only interact weakly with matter. It has remained one of the most enduring unsolved problems in supernova physics, how a fraction ($\sim 1\%$) of this energy can be deposited just below the accretion shock and be allowed to accumulate till enough energy is available to drive an explosion. In the presently favored model of *delayed neutrino-driven explosions*, this may require more than 500 ms, which is extremely long compared to the dynamical timescale

of the proto-neutron star (~ 1 ms). If this mechanism fails, matter will continue to fall onto the proto-neutron star and ultimately convert it into a black hole.⁴

2.2 Thermonuclear Explosions

The second important explosion mechanism has nothing to do with massive stars but is generally believed to occur in accreting CO white dwarfs when their mass approaches the Chandrasekhar mass. When the mass reaches $\sim 1.37 M_{\odot}$, carbon is ignited in or near the center of the white dwarf. Initially, this drives convection in the core, transporting the energy outward and radiating it away in the form of neutrinos (this phase of low-level carbon burning, referred to as the *simmering phase*, can last for up to $\sim 10^3$ year). But, there comes a point when the core is unable to rid itself of the excess nuclear energy, and the burning process becomes explosive. The reason for this nuclear runaway is that the core material is highly degenerate. This means that the core pressure is independent of temperature. Therefore, a rise in central temperature (due to the carbon burning) does not produce an increase in pressure which would limit the increase in temperature (the valve mechanism that keeps burning in ordinary stars, supported by thermal pressure, stable). The further increase in temperature increases the nuclear burning further, producing a runaway process which incinerates a large fraction of the white dwarf and ultimately destroys it completely. In the case of a thermonuclear explosion unlike the case of core collapse, no remnant is expected, and the energy source is purely nuclear energy ($\sim 10^{44}$ J). The fact that the energy in the two types of explosion is comparable ($\sim 10^{44}$ J) is not a coincidence, since, in both cases, the energy scale is set by the binding energy of the core (the CO core in the case of the thermonuclear explosion, and the binding energy of the outer Fe core in the core-collapse case), which are comparable (they are ultimately determined by the same physics of electron degeneracy, which determines the immediate pre-supernova structure).

In the ensuing explosion, a large fraction of the white dwarf is burned, in the inner part completely to *nuclear statistical equilibrium* (NSE), which means mainly to iron-group elements, mostly ^{56}Ni , and incompletely further out, producing mainly intermediate-mass elements, such as ^{28}Si and ^{32}S . The radioactive ^{56}Ni will subsequently decay to ^{56}Co (with a half-life of 6.1 d), powering the supernova lightcurve, and ultimately to ^{56}Fe (with a half-life of 77.3 d). A typical supernova of this type produces $\sim 0.7 M_{\odot}$ of ^{56}Fe ; hence, these supernovae are believed to be the dominant producers of iron in the universe. Since most of them produce very similar amounts of radioactive ^{56}Ni , the resulting supernova lightcurves are quite similar, which means that they can be used as standard distance candles (strictly speaking “standardizable” distance candles; see [Sect. 5.1](#)).

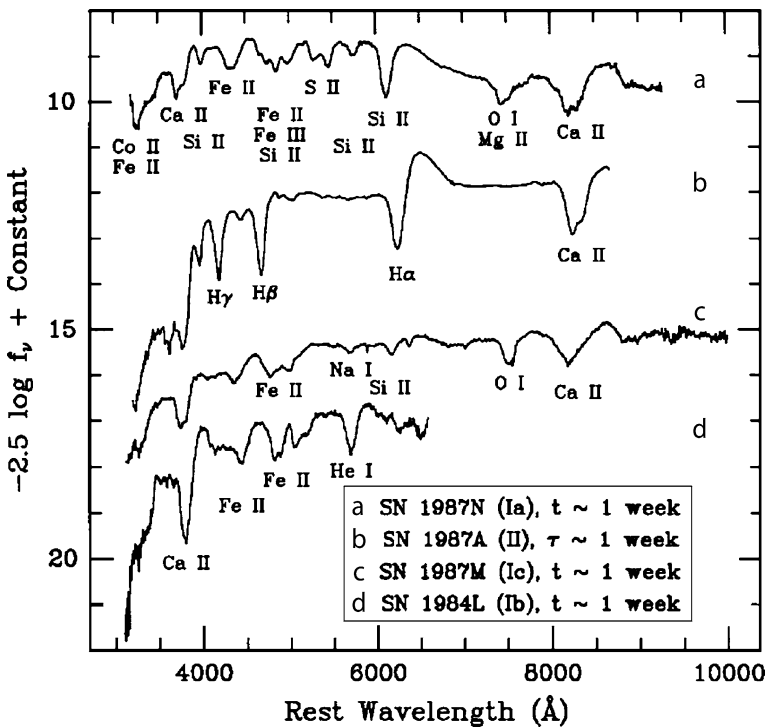
Unlike core-collapse supernovae, the physics of thermonuclear explosions is reasonably well understood. One of the lingering uncertainties is how the carbon burning front, which starts as a *deflagration* (i.e., a sub-sonic burning front), is accelerated into a *detonation* (i.e., a super-sonic burning front), which seems to be favored by observations for the majority of thermonuclear explosions.

⁴There are other ideas of how to generate a supernova explosion e.g., involving jet-driven explosions, or very strong magnetic fields. The latter also requires a very rapidly rotating pre-supernova core. While present pre-supernova models do not predict sufficiently rapidly rotating cores, there may be special circumstances in which this is the case, and this may be the origin of *magnetars*, neutron stars with very large magnetic fields, or even gamma-ray bursts (see [Sect. 7.2](#)).

The main uncertainty, even controversy, is the question of their progenitors, the type of stellar systems in which a CO white dwarf can grow toward the Chandrasekhar mass. We will return to this issue in [● Sect. 5.2](#).

3 Supernova Classification

The basic classification of supernovae is quite simple: they are classified as Type I or Type II supernovae, depending on whether they have hydrogen lines in the spectrum (Type II) or lack hydrogen lines (Type I). For a long time, it was thought that these two observational classes may have a one-to-one relation to the two explosion mechanisms discussed in the last section, core collapse supernovae (Type II) and thermonuclear explosions (Type I), respectively. However, over the last three decades, it has become clear that this is not the case and that, in principle, both explosion types could come in both observational varieties. As a consequence, the basic classification has become much more complex, requiring the introduction of more and more subtypes ([● Fig. 14-2](#)).



■ Fig. 14-2

SN spectral classification. SN Ia (a), SN II (b), SN Ic (c), and SN Ib (d) (From Filippenko ([1997](#)))

3.1 Main Classification Scheme

The thermonuclear explosion of a CO white dwarf is now believed to be associated with a *Type Ia supernova* (SN Ia). These supernovae have no hydrogen but strong Si lines. Si and also S are intermediate-mass nuclei which are produced in abundance in the part of the exploding white dwarf that does not burn completely to NSE and therefore provides a very characteristic signature for a thermonuclear explosion (● [Table 14-1](#)).

In addition to SNe Ia, there are two other subtypes of Type I supernovae, Type Ib and Type Ic. These types are also defined on the basis of their spectroscopic characteristics, both lack hydrogen, but *Type Ib supernovae* (SNe Ib) show He lines, while *Type Ic supernovae* (SNe Ic) lack both Si and He lines. Unlike SNe Ia, they produce fairly little ^{56}Ni and are found predominantly in or near star-forming regions and are therefore believed to be connected with core-collapse supernovae, i.e., the explosions of massive stars that have lost their H-rich envelopes and, in the case of SNe Ic, their He-rich layers as well.⁵

There are also several different subtypes of Type I supernovae. Unlike SNe I, they are not always defined by their spectroscopic properties but by their lightcurves, i.e., their luminosity, measured in a particular waveband (typically *B* or *V*) as a function of time. The lightcurves of *Type II-P supernovae* (SNe II-P), where the “P” stands for “plateau,” show a long phase, lasting up to a ~ 100 d where the lightcurve is nearly constant (the *plateau* phase). Their progenitors are most likely massive red supergiants (with a typical mass $\lesssim 20 M_{\odot}$) that experience core collapse. The second much less common variety, *Type II-L supernovae* (SNe II-L), do not show this plateau but drop off more or less linearly (on a logarithmic scale) after their lightcurve has peaked (hence the letter “L” for “linear”). These are almost certainly also core-collapse supernovae, but in this case, the progenitors must have already lost a large fraction of their H-rich envelopes.

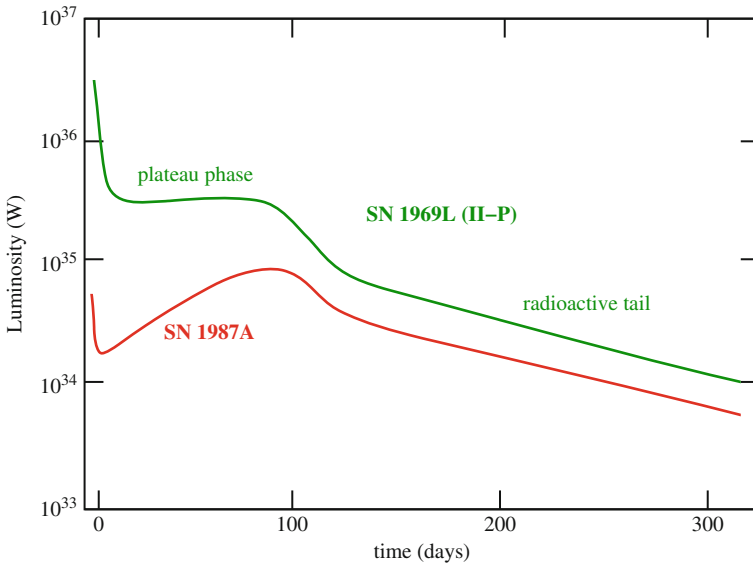
3.2 Complications

Unfortunately, there are many further complications going beyond this simple scheme. The progenitor of supernova 1987A (SN 1987A, see ● [Sect. 4](#)) had a large H-rich envelope but did not have an extended plateau phase (see ● [Fig. 14-3](#)), and therefore, SN 1987A defines a class of its own. Other supernovae appear to change their type. Supernova 1993J initially looked like a

■ **Table 14-1**
Supernova spectral classification scheme

Type I no H			Type II H		
Ia Si (S) lines	Ib No Si, but He	Ic No Si, no He	II-P Plateau lightcurve	II-L Linear lightcurve	IIb Change type II \rightarrow Ib
Thermonuclear			Core collapse		

⁵It is presently not entirely clear how much He could be present in a SN Ic. Since He is non-thermally excited, it requires the presence of a source of energetic photons, e.g., from the radioactive decay of ^{56}Ni . If the He layer is shielded from this radioactive source, it is possible in principle to hide significant amounts of He. However, the most recent estimates (Hachinger et al. 2012) suggest that at most, $0.2 M_{\odot}$ can be hidden.



■ Fig. 14-3


Bolometric supernova lightcurves for two Type II supernovae. In both cases, the progenitor was a star with a massive hydrogen envelope, but in the case of SN 1969L, the progenitor was a red supergiant, while for SN 1987A, it was a blue supergiant with $R \approx 40 R_{\odot}$


type II supernova but soon transformed into a supernova resembling a SN Ib. As a consequence, this supernova type is now referred to as a *Type IIb supernova* (SN IIb). Other subtypes are not directly related to a particular supernova mechanism but to a supernova-related phenomenon. For example, *Type IIn supernovae* stand for supernovae that show narrow H lines ($H\alpha$) in emission. These must come from H-rich material in the immediate neighborhood of the supernova, most likely ejected by the progenitor in the not-too-distant past, that was flash-ionized by the first light from the supernova. This is not necessarily related to a particular explosion type; it just implies a particular mass-loss history of the progenitor. In a more extreme version, there may be so much material around the exploding star that the supernova ejecta are rapidly slowed down by the interaction with this material, converting kinetic energy into thermal energy and ultimately radiation. In this case, the lightcurve shape itself is determined by this interaction with the circumstellar material. Supernovae (e.g., ► Sect. 8.3) that show evidence for such interactions are sometimes referred to as *Type Iia supernovae* (SNe Iia), though how this fits into the overall supernova scheme and, in particular, its relation to SNe IIb devies any obvious logic.

Indeed, as this previous discussion shows, the supernova classification scheme has become too complicated and convoluted to be very useful. In fact, sometimes even supernova experts get confused. The problem is that the main scheme is a discrete one, while the supernova properties clearly vary in a continuous manner. What one would like from a theoretical point of view is a scheme that first specifies the explosion mechanism and then has one or more continuous parameters that relate to the key properties of the progenitor that vary from supernova to supernova (such as the envelope mass; see ► Sect. 3.3). In addition, one needs a parameter that describes the environment in which the supernova occurs, in particular, the circumstellar material that may be a direct result of the mass-loss history of the progenitor.

3.3 The Diversity of Core-Collapse Supernova Lightcurves




An observable supernova requires the ejection of material (the envelope of a collapsing star or a completely disrupted star) that cools and radiates away its energy. In general, there are two main sources for the observed radiation: one is the *shock energy* that was deposited in the ejecta by the initial shock wave. The second is radioactive energy from the decay of radioactive ^{56}Ni that is produced in the explosion itself that decays first to ^{56}Co (with a half-life of 6.1 d) and then to ^{56}Fe (with a half-life of 77.3 d). In a Type II supernova, the main lightcurve is generally dominated by the release of shock energy. It is only at late times that the lightcurves are powered by radioactivity, leading to a characteristic exponential lightcurve tail. The latter allows an estimate for the amount of ^{56}Ni that was produced in the explosion. In the case of Type I supernovae, which generally have very compact progenitors, the shock energy escapes very quickly and is generally not observed, and the main lightcurve peak is powered by radioactivity. In both cases, the width of the main peak of the lightcurve is determined by the diffusion timescale on which photons escape from the expanding ejecta.

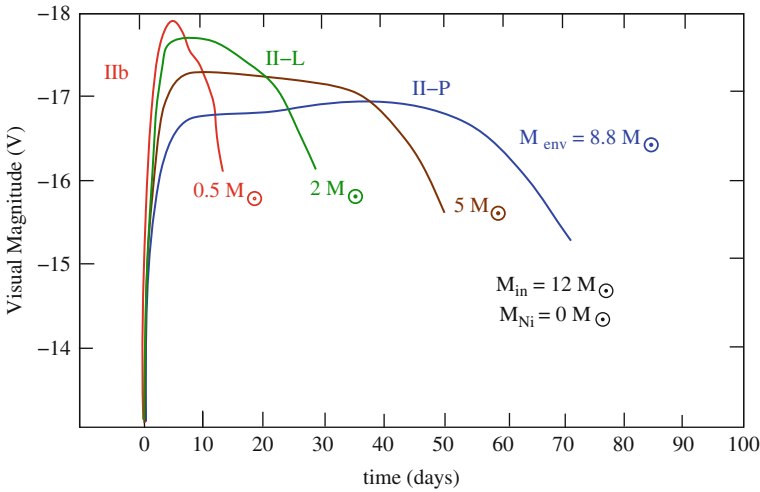
Even though there is an enormous diversity of supernova lightcurves of core-collapse supernovae, a lot of this diversity is relatively easy to understand as a variation of the envelope properties, specifically its radius and mass, even if the explosions at the center are identical.  [Figure 14-3](#) compares the lightcurve of a typical SN II-P (SN 1969L) to the lightcurve of SN 1987A. Both supernovae probably had massive progenitors with large H-rich envelopes. But the difference is that the progenitor of SN 1969L almost certainly was a red supergiant, with a radius of $\sim 1,000\text{--}2,000 R_{\odot}$, while the progenitor of SN 1987A only had a radius of $\sim 40 R_{\odot}$. In the latter case, a significant fraction of the explosion energy was consumed in expanding the ejecta from its more compact, more tightly bound configuration, making it less luminous, at least initially.

An even more important parameter is the envelope mass. This is illustrated in  [Fig. 14-4](#) which shows theoretical lightcurves (in this case, V-band magnitude versus time) for a star of initially $12 M_{\odot}$ that was assumed to have lost different amounts of envelope mass before the explosion (in all cases, the star is a red supergiant at the time of explosion). All simulations start by depositing the same amount of explosion energy (10^{44} J) in the center. For the case, where the star has lost none of its envelope and has a final presupernova H-rich envelope mass of $8.8 M_{\odot}$ (as labeled), the lightcurve is that of a typical SN II-P with a plateau phase lasting ~ 50 d.⁶ As the envelope mass is reduced, the plateau shortens and the overall luminosity increases (as the total energy input is the same). Once the envelope mass has dropped below $2 M_{\odot}$, the plateau has disappeared and this lightcurve would be classified as a SN II-L. As the envelope mass is reduced further to $0.5 M_{\odot}$, the lightcurve only has a short peak where hydrogen would be visible in the spectrum; this would be classified as a SN IIb. If the hydrogen envelope is lost completely, one would have a SN Ib and, if the He layer is lost as well, a SN Ic. Thus, the whole sequence,

$$\text{SN II-P} \rightarrow \text{SN II-L} \rightarrow \text{SN IIb} \rightarrow \text{SN Ib} \rightarrow \text{SN Ic},$$

is just a sequence of increased envelope loss, first of the H-rich envelope and then the He-rich layer. The immediate physical question is what causes this mass loss. While stellar winds may play an important role in some cases, binary interactions are almost certainly even more

⁶The reason why this lightcurve does not show the initial peak seen for SN 1969L in  [Fig. 14-3](#) is that  [Fig. 14-3](#) shows a bolometric lightcurve, while  [Fig. 14-4](#) shows a V-band lightcurve. The V-band lightcurve peaks later, as initially most of the radiation from the supernova escapes in the UV.



■ Fig. 14-4

Theoretical visual lightcurves for red-supergiant progenitors with varying masses of hydrogen (as indicated) in the red-supergiant envelope, illustrating the dependence of the lightcurves on envelope mass (Adapted from Hsu (1991))

important since a large fraction, if not the majority, of all massive stars are in relatively close binaries where the components can interact directly (e.g., by mass transfer causing mass loss, mass accretion, and, in the most extreme case, by the complete merger of the binary components (Podsiadlowski et al. 1992); see ● Sect. 7.1). These interactions particularly affect the envelope properties of the massive progenitors and hence help to determine the shapes of the resulting lightcurves.

3.3.1 Understanding SN II-P Lightcurves

Most of the key properties of a II-P lightcurve are relatively easy to understand from basic principles. A supernova is basically a huge explosion where a certain amount of energy, the explosion energy (E_{exp}), is suddenly put in the center of a large stellar envelope. The explosion energy is determined by the details of the explosion mechanism; in the presently favored delayed neutrino-driven explosion, this energy accumulates till it is sufficient to unbind the parts of the core that do not become part of the central neutron star; this sets a natural energy scale for $E_{\text{exp}} \sim 10^{44}$ J. Since this energy is injected on a timescale of $\lesssim 1$ s, which is short compared to the dynamical timescale of even the core of the star, this will drive a shock front, propagating outward through the rest of star and accelerating it in the process. The first visible sign of the supernova (apart from the neutrino signal from the initial collapse that propagates with the speed of light) appears when the shock reaches the photosphere which finds itself suddenly heated to $0.5 - 1 \times 10^6$ K, the temperature behind the shock.⁷ From this point onward,

⁷Even before the shock reaches the surface of the star, some radiation generated in the shocked region will diffuse outward faster than the shock front itself moves; when this radiation escapes from the progenitor, it

the supernova can be considered an expanding and cooling sphere of gas. Its luminosity is determined by the location of the photosphere,⁸ which recedes in mass coordinates but may either expand or contract in radius coordinates, and the characteristic temperature at the photosphere. Since this temperature is initially very high ($\sim 10^6$ K), the spectrum will first peak in soft X-rays and the UV (producing a short-lived *UV flash* that will ionize any nearby circumstellar material), moving steadily toward the B and V, as the ejecta cool, mainly due to adiabatic expansion.

The supernova shock accelerates as it moves through the envelope toward lower-density regions. As a result, the ejecta reach their highest velocities in the very outermost layers (this velocity may easily exceed $0.1 c$, where c is the speed of light). The characteristic velocity of the ejecta on the whole can be estimated simply from energy conservation,

$$E_{\text{exp}} \sim \frac{1}{2} M_{\text{ej}} v_{\text{ej}}^2, \quad (14.2)$$

where M_{ej} and v_{ej} are the mass and characteristic velocity of the ejecta, respectively (this ignores the binding energy of the envelope which is not significant for red supergiants). For typical values, this yields

$$v_{\text{ej}} \sim 3,000 \left(\frac{E_{\text{exp}}}{10^{44} \text{ J}} \right)^{1/2} \left(\frac{M_{\text{ej}}}{10 M_{\odot}} \right)^{-1/2} \text{ km s}^{-1}. \quad (14.3)$$

As the energy that was deposited by the initial shock has to diffuse out of this expanding sphere of gas, the overall length of the lightcurve is determined by the *diffusion time*, which is given by

$$t_{\text{diff}} = \frac{R^2}{lc}, \quad (14.4)$$

where R is the radius of the sphere and l the mean free path of the diffusing photons (this formula can be derived from a simple random-walk process). The mean free path itself can be related to the opacity κ and density ρ of the ejecta according to $l = 1/\kappa\rho$. Approximating ρ by $M_{\text{ej}}/4R^3$, (14.4) can be rewritten as

$$t_{\text{diff}} \sim \frac{M_{\text{ej}} \kappa}{4Rc}. \quad (14.5)$$

However, R itself increases with time t : $R(t) \sim v_{\text{ej}} t$. Substituting this into (14.5) and setting $t = t_{\text{diff}}$, one can solve the resulting equation for t_{diff} to obtain

$$t_{\text{diff}} \sim \frac{M_{\text{ej}}^{3/4} \kappa^{1/2}}{2(2E_{\text{exp}})^{1/4} c^{1/2}} \simeq 150 \text{ d} \quad (14.6)$$

(for $E_{\text{exp}} = 10^{44}$ J, $M_{\text{ej}} = 10 M_{\odot}$, $\kappa = 0.034 \text{ m}^2 \text{ kg}^{-1}$ for pure electron scattering in a solar-type plasma). This estimate is in reasonable agreement with detailed numerical calculations. The flat portion of the lightcurve in Fig. 14-3, the plateau phase, starts when the temperature at the photosphere has fallen sufficiently so that H starts to recombine (at a typical temperature $T \sim 6,000$ K). Since the recombination releases the ionization energy of H, this helps to keep

produces a radiative precursor which precedes the actual shock breakout. Such a precursor has been observed for the II-P supernova SNLS-04D2dc (Schawinski et al. 2008).

⁸Strictly speaking, there is no well-defined photosphere, as the point where the optical depth is of order unity is a strong function of wavelength.

the photosphere hotter than it otherwise would be (H recombination acts like a thermostat keeping the temperature at $\sim 6,000$ K). Since the photospheric radius changes only slowly in this phase, even though the H recombination front moves inward in mass coordinates, the overall luminosity remains roughly constant.

Once the photosphere has passed through the H-rich ejecta, the lightcurve tends to drop sharply and at late times shows the characteristic exponential tail due to the radioactive decay of ^{56}Co , which itself is the decay product of ^{56}Ni , produced in the explosion.

4 Supernova 1987A in the Large Magellanic Cloud

Supernova 1987A (SN 1987A) was the first supernova visible with the naked eye since Kepler's supernova in 1604. It occurred on February 23, 1987 in the Large Magellanic Cloud, a satellite galaxy of our own galaxy, the Milky Way, about 180,000 lightyears away. Since it was the first nearby supernova seen in almost 400 years, it has long been awaited by astronomers and therefore was one of the major astronomical events of the 1980s. The observations of neutrinos from the supernova confirmed that this type of supernova is produced by the collapse of a massive star, leading to the formation of a neutron star.

4.1 The Mystery of the Progenitor Star

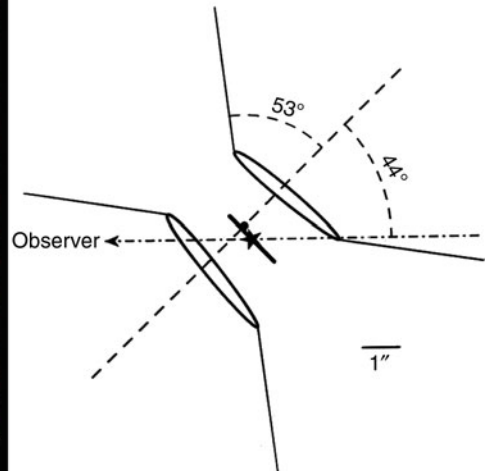
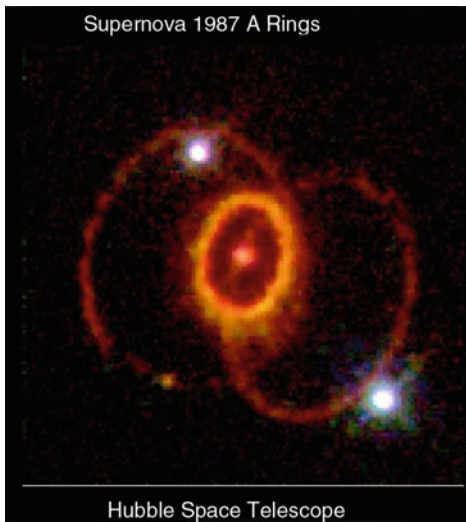
While SN 1987A confirmed some long-held beliefs about supernovae, the star that exploded defied expectations and provided a major mystery. Images of the region taken before the explosion (🔍 [Fig. 14-5](#)) showed that the star that exploded was a blue supergiant with an initial mass around $20 M_{\odot}$ and a radius at the time of the explosion of $40 R_{\odot}$. Theoretical models of the evolution of massive stars had predicted that these end their evolution as red supergiants with a radius of at least $10^3 R_{\odot}$. In addition, the composition of the star that exploded was very unusual; most importantly, the outer layers of the star had an abundance of helium (as a fraction of the total composition) that was about a factor of 2 larger than the expected abundance, as if part of the material from the core, where helium has been produced during the previous evolution of the star, was mixed into the outer layers by some unexpected mixing process. Why was the star blue and not red and why were there these chemical anomalies?

A major clue to resolve this mystery was the discovery of the triple-ring nebula around the supernova (Burrows et al. 1995). The nebula is composed of three overlapping rings, seen in projection (see the left panel of 🔍 [Fig. 14-6](#)), and consists of material that was ejected from the progenitor star some 20,000 years before the explosion. The supernova occurred right at the center of the inner ring, while the outer rings lie in planes below and above the central ring plane. The whole nebula was flash-ionized by the supernova, and we are now seeing the afterglow of this flash. The nebula is almost axisymmetric. This generally suggests that rotation may have played an important role in the shaping of the nebula. However, any star that was rapidly rotating early in its evolution could not have been rapidly rotating as a supergiant. This essentially ruled out that the progenitor star could have been a normal single star and strongly suggested that the progenitor was a member of a binary system, specifically a system where the two binary components merged completely 20,000 years before the explosion.



■ Fig. 14-5

Before and after images (*right and left*), taken with the AAT of SN 1987A in the Large Magellanic Cloud. The *arrow* in the before picture points at the progenitor, a blue supergiant

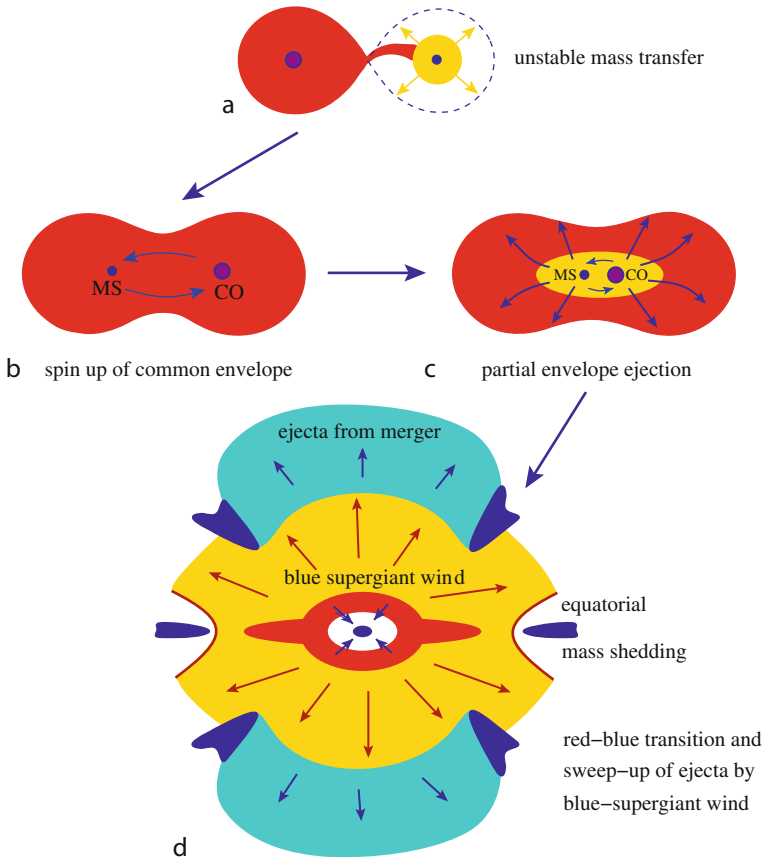


■ Fig. 14-6

The triple-ring nebula around SN 1987A as observed with the HST (*left*) and a geometric model of the three rings (*right*). The material in the three rings was ejected by the supernova progenitor $\gtrsim 20,000$ years before the explosion

4.2 A Binary Merger Model for the Progenitor of SN 1987A

The idea that the anomalous properties of the progenitor of SN 1987A are the result of the merger had already been proposed long before the discovery of the triple-ring nebula (Hillebrandt and Meyer 1989; Podsiadlowski et al. 1990). In this model, the system consisted initially of two massive stars, one with a mass of $15\text{--}20 M_{\odot}$ and a less massive companion with a mass of $2\text{--}5 M_{\odot}$, orbiting with an orbital period of at least 10 years. Because of the relatively long orbital period, these stars would only interact late during the evolution of the more massive star, in fact about 20,000 years before the explosion. When the more massive star starts to transfer mass to the less massive star, mass transfer is unstable, and the more massive star engulfs its less massive companion (see ► Fig. 14-7 for a schematic diagram). The less massive star then orbits the core of the more massive star inside the envelope of the more massive component. Because of friction with the envelope, the two stars spiral toward each other and ultimately merge completely, spinning up the envelope in the process and dredging up part of the helium



■ Fig. 14-7

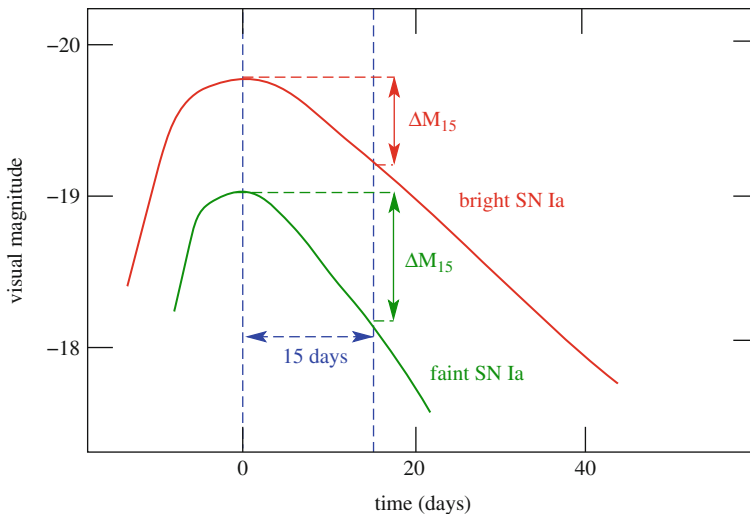
Cartoon illustrating the probable evolution of the SN 1987A progenitor and the formation of the triple ring by a merger $\sim 20,000$ years before the explosion (From Morris and Podsiadlowski (2007))

core, explaining the unusual abundances of the supernova (Ivanova and Podsiadlowski 2003). Because of the dredge-up of the helium and the increased envelope mass, the now single object shrinks to become a blue supergiant. In this transition as much as several solar masses are spun off in the equatorial direction producing a disk-like outflow. In the last 20,000 years before the star explodes, it is a blue supergiant. Such stars have strong stellar winds. This wind sweeps up all the structures that have been produced previously to produce the observed triple-ring nebula (see Morris and Podsiadlowski (2007) for details).

5 Type Ia Supernovae

5.1 Type Ia Supernovae as Cosmological Distance Candles

Type Ia supernovae (SNe Ia) have been very successfully used as standardizable distance candles and have provided the first indication for an accelerating universe (Riess et al. 1998; Perlmutter et al. 1999). It is sometimes said that SNe Ia are “standard” candles because they occur in Chandrasekhar-mass white dwarfs with standard properties. This is actually not true; the range of observed peak luminosities is a factor of 10, reflecting the vastly different amounts of ^{56}Ni produced in different supernovae ($0.1\text{--}1 M_{\odot}$), a variation that is not at all understood theoretically. However, SN Ia lightcurves vary in a very systematic fashion, where the brighter supernovae have broader lightcurves than their fainter counterparts (as illustrated in Fig. 14-8). Indeed, empirically there seems to be a one-to-one relationship between the peak magnitude and the lightcurve width which can be measured by the change in magnitude (e.g., in the B band) of the lightcurve between the peak and 15 days after the peak (defining the



■ Fig. 14-8

Visual lightcurves for SNe Ia, illustrating the relationship between peak magnitude and lightcurve width and their use as standardizable cosmological distance candles

ΔM_{15} parameter). This relationship, known as the “Phillips relation” (Phillips 1993), can be used to determine the true peak magnitude of any SN Ia and makes them useful as cosmological distance candles. The physical origin of the Phillips relation that links the peak SN Ia luminosity to the lightcurve width is now reasonably well understood (see, e.g., Mazzali et al. 2001). The supernova lightcurve is powered by the radioactive decay of ^{56}Ni into ^{56}Co and ^{56}Fe , and hence, the peak luminosity is directly proportional to the ^{56}Ni mass (Arnett 1982). The lightcurve width on the other hand is determined by the diffusion time, i.e., the timescale on which photons generated by the decay escape from the ejecta. This in turn depends on the opacity, which under these conditions is dominated by iron-group elements (in particular, ^{56}Ni , ^{58}Ni , ^{54}Fe). Since the dominant iron-group element is ^{56}Ni , which also determines the lightcurve peak, this introduces a one-to-one relationship between the lightcurve peak and the lightcurve width. However, the opacity is also dependent on other iron-group elements, such as ^{58}Ni and ^{54}Fe , which are stable elements and do not provide a radioactive energy source. The ratio of ^{58}Ni to ^{56}Ni depends on the neutron excess η in the exploding white dwarf, where η can be related to the electron number fraction Y_e by $\eta = 1 - 2Y_e$. Most importantly, Y_e depends on the initial metallicity of the progenitor and, in particular, on the initial oxygen abundance. Timmes et al. (2003) have estimated the final ^{56}Ni mass fraction as $X(^{56}\text{Ni}) \approx 1 - 0.054Z/Z_\odot$, where Z/Z_\odot is the metallicity relative to solar metallicity. This introduces metallicity as a second, necessary parameter that affects the lightcurve shape. For a reasonable range of metallicity, this may account for the observed spread in the Phillips relation. It is clear now that this metallicity effect needs to be taken into account, in particular, if one wants to improve the use of SNe Ia as cosmological distance candles.

5.2 The Progenitors of SNe Ia

One of the weakest points in our understanding of SNe Ia is the nature of their progenitors. There is now broad agreement that most SNe Ia are caused by a thermonuclear explosion of a CO white dwarf when its mass approaches the Chandrasekhar mass. What is still uncertain and is indeed controversial is the evolution that produces these Chandrasekhar-mass white dwarfs. The most popular progenitor models fall broadly into two classes, the single-degenerate (SD) model and the double-degenerate (DD) model.

5.2.1 The Single-Degenerate Model

In the SD model, the white dwarf grows in mass by accreting from a non-degenerate companion star (Whelan and Iben 1973; Nomoto 1982), where the companion star can either be a main-sequence star, a helium star, a subgiant, or even a giant. The main problem with this class of models is that it is generally difficult to increase the mass of a white dwarf by accretion: if the mass-accretion rate is too low, this causes nova explosions and/or helium flashes (Nomoto 1982) which may eject most of the accreted mass. If the mass-accretion rate is too high, most of the transferred mass must be lost in a disc wind to avoid a merger of the binary, again leading to a low accretion efficiency. There is only a very narrow parameter range where a white dwarf can accrete hydrogen-rich material and burn in a stable manner. This parameter range may be increased if differential rotation affects the accretion process (Yoon and Langer 2004). One promising channel that has been identified in recent years relates them to

supersoft X-ray sources (van den Heuvel et al. 1992). However, it is not clear whether this channel produces a sufficient number of systems to explain the observed SN Ia rate in our Galaxy ($\sim 3 \times 10^{-3} \text{ year}^{-1}$; Cappellaro and Turatto 1997). On the plus side, a number of binary systems are known that are excellent candidates for SN Ia progenitors: U Sco, RS Oph, and TCrB all contain white dwarfs that are already close to the Chandrasekhar mass, where the latter two systems are symbiotic binaries containing a giant companion (see Hachisu et al. 1999 for a discussion of this channel). However, in none of these cases, is it clear whether the massive white dwarf is a CO or an ONeMg white dwarf (the latter is not expected to produce a SN Ia).

5.2.2 The Double-Degenerate Model

In contrast to the SD model, the DD model (Iben and Tutukov 1984; Webbink 1984) involves the merger of two CO white dwarfs with a combined mass in excess of the Chandrasekhar mass. This model has the advantage that the theoretically predicted merger rate is quite high (see, e.g., Yungelson et al. 1994; Han et al. 1995; Nelemans et al. 2001), consistent with the observed SN Ia rate. The main problem with this scenario is that it seems more likely that the disruption of the lighter white dwarf and the accretion of its debris onto the more massive one leads to the transformation of the surviving CO white dwarf into an ONeMg white dwarf which subsequently collapses to form a neutron star (i.e., undergoes accretion-induced collapse) rather than experiencing a thermonuclear explosion (e.g., Nomoto and Iben 1985), although there may be circumstances under which this can be avoided (Yoon et al. 2007).

In recent years, numerous observational tests have been devised to test these various progenitor models, sometimes with strong claims (“this rules out . . .”), but more often with claims that contradict others, and the whole discussion has remained fluid. Some of these tests involve the detection of circumstellar material, as might be produced by the wind of a single-degenerate donor star, the detection of a surviving companion in a historical supernova remnant in our own galaxy, the direct detection of a progenitor on old plates, and many more.

Two recent SNe Ia have provided important new constraints. The first one was one of the closest SNe Ia in many decades, SN 2011fe/PTF11kly, in M101, the pinwheel galaxy, which reached a peak apparent magnitude of 10 and will remain observable for many years to come. Because M101 is a well-studied galaxy, there are deep archival HST images of the supernova field. Since these do not show any counterpart at the explosion site, this rules out a single-degenerate progenitor where the donor star is a typical red giant, such as in RS Oph (Li et al. 2011). In contrast, a second recent supernova, PTF 11kx (Dilday et al. 2012) shows clear evidence for hydrogen in the immediate vicinity of the supernova, which must have been ejected from the progenitor system. The most likely candidate is a red-giant donor with a somewhat larger mass-loss rate than is observed in RS Oph. This clearly demonstrates that there is more than one channel that can produce a normal-looking SN Ia.

