

Second
Edition

I. Neill Reid
Suzanne L. Hawley

New Light on Dark Stars

**Red Dwarfs,
Low-Mass Stars,
Brown Dwarfs**



 Springer

PRAxis[®]

New Light on Dark Stars

Red Dwarfs, Low-Mass Stars, Brown Dwarfs

I. Neill Reid and Suzanne L. Hawley

New Light on Dark Stars

Red Dwarfs, Low-Mass Stars, Brown Dwarfs



Published in association with
Praxis Publishing
Chichester, UK



Dr I. Neill Reid
Space Telescope Science Institute
Baltimore
Maryland
USA

Professor Suzanne L. Hawley
Professor, Department of Astronomy
University of Washington
Seattle
USA

SPRINGER-PRAXIS BOOKS IN ASTROPHYSICS AND ASTRONOMY
SUBJECT ADVISORY EDITOR: John Mason B.Sc., M.Sc., Ph.D.

ISBN 3-540-25124-3 Springer-Verlag Berlin Heidelberg New York

Springer is part of Springer-Science + Business Media (springeronline.com)
Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available from the Internet at <http://dnb.ddb.de>

Library of Congress Control Number: 2005924587

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

© Praxis Publishing Ltd, Chichester, UK, 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, 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.

Cover design: Jim Wilkie
Project management: Originator Publishing Services, Gt Yarmouth, Norfolk, UK

Printed on acid-free paper

In memory of Iain D. Reid and Elizabeth N. Reid

Contents

Preface to the second edition	xv
Preface to the first edition	xvii
List of figures	xxi
List of tables	xxxii
Acknowledgements	xxxiii
1 Astronomical concepts	1
1.1 Introduction	1
1.2 The electromagnetic spectrum	1
1.3 Positional astronomy	3
1.3.1 Co-ordinate systems	4
1.3.2 Stellar astrometry	8
1.3.3 Stellar kinematics	10
1.4 Telescopes and detectors	12
1.4.1 Telescopes	12
1.4.2 Detectors	15
1.5 Stellar photometry	18
1.5.1 The magnitude system	19
1.5.2 Measuring magnitudes	21
1.5.3 Bolometric magnitudes and effective temperatures	23
1.5.4 Interstellar absorption	24
1.6 Sky surveys	25
1.6.1 Photographic surveys	25
1.6.2 Infrared surveys	26

1.6.3	Deep optical imaging surveys	29
1.6.4	X-ray surveys	29
1.6.5	Digitised surveys and virtual observatories	30
1.7	Spectrographs and spectroscopy	31
1.8	Improving image quality	35
1.8.1	Speckle interferometry	37
1.8.2	Adaptive optics and interferometry	38
1.9	Summary	40
1.10	References	40
1.11	Homework problems	42
2	Observational properties of low-mass dwarfs	45
2.1	Introduction	45
2.2	Spectral classification: What are M, L and T dwarfs?	46
2.2.1	Early observations	46
2.2.2	An M dwarf classification system	48
2.2.3	Beyond M: Spectral type L	52
2.2.4	... and spectral type T	55
2.2.5	... and carbon dwarfs	56
2.2.6	Near-infrared spectroscopy	58
2.2.7	Mid-infrared and longer wavelengths	62
2.2.8	Spectroscopic dwarf/giant indicators	64
2.3	Broadband photometry of M dwarfs	66
2.3.1	Spectral energy distributions, temperatures and radii	66
2.3.2	Empirical colour–magnitude diagrams	70
2.4	Abundance effects on the H–R diagram	74
2.4.1	Spectroscopic bandstrengths, absolute magnitudes and abundance	75
2.4.2	Effects on broadband photometry	77
2.5	Activity	79
2.5.1	Chromospheric activity	80
2.5.2	The corona	83
2.5.3	Flares	84
2.6	Masses of M dwarfs	86
2.7	Stellar statistics	88
2.7.1	Binary and multiple star systems	88
2.7.2	The stellar luminosity and mass functions	92
2.8	Summary	94
2.9	References	95
2.10	Homework problems	99
3	The structure, formation and evolution of low-mass stars and brown dwarfs	101
3.1	Introduction	101
3.2	A brief history	101

3.3	General principles of stellar structure	104
3.3.1	The fundamental equations	104
3.3.2	Gas polytropes and the Lane–Emden equations	105
3.3.3	Energy generation	108
3.3.4	Energy transport	111
3.3.5	The equation of state	113
3.4	Models of low-mass dwarfs	115
3.4.1	The development of low-mass models	115
3.4.2	The minimum mass for hydrogen burning	118
3.4.3	Properties of current models of low-mass dwarfs	122
3.5	Matching models and observations	125
3.5.1	The Hertzsprung–Russell diagram	126
3.5.2	The mass–luminosity relationship	131
3.5.3	The mass–radius relationship	132
3.6	The formation of low-mass stars	134
3.6.1	Molecular clouds: the initial phase	134
3.6.2	Brown dwarf formation mechanisms	136
3.6.3	Young stellar objects in the field	137
3.7	Post-main sequence evolution and the death of the Milky Way	139
3.8	Summary	142
3.9	References	143
3.10	Homework problems	145
4	The photosphere	147
4.1	Introduction	147
4.2	Historical perspective	148
4.3	Theoretical considerations	153
4.4	Treatment of opacity	157
4.5	Photospheric models	161
4.6	Photospheric diagnostics	165
4.6.1	The comparison with observations	165
4.6.2	Temperature determinations	167
4.6.3	Gravity and radius determinations	172
4.6.4	Metallicity	174
4.7	Dusty atmospheres in the lowest-mass stars and brown dwarfs	178
4.8	Summary	181
4.9	References	181
4.10	Homework problems	182
5	Stellar activity	185
5.1	Introduction	185
5.2	The magnetic field	186
5.2.1	Dynamo theory	186
5.2.2	Magnetic field observations	189
5.3	The outer atmosphere	194

5.3.1	The chromosphere.	195
5.3.2	The transition region.	196
5.3.3	The corona	198
5.4	Spots and flares	201
5.4.1	Spots and spot cycles.	202
5.4.2	Flares	206
5.5	Global properties of chromospheric activity in M dwarfs	213
5.5.1	Activity effects on colours, magnitudes and bandstrengths	213
5.5.2	Empirical relationships between age and activity	216
5.5.3	Emission in other wavelength regions	220
5.5.4	The connection between rotation and activity	223
5.6	Activity properties of the lowest-mass stars and brown dwarfs	224
5.6.1	H α emission in VLM dwarfs.	225
5.6.2	Spots and flares	226
5.7	Compilation and synthesis of activity results.	228
5.8	Summary	229
5.9	References	230
5.10	Homework problems	234
6	Brown dwarfs	237
6.1	Introduction	237
6.2	How to recognise a brown dwarf	238
6.3	First steps: some interesting mistakes.	240
6.3.1	VB 8B – the disappearing brown dwarf	240
6.3.2	G 29-38 – dust or a brown dwarf?	241
6.4	First generation surveys for brown dwarfs	244
6.5	Brown dwarfs revealed	246
6.5.1	GD 165B	246
6.5.2	HD 114762B	248
6.5.3	Gl 229B	250
6.5.4	PPI 15	252
6.5.5	Field brown dwarfs	253
6.6	Second generation surveys for brown dwarfs.	254
6.6.1	L dwarfs in the field	254
6.6.2	T dwarfs in the field	256
6.7	Characteristics of L and T dwarfs.	257
6.7.1	Photometric properties and the H–R diagram	257
6.7.2	Bolometric corrections and the temperature scale	259
6.7.3	Atmospheres, dust and the L/T transition	264
6.7.4	Gravities and masses	269
6.7.5	Activity and rotation	271
6.8	Brown dwarfs as companions.	275
6.8.1	Brown dwarf companions of main sequence stars	276
6.8.2	Ultracool binary systems	277
6.9	Beyond T: Y dwarfs and the future	281