6 Gamma-Ray Bursts, Collapsars, and Hypernovae

6.1 The History of Gamma-Ray Bursts

Gamma-ray bursts (GRBs) are short flashes of gamma rays, lasting from a fraction of a second to hundreds, sometimes thousands of seconds. They were discovered accidentally in 1967 by a US

spy satellite that was scanning the Earth's atmosphere for gamma rays from Russian atmospheric nuclear tests in violation of the atmospheric test-ban treaty. Instead of finding gamma rays from the Russians, the satellite found GRBs from all over the sky. This discovery was kept secret till 1973 (Klebesadel et al. 1973) and remained one of the biggest mysteries in astronomy until 1998. At some point, there were more theories for GRBs than actual bursts. Initially, most people thought that these were probably associated with events on neutron stars in our own galaxy. The mystery only deepened when in the early 1990s a new satellite was launched, NASA's Compton-Ray Observatory, and BATSE, one of its onboard experiments, found that GRBs were not concentrated toward the Galactic disc and center, as most people had expected, but that their locations were distributed perfectly isotropically all over the sky. This ruled out one of the leading models, neutron stars in the Galactic disc, as sources of these events. It still did not fix their distance scale as one can imagine various isotropic distributions. The closest is the Oort cloud of comets around our own solar system, although it is far from clear how these could produce GRBs. A more realistic alternative were sources in the halo of the Milky Way; if the characteristic distance scale of a halo population (e.g., of a special subset of neutron stars or other compact objects) is sufficiently large, it could potentially produce an isotropic sky distribution. The most dramatic possibility was that GRBs originated from the distant universe and that the GRB isotropy reflected the large-scale isotropy of the universe.⁹ Depending on the distance scale, GRBs could be associated with feeble events in the outer solar system or with some of the most energetic events in the universe. One problem with BATSE was that it was not able to localize GRBs very well, at best only to within a few degrees. Such a large field of sky contains far too many sources and makes it impossible to find GRB counterparts in other wavebands.

The resolution of this mystery came with the launch of another high-energy satellite in 1996, BeppoSAX, an Italian-Dutch collaboration, which was able to localize GRBs much better than BATSE (within about 1 arcmin) and, moreover, very quickly, allowing the rapid follow-up with optical telescopes. This quickly led to the first detection of a GRB in the optical, an *optical afterglow* associated with a GRB. Moreover, optical spectra of these afterglows showed highly redshifted absorption lines, which must be caused by galaxies in front of the GRBs. This proved beyond doubt that they must originate from cosmological distances, connecting them with the most energetic events in the universe (apart from the Big Bang itself). Indeed, GRBs are so energetic that they can be detected throughout the observable universe and can be used as a probe of the early universe. At present, the most distant explosion in the universe, and arguably the most distant object known in the Universe, is a GRB at a redshift of 9.4 (Cucchiara et al. 2011).

6.2 The Main Properties of GRBs

GRBs come in many different shapes; some show substructure on sub-millisecond scales, others have a smooth time profile lasting tens to hundreds of seconds, perhaps with some short spikes superimposed. Others seem to turn off, only to produce another burst-like event 100s, sometimes 1,000s of seconds later. Based on their duration, GRBs are divided into two

⁹Before the advent of BATSE, one of the few astrophysicists who strongly advocated a cosmological origin for GRBs was Bohdan Paczyński whose arguments at the time, however, were not taken very seriously by most people in the field.

groups, short-duration GRBs (SGRBs) with an average duration of 0.3 s and long-duration GRBs (LGRBs) with a median duration of around 20 s, where the spectra of the former group tend to be significantly harder. At present, optical afterglows with intervening absorption lines have only been discovered for LGRBs, though a few SGRBs are found close to big galaxies, making an association likely. It is generally believed that the two groups of GRBs are associated with very different events. SGRBs are probably caused by the merger of two compact objects, in particular, two neutron stars or a neutron star and a black hole. In these systems, the two compact objects are driven toward each other by the loss of angular momentum by gravitational radiation, ultimately leading to a cataclysmic merger and a burst of gravitational radiation and gamma rays. In contrast, LGRBs are probably connected with the death of massive stars, as discussed in more detail in [Sect. 6.4](#). To complicate things further, there is another class of GRBs, so-called *soft gamma-ray repeaters* that have a Galactic origin. They tend to have a softer spectrum than classical GRBs (i.e., SGRBs and LGRBs) and appear to be associated with repeating, nondestructive events occurring on magnetic neutron stars, so-called magnetars.¹⁰

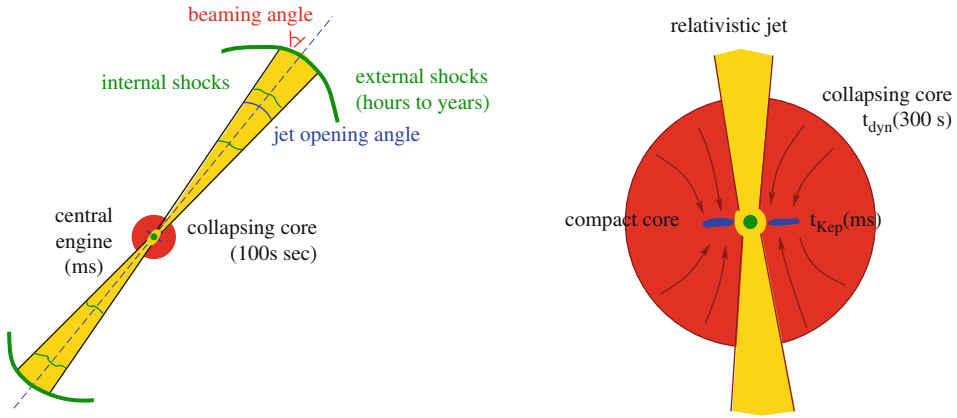
The spectra of GRBs tend to be quite hard with most of the energy being emitted at $\gtrsim 200$ keV, and the spectra themselves can best be described by a power law, indicating a mainly nonthermal origin, such as synchrotron radiation and/or inverse Compton radiation.

The apparent range of total energy coming out in gamma rays can be large, ranging from 10^{44} to several 10^{47} J. However, this assumes that the gamma-ray emission is isotropic. In fact, it is generally accepted that GRBs are highly beamed, both relativistically and geometrically.

6.2.1 Relativistic Beaming of GRBs

Detailed models of GRBs indicate that the material that emits the gamma rays moves relativistically with Lorentz factors $\gamma \gtrsim 100$. This implies that the radiation is *relativistically beamed* with most of the emission confined to a narrow beam in the direction of the flow with a beaming angle $\theta_{\text{rel}} \lesssim 1/\gamma$. Since the flow itself is believed to be confined to a narrow jet with an opening angle $\theta_{\text{jet}} \sim 1^\circ - 10^\circ$, only an observer within the beam will be able to observe a GRB (see [Fig. 14-9](#)). This has two important implications: (1) since the energy is not isotropic, the true energy in a GRB will be reduced by a factor of $4\pi/\Delta\Omega$ over the naive estimate based on isotropic emission, where $\Delta\Omega$ is the effective solid angle of the beam. (2) Since the beaming angle is small, most GRBs are not observable (at least not in gamma rays). This implies that the true GRB event rate has to be increased by the same factor $4\pi/\Delta\Omega$. Initially, θ_{rel} will generally be less than θ_{jet} . However, as the material in the jet is slowed down, most likely because of an interaction with the circumstellar medium, there will come a point when θ_{rel} becomes larger than θ_{jet} , which means that the afterglow emission is spread over a larger solid angle. This produces a characteristic break in the afterglow lightcurve and can be used, in combination with detailed modeling of the afterglow, to estimate the jet-opening angle and hence the beaming correction (typically a factor of a few hundred). Indeed, when this is done, this reduces the estimate of GRB energies to a much narrower range around 10^{44} J (Frail et al. 2001), similar to the

¹⁰BATSE has also detected many GRB-like events from the Earth's atmosphere; these are associated with thunderstorms in the upper atmosphere. While they cannot be mistaken for classical GRBs because of their location, it is amusing that nature produces events that look very similar in terms of durations and observable gamma-ray fluxes but have very different underlying mechanisms on hugely varying energy scales.



■ Fig. 14-9

Cartoon illustrating the origin of long-duration GRBs in the collapsar model

energy in a normal supernova.¹¹ However, this accounts only for the energy in gamma rays; the total energy in the GRB event, including the energy in the jet itself and any possibly associated supernova event, is probably at least an order of magnitude larger (see ▶ Sect. 6.5). It also means that, while there are about 3 GRBs in the universe per day that we can detect in principle, the true rate is at least a few hundred times larger. This still makes GRBs extremely rare events; for comparison, core-collapse supernovae are about a factor of 1,000 more common. Therefore, whatever produces a GRB clearly requires very special circumstances.

6.3 The Relativistic Fireball Model

6.3.1 The Compactness and the Baryon Loading Problem

GRBs are variable on timescales of $t_{\min} \sim 10^{-3}$ s. This implies that the emitting region, which should scale like $c t_{\min}$ (where c is the speed of light) is $\sim 10^5$ m. However, this causes an immediate problem; if one confined all the photons emitted in a GRB to a sphere of that radius, this photon gas would be optically thick to the production of electron/positron pairs (since most photons have energies larger than 0.5 MeV). This has several consequences. (1) The typical photon energy is rapidly downgraded to energies < 0.5 MeV. (2) The gamma-ray energy is converted into kinetic energy (of the e^+e^- plasma). (3) The resulting spectrum becomes that of a modified blackbody, which is not what is observed.

In addition, this emitting region has to be essentially baryon-free. Otherwise, the gamma-ray energy would mainly go into accelerating the baryons. What is needed is an essentially baryon-free *relativistic fireball* consisting only of electrons, positrons, and gamma-ray photons.

Having such a fireball confined to a jet alleviates the problem slightly as it reduces the rate of pair production but does not solve it.

¹¹NASA's Swift satellite, launched in 2004, has only been able to detect clear jet breaks in a few GRBs, despite its much better sensitivity. It is presently not clear what implications this has for estimates of the jet opening angle and the beaming correction that needs to be applied.

6.3.2 The Role of Shocks

The solution of this conundrum is that indeed the energy that is produced in the central GRB engine goes first into the kinetic energy of the relativistic jet but that this energy is later reconverted into photons by shocks *after* the flow has become optically thin (Meszaros and Rees 1993), e.g., emitted in the form of synchrotron radiation in the shocked material. Two types of shocks may play a role in this picture: *external shocks* caused by the fireball running into an external medium and *internal shocks* in the relativistic flow itself, as faster portions of the flow catch up with slower ones (see ► Fig. 14-9). It is easy to imagine that a combination of the two types of shocks can account for the large diversity of observed GRB profiles, where the internal shocks may account for a lot of the fine structure of the bursts, while the external shocks are more likely to produce the broader structures. Even though this fireball model does not depend on the details of the central engine that actually drives the GRB, the engine still has to be variable on very short timescales to produce the observed variations in the flow and its interaction with the surrounding medium.

6.4 The Collapsar Model and the Central Engine

Observationally, long-duration GRBs involve processes that occur on timescales that vary by many orders of magnitude (see ► Fig. 14-9). The central engine that drives the GRB operates on a characteristic timescales of milliseconds with sub-millisecond variations, while the main GRB must last for $10\text{--}10^3$ s. The longest timescale associated with the optical afterglow (hours to years) is probably the easiest to understand; in the framework of the relativistic fireball model, it is just determined by the interaction of the relativistic jet with the surrounding medium (► Sect. 6.3) and any supernova event possibly associated with the GRB (► Sect. 6.5). The presently favored model for the GRB itself is the *collapsar model* (Woosley 1993), which connects a long-duration GRB with the collapse of the rapidly rotating core of a massive star to a black hole.¹² Because of its rapid rotation, such a core cannot collapse directly to a black hole since angular momentum has to be conserved in a dynamical collapse. While the low-angular-momentum material along the polar region can collapse directly to form a proto-neutron star or black hole at the center, most of the high-angular-momentum material first has to first form a centrifugally supported disc orbiting the central object. Viscous processes within the disc will then allow accretion of this material onto the black hole. The characteristic timescale associated with this is the Keplerian timescale near the central compact object, which for a neutron star or low-mass black hole is of order milliseconds. On the the hand, the overall duration of the accretion activity is determined by the collapse timescale of the core which is of order its dynamical timescale, which can be estimated as $t_{\text{dyn}} \sim 1/\sqrt{4G\rho}$, where ρ is the mean density of the core. Taking the typical density of a collapsing helium star as $\rho_c \simeq 4 \times 10^4 \text{ kg m}^{-3}$ then yields a characteristic collapse timescale of $t_{\text{dyn}} \sim 300$ s, very much consistent with the characteristic observed duration of LGRBs. Thus, the collapsar model naturally explains both the

¹²The collapsar model is not the only model presently under consideration. One promising alternative involves the formation of a rapidly rotating neutron star with a very strong magnetic field, a magnetar, which is spun down on a timescale of a few seconds, extracting a large fraction of the rotational energy and powering a GRB (see, e.g., Metzger et al. 2011). Similar to the collapsar model, it requires a rapidly rotating progenitor core, but the final object is more likely a neutron star.

overall duration of an LGRB and its short-term variability. Note that the accretion rate onto the central object is generally determined by the viscous processes in the disc and not the collapse of the core itself. If the central accretion rate is lower than the overall infall rate, as is particularly likely to happen in the early phase of collapse, the mass in the disc will increase until the disc itself becomes gravitationally unstable (via the Toomre instability; Toomre 1964). The disc then becomes globally unstable and will develop spiral modes that very efficiently redistribute angular momentum in the disc (on a dynamical timescale and not the natural viscous timescale of the disc). This leads to a sudden increase in the central accretion rate ($\gtrsim 1 M_{\odot} \text{ s}^{-1}$), reflected in a sharp increase in GRB activity (for details, see Taylor et al. 2011). In addition, thermal instabilities in such an unstable disc cause clumping of disc material and, as these clumps are accreted, could cause variability on sub-millisecond timescales.

A necessary requirement for the collapsar model is that the collapsing core has enough angular momentum to be able to form a disc. We can estimate the amount required by considering the case of forming a disc around a black hole of mass M_{BH} that formed in the initial collapse. Treating this in a completely Newtonian manner, the specific angular momentum of a Keplerian disc is given by $j = \sqrt{GM_{\text{BH}} R_{\text{in}}}$, where R_{in} is the inner radius of the disc. Taking this radius to be the last stable orbit around a Schwarzschild black hole, $R_{\text{in}} = 6GM_{\text{BH}}/c^2$, the critical j for disc formation can be written as $j_{\text{crit}} = \sqrt{6} GM_{\text{BH}}/c \simeq 2 \times 10^{12} \text{ m}^2 \text{ s}^{-1} (M_{\text{BH}}/2M_{\odot})$. This estimate gives the specific angular momentum required in the outer parts of the progenitor's core.

6.4.1 The GRB Energy Source

In order to produce an energetic GRB, more energy is required than in an ordinary supernova (taking into account that only a fraction of the total energy emerges in gamma rays). There are two potential energy sources. The first is the gravitational binding energy of the disc that is released as material is accreted onto the central black hole. In terms of the rest-mass energy of the accreted material, M_{acc} , this can be written as

$$E = \eta M_{\text{acc}} c^2. \quad (14.7)$$

For a non-rotating Schwarzschild black hole, the efficiency factor $\eta \simeq 0.06$, as determined by the binding energy at the last stable orbit. For a rotating Kerr black hole, η can, in principle, be as high as 42%, although in practice it is unlikely to be larger than $\sim 30\%$. For a disc mass of $\sim 1 M_{\odot}$, this means that more than 10^{46} J could potentially be extracted from the disc.

The second source of energy is the spin energy of the black hole itself. If the material that is accreted by a rotating black hole is magnetically connected to the disc surrounding the black hole (provided the disc has a strong poloidal magnetic field), this causes a magnetic coupling of the black hole's ergosphere with the disc, which allows black-hole spin energy to be extracted. This is known as the *Blandford-Znajek mechanism* (Blandford and Znajek 1977). Up to 29% of the spin energy can in principle be extracted, where more generally the extractable energy is given by

$$E_{\text{BZ}} = 5 \times 10^{46} \text{ J } f(a) (M_{\text{BH}}/M_{\odot}), \quad (14.8)$$

where $f(a) = 1 - ([a + \sqrt{1 - a^2}]/2)^{1/2} \leq 0.29$ and a is the angular momentum parameter that can vary between 0 and 1.

For both energy sources, there is more than enough energy to power even a very powerful GRB. The key is that, because of the rotation, a much larger fraction of the gravitational binding

energy of the compact object being formed can be used than in a classical neutrino-driven explosion, where most of this energy just escapes in the form of neutrinos and is wasted.

6.4.2 The Production of a Relativistic Jet

Relativistic jets are a generic feature of most GRB models; how they are produced is less clear. One possibility is that they are *neutrino-driven jets*. Just, as in the case of a normal core-collapse supernova, neutrinos are copiously produced and in fact provide the main cooling process. Most of these neutrinos will escape from the system, but those that interact with other neutrinos will do so preferentially along the polar axis (for geometrical reasons) producing e^+e^- pairs. Since this is also the region with low baryon loading, this could potentially drive a relativistic jet. It is not clear whether this process is efficient enough in the case of LGRBs but could be in the case of merging neutron stars in the context of SGRBs. For LGRBs, it appears more likely that the relativistic jets are produced by some, presently not well understood MHD process as matter is accreted at a very high rate.

6.5 Hypernovae: The LGRB – Supernova Connection

In the late 1990s, it was realized that, in addition to the normal core-collapse and thermonuclear explosions, there are more energetic supernovae with an energy output $\gtrsim 10^{45}$ J, i.e., they are at least ten times as energetic as a normal supernova. These are now often referred to as *hypernovae* (HNe) or alternatively as *broad-lined supernovae* since they have very broad lines, reflecting their more energetic nature. The prototype is SN 1998bw (Iwamoto et al. 1998). SN 1998bw was also observed as an LGRB, establishing the first connection between a GRB and a supernova, the death of a massive star. Admittedly, the GRB associated with SN 1998bw was rather feeble, causing some arguments in the literature about the validity of this link, but there have been many more GRB/SN connections since, in some cases involving very typical GRBs (e.g., SN 2009nz/GRB 091127; Berger et al. 2011). This leaves no doubt that at least some LGRBs are associated with the explosions of massive stars. In addition, many GRB afterglows show bumps in the lightcurve that are consistent with an underlying hypernova-like event. Interestingly, at present, all GRB supernovae are classified as SNe Ic, i.e., supernovae that have lost both their hydrogen and their helium envelopes, providing a potentially important clue to the nature of their progenitors. Hypernovae are just as rare as LGRBs with a characteristic rate for a galaxy like the Milky Way of 10^{-5} year $^{-1}$ (Podsiadlowski et al. 2004b). This, however, does not mean that all LGRBs are accompanied by a hypernova. In order to see a hypernova, the event has to produce a large amount of ^{56}Ni since it is the radioactive decay of ^{56}Ni that powers a SN Ic lightcurve. However, Ni production is not required in the collapsar model; indeed, Woosley (1993) originally suggested that this model would produce a faint supernova or no supernova at all. At present, it is not clear where Ni is produced in the collapsar model, and it is conceivable that in some cases, no or very little Ni is ejected, leading to a dark GRB burst, i.e., a GRB not accompanied by a bright afterglow or hypernova.

One would also not necessarily expect every hypernova to be associated with a GRB, even if the hypernova is created by a central engine. In order to observe a GRB, a relativistic jet has not only to form but also has to escape from the collapsing star, i.e., be able to penetrate the

remaining envelope. If the jet fails to escape, it may still drive a powerful explosion but may not be accompanied by a GRB: it would be a *failed GRB*.

Even though hypernovae are rare today, if they were more common in the past (e.g., because a lower metallicity makes them more common), they could be important contributors to the global nucleosynthesis as they produce a lot of Fe (from the decay of Ni) at very early times and other elements like Zn, that are overproduced in such energetic explosions.

6.6 The Progenitors of GRBs

For the case of short-duration GRBs, there is broad agreement that they are probably associated with the merger of two compact objects, two neutron stars, a neutron star and a black hole, or perhaps a neutron star and a white dwarf. The situation is very different for long-duration GRBs, where there is not even agreement whether they involve single stars or binaries. In either case, as GRBs are rare, there have to be special circumstances that produce a GRB progenitor.

6.6.1 Single-Star Progenitor Models for LGRBs

The problem with single stars is that they very efficiently lose their angular momentum in stellar winds. Moreover, even very weak seed magnetic fields efficiently redistribute angular momentum in a star (trying to erase any differential rotation), which prevents the formation of a rapidly rotating core (see Heger et al. 2005). This problem is less severe at lower metallicity where the wind mass-loss rates are much lower. Moreover, if stars rotate extremely rapidly, they may evolve homogeneously and avoid a red-supergiant phase altogether (see ▶ Sect. 7.3) and the associated strong mass loss. Yoon and Langer (2005b) and Woosley and Heger (2006) have proposed that such *homogeneous evolution* can produce rapidly rotating progenitor cores, provided that the metallicity is quite low, typically less than 1/5 solar (Yoon et al. 2006). Indeed, there has been some observational evidence that LGRBs are preferentially found in small galaxies with low metallicity, although in recent years, it has become clear that LGRBs can also occur at solar or even super-solar metallicity (e.g., Levesque et al. 2010). Therefore, it seems from observations that there is no strong metallicity cutoff for GRBs but perhaps just a metallicity bias.

6.6.2 Binary Progenitor Models

In some sense, it seems much easier to produce a GRBs in a binary as there always is an overabundance of stored angular momentum in the form of the orbital angular momentum of the binary: one just has to be able to tap that orbital angular momentum at the right time to produce a rapidly rotating stellar core (within the framework of the collapsar model). But even this is not so easy, as the stars in binaries are subject to the same processes that spin-down single stars, such as stellar winds and magnetic coupling.

Considering the complexities of binary evolution, it is perhaps not surprising that there is large number of proposed binary scenarios which we can only sketch here briefly (for a detailed review, see Fryer et al. 2007).

In many respects, the simplest binary process that can produce a rapidly rotating helium star is *tidal spin-up* since, in a tidally locked binary, a star can be spun up (or down) until its

spin angular velocity is equal to the orbital angular velocity (e.g., Izzard et al. 2004; Detmers et al. 2008). Simple angular-momentum estimates suggest that this requires an orbital period shorter than ~ 10 h. In practice, this means that the companion is most likely a compact object (a neutron star or a black hole), similar to the X-ray binary Cygnus X-3. However, Detmers et al. (2008) found that, at solar metallicity, the expected strong wind from the Wolf-Rayet star leads to a significant widening of the binary and the ultimate *spin-down* of the companion. As a consequence, this channel is only likely to work at low metallicity.

Most binary models for LGRBs proposed to date involve the *merger of two stars* of some type, e.g., a compact object (neutron star or black hole) with a helium core, two helium cores, a helium core, and a CO core (see Fryer et al. 2007 for detailed discussions and references). This is a particularly efficient way for converting *orbital* angular momentum into *spin* angular momentum. These models work best if the interaction occurs late in the evolution of the progenitor (so-called Case C mass transfer; see ▶ Sect. 7.1) as there is little remaining time after the merger to spin down the merger product in a strong Wolf-Rayet wind.

Observationally all GRB-related supernovae to date have been identified as SNe Ic, i.e., involve progenitors that have lost both their hydrogen and their helium envelopes, an issue that it is not usually addressed in these models. There has been some debate of how much helium could be hidden in a SN Ic. The most recent estimates suggest that it could be only very little (less than a few $0.1 M_{\odot}$; Hachinger et al. 2012); this provides a serious challenge to all the models mentioned so far, single and binary.

This problem could potentially be avoided if the merger itself is explosive. Podsiadlowski et al. (2010) proposed an *explosive common-envelope ejection* scenario in certain types of late binary mergers, where the merging of a hydrogen-rich low-mass star with an evolved massive helium core triggers a thermonuclear runaway in the helium layer, leading to the ejection of both the hydrogen envelope and the helium layer. As the CO core is moderately spin up in the process, this could provide a viable (though probably not very common) channel for a GRB associated with a SN Ic.

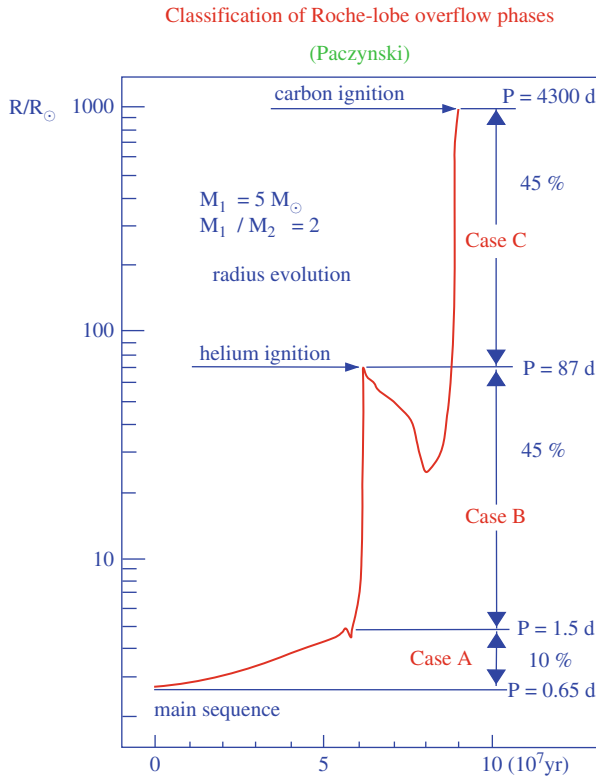
7 The Diversity of Supernova Explosions

7.1 The Role of Binarity

While it has been clear for many years that binary interactions strongly affect the structures of stellar envelopes, both by mass loss and by mass accretion, and hence are likely to be a major cause for the observed diversity of supernova subtypes, it has only recently become clear that they can also alter the core evolution and, in fact, the final fate of a star. Generically, one expects that, if mass loss/accretion occurs during an early evolutionary phase, the core continues to evolve subsequently like a less or more massive star. However, this is not true if mass loss occurs after the main-sequence phase.

7.1.1 Types of Binary Interactions

While most stars in the sky are probably in binary system, the only ones we are interested here are those where at least one of the components transfers mass to the other one by Roche-lobe



■ Fig. 14-10

The evolution of the radius of a $5 M_{\odot}$ star as a function of its lifetime to illustrate the ranges in radius and orbital period for the different cases of RLOF phases, as indicated, assuming a $2 M_{\odot}$ companion

overflow (RLOF). The fraction of massive binaries that interact at some point during their evolution has been estimated to be in the range of 30–50% of all stellar systems, where the more massive stars are likely to interact more frequently. Indeed, there is even some evidence that for binaries containing O stars, the fraction of interacting binaries may be as high as 75 %. For the first phase¹³ of mass transfer for one of the stars, one distinguishes three cases of mass transfer depending on the nuclear evolutionary state of the star: *case A* (the star is on the main sequence burning hydrogen), *case B* (the star has finished hydrogen burning but not helium burning in the core), and *case C* (the star has completed core helium burning). ● Figure 14-10 shows the radius evolution of a $5 M_{\odot}$ star as a function of time and indicates the range where the different cases occur. Since the radius of the star expands only very little (a factor of ~ 2) on the main sequence but a factor of more than ten before helium ignition and again after helium burning, it is much more likely that RLOF starts after the star has completed its main-sequence phase (this assumes a logarithmically flat initial period distribution). On the other hand, since a star spends most of its life on the main sequence, it follows that most binaries observed in the sky

¹³If a star experiences more than one mass-transfer phase, the nomenclature quickly becomes complicated, and there is no established standard notation.

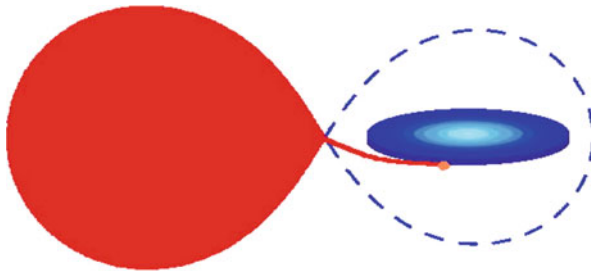
have not yet had a strong binary interaction, but many of them will do so in the future. This is particularly important when studying the end states of stars and supernovae that probe the late evolutionary phases of a star. Note also that quite massive stars ($\gtrsim 20 M_{\odot}$) tend to expand only moderately after helium core burning, and hence, for massive stars, case C mass transfer tends to be much less important than case B mass transfer, where most of the expansion occurs.

When RLOF occurs, one has to distinguish between different modes of mass transfer, depending on whether mass transfer is stable or unstable with very different outcomes.

Stable Mass Transfer

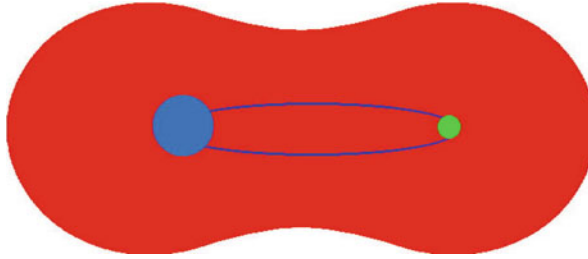
Stable, (quasi-)conservative mass transfer (as illustrated in ► Fig. 14-11) is the easiest type of mass transfer to understand. In this case, most, but not necessarily all, of the transferred mass is accreted by the companion star, generally leading to a widening of the binary. Mass transfer ends when most of the hydrogen-rich envelope of the donor star has either been transferred to the companion or been lost from the system. The end product will be a hydrogen-exhausted helium star with at most a small hydrogen-rich envelope.¹⁴ Mass accretion will also change the structure of the accreting star. If it is still on the main sequence, the accretor tends to be *rejuvenated* and then behaves like a more massive normal main-sequence star. On the other hand, if it has already left the main sequence, its evolution can be drastically altered, and the star may never evolve to become a red supergiant but explode as a blue supergiant (if it is a massive star; Podsiadlowski and Joss 1989).

This is illustrated in ► Fig. 14-12 which shows the evolutionary tracks of both components in a binary in a model that reproduces the properties of the progenitor system of SN 1993J, a Type IIb supernova in M81 (Maund et al. 2004). The primary with an initial mass of $15 M_{\odot}$ starts to experience stable mass transfer near the end of its evolution (late case C mass transfer) but still has a small envelope of $0.3 M_{\odot}$ left at the time of the explosion, consistent with the requirement for a SN IIb (see ► Sect. 3.3). The companion accretes a large fraction of this mass, but because its initial mass ($14 M_{\odot}$) is close to the mass of the primary (in order to ensure stable mass transfer), it has already left the main sequence and because of the accretion never becomes a red supergiant. It will spend the rest of its evolution as a blue supergiant and finally explode in



■ Fig. 14-11
Cartoon illustrating stable mass transfer

¹⁴Stable mass transfer can also occur for an expanding hydrogen-exhausted helium star (so-called *case BB* mass transfer). In this case, the star is likely to lose a large fraction/most of its helium envelope. This can produce a SN Ic progenitor with very low ejecta mass.



■ Fig. 14-12

Cartoon illustrating unstable mass transfer

a supernova similar to SN 1987A (● Sect. 4). Thus, this single observed system illustrates two ways in which binary evolution produces a particular supernova subtype.

Unstable Mass Transfer and Common-Envelope Evolution

Mass transfer is unstable when the accreting star cannot accrete all off the material transferred from the donor star. The transferred material then piles up on the accretor and starts to expand, ultimately filling and overfilling the accretor's Roche lobe. This leads to the formation of a common-envelope (CE) system, where the core of the donor and the companion form a binary immersed in the envelope of the donor star (see ● Fig. 14-13). Friction between the immersed binary and the envelope will make the two components spiral toward each other until enough orbital energy has been released to eject the envelope (Paczynski 1976). This ends the spiral-in phase and leaves a much closer binary consisting of the core of the giant and a normal-star secondary. If the core is massive enough to explode in a supernova, it will produce a SN Ib or Ic. Since this spiral-in phase is very short-lived, the immersed companion star will not be able to accrete much matter and will emerge little changed from the CE phase.

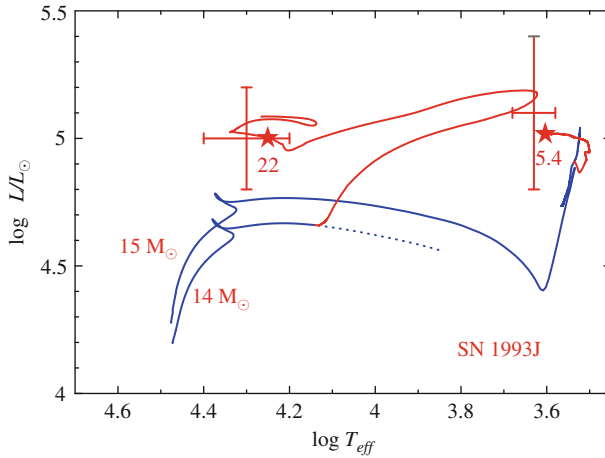
Binary Mergers

The most dramatic consequence of a CE phase is that the orbital energy that is released in the spiral-in phase is not sufficient to eject the envelope. In this case, the spiral-in process continues till the core of the donor has merged with the companion, producing a *single*, initially rapidly rotating star (such as FK Comae stars).

Binary mergers are one of the least studied phases of binary evolution. Despite of their lack of attention, binary mergers are by no means rare events: estimates based on binary population synthesis (BPS) studies suggest that ~5–10% of *all stars* experience a complete merger with a companion star during their evolution and this fraction could be substantially higher for binaries containing O stars.

7.1.2 Black Hole or Neutron Star?

Binary interactions can also drastically change the core evolution of a massive star, altering its final fate. This is particularly true if a star loses its hydrogen envelope just before or early during He-core burning (early case B mass transfer).



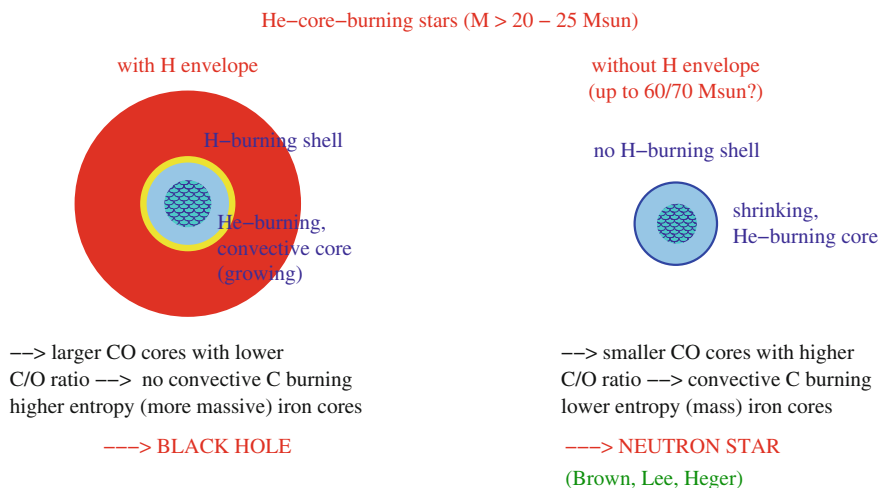
■ Fig. 14-13

Hertzsprung–Russell diagram showing the evolution of the two components of a massive binary that experiences stable case C mass transfer. At the time of the explosion, the primary is a $5.4 M_{\odot}$ star with a small $0.3 M_{\odot}$ H-rich envelope. Because of the accreted mass, the secondary never becomes a red supergiant and explodes as a blue supergiant similar to SN 1987A. The locations of the two components in the observed progenitor system of the IIb supernova SN 1993J are indicated by large error bars (From Maund et al. 2004)

First, consider the case of a massive star ($>20\text{--}25 M_{\odot}$) that, if single, would be likely to eventually collapse to a black hole (Fryer and Kalogera 2001). However, if it loses its H-rich envelope early enough, its evolution is quite different (see ● Fig. 14-14). Because of the lack of a H-burning shell, the convective core does not grow during helium core burning, and stars end up with much smaller CO and ultimately iron cores (Brown et al. 2001). Indeed, because of this, such H-deficient stars formed in case B binaries are expected to end their evolution as neutron stars rather than as black holes (Brown et al. 2001), even for initial masses as high as $\sim 60 M_{\odot}$. In contrast, single stars are believed to produce black holes for much lower initial masses ($\sim 20\text{--}25 M_{\odot}$; e.g., Fryer and Kalogera 2001). On the other hand, single stars only become Wolf-Rayet stars if their initial mass is larger than $\sim 25\text{--}35 M_{\odot}$ (depending on the exact mass-loss rate). Since the formation of a slowly rotating black hole is not apriori expected to be associated with a bright supernova (as the whole star can just collapse into a black hole), this has the important implication that all normal H-deficient core-collapse supernovae (SNe Ib/Ic) may require a close binary companion.

7.1.3 Electron-Capture Supernovae in Close Binaries

Another mass range where binary interactions can drastically change the final fate of a massive star is near the minimum mass for stars to explode as supernovae (around $7 M_{\odot}$, where the exact value depends on the amount of convective overshooting and the metallicity of the star). Single stars in this mass range experience a second dredge-up phase when they ascend the asymptotic giant branch (AGB), where a large fraction of the H-exhausted core is dredge-up and mixed with



■ Fig. 14-14

Schematic comparison of the helium-core-burning phase of stars with (*left*) and without a hydrogen envelope (*right*) for massive stars ($M \gtrsim 20 M_{\odot}$). Because of the lack of a H-burning shell, stars without a hydrogen envelope produce smaller He-exhausted cores with a larger C/O fraction. This makes them more likely to ultimately collapse to a neutron star instead of a black hole

the envelope. This reduces the core mass at the end of the AGB phase; as a consequence, single stars as massive as $10/11 M_{\odot}$ probably produce ONeMg white dwarfs rather than a supernova. In contrast, if such stars lose their H-rich envelopes due to a binary interaction before reaching the AGB, they end up with much larger He cores and are likely to produce an electron-capture (e-capture) supernova (see 🔍 Sect. 7.2 for further discussion) (🔍 Fig. 14-15).

To summarize the effects of binary interactions on the final fate of stars, 🔍 Table 14-2 contrasts the expected differences for stars in single systems/wide binaries (which includes case C systems) to stars in close binaries.