6.10	Summary	283
6.11	References	283
6.12	Homework problems	287
7	A Galactic structure primer	289
7.1	Introduction	289
7.2	Stellar populations	290
7.3	The Galactic Halo	294
7.3.1	The metal-poor main sequence	296
7.3.2	The density distribution	298
7.3.3	Halo kinematics	301
7.3.4	The abundance distribution of the halo	302
7.3.5	The age of the Galactic Halo	308
7.3.6	Forming the halo	312
7.4	The Galactic Bulge	314
7.5	The Galactic Disk	316
7.5.1	The kinematics of disk stars	316
7.5.2	The disk density distribution	320
7.5.3	The abundance distribution	323
7.5.4	The age of the Galactic Disk	324
7.5.5	The star-formation history of the disk	326
7.6	A quick tour of the wider Solar Neighbourhood	328
7.7	Summary	332
7.8	References	333
7.9	Homework problems	337
8	The stellar luminosity function	339
8.1	Introduction	339
8.2	The early years	339
8.2.1	The method of mean parallaxes	341
8.2.2	The method of trigonometric parallaxes	342
8.2.3	The method of mean absolute magnitudes	344
8.2.4	Kuiper's nearby-star luminosity function	348
8.3	The 1970s: A plethora of M dwarfs?	349
8.4	Systematic biases	354
8.4.1	Lutz–Kelker corrections	354
8.4.2	Malmquist bias	357
8.5	Modern times I: Field-star surveys	360
8.5.1	A census of the Solar Neighbourhood	360
8.5.2	Photometric surveys	365
8.6	Modern times II: The digital era	369
8.6.1	M dwarfs	370
8.6.2	Ultracool M and L dwarfs	372
8.6.3	T dwarfs	375
8.7	Specialised survey methods	375

8.7.1	M dwarfs in front of dark clouds	375
8.7.2	M dwarf companions to white dwarfs	376
8.8	M dwarfs in open clusters	378
8.8.1	Cluster membership criteria	379
8.8.2	Dynamical effects and the present-day luminosity function in clusters	383
8.9	$\Phi(M)$ for nearby clusters	385
8.9.1	The Hyades	385
8.9.2	The Pleiades	388
8.9.3	Praesepe	391
8.9.4	α Persei	392
8.9.5	Summary	393
8.10	The luminosity function of the Galactic Disk	393
8.11	References	395
8.12	Homework problems	400
9	The mass function	403
9.1	Introduction	403
9.2	Defining the mass function	404
9.2.1	Basic terminology	404
9.2.2	The Salpeter mass function	405
9.2.3	The mass function for intermediate- and high-mass stars	406
9.3	Measuring stellar masses	408
9.3.1	Astrometric binaries	408
9.3.2	Spectroscopic binaries	411
9.3.3	Mass determination from gravitational lensing	413
9.3.4	Summary	416
9.4	Mass–luminosity relationship	416
9.4.1	The mass–luminosity relation for field stars	417
9.4.2	The mass–luminosity relation for pre-main sequence stars	422
9.5	The field-star mass function	424
9.5.1	Statistical analyses of $\Psi(M)$	424
9.5.2	The nearby stars	426
9.5.3	The substellar mass function in the field	427
9.5.4	The initial mass function of the Galactic Disk	428
9.6	Mass functions of open clusters	432
9.6.1	The Hyades, Praesepe and older open clusters	433
9.6.2	The Pleiades and intermediate-age open clusters	433
9.7	Protoclusters and star-forming regions	436
9.7.1	Infrared imaging surveys	437
9.7.2	H–R diagram analyses	438
9.7.3	Two case studies: IC 348 and the Orion Nebula Cluster	438
9.7.4	The IMF in young clusters and associations	443
9.8	Theoretical interpretations of the IMF	447
9.9	The star-forming mass distribution and the local mass density	449