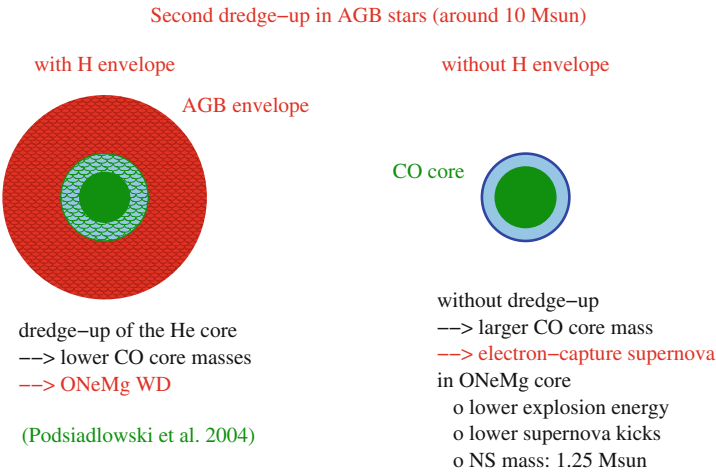
7.2 The Diversity of Supernova Explosions

As has already become clear from the discussion so far, there is more to supernovae than just two standard explosion types. This section summarizes all the main explosion types presently being considered (🔍 Table 14-3).

7.2.1 Neutron Stars

Iron Core Collapse

The collapse of a massive iron core to a neutron star, leading to a neutrino-driven explosion (see 🔍 Sect. 2.1), is the expected standard fate for the majority of massive stars, both single and binary. The typical supernova energy of these explosions is $\sim 10^{44}$ J, as set by the binding energy of the inner part of the ejecta.



■ Fig. 14-15
Schematic comparison of the late evolution of stars with an initial mass around $10 M_{\odot}$ with and without a H-rich envelope. The former experience a so-called second dredge-up phase at the beginning of the AGB phase reducing the size of the helium core; these stars are therefore more likely to end up as ONeMg white dwarfs. On the other hand, stars without the second dredge-up may eventually collapse and experience an *electron-capture supernova*

■ Table 14-2
Final fate of stars: single versus binary

	Single/wide binary	Close binary
CO white dwarf	$<7 M_{\odot}$	$<7-17 M_{\odot}$
ONeMg white dwarf	$7-10 M_{\odot}$	$7-8 M_{\odot}$
Neutron star:		
Electron-capture	$\sim 10 M_{\odot}$	$7/8-10 M_{\odot}$
Iron core collapse	$10-20/25 M_{\odot}$	$10-60 M_{\odot}$
Black hole:		
Two-step	$20/25-40(?) M_{\odot}$	$\gtrsim 60 M_{\odot} (?)$
Prompt	$\gtrsim 40 M_{\odot} (?)$	
No remnant (Z?)	$\gtrsim 140 M_{\odot}$	

Electron-Capture (E-Capture) Supernovae

A neutron star can also form by the collapse of a very degenerate ONeMg core, long before an iron core has developed. This is the expected fate for stars around $\sim 10 M_{\odot}$. In this case, the collapse is triggered by the sudden capture of electrons onto Ne and Mg nuclei taking away the hydrostatic support provided by the degenerate electrons (Nomoto 1984). This occurs at a characteristic density ($\sim 4.5 \times 10^{12} \text{ kg m}^{-3}$; Podsiadlowski et al. 2005), which can be related to a critical pre-collapse mass for the ONeMg core of $\sim 1.37 M_{\odot}$. Since, in this case, most of the core collapses to form a low-mass neutron star (around $1.25 M_{\odot}$), relatively little energy is needed to eject the remainder of the star; the resulting supernova is therefore expected to be of relatively low energy ($\sim 10^{43} \text{ J}$) and may produce a neutron star with a relatively low kick velocity

■ Table 14-3

Summary of explosion types

Neutron Star	Fe core collapse	10^{44} J	Large SN kick
	Electron capture	$\lesssim 10^{43}$ J	Small SN kick
	Magnetar	$\gg 10^{44}$ J	
Black Hole	Prompt	No SN	
	Fallback	Faint SN	
	Collapsar	$\gtrsim 10^{45}$ J	GRB related
	(Engine-driven)		
Thermonuclear	Runaway C burning	10^{44} J	No remnant
	He detonation on WD	$\lesssim 10^{43}$ J	WD remnant (?)
Pair instability	Runaway O burning	$\gg 10^{44}$ J	No remnant

(unlike the case of Fe-core collapse; see ● Sect. 8.6 for further discussion). The supernova that produced the Crab supernova in 1054 may provide an example for this type of explosion (Nomoto et al. 1982).

It should also be noted that an e-capture supernova can also occur in an accreting binary, where an accreting ONeMg white dwarf passes the threshold for the e-capture instability (so-called *accretion-induced collapse* [AIC]), or in the case of the merger of two white dwarfs of various different types (CO+CO, ONeMg+He, ONeMg+CO or even ONeMg+ONeMg [if possible]; so-called *merger-induced collapse* [MIC]). These channels could produce as many as ~10% of all neutron stars.

Magnetars

If the collapse of a massive star produces a rapidly rotating, magnetic neutron star, a *magnetar*, the rapid spin-down of this magnetar could potentially power a very energetic explosion on timescales of days to weeks (this is the neutron-star analogue to the collapsar model). As this would allow the extraction of a large fraction of the rotational energy of a neutron star, this could produce an extremely energetic supernova (see, e.g., Kasen and Bildsten 2010).

7.2.2 Black Holes

Prompt or Fallback Collapse

The formation of a black hole is probably rather unspectacular if no rotation is involved. If a black hole forms promptly, essentially most of the star is expected to collapse into the black hole. If a weak outward going shock is formed initially but does not succeed to eject the whole core, a black hole can form in a delayed manner by fallback of the non-ejected core. In the first case, no supernova is expected, while the latter may produce a faint supernova if some of the envelope can be ejected. As this is the expected fate for massive single stars that form black holes, this implies that as many as 10% of core collapses may not be accompanied by a bright supernova event.

Collapsars and Jet-Driven Explosions

On the other hand, in the collapse of a rapidly rotating core (i.e., in the collapsar model; ● Sect. 6.6.1), a large fraction of the binding energy of the compact object can be

extracted driving the powerful explosions often associated with long-duration GRBs. Even without a GRB, such *jet-driven explosions* could produce very energetic explosions (i.e., hypernovae or broad-lined SNe Ic).

7.2.3 Thermonuclear Explosions

Type Ia Supernovae

The thermonuclear runaway in a CO white dwarf as it approaches the Chandrasekhar mass is the standard model for SNe Ia (🔗 [Sect. 2.2](#)).

He Detonations

If a helium shell detonates in an accreting white dwarf, this can lead to a rather bright explosion and could easily be mistaken for a faint supernova (Nomoto 1982). Indeed, as such detonations produce elements similar to a SN Ia, they may resemble a SN Ia and are therefore now sometimes referred to as SNe .Ia (note the period). The main difference is that the underlying CO white dwarf is likely to survive. On the other hand, if the helium detonation drives a sufficiently strong shock into the CO core, this could trigger a second detonation near the core, a carbon detonation, which would then destroy the whole white dwarf and produce a SN Ia-like event. Such *double-detonation* models are an alternative scenario for SNe Ia; they have the additional advantage that they could also occur in CO white dwarfs below the Chandrasekhar mass, which are statistically much more common (see Woosley and Weaver 1994; Fink et al. 2010).

Pair-Instability Supernovae

There is another type of thermonuclear runaway that can occur in very massive oxygen cores when the core temperature becomes high enough ($\gtrsim 10^9$ K) that electron-positron pairs form abundantly. This leads to the conversion of internal energy into the rest-mass energy of the pairs and reduces the central thermal pressure support. As such cores are radiation-pressure supported and are only marginally stable to start out with, this decreases the average adiabatic index below the critical value of $4/3$, triggering a partial collapse of the core. This accelerates the rate of oxygen burning and leads to a thermonuclear runaway that in most cases is expected to completely destroy the star. This type of supernova requires a very massive helium core after H-core burning ($\gtrsim 60 M_\odot$) which corresponds to an initial main-sequence mass of $\sim 140 M_\odot$ (Heger and Woosley 2002), if there is no mass loss. If the He core is too massive ($\gtrsim 130 M_\odot$ corresponding to a main-sequence mass of $\sim 260 M_\odot$), the photodisintegration of heavy nuclei becomes important. This involves endothermic reactions that completely use up the energy produced by oxygen burning, preventing it from reversing the initial collapse and ultimately leading to a collapse of the whole core to a rather massive black hole.

Even though the physics of this explosion mechanism has been well understood for a long time (Rakavy et al. 1967), pair-instability supernovae have traditionally not been considered important at solar metallicity. Even if such massive progenitor stars were able to form, the wind mass loss on the main sequence would prevent the formation of a sufficiently massive He core. Langer et al. (2007) estimate that the maximum allowed metallicity for a pair-instability supernova is $1/3$ solar. A pair-instability supernova could be extremely bright as several M_\odot (perhaps even more than $10 M_\odot$) of radioactive Ni can be produced in such an event.



At Population III metallicity, these could, in principle, be the dominant source of Fe. However, detailed nucleosynthesis models also predict a strong odd-even effect in the nuclear

abundance pattern, where nuclei with even nuclear charge are much more abundant than those with odd charge (Heger and Woosley 2002). This would produce a nucleosynthetic imprint that is not actually observed. This implies that pair-instability supernovae cannot be so common to make a major contribution to the global nucleosynthesis.

A good candidate for a pair-instability supernova is SN 2007bi (Gal-Yam et al. 2009), a SN Ic with an ejecta mass estimated as $\sim 100 M_{\odot}$ and a Ni mass $\gtrsim 3 M_{\odot}$. It also occurred in a small dwarf galaxy, consistent with a low-metallicity progenitor.

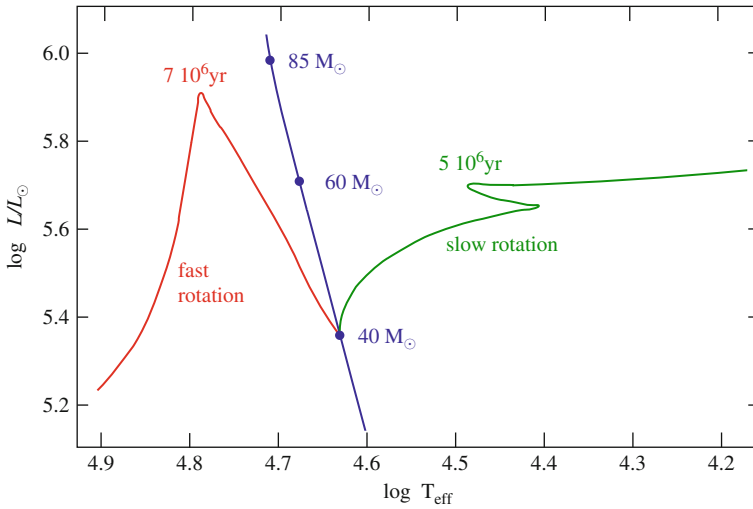
7.3 Rotation and Magnetic Fields

Massive stars tend to be rapid rotators. In order for rotation to be dynamically important, the rotation rate has to be quite close to critical rotation (i.e., close to break-up) which stars rarely achieve except sometimes in their outer layers. This is particularly true for stars in binaries which can be spun up by accretion from a companion star or by merging with a companion. Apart from changing the appearance of a star (e.g., due to rotational flattening causing a nonuniform surface temperature distribution), the main effect of rotation is to cause additional mixing within the star (e.g., due to baroclinic instabilities). This can mix nuclear-processed material to the surface, a tell-tale sign of rapid rotation. For single stars, rotation is most important on the main sequence, as massive stars are very efficient in losing their angular momentum, mainly in the form of stellar winds that carry away mass and angular momentum. In addition, it is now believed that, if a star is differentially rotating, even small seed magnetic fields are amplified causing efficient redistribution of angular momentum inside a star and driving the rotation profile toward solid-body rotation (Heger et al. 2005). This is the reason why the cores of most massive stars are unlikely to be rotating rapidly at the time of core collapse, consistent with the fact that newborn radio pulsars appear not to be rotating near breakup. Therefore, the main effect of rotation in single stars is to increase the final mass of the H-exhausted core at the end of the main sequence. Effectively, after its main-sequence phase, a rapidly rotating star will behave similar to a more massive star that has lost a larger fraction of its envelope (i.e., has a larger ratio of core mass to total mass).

However, if a main-sequence star is rotating extremely rapidly, rotational mixing can be so efficient that it keeps the star almost homogeneously mixed; this will drastically change its evolution as such a star evolves toward the blue and is effectively converted into a Wolf-Ray-type star even without any mass loss (Maeder 1987). This is shown in  Fig. 14-16, which shows the evolution of a $40 M_{\odot}$ with slow and with rapid rotation. Since such homogeneously evolving stars avoid a red-supergiant phase and the strong mass loss experienced in this phase, they may retain a much larger fraction of their initial angular momentum and may still have rapidly rotating cores at the time of the supernova. This makes them potential candidates for LGRBs in the collapsar model (see  Sect. 6.6.1).

7.4 Metallicity

The main effects of metallicity are to change the opacity and the mass-loss rate of a star. A lower metallicity will generally lower the opacity of the star (this is particularly true for low- and intermediate-mass stars), making them more luminous and shortening their lifetimes. One consequence of this is that it changes the thermodynamic structure of the core



■ Fig. 14-16

Hertzsprung–Russell diagram illustrating the bifurcation of the evolution of a $40 M_{\odot}$ with slow and rapid rotation. Because of its almost homogeneous evolution, the rapidly rotating star evolves to the blue and avoids a red-supergiant phase. The *blue curve* shows the zero-age main sequence for massive stars with several masses indicated (Figure adapted from Maeder (1987))

of a star. For example, the core of a $5 M_{\odot}$ AGB star at a typical globular-cluster metallicity (with $Z \simeq 0.001$) is similar to the core of a $7 M_{\odot}$ star at solar metallicity. This implies that the initial-mass limits, shown in ● Table 14-2, have to be shifted accordingly.

Since heavy elements play an important role in driving mass loss from stars, the wind mass-loss rate is a strong function of metallicity (typically $\dot{M} \propto Z^{0.7}$). One implication of this is that lower- Z stars do not only lose less mass but also less angular momentum (see ● Sect. 6.6.1); this makes it more likely that a low- Z star still has a rapidly rotating core at the time of explosion, which could explain why LGRBs appear to favor low- Z environments. Also, because of the lower mass loss, the most massive stars may end with much more massive cores. This makes pair-instability supernovae more likely at low Z as it is mainly determined by the core mass at the beginning of O burning.

Finally, for metal-free stars (so-called population III stars), the whole evolution of a massive star is drastically altered, as, e.g., on the main sequence nuclear burning has to take place via the pp cycle rather than the standard CNO cycle.

7.5 Dynamical Interactions

A significant fraction of stars are found in globular clusters which means that direct dynamical interactions can be important. This may be particularly true for massive stars that are often born in a cluster-type environment. Dynamical interactions (such as tidal captures, direct collisions, and exchange interactions) are likely to enhance certain evolutionary channels (e.g., mergers of degenerate objects) and may even produce new types of stars that cannot be found elsewhere

(not even in binaries). A particular example are runaway collisions at the center of clusters that may produce much more massive stars than can ordinarily be formed; the Pistol star in the Quintuplet cluster near the Galactic center, one of the most massive stars known in our Galaxy, is a potential candidate. The maximum mass that can be built up in this way is a strong function of the mass loss the object experiences between merger events, which will be strongly enhanced. With realistic mass loss prescriptions, it presently seems difficult to form extremely massive stars at least at solar metallicity (see Gleebeek et al. 2009).

8 Current Topics

At the moment, we are in a golden epoch for supernova research as current all-sky surveys, such as the Palomar Transient Factory (PTF), Pan-STARRS, and Skymapper, continue to discover more and more supernova types and subtypes. In this last section, I will discuss a selection of the topics that are considered “hot” at the present time, a list that undoubtedly will grow in the near future.

8.1 The Diversity of SN Ia Progenitors

Some of the recent PTF supernovae have provided some important constraints on the progenitors of fairly normal SNe Ia, strongly suggesting that there is more than one progenitor channel: while PTF 11kx clearly must have occurred in a binary with a hydrogen-rich donor star (most likely a giant; Dilday et al. 2012), such a companion can firmly be ruled out in the case of SN 2011fe/PTF 11kly (Li et al. 2011). But there are even more unusual SNe Ia. The supernova SNLS-03D3bb provided the first example of a new class of SNe Ia where the exploding white dwarf appears to have a mass well in excess of the Chandrasekhar mass (once the mass in all the elements in the ejecta has been added up; Howell et al. 2006). This is possible in principle if the white dwarf is rotating rapidly as this can increase the critical mass up to $2 M_{\odot}$ and above (e.g., Yoon and Langer 2005a). Alternatively, these “super-Chandrasekhar” SNe Ia could be the result of the direct dynamical merger of two rather massive CO white dwarfs.

8.2 Superluminous Supernovae

One of the major recent surprises has been the discovery of a class of unusually luminous supernovae. The first one of these reported was SN 2006gy which had a peak absolute magnitude of -22 , much brighter than a typical SN Ia with a peak magnitude of -19 . There is now a whole class of such supernovae, most of them hydrogen-free, reaching peak magnitudes of almost -23 (see, e.g., Quimby et al. 2011). One early suggestion was that their lightcurves could be powered by the decay of several solar masses of radioactive Ni, as can be produced in a pair-instability supernova (see [Sect. 7.2](#)). However, their late lightcurves do not show the expected exponential decay ruling out this possibility. It should be noted that a very luminous lightcurve does not necessarily imply a very energetic explosion as, in a typical explosion, only a small fraction of the kinetic energy in the explosion is radiated away in the first few years after a supernova.

Indeed, if a supernova is surrounded by a very dense medium or a shell with a mass comparable to the ejecta mass, the ejecta can, in principle, be slowed down on a timescale of less than a year, rapidly converting kinetic energy into thermal energy and radiation, i.e., power the lightcurve by the interaction with the medium. Alternatively, such supernovae could be powered by a long-lived central engine, such as a magnetar (see [Sect. 7.2](#)). These superluminous supernovae, which one should be able to detect up to a large redshift ($z \gtrsim 4$), must be very rare and therefore require very special circumstances. Probably, only 1 in 10^4 massive stars end their lives in a superluminous supernova.

8.3 Supernovae with a Circumstellar Medium

A significant fraction of supernovae (perhaps as many as 6%) appear to explode surrounded by a dense medium, indicating that they ejected a large amount of mass very recently before the explosion. These are generally classified as II_n or II_a supernovae (see [Sect. 3.2](#)) and may include both core-collapse and thermonuclear explosions. More surprisingly, there may even be some evidence that some explode in a luminous-blue variable (LBV) phase (reminiscent of the great outburst of η Carinae; e.g., SN 2005gj (Trundle et al. 2008)). The latter would be very surprising as this is not consistent with the theoretically expected evolution of very massive stars that are expected to encounter the LBV instability near the end or just after the main-sequence phase. At this stage, they are assumed to lose their H-rich envelopes in a series of LBV outbursts and spend most of their remaining evolution (at least several 10^5 year) as Wolf-Rayet stars. If single massive stars were able to explode in an LBV phase, all evolutionary models for massive stars existing today would have to be discarded. However, as most massive stars are in relatively close binaries, there may be a simple alternative solution. As discussed in [Sect. 7.1](#), if a massive star that has left the main sequence before it accretes matter or merges with a massive companion star, its subsequent evolution can be drastically altered. In particular, it will spend most of its helium core-burning phase as a blue supergiant and may only encounter the LBV instability after this phase, when its remaining lifetime will typically be less than a few 10^3 year; such a star could easily still be in an LBV phase at the time of the explosion.

8.4 Supernovae in the Middle of Nowhere

Traditionally, supernova searches have targeted large galaxies introducing some serious observational biases. This has changed somewhat with the recent advent of untargated, all-sky surveys. Supernovae are now also regularly found in small galaxies and – more surprisingly – far away from any galaxy. These supernovae often have quite distinct properties. One example is SN 2005E, a faint Ib supernova in the halo of a nearby galaxy (Perets et al. 2010). This supernova is also very rich in Ca, making it a prototype for a new supernova class. Since the estimated ejecta mass is very small (only $\sim 0.3 M_{\odot}$), it is not even clear whether this is a proper supernova or related to explosive helium burning on a white dwarf (see [Sect. 7.2](#)). Alternatively, it could be a hybrid supernova, part core collapse, part thermonuclear. For example, the merger of an ONeMg white dwarf with a He white dwarf could lead to the collapse the ONeMg core to a neutron star (in an e-capture supernova; [Sect. 7.2](#)) which triggers explosive burning of the

helium envelope which in turn powers the explosion. Even though this would technically be classified as a core-collapse supernovae, the explosion would be driven by the thermonuclear runaway.

A second example is PTF10ops, a subluminal SN Ia, that occurred in the middle of nowhere (Maguire et al. 2011). Even though clearly a SN Ia, it is an unusual one that does not fit the standard SN Ia peak magnitude – lightcurve width relation.

Why are these supernovae so far away from a major galaxy? Just because these channels require “old” progenitors? Or were the explosions delayed because they involved the merger of two white dwarfs where the delay time is determined by the gravitational radiation timescale that drives the system together? One should also consider the possibility that these events occur in globular clusters in the halo of galaxies, which for these supernovae would be undetectable, and that these unusual supernovae are somehow connected to their unusual dynamical environment.

8.5 Faint and Failed Supernovae

Single massive stars ($\gtrsim 20\text{--}25 M_{\odot}$) are expected to form black holes, either promptly or by fallback. These are unlikely to be associated with a bright supernova event. Perhaps in many cases, there is no supernova at all (a “failed” supernova). This means that a star, in most cases probably a Wolf-Rayet star, should suddenly disappear. While it is difficult to detect such “disappearing stars,” with modern large-scale surveys this is not impossible (see, e.g., Kochanek et al. 2008).

8.6 Supernova Kicks

Young, single radio pulsars are observed to have a large space motion relative to their parent populations which is best described by a Maxwellian distribution with a velocity dispersion of 265 km s^{-1} (Hobbs et al. 2005). This implies that these single neutron stars must have received a large kick when they were born in a supernova, which provides an important clue to the supernova mechanism. At present, the most promising explanation for these kicks in the delayed neutrino-driven explosion scenario (▶ Sect. 2.1) is an instability in the accretion shock around a proto-neutron star, the so-called SASI instability (“standing accretion-shock instability”), which causes a wobbling of the core, imparting momentum in the process (see, e.g., Blondin and Mezzacappa 2007; Foglizzo et al. 2007). In order for this instability to be able to grow sufficiently, the delay between the initial formation of the proto-neutron star and the initiation of a “successful” explosion has to be $\gtrsim 500 \text{ ms}$ (i.e., hundreds of dynamical times), consistent with the most promising models of Fe core collapse to date. On the other hand, in the case of an e-capture supernova (▶ Sect. 7.2), where the binding energy of the inner part of the ejecta is very small, the explosion is expected to occur with a much shorter delay (e.g., Kitaura et al. 2006), suggesting that this will produce at best a moderate kick. Since e-capture supernovae are more likely to occur in binary systems, this could explain why neutron stars in close binary sometimes appear to have received a much smaller kick than the majority of their single counterparts (Podsiadlowski et al. 2004b).

References

- Arnett, D. 1982, *ApJ*, 253, 785
- Baade, W., & Zwicky, F. 1934, *Phys. Rev.*, 4676
- Berger, E., et al. 2011, *ApJ*, 743, 204
- Blandford, R. D., & Znajek, R. L. 1977, *Month. Not. R. Astron. Soc.*, 179, 433
- Blondin, J. M., & Mezzacappa, A. 2007, *Nature*, 445, 58
- Brown, G. E., Heger, A., Langer, N., Lee, C.-H., Wellstein, S., & Bethe, H. 2001, *New Astron.*, 6, 457
- Burrows, C. J., et al. 1995, *ApJ*, 452, 680
- Cappellaro, E., & Turatto, M. 1997, in *Thermonuclear Supernovae*, eds. P. Ruiz-Lapuente et al. (Kluwer, Dordrecht), 77
- Cucchiara, A., et al. 2011 *ApJ*, 736, 7
- Detmers, R. G., Langer, N., Podsiadlowski, Ph., & Izzard, R. G. 2008, *A&A*, 484, 831
- Dilday, B., et al. 2012, submitted
- Filippenko, A. V. 1997, *Ann. Rev. Astron. Astrophys.*, 35, 309
- Fink, M., et al. 2010, *A&A*, 514, 53
- Foglizzo, T., Galletti, P., Scheck, L., & Janka, H.-T. 2007, *ApJ*, 654, 1006
- Frail, D. A., et al. 2001, *ApJ*, 562, 55
- Fryer, C. L., & Kalogera, V. 2001, *ApJ*, 554, 548
- Fryer, C. L., et al. 2007, *Publ. Astron. Soc. Pacific*, 119, 1211
- Gal-Yam, A., et al. 2009, *Nature*, 462, 624
- Gamow, G., & Schoenberg, M. 1941, *Phys. Rev.*, 59, 539
- Gleebeek, E., Gaburov, E., de Mink, S. E., Pols, O. R., & Portegies Zwart, S. F. 2009, *A&A*, 497, 255
- Hachinger, S., Mazzali, P. A., Taubenberger, S., Hillebrandt, W., Nomoto, K., & Sauer, D. N. 2012, *arXiv:1201.1506*
- Hachisu, I., Kato, M., Nomoto, K., & Umeda, H. 1999, *ApJ*, 519, 314
- Han, Z., Podsiadlowski, Ph., & Eggleton, P. P. 1995, *Month. Not. R. Astron. Soc.*, 272, 800
- Heger, A., & Woosley, S. E. 2002, *ApJ*, 567, 532
- Heger, A., Woosley, S. E., & Spruit H. 2005, *ApJ*, 626, 350
- Hillebrandt, W., & Meyer, F. 1989, *A&A*, 219, 3
- Hobbs, G., Lorimer, D. R., Lyne, A. G., & Kramer, M. 2005, *Month. Not. R. Astron. Soc.*, 360, 974
- Howell, D. A., et al. 2006, *Nature*, 443, 308
- Hsu, J. J. L. 1991, PhD Thesis (M.I.T.)
- Iben, I., Jr., & Tutukov, A. V. 1984, *ApJ Suppl.*, 54, 335
- Ivanova, N., & Podsiadlowski, Ph., 2003, in *From Twilight to Highlight: The Physics of Supernovae*, eds. W. Hillebrandt & B. Leibundgut (Berlin, Springer), 19
- Iwamoto, K., et al. 1998, *Nature*, 395, 672
- Izzard, R. G., Ramirez-Ruiz, E., & Tout, C. A. 2004, *Month. Not. R. Astron. Soc.*, 348, 1215
- Kasen, D., & Bildsten, L. 2010, *ApJ*, 717, 245
- Kitaura, F. S., Janka, H.-Th., & Hillebrandt, W. 2006, *A&A*, 450, 345
- Klebesadel, R. W., Strong, I. B., Olson, & R. A. 1973, *ApJ*, 182, 85
- Kochanek, C. S., et al. 2008, *ApJ*, 684, 1336
- Langer, N., Norman, C. A., de Koter, A., Vink, J. S., Cantiello, M., & Yoon, S.-C. 2007, *A&A*, 475, 19
- Levesque, E. M., Kewley, L. J., Graham, J. F., & Fruchter, A. S. 2010, *ApJ*, 712, 26
- Li, W., et al. 2011, *Nature*, 480, 348
- Maeder, A. 1987, *A&A*, 178, 159
- Maguire, K., et al. 2011, *Month. Not. R. Astron. Soc.*, 418, 747
- Maund, J. R., Smartt, S. J., Kudritzki, R. P., Podsiadlowski, Ph., & Gilmore, G. F. 2004, *Nature*, 427, 129
- Mazzali, P. A., Nomoto, K., Cappellaro, E., Nakamura, T., Umeda, H., & Iwamoto, K. 2001, *ApJ*, 547, 988
- Meszáros, P., & Rees, M. J. 1993, *ApJ*, 405, 278
- Metzger, B. D., Giannios, D., Thompson, T. A., Bucciantini, N., & Quataert, E. 2011, *Month. Not. R. Astron. Soc.*, 413, 2031
- Morris, T., & Podsiadlowski, Ph. 2007, *Science*, 315, 1103
- Nelemans, G., Yungelson, L. R., Portegies Zwart, S. F., & Verbunt, F. 2001, *A&A*, 365, 491
- Nomoto, K. 1982, *ApJ*, 253, 798
- Nomoto, K. 1984, *ApJ*, 277, 791
- Nomoto, K., & Iben, I., Jr. 1985, *ApJ*, 297, 531
- Nomoto, K., Sugimoto, D., Sparks, W. M., Fesen, R. A., Gull, T. R., & Miyaji, S. 1982, *Nature*, 299, 803
- Paczynski, B. 1976, in *Structure and Evolution in Close Binary Systems*, eds. P. P. Eggleton, S. Mitton, & J. Whelan (Dordrecht, Reidel), 75
- Perets, H. B., et al. 2010, *Nature*, 465, 322
- Perlmutter, S., et al. 1999, *ApJ*, 517, 565
- Phillips, M. M. 1993, *ApJ*, 413, L105
- Podsiadlowski, Ph., & Joss, P. C. 1989, *Nature*, 338, 401
- Podsiadlowski, Ph., Joss, P. C., & Rappaport, S. 1990, *A&A*, 227, L9
- Podsiadlowski, Ph., Joss, P. C., & Hsu, J. J. L. 1992, *ApJ*, 391, 245
- Podsiadlowski, Ph., Langer, N., Poelarends, A. J. T., Rappaport, S., Heger, A., & Pfahl, E. 2004a, *ApJ*, 612, 1044
- Podsiadlowski, Ph., Mazzali, P. A., Nomoto, K., Lazzati, D., & Cappellaro, E. 2004b, *ApJ*, 607, L17

- Podsiadlowski, Ph., Dewi, J. D. M., Lesaffre, P., Miller, J. C., Newton, W. G., & Stone, J. R. 2005, *Month. Not. R. Astron. Soc.*, 361, 1243
- Podsiadlowski, Ph., Ivanova, N., Justham, S., & Rappaport, S. 2010, *Month. Not. R. Astron. Soc.*, 406, 840
- Quimby, R. M., et al. 2011, *Nature*, 474, 487
- Rakavy, G., Shaviv, G., & Zinamon, Z. 1967, *ApJ*, 150, 131
- Riess, A. G., et al. 1998, *Astron. J.*, 116, 1009
- Schawinski, K., et al. 2008, *Science*, 321, 223
- Taylor, P. A., Miller, J. C., & Podsiadlowski, Ph. 2011, *Month. Not. R. Astron. Soc.*, 410, 2385
- Timmes, F. X., Brown, Edward F., Truran, J. W. 2003, *ApJ*, 590, 83
- Toomre, A. 1964, *ApJ*, 139, 1217
- Trundle, C., Kotak, R., Vink, J. S., & Meikle, W. P. S. 2008, *A&A*, 483, L47
- van den Heuvel, E. P. J., Bhattacharya, D., Nomoto, K., & Rappaport, S. 1992, *A&A*, 262, 97
- Webbink, R. F. 1984, *ApJ*, 277, 355
- Whelan, J., & Iben, I., Jr. 1973, *ApJ*, 186, 1007
- Woosley, S. E. 1993, *ApJ*, 405, 273
- Woosley, S. E., & Heger, A. 2006, *ApJ*, 637, 914
- Woosley, S. E., & Weaver, T. A. 1994, *ApJ*, 423, 371
- Yoon, S.-C., & Langer, N. 2004, *A&A*, 419, 623
- Yoon, S.-C., & Langer, N. 2005a, *A&A*, 435, 967
- Yoon, S.-C., & Langer, N. 2005b, *A&A*, 443, 643
- Yoon, S.-C., Langer, N., & Norman, C. 2006, *A&A*, 460, 199
- Yoon, S.-C., Podsiadlowski, Ph., & Rosswog, S. 2007, *Month. Not. R. Astron. Soc.*, 390, 933
- Yungelson, L. R., Livio, M., Tutukov, A. V., & Saffer, R. 1994, *ApJ*, 420, 336

15 Stellar Winds

Stan Owocki

Bartol Research Institute, Department of Physics and Astronomy,
University of Colorado, Newark, DE, USA

1	<i>Introduction and Background</i>	737
2	<i>Observational Diagnostics and Inferred Properties</i>	740
2.1	Solar Corona and Wind	740
2.2	Spectral Signatures of Dense Winds from Hot and Cool Stars	742
2.2.1	Opacity and Optical Depth	742
2.2.2	Doppler-Shifted Line Absorption	744
2.2.3	Asymmetric P-Cygni Profiles from Scattering Lines	745
2.2.4	Wind-Emission Lines	748
2.2.5	Continuum Emission in Radio and Infrared	750
3	<i>General Equations and Formalism for Stellar Wind Mass Loss</i>	751
3.1	Hydrostatic Equilibrium in the Atmospheric Base of Any Wind	751
3.2	General Flow Conservation Equations	752
3.3	Steady, Spherically Symmetric Wind Expansion	753
3.4	Energy Requirements of a Spherical Wind Outflow	753
4	<i>Coronal Expansion and Solar Wind</i>	754
4.1	Reasons for Hot, Extended Corona	754
4.1.1	Thermal Runaway from Density and Temperature Decline of Line-Cooling	754
4.1.2	Coronal Heating with a Conductive Thermostat	756
4.1.3	Incompatibility of Hot Extended Hydrostatic Corona with ISM Pressure	756
4.2	Isothermal Solar Wind Model	757
4.2.1	Temperature Sensitivity of Mass Loss Rate	759
4.2.2	The Solar Wind as a Thermostat for Coronal Heating	760
4.3	Driving High-Speed Streams	761
4.4	Relation to Winds from other Cool Stars	761
4.4.1	Coronal Winds of Cool Main-Sequence Stars	761
4.4.2	Alfvén Wave Pressure Driving of Cool-Giant Winds	762
4.4.3	Superwind Mass Loss from Asymptotic Giant Branch Stars	763
4.5	Summary for the Solar Wind	764
5	<i>Radiatively Driven Winds from Hot, Massive Stars</i>	765
5.1	Radiative Acceleration	765
5.1.1	Electron Scattering and the Eddington Limit	765
5.1.2	Driving By Doppler-Shifted Resonant Line Scattering	766
5.1.3	Radiative Acceleration from a Single, Isolated Line	767

5.1.4 Sobolev Localization of Line-Force Integrals for a Point Star 767

5.2 The CAK Model for Line-Driven Winds 768

5.2.1 The CAK Line-Ensemble Force 768

5.2.2 CAK Dynamical Solution for Mass Loss Rate and Terminal Speed 769

5.3 Extensions of Idealized CAK Model 771

5.3.1 Finite-Size Star 771

5.3.2 Radial Variations in Ionization 771

5.3.3 Finite Gas Pressure and Sound Speed 772

5.4 Wind Instability and Variability 772

5.5 Effect of Rotation 774

5.6 Summary for Radiatively Driven Massive-Star Winds 776

6 Wolf-Rayet Winds and Multiline Scattering 778

6.1 Example of Multiple Momentum Deposition in a Static Gray Envelope 778

6.2 Multiline Transfer in an Expanding Wind 780

6.3 Wind Momentum–Luminosity Relation for WR Stars 781

6.4 Cumulative Comoving-Frame Redshift from Multiline Scattering 782

6.5 Role of Line Bunches, Gaps, and Core Thermalization 783

6.6 Summary for Wolf-Rayet Winds 785

References 786

Abstract: A “stellar wind” is the continuous, supersonic outflow of matter from the surface layers of a star. Our sun has a solar wind, driven by the gas-pressure expansion of the hot ($T > 10^6$ K) solar corona. It can be studied through direct *in situ* measurement by interplanetary spacecraft; but analogous coronal winds in more distant solar-type stars are so tenuous and transparent that they are difficult to detect directly. Many more luminous stars have winds that are dense enough to be opaque at certain wavelengths of the star’s radiation, making it possible to study their wind outflows remotely through careful interpretation of the observed stellar spectra. Red giant stars show slow, dense winds that may be driven by the pressure from magnetohydrodynamic waves. As stars with initial mass up to $8 M_{\odot}$ evolve toward the Asymptotic Giant Branch (AGB), a combination of stellar pulsations and radiative scattering off dust can culminate in “superwinds” that strip away the entire stellar envelope, leaving behind a hot white dwarf stellar core with less than the Chandrasekhar mass of $\sim 1.4 M_{\odot}$. The winds of hot, luminous, massive stars are driven by *line-scattering* of stellar radiation, but such massive stars can also exhibit superwind episodes, either as Red Supergiants or Luminous Blue Variable stars. The combined wind and superwind mass loss can strip the star’s hydrogen envelope, leaving behind a Wolf-Rayet star composed of the products of earlier nuclear burning via the CNO cycle. In addition to such direct effects on a star’s own evolution, stellar winds can be a substantial source of mass, momentum, and energy to the interstellar medium, blowing open large cavities or “bubbles” in this ISM, seeding it with nuclear processed material, and even helping trigger the formation of new stars, and influencing their eventual fate as white dwarves or core-collapse supernovae. This chapter reviews the properties of such stellar winds, with an emphasis on the various dynamical driving processes and what they imply for key wind parameters like the wind flow speed and mass loss rate.

1 Introduction and Background

The Sun and other stars are commonly characterized by the radiation they emit. But one of the great astronomical discoveries of the latter half of the past century was the realization that nearly all stars lose mass through a more or less continuous surface outflow called a “stellar wind.” While it was long apparent that stars could eject material in dramatic outbursts like novae or supernovae, the concept of continuous mass loss in the relatively quiescent phases of a star’s evolution stems in large part from the direct *in situ* detection by interplanetary spacecraft of a high-speed, supersonic outflow from the sun. For this *solar* wind, the overall rate of mass loss is quite modest, roughly $10^{-14} M_{\odot}/\text{yr}$, which, even if maintained over the sun’s entire main sequence lifetime of ca. 10^{10} year, would imply a cumulative loss of only a quite negligible 0.01% of its initial mass. By contrast, other stars – and indeed even the sun in its later evolutionary stages as a cool giant – can have winds that over time substantially reduce the star’s mass, with important consequences for its evolution and ultimate fate (de Koter et al. 2008; Maeder and Meynet 2000; Leer et al. 1982; Vink 2008; Willson 2008). Moreover, the associated input of mass, momentum, and energy into the interstellar medium can have significant consequences, forming visually striking nebulae and windblown “bubbles” (Castor et al. 1975), and even playing a feedback role in bursts of new star formation that can influence the overall structure and evolution of the parent galaxy (Oey and Clark 2007). In recent years, the general concept of a continuous “wind” has even been extended to describe outflows with a diverse

range of conditions and scales, ranging from stellar accretion disks (Calvet 2004), to Active Galactic Nuclei (Proga 2007), to even whole galaxies (Breitschwerdt and Komossa 2000).

In the half-century since this concept of a wind from the sun or stars took hold, there has amassed a vast literature – consisting of thousands of journal papers, roughly a hundred conference proceedings, dozens of review articles, and even a handful of books – on various aspects of solar and stellar wind mass loss. It is neither feasible nor desirable to attempt any comprehensive survey of this literature, and interested readers looking for entry points beyond the topical coverage here are encouraged to begin with a few key complementary reviews and books. For the solar wind, the original monograph *Interplanetary Dynamics* by Parker (1963) provides a still relevant and fundamental basis, while the recent *Basics of the Solar Wind* by Meyer-Vernet (2007) gives an accessible modern summary. Particularly insightful reviews include those by Leer et al. (1982), Parker (1991), and Cranmer (2009). Proceedings of the regular series of dedicated solar wind conferences, the most recent being *Solar Wind 12* (Maksimovic et al. 2010), also make a good entry point into the evolving solar wind literature. For stellar winds, the text by Lamers and Cassinelli (1999) provides a good general introduction, while more targeted reviews tend to split between those focused on hot, massive stars (Owocki 2001, 2004; Puls et al. 2008) vs. cool, low-mass giants (Holzer and MacGregor 1985; Holzer 1987; Willson 2000; Dupree 2004).

This chapter aims to give a broad physical overview of the properties and processes involved in solar and stellar wind mass loss. A general theme is to identify the *forces* and *energies* that can overcome the gravitational binding that holds material onto a nearly hydrostatic stellar surface, and thereby lift and accelerate the outermost layers into a sustained outflow through which material ultimately escapes entirely the star’s gravitational potential. This competition between outward driving and inward gravity is key to determining two basic wind characteristics, namely the mass loss rate \dot{M} and terminal flow speed v_∞ .

Stellar wind mass loss rates span many decades, reaching more than a billion times that of the solar wind, $\dot{M} > 10^{-5} M_\odot/\text{yr}$, for luminous giant stars; the exact values depend crucially on the specific wind driving mechanism. But for a wide range of driving mechanisms, a quite general rule of thumb is that the wind terminal speed v_∞ scales in proportion to the escape speed v_e from the stellar surface. One intuitive way to think of the process is that in general the outward driving mechanism tends to tune itself to siphon off just enough mass from the star to keep the outward force an order-unity factor above gravity; the net effective “antigravity” means wind material effectively “falls off” the star, reaching a final speed comparable to v_e .

In practice, the proportionality factor can vary from about a third (for red giants) to about factor three (for blue supergiants). Note that in the former case 90% of the energy to drive the wind is expended in lifting the material out of the star’s gravitational potential, with the remaining 10% expended in acceleration to the terminal flow speed; in the latter case, this relative allotment of energy is reversed. The total wind energy loss rate, $L_{\text{wind}} \equiv \dot{M}(v_e^2 + v_\infty^2)/2$, is typically only a tiny fraction of the star’s radiative luminosity, about 10^{-6} for the solar wind, and ranging up to a few percent for winds from luminous, massive stars. During brief “superwind” episodes of massive “Luminous Blue Variable” (LBV) stars, the radiative and wind luminosity can even become comparable.

The specific driving mechanisms vary with the various types of stellar wind, which can be broadly organized into three general classes:

1. *Coronal Winds* from the Sun and Other Cool, Main-Sequence Stars. These can be characterized as thermally or *gas pressure driven*, with the main outward force to overcome gravity stemming from the gas-pressure gradient associated with the high-temperature corona. The

key issues thus lie in understanding how mechanical energy generated in the near-surface convection zone is transmitted upward (e.g., via wave oscillations in the magnetic field) to heat and drive a high-temperature coronal expansion. While important as a prototype, and because the earth itself is embedded in the solar coronal wind, the low \dot{M} means such coronal winds have negligible direct effect on the star's evolution. (On the other hand, the loss of angular momentum in magnetized coronal winds does seem sufficient to cause an evolutionary decline of rotation rates in the sun and other cool stars; Weber and Davis 1967; Scholz 2009).

2. *Winds from Cool Giants and Supergiants.* The lower surface gravity of giant stars facilitates much stronger, but slower wind mass loss. For example, the slow ($v_\infty \sim 10\text{--}50$ km/s), moderately strong ($\dot{M} \sim 10^{-8} M_\odot/\text{yr}$) winds of Red Giants may be driven by the direct momentum (vs. energy) addition of magnetic Alfvén waves. But in later evolutionary stages, stellar pulsations, perhaps augmented by radiative *scattering* on dust, can apparently induce a much stronger, even runaway mass loss. Over the final few thousand years, this can reduce stars with initial mass as high as $8 M_\odot$ to a remnant white dwarf below the Chandrasekhar limit of $\sim 1.4 M_\odot$! Similarly, initially hot stars with $M \approx 8\text{--}40 M_\odot$ can evolve through a Red Supergiant phase, with a strong mass loss that has important consequences for their subsequent evolution and eventual demise as core-collapse supernovae (SN).
3. *Winds from Hot, Massive, Luminous Stars.* The high luminosity of hot, massive stars means that the momentum of the light scattering off the electrons and ions in the atmosphere is by itself able to overcome the gravity and drive winds with $v_\infty \sim (1\text{--}3)v_e \sim 1,000\text{--}3,000$ km/s and $\dot{M} \sim 10^{-10}\text{--}10^{-5} M_\odot/\text{yr}$. As quantified in the so-called CAK (after Castor et al. 1975) formalism, for winds from OB-type stars the essential coupling between radiative momentum and gas is via *line scattering* by heavy, minor ions; regulation of the driving by line saturation leads to relatively simple, analytic scaling laws for the dependence of mass loss rate and flow speed on stellar parameters like luminosity, mass, and radius. In the subclass known as Wolf-Rayet (WR) stars, the mass loss is so high that the stellar “photosphere” lies in the wind itself, characterized now by strong wind-broadened emission lines of helium and abundant heavy elements like carbon, nitrogen, or oxygen (CNO); in evolved WR stars the depletion or absence of hydrogen indicates the mass loss has actually stripped away the star's original hydrogen envelope, exposing a core of material processed by various stages of nuclear burning. This mass loss may also be augmented by Luminous Blue Variable (LBV) phases characterized by eruptions or “superwinds” with mass loss rates up to $0.01\text{--}1 M_\odot/\text{yr}$ lasting several years or decades. The 1840-60 giant eruption of η Carinae provides a key prototype.

Of these, the first and third have the more extensively developed theory, and as such will constitute the major focus of detailed discussion below. For the second class, the reader is referred to previous reviews (Holzer and MacGregor 1985; Holzer 1987; Willson 2000; Dupree 2004) and references therein. (See, however, [Sect. 4.4](#) below.)

The next section ([Sect. 2](#)) summarizes the diagnostic methods and resulting inferred general properties for these various classes of stellar wind. The following section ([Sect. 3](#)) introduces the basic dynamical conservation equations governing wind outflow, which are then applied ([Sect. 4](#)) to the gas-pressure-driven solar wind. The last two sections give a quite extensive discussion of radiatively driven mass loss from hot, massive, luminous stars, including the CAK line-driving of OB winds ([Sect. 5](#)), and the multiline-scattered winds of Wolf-Rayet stars ([Sect. 6](#)).

2 Observational Diagnostics and Inferred Properties

2.1 Solar Corona and Wind

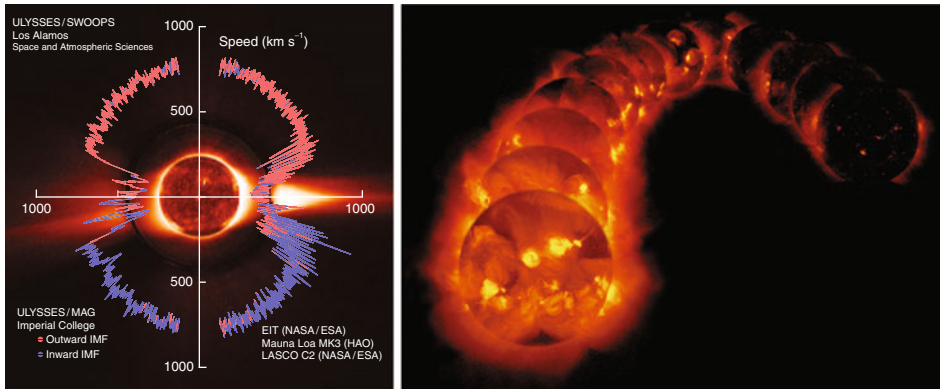
The close proximity of the sun makes it possible to obtain high-resolution observations of the complex magnetic structure in the solar atmosphere and corona, while also flying interplanetary spacecraft to measure in situ the detailed plasma and magnetic field properties of the resulting solar wind. Even before the spacecraft era, early visual evidence that the sun has an outflowing wind came from antisolar deflection of comet tails (Biermann 1951; ► [Fig. 15-1a](#)), and from the radial striations seen in eclipse photographs of the solar corona (Newkirk 1967; ► [Fig. 15-1b](#)). Nowadays, modern *coronagraphs* use an occulting disk to artificially eclipse the bright solar surface, allowing ground- and space-based telescopes to make routine observations of the solar corona.

► [Figure 15-2a](#) (from McComas et al. 1998) combines such coronagraphic images of the inner and outer corona (taken respectively from the SOHO spacecraft, and from the solar observatory on Mauna Loa, Hawaii) with an extreme ultraviolet image of the solar disk (also from SOHO), and a polar line plot of the latitudinal variation of solar wind speed (as measured by the *Ulysses* spacecraft during its initial polar orbit of the sun). The equatorial concentration of coronal brightness stems from confinement of coronal plasma by closed loops of magnetic field, with the outer loops pulled into the radial “helmet” streamers by the outflow of the solar wind. By contrast, the open nature of magnetic field lines over the poles allow material to flow freely outward, making for reduced density *coronal holes* (Zirker 1977). A key point from ► [Fig. 15-2a](#) is that the solar wind arising from these polar coronal holes is much steadier and faster (about 750 km/s) than the equatorial, streamer-belt wind measured near the ecliptic (with speeds from



■ Fig. 15-1

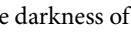
Two telescopic clues to the existence of the solar wind. *Left*: The dual tails of Comet Hale–Bopp. The upper one here consists mostly of *dust* slowly driven away from the comet by solar radiation; it is tilted from the antisun (radial) direction by the comet’s own orbital motion. The lower *plasma* tail comes from cometary ions picked up by the solar wind; its more radial orientation implies that the radial outflow of the solar wind must be substantially faster than the comet’s orbital speed. *Right*: White-light photograph of 1980 solar eclipse, showing how the million-degree solar corona is structured by the solar magnetic field. The closed magnetic loops that trap gas in the inner corona become tapered into pointed “helmet” streamers by the outward expansion of the solar wind. Eclipse image courtesy Rhodes College, Memphis, Tennessee, and High Altitude Observatory (HAO), University Corporation for Atmospheric Research (UCAR), Boulder, Colorado. UCAR is sponsored by the National Science Foundation



■ Fig. 15-2

Left: Polar plot of the latitudinal variation of solar wind speed (*jagged lines*) and magnetic polarity (indicated by *red vs. blue* line color), as measured by the *Ulysses* spacecraft during its initial polar orbit of the sun. This is overlaid on a composite of coronagraphic images of the inner and outer corona, and an extreme ultraviolet image of the solar disk, all made on August 17, 1996, near the minimum of the solar sunspot cycle. **Right:** X-ray images of the sun made by the *Yokohoh* satellite from the 1991 solar maximum (*left foreground*) to the 1996 solar minimum (*right background*)

400 to 750 km/s). This provides a vivid illustration of how the magnetic structure of the corona can have a profound influence on the resulting solar wind.

Early direct evidence for the very high coronal temperature came from eclipse observations of emission lines from highly stripped ions of calcium and iron, the most prominent being the Fe^{+13} “green” line at 5,303Å. This stripping of 13 electrons from iron arises from high-energy collisions with free electrons, implying a characteristic electron temperature around one and half million Kelvin (1.5 MK) in the dense, coronal loops that dominate the line emission. Nowadays the high temperature of these coronal loops is most vividly illustrated by soft X-ray images, such as those shown in  Fig. 15-2b from the observations by the *Yokoh* satellite made over the 1991–1996 declining phase of the sun’s 11-year magnetic activity cycle. Unfortunately, the relative darkness of coronal holes means it is much more difficult to use emission lines and X-rays to infer the temperature or other properties of these important source regions of the solar wind.

Complementary studies with space-borne coronal spectrographs (such as the UVCS instrument on SOHO; Kohl et al. 1995) use the strength and width of ultraviolet *scattering* lines to infer the flow speed and kinetic temperatures of various ion species throughout the corona. Results show the outflow speed in coronal holes already exceeds 100 km/s within a solar radius from the surface; moreover, ion temperatures are even *higher* than for electrons, up to 4 MK for protons, and 100 MK for oxygen! The recent, very comprehensive review of coronal holes by Cranmer (2009) provides an extensive discussion of how these different temperatures may be linked to the detailed mechanisms for heating the corona, e.g., by magnetic waves excited by the complex convective motions of the solar atmosphere.

In situ measurements by *Ulysses* and many other interplanetary spacecraft show that the solar wind mass flux (given by $\rho v r^2$, the product of density, flow speed, and the square of the local heliocentric radius) is remarkably constant, varying less than a factor two in space and time

(Withbroe 1989). If extended over the full sphere around the sun, the corresponding global mass loss rate works out to $\dot{M} \approx 2 - 3 \times 10^{-14} M_{\odot}/\text{year}$. In the ecliptic plane, the typical flow speed is $v_{\infty} \approx 400 \text{ km/s}$, but it extends up to $v_{\infty} \approx 750 \text{ km/s}$ in both ecliptic and polar high-speed streams that are thought to originate from coronal holes.

Finally, while many other cool, solar-type stars show clear signatures of magnetic activity and hot, X-ray emitting coronae, any associated coronal type winds are simply too tenuous and transparent to be detected directly by any absorption signatures in the spatially unresolved stellar flux. In a handful of stars, there is evidence for an “astrosphere” (Wood 2004), the stellar analog of the bullet-shaped heliospheric cavity carved out by the solar wind as the sun moves through the local interstellar medium. A wall of neutral hydrogen that forms around the nose of this bullet can lead to a detectable hydrogen line absorption in the observed spectrum of stars positioned so that their light passes through this wall on its way to earth. But apart from such indirect evidence, most of what is inferred about coronal winds from other stars is based on analogy or extrapolation of conditions from the much more detailed measurements of the sun’s coronal wind.

2.2 Spectral Signatures of Dense Winds from Hot and Cool Stars

2.2.1 Opacity and Optical Depth

The much higher mass loss rates in stellar winds from both hot and cool giants make them dense enough to become opaque in certain spectral bands, imparting detectable observational signatures in the star’s flux spectrum that can be used to infer key wind conditions. The coupling of wind material to stellar light can take various forms (Mihalas 1978), including:

- Free-electron (Thomson) scattering
- Free-free absorption and emission by electrons near ions
- Bound-free (ionization) absorption and free-bound (recombination) emission by atoms or molecules
- Bound-bound (line) scattering, absorption, and/or emission by atoms or molecules
- Scattering, absorption, and emission by dust grains

In parsing this list, it is helpful to focus on some key distinctions. *Scattering* involves a change in direction of light without true absorption of its energy, while *emission* creates new light, e.g., from the energy within the gas. *Bound-bound* processes lead to narrow wavelength features in the spectrum, known as (absorption or emission) lines. Most other processes involve the broad spectral *continuum* of stellar radiation, albeit with bound-free or free-bound *edges* at the wavelength associated with the energies for ionization/recombination of the ion, or broad variations due to wavelength dependence of the cross section, e.g., for dust.

In the context of determining stellar wind mass loss rates, a particularly key distinction lies between *single-body* processes (e.g., electron scattering, bound-free absorption, or dust scattering and absorption), and *two-body* processes involving both an ion and a free electron (e.g., free-free absorption, and free-free or free-bound emission). The strength of the former scales in proportion to the single density of the species involved, whereas the latter depend on the product of the electron and ion density, and so have an overall *density-squared* dependence. In a wind with a high degree of spatial clumping, such density-squared processes will be enhanced over what would occur in a smooth flow by a clumping factor $C_f \equiv \sqrt{\langle \rho^2 \rangle} / \langle \rho \rangle$,

where the square brackets represent volume averaging on a scale large compared to the clump size. For a wind in which material is mostly in clumps filling only a fraction f_v of the total volume, any mass loss rate diagnostic that scales with ρ^2 will tend to *overestimate* the true mass loss rate by about a factor $C_f = 1/\sqrt{f_v}$. Further details can be found in the proceedings of a recent conference on wind clumping (Hamann et al. 2008).

Single-density continuum processes (e.g., electron scattering, bound-free absorption, or dust scattering and absorption) can be characterized by the opacity, κ , which is effectively a cross section per unit mass, with CGS units cm^2/g . The overall effectiveness of continuum absorption or scattering by the wind depends on the associated *optical depth* from some surface radius R to the observer. Along the radial direction r , this is given by the integral of the wind opacity κ times density $\rho(r)$,

$$\tau_* = \int_R^\infty \kappa \rho(r) dr \approx \frac{\kappa \dot{M}}{4\pi v_\infty R} = 2.5 \frac{\kappa R_\odot}{\kappa_e R} \frac{\dot{M}_{-5}}{v_{1000}}. \quad (15.1)$$

The second, approximate equality here applies to the simple case of a steady, spherically symmetric wind with mass loss rate \dot{M} , and approximated for simplicity to flow at a constant speed v_∞ . The last equality uses the shorthand $\dot{M}_{-5} = \dot{M}/(10^{-5} M_\odot/\text{yr})$ and $v_{1000} = v_\infty/(1,000 \text{ km/s})$ to give numerical scalings in terms of the opacity for electron scattering, which for a fully ionized medium with standard hydrogen mass fraction, $X = 0.7$, has a CGS value $\kappa_e = 0.2(1+X) = 0.34 \text{ cm}^2/\text{g}$. For example, electron scattering in the solar wind has $\tau_* \approx 10^{-8}$, confirming it is very transparent. But the winds from Wolf-Rayet stars, with a billion times higher \dot{M} , can become optically thick, even to electron scattering.

For the strong, slow winds from cool giants, the fraction of free electrons becomes small, but the cool, dense conditions often allow molecules and even dust to form. Spherical dust of radius a , mass m_d , density ρ_d has a physical cross section $\sigma = 3m_d/4a\rho_d$; conversion of a mass fraction X_d of wind material into dust thus implies an overall opacity $\kappa_d = 3X_d/4a\rho_d$. For standard (solar) abundances, $X_d \approx 2 \times 10^{-3}$, and since $\rho_d \sim 1 \text{ g/cm}^3$ (most dust would almost float in water), it is seen that for dust grains with typical size $a \approx 0.1 \mu$, the corresponding opacity is of order $\kappa_d \approx 100 \text{ cm}^2/\text{g}$, several hundred times greater than for free-electron scattering.

This geometric cross section only applies to wavelengths much smaller than the dust size, $\lambda \ll a$; for $\lambda \geq a$, the associated dust opacity decreases as $\kappa_d(\lambda) \sim (\lambda/a)^{-\beta}$, where $\beta = 4$ for simple Rayleigh scattering from smooth spheres of fixed size a . In practice, the complex mixtures in sizes and shapes of dust typically lead to a smaller effective power index, $\beta \approx 1-2$. Still, the overall inverse dependence on wavelength means that winds that are optically thick to dust absorption and scattering can show a substantially reddened spectrum.

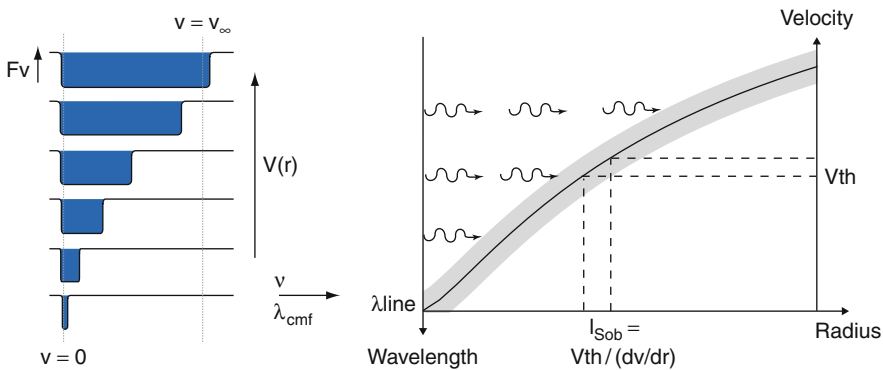
Moreover, the energy from dust-absorbed optical light is generally reemitted in the infrared (IR), at wavelengths set by the dust temperature through roughly the standard Wien's law for peak emission of a blackbody, $\lambda = 2.90 \mu\text{m}/(T/1,000 \text{ K})$. For typical dust temperatures of several hundreds to a thousand degrees, this gives a strong IR excess in emission peaking in the 3–10 μm range. Such IR excess signatures of wind dust are seen in very cool-giant stars, particularly the so-called Carbon stars, but also in the much hotter Carbon-type Wolf-Rayet stars, known as WC stars. A good overview of dust opacity and emission is given by Li (2005).

2.2.2 Doppler-Shifted Line Absorption

It is much more common for winds to become optically thick to *line* absorption or scattering that results from transitions between two bound levels of an atom (including ions), or molecule. Because this is a *resonant* process, the associated cross sections and opacities can be much larger than from free-electron scattering, enhanced by factor that depends on the “quality” of the resonance.¹ For atomic transitions, classical calculations (i.e., without quantum mechanics) show the resonance enhancement factor is given roughly by $q = A\lambda_o/r_e$, where A is the mass abundance fraction of the atom, λ_o is the photon wavelength for the line transition, and $r_e = e^2/m_e c^2 = 2.6 \times 10^{-13}$ cm is the classical radius for electron charge e and mass m_e (Gayley 1995). For ultraviolet line transitions with $\lambda_o \approx 250$ nm and from abundant ions with $A \approx 10^{-5}$, we find $q \approx 10^3$.

As measured in the *rest frame* of the ion, this line opacity is confined to a very narrow range around the resonant wavelength λ_o . But the Doppler shift from the ion’s random thermal motion gives lines a finite Doppler width $\Delta\lambda_D = \lambda_o v_{th}/c$, where for ions with atomic mass m_i and temperature T , the thermal speed $v_{th} = \sqrt{kT/m_i}$. For typical observed ions of carbon, nitrogen, and oxygen (CNO) in a hot-star wind with temperatures a few times 10^4 K, we find $v_{th} \approx 10$ km/s.

In a stellar wind with flow speed $v \approx 1,000$ km/s $\gg v_{th}$, the Doppler shift from the wind flow completely dominates this thermal broadening. As illustrated in ► Fig. 15-3, a stellar photon of wavelength $\lambda < \lambda_o - \Delta\lambda_D$ emitted radially from the star thus propagates freely through the wind until reaching a location where the wind speed v has shifted the line into resonance with the



► Fig. 15-3

Two perspectives for the Doppler-shifted line resonance in an accelerating flow. *Right*: Photons with a wavelength just shortward of a line propagate freely from the stellar surface up to a layer where the wind outflow Doppler shifts the line into a resonance over a narrow width (Represented here by the shading) equal to the Sobolev length, set by the ratio of thermal speed to velocity gradient, $l_{Sob} \equiv v_{th}/(dv/dr)$. *Left*: Seen from successively larger radii within the accelerating wind, the Doppler shift sweeps out an increasingly broadened line-absorption trough blueward of line center in the stellar spectrum

¹The effect is somewhat analogous to blowing into a whistle vs. just into open air. Like the sound of whistle, the response occurs at a well-tuned frequency, and has a greatly enhanced strength.

photon,

$$\lambda = \lambda'_o \equiv \lambda_o(1 - v/c). \quad (15.2)$$

The width of this resonance is given by $l_{Sob} = v_{th}/(dv/dr)$, known as the Sobolev length, after the Russian astrophysicist V. V. Sobolev, who first developed the theory for line radiation transport in such high-speed outflows (Sobolev 1957, 1960). For a wind of terminal speed v_∞ from a star of radius R , a characteristic velocity gradient is $dv/dr \approx v_\infty/R$, implying $l_{Sob}/R \approx v_{th}/v_\infty \approx 0.01 \ll 1$, and so showing this Sobolev line resonance is indeed quite sharp and narrow.

The *Sobolev approximation* assumes a narrow-line limit $v_{th} \rightarrow 0$ to derive localized, analytic solutions for the line transport, using the fact that the opacity then becomes nearly a δ -function about the local resonant wavelength, $\kappa(\lambda) = q\kappa_e \delta(\lambda/\lambda_o - 1 + v/c)$. For wavelengths from line center λ_o to a blue-edge $\lambda_\infty \equiv \lambda_o(1 - v_\infty/c)$, the associated wind optical depth then becomes

$$\tau(\lambda) = \int_R^\infty \kappa(\lambda) \rho dr = \left[\frac{q\kappa_e \rho c}{dv/dr} \right]_{res} \equiv \tau_{Sob}, \quad (15.3)$$

where in the second equality the wind quantities are to be evaluated at the relevant resonance radius where $v(r_{res}) \equiv c(1 - \lambda/\lambda_o)$.

As a simple explicit example, for a steady wind with a canonical wind velocity law $v(r) = v_\infty(1 - R/r)$, we find

$$\tau(\lambda) = \frac{\tau_o}{1 - R/r_{res}} = \frac{\tau_o v_\infty}{c(1 - \lambda/\lambda_o)}, \quad (15.4)$$

where

$$\tau_o \equiv \frac{q\kappa_e \dot{M}c}{4\pi R v_\infty^2} = \frac{qc}{v_\infty} \tau_{*,e}. \quad (15.5)$$

The latter equality in (15.5) emphasizes that this *line* optical depth is a factor qc/v_∞ larger than the *continuum* optical depth for electron scattering, as given by (15.1). For strong lines ($q \approx 1,000$), and even for moderately fast winds with $v_\infty \approx 1,000$ km/s, this enhancement factor can be several *millions*. For winds with $v_\infty \approx 1,000$ km/s and $R \approx 10^{12}$ cm, such strong lines would be optically thick for mass loss rates as low as $\dot{M} \sim 10^{-10} M_\odot/\text{yr}$; for dense winds with $\dot{M} \sim 10^{-6} M_\odot/\text{yr}$, strong lines can be *very* optically thick, with $\tau_o \sim 10^4$!

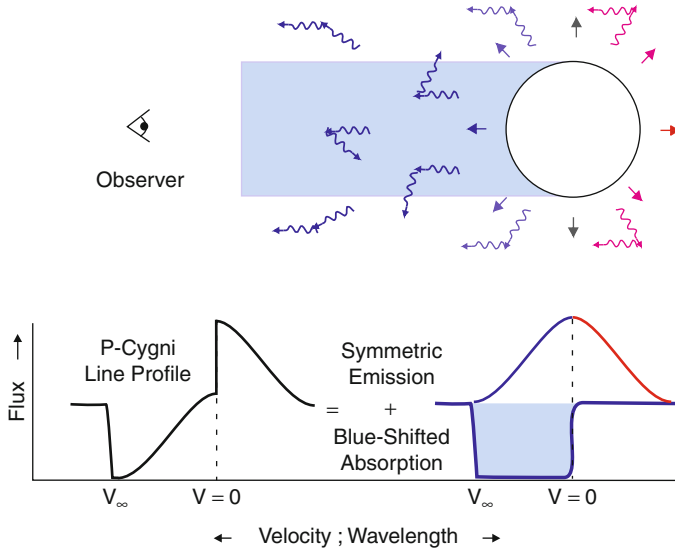
Defining a wind-scaled wavelength displacement from line center, $x \equiv (c/v_\infty)(\lambda/\lambda_o - 1)$, (15.4) gives the very simple wavelength scaling, $\tau(x) = \tau_o/(-x)$. For a simple “point-source” model (which ignores the finite size of the star and so assumes that all radiation is radially streaming), this predicts the observed stellar flux spectrum should show a blueward line-absorption trough of the form,

$$F(x) = F_c e^{-\tau_o/|x|}; \quad -1 < x < 0, \quad (15.6)$$

where F_c is the stellar continuum flux far away from the line. (For the somewhat more general velocity law $v(r) = v_\infty(1 - R/r)^\beta$, the optical depth scaling becomes $\tau(x) = \tau_o/\beta|x|^{2-1/\beta}$.)

2.2.3 Asymmetric P-Cygni Profiles from Scattering Lines

Strong spectral lines often tend to *scatter* rather than fully absorb stellar radiation, and this leads to the emergent line profile developing a very distinctive form known as a “P-Cygni profile,”



■ Fig. 15-4

Schematic illustration of the formation of a P-Cygni type line profile in an expanding wind outflow

after the star P Cygni for which its significance as a signature of mass outflow was first broadly recognized.

► *Figure 15-4* illustrates the basic principles for formation of such a P-Cygni line profile. Wind material approaching an observer within a column in front of the star has its line resonance blueshifted by the Doppler effect. Thus scattering of stellar radiation out of this direction causes an absorption trough on the blue side of the line profile, much as described in the simple, pure-absorption analysis given above. However, from the lobes on either side of this absorption column, wind material can also scatter radiation *toward* the observer. Since this can occur from either the approaching or receding hemisphere, this scattered radiation can be either blueshifted or redshifted. The associated extra flux seen by the observer thus occurs as a symmetric emission component on both sides of the line center. Combined with the reduced blue-side flux, the overall profile has a distinctly asymmetric form, with apparent net blueward absorption and redward emission.

Derivations of the full equations are given, e.g., in Mihalas (1978) and Lamers and Cassinelli (1999), but a brief outline of the quantitative derivation is as follows.

First, the above point-star computation of the absorption trough from stellar light scattered out of the line of sight to the observer must now be generalized to account for rays parallel to the radial ray to the observer, but offset by a distance p , with $0 \leq p \leq R$. Denoting z as the coordinate along such rays, with $z = 0$ at the midplane where the rays come closest the star's center, then the optical depth from the star (at $z_* = \sqrt{R^2 - p^2}$) to the observer (at $z \rightarrow +\infty$) is (cf. [15.3])

$$\tau(p, \lambda) = \int_{z_*}^{\infty} \kappa(\lambda) \rho dz = \left[\frac{q \kappa_e \rho c}{dv_z/dz} \right]_{res}. \quad (15.7)$$

The resonance condition is now $v_z(z_{res}) \equiv c(1 - \lambda/\lambda_o) = -v_\infty x$, with the ray projection of the radial wind speed given by $v_z = \mu v$, where $\mu = z/\sqrt{z^2 + p^2}$. Written in terms of the wind-scaled wavelength $x \equiv (c/v_\infty)(\lambda/\lambda_o - 1)$, the observed *direct* intensity is thus attenuated from the stellar surface value,

$$I_{dir}(p, x) = I_c e^{-\tau(p, x)}, \quad (15.8)$$

which upon integration over the stellar core gives for the observed direct flux (cf. [15.6])

$$F_{dir}(x) = F_c \int_0^R e^{-\tau(p, x)} \frac{2p dp}{R^2}. \quad (15.9)$$

For the scattered, emission component, the formal solution of radiative transfer (Mihalas 1978) takes a purely local form within the Sobolev approximation,

$$I_{scat}(p, x) = S \left(1 - e^{-\tau(p, x)}\right), \quad (15.10)$$

where the source function S for pure-scattering is given by the wavelength- and angle-averaged mean intensity, customarily denoted \bar{J} . In static atmospheres, computing this source function requires a *global* solution of the line scattering, but in an expanding wind, methods developed by Sobolev (1960) allow it to be computed in terms of purely *local* “escape probabilities.” Overall, the radial scaling follows roughly the optically thin form for diluted core radiation $S_{thin} = WI_c$, where the dilution factor $W \equiv (1 - \mu_*)/2$, with $\mu_* \equiv \sqrt{1 - R^2/r^2}$. The observed scattered flux is then computed from integration over all suitable rays,

$$F_{scat}(x) = \frac{2\pi}{r^2} \int_{p_{min}}^\infty I_{scat}(p, x) p dp, \quad (15.11)$$

where $p_{min} = R$ for red-side wavelengths ($\lambda > \lambda_o$) that scatter from the back hemisphere that is occulted by the star, and $p_{min} = 0$ otherwise.

The total observed flux, $F(x) = F_{dir}(x) + F_{scat}(x)$, combines the blueshifted absorption of the direct component plus the symmetric emission of the scattered component, yielding then the net asymmetry of the P-Cygni profile shown in [15.4] and [15.5].

The quantitative fitting of observed line profiles often uses the so-called SEI (Sobolev with Exact Integration) method (Lamers et al. 1987; Groenewegen and Lamers 1989), in which the Sobolev approximation is only used for computation of the scattering source function, which as a wavelength- and angle-averaged quantity tends to cancel errors associated with the finite width of the line resonance. However, because *microturbulence* in a stellar wind can effectively broaden that resonance to many (factor ten or more) times the thermal value $v_{th} \approx 10$ km/s, an exact integration over depth is used for the formal solution of line transfer.

Measurements of the blue-edge of a strong absorption trough immediately gives the wind terminal speed. Moreover, if the line is detectable but not saturated (i.e., $0.1 < \tau_o < 3$), then the observed depth of the trough can in principal be used to infer \dot{M} , assuming one knows the line opacity. In practice, uncertainties in the ionization fraction of prominent ions, which feed directly into the opacity strength factor q , often limit the utility of this method.

But recent *FUSE* observations of UV lines from P^{+4} (a k a PV) have proven particularly useful for inferring mass loss rates of O stars (Fullerton et al. 2006). Because PV is often nearly the dominant ionization stage, the ionization correction is less uncertain. Moreover, the low elemental abundance of phosphorus means that the lines, unlike those from more abundant elements, are not generally saturated. Finally, in contrast to density-squared diagnostics, line

scattering is a single-density process, and so is not affected by the degree of wind clumping. Indeed, mass loss rates inferred by this method tend to be substantially lower, by a factor of 10 or more, compared to those inferred from the density-squared diagnostics discussed in the next subsections (Fullerton et al. 2006). The resolution of this mass loss discrepancy is matter of much current research and discussion (see, e.g., Puls et al. 2006 and Hamann et al. 2008).

2.2.4 Wind-Emission Lines

In WR stars with very dense winds, the emergent spectrum is characterized by very broad ($>1,000$ km/s Doppler width) emission lines from various ionization stages of abundant elements like He, N, C, and O. The recombination of an electron and ion generally occurs through a downward cascade through the many bound levels of the ion, giving the bound-bound line emission. The involvement of an ion plus free electron means this is one of the processes that scale with the *square* of the gas density. In the narrow-line (Sobolev) approximation, the *line emissivity* (energy per unit time, volume, and solid angle) at wind-scaled wavelength x and from some wind radius with velocity $v(r)$ is proportional to

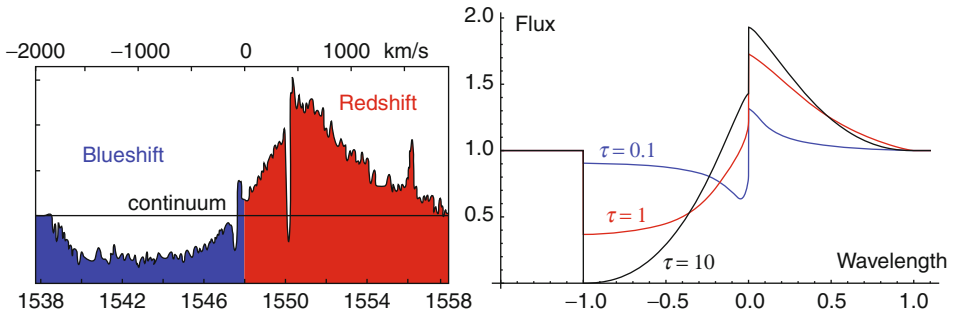
$$\eta(x, r, \mu) \propto \rho^2 \delta(x + \mu v(r)/v_\infty), \quad (15.12)$$

where μ is the direction cosine for the emitted radiation relative to the local radial direction. Integration over all directions μ then gives for the emission luminosity contribution from a narrow shell between r and $r + dr$,

$$\frac{dL(x)}{dr} \propto r^2 \rho^2 \int_{-1}^{+1} d\mu \delta(x + \mu v(r)/v_\infty) \quad (15.13)$$

$$\propto r^2 \rho^2 \frac{v_\infty}{v(r)} \quad \text{if } |x| \leq \frac{v(r)}{v_\infty} \quad (15.14)$$

$$= 0 \quad \text{otherwise.} \quad (15.15)$$



■ Fig. 15-5

Left: Observed P-Cygni line profile for C⁺3 from NGC 6543, the central star in the “Cat’s Eye” nebula. **Right:** Theoretical P-Cygni profile for a wind with velocity law $v(r) = v_\infty (1 - R/r)$, plotted as flux vs. wind-scaled wavelength $x = (c/v_\infty)(\lambda/\lambda_o - 1)$; the three curves show results for a weak, moderate, and strong line, with characteristic line strengths (see (15.5)) $\tau_o = 0.1, 1$, and 10

Thus each such shell produces a *flat-top* profile extending to $x = \pm v/v_\infty$ about line center. If the ionization stage to produce the emission does not form until near the wind terminal velocity, then the observed profile is indeed flat topped, with a half-width given by v_∞ .

But if the ion already forms at a lower radius with a velocity $v_{min} < v_\infty$, then the profile exhibits a flat-topped center for $|x| < v_{min}/v_\infty$, with tapered wings extending to $x = \pm 1$ (see, e.g., Dessart and Owocki 2003). For example, for the case of the standard velocity law $v(r) = v_\infty(1 - R/r)$, the emission at $|x| > v_{min}/v_\infty$ depends on the integral,

$$L(x) \propto \int \frac{\rho^2 r^2}{v} dr \propto \int \frac{dr}{v^3 r^2} \propto \int \frac{dx}{x^3} \propto \frac{1}{2x^2}, \quad (15.16)$$

where the second integrand follows from mass continuity, and the third uses the fact that $Rdr/r^2 = dv/v_\infty = -dx$. We thus obtain the simple analytic result that the line wing tapers off as $\log|x|$ from each side of the inner flat top over $|x| < v_{min}/v_\infty$.

Using this simple model, the right panel of **Fig. 15-6** plots a sample flat-top profile for $v_{min}/v_\infty = 0.5$. The left panel of **Fig. 15-6** (taken from Lepine and Moffat 1999) shows a corresponding observed profile, for the CIII line from the star WR 134. The observed half-width of this emission line directly implies a wind speed of nearly 2,000 km/s for this star. Other WR stars show half-widths that imply wind speeds in the range 1,000–3,000 km/s.

Since the emission strength is proportional to $\rho^2 \sim \dot{M}^2$, careful modeling of the expanding wind and atmosphere can in principal enable one to use the observed strength of emission lines to also estimate the stellar mass loss rate. But such estimates depend on the wind velocity law, and, as noted above, on the degree of wind *clumping* (Hamann 1996; Hillier and Miller 1999; Hamann et al. 2003; Hillier 2003; Puls et al. 2006).

For O stars, a key mass loss rate diagnostic is the Balmer ($H\alpha$) line emission from the $n = 3$ to $n = 2$ transition of recombining Hydrogen (Puls et al. 1996). It tends to be dominated by emission very near the star, and so is not as sensitive to the wind terminal speed. As a ρ^2 diagnostic, it is also sensitive to wind clumping, but in this case, in the relatively low-speed flow near the stellar surface.