9.10	Summary	450
9.11	References	451
9.12	Homework problems	454
10	M dwarfs in the Galactic Halo	457
10.1	Introduction	457
10.2	Halo subdwarfs in the field	458
10.2.1	Photometric properties of M subdwarfs	458
10.2.2	Searching for local subdwarfs	459
10.2.3	The reduced proper-motion diagram	462
10.3	Classifying late-type subdwarfs	463
10.3.1	Cool subdwarfs	463
10.3.2	Ultracool subdwarfs	466
10.4	The field subdwarf luminosity function	468
10.4.1	Schmidt's derivation of the halo luminosity function	470
10.4.2	$\Phi_{\text{halo}}(M_V)$ from proper-motion star samples	472
10.4.3	Φ_{halo} from star-count analyses	474
10.4.4	The field halo luminosity function	475
10.5	Luminosity function in globular clusters	476
10.6	The mass function of the stellar halo	481
10.6.1	$\Psi_{\text{halo}}(M)$ in the field	481
10.6.2	$\Psi_{\text{halo}}(M)$ in globular clusters	484
10.6.3	Summary	486
10.7	M subdwarfs and the distance scale	487
10.8	Summary	488
10.9	References	489
10.10	Homework problems	491
11	Extrasolar planets	493
11.1	Introduction	493
11.2	Planet formation	495
11.2.1	Classical scenarios	495
11.2.2	Disk instability models	497
11.2.3	Definitions: brown dwarfs vs. planets	498
11.3	Searches for extrasolar planets	499
11.3.1	Astrometric surveys	499
11.3.2	Photometric detections and direct imaging	501
11.3.3	Detection using Doppler techniques	508
11.4	Interpreting the results	513
11.4.1	Alternative interpretations: pulsations, brown dwarfs or spots?	513
11.4.2	Formation conundra: Jovian planets in Mercurian orbits	515
11.5	What do we know about extrasolar planets?	516
11.5.1	The mass distribution	518

11.5.2	The orbits: semi-major axes, eccentricities and multiple systems	522
11.5.3	Radii and atmospheric composition	523
11.5.4	Metallicities and kinematics of the host stars	524
11.6	The future	528
11.6.1	<i>Kepler</i> and COROT	528
11.6.2	Ground-based interferometry and the Space Interferometry Mission	530
11.6.3	<i>Darwin</i> and the Terrestrial Planet Finder	533
11.7	Summary	534
11.8	References	534
11.9	Homework problems	538
Appendix	541
Index	551

Preface to the second edition

Comedy isn't the only field where timing is everything. The first edition of this book came out in 2000, as initial results from the DENIS, 2MASS and SDSS surveys were yielding the first extensive samples of L and T-type brown dwarfs, while the handful of extrasolar planets were still regarded with skepticism in some quarters. The ensuing five years have seen a spate of discoveries, with an order of magnitude increase in the number of known L and T dwarfs, new infrared surveys of most of the nearest star-forming regions, and a thorough census of cool dwarfs in the immediate Solar Neighbourhood. Detailed spectroscopic and photometric observations of ultracool dwarfs, spanning the full wavelength range from radio to X-ray, have produced new insight into their atmospheric structure, magnetic properties and circumstellar environments. At the same time, intense scrutiny of solar-type stars in the vicinity of the Sun has resulted in a fivefold increase in the number of known extrasolar planets, with the catalogue now extending to companions lying more than 4 AU from the central star, and with masses comparable to that of Neptune.

Recently, the frantic pace of discovery has fallen off to some extent – most of the easy questions have been answered. We have taken advantage of this comparative lull in activity to revise and update this textbook, incorporating as many as possible of the new results. Not surprisingly, the most extensive changes centre on the brown dwarf and extrasolar planet chapters, which have been essentially rewritten. We have also re-ordered the chapters, to take account of the fact that brown dwarfs are now very much in the main stream of astronomy, rather than a radical new discovery. There are substantial modifications to the chapters on the empirical characteristics of cool dwarfs (Chapter 2), magnetic activity (Chapter 5), the luminosity function (Chapter 8) and the mass function (Chapter 9); in addition, we have updated our brief survey of the Galactic structure (Chapter 7) to take into account recent results in that area. Throughout, we have corrected errors and inaccuracies in the first

edition, without, we trust, introducing too many new infelicities in the present copy. Finally, we have added some example questions to each chapter to aid the use of the book as a text for college and graduate courses.

As with the original edition, many of our colleagues provided useful information and helpful comments during the production of this textbook. In particular, we would like to thank the members of the 2MASS NStars project, Kelle Cruz, Peter Allen, Jim Liebert, Davy Kirkpatrick, Adam Burgasser, Patrick Lowrance and John Gizis, and students and postdocs at the University of Washington, Andrew West, Kevin Covey, John Bochanski, Lucianne Walkowicz and Nicole Silvestri. Ian Bond and Mark McGovern kindly provided figures at short notice. Able help on stellar flare topics was provided by early mentors Sandy Sandmann and Bjorn Pettersen, and colleagues Rachel Osten, Joel Allred and Edo Berger. Thanks to Clive Horwood of Praxis for the opportunity to produce this second edition, and to Neil Shuttlewood for shepherding us through the production process. Finally, as before, we would like to thank our families for their forbearance: Jamie and Catherine Mairi in Maryland; and Jim, Rosa and Tom in Seattle.

Preface to the first edition

Perhaps the most common question that a child asks when he or she sees the night sky from a dark site for the first time is: ‘How many stars are there?’ This happens to be a question which has exercised the intellectual skills of many astronomers over the course of most of the last century, including, for the last two decades, one of the authors of this text. Until recently, the most accurate answer was ‘We are not certain, but there is a good chance that almost all of them are M dwarfs.’ Within the last three years, results from new sky surveys – particularly the first deep surveys at near-infrared wavelengths – have provided a breakthrough in this subject, solidifying our census of the lowest mass stars and identifying large numbers of the hitherto almost mythical substellar-mass brown dwarfs.

These extremely low-luminosity objects are the central subjects of this book, and the subtitle should be interpreted accordingly. The expression ‘low-mass stars’ carries a wide range of meanings in the astronomical literature, but is most frequently taken to refer to objects with masses comparable with that of the Sun – F and G dwarfs, and their red giant descendants. While this definition is eminently reasonable for the average extragalactic astronomer, our discussion centres on M dwarfs, with masses of no more than 60% that of the Sun, and extends to ‘failed stars’ – objects with insufficient mass to ignite central hydrogen fusion. From this perspective, the Sun is a high-mass star.

The physical quantity underlying our curious child’s question is the contribution made by stars (and brown dwarfs) to the mass density in the Galaxy. This quantity is derived by integrating the stellar and substellar mass function – the number of objects per unit volume and unit mass. Rephrasing matters, a crucial question is whether low-mass stars and/or brown dwarfs can contribute enough dark matter to solve any of the missing-mass problems present in current astronomy. The answer happens to be ‘no’, but the question is sufficiently important, and the route to the

answer, pitfalls and all, is sufficiently interesting to merit more than the brief discussion afforded these objects in most textbooks.

M dwarfs are also of interest in their own right, with low-temperature atmospheres dominated by complex molecular opacities, semi-degenerate but fully convective interiors, and a tendency towards violent activity in their outer chromospheres and coronae. Beyond spectral class M are the cooler L dwarfs, comprising both the lowest mass stars and brown dwarfs. Identified and recognised as a separate class only within the last two years, they exhibit quite remarkable spectra induced by the combined effects of dust formation and ‘weather’ within their atmospheres. At even lower temperatures, brown dwarfs of spectral class T are closer cousins to the planetary gas giants of our Solar System than to the Sun.