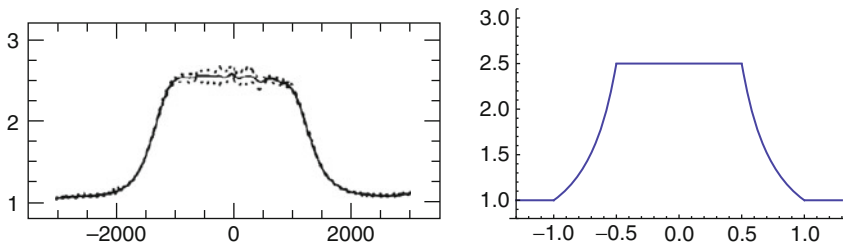


Fig. 15-6

Left: Observed flat-top line profiles of twice ionized carbon CIII ($\lambda 5,696 \text{ \AA}$) from the Wolf-Rayet star WR 137 (From Lepine and Moffat 1999), plotted vs. velocity-unit wavelength shift from line center. **Right:** Simple analytic flat-top emission line profile plotted vs. wind-scaled wavelength $x = (c/v_\infty)(\lambda/\lambda_o - 1)$; the model assumes a velocity law $v(r) = v_\infty(1 - R/r)$ with emission starting at a minimum speed $v_{min} = v_\infty/2$

2.2.5 Continuum Emission in Radio and Infrared

The dense, warm stellar winds from hot supergiants can also produce substantial continuum emission at radio and infrared (IR) wavelengths. Those with spectra showing a flux that increases with frequency are identified as “non-thermal,” and are now associated with binary systems wherein electrons accelerated in the shock collision between the two winds emit radio waves through gyration about the local magnetic field (van Loo et al. 2005).

More useful as a wind diagnostic are the *thermal* emitters, wherein the radio/IR arises from free-free emission by electrons with thermal energies characteristic of the wind temperature. A landmark analysis by Wright and Barlow (1975) showed how the observed flux of thermal radio/IR emission could be used as a quite direct diagnostic of the wind mass loss rate.

Both free-free emission and absorption are density-squared processes. For absorption in a wind moving at constant speed v_∞ , the radial optical depth from a radius r to the observer is thus given by

$$\tau_\nu(r) = K(\nu, T) \int_r^\infty \rho^2(r') dr' = K(\nu, T) \frac{\dot{M}^2}{48\pi^2 v_\infty^2 r^3}, \quad (15.17)$$

where the dependence on frequency ν and temperature T is set by $K(\nu, T) \sim \nu^{-2} T^{-3/2}$ (Allen 1973; Wright and Barlow 1975). Solving $\tau_\nu(R_\nu) = 1$ allows us to define a characteristic radius for the “radio/IR photosphere” at frequency ν ,

$$R_\nu = \left[K(\nu, T) \frac{\dot{M}^2}{48\pi^2 v_\infty^2} \right]^{1/3}. \quad (15.18)$$

For hot supergiants, the wind density is so high that this radio/IR photosphere can be at *hundreds* of stellar radii, $R_\nu > 100 R$.

The radio/IR luminosity can then be estimated by a simple model that accumulates the total outward emission from a sphere with this photospheric radius; since free-free emission is a thermal process, the surface brightness of this photosphere is set by the Planck function $B_\nu(T)$, giving

$$L_\nu = 2\pi B_\nu 4\pi R_\nu^2 = 8\pi^2 B_\nu \left[K(\nu, T) \frac{\dot{M}^2}{48\pi^2 v_\infty^2} \right]^{2/3} \quad (15.19)$$

For radio/IR frequencies and typical stellar wind temperatures, the Planck function is well approximated by its Rayleigh–Jeans form, $B_\nu \approx 2\nu^2 kT/c^2$. Using the above scaling for $K(\nu, T) \sim \nu^{-2} T^{-3/2}$, it is then seen that the radio/IR luminosity scales with frequency as $L_\nu \sim \nu^{2/3}$, but becomes *independent* of temperature.

The luminosity scales with wind parameters as $L_\nu \sim (\dot{M}/v_\infty)^{4/3}$, but otherwise depends just on known atomic constants. As such, if the wind terminal speed is known from other diagnostics, such as P-Cygni line profiles, measurements of the radio/IR flux can provide (for known stellar distance) a relatively robust diagnostic of the wind mass loss rate \dot{M} . But again, because of the scaling with density squared, in a clumped wind such radio/IR inferred mass loss rates are overestimated by a factor $C_f = 1/\sqrt{f_v}$, where now f_v is the clump volume filling factor in the distant, *outer* wind. Comparison with mass loss rates derived from $H\alpha$, which also scales with density squared but is formed near the star, can thus give a clue to the radial evolution of wind clumping (Puls et al. 2006).

3 General Equations and Formalism for Stellar Wind Mass Loss

With this background on wind diagnostics, let us now turn to theoretical modeling of stellar winds. This section lays the groundwork by defining the basic dynamical equations for conservation of mass, momentum, and energy. The following sections review the mechanisms for wind driving for the different types of stellar winds.

3.1 Hydrostatic Equilibrium in the Atmospheric Base of Any Wind

A stellar wind outflow draws mass from the large reservoir of the star at its base. In the star's atmosphere, the mass density ρ becomes so high that the net mass flux density ρv for the overlying steady wind requires only a very slow net drift speed v , much below the local sound speed. In this nearly static region, the gravitational acceleration \mathbf{g}_{grav} acting on the mass density ρ is closely balanced by a pressure gradient ∇P ,

$$\nabla P = \rho \mathbf{g}_{grav}, \quad (15.20)$$

a condition known as *hydrostatic equilibrium*. In luminous stars, this pressure can include significant contributions from the stellar radiation, but for now let us assume (as is applicable for the sun and other cool stars) it is set by the ideal gas law

$$P = \rho k T / \mu = \rho a^2. \quad (15.21)$$

The latter equality introduces the isothermal sound speed, defined by $a \equiv \sqrt{kT/\mu}$, with T the temperature, μ the mean atomic weight, and k Boltzmann's constant. The ratio of (15.21) to (15.20) defines a characteristic pressure scale height,

$$H \equiv \frac{P}{|\nabla P|} = \frac{a^2}{|g_{grav}|}. \quad (15.22)$$

In the simple ideal case of an isothermal atmosphere with constant sound speed a , this represents the scale for exponential stratification of density and pressure with height z

$$\frac{P(z)}{P_*} = \frac{\rho(z)}{\rho_*} = e^{-z/H}, \quad (15.23)$$

where the asterisk subscripts denote values at some surface layer where $z \equiv 0$. In practice, the temperature variations in an atmosphere are gradual enough that quite generally both pressure and density very nearly follow such an exponential stratification.

As a typical example, in the solar photosphere $T \approx 6,000$ K and $\mu \approx 10^{-24}$ g, yielding a sound speed $a \approx 9$ km/s. For the solar surface gravity $g_{grav} = GM_\odot/R_\odot^2 \approx 2.7 \times 10^4$ cm/s², this gives a pressure scale height of $H \approx 300$ km, which is very much less than the solar radius $R_\odot \approx 700,000$ km. This implies a sharp edge to the visible solar photosphere, with the emergent spectrum well described by a planar atmospheric model fixed by just two parameters – typically effective temperature and gravity – and not dependent on the actual solar radius.

This relative smallness of the atmospheric scale height is a key general characteristic of static stellar atmospheres, common to all but the most extremely extended giant stars. In general, for stars with mass M , radius R , and surface temperature T , the ratio of scale height to radius

can be written in terms of the ratio of the associated sound speed a_* to surface escape speed $v_e \equiv \sqrt{2GM/R}$,

$$\frac{H_*}{R} = \frac{2a_*^2}{v_e^2} \equiv 2s. \quad (15.24)$$

One recurring theme of this chapter is that the value of this ratio is also of direct relevance to stellar winds. The parameter $s \equiv (a_*/v_e)^2$ characterizes roughly the ratio between the gas internal energy and the gravitational escape energy. For the solar photosphere, $s \approx 2.2 \times 10^{-4}$, and even for very hot stars with an order of magnitude higher photospheric temperature, this parameter is still quite small, $s \sim 10^{-3}$. However, as discussed further in the next section, for the multimillion-degree temperature of the solar corona, this parameter is much closer to unity, and that is a key factor in the capacity for the thermal gas pressure to drive the outward coronal expansion that is the solar wind.

3.2 General Flow Conservation Equations

Let us now consider a case wherein a nonzero net force leads to a net acceleration,

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{\nabla P}{\rho} - \frac{GM}{r^2} \hat{\mathbf{r}} + \mathbf{g}_x. \quad (15.25)$$

Here \mathbf{v} is the flow velocity, \mathbf{g}_x is some yet-unspecified force-per-unit-mass, and the gravity is now cast in terms of its standard dependence on gravitation constant G , stellar mass M , and local radius r , with $\hat{\mathbf{r}}$ a unit radial vector.

Conservation of mass requires that any temporal change in the local density ρ must arise from a net divergence of the mass flux density $\rho \mathbf{v}$,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{v} = 0. \quad (15.26)$$

Conservation of internal energy takes a similar form, but now accounting also for any local sources or sinks of energy,

$$\frac{\partial e}{\partial t} + \nabla \cdot e \mathbf{v} = -P \nabla \cdot \mathbf{v} - \nabla \cdot \mathbf{F}_c + Q_x. \quad (15.27)$$

For an ideal gas with ratio of specific heats γ , the internal energy density e is related to the pressure through

$$P = \rho a^2 = (\gamma - 1)e. \quad (15.28)$$

On the right-hand-side of (15.27), Q_x represents some still unspecified, net volumetric heating or cooling, while \mathbf{F}_c is the conductive heat flux density, taken classically to depend on the temperature as

$$\mathbf{F}_c = -K_o T^{5/2} \nabla T, \quad (15.29)$$

where for electron conduction the coefficient $K_o = 5.6 \times 10^{-7}$ erg/s/cm/K^{7/2} (Spitzer 1962).

Collectively, (15.25)–(15.28) represent the general equations for a potentially time-dependent, multidimensional flow.

3.3 Steady, Spherically Symmetric Wind Expansion

In application to stellar winds, first-order models are commonly based on the simplifying approximations of steady-state ($\partial/\partial t = 0$), spherically symmetric, radial outflow ($\mathbf{v} = v\hat{r}$). The mass conservation requirement (● 15.26) can then be used to define a constant overall mass loss rate,

$$\dot{M} \equiv 4\pi\rho v r^2. \quad (15.30)$$

Using this and the ideal gas law to eliminate the density in the pressure gradient term then gives for the radial equation of motion

$$\left(1 - \frac{a^2}{v^2}\right) v \frac{dv}{dr} = -\frac{GM}{r^2} + \frac{2a^2}{r} - \frac{da^2}{dr} + g_x. \quad (15.31)$$

In general, evaluation of the sound-speed terms requires simultaneous solution of the corresponding, steady-state form for the flow energy, (● 15.27), using also the ideal gas law (● 15.28). But in practice, this is only of central importance for proper modeling of the pressure-driven expansion of a high-temperature (million-degree) corona, as discussed in the next section for the solar wind.

For stars without such a corona, any wind typically remains near or below the stellar photospheric temperature T , which as noted above implies a sound speed a_* that is much less than the surface escape speed v_e . This in turn implies that the sound-speed terms on the right-hand side of (● 15.31) are quite negligible, of order $s \equiv a_*^2/v_e^2 \sim 10^{-3}$ smaller than the gravity near the stellar surface. These terms are thus of little dynamical importance in determining the overall wind properties, like the mass loss rate or velocity law (see ● Sect. 4). However, it is still often convenient to retain the sound-speed term on the left-hand side, since this allows for a smooth mapping of the wind model onto a hydrostatic atmosphere through a subsonic wind base.

In general, to achieve a supersonic flow with a net outward acceleration, (● 15.31) shows that the net forces on the right-hand side must be positive. For coronal winds, this occurs by the sound-speed terms becoming bigger than gravity. For other stellar winds, overcoming gravity requires the additional body force represented by g_x , for example from radiation (● Sect. 5).

3.4 Energy Requirements of a Spherical Wind Outflow

In addition to these force or momentum conditions for a wind outflow, it is instructive to identify explicitly the general energy requirements. Combining the momentum and internal energy equations for steady, spherical expansion, we can integrate to obtain the *total energy* change from the base stellar radius R to some given radius r

$$\dot{M} \left[\frac{v^2}{2} - \frac{GM}{r} + \frac{\gamma}{\gamma-1} \frac{P}{\rho} \right]_R^r = \int_R^r (\dot{M} g_x + 4\pi r'^2 Q_x) dr' - 4\pi [r'^2 F_c]_R^r. \quad (15.32)$$

On the left-hand side, the terms represent the kinetic energy, gravitational potential energy, and internal enthalpy. To balance this, on the right-hand side are the work from the force g_x , the net volumetric heating Q_x , and the change in the conductive flux F_c . Even for the solar wind, the enthalpy term is generally much smaller than the larger of the kinetic or potential energy. Neglecting this term, evaluation at arbitrarily large radius $r \rightarrow \infty$ thus yields the approximate

requirement for the total energy per unit mass,

$$\frac{v_\infty^2}{2} + \frac{v_e^2}{2} \approx \int_R^\infty (g_x + 4\pi r'^2 Q_x / \dot{M}) dr' + \frac{4\pi R^2 F_{c*}}{\dot{M}}, \quad (15.33)$$

where F_{c*} is the conductive heat flux density at the coronal base, and it is assumed the outer heat flux vanishes far from the star. This equation emphasizes that a key general requirement for a wind is to supply the combined kinetic plus potential energy. For stellar winds, this typically occurs through the direct work from the force g_x . For the solar wind, it occurs through a combination of the volume heating and thermal conduction, as discussed next.

4 Coronal Expansion and Solar Wind

4.1 Reasons for Hot, Extended Corona

4.1.1 Thermal Runaway from Density and Temperature Decline of Line-Cooling

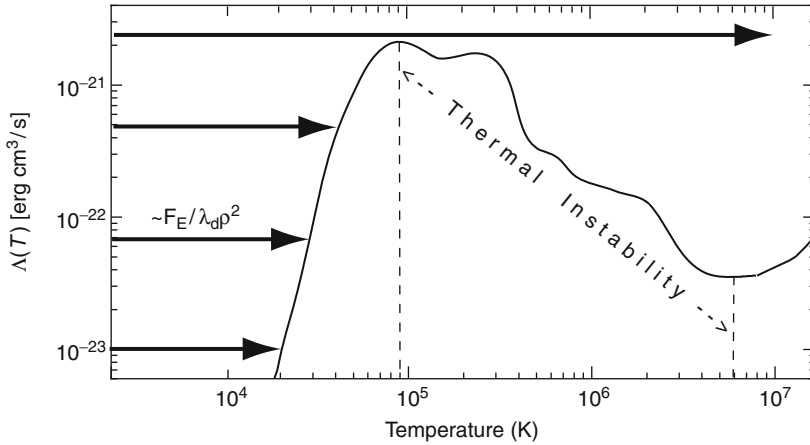
In the sun and other relatively cool stars, the existence of a strong near-surface convection zone provides a source of mechanical energy to heat the upper atmosphere and thereby *reverse* the vertical decrease of temperature of the photosphere. From photospheric values characterized by the effective temperature $T_{\text{eff}} \approx 6,000$ K, the temperature declines to a minimum $T_{\text{min}} \approx 4,000$ K, but then rises through a layer (the chromosphere) extending over many scale heights to about 10,000 K. Above this, it then jumps sharply across a narrow *transition region* of less than scale height, to values in the *corona* of order a *million* degrees! The high pressure associated with this high temperature eventually leads to the outward coronal expansion that is the solar wind.

This sharp jump in temperature is a direct result of the inability of the increasingly rarefied material in the upper atmosphere to radiate away any mechanical heating, for example associated with dissipation of some generic wave energy flux F_E over some damping length λ_d . To maintain a steady state with no net heating, this energy deposition must be balanced by radiative cooling, which can be taken to follow the scaling (e.g., Cox and Tucker 1969; Raymond et al. 1976),

$$\frac{F_E}{\lambda_d} = n_H n_e \Lambda(T). \quad (15.34)$$

Here $\Lambda(T)$ is known as the optically thin cooling function, which can be readily tabulated from a general atomic physics calculation for any assumed abundance of elements, with the standard result for solar (or “cosmic”) abundances plotted schematically in [Fig. 15-7](#). The overall scaling with the product of the hydrogen and electron number densities, n_H and n_e , reflects the fact that the radiative cooling arises from collisional excitation of ions by electrons, for the assumed abundance of ion species per hydrogen atom.

Consider then the nature of this energy balance within a hydrostatically stratified atmosphere, for which all the densities are declining *exponentially* in height z , with a scale height $H \ll R$. If the damping is linked to material absorption, it might scale inversely with density, but even so cooling would still scale with one higher power of density. Using ([15.34](#)), one



■ Fig. 15-7

Radiative cooling function $\Lambda(T)$ plotted vs. temperature T on a log–log scale. The arrows from the left represent the density-scaled rate of energy deposition, which increases with the exponential decrease in density with height. At the point where this scaled heating exceeds the maximum of the cooling function, the radiative equilibrium temperature jumps to a much higher value, in excess of 10^7 K

thus finds the radiative cooling function needs to increase exponentially with height

$$\Lambda[T(z)] \sim \frac{F_E}{\lambda_d \rho^2} \sim F_E e^{+z/H}. \quad (15.35)$$

As illustrated in ● Fig. 15-7, for any finite wave flux this required increase should lead to a steadily higher temperature until, upon reaching the local maximum of the cooling function at a temperature of ca. 10^5 K, a radiative balance can no longer be maintained without a drastic jump to very much higher temperature, above 10^7 K.

This *thermal instability* is the direct consequence of the decline in cooling efficiency above 10^5 K, which itself is an intrinsic property of radiative cooling, resulting from the progressive ionization of those ion stages that have the bound electrons needed for line emission. The characteristic number density at which runaway occurs can be estimated as

$$n_{run} \approx \sqrt{\frac{F_E}{\lambda_d \Lambda_{max}}} \approx 3.8 \times 10^7 \text{ cm}^{-3} \sqrt{\frac{F_5 R_\odot}{\lambda_h}} \quad (15.36)$$

where $\Lambda_{max} \approx 10^{-21} \text{ erg cm}^3/\text{s}$ is the maximum of the cooling function (see ● Fig. 15-7). The latter equality evaluates the scaling in terms of typical solar values for damping length $\lambda_d \approx R_\odot$ and wave energy flux density $F_5 = F_E/10^5 \text{ erg/cm}^2/\text{s}$. This is roughly comparable to the inferred densities at the base of the solar corona, that is, just above the top of the transition region from the chromosphere.

4.1.2 Coronal Heating with a Conductive Thermostat

In practice, the outcome of this temperature runaway of radiative cooling tends to be tempered by conduction of heat back into the cooler, denser atmosphere. Instead of the tens of million degrees needed for purely radiative restabilization, the resulting characteristic coronal temperature is “only” a few million degrees. To see this temperature scaling, consider a simple model in which the upward energy flux F_E through a base radius R is now balanced at each coronal radius r purely by downward conduction,

$$4\pi R^2 F_E = 4\pi r^2 K_o T^{5/2} \frac{dT}{dr}. \quad (15.37)$$

Integration between the base radius R and an assumed energy deposition radius R_d yields a characteristic peak coronal temperature

$$T \approx \left[\frac{7}{2} \frac{F_E}{K_o} \frac{R_d - R}{R_d/R} \right]^{2/7} \approx 2 \times 10^6 K F_5^{2/7}, \quad (15.38)$$

where the latter scaling applies for a solar coronal case with $R = R_\odot$ and $R_d = 2R_\odot$. This is in good general agreement with observational diagnostics of coronal electron temperature, which typically give values near 2 MK.

Actually, observations (Cranmer et al. 1999; Kohl et al. 1999) of the “coronal hole” regions thought to be the source of high-speed solar wind suggest that the temperature of *protons* can be significantly higher, about 4–5 MK. Coronal holes are very low density regions wherein the collisional energy coupling between electrons and protons can be insufficient to maintain a common temperature. For complete decoupling, an analogous conductive model would then require that energy added to the proton component must be balanced by its *own* thermal conduction. But because of the higher mass and thus lower thermal speed, proton conductivity is reduced by the root of the electron/proton mass ratio, $\sqrt{m_e/m_p} \approx 43$, relative to the standard electron value used above. Application of this reduced proton conductivity in (15.38) thus yields a proton temperature scaling

$$T_p \approx 5.8 \times 10^6 K F_5^{2/7}, \quad (15.39)$$

where now $F_5 = F_{Ep}/10^5 \text{ erg/cm}^2/\text{s}$, with F_{Ep} the base energy flux associated with proton heating. Equation 15.39 matches better with the higher inferred proton temperature in coronal holes, but more realistically, modeling the proton energy balance in such regions must also account for the energy losses associated with coronal expansion into the solar wind (see Sect. 3.3).

4.1.3 Incompatibility of Hot Extended Hydrostatic Corona with ISM Pressure

Since the thermal conductivity increases with temperature as $T^{5/2}$, the high characteristic coronal temperature also implies a strong *outward* conduction flux F_c . For a conduction-dominated energy balance, this conductive heat flux has almost zero divergence,

$$\nabla \cdot \mathbf{F}_c = \frac{1}{r^2} \frac{d(r^2 K_o T^{5/2} (dT/dr))}{dr} \approx 0. \quad (15.40)$$

Upon double integration, this gives a temperature that declines only slowly outward from its coronal maximum, that is, as $T \sim r^{-2/7}$. The overall point thus is that, once a coronal base

is heated to a very high temperature, thermal conduction should tend to extend that high temperature outward to quite large radii (Chapman 1961).

This radially extended high temperature of a corona has important implications for the dynamical viability of maintaining a hydrostatic stratification. First, for such a high temperature, (☛ 15.24) shows that the ratio of scale height to radius is no longer very small. For example, for the typical solar coronal temperature of 2 MK, the scale height is about 15% of the solar radius. In considering a possible hydrostatic stratification for the solar corona, it is thus now important to take explicit account of the radial decline in gravity,

$$\frac{d \ln P}{dr} = -\frac{GM}{a^2 r^2}. \quad (15.41)$$

Motivated by the conduction-dominated temperature scaling $T \sim r^{-2/7}$, let us consider a slightly more general model for which the temperature has a power-law radial decline, $T/T_* = a^2/a_*^2 = (r/R)^{-q}$. Integration of (☛ 15.41) then yields

$$\frac{P(r)}{P_*} = \exp \left(\frac{R}{H_*(1-q)} \left[\left(\frac{R}{r} \right)^{1-q} - 1 \right] \right), \quad (15.42)$$

where $H_* \equiv a_*^2 R^2 / GM$. A key difference from the exponential stratification of a nearly planar photosphere (cf. [☛ 15.23]) is that the pressure now approaches a *finite* value at large radii $r \rightarrow \infty$,

$$\frac{P_\infty}{P_*} = e^{-R/H_*(1-q)} = e^{-14/T_6(1-q)}, \quad (15.43)$$

where the latter equality applies for solar parameters, with T_6 the coronal base temperature in units of 10^6 K. This gives $\log(P_*/P_\infty) \approx 6/T_6(1-q)$.

To place this in context, note that a combination of observational diagnostics give $\log(P_{TR}/P_{ISM}) \approx 12$ for the ratio between the pressure in the transition region base of the solar corona and that in the interstellar medium. This implies that a hydrostatic corona could only be contained by the interstellar medium if $(1-q)T_6 < 0.5$. Specifically, for the conduction-dominated temperature index $q = 2/7$, this requires $T_6 < 0.7$. Since this is well below the observational range $T_6 \approx 1.5\text{--}3$, the implication is that a conduction-dominated corona cannot remain hydrostatic, but must undergo a continuous expansion, known of course as the solar wind.

But it is important to emphasize here that this classical and commonly cited argument for the “inevitability” of the solar coronal expansion depends crucially on *extending* a high temperature at the coronal base far outward. For example, a base temperature in the observed range $T_6 = 1.5\text{--}3$ would still allow a hydrostatic match to the interstellar medium pressure if the temperature were to decline with just a somewhat bigger power index $q = 2/3 - 5/6$.

4.2 Isothermal Solar Wind Model

These problems with maintaining a hydrostatic corona motivate consideration of dynamical wind solutions (Parker 1963). A particularly simple example is that of an *isothermal*, steady-state, spherical wind, for which the equation of motion [from (☛ 15.31), without external driving ($g_x = 0$) or sound-speed gradient ($da^2/dr = 0$)] becomes

$$\left(1 - \frac{a^2}{v^2} \right) v \frac{dv}{dr} = \frac{2a^2}{r} - \frac{GM}{r^2}, \quad (15.44)$$

Recall that this uses the ideal gas law for the pressure $P = \rho a^2$, and eliminates the density through the steady-state mass continuity; it thus leaves *unspecified* the constant overall mass loss rate $\dot{M} \equiv 4\pi\rho v r^2$.

The right-hand side of (15.44) has a zero at the critical radius

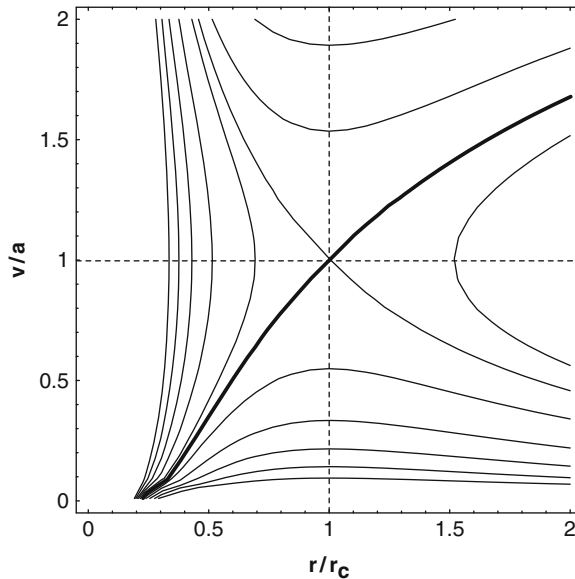
$$r_c = \frac{GM}{2a^2}. \quad (15.45)$$

At this radius, the left-hand side must likewise vanish, either through a zero velocity gradient $dv/dr = 0$, or through a sonic flow speed $v = a$. Direct integration of (15.44) yields the general solution

$$F(r, v) \equiv \frac{v^2}{a^2} - \ln \frac{v^2}{a^2} - 4 \ln \frac{r}{r_c} - \frac{4r_c}{r} = C, \quad (15.46)$$

where C is an integration constant. Using a simple contour plot of $F(r, v)$ in the velocity-radius plane, Fig. 15-8 illustrates the full “solution topology” for an isothermal wind. Note for $C = -3$, two contours cross at the critical radius ($r = r_c$) with a sonic flow speed ($v = a$). The positive slope of these represents the standard solar wind solution, which is the only one that takes a subsonic flow near the surface into a supersonic flow at large radii.

Of the other initially subsonic solutions, those lying above the critical solution fold back and terminate with an infinite slope below the critical radius. Those lying below remain subsonic everywhere, peaking at the critical radius, but then declining to arbitrarily slow, very subsonic



■ Fig. 15-8

Solution topology for an isothermal coronal wind, plotted via contours of the integral solution (15.46) with various integration constants C , as a function of the ratio of flow speed to sound speed v/a , and the radius over critical radius r/r_s . The heavy curve drawn for the contour with $C = -3$ represents the transonic solar wind solution

speeds at large radii. Because such subsonic “breeze” solutions follow a nearly *hydrostatic* stratification (Chamberlain 1961), they again have a large, finite asymptotic pressure that does not match the required interstellar boundary condition.

In contrast, for the solar wind solution the supersonic asymptotic speed means that, for any finite mass flux, the density, and thus the pressure, asymptotically approaches zero. To match a small, but finite interstellar medium pressure, the wind can undergo a shock jump transition onto one of the declining subsonic solutions lying above the decelerating critical solution.

Note that, since the density has scaled out of the controlling equation of motion (☛ 15.44), the wind mass loss rate $\dot{M} \equiv 4\pi\rho v r^2$ does not appear in this isothermal wind solution. An implicit assumption hidden in such an isothermal analysis is that, no matter how large the mass loss rate, there is some source of heating that counters the tendency for the wind to cool with expansion. As discussed below, determining the overall mass loss rate requires a model that specifies the location and overall level of this heating.

4.2.1 Temperature Sensitivity of Mass Loss Rate

This isothermal wind solution does nonetheless have some important implications for the *relative* scaling of the wind mass loss rate. To see this, note again that within the *subsonic* base region, the inertial term on the left side of (☛ 15.44) is relatively small, implying the subsonic stratification is nearly hydrostatic. Thus, neglecting the inertial term v^2/a^2 in the isothermal solution (☛ 15.46) for the critical wind case $C = -3$, one can solve approximately for the *surface* flow speed

$$v_* \equiv v(R) \approx a \left(\frac{r_c}{R} \right)^2 e^{3/2 - 2r_c/R}. \quad (15.47)$$

Here “surface” should really be interpreted to mean at the base the hot corona, that is, just above the chromosphere–corona transition region.

In the case of the sun, observations of transition region emission lines provide a quite tight empirical constraint on the gas *pressure* P_* at this near-surface coronal base of the wind (Withbroe 1988). Using this and the ideal gas law to fix the associated base density $\rho_* = P_*/a^2$, one finds that (☛ 15.47) implies a mass loss scaling

$$\dot{M} \equiv 4\pi\rho_* v_* R^2 \approx 56 \frac{P_*}{a} r_c^2 e^{-2r_c/R} \propto \frac{P_*}{T_6^{5/2}} e^{-14/T_6}. \quad (15.48)$$

With $T_6 \equiv T/10^6$ K, the last proportionality applies for the solar case, and is intended to emphasize the steep, exponential dependence on the inverse temperature. For example, assuming a fixed pressure, just doubling the coronal temperature from one to two million degrees implies nearly a factor 200 increase in the mass loss rate!

Even more impressively, decreasing from such a 1 MK coronal temperature to the *photospheric* temperature $T \approx 6,000$ K would decrease the mass loss rate by more than a *thousand orders of magnitude*! This reiterates quite strongly that thermally driven mass loss is completely untenable at photospheric temperatures.

The underlying reason for this temperature sensitivity stems from the exponential stratification of the subsonic coronal density between the base radius R and the sonic/critical radius r_c . From (☛ 15.24) and (☛ 15.45) it is clear that this critical radius is closely related to the ratio of

the base scale height to stellar radius

$$\frac{r_c}{R} = \frac{R}{2H} = \frac{v_e^2}{4a^2}, \quad (15.49)$$

where the latter equality also recalls the link with the ratio of sound speed to surface escape speed. Application in (15.48) shows that the argument of the exponential factor simply represents the number of base scale heights within a critical radius.

As noted in Sect. 2.1, in situ measurements by interplanetary spacecraft show that the solar wind mass flux is actually quite constant, varying only by about a factor of 2 or so. In conjunction with the predicted scalings like (15.48), and the assumption of fixed based pressure derived from observed transition region emission, this relatively constant mass flux has been viewed as requiring a sensitive *fine-tuning* of the coronal temperature (e.g., Leer and Holzer 1979; Withbroe 1989).

4.2.2 The Solar Wind as a Thermostat for Coronal Heating

But a more appropriate perspective is to view this temperature-sensitive mass loss as providing an effective way to *regulate* the temperature resulting from coronal heating. This can be seen by considering explicitly the energy requirements for a thermally driven coronal wind. From (15.32) (with $g_x = 0$), the total energy change from a base radius R to a given radius r is

$$\dot{M} \left[\frac{v^2}{2} - \frac{v_e^2}{2} \frac{R}{r} + \frac{\gamma a^2}{\gamma - 1} \right]_R^r = 4\pi \int_R^r r'^2 Q_x dr' + 4\pi [R^2 F_{c*} - r^2 F_c]. \quad (15.50)$$

The expression here of the gravitational and internal enthalpy in terms of the associated escape and sound speed v_e and a allows convenient comparison of the relative magnitudes of these with the kinetic energy term, $v^2/2$. Even at typical coronal temperature of 2 MK, the enthalpy term is only about a third of the gravitational escape energy from the solar surface, implying that the initially slow, subsonic flow at the wind base has negative total energy set by gravity. Far from the star, this gravitational term vanishes, and the kinetic energy associated with the supersonic solar wind dominates. Since enthalpy is thus is not important in either limit, let us for convenience ignore its relatively minor role in the global wind energy balance, giving then

$$E_w \equiv \dot{M} \left(\frac{v_\infty^2}{2} + \frac{v_e^2}{2} \right) \approx 4\pi \int_R^\infty r'^2 Q_x dr' + 4\pi R^2 (F_{c*}). \quad (15.51)$$

Here the outer heat flux is assumed to vanish asymptotically far from the star, while $F_{c*} < 0$ is the (inward) conductive heat flux density at the coronal base. Equation 15.51 shows then that the total energy of the solar wind is just given by the total, volume-integrated heating of the solar corona and wind, minus the lost energy by conduction back into the solar atmosphere. The analysis in Sect. 4.1.3 discussed how the heating of a strictly static corona can be balanced by the inward conductive flux back into the sun. But (15.51) now shows that an outflowing corona allows a net difference in heating minus conduction to be carried outward by the solar wind.

In this sense, the problem of “fine-tuning” the coronal temperature to give the observed, relatively steady mass flux is resolved by a simple change of perspective, recognizing instead that for a given level of coronal heating, the strongly temperature-sensitive mass and energy loss of the solar wind provides a very effective “thermostat” for the coronal temperature.

4.3 Driving High-Speed Streams

More quantitative analyses solve for the wind mass loss rate and velocity law in terms of some model for both the level and spatial distribution of energy addition into the corona and solar wind. The specific physical mechanisms for the heating are still a matter of investigation, but one quite crucial question regards the relative fraction of the total base energy flux deposited in the subsonic vs. supersonic portion of the wind. Models with an explicit energy balance generally confirm a close link between mass loss rate and energy addition to the subsonic base of coronal wind expansion. By contrast, in the supersonic region this mass flux is essentially fixed, and so any added energy there tends instead to increase the energy per mass, as reflected in asymptotic flow speed v_∞ (Leer et al. 1982).

A important early class of solar wind models assumed some localized deposition of energy very near the coronal base, with conduction then spreading that energy both downward into the underlying atmosphere and upward into the extended corona. As noted, the former can play a role in regulating the coronal base temperature, while the latter can play a role in maintaining the high coronal temperature needed for wind expansion. Overall, such conduction models of solar wind energy transport were quite successful in reproducing interplanetary measurements of the speed and mass flux of the “quiet,” low-speed ($v_\infty \approx 350\text{--}400$ km/s) solar wind.

However, such models generally fail to explain the high-speed ($v_\infty \approx 700$ km/s) wind streams that are inferred to emanate from solar “corona holes.” Such coronal holes are regions where the solar magnetic field has an open configuration that, in contrast to the closed, nearly static coronal “loops,” allows outward, radial expansion of the coronal gas. To explain the high-speed streams, it seems that some substantial fraction of the mechanical energy propagating upward through coronal holes must not be dissipated as heat near the coronal base, but instead must reach upward into the supersonic wind, where it provides either a direct acceleration (e.g., via a wave pressure that gives a net outward g_x) or heating ($Q_x > 0$) that powers extended gas-pressure acceleration to high speed.

Observations of such coronal hole regions from the SOHO satellite (Cranmer et al. 1999; Kohl et al. 1999) show temperatures of $T_p \approx 4\text{--}5$ MK for the protons, and perhaps as high as 100 MK for minor ion species like oxygen. The fact that such proton/ion temperatures are much higher than the ca. 2 MK inferred for electrons shows clearly that electron heat conduction does not play much role in extending the effect of coronal heating outward in such regions. The fundamental reasons for the differing temperature components are a topic of much current research; one promising model invokes ion cyclotron resonance damping of magnetohydrodynamics waves (Cranmer 2000, 2009).

4.4 Relation to Winds from other Cool Stars

4.4.1 Coronal Winds of Cool Main-Sequence Stars

Like the sun, other cool stars with surface temperatures $T \lesssim 10,000$ low enough for hydrogen recombination are expected to have vigorous subsurface convection zones that generate a strong upward flux of mechanical energy. For *main-sequence* stars in which gravitational stratification implies a sharp, exponential drop in surface density over small-scale height $H \ll R$, this should lead to the same kind of thermal runaway as in the sun (► Sect. 4.1; ► Fig. 15-7), generating a hot corona, and so also an associated, pressure-driven coronal wind expansion. As

noted in ● Sect. 2.1, observations of X-ray emission and high-ionization UV lines do indeed provide strong evidence for such stellar coronae, but the low density of the very low mass loss rate means there are few direct diagnostics of actual outflow in stellar coronal winds.

4.4.2 Alfvén Wave Pressure Driving of Cool-Giant Winds

By contrast, cool *giants* do show much direct evidence for quite dense, cool winds, without any indication of a hot corona that could sustain a gas-pressure-driven expansion. However, the mechanical energy flux density needed to drive such winds is actually quite comparable to what is inferred for the solar surface, $F_E \approx 10^5 \text{ erg/cm}^2/\text{s}$, which likely could again be readily supplied by their H-recombination convection zones. But a key difference now is that, owing to the much weaker gravitational stratification, this need no longer lead to a thermal runaway.

In particular, consider again the general scaling of the ratio of the gravitational scale height to stellar radius (cf. ● 15.24),

$$\frac{H}{R} = \frac{2a^2}{v_e^2} = \frac{kTR}{G\mu M} = 4 \times 10^{-4} \frac{(T/T_\odot)(R/R_\odot)}{M/M_\odot}. \quad (15.52)$$

In cool, solar-mass giant stars, the much larger radius can increase the surface value of this ratio by a factor hundred or more over its (very small) solar value, that is, to several percent. Moreover, at the $>1 \text{ MK}$ coronal temperature of a thermal runaway, the ratio would now be well above unity, instead of ca. 10% found for the solar corona. This would formally put the critical/sonic radius *below* the stellar radius, representing a kind of free expansion of the coronal base, at a supersonic speed well above the surface escape speed, instead of critical transition from subsonic to supersonic outflow. It seems unlikely thermal runaway could be sustained in such an expanding vs. static medium.

But in lieu of runaway heating, it seems instead that the upward mechanical energy flux could readily provide the direct *momentum* addition to drive a relatively slow wind outflow against the much weaker stellar gravity, operating now via a gradient in the pressure associated with *waves* instead of the gas. In general, the mechanical stress or pressure exerted by waves scales, like gas pressure, with the gas density ρ ; but now this is multiplied by the square of a wave velocity amplitude δv , instead of the sound speed a , that is, $P_w = \rho \delta v^2$.

Although the details of how such a wave pressure is generated and evolves depend on the specific type of wave, a general energetic requirement for any waves to be effective in driving a wind outflow is that the fluctuating velocity δv must become comparable to the local gravitational escape speed somewhere near the stellar surface. This already effectively rules out wind driving by sound waves, because to avoid strong damping as shocks, their velocity amplitude must remain subsonic, $\delta v < a$; this implies $P_w < P_{\text{gas}}$, and thus that sound waves would be even less effective than gas pressure in driving such cool winds (Koninx and Pijpers 1992).