Our text is written at a level comprehensible to an advanced undergraduate physics major, and is intended as a reference book for post-graduate students and researchers – including the authors themselves. In these chapters we consider low-mass stars and brown dwarfs both as individual objects and as representatives of the parent Galactic populations. Our perspective is that of the observational astronomer rather than the theorist. Although theory plays an indispensable role both in interpreting the available data and in suggesting new avenues for future work, this subject had been driven empirically. With that consideration in mind, the first chapter provides a broad introduction to the range of observational techniques employed in studying these systems. It also happens to collect together most of the formulae (normally dispersed in a variety of sources) which the authors have found useful in preparing for their own observations.

The succeeding four chapters discuss the intrinsic properties of low-mass stars and brown dwarfs. Chapter 2 sets the stage with a thorough discussion of their observational characteristics; Chapter 3 outlines their internal structure; Chapter 4 reviews analyses of their complex atmospheres; and Chapter 5 describes observations and theoretical models of the magnetically driven activity generated in the high-temperature chromosphere and coronal regions.

Later chapters consider M, L and T dwarfs as members of the Galaxy. Chapter 6 provides the context, with an overview of our current understanding of Galactic structure. Chapters 7 and 8 address the numbering and weighing of stars within the Galactic Disk population. Chapters 9 and 10 extend discussion to the search for brown dwarfs and for planetary-mass companions to main-sequence stars – two long-running quests which finally met with success in the closing years of the 20th century. Finally, Chapter 11 describes the nature and number of the low-mass stars in the metal-poor Galactic Halo.

Our aim throughout is to provide a thorough discussion of current knowledge in each of these areas, and to that end each chapter includes an historical overview of the development of that knowledge. Some topics must necessarily receive only cursory discussion within the limits of a single volume. We have, however, appended to each chapter an extensive, although not exhaustive, list of references to papers in the scientific literature which can serve as a starting point for more

detailed bibliographic research. Of necessity, these references – and the theories and observations outlined in the book – are current to early 1999. Given the explosive rate of discoveries in brown dwarf and extrasolar planet research between 1995 and the present date, it would be rash to predict what may emerge in the early years of the 21st century. The interested reader will be well advised to keep a weather eye on contemporary reviews.

Figures

1.1	Atmospheric transmission at an altitude of 4 km, equivalent to the altitude of Mauna Kea Observatory	2
1.2	Great circles and small circles	4
1.3	Celestial co-ordinate systems	5
1.4	Spherical triangles	7
1.5	Stellar parallax. (<i>Left</i>) the change in perspective of an Earth-bound observer; (<i>right</i>) the parallactic ellipse traced on the sky by an idealised nearby star (four-point star) against the reference frame of (five-point) background stars	9
1.6	Images of a 5×5 arcmin region from the STScI digitisation of POSS I (<i>left</i> , 1950) and POSS II (<i>right</i> , 1992) photographic plates	11
1.7	Schematics of optical telescopes: (<i>left</i>), a conventional Cassegrain reflector; (<i>right</i>), a Schmidt telescope	14
1.8	A schematic of an X-ray telescope	16
1.9	A schematic of a photon-counting detector	17
1.10	A schematic of a CCD array	18
1.11	A schematic of a diffraction-grating spectrograph.	32
1.12	Spectral orders produced by a diffraction grating	33
1.13	The stellar point-spread function	36
1.14	The speckle pattern	37
1.15	Speckle imaging analysis of two binaries in the Hyades	38
2.1	The Hertzsprung–Russell diagram.	48
2.2	Blue–green optical spectra of M dwarfs, illustrating the main features used to calibrate spectral type	50
2.3	Red optical spectra of M dwarfs, illustrating the main features used to calibrate spectral types	51
2.4	The L dwarf spectral sequence	53
2.5	Key features defining spectral class L	54
2.6	Far-red optical spectra of T dwarfs	56
2.7	Optical spectra of representative dC dwarfs, including G77-61, the prototype	58