In this respect, the prospects are much more favorable for magnetohydrodynamic waves, particularly the Alfvén mode. For the simple case of a radially oriented stellar magnetic field $\mathbf{B} = B\hat{r}$, the mean field can impart no outward force to propel a radial wind. But Alfvén wave fluctuations δB *transverse* to the mean field B give the field a wave pressure $P_w = \delta B^2/8\pi$, for which any radial gradient dP_w/dr does have an associated radial force. Moreover, since the fractional amplitude $f = \delta B/B$ of Alfvén waves can be up to order unity, the associated velocity

amplitude δv of these transverse fluctuations can now extend up to the Alfvén speed,

$$v_A \equiv \frac{B}{\sqrt{4\pi\rho}}. \quad (15.53)$$

Unlike for sound waves, this can readily become comparable to the stellar escape speed v_e , implying a ready potential for Alfvén waves to drive a strong stellar wind.

In fact, for Alfvén waves with a fractional amplitude $f = \delta v/v_A = \delta B/B$, requiring $\delta v = v_e = v$ at a wind critical point near the stellar surface radius R gives a direct estimate for the potential wind mass loss rate,

$$\dot{M} \approx \frac{\delta B^2 R^2}{v_e} = 1.2 \times 10^{-7} \frac{M_\odot}{\text{yr}} \frac{\delta B^2 R_{100}^{5/2}}{\sqrt{M/M_\odot}}, \quad (15.54)$$

where $R_{100} \equiv R/(100R_\odot)$ and $\delta B = fB$ is in Gauss. Indeed, note that for the sun, taking $f = 0.1$ and a mean field of one Gauss gives $\dot{M} = 1.2 \times 10^{-14} M_\odot/\text{year}$, which is somewhat fortuitously consistent with the mass flux measured by interplanetary spacecraft (► Sect. 2.1). More significantly, for giant stars with fluctuating fields $\delta B = fB$ of a few Gauss, the scaling in (► 15.54) matches quite well their much higher, observationally inferred mass loss rates. For a wind with this mass loss scaling and terminal speed comparable to the surface escape $v_\infty \approx v_e$, the energy flux density required is

$$F_E = \frac{\dot{M} v_e^2}{4\pi R^2} \approx 5 \times 10^4 \frac{\text{erg}}{\text{cm}^2 \text{s}} f^2 B^2 \sqrt{\frac{M/M_\odot}{R_{100}}}, \quad (15.55)$$

which, as noted above, for typical cool-giant parameters is quite close to solar energy flux density.

Actually, observations indicate such cool-giant winds typically have terminal speeds of only 10–20 km/s, significantly below even their relatively low surface escape speeds $v_e \approx 50$ km/s, and implying that most all (~90%) of the input energy goes toward lifting the wind material out of the star’s gravitational potential. Explaining this together with the large mass loss rate demands a delicate fine-tuning in any wind model, for example that the Alfvén waves are damped over a length scale very near a stellar radius (Hartmann and MacGregor 1980; Holzer et al. 1983). For further details, the reader is referred to the very lucid reviews by Holzer (1987) and Holzer and MacGregor (1985).

4.4.3 Superwind Mass Loss from Asymptotic Giant Branch Stars

Stars with initial mass $M < 8 M_\odot$ are all thought to end up as white dwarfs with mass below the Chandrasekhar limit $M < 1.4 M_\odot$. This requires copious mass loss ($\dot{M} > 10^{-4} M_\odot/\text{year}$) to occur in a relatively brief ($\sim 10^4$ year) “superwind” phase at the end of a star’s evolution on the Asymptotic Giant Branch (AGB). It seems unlikely that mere wave pressure generated in near-surface convection zones could drive such enormous mass loss, and so instead, models have emphasized tapping the more deep-seated driving associated with global stellar *pulsations*. The most effective seem to be the radial pressure or p-modes thought to be excited by the so-called kappa mechanism (Hansen et al. 2004) within interior ionizations zones (e.g., of Helium). Because of the much higher interior temperature and sound speed, such pulsations can have much higher velocity amplitude than surface sound waves. Indeed, as the pulsations propagate

upward through the declining density, their velocity amplitude can increase, effectively lifting gas parcels near the surface a substantial fraction of the stellar radius.

If the subsequent infall phase of these near-surface parcels meets with the next outward pulsation before returning to its previous minimum radius, the net effect is a gradual lifting of material out of the much weakened gravitational potential. Moreover, as the gas cools through the expansion into ever larger radii and away from the stellar photosphere, dust can form, at which point the additional driving associated with scattering and absorption of the bright stellar radiation can propel the coupled medium of gas and dust to full escape from the star. But such dust-driving is only possible when the pulsations are able to effectively “levitate” stellar material in a much weakened gravity.

As reviewed by Willson (2000), models suggest the net mass loss can eventually become so effective that the associated mass loss timescale $\tau_M = M/\dot{M}$ becomes shorter than the timescale for interior evolution, representing then a kind of “death line” at which the entire stellar envelope is effectively lost, leaving behind only the hot, dense, degenerate stellar core to become a white dwarf remnant. Overall, this process seems arguably a largely interior evolution issue, more akin to a “critical overflow” (e.g., like the “Roche lobe” overflow in mass exchange binaries) than a steady, supersonic outflow from a fixed stellar surface. As such, the kind of simplified steady-state wind formalism outlined here is perhaps of limited relevance. The reader is thus referred to the review by Willson (2000) and references therein for a more detailed discussion of this extreme final stage of cool-star mass loss.

4.5 Summary for the Solar Wind

- The solar wind is driven by the gas-pressure gradient of the high-temperature solar corona.
- The hot corona is the natural consequence of heating from mechanical energy generated in the solar convection zone. At low density radiative cooling cannot balance this heating, leading to a thermal runaway up to temperatures in excess of a million degrees, at which inward thermal conduction back into the atmosphere can again balance the heating.
- Outward thermal conduction can extend this high temperature well away from the coronal base.
- For such a hot, extended corona, hydrostatic stratification would lead to an asymptotic pressure that exceeds the interstellar value, thus requiring a net outward coronal expansion that becomes the supersonic solar wind.
- Because the wind mass flux is a sensitive function of the coronal temperature, the energy loss from wind expansion acts as an effective coronal thermostat.
- While the mass loss rate is thus set by energy deposition in the subsonic wind base, energy added to the supersonic region increases the flow speed.
- The high-speed wind streams that emanate from coronal holes require extended energy deposition. Coronal-hole observations showing the proton temperature significantly exceeds the electron temperature rule out electron heat conduction as the mechanism for providing the extended energy.
- Magnetic fields lead to extensive structure and variability in the solar corona and wind.
- Other cool main-sequence stars likely also have hot, pressure-driven coronal winds, but the dense, cool winds of cool giants could instead be driven by the pressure from Alfvén waves.

5 Radiatively Driven Winds from Hot, Massive Stars

Among the most massive stars – which tend also to be the hottest and most luminous – stellar winds can be very strong, with important consequences for both the star’s own evolution, and for the surrounding interstellar medium. In contrast to the *gas-pressure-driven* solar wind, such hot-star winds are understood to be driven by the pressure gradient of the star’s emitted *radiation*.

The sun is a relatively low-mass, cool star with a surface temperature about 6,000 K; but as discussed above, its wind arises from pressure-expansion of the very hot, million-degree solar corona, which is somehow superheated by the mechanical energy generated from convection in the sun’s subsurface layers. By contrast, high-mass stars with much hotter surface temperatures (10,000–100,000 K) are thought to lack the strong convection zone needed to heat a circum-stellar corona; their stellar winds thus remain at temperatures comparable to the star’s surface, and so lack the very high gas pressure needed to drive an outward expansion against the stellar gravity.

However, such hot stars have a quite high radiative flux, since by the Stefan–Boltzmann law this scales as the fourth power of the surface temperature. Because light carries momentum as well as energy, this radiative flux imparts a force to the atoms that scatter the light. At the level of the star’s atmosphere where this force exceeds the inward force of the stellar gravity, material is accelerated upward and becomes the stellar wind.

An important aspect of this radiative driving process is that it often stems mostly from *line* scattering. As discussed in [Sect. 2.2](#), in a static medium such line interactions are confined to radiation wavelengths within a narrow thermal width of line center. However, in an accelerating stellar wind flow, the Doppler effect shifts the resonance to increasingly longer wavelengths, allowing the line scattering to sweep gradually through a much broader portion of the stellar spectrum (see [Fig. 15-3](#)). This gives the dynamics of such winds an intricate feedback character, in which the radiative driving force that accelerates the stellar wind depends itself on that acceleration.

5.1 Radiative Acceleration

5.1.1 Electron Scattering and the Eddington Limit

Before analyzing the radiative acceleration from line scattering, let us first consider the simpler case of scattering by free electrons, which is a “gray,” or frequency-independent process. Since gray scattering cannot alter the star’s total luminosity L , the radiative energy flux at any radius r is simply given by $F = L/4\pi r^2$, corresponding to a radiative *momentum* flux of $F/c = L/4\pi r^2 c$. As noted in [Sect. 2.2.1](#), the opacity for electron scattering is $\kappa_e = 0.2(1 + X) = 0.34 \text{ cm}^2/\text{g}$, where the latter result applies for the standard (solar) hydrogen mass fraction $X = 0.72$. The product of this opacity and the radiative momentum flux yields the radiative acceleration (force-per-unit-mass) from free-electron scattering,

$$g_e(r) = \frac{\kappa_e L}{4\pi r^2 c}. \quad (15.56)$$

It is of interest to compare this with the star’s gravitational acceleration, GM/r^2 . Since both accelerations have the same $1/r^2$ dependence on radius, their ratio is spatially constant, fixed by

the ratio of luminosity to mass,

$$\Gamma_e = \frac{\kappa_e L}{4\pi G M c}. \quad (15.57)$$

This ratio, sometimes called the Eddington parameter, thus has a characteristic value for each star. For the sun it is very small, of order 2×10^{-5} , but for hot, massive stars it can approach unity. As noted by Eddington, electron scattering represents a basal radiative acceleration that effectively counteracts the stellar gravity. The limit $\Gamma_e \rightarrow 1$ is known as the Eddington limit, for which the star would become gravitationally unbound.

It is certainly significant that hot stars with strong stellar winds have Γ_e only a factor two or so below this limit, since it suggests that only a modest additional opacity could succeed in fully overcoming gravity to drive an outflow. But it is important to realize that a stellar wind represents the outer envelope outflow from a nearly static, gravitationally bound base, and as such is not consistent with an *entire* star exceeding the Eddington limit. Rather the key requirement for a wind is that the driving force increases naturally from being smaller to larger than gravity at some radius near the stellar surface. How the force from line-scattering is ideally suited for just such a spatial modulation is described next.

5.1.2 Driving By Doppler-Shifted Resonant Line Scattering

As discussed in [Sect. 2.2](#), the *resonant strength* and sensitive *wavelength dependence* of bound-bound *line* opacity leads to strong, characteristic interactions of stellar radiation with the expanding wind material, giving rise to distinct features in the emergent stellar spectrum, like asymmetric P-Cygni line profiles, which provide key signatures of the wind expansion. These same basic properties – resonance strength and Doppler shift of the line resonance in the expanding outflow – also make the radiative force from line scattering the key factor in driving most massive-star winds. Relative to free-electron scattering, the overall amplification factor for a broadband, untuned radiation source is set by the *quality* of the resonance, $Q \approx \nu_o/A$, where ν_o is the line frequency and A is decay rate of the excited state. For quantum mechanically allowed atomic transitions, this can be very large, of order 10^7 . Thus, even though only a very small fraction ($\sim 10^{-4}$) of electrons in a hot-star atmosphere are bound into atoms, illumination of these atoms by an *unattenuated* (i.e., optically thin), broadband radiation source would yield a collective line-force which exceeds that from free electrons by about a factor $\bar{Q} \approx 10^7 \times 10^{-4} = 1,000$ (Gayley 1995). For stars within a factor two of the free-electron Eddington limit, this implies that line scattering is capable, in principle, of driving material outward with an acceleration on order a *thousand* times the inward acceleration of gravity!

In practice, of course, this does not normally occur, since any sufficiently large collection of atoms scattering in this way would readily block the limited flux available within just the narrow frequency bands tuned to the lines. Indeed, in the static portion of the atmosphere, the flux is greatly reduced at the line frequencies. Such line “saturation” keeps the overall line force quite small, in fact well below the gravitational force, which thus allows the inner parts of the atmosphere to remain gravitationally bound.

This, however, is where the second key factor, the *Doppler effect*, comes into play. As illustrated in [Fig. 15-3](#), in the outward-moving portions of the outer atmosphere, the Doppler effect redshifts the local line resonance, effectively desaturating the lines by allowing the atoms to resonate with relatively unattenuated stellar flux that was initially at slightly higher frequencies. By effectively sweeping a broader range of the stellar flux spectrum, this makes it possible

for the line force to overcome gravity and accelerate the very outflow it itself requires. As quantified within the CAK wind theory described below, the amount of mass accelerated adjusts such that the self-absorption of the radiation reduces the overall line-driving to being just somewhat (not a factor thousand) above what is needed to overcome gravity.

5.1.3 Radiative Acceleration from a Single, Isolated Line

For weak lines that are optically thin in the wind (i.e., with a characteristic optical depth from [● 15.5], $\tau_o \ll 1$), the line-acceleration takes a form similar to the simple electron scattering case (cf. [● 15.56]),

$$g_{thin} \equiv \frac{\kappa v_{th} v_o L_v}{4\pi r^2 c^2} = w_{v_o} q g_e, \quad (15.58)$$

where κ characterizes the opacity near line center, $q \equiv \kappa v_{th}/\kappa_e c$ is a dimensionless measure of the frequency-integrated line opacity (cf. ● Sect. 2.2.2), and $w_{v_o} \equiv v_o L_v/L$ weights the placement of the line within the luminosity spectrum L_v .

For stronger lines, the absorption and scattering of radiation can be treated within the Sobolev approximation in terms of the optical thickness of the *local* Sobolev resonance zone, as given in ● Sect. 2.2 for nonradial rays integrated over the finite-size star (see [● 15.7–15.9]). In the further approximation of purely radial streaming radiation from a point-like star, the radial acceleration takes the form,

$$g_{line} \approx g_{thin} \frac{1 - e^{-\tau_{Sob}}}{\tau_{Sob}}. \quad (15.59)$$

wherein the reduction from the optically thin case depends only on the local radial Sobolev optical depth given in (● 15.3), which is rewritten here as

$$\tau_{Sob} = \rho \kappa l_{Sob} = \frac{\rho \kappa v_{th}}{dv/dr} = \frac{\rho q \kappa_e c}{dv/dr} \equiv q t. \quad (15.60)$$

In the final equality, $t \equiv \kappa_e \rho c/(dv/dr)$ is the Sobolev optical depth for a line with integrated strength equal to free-electron scattering, that is, with $q = 1$.

In the limit of a very optically thick line with $\tau_s \gg 1$, the acceleration becomes,

$$g_{thick} \approx \frac{g_{thin}}{\tau_s} = w_{v_o} \frac{L}{4\pi r^2 \rho c^2} \frac{dv}{dr} = w_{v_o} \frac{L}{\dot{M} c^2} v \frac{dv}{dr}, \quad (15.61)$$

where the last equality uses the definition of the wind mass loss rate, $\dot{M} \equiv 4\pi \rho v r^2$. A key result here is that the optically thick line force is *independent* of the line strength q , and instead varies in proportion to the velocity gradient dv/dr . The basis of this is illustrated by ● Fig. 15-3, which shows that the local *rate* at which stellar radiation is redshifted into a line resonance depends on the *slope* of the velocity. By Newton's famous equation of motion, a force is normally understood to *cause* an acceleration. But here it is seen that an optically thick line-force also *depends* on the wind's advective rate of acceleration, $v dv/dr$.

5.1.4 Sobolev Localization of Line-Force Integrals for a Point Star

To provide a more quantitative illustration of this Sobolev approximation, let us now derive these key properties of line-driving through the localization of the spatial optical depth integral.

Under the simplifying approximation that the stellar radiation flux is purely radial (as from a central point-source), the force-per-unit-mass associated with direct absorption by a single line at a radius r is given by

$$g_{line}(r) = g_{thin} \int_{-\infty}^{\infty} dx \phi(x - u(r)) e^{-\tau(x,r)}. \quad (15.62)$$

The integration is over a scaled frequency $x \equiv (\nu/\nu_o - 1)(c/\nu_{th})$, defined from line center in units of the frequency broadening associated with the ion thermal speed ν_{th} , and $u(r) \equiv \nu(r)/\nu_{th}$ is the radial flow speed in thermal-speed units. The integrand is weighted by the line profile function $\phi(x)$, which for thermal broadening typically has the Gaussian form $\phi(x) \sim e^{-x^2}$. The exponential reduction takes account of absorption, as set by the frequency-dependent optical depth to the stellar surface radius R ,

$$\tau(x, r) \equiv \int_R^r dr' \kappa \rho(r') \phi(x - u(r')). \quad (15.63)$$

A crucial point in evaluating this integral is that in a supersonic wind, the variation of the integrand is dominated by the velocity variation within the line-profile. As noted above, this variation has a scale given by the Sobolev length $l_{sob} \equiv \nu_{th}/(d\nu/dr)$, which is smaller by a factor ν_{th}/ν than the competing density/velocity scale, $H \equiv |\rho/(d\rho/dr)| \approx \nu/(d\nu/dr)$. A key step in the Sobolev approximation is thus to recast this spatial integration as an integration over the comoving-frame frequency $x' \equiv x - u(r')$,

$$\tau(x, r) = - \int_{x'(R)}^{x'(r)} \frac{dx' \nu_{th}}{d\nu/dr'} \kappa \rho(x') \phi(x') \approx \tau_{sob} \Phi(x - u(r)), \quad (15.64)$$

where the integrated profile

$$\Phi(x) \equiv \int_x^{\infty} dx' \phi(x'), \quad (15.65)$$

and the latter approximation in (15.64) uses the assumption that $\nu(r) \gg \nu_{th}$ to formally extend the surface frequency $x'(R)$ to infinity relative to the local resonance $x'(r)$. As defined in (15.60), the Sobolev optical thickness $\tau_s = qt$ arises here as a collection of spatial variables that are assumed to be nearly constant over the Sobolev resonance zone, and thus can be extracted outside the integral.

Finally, a remarkable, extra bonus from this approximation is that the resulting optical depth (15.64) now has precisely the form needed to allow analytic evaluation of the line-force integral (15.62), yielding directly the general Sobolev force expression given in (15.59).

5.2 The CAK Model for Line-Driven Winds

5.2.1 The CAK Line-Ensemble Force

In practice, a large number of lines with a range of frequencies and strengths can contribute to the wind driving. Castor et al. (1975; hereafter CAK) introduced a practical formalism for computing the cumulative line-acceleration from a parameterized line-ensemble, under the simplifying assumption that the spectral distribution keeps the individual lines nearly independent.

For the point-star model used above, direct summation of the individual line-accelerations as given by (15.59) yields

$$g_{tot} = g_e \sum w_{v_o} q \frac{1 - e^{-qt}}{qt} \approx g_e \int_0^\infty q \frac{dN}{dq} \frac{1 - e^{-qt}}{qt} dq, \quad (15.66)$$

where the latter equality approximates the discrete sum as a continuous integral over the flux-weighted number distribution dN/dq . Following CAK, a key further simplification is to approximate this number distribution as a simple power law in the line strength q ,

$$q \frac{dN}{dq} = \frac{1}{\Gamma(\alpha)} \left(\frac{q}{\bar{Q}} \right)^{\alpha-1}, \quad (15.67)$$

where $\Gamma(\alpha)$ is the complete Gamma function, the CAK power-law index satisfies $0 < \alpha < 1$, and the above mentioned cumulative line strength \bar{Q} provides a convenient overall normalization (cf. Sect. 4.2.2). Application of (15.67) in (15.66) then yields the CAK, point-star, radiative acceleration from this line-ensemble,

$$\begin{aligned} g_{CAK} &= \frac{\bar{Q}^{1-\alpha}}{(1-\alpha)} \frac{g_e}{t^\alpha} \\ &= \frac{1}{(1-\alpha)} \frac{\kappa_e L \bar{Q}}{4\pi r^2 c} \left(\frac{dv/dr}{\rho c \bar{Q} \kappa_e} \right)^\alpha. \end{aligned} \quad (15.68)$$

Note that this represents a kind of “geometric mean” between the optically thin and thick forms (15.58) and (15.61) for a single line.²

5.2.2 CAK Dynamical Solution for Mass Loss Rate and Terminal Speed

The simple, local CAK/Sobolev expression (15.68) for the cumulative line-driving force in the point-star approximation provides a convenient basis for deriving the basic scalings for a line-driven stellar wind. Since, as noted above, the gas pressure is not of much importance in the overall wind driving, let us just consider the steady-state equation of motion (15.31) in the limit of zero sound speed, $a = 0$. This simply requires that the wind acceleration must equal the line-acceleration minus the inward acceleration of gravity,

$$v \frac{dv}{dr} = g_{CAK} - \frac{GM(1 - \Gamma_e)}{r^2}, \quad (15.69)$$

wherein we have also taken into account the effective reduction of gravity by the free-electron scattering factor Γ_e . Note from (15.68) that the CAK line-force g_{CAK} depends itself on the flow acceleration it drives.

²CAK dubbed the ratio g_{CAK}/g_e the “force multiplier,” written as $M(t) = kt^{-\alpha}$, with k a normalization constant. Note however that their original formulation used a fiducial thermal speed v_{th} in the definition of the optical depth parameter $t = \kappa_e \rho v_{th} / (dv/dr)$, which then gives an artificial thermal speed dependence to the CAK normalization, $k = (v_{th}/c)^\alpha \bar{Q}^{1-\alpha} / (1-\alpha)$. In the Sobolev approximation, the line-force has no physical dependence on the thermal-speed, and so the formal inclusion of a fiducial v_{th} in the CAK parameterization has sometimes led to confusion, e.g., in applying tabulations of the multiplier constant k . The formulation here avoids this problem, since the parameters q , \bar{Q} , and $t = \kappa_e \rho c / (dv/dr)$ are all independent of v_{th} . A further advantage is that, for a wide range of hot-star parameters, the line normalization has a relatively constant value $\bar{Q} \approx 2000Z$, where Z is the metallicity relative to the standard solar value (Gayley 1995).

Since this feedback between line-driving and flow acceleration is moderated by gravity, let us define the gravitationally scaled inertial acceleration,

$$w' \equiv \frac{r^2 v v'}{GM(1 - \Gamma_e)}. \quad (15.70)$$

In terms of an inverse-radius coordinate $x \equiv 1 - R/r$, note that $w' = dw/dx$, where $w \equiv v^2/v_{esc}^2(1 - \Gamma_e)$ represents the ratio of wind kinetic energy to the (electron-force-reduced) effective gravitational binding from the surface. Using (15.68), the equation of motion can then be rewritten in the simple, dimensionless form,

$$w' = C w'^\alpha - 1, \quad (15.71)$$

where the constant is given by

$$C = \frac{1}{1 - \alpha} \left(\frac{\bar{Q}\Gamma_e}{1 - \Gamma_e} \right)^{1-\alpha} \left(\frac{L}{\dot{M}c^2} \right)^\alpha. \quad (15.72)$$

Figure 15-9 illustrates the graphical solution of this dimensionless equation of motion for various values of the constant C . For fixed stellar and opacity-distribution parameters, this corresponds to assuming various values of the mass loss rate \dot{M} , since $C \propto 1/\dot{M}^\alpha$. For high \dot{M} (low C), there are no solutions, while for low \dot{M} (high C), there are two solutions. The two limits are separated by a critical case with one solution – corresponding to the *maximal* mass loss rate – for which the function Cw'^α intersects the line $1 + w'$ at a tangent. For this critical case,

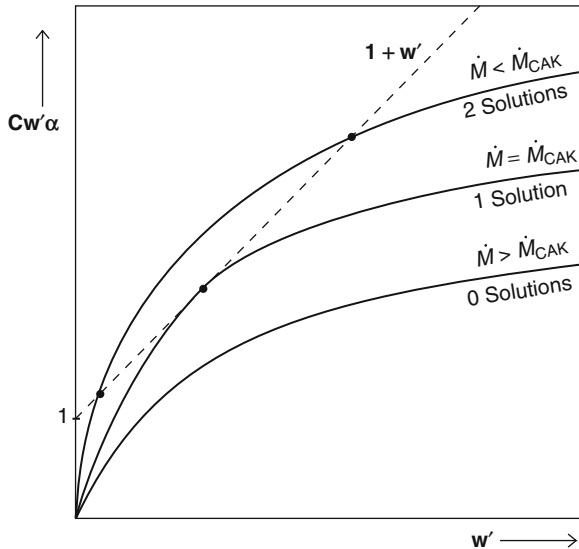


Fig. 15-9

Graphical solution of the dimensionless equation of motion (15.71) representing a 1D, point-star, zero-sound-speed CAK wind, as controlled by the constant $C \sim 1/\dot{M}^\alpha$. If \dot{M} is too big, there are no solutions; if \dot{M} is small there are two solutions. A maximal value $\dot{M} = \dot{M}_{CAK}$ defines a single, “critical” solution

the tangency requirement implies $\alpha C_c w_c^{\alpha-1} = 1$, which together with the original (► 15.71) yields the critical conditions $w'_c = \alpha/(1-\alpha)$ and $C_c = 1/\alpha^\alpha (1-\alpha)^{1-\alpha}$.

From (► 15.72), this critical value of C_c defines the maximal CAK mass loss rate

$$\dot{M}_{CAK} = \frac{L}{c^2} \frac{\alpha}{1-\alpha} \left[\frac{\bar{Q}\Gamma_e}{1-\Gamma_e} \right]^{(1-\alpha)/\alpha}. \quad (15.73)$$

Moreover, since (► 15.71) has no explicit spatial dependence, these critical conditions hold at all radii. By spatial integration of the critical acceleration w'_c from the surface radius R , we thus obtain a specific case of the general “beta”-velocity law,

$$v(r) = v_\infty \left(1 - \frac{R_*}{r} \right)^\beta, \quad (15.74)$$

where $\beta = 1/2$, and the wind terminal speed $v_\infty = v_{esc} \sqrt{\alpha(1-\Gamma_e)/(1-\alpha)}$.

An important success of these CAK scaling laws is the theoretical rationale they provide for an empirically observed “wind momentum–luminosity” (WML) relation (Kudritzki et al. 1995). Combining the CAK mass-loss law (► 15.73) together with the scaling of the terminal speed with the effective escape, we obtain a WML relation of the form,

$$\dot{M} v_\infty \sqrt{R_*} \sim L^{1/\alpha} \bar{Q}^{1/\alpha-1} \quad (15.75)$$

wherein we have neglected a residual dependence on $M(1-\Gamma_e)$ that is generally very weak for the usual case that α is near $2/3$. Fits for galactic OB supergiants (Puls et al. 1996) give a luminosity slope consistent with $\alpha \approx 0.6$, with a normalization consistent with $\bar{Q} \approx 10^3$. Note that the direct dependence $\bar{Q} \sim Z$ provides the scaling of the WML with metallicity Z .

5.3 Extensions of Idealized CAK Model

5.3.1 Finite-Size Star

These CAK results strictly apply only under the idealized assumption that the stellar radiation is radially streaming from a point-source. If one takes into account the finite angular extent of the stellar disk, then near the stellar surface the radiative force is reduced by a factor $f_{d*} \approx 1/(1+\alpha)$, leading to a reduced mass loss rate (Friend and Abbott 1986; Pauldrach et al. 1986),

$$\dot{M}_{fd} = f_{d*}^{1/\alpha} \dot{M}_{CAK} = \frac{\dot{M}_{CAK}}{(1+\alpha)^{1/\alpha}} \approx \dot{M}_{CAK}/2. \quad (15.76)$$

Away from the star, the correction factor increases back toward unity, which for the reduced base mass flux implies a stronger, more extended acceleration, giving a somewhat higher terminal speed, $v_\infty \approx 3v_{esc}$, and a flatter velocity law, approximated by replacing the exponent in (► 15.74) by $\beta \approx 0.8$.

5.3.2 Radial Variations in Ionization

The effect of a radial change in ionization can be approximately taken into account by correcting the CAK force (► 15.68) by a factor of the form $(n_e/W)^\delta$, where n_e is the electron density,

$W \equiv 0.5(1 - \sqrt{1 - R_*/r})$ is the radiation “dilution factor,” and the exponent has a typical value $\delta \approx 0.1$ (Abbott 1982). This factor introduces an additional density dependence to that already implied by the optical depth factor $1/t^\alpha$ given in (► 15.68). Its overall effect can be roughly taken into account with the simple substitution $\alpha \rightarrow \alpha' \equiv \alpha - \delta$ in the power exponents of the CAK mass loss scaling law (► 15.73). The general tendency is to moderately increase \dot{M} , and accordingly to somewhat decrease the wind speed.

5.3.3 Finite Gas Pressure and Sound Speed

The above scalings also ignore the finite gas pressure associated with a small but nonzero sound speed a . If we apply the full gas-pressure equation of motion (► 15.31) to the case with line-driving ($g_x = g_{\text{CAK}}$), then through a perturbation expansion in the small parameter a/v_{esc} , it is possible to derive simple scalings for the fractional corrections to the mass loss rate and terminal speed (Owocki and ud-Doula 2004),

$$\delta m \approx \frac{4\sqrt{1-\alpha}}{\alpha} \frac{a}{v_{\text{esc}}} ; \quad \delta v_\infty \approx \frac{-\alpha \delta m}{2(1-\alpha)} \approx \frac{-2}{\sqrt{1-\alpha}} \frac{a}{v_{\text{esc}}} . \quad (15.77)$$

For a typical case with $\alpha \approx 2/3$ and $a \approx 20 \text{ km/s} \approx v_{\text{esc}}/30$, the net effect is to increase the mass loss rate and decrease the wind terminal speed, both by about 10%.

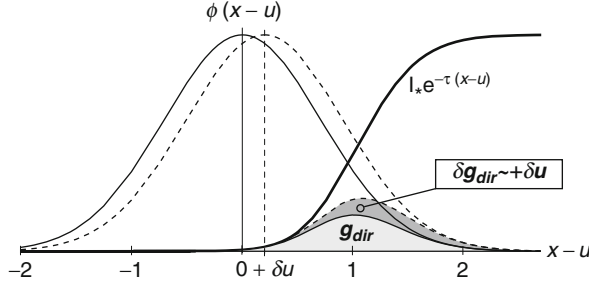
5.4 Wind Instability and Variability

The steady, spherically symmetric, line-driven-wind models described above have had considerable success in explaining the inferred general properties of OB-star winds, like the total mass loss rate and mean velocity law. But when viewed more carefully there is substantial evidence that such winds are actually quite highly structured and variable on a broad range of spatial and temporal scales.

Large-scale variations give rise to *Discrete Absorption Components* (DACs) seen in UV spectra lines (Howarth and Prinja 1989; Fullerton 2003). These begin as broad absorption enhancement in the inner part of the blue absorption trough of an unsaturated P-Cygni line, which then gradually narrow as they drift, over a period of days, toward the blue-edge of the profile. These DACs likely result from disturbances (e.g., localized magnetic fields or nonradial pulsations) at the photospheric base of the stellar wind, which through stellar rotation evolve in the wind into semi-regular “Co-rotating Interaction Regions” (CIRs: Mullan 1984; Cranmer and Owocki 1996), much as detected directly in the solar wind at boundaries between fast and slow wind streams (see ► Sect. 2.1).

But there is also strong, though indirect evidence that OB winds have an extensive, small-scale, turbulent structure. Saturated P-Cygni lines have extended black troughs thought to be a signature that the wind velocity is highly nonmonotonic (Lucy 1982), and such stars commonly exhibit soft X-ray emission thought to originate from embedded wind shocks (Cohen 2008).

This small-scale flow structure likely results from the strong *Line-Deshadowing Instability* (LDI) that is intrinsic to the radiative driving by line scattering (MacGregor et al. 1979; Owocki and Rybicki 1984, 1985). As noted above, there is a strong hidden potential in line scattering to drive wind material with accelerations that greatly exceed the mean outward acceleration. For any small velocity perturbations with a length scale near or below the Sobolev length l_{Sob} ,



■ Fig. 15-10

Illustration of the physical origin of the line deshadowing instability. At an arbitrary point in the expanding wind, it shows the line profile ϕ and direct intensity from the star plotted vs. comoving-frame frequency $x - u$. The light shaded overlap area is proportional to the component line-acceleration g_{dir} . The dashed profile shows the effect of the Doppler shift from a perturbed velocity δu (in units of thermal speed v_{th}), with the resulting extra area in the overlap with the blue-edge intensity giving a perturbed line-force δg that scales in proportion to this perturbed velocity

linear stability analyses show that the perturbed radiative acceleration δg no longer scales with the velocity gradient $\delta v'$, as expected from the above Sobolev analysis (Abbott 1980), but rather in direct proportion to the perturbed velocity itself, $\delta g \sim \delta v$ (MacGregor et al. 1979; Owocki and Rybicki 1984). This leads to a strong instability, in which the stronger acceleration leads to a faster flow, which in turns leads to an even stronger acceleration (see ► Fig. 15-10).

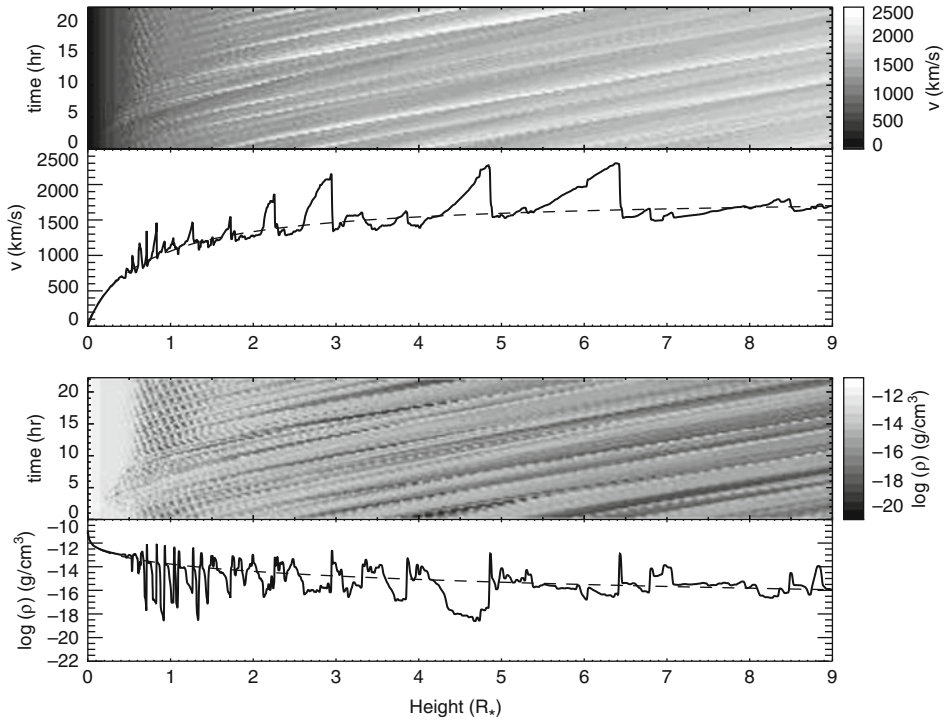
The growth rate of this instability scales with the ratio of the mean driving force to the ion thermal speed, $\omega_g \approx g_{CAK}/v_{th}$. Since the CAK line-driving also sets the flow acceleration, that is, $g_{CAK} \approx v dv/dr$, it is readily found that the growth is approximately at the flow rate through the Sobolev length, $\omega_g \approx v/l_{Sob}$. This is a large factor $v/v_{th} \approx 100$ bigger than the typical wind expansion rate $dv/dr \approx v/R$, implying then that a small perturbation at the wind base would, within this lineary theory, be amplified by an enormous factor, of order $e^{v/v_{th}} \approx e^{100}$!

In practice, a kind of “line-drag” effect of the *diffuse* component of radiation scattered within the lines reduces this stability (Lucy 1984; Owocki and Rybicki 1985), giving a net growth rate that scales roughly as

$$\omega_g(r) \approx \frac{g_{CAK}}{v_{th}} \frac{\mu_*(r)}{1 + \mu_*(r)}. \quad (15.78)$$

where $\mu_* \equiv \sqrt{1 - R^2/r^2}$. This net growth rate vanishes near the stellar surface, where $\mu_* = 0$, but it approaches half the pure-absorption rate far from the star, where $\mu_* \rightarrow 1$. This implies that the outer wind is still very unstable, with cumulative growth of ca. $v_\infty/2v_{th} \approx 50$ e-folds.

Because the instability occurs at spatial scales near and below the Sobolev length, hydrodynamical simulations of its nonlinear evolution cannot use the Sobolev approximation, but must instead carry out *nonlocal* integrals for both direct and diffuse components of the line radiation transport (Owocki et al. 1988; Feldmeier 1995; Owocki and Puls 1996, 1999). The right panel of ► Fig. 15-11 illustrates typical results of a 1-D time-dependent simulation. Because of the stabilization from line-drag, the flow near the photospheric wind base follows the steady, smooth CAK form (dashed curve), but beginning about $r \approx 1.5R$, the strong intrinsic instability lead to extensive wind structure, characterized by high-speed rarefactions in between slower dense, compressed shells.



■ Fig. 15-11

Results of 1-D, time-dependent numerical hydrodynamics simulation of the nonlinear growth of the instability. The line plots show the spatial variation of velocity (*upper*) and density (*lower*) at a fixed, arbitrary time snapshot. The corresponding gray scales show both the time (*vertical axis*) and height (*horizontal axis*) evolution. The dashed curve shows the corresponding smooth, steady CAK model. from the vertical in intervals of 4,000 s from the CAK initial condition

Because of the computational expense of carrying out the nonlocal integrals for the line-acceleration, 2-D simulations of this instability have so far been limited, either ignoring or using simplified approximations for the lateral components of the diffuse line-drag (Owocki 1999; Dessart and Owocki 2003, 2005). But results so far indicate the compressed shells of 1-D simulations should break up into small-scale clumps, with a characteristic size of a few percent of the local radius, and a volume filling factor $f_v \approx 0.1$. As noted in ► Sect. 2.2.1, this has important implications for interpreting diagnostics that scale with the square of wind density, suggesting that mass loss rates based on such diagnostics may need to be reduced by factors of a few, that is, $1/\sqrt{f_v} \approx \sqrt{10}$.

5.5 Effect of Rotation

The hot, luminous stars that give rise to line-driven stellar winds tend generally to have quite rapid stellar rotation. This is most directly evident through the extensive broadening of their

photospheric absorption lines, which suggest projected equatorial rotation speeds $V_{rot} \sin i$ of hundreds of km/s, where $\sin i \leq 1$ accounts for the inclination angle i of the observer line of sight to the stellar rotation axis. In some of the most rapid rotators, e.g., the Be stars, the surface rotation speed V_{rot} is inferred to be a substantial fraction, perhaps 70–80% or more, of the so-called critical rotation speed, $V_{crit} \equiv \sqrt{2GM/R}$, for which material at the equatorial surface would be in Keplerian orbit (Townsend et al. 2004).