2.8	Near-infrared spectra of three late-type dwarfs: Gl 411 (M2), Gl 406 (M6) and VB 10 (M8)	59
2.9	Near-infrared spectra of L dwarfs	60
2.10	Near-infrared spectra of T dwarfs	61
2.11	Spitzer LRS 5–15- μ m spectra of an M dwarf, GJ1001A (LHS 102A), a late-type L dwarf, 2M0255-4700, and the T dwarf binary, ε Indi Bab	63
2.12	Luminosity-sensitive features of M-type stars	64
2.13	An M giant spectral sequence	65
2.14	The relationship between absolute visual magnitude, and spectral type, defined by stars with accurate trigonometric parallax data	67
2.15	Broadband filter response curves	68
2.16	The (M_V , ($B-V$)) and (M_V , ($V-I$)) colour-magnitude diagrams for nearby single dwarfs with accurate trigonometric parallaxes and reliable photometry .	70
2.17	Optical/near-infrared colour-magnitude diagrams for nearby stars and brown dwarfs with accurate photometry and reliable trigonometric parallax measurements	72
2.18	Bolometric corrections in the V , I and K bands as a function of ($V-I$), ($I-K$) and spectral type	74
2.19	The subdwarf sequence in the (M_V , ($B-V$)) and (M_V , ($V-I$)) planes	75
2.20	Red spectra of the near-solar abundance disk M dwarf Gl 166C and of the high-velocity halo subdwarf LHS 1174	76
2.21	The ($(B-V)/(V-I)$) and ($(V-R)/(R-I)$) two-colour diagrams for main sequence stars	78
2.22	The near-infrared ($J-H$)/($H-K$) diagram	79
2.23	The log of the ratio between the luminosity emitted in the H α line and the total bolometric luminosity as a function of M_{bol} and spectral type for field stars in the Solar Neighbourhood, members of the Hyades cluster, Pleides cluster and the TW Hydreae Association	82
2.24	The logarithmic ratio between X-ray flux and bolometric flux plotted as a function of spectral type for nearby stars and members of the Hyades, Pleiades and TW Hydreae Association	84
2.25	(a) Light curves of AD Leo during its outburst of 12 April 1985 (b) A comparison between the ultraviolet-blue spectrum of AD Leo in outburst and in quiescence	85
2.26	The (mass, M_V) relationship defined by stars in binaries with well-determined orbits	87
2.27	Binary component separation plotted against the total system mass for a large sample of binaries	90
2.28	The distribution of mass ratios for binaries in the 8-parsec sample	91
2.29	$\Phi(M_V)$ and $\Phi(M_J)$ for the northern 8-parsec sample and the corresponding mass function, $\xi(\log M) \log M$	94
2.30	Far-red optical and near-infrared spectra of late-type dwarfs	99
3.1	The relative energy-generation rates of the P-P chain, the CNO cycle and triple- α helium fusion	109
3.2	The (density, temperature) plane	114
3.3	The effect of dissociation of the H ₂ molecule on the form of the lower main sequence	116