Initial investigations (Friend and Abbott 1986; Pauldrach et al. 1986) of the effect of rotation on radiatively driven winds derived 1-D models based on the standard CAK line-driving formalism, but now adding the effect of a centrifugally reduced, *effective* surface gravity as a function of colatitude θ ,

$$g_{eff}(\theta) = \frac{GM}{R^2} (1 - \Omega^2 \sin^2 \theta), \quad (15.79)$$

where $\Omega^2 \equiv V_{rot}^2 R/GM$, and for simplicity, we have ignored any rotational distortion of the surface radius R . This allows one to write the standard CAK mass loss rate scaling law (cf. [15.73]) in terms of *surface* values of the mass flux $\dot{m} = \rho v$, radiative flux F , and effective gravity g_{eff} , relative to corresponding polar ($\theta = 0$) values \dot{m}_o , F_o , and $g_o = GM/R^2$,

$$\frac{\dot{m}(\theta)}{\dot{m}_o} = \left[\frac{F(\theta)}{F_o} \right]^{1/\alpha} \left[\frac{g_{eff}(\theta)}{g_o} \right]^{1-1/\alpha}. \quad (15.80)$$

If, as was initially assumed, the surface flux is taken to be constant in latitude, $F(\theta) = F_o$, then we obtain the scaling

$$\frac{\dot{m}(\theta)}{\dot{m}_o} = [1 - \Omega^2 \sin^2 \theta]^{1-1/\alpha}; \quad F(\theta) = F_o. \quad (15.81)$$

Since $\alpha < 1$, the exponent $1-1/\alpha$ is negative, implying that the mass flux increases monotonically from pole ($\sin \theta \rightarrow 0$) toward the equator ($\sin \theta \rightarrow 1$).

However, as first demonstrated by von Zeipel (1924), for a radiative stellar envelope undergoing solid body rotation, the emergent radiative flux at any latitude is predicted to vary in proportion to the centrifugally reduced effective gravity, $F(\theta) \sim g_{eff}$. Applying this equatorial “gravity darkening” for the surface flux, we obtain

$$\frac{\dot{m}(\theta)}{\dot{m}_o} = 1 - \Omega^2 \sin^2 \theta; \quad F(\theta) \sim g_{eff}(\theta) \quad (15.82)$$

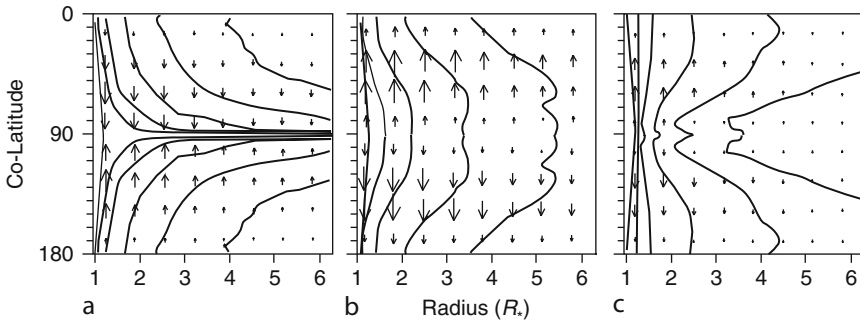
so that the mass flux now *decreases* toward the equator, with a maximum at the *pole*!

Recalling that the wind terminal speed tends to scale with the surface gravity through the escape speed, this 1-D analysis also predicts a latitudinally varying wind speed that is proportional to a centrifugally reduced, effective escape speed,

$$v_\infty(\theta) \sim v_{esc} \sqrt{1 - \Omega^2 \sin^2 \theta}. \quad (15.83)$$

The latitudinal variation of wind density is then obtained from $\rho(\theta) \sim \dot{m}(\theta)/v_\infty(\theta)$.

More generally, the wind from a rotating star can also flow in latitude as well as radius, requiring then a 2-D model. In this regard, a major conceptual advance was the development of the elegantly simple “Wind Compressed Disk” (WCD) paradigm by Bjorkman and Cassinelli (1993). They noted that, like satellites launched into earth orbit, parcels of gas gradually driven radially outward from a rapidly rotating star should remain in a tilted “orbital plane” that brings them over the equator, where they collide to form compressed disk. Initial 2-D hydrodynamical simulations (Owocki et al. 1994) generally confirmed the basic tenets of the WCD model



■ Fig. 15-12

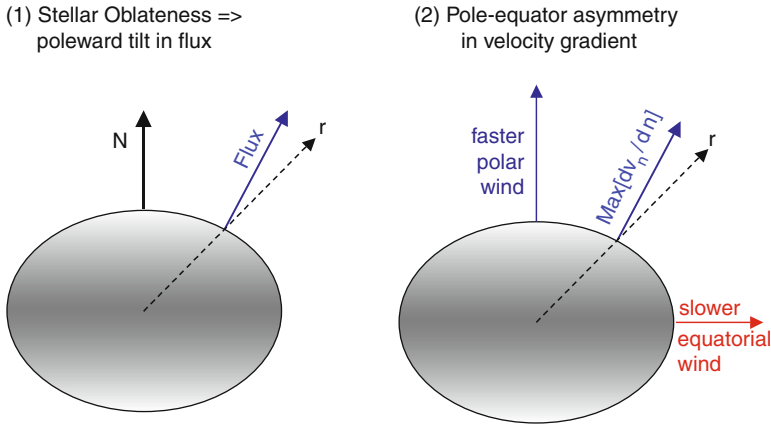
Contours of density in stellar wind from a star rotating at 75% of critical rate, plotted vs. colatitude θ and radius r , spaced logarithmically with two contours per decade. The superposed vectors represent the latitudinal velocity, with the maximum length corresponding to a magnitude of $v_\theta = 100$ km/s. The three panels show the cases (a) without nonradial forces or gravity darkening; (b) with nonradial forces but no gravity darkening, and; (c) with both nonradial forces and gravity darkening

(► Fig. 15-12a), with certain detailed modifications (e.g., infall of inner disk material). But later simulations (Owocki et al. 1996) that account for a net poleward component of the line-driving force (► Fig. 15-13) showed that this can effectively *reverse* the equatorward drift, and so completely *inhibit* formation of any equatorial compressed disk (► Fig. 15-12b). Indeed, when equatorial gravity darkening is taken into account, the lower mass flux from the equator makes the equatorial wind have a reduced, rather than enhanced, density (► Fig. 15-12c).

The net upshot then is that a radiatively driven wind from a rapidly rotating star is predicted to be both faster and denser over the poles, instead of the equator. This may help explain spectroscopic and interferometric evidence that the current-day wind of the extreme massive star η Carinae is faster and denser over the poles, leading to a prolate shape for its dense, optically thick “wind photosphere” (Smith et al. 2003; van Boekel et al. 2003; Groh et al. 2010). Extensions of the rotational scalings to the case of continuum-driven mass loss could also explain the bipolar shape of the Homunculus nebula (Owocki et al. 2004).

5.6 Summary for Radiatively Driven Massive-Star Winds

- The large ratio of luminosity to mass means OB stars are near the “Eddington limit,” for which the radiative force from just scattering by free electrons nearly cancels the stellar gravity.
- The resonance nature of line absorption by bound electrons makes their cumulative effective opacity of order a thousand times larger than free electrons, but the discrete energies of bound states means the opacity is tuned to very specific photon wavelengths.
- In a static atmosphere, saturation of line absorption keeps the associated force smaller than gravity; but in an accelerating wind, the associated Doppler shift of the line frequency exposes it to fresh continuum radiation, allowing the line-force to become strong enough to overcome gravity and drive the outflow.



■ Fig. 15-13

Illustration of the origin of poleward component of the radiative force from a rotating star. *Left:* The poleward tilt of the radiative flux arising from the oblateness of the stellar surface contributes to a poleward component of the driving force. *Right:* Since the wind speed scales with surface escape speed, the lower effective gravity of the equator leads to a slower equatorial speed. The associated poleward increase in speed leads to a poleward tilt in the velocity gradient, and this again contributes to a poleward component of the line force

- Because lines have a thermally broadened velocity width ($v_{th} \lesssim a$) much less than the wind flow speed ($v \approx v_{esc}$), the absorption or scattering of photons in such an accelerating wind occurs over a narrow resonance layer (of width $l_{Sob} = v_{th}/(dv/dr) \approx (v_{th}/v_{esc})R$); following methods introduced by V. V. Sobolev, this allows nearly local solution of the line transfer in such accelerating flow.
- Within the CAK model for a power-law distribution of line strengths, the cumulative line-force scales with a power of the local velocity gradient divided by the density.
- For the idealized case of negligible gas pressure and a radially streaming point-source of stellar radiation, application of this CAK line-force within a steady-state radial equation of motion yields analytic scalings for the mass loss rate and wind velocity law.
- Accounting for the finite cone-angle of the stellar disk and radial ionization balance variations in the driving opacity yield order-unity corrections to the point-star scalings. Corrections for a finite gas pressure are smaller, of order 10%.
- Overall, the predictions of the CAK scalings agree well with an observationally inferred wind momentum–luminosity relation for OB supergiants.
- Careful analysis indicates that hot-star winds have extensive structure and variability, with small-scale turbulent structure likely arising from intrinsic instability of line-driving. The associated clumping has important implications for interpreting mass-loss rate diagnostics that scale with density squared.
- Rotation can also significantly affect the mass loss and wind speed, with the gravity darkening from rapid rotation leading to a somewhat surprising polar enhancement of wind density and flow speed.


6 Wolf-Rayet Winds and Multiline Scattering

The above CAK model is based on a simplified picture of absorption of the star's radiative momentum by many *independent* lines, effectively ignoring overlap effects among optically thick lines. Since each such thick line sweeps out a fraction v_∞/c of the star's radiative momentum L/c , the ratio of wind to radiative momentum is

$$\eta \equiv \frac{\dot{M}v_\infty}{L/c} = N_{thick} \frac{v_\infty}{c}, \quad (15.84)$$

which is sometimes termed the wind “momentum” or “performance” number. To avoid overlap, we must require $N_{thick} < c/v_\infty$, which in turn implies the so-called single-scattering limit for the wind momentum, $\eta < 1$. Since most OB winds are inferred to be below this limit, the basic CAK formalism ignoring overlap still provides a reasonably good model for explaining their overall properties.

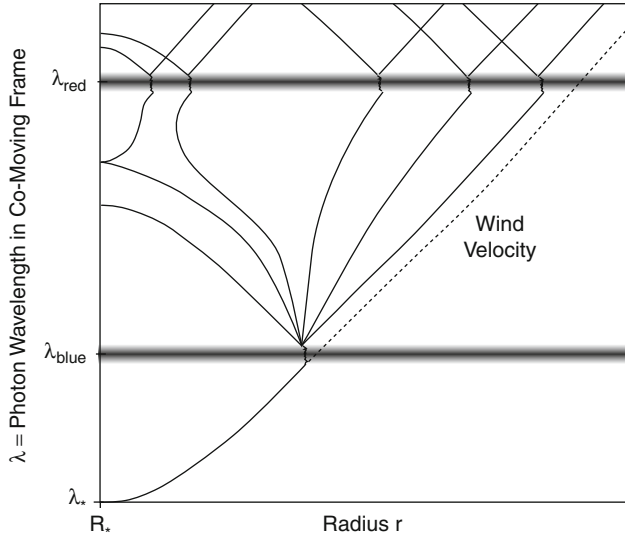
However, such a single-scattering formalism seems quite inadequate for the much stronger winds of Wolf-Rayet (WR) stars. Wolf-Rayet stars are evolved, massive, hot stars for which the cumulative mass loss has led to depletion of the original hydrogen envelope. They typically show broad wind-emission lines of elements like carbon, nitrogen, and/or oxygen that are the products of core nucleosynthesis. Overall, observations indicate that WR winds are especially strong, and even optically thick to continuum scattering by electrons. Notably, inferred WR wind momenta $\dot{M}v_\infty$ are generally substantially higher than for OB stars of comparable luminosity, placing them well above the OB-star line in the above wind momentum–luminosity relation. In fact, in WR winds the inferred momentum numbers are typically well above the single-scattering value $\eta = 1$, sometimes as high as $\eta = 10\text{--}50$!

This last property has often been cast as posing a WR wind “momentum problem,” sometimes with the implication that it means WR winds cannot be radiatively driven. In fact, it merely means that, unlike for OB stars, WR winds cannot be treated in the standard, single-scattering formalism. However, momentum ratios above unity can, in principle, be achieved by *multiple scattering* between overlapping thick lines with a velocity-unit frequency separation $\Delta v < v_\infty$.  [Figure 15-14](#) illustrates this for the simple case of two overlapping lines. But, as discussed below, the large momentum of WR winds requires much more extensive overlap, with thick lines spread densely throughout the spectrum without substantial gaps that can allow radiation to leak out. An overall theme here is that understanding WR mass loss represents more of an “opacity” than a “momentum” problem.

6.1 Example of Multiple Momentum Deposition in a Static Gray Envelope

To provide the basis for understanding such multiline scattering, it is helpful first to review the momentum deposition for *continuum* scattering. The total radial momentum imparted by radiation on a spherically symmetric circumstellar envelope can be expressed in terms of the radiative force density ρg_{rad} integrated over volume, outward from the wind base at radius R ,

$$\dot{p}_{rad} = \int_R^\infty 4\pi r^2 \rho g_{rad} dr, \quad (15.85)$$



■ Fig. 15-14

The Doppler-shifted line resonance in an accelerating flow, for two lines with relative wavelength separation $\Delta\lambda/\lambda < v_\infty/c$ close enough to allow multiline scattering within the wind. Photons scattered by the bluer line are reemitted with nearly equal probability in the forward or backward directions, but then further redshifted by the wind expansion into resonance with the redder line. Because of the gain in radial direction between the line resonances, the red-line scattering imparts an additional component of outward radial momentum

where the radiative acceleration is given by a frequency integral of the opacity κ_ν over the stellar flux spectrum, F_ν ,

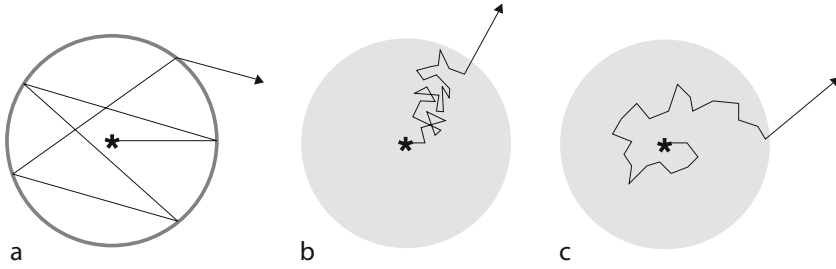
$$g_{rad} = \int_0^\infty \frac{\kappa_\nu F_\nu}{c} d\nu \quad (15.86)$$

A particularly simple way to illustrate the requirements of multiple momentum deposition is in terms of an envelope with a gray opacity κ . In this case, the flux is just a constant in frequency, set at each radius by the bolometric luminosity through $F = L/4\pi r^2$. This implies $g_{rad} = \kappa L/4\pi r^2 c$, and so yields

$$\dot{p}_{rad} = \frac{L}{c} \int_R^\infty \kappa \rho dr = \frac{L}{c} \tau, \quad (15.87)$$

where τ is the total wind optical depth. It is thus seen that the requirement for exceeding the single-scattering limit for radiative momentum deposition, $\dot{p}_{rad} > L/c$, is simply that the envelope be optically thick, $\tau > 1$.

► Figure 15-15 provides a geometric illustration of how multiple momentum deposition occurs in an optically thick envelope. ► Figure 15-15a shows the case of a hollow shell with optical thickness $\tau = 5$, wherein a photon is backscattered within the hollow sphere roughly τ times before escaping, having thus cumulatively imparted τ times the single photon momentum, as given by (► 15.87).



■ Fig. 15-15

Schematic photon trajectories in (a) hollow and; (b) filled gray spheres, with part; (c) illustrating the net “winding” that makes even the filled case have a persistent net outward push that implies multiple radial momentum deposition (see text for details)

However, the same momentum deposition also occurs in a *solid* sphere with the same total optical depth, even though, as shown in ● Fig. 15-15b, photons in this case undergo a much more localized diffusion without hemispheric crossing.

● Figure 15-15c illustrates how these diffusive vs. direct-flight pictures of multiple momentum deposition can be reconciled by thinking in terms of an effective “winding” around the envelope. For each scattering within a spherical envelope, the radial momentum deposition is unchanged by arbitrary rotations about a radius through the scattering point. For ● Fig. 15-15c the rotations are chosen to bring all the scattered trajectories into a single plane, with the azimuthal component always of the same sense, here clockwise viewed from above the plane. In this artificial construction, scattering thus leads to a *systematic* (vs. random walk) drift of the photon along one azimuthal direction, implying a cumulative *winding* trajectory for which the systematic *outward* push of the scattered photon is now apparent.

6.2 Multiline Transfer in an Expanding Wind

The multiple scattering by a dense ensemble of lines can actually be related to the above gray envelope case, *if* the spectral distribution of lines is Poisson, and so is spread throughout the spectrum without extended bunches or gaps. As first noted by Friend and Castor (1983), and later expanded on by Gayley et al. (1995), a wind driven by such an effectively gray line-distribution can be analyzed through an extension of the above standard CAK formalism traditionally applied to the more moderate winds of OB stars. In the case of dense overlap, wherein optically thick lines in the wind have a frequency separation characterized by a velocity Δv much less than the wind terminal speed v_∞ , the mean-free-path between photon interactions with separate lines is given by

$$\frac{1}{\rho \kappa^{\text{eff}}} = \frac{\Delta v}{dv_n/dn}, \quad (15.88)$$

where $dv_n/dn \equiv \hat{\mathbf{n}} \cdot \nabla (\hat{\mathbf{n}} \cdot \mathbf{v})$ is the projected wind velocity gradient along a given photon direction $\hat{\mathbf{n}}$. The directional dependence of this velocity gradient implies an inherent *anisotropy* to the associated effectively gray line-ensemble opacity κ^{eff} . However, a full non-isotropic diffusion analysis (Gayley et al. 1995) indicates that the overall wind dynamics is not too sensitive to this anisotropy. In particular, the total wind momentum can still be characterized by the effective

radial optical depth, τ_r^{eff} , which in this case yields,

$$\dot{p}_{\text{rad}} \approx \frac{L}{c} \tau_r^{\text{eff}} = \frac{L}{c} \int_R^\infty \rho \kappa_r^{\text{eff}} dr = \frac{L}{c} \int_R^\infty \frac{1}{\Delta v} \frac{dv_r}{dr} dr = \frac{L}{c} \frac{v_\infty}{\Delta v}, \quad (15.89)$$

from which we identify $\tau_r^{\text{eff}} = v_\infty / \Delta v$. Neglecting a modest correction for gravitational escape, global momentum balance requires $\dot{p}_{\text{rad}} \approx \dot{M} v_\infty$, thus implying

$$\eta \approx \frac{v_\infty}{\Delta v}. \quad (15.90)$$

For winds driven by a gray ensemble of lines, it is thus seen that large momentum factors $\eta \gg 1$ simply require that there be a large number of optically thick lines overlapping within the wind, $v_\infty \gg \Delta v$.

Note that (► 15.90) implies that the mass loss scales as

$$\dot{M} \approx \frac{L}{c^2} \frac{c}{\Delta v}, \quad (15.91)$$

wherein it is noted that $c / \Delta v$ just represents the total, spectrum-integrated number of thick lines, N_{thick} . It is important to realize, however, that this number of thick lines is not fixed a priori, but is itself dependent on wind properties like the mass loss rate and velocity law.

Self-consistent solution is again possible though through an extension of the standard CAK formalism. The key is to account for the fact that radiation entering into resonance of each line does not in general come directly from the stellar core, but instead has been previously scattered by the next blueward overlapping line. In the limit of strong overlap, the transfer between lines can be treated as local diffusion, using however a non-isotropic diffusion coefficient to account for the directional variation of the velocity gradient. In analogy to the finite-disk correction for point-star CAK model, one can then derive a “non-isotropic diffusion” correction factor, f_{nid} , to account for the diffuse angle distribution of the previously line-scattered radiation as it enters into resonance with then next line. With this factor, one again finds that the mass loss follows a standard CAK scaling relation (► 15.73)

$$\dot{M} = f_{\text{nid}}^{1/\alpha} \dot{M}_{\text{cak}} = f_{\text{nid}}^{1/\alpha} \frac{L}{c^2} \frac{\alpha}{1-\alpha} \left[\frac{\bar{Q}\Gamma}{1-\Gamma} \right]^{(1-\alpha)/\alpha}. \quad (15.92)$$

where $f_{\text{nid}}^{1/\alpha}$ is typically about one half (see ► Fig. 15-6 of Gayley et al. 1995).

The wind velocity law in such multi-scattering models is found to have a somewhat more extended acceleration than in standard finite-disk CAK models, with velocity law indexes of roughly $\beta = 1.5$ – 2 . The terminal speed again scales with the effective escape speed, $v_\infty \approx 3v_{\text{esc}}$.

This analysis shows that, within such “effectively gray” distribution of overlapping lines, multiline scattering can, in principle, yield momentum numbers $\eta \gg 1$ well in excess of the single-scattering limit. In this sense, there is thus no fundamental “momentum” problem for understanding WR winds.

6.3 Wind Momentum–Luminosity Relation for WR Stars

As noted above (► Sect. 5.2), this CAK mass-loss law together with the tendency for the terminal speed v_∞ to scale with the effective escape speed v_e implies the wind momentum–luminosity relation (► 15.75). For WR stars, such comparisons of wind momentum vs. luminosity show a

much greater scatter, but with momenta consistently above those inferred for OB stars, typically reflecting more than a factor 10 higher mass loss rate for the same luminosity (Hamann et al. 1995). If we assume the same $\alpha \approx 0.6$ that characterizes OB winds, (► 15.75) suggests that WR winds must have a line-opacity normalization \bar{Q} that is more than a factor $\approx 10^{\alpha/(1-\alpha)} \approx 30$ higher! Alternatively, this enhanced mass loss could also be obtained through a slightly lower α , representing a somewhat flatter number distribution in line opacity. For example, for stars with Eddington parameter $\Gamma \approx 1/2$ and O-star value for $\bar{Q} \approx 10^3$, simply decreasing from $\alpha = 0.6$ to $\alpha = 0.5$ yields the required factor $\bar{Q}^{1/0.5-1/0.6} \approx 10$ increase in \dot{M} .

While such a modest reduction in α may seem more plausible than a large increase in \bar{Q} , it is generally not clear what basic properties of WR winds could lead to either type of change in the line-opacity distribution. In this context, it thus seems useful here to distinguish the classical *momentum* problem of achieving the large inferred WR momentum numbers, from an *opacity* problem of understanding the underlying sources of the enhanced line opacity needed to drive the enhanced mass loss of WR winds. The wind momentum vs. luminosity relation given in (► 15.75) yields a scaling of wind momentum numbers with $\eta \sim L^{1/\alpha-1} \sim L^{2/3}$, implying that even OB winds should have large momentum numbers, even exceeding the single-scattering limit, for a sufficiently large luminosity. Viewed in this way, the fundamental distinguishing characteristic of WR winds is not so much their large momentum number, but rather their enhanced mass loss compared to OB stars with similar luminosity. Identifying the sources of the enhanced opacity required to drive this enhanced mass loss thus represents a fundamental, unsolved “opacity problem.”

6.4 Cumulative Comoving-Frame Redshift from Multiline Scattering

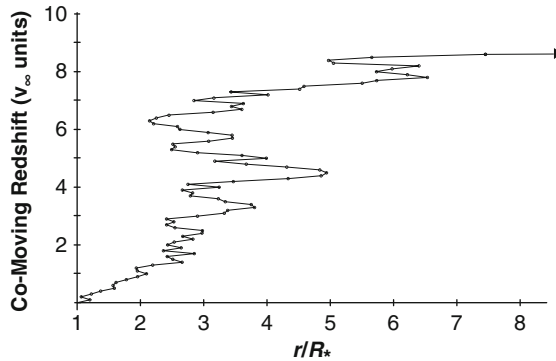
One can also view this multiline scattering that occurs in a WR wind as a random walk over wind *velocity*. To achieve the rms target velocity v_∞ in random increments of Δv requires stepping through $(v_\infty/\Delta v)^2$ lines. Since the expected net redshift from each line interaction is of order $\Delta v/c$, photons undergo a cumulative redshift $\Delta E/E \approx (v_\infty^2/\Delta v^2)\Delta v/c$ over the course of their escape. The associated radial momentum deposition factor is $\eta \approx (\Delta E/E)(c/v_\infty) \approx v_\infty/\Delta v$, as found above.

► Figure 15-16 illustrates the comoving-frame redshift for a typical photon as it diffuses in radius due to multiline scattering. The photon track shown is a characteristic result of a simple Monte Carlo calculation for a wind velocity law $v(r) = v_\infty(1 - R/r)$ and a constant line spacing $\Delta v = v_\infty/10$. Each of the nodes shown represents scattering in one line. In this specific case, the photon escapes only after interacting with nearly 90 lines, resulting in cumulative redshift of nearly $9v_\infty/c$. These are near the statistically expected values of $(v_\infty/\Delta v)^2 = 100$ line scatterings resulting in a total redshift $v_\infty^2/c\Delta v = 10v_\infty/c$, as given by the above random walk arguments.

This systematic photon redshift can also be related to the energy loss – or photon “tiring” – that results from the *work* the radiation does to accelerate the wind to its terminal speed v_∞ . The ratio of wind kinetic energy to the radiative energy represents a “kinetic tiring number”

$$m_{kin} = \frac{\dot{M}v_\infty^2}{2L} = \eta \frac{v_\infty}{2c}. \quad (15.93)$$

Since typically $v_\infty/c < 0.01$, it is seen that for WR winds with momentum numbers of order $\eta \approx 10$, photon tiring is only a $\sim 5\%$ effect. Although thus not of much significance for either OB



■ Fig. 15-16

Simulated photon redshift in the comoving frame as a photon executes a random walk through the wind. The adopted line density is $v_\infty/\Delta v = 10$, which because of the random walk character, implies the photon will interact with roughly 10^2 different lines during escape. Since the comoving redshift between scatterings is $v_\infty/10$, this implies a total comoving redshift of $10 v_\infty$.

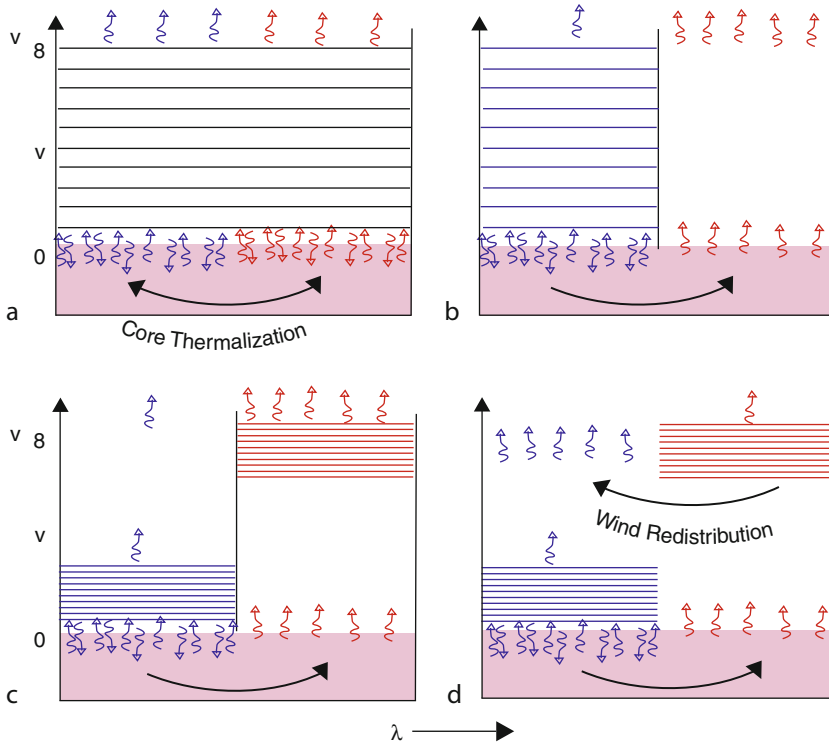
or WR winds, photon tiring does represent a fundamental limit to mass loss, with implications for understanding the giant eruptions from Luminous Blue Variable (LBV) stars.

6.5 Role of Line Bunches, Gaps, and Core Thermalization

An essential complication for developing realistic models of WR winds stems from the inherently non-Poisson character of the spectral-line distributions derived from atomic databases of line lists. At any given wind radius, the dominant contribution to the line opacity stems from a surprisingly small number of specific ionization stages of abundant heavy metals, primarily iron and iron-group elements. Moreover, for any given ion stage, the term “structure” is such that the lines are notably “bunched” into relatively restricted ranges of wavelength. With just a small number of distinct ion stages, the cumulative line spectrum thus likewise exhibits extensive wavelength bunching. Within the gaps between these bunches, the radiation can propagate relatively unimpeded by line scattering, thus representing a potential preferential “leakage” that can significantly reduce the global radiative momentum deposition.

To provide a physical perspective it is helpful to return to the above concept that the line-ensemble constitutes an effective *continuum* opacity, but now allowing this to be frequency dependent to account for the relative bunches and gaps in the spectral distribution of lines. As seen from (► 15.86), the radiative acceleration for such a non-gray opacity depends on the spectral integral of the opacity times the *frequency-dependent* radiation flux. In general, the flux spectrum at any given location in an atmosphere or wind depends on self-consistent solution of a global radiation transport problem, with overall characteristics depending critically on the *thermalization* and frequency *redistribution* properties of the medium.

► Figure 15-17a–d illustrate this key role of photon thermalization and redistribution for simple non-gray line-distributions that are divided locally into two distinct spectral regions, representing either a line “bunch” or “gap.” ► Figure 15-17a first recaps the effectively gray case



■ Fig. 15-17

Schematic diagram illustrating the role of line gaps, line bunches, and photon thermalization for WR wind momentum deposition. The horizontal lines represent the velocity/frequency spacing of optically thick lines in the wind. The four parts represent: (a) an effectively gray model; (b) a wind with fixed ionization and extensive gaps; (c) a wind with ionization stratification that fills gaps; (d) the importance of limiting energy redistribution in the wind

in which the entire spectrum is covered by lines at fixed velocity separation $\Delta v = v_\infty/10$. Since this represents a total effective optical depth $\tau = 10$ for photons to escape from the surface ($v = 0$) to infinity ($v = v_\infty$), the global radiative momentum is $\dot{p}_{rad} = 10L/c$, as follows from (► 15.87).

► Figure 15-17b represents the case when lines have the same concentration $\Delta v = v_\infty/10$ in half the spectrum (the bunch), with the other half completely line-free (the gap). Photons blocked in the bunched region are then rethermalized in the stellar core, and so tend to escape through the gap. Simple statistical arguments show that only about $0.5/(1 + \tau/2) = 1/12$ of the total flux now makes it out through the half of the spectrum covered by the bunch, implying from (► 15.87) that the total radiative momentum is now only $\dot{p}_{rad}/(L/c) \approx 10 \times 1/12 + 0 \times 11/12 = 5/6 < 1$! This roughly represents the circumstance applicable to OB star winds, which have radially constant ionization, and so a radially fixed line spectrum. For such winds, it can be seen that any significant spectral gaps keep the momentum number to near the single-scattering limit $\dot{p}_{rad} \approx L/c$, even if there is very extensive line overlap in spectral line bunches.

► *Figure 15-17c* represents the case wherein there are again $\tau = 10$ lines covering the full spectrum, but now over a limited spatial range, occurring very close to the star in one spectral region, and very far away in the other. In the region where line-blocking occurs near the star, core thermalization again channels photons to the other spectral region. But in this second spectral region, the blocking occurs far away from the star, with greatly reduced probability of backscattering to thermal redistribution in the stellar core. Again assuming purely coherent line scattering in the wind, this means the flux is nearly independent of the outer wind blocking. Since all the stellar flux must diffuse through a layer somewhere with $\tau = 10$ lines, the global momentum deposition is now again, as in case a, simply $\dot{p}_{rad} = 10L/c$. However, note that the flux distribution is still similar to case b, that is, in proportion 1/12 and 11/12 to the spectral regions corresponding respectively to the inner and outer wind blocking. This thus implies the same ratio for the relative deposition of radiative momentum. Overall it can be seen that this example, intended to represent the case of an optically thick WR wind with ionization stratification, does indeed illustrate how it is possible to get a large global momentum deposition even when the line-opacity spectrum is locally divided into gaps and bunches. However, this radiative momentum tends to be deposited more in the outer wind, leaving a net deficit in driving needed to initiate the outflow in the lower wind. This may play a factor in inducing the inferred structure and variability of WR winds.

The final example in ► *Fig. 15-17d* illustrates the crucial importance of the assumption that the radiative transfer within the wind itself is through pure, coherent scattering. If there is significant thermalization or any other type of spectral energy redistribution within the wind itself, then radiation will always tend to be channeled into local gaps, thus again limiting the momentum deposition to a level roughly characterized by the single-scattering limit.

6.6 Summary for Wolf–Rayet Winds

- The classical “momentum problem” (to explain the large inferred ratio of wind to radiative momentum, $\eta \equiv \dot{M}v_{\infty}/(L/c) \gg 1$) is in principle readily solved through multiple scattering of radiation by an opacity that is sufficiently “gray” in its spectral distribution. In this case, one simply obtains $\eta \approx \tau$, where τ is the wind optical depth.
- Lines with a Poisson spectral distribution yield an “effectively gray” cumulative opacity, with multilines scattering occurring when the velocity separation between thick lines Δv is less than the wind terminal speed v_{∞} . In this case, one obtains $\eta \approx v_{\infty}/\Delta v$.
- However, realistic line lists are not gray, and leakage through gaps in the line spectral distribution tends to limit the effective scattering to $\eta \lesssim 1$.
- In WR winds, ionization stratification helps spread line bunches and so fill in gaps, allowing for more effective global trapping of radiation, and thus $\eta > 1$.
- However, photon thermalization can reduce the local effectiveness of line-driving near the stellar core, making it difficult for radiation alone to initiate the wind.
- The relative complexity of WR wind initiation may be associated with the extensive turbulent structure inferred from observed variability in WR wind-emission lines.
- Overall, the understanding of WR winds is perhaps best viewed as an “opacity problem”, that is, identifying the enhanced opacity that can adequately block the radiation flux throughout the wind, and thus drive a WR mass loss that is much greater than from OB stars of comparable luminosity.

References

- Abbott, D. C. 1980, *ApJ*, 242, 1183
- Abbott, D. C. 1982, *ApJ*, 259, 282
- Allen, C. W. 1973, (3rd ed, London: University of London, Athlone Press)
- Biermann, L. 1951, *Kometenschweife und solare Korpuskularstrahlung. Zeitschrift für Astrophysik*, 29, 274
- Bjorkman, J. E., & Cassinelli, J. 1993, *ApJ*, 409, 429
- Breitschwerdt, D., & Komossa, S. 2000, *Galactic fountains and galactic winds. Ap&SS*, 272, 3–13
- Calvet, N. 2004, *Outflows and accretion in young stellar objects*, in *Stars as Suns : Activity, Evolution and Planets*, Proceedings of the 219th Symposium of the International Astronomical Union Held During the IAU General Assembly XXV, Sydney, ed. A. K. Dupree, & A. O. Benz (San Francisco, CA: ASP), 599
- Castor, J. I., Abbott, D. C., & Klein, R. 1975, *ApJ*, 195, 157
- Chamberlain, J. W. 1961, *ApJ*, 133, 675
- Chapman, S. 1961, *Space Astrophys*, 133, 675
- Cohen, D. H. 2008, *IAU Symposium*, 250, 17
- Cox, D. P., & Tucker, W. H. 1969, *ApJ*, 157, 1157
- Cranmer, S. R., et al. 1999, *ApJ*, 511, 481
- Cranmer, S. R. 2000, *ApJ*, 532, 1197
- Cranmer, S. R. 2009, *Coronal holes. Living Rev Solar Phys*, 6, 3. URL: <http://www.livingreviews.org/lrsp-2009-3>
- Cranmer, S. R., & Owocki, S. P. 1996, *Hydrodynamical simulations of corotating interaction regions and discrete absorption components in rotating O-star winds. ApJ*, 462, 469
- Cranmer, S. R., & Owocki, S. P. 1999, *ApJ*, 440, 308
- de Koter, A., Vink, J. S., & Muijres, L. 2008, in *Clumping in Hot-Star Winds*, 47, Proceedings of an International Workshop Held in Potsdam, ed. W. -R. Hamann, A. Feldmeier, L. M. Oskinova (Potsdam). ISBN 978-3-940793-33-1
- Dessart, L., & Owocki, S. P. 2003, *Two-dimensional simulations of the line-driven instability in hot-star winds. A&A*, 406, L1–L4
- Dessart, L., & Owocki, S. P. 2005, *2D simulations of the line-driven instability in hot-star winds. II. Approximations for the 2D radiation force. A&A*, 437, 657–666
- Dupree, A. K. 2004, *Winds from cool stars*, in *Stars as Suns : Activity, Evolution and Planets*, Proceedings of the 219th Symposium of the International Astronomical Union Held During the IAU General Assembly XXV, Sydney, ed. A. K. Dupree, & A. O. Benz (San Francisco, CA: ASP), 623
- Feldmeier, A. 1995, *A&A*, 299, 523
- Friend, D. B., & Abbott, D. C. 1986, 311, 701
- Friend, D. B., & Castor, J. C. 1983, 272, 259
- Fullerton, A. W. 2003, *Cyclical wind variability from O-type stars*, in *Magnetic Fields in O, B and A Stars: Origin and Connection to Pulsation, Rotation and Mass Loss*, Proceedings of the Conference Held 27 November - 1 December, 2002 at University of North-West, Mmabatho, Vol. 305, ed. L. A. Balona, H. F. Henrichs, & R. Medupe (San Francisco: ASP), 333
- Fullerton, A. W., Massa, D. L., & Prinja, R. K. 2006, *The discordance of mass-loss estimates for galactic O-type stars. ApJ*, 637, 1025–1039
- Gayley, K. G. 1995, *ApJ*, 454, 410
- Gayley, K. G., Owocki, S. P., & Cranmer, S. R. 1995, *ApJ*, 442, 296
- Groenewegen, M. A. T., & Lamers, H. J. G. L. M. 1989, *The winds of O-stars. I – an analysis of the UV line profiles with the SEI method. A&AS*, 79, 359
- Groh, J. H., Madura, T. I., Owocki, S. P., Hillier, D. J., & Weigelt, G. 2010, *Is Eta Carinae a fast rotator, and how much does the companion influence the inner wind structure? ApJL*, 716, L223–L228
- Hamann, W.-R. 1996, *Spectral analysis and model atmospheres of WR central stars (Invited paper). Ap&SS*, 238, 31–42
- Hamann, W. R., Koesterke, L., & Wessolowski, U. 1995, *A&A*, 299, 151
- Hamann, W.-R., Gräfener, G., & Koesterke, L. 2003, *WR central stars (invited review)*, in *Planetary Nebulae: Their Evolution and Role in the Universe*, Proceedings of the 209th Symposium of the International Astronomical Union held at Canberra, ed. S. Kwok, M. Dopita, & R. Sutherland (ASP)
- Hamann, W.-R., Feldmeier, A., & Oskinova, L. M. 2008, *Clumping in hot-star winds*, in *Clumping in Hot-Star Winds: Proceedings of an International Workshop Held in Potsdam*, ed. W. R. Hamann, A. Feldmeier, & L. M. Oskinova. ISBN 978-3-940793-33-1
- Hansen, C. J., Kawaler, S. D., & Trimble, V. 2004, *Stellar Interiors: Physical Principles, Structure, and Evolution* (2nd ed.; New York: Springer)
- Hartmann, L., & MacGregor, K. B. 1980, *Momentum and energy deposition in late-type stellar atmospheres and winds. ApJ*, 242, 260–282
- Howarth, I. D., & Prinja, R. K. 1989, *The stellar winds of 203 Galactic O stars – A quantitative ultraviolet survey. ApJS*, 69, 527–592
- Holzer, T. E., & MacGregor, K. B. 1985, *Mass loss mechanisms for cool, low-gravity stars*, in *IN: Mass Loss From Red Giants; Proceedings of the*

- Conference, Los Angeles (Dordrecht: D. Reidel Publishing Co.), 229–255
- Hillier, D. J. 2003, *Advances in modeling of Wolf-Rayet stars*, in *A Massive Star Odyssey: From Main Sequence to Supernova*, Proceedings of IAU Symposium #212, Held 24–28 June 2001 in Lanzarote, Canary island, ed. K. van der Hucht, A. Herrero, & C. Esteban (San Francisco: ASP), 70
- Hillier, D. J., & Miller, D. L. 1999, Constraints on the evolution of massive stars through spectral analysis. I. The WC5 star HD 165763. *ApJ*, 519, 354–371
- Holzer, T. E., Fla, T., & Leer, E. 1983, Alfven waves in stellar winds. *ApJ*, 275, 808–835
- Holzer, T. E., & MacGregor, K. B. 1985, Mass Loss from Red Giants, 117, 229
- Holzer, T. E. 1987, Theory of winds from cool stars, in *IN: Circumstellar Matter; Proceedings of the IAU Symposium, Heidelberg* (Dordrecht: D. Reidel Publishing Co.), 289–305
- Kohl, J. L., Esser, R., Gardner, L. D., Habbal, S., Daigneau, P. S., Dennis, E. F., Nystrom, G. U., Panasyuk, A., Raymond, J. C., Smith, P. L., Strachan, L., van Ballegooijen, A. A., Noci, G., Fineschi, S., Romoli, M., Ciaravella, A., Modigliani, A., Huber, M. C. E., Antonucci, E., Benna, C., Giordano, S., Tondello, G., Nicolosi, P., Naletto, G., Pernechele, C., Spadaro, D., Poletto, G., Livi, S., von der Lühe, O., Geiss, J., Timothy, J. G., Gloeckler, G., Allegra, A., Basile, G., Brusa, R., Wood, B., Siegmund, O. H. W., Fowler, W., Fisher, R., & Jhabvala, M. 1995, The ultraviolet coronagraph spectrometer for the solar and heliospheric observatory. *Solar Phys*, 162, 313–356
- Kohl, J. L., Esser, R., Cranmer, S. R., Fineschi, S., Gardner, L. D., Panasyuk, A. V., Strachan, L., Suleman, R. M., Frazin, R. A., & Noci, G. 1999, *ApJ*, 510, L59
- Koninx, J. -P. M., & Pijpers, F. P. 1992, The applicability of the linearized theory of sound-wave driven winds. *A&A*, 265, 183–194
- Kudritzki, R. P., Lennon, D. J., & Puls, J. 1995, Quantitative Spectroscopy of luminous blue stars in distant galaxies. In *Science with the VLT*, eds. J. R. Walsh & I. J. Danziger (Berlin: Springer), 246
- Lamers, H. J. G. L. M., & Cassinelli, J. 1999, *Introduction to Stellar Winds* (Cambridge/New York: Cambridge University Press)
- Lamers, H. J. G. L. M., Cerruti-Sola, M., & Perinotto, M. 1987, The ‘SEI’ method for accurate and efficient calculations of line profiles in spherically symmetric stellar winds. *ApJ*, 314, 726–738
- Leer, E., & Holzer, T. E. 1979, *Solar Phys.*, 63, 143
- Leer, E., Holzer, T. E., & Fla, T. 1982, Acceleration of the solar wind. *SSRv*, 33, 161–200
- Lépine, S., & Moffat, A. F. J. 1999, *ApJ*, 514, 909
- Li, A., in “The Spectral Energy Distribution of Gas-Rich Galaxies: Confronting Models with Data”, edited by C.C. Popescu & R.J. Tuffs, AIP Conf. Proc. 761, pp. 123–133
- Lucy, L. B. 1982, *ApJ*, 255, 278
- Lucy, L. B. 1984, 284, 351
- MacGregor, K. B., Hartmann, L., & Raymond, J. C. 1979, *ApJ*, 231, 514
- Maeder, A., & Meynet, G. 2000, *ARA&A*, 38, 143
- Maksimovic, M., Issautier, K., Meyer-Vernet, N., Moncuquet, M., & Pantellini, F. 2010, in *Twelfth International Solar Wind Conference*. AIP Conference Proceedings, Vol. 1216
- McComas, D. J., Bame, S. J., Barraclough, B. L., Feldman, W. C., Funsten, H. O., Gosling, J. T., Riley, P., Skoug, R., Balogh, A., Forsyth, R., et al. 1998, Ulysses’ return to the slow solar wind. *GeoRL*, 25, 1–4
- Meyer-Vernet, N. 2007, *Basics of the Solar Wind* (Cambridge, UK: Cambridge University Press). ISBN-10 0-521-81420-0 (HB); ISBN-13 978-0-521-81420-1 (HB)
- Mihalas, D. 1978, (San Francisco: W. H. Freeman and Co.), 650
- Mihalas, D. 1978, *Stellar Atmospheres* (San Francisco: Freeman)
- Mullan, D. J. 1984, Corotating interaction regions in stellar winds. *ApJ*, 283, 303–312
- Newkirk, G. 1967, Structure of the solar corona. *Ann Rev Astron Astrophys*, 5, 213–266
- Oey, M. S., & Clarke, C. J. 2007, Massive stars: feedback effects in the local universe. eprint arXiv:astro-ph/0703036
- Owocki, S. P. 1999, Co-rotating interaction regions in 2D hot-star wind models with line-driven instability, in *IAU Colloq. 169: Variable and Non-spherical Stellar Winds in Luminous Hot Stars*. Lecture Notes in Physics, Vol 523, ed. B. Wolf, O. Stahl, & A. W. Fullerton (Berlin: Springer), 294
- Owocki, S. 2001, Radiatively driven stellar winds from hot stars, in *Encyclopedia of Astronomy and Astrophysics*, ed. P. Murdin, article 1887 (Bristol: Institute of Physics Publishing)
- Owocki, S. 2004, Stellar wind mechanisms and instabilities, in *Evolution of Massive Stars, Mass Loss and Winds*, Held in Aussois and Oléron. EAS Publications Series, Vol. 13. doi:10.1051/eas:2004055
- Owocki, S. P., & Puls, J. 1996, Nonlocal escape-integral approximations for the line force

- in structured line-driven stellar winds. *ApJ*, 462, 894
- Owocki, S. P., & Puls, J. 1999, Line-driven stellar winds: the dynamical role of diffuse radiation gradients and limitations to the Sobolev approach. *ApJ*, 510, 355–368
- Owocki, S. P., & ud-Doula, A. 2004, The effect of magnetic field tilt and divergence on the mass flux and flow speed in a line-driven stellar wind. *ApJ*, 600, 1004–1015
- Owocki, S. P., Castor, J. I., & Rybicki, G. B. 1988, *ApJ*, 335, 914
- Owocki, S. P., Cranmer, S. R., & Blondin, J. 1994, *ApJ*, 424, 887
- Owocki, S. P., Cranmer, S. R., & Gayley, K. G. 1996, *ApJ*, 472, L115
- Owocki, S. P., Gayley, K. G., & Shaviv, N. J. 2004, A porosity-length formalism for photon-tiring-limited mass loss from stars above the Eddington limit. *ApJ*, 616, 525–541
- Owocki, S. P., & Rybicki, G. B. 1984, *ApJ*, 284, 337
- Owocki, S. P., & Rybicki, G. B. 1985, *ApJ*, 299, 265
- Parker E. N. 1963, *Interplanetary Dynamical Processes* (New York: Interscience Publishers)
- Pauldrach, A. W. A., Puls, J., & Kudritzki, R. P. 1986, *A&A*, 164, 86
- Proga, D. 2007, Theory of winds in AGNs, in *The Central Engine of Active Galactic Nuclei*, ASP Conference Series, Vol. 373, Proceedings of the Conference Held 16-21 October, 2006 at Xi'an Jiaotong University, Xi'an, ed. L. C. Ho, & J. -M. Wang (ASP), 267
- Puls, J., Kudritzki, R. P., Herrero, A., Pauldrach, A. W. A., Haser, S. M., Lennon, D. J., Gabler, R., Voels, S., Vilchez, J. M., & Feldmeier, A. 1996, *A&A*, 305, 171
- Puls, J., Markova, N., Scuderi, S., Stanghellini, C., Taranova, O. G., Burnley, A. W. & Howarth, I. D. 2006, Bright OB stars in the Galaxy. III. Constraints on the radial stratification of the clumping factor in hot star winds from a combined H_α , IR and radio analysis. *A&A*, 454, 625–651
- Puls, J., Vink, J. S., & Najarro, F. 2008, Mass loss from hot massive stars. *A&ARv*, 16, 209–325
- Raymond, J. C., Cox, D. P., & Smith, B. W. 1976, *ApJ*, 204, 290
- Scholz, A. 2009, Stellar spindown: from the ONC to the sun, in *Cool Stars, Stellar Systems And The Sun: Proceedings of the 15th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*. AIP Conference Proceedings, Vol. 1094. doi:10.1063/1.3099189
- Smith, N., Davidson, K., Gull, T. R., Ishibashi, K., Hillier, D. J. 2003, *ApJ*, 586, 432
- Sobolev, V. V. 1960, *Moving Envelopes of Stars* (Cambridge, MA: Harvard Univ. Press)
- Spitzer, L. 1962, *Physics of Fully Ionized Gases* (2nd ed; New York: Interscience)
- Townsend, R. H. D., Owocki, S. P., & Howarth, I. D. 2004, Be-star rotation: how close to critical? *MNRAS*, 350, 189–195
- Van Loo, S. 2005, Non-thermal radio emission from single hot stars. PhD thesis, Katholieke Universiteit Leuven
- van Boekel, R., Kervella, P., Schoeller, M., Herbst, T., Brandner, W., de Koter, A., Waters, L., Hillier, D. J., Paresce, F., Lenzen, R., & Lagrange, A. -M. 2003, *A&A*, 410, L37
- von Zeipel, H. 1924, *MNRAS*, 84, 665
- Vink, J. S. 2008, Mass loss and evolution of hot massive stars, in *The Art of Modeling Stars in the 21st Century*, Proceedings of the International Astronomical Union. IAU Symposium, Vol. 252, 271–281. doi:10.1017/S1743921308023016
- Weber, E. J., & Davis, L., Jr. 1967, The angular momentum of the solar wind. *ApJ*, 148, 217–227
- Willson, L. A. 2000, Mass loss from cool stars: impact on the evolution of stars and stellar populations. *ARA&A*, 38, 573–611
- Willson, L. A. 2008, Deathzones and exponents: a different approach to incorporating mass loss in stellar evolution calculations, in *The Art of Modeling Stars in the 21st Century*, Proceedings of the International Astronomical Union. IAU Symposium, Vol. 252, 189–195. doi:10.1017/S1743921308022758
- Withbroe, G. L. 1988, The temperature structure, mass, and energy flow in the corona and inner solar wind. *ApJ*, 325, 442–467
- Withbroe, G. L. 1989, The solar wind mass flux. *ApJ*, 337, L49–L52
- Wood, B.E., 2004, Astrospheres and solar-like stellar winds. *Living Rev Solar Phys*, 1, Irs-2004-2
- Wright, A. E., & Barlow, M. J. 1975, The radio and infrared spectrum of early-type stars undergoing mass loss. *MNRAS*, 170, 41–51
- Zirker, J. 1977, *Coronal Holes & High-speed Wind Streams* (Boulder : Colorado Associated University Press)

Index

A

Absorption, 56–58, 63, 64, 67, 68, 71, 75
Absorption coefficient, 57, 67
Abundances, 15, 19, 28, 29, 31–35, 37, 40–42, 398,
403, 426, 432, 449, 454, 456, 457, 464,
468–473, 478, 479
Accelerated lambda iteration (ALI), 70–75
Accretion, 591–594, 596, 598, 599, 615, 616,
618–631, 633, 635, 637–640, 643,
646–649
Accretion shocks, 298, 318, 327
Active region moss, 161
Active regions (ARs), 93, 97–104, 110, 111, 113, 119,
124, 125, 134, 144, 147, 154, 156,
159–169, 173, 176, 182, 188, 197, 201
Adaptive optics, 663, 666, 684
Ae Herbig stars, 326
ALI. *See* Accelerated lambda iteration (ALI)
Astrometric binaries, 656
Astrometric techniques, 657
ATLAS, 73, 74
Astrophysics, 676–680
Atmospheres, 10, 36, 38–41, 44, 45, 47

B

Balance, 117, 136, 138–140, 145, 168, 188, 190
B stars, 454, 455, 472–473, 475
Binaries, 208, 236–237, 287, 288, 290, 317
 astrometric, 656
 cataclysmic, 658, 688
 close, 619, 634, 644, 645
 eclipsing, 208, 236–237, 657, 673–675, 677
 spectroscopic, 669–673, 676, 678, 684, 685, 687
 visual, 655–657, 659–668, 672, 675, 679–681, 683
 x-ray, 658, 688–689
Binary interactions, 702, 718–723
Binary mergers, 707–708, 718, 721
Binary star orbits, 661
Binary stars, 208, 267
Binary systems, 344, 349, 353, 362–364, 368, 374,
376, 377, 383–385, 387–393
Binary white dwarfs, 577, 579, 604, 607–608
Black holes, 615–649, 698, 712, 714, 715, 717, 718,
721–726, 731
Blandford-Znajek mechanism, 715
Blue supergiants (BSGs), 452, 464, 468, 475–477, 481
Brown dwarf, 246, 247, 250, 251, 256, 257, 259, 260,
262–267, 337–394, 490, 491, 502, 519,
520, 530

desert, 386–388, 392
formation, 342, 353
rotation, 369–371
survey, 338
weather, 374, 381

BSGs. *See* Blue supergiants (BSGs)

BSTAR2006, 74, 75, 79

Butterfly diagram, 116

BY Dra-type stars, 491

C

Cataclysmic binaries, 658, 688
Cepheids, 209–211, 450, 452, 454, 468, 469
Chandrasekhar mass, 561, 563, 569, 572, 573, 696,
698, 699, 708–710, 726, 729
Chemical abundances, 90, 117, 119, 125, 130, 133,
137, 144, 149, 152–159, 167, 188
Chemical evolution, 419, 426, 428, 429
CHIANTI atomic database, 150
Chromosphere, 90, 95–97, 103, 104, 110, 112,
117–124, 127, 129, 130, 136, 156–158,
160, 162, 164, 171–175, 180, 191,
488–503, 509–511, 513, 524, 525, 537,
539, 540, 542, 545
Chromospheric activity, 339, 372, 373, 381
Clouds, 339, 340, 343, 344, 366, 374, 380–381, 392
Clumping, 743, 748–750, 777
Clusters, 236
CMEs. *See* Coronal mass ejections (CMEs)
CMF. *See* Co-moving frame (CMF)
CMFGEN, 62, 73
Collapsar model, 713–717, 725, 727
Collisional ionization, 140
Collision strengths, 139, 149, 151
Co-moving frame (CMF), 768, 773, 782–783
Complete linearization, 69–73
Comptonization, 622–623
Conduction, 10–12, 37, 42, 47
Conductive, 10, 11, 37
Conductivities, 4, 17, 34–37, 47
Continuum emission, 136, 149
Contribution function, 138, 143, 144
Convection, 11, 13–16, 41, 42, 46, 47, 210, 211, 217,
225, 231, 232, 235–237, 239
Convection zone, 96, 125, 129, 131, 134, 135
Convective, 11, 13–16, 31, 33, 41, 46–47
 mixing, 579, 595, 597–598
 overshooting, 233

Corona, 90–92, 94, 95, 97, 98, 100, 103–105, 112, 114, 119, 121, 123, 124, 129, 135–152, 158–170, 172–177, 182, 183, 185, 187–189, 197, 200, 201, 488–492, 498, 500, 502, 503, 511, 513, 515, 517–533, 535–537, 539, 540, 542–543, 545, 546
 Coronagraph, 94, 96, 100, 103, 105, 107, 113, 150, 174, 176, 177, 187, 200
 Coronal abundances, 101, 152–155, 167
 Coronal green line, 95, 96, 147
 Coronal heating, 97, 112, 135, 162–170, 201, 756, 760, 761
 Coronal holes (CHs), 97–99, 105, 119, 122, 123, 144, 149, 154, 159, 160, 162, 163, 187–189, 741, 742, 756, 761, 764
 Coronal mass ejections (CMEs), 91–93, 98, 105, 112, 113, 125, 135, 155, 160, 169–182, 188, 202
 Coronal wind, 738–739, 742, 753, 758, 760–762, 764
 CORONAS, 108–110, 156, 185

D

Dalton minimum, 193
 Debris disks, 591, 598–600, 609
 Delayed lunar occultation interferometry, 668
 Delta Scuti stars, 231–232
 Deuterium fusion, 342
 Dielectronic recombination, 101, 102, 140, 141
 Differential emission measure, 138, 143, 151
 Diffusion, 6, 9–11, 15, 16, 34, 38, 39, 42, 576, 579, 587–600
 Disk, 279–331
 chemistry, 296, 297, 308
 evolution, 281, 289, 290, 299, 305, 310–312, 329
 formation, 281, 290, 298–300
 stability, 303, 310
 winds, 318–320, 322, 324–326
 Disk-star interaction, 302, 324
 Doppler, 504, 505, 509–511, 513, 514, 517
 blue-shift, 746
 effect, 746, 765
 imaging, 504, 505, 509, 511, 517
 red-shift, 766
 shift, 744–745, 773, 776, 779
 Dust, 281, 283, 286, 289–292, 294–297, 304, 306–308, 310, 339, 347, 357, 366, 367, 379–381, 386, 393
 Dwarf galaxies, 398
 Dynamo, 488, 489, 491, 498, 502, 506, 509, 512, 521, 533, 539

E

Eclipsing binaries, 657, 673–675, 677
 Effective temperature, 55, 61, 73, 75, 82, 414
 surface gravity, 61, 82

Electron density, 101, 115, 136–138, 140, 141, 145–150, 160, 169, 180
 Electron temperature, 115, 148–150, 168, 186
 Emission, 57, 58, 60, 67, 68
 coefficients, 56, 58, 67
 infra-red, 750
 line, 739, 741, 742, 748–749, 755, 759, 778
 measure, 108, 138, 142–146, 151, 154, 159, 162, 167, 181
 measure loci, 144
 radio, 750
 Equation of state, 4–7, 12, 14, 16–25, 34, 47
 Escape speed, 738, 752, 753, 760, 762, 763, 775, 777, 781
 EUV. *See* Extreme ultraviolet (EUV)
 Evolution, 3, 7, 14, 15, 23, 29, 44–46, 208, 455, 457–464, 468, 470, 472, 474, 481
 Evolutionary, 3, 5, 6, 14, 30, 32, 37, 38, 43–47
 Exoplanets, 490, 528, 535, 543–547
 Extreme Ultraviolet (EUV), 490, 491, 498, 511, 524, 526, 528, 532, 535, 537, 545, 546
 emission, 537

F

F-corona, 94
 FGS. *See* Fine guidance sensors (FGS)
 Filaments, 98, 119, 123, 148, 171, 173, 175–178
 Fine guidance sensors (FGS), 663, 666–668, 679
 FIP effect, 154, 155, 158, 159
 Flares, 489–492, 510, 515, 520, 521, 523, 525, 526, 528–531, 536–543, 545
 Flash spectrum, 95, 118
 f-modes, 130
 Fragmentation, 677–679, 686
 Fraunhofer, 94, 95, 152
 Frequency analysis, 220

G

Galactic cosmic rays (GCRs), 185, 189, 190, 194
 Gamma-ray bursts (GRBs)
 long-duration, 712–714, 717, 726
 short-duration, 712, 717
 GCRs. *See* Galactic cosmic rays (GCRs)
 GI 229B, 338, 339, 379, 386
 Globular clusters, 398, 432
 g-modes, 130
 Granules, 116, 117
 Gravitational settling, 563, 565, 587–589, 598
 Gravity acceleration at the stellar surfaces, 55
 Gravity darkening, 775–777
 GRBs. *See* Gamma-ray bursts (GRBs)
 Group sunspot number (GSN), 115, 191, 193
 GSN. *See* Group sunspot number (GSN)

H

HAeBe stars. *See* Herbig AeBe stars
 Helioseismology, 92, 103, 105, 130–135, 152, 157
 Heliosphere, 91, 102, 113, 160, 174, 185, 188, 189, 199–201
 Herbig AeBe stars, 290–292, 295, 296, 306, 326
 Herbig-Haro objects (HHOs), 282, 290, 294–296, 314, 316, 318
 Hertzsprung-Russell (HR) diagram, 209, 211, 224, 230, 232, 235, 236
 HHOs. *See* Herbig-Haro objects (HHOs)
 High-speed streams, 742, 761
 Hinode, 90, 99, 110–113, 115–120, 124, 125, 147, 161, 162, 165, 169, 172, 174, 187, 199
 Hydrodynamic models, 121, 166–168
 Hydrodynamics, 227
 Hydrogen-burning limit, 339, 342, 344, 349, 351, 366, 369, 376, 386, 388, 389
 Hydrostatic models, 165–166
 Hypernovae, 695, 710–718, 726

I

IFMR. *See* Initial-final mass relation (IFMR)
 IMF. *See* Initial mass function (IMF)
 Initial-final mass relation (IFMR), 584–587, 604, 605, 607
 Initial mass function (IMF), 247, 259–263, 265, 266, 275, 348, 350, 394
 Interacting binaries, 686–689
 Interiors, 3, 4, 9, 22, 25, 26, 28, 34, 38
 Intermediate mass black holes, 645–648
 InterStellarMedium (ISM), 282, 290, 294–296
 Ionization, 118, 136–138, 140, 142, 145, 147, 149, 151, 153, 157–159, 168, 199
 ISM. *See* InterStellarMedium (ISM), 295

J

Jets, 313, 314, 316–318, 320–326, 329, 331

K

K-corona, 94

L

LBVs. *See* Luminous blue variables (LBVs)
 LDI. *See* Line-deshadowing instability (LDI)
 Levitation, 6, 38, 39
 Line, 209, 213, 215, 219, 221–223, 225–228, 230, 234 profiles, 226
 Line-deshadowing instability (LDI), 772, 773
 Line-driving, 740, 767, 769, 770, 772, 773, 775–777, 785
 Line flux, 137–138

Line widths, 112, 115, 150, 162, 165
 Lithium fusion, 342, 343, 366
 Local thermodynamic equilibrium (LTE), 54, 59–61, 64, 65, 70, 73–74, 83
 Loops, 97, 99, 121, 123–125, 135, 144, 153, 154, 160–162, 164–174, 176, 182
 LTE. *See* Local thermodynamic equilibrium (LTE)
 Luminosity function, 350, 351, 353, 360, 362, 382
 Luminous blue variables (LBVs), 452, 454, 456–457, 467, 468, 470, 481, 738, 739, 783

M

Magnetar, 698, 712, 714, 725, 730
 Magnetic, 488–492, 495, 498–500, 502, 505–509, 511, 512, 515–518, 520, 523, 526–533, 536, 537, 539, 540, 542–547
 Magnetic fields, 219, 230, 234, 239, 281, 282, 298, 301, 305, 309, 318, 320, 322–326, 488–490, 498, 502, 505, 506, 509, 515, 518, 521, 526, 528–531, 536, 537, 539, 544, 545, 561, 592, 600–605, 609
 Magnetospheres, 283, 306, 319, 320, 324, 326–329, 331
 MARCS, 73
 Mass, 53–55, 57, 59, 61, 66, 68, 74, 79, 82
 Mass accretion, 703, 709, 718, 720
 Masses techniques, 655
 Massive stars, 448–452, 454–457, 462, 464, 466–468, 470–472, 474, 478, 480, 481
 Mass loss, 448, 450, 456, 457, 459, 464, 466–468, 472, 474, 476–478, 480, 481
 Mass loss rate, 738, 739, 742, 743, 745, 747–750, 753, 758, 759, 761–764, 767, 769–772, 774, 775, 777, 781, 782
 Mass-luminosity relation (MLR), 668, 676–677, 682, 684
 Mass-radius relation (MRR), 561, 563, 569, 572–574, 577, 582–583, 586, 609, 676–677
 Mass transfer, 703, 707, 718–722 cases of, 719
 Maunder minimum, 93, 193, 195
 Meridional flow, 131, 132
 Metallicity, 398, 406, 417, 425, 426, 440, 709, 717, 718, 722, 726–729
 Metal line-blanketed, 74
 Metal line blanketing, 83
 Metal lines, 63, 64, 74, 75, 82, 83
 Methane, 339, 356, 357, 359, 361, 362, 365, 379, 380, 383, 386
 Methods, 209, 219–224, 226, 227, 232, 233, 237 data analysis, 237
 Microflares, 108, 112, 164, 169–170
 MLR. *See* Mass-luminosity relation (MLR)
 Mode identification, 224–226, 229, 233, 234, 236, 238

Model atmospheres, 52, 61–83
 Models, 212, 215, 216, 218, 219, 222, 227, 233, 236
 Model stellar atmospheres, 52, 58, 61, 63, 66, 71–82
 Molecular clouds, 245–247, 249–254
 Molecular cores, 261
 Molecular gas, 284, 289, 296–297, 328, 330, 331
 Molecular outflows, 312, 314
 MRR. *See* Mass-radius relation (MRR)

N

Nanoflares, 169–170
 Neutrinos, 4, 6, 18, 37, 38, 42, 47, 195, 196, 697, 698, 703, 705, 716, 723, 731
 Neutron stars, 615–649, 697, 698, 703, 705, 710–712, 714, 716–718, 721–725, 730, 731
 Non-LTE (NLTE), 54, 58–61, 64, 65, 71, 73–75, 81, 83
 Nonthermal velocities, 112, 150, 151
 Normal modes, 212, 229, 238
 Nuclear reactions, 4, 6, 8, 9, 13, 17, 25–34, 42, 46, 47, 211, 231
 Nucleosynthesis, 398–400, 402, 408, 410, 417, 418, 421–430, 432, 441, 442, 449, 450, 456
 Numerical methods, 52, 53, 58, 69–74

O

OB stars, 455, 456, 462, 472
 Opacity, 4, 6, 10, 11, 14, 17, 34–37, 40, 45, 47
 bound-bound (line), 742, 766
 bound-free, 742
 dust, 742, 743
 free-free, 742
 Optical depth, 742–743, 745, 746, 750, 767–769, 772, 779–781, 784, 785
 Optical interferometry, 663, 665–666, 668
 Oscillations, 210–214, 217, 219, 221, 222, 224, 226, 227, 231, 232, 234–238
 OSTAR2002, 74–78
 grid, 74, 75

P

P-Cygni profile, 745–748
 PDRs. *See* PhotoDissociation Regions (PDRs)
 PHOENIX, 63, 73, 74
 Photodissociation, 295, 296, 306, 309
 PhotoDissociation Regions (PDRs), 304
 Photometry, 209, 219, 223, 225–227, 233–238
 Photospheres, 55, 56, 61, 62, 64, 65, 91, 94–99, 105, 110, 112, 115–119, 122, 125, 127, 129, 130, 135, 153, 154, 157, 164, 165
 Photospheric abundances, 134, 152–154, 156, 158, 159
 Physical data and processes, 6, 18, 38–40, 47

Plages, 489, 492, 493, 495, 509–512, 543
 p-modes, 121, 128, 130
 PMS. *See* Pre-main sequence (PMS) star
 PMS evolution. *See* Pre-main sequence (PMS) evolution
 Populations, 398
 Pre-main sequence (PMS) evolution, 281–328
 Pre-main sequence (PMS) star, 246, 254, 281–328
 Pressure
 Alfvén wave, 762–763, 766
 gas, 738, 739, 752, 758, 759, 761, 762, 764, 765, 769, 772, 777
 MHD wave, 762
 radiation, 751, 765
 Prominences, 92, 94, 95, 118, 119, 121, 160, 173, 175, 177
 Prompt fragmentation, 677, 678
 Protostars, 246, 247, 254–256, 258, 263, 268
 Protostellar disks, 279–331
 evolution, 281, 289, 290, 299, 305, 310–312, 329
 model, 304
 Pulsars, 615, 628, 633, 636, 637, 641–645, 648, 649
 general, 615, 633, 636, 649
 Pulsations, 209–217, 219, 221–227, 229, 231–239

Q

Quiet Sun (QS), 91, 97, 98, 114, 116, 119, 122, 124, 143, 151, 154, 157, 159, 160, 162, 163, 197

R

Radial velocities, 219, 221, 225, 226, 233, 234, 237, 239
 Radiation, 281, 282, 284, 289, 290, 292, 293, 296, 298, 300–312, 314, 318, 320, 322, 324–327, 329
 Radiative loss, 124, 149, 166, 167
 Radiative recombination, 140, 149
 Radiative transfer, 10, 41, 53, 56, 61, 63, 65–67, 73
 Radiative transfer equation, 56, 61, 65–67
 Radiative zone, 131
 Radio, 489–491, 509, 518, 528–531, 537, 539–542
 emission, 491, 518, 528–531, 539–541
 Ramaty high-energy solar spectroscopic imager (RHESSI), 108, 109, 152, 169, 171
 Red supergiants (RSGs), 450, 452, 464, 467, 468, 470, 474–478
 Relativistic beaming, 712–713
 Relativistic fireball, 713–714
 Relativistic jets, 624–625, 631, 638
 RHESSI. *See* Ramaty high-energy solar spectroscopic imager (RHESSI)
 RLOF. *See* Roche-lobe overflow (RLOF)
 Roche-lobe overflow (RLOF), 718–720

Rotation, 218, 219, 448, 449, 454, 455, 457–466, 468,
470–472, 474, 475, 480, 481, 488, 491,
495, 498, 499, 503–505, 508–512, 517,
518, 520–522, 527, 530, 532–535, 537,
543, 545, 705, 709, 714, 715, 717, 725,
727, 728, 739, 772, 774–777, 780
RS CVn-type stars, 491, 504–507, 510, 511, 515, 518,
520, 524, 527–530, 537–540, 542
RSGs. *See* Red supergiants (RSGs)

S

Satellites, 3, 75, 90, 100–103, 110, 112, 127, 141, 159,
170, 172, 178, 180, 181, 184–186, 190,
192, 196, 197, 199, 200, 237, 238, 356,
373, 398, 494, 500, 502, 507, 510, 512,
518, 519, 526, 542, 563, 655, 662, 666,
668, 681, 683, 705, 711, 713, 741, 761, 775
Scale height, 751, 754, 757, 760–762
Scaling law, 165–166
Scattering
dust, 739, 742, 743
electron, 739, 742–745, 765–767, 769, 778
line, 739–742, 745–748, 765–767, 772, 782,
783, 785
multiline, 740, 778–785
SDO. *See* Solar dynamics observatory (SDO)
SEP. *See* Solar energetic particle (SEP)
Shocks, 286, 298, 303, 304, 314, 316–318, 320,
326, 327
Skylab, 98–100, 121–124, 145, 151, 154, 159–161,
165, 166, 174, 187
SMM. *See* Solar maximum mission (SMM)
SN. *See* Supernovae (SN)
Sobolev
approximation, 745, 747, 748, 767–769, 773
length, 744, 745, 768, 772, 773
SoHO. *See* Solar and Heliospheric Observatory
(SoHO)
Solar dynamics observatory (SDO), 90, 114, 115,
127, 134, 136, 199
Solar dynamo, 135
Solar energetic particle (SEP), 152, 155, 156, 174,
190, 200
Solar flares, 91, 93, 98–100, 102, 108–110, 155, 156,
160, 169–182
Solar and Heliospheric Observatory (SoHO), 90, 91,
95, 103–109, 112, 113, 115, 116, 121–124,
127, 131–134, 147, 149, 150, 152, 154,
155, 157, 159–164, 174–177, 185–188,
191, 197–199
Solar irradiance, 127, 189–199
Solar maximum mission (SMM), 100–102, 154–156,
162, 174, 190, 191
Solar orbiter, 199, 200
Solar oscillations, 126–130

Solar Probe+, 199–201
Solar rotation, 116, 130, 132, 135, 185
Solar TERrestrial Relations Observatory (STEREO),
90, 113, 115, 176, 186
Solar wind (SW), 91, 92, 102, 103, 113, 125, 135, 152,
153, 155, 156, 158, 182–189, 200, 201,
737–743, 752–765, 772
Sound speed, 751–753, 757, 758, 760, 762, 763, 769,
770, 772
Speckle, 661–666, 676, 681, 682, 684
Spectra, 53, 61, 73
Spectral classification, 355, 356, 394
Spectroscopic, 656, 657, 665, 667, 669–673, 675–682,
684, 685, 688
Spectroscopic binaries, 669–673, 676, 678, 684,
685, 687
Spectroscopic diagnostics, 52
Spectroscopic filling factors, 147–148
Spectroscopy, 219, 225–227, 233–238
Spherical harmonics, 213, 234, 2218
Spicules, 105, 118, 119, 121, 123
Standard solar model, 125, 126, 133, 153
Stars, 243–275
atmospheres, 38, 231
chemically peculiar, 232
clusters, 245–248, 252, 260, 266–275, 657,
672, 689
early-type, 233
evolution, 3, 17, 44, 208, 212, 217, 221, 230, 233
formation, 243–275, 678, 684, 686
fundamental parameters, 3, 4, 217, 236
horizontal branch, 209
interiors, 3, 4, 208, 211, 212, 219, 229, 230, 232,
233, 239
magnetic fields, 234
oscillations (including pulsations), 201–212,
214, 217, 219, 221, 223, 230–238
premain sequence, 232
rotation, 213, 218, 219, 224, 229, 233, 234
structure, 3, 4, 40–42, 44
Starspot, 489, 491, 492, 503–509, 534, 537, 545
Statistical equilibrium, 138
STEREO. *See* Solar TERrestrial Relations
Observatory (STEREO)
Structures, 1–47, 212, 218, 229, 232
Structure, stellar, 1–47
Supergranular pattern, 117
Supernovae (SN), 448, 456, 464, 465, 468, 476, 481
classification, 699–705
core-collapse, 696–700, 702–705, 713, 716, 722,
730, 731
electron-capture, 722–725, 730, 731
kicks, 724, 731
lightcurves, 698, 700–705, 708, 716, 729
pair-instability, 725–729
superluminous, 729–730

supernova 1987A (SN 1987A), 695, 700, 705–708
 thermonuclear, 698–700, 726
 type Ia supernova (SN Ia)
 cosmological distance candles, 658, 708–709
 double-degenerate progenitors, 709, 710
 single-degenerate progenitors, 710
 type II supernova (SN II), 699, 701, 702
 Surface gravity, 61, 82

T

Techniques

 photometric, 226
 radial velocity, 226, 237
 spectroscopic, 226, 233, 234
 Terminal speed, 738, 745, 747, 749, 750, 763,
 769–772, 775, 780–782, 785
 Thermal conduction, 754, 756, 757, 764
 Thermal speed, 744, 756, 768, 769, 773
 Thermonuclear reaction, 4, 17, 26–28
 TLUSTY, 69, 73, 74
 TMAP, 73, 74
 TRACE. *See* Transition region and coronal explorer
 (TRACE)
 Transition, 488–491, 498–503, 509–518, 525, 527,
 530, 533, 542
 Transition region, 90, 97, 99, 103, 104, 108, 122–125,
 129, 136, 148, 158, 165, 180, 197,
 488–491, 498–503, 509–518, 527, 530,
 533, 542
 Transition region and coronal explorer (TRACE),
 90, 99, 107, 108, 115, 125, 159–162, 164,
 169, 170, 172, 200
 T Tau/T Tauri stars (TTs), 281, 282, 287, 289–292,
 294–296, 300, 306, 307, 309, 310, 317,
 318, 320, 324, 326–328
 TTs. *See* T Tau/T Tauri stars (TTs)
 TTs magnetospheres, 327, 328
 Turbulence, 245, 247, 248, 252–254, 261–263,
 265, 268
 Type I supernovae. *See also* Supernovae (SN), 658

U

Ultracool dwarf, 339, 342, 352, 353, 356, 358–361,
 363–365, 367–369, 371, 373, 374, 376,
 378, 380, 385, 386, 388–390, 393
 Ultracool pulsar, 374
 Ultracool subdwarf, 368–369

Ultraviolet (UV) Cet-type stars, 491, 530, 537,
 538, 540
 Ultraviolet (UV) emission, 498, 511, 533, 537
 Ultraviolet (UV) radiation, 282, 296, 303–309, 311,
 325, 326, 329
 Ulysses, 91, 102, 155, 156, 158, 185, 188–189
 UV radiation. *See* Ultraviolet (UV) radiation

V

Variables, other, 208
 Variable stars, 208, 209
 Variable white dwarfs, 210, 604–605
 Visual binaries, 655–657, 659–668, 672, 675,
 679–681, 683
 Visual orbit, 656, 659–662, 665, 666, 670, 673, 675,
 676, 679

W

White dwarfs, 561–609
 evolution, 574, 576, 577, 580, 581, 584, 586, 603,
 608, 609
 spectra, 562, 563, 576, 590, 593–597
 Wind-momentum, 771, 777, 778, 780–782, 784
 Winds, 489, 490, 492, 500, 502, 517, 531–532
 Wolf-Rayet (WR) star, 449, 452, 454–457, 459, 464,
 470, 478–481, 739, 740, 743, 749, 778
 WR star. *See* Wolf-Rayet (WR) star

X

X-rays, 489–491, 498, 500, 502, 503, 511, 516–546
 binaries, 615, 619–632, 634–640, 643–649, 658,
 688–689
 bursts, 627, 628, 635, 636, 638, 647
 emission, 500, 503, 518–528, 530, 537, 539, 540,
 542, 545

Y

Yohkoh, 90, 91, 98, 102–103, 154, 156, 159–161, 168,
 169, 178, 199
 Young stellar objects (YSOs), 279–331
 disk, 293, 294, 296, 303, 304, 310, 313, 319
 jets, 317
 X-ray, 305–306
 YSOs. *See* Young stellar objects (YSOs)