3.4	The mass–luminosity relationship derived by Grossman <i>et al.</i> and D’Antona and Mazzitelli	118
3.5	The Coulomb force as a function of the radial viation in the degeneracy parameter for the 0.5, 0.3, 0.2, 0.1 and $0.085 M_{\odot}$ models computed by [G8]	119
3.6	The evolution of core temperature with time predicted by Burrows <i>et al.</i> for dwarfs bridging the transition from main sequence, core-hydrogen-burning stars to brown dwarfs	120
3.7	The fractional luminosity contributed by nuclear fusion reactions as a function of time for the models plotted in Figure 3.6.	121
3.8	Luminosity evolution of low-mass dwarfs as computed by the Tucson group	123
3.9	The evolution of effective (surface) temperature with time, for the same models illustrated in Figure 3.8	124
3.10	The variation with time of radii of low-mass dwards, as precited by the Tuscon models	125
3.11	The mass–luminosity relationship at various ages, as predicted by the Tuscon low-mass dwarf models	126
3.12	Evolutionary tracks of low-mass dwarfs	127
3.13	Comparison of the theoretical and observations H–R diagrams	128
3.14	The (M_V , ($V-I$)) and (M_I , ($I-K$)) theoretical sequences compared with observations of nearby stars with parallax measurements	129
3.15	Theoretical main sequence relationships for subdwarfs, as predicted by D’Antona and Mazzitelli and Baraffe <i>et al.</i>	130
3.16	The mass–luminosity relationship for low-mass stars	131
3.17	The mass–radius relationship	133
3.18	Hubble Space Telescope images of gaseous jets from Herbig–Haro objects . . colour	
3.19	A protoplanetary disk, or ‘proplyd’, set against the Orion Nebula colour	
3.20	Hubble Space Telescope images of disks around young stars. . . . colour	
3.21	The evolution of a $0.1 M_{\odot}$ M dwarf as predicted by Laughlin <i>et al.</i>	140
3.22	Evolutionary tracks for low-mass dwarfs.	141
4.1	Schematic of energy levels of the hydrogen atom, illustrating bound–bound (absorption) and bound–free (ionisation) transitions in the Lyman and Balmer series	149
4.2	The line absorption profiles due to random thermal motions and finite energy width	150
4.3	The area under the continuum of the absorption line is the same as the area within the rectangle of width and height equal to the continuum flux.	152
4.4	A schematic curve of growth, showing the progression in equivalent width that an absorption line would experience as the number of atoms increases.	152
4.5	The observed spectrum of the M8 dwarf VB 10 is shown together with a model fit.	158
4.6	Wavelength dependence of opacity sources for solar and $[m/H] = -2.5$ models with $T = 2,800$ K and $\log g = 5.0$	160
4.7	The effect of changes in model computations	163
4.8	Comparison of a standard model computed by several authors, as indicated	165
4.9	The NextGen models [H1], illustrating structural changes with temperature, gravity and metallicity	166
4.10	Spectra from the NextGen models shown in Figure 4.9, illustrating the changes that occur when the temperature, gravity and metallicity are varied	168

4.11	Optical spectra for the prototype stars with spectral types M0, M2, M4 and M6 in Table 4.1, together with NextGen model spectra	170
4.12	Gliese 411 (M2), Gliese 406 (M6) and VB 10 (M8) spectra from 0.6–3.0 μm are shown together with NextGen model spectra of 3,400 K, 2,600 K and 2,200 K respectively	171
4.13	The ($V-I$) colour–temperature relationship from Table 4.1	172
4.14	Higher resolution spectra from the NextGen models of Figure 4.9, showing the detailed changes in the optica CaH and TiO bands with varying gravity and metallicity	173
4.15	The sdM1.5 subdwarf LHS 482 spectrum compared with model fits with differing temperature, gravity and metallicity	176
4.16	The abundance ratio is shown as a function of [Fe/H] for K and M dwarfs and subdwarfs	177
4.17	A series of late M and L dwarf spectra showing the increasingly strong alkali metal resonance lines with decreasing temperature	179
5.1	The Babcock model of the magnetic dynamo	188
5.2	The splitting of a degenerate atomic level into its magnetic sub-levels in the presence of a magnetic field	190
5.3	The Zeeman–broadened Fe I 8468 Å line observed in the active stars EV Lac and Gl 729	192
5.4	A single field strength and filling factor are insufficient to characterise the observed, Zeeman-broadened, line profile in AD Leo	193
5.5	The VAL3C semi-empirical model of the solar chromosphere is the solid line	194
5.6	Radiative losses are shown as a function of temperature for a hot, optically thin plasma	197
5.7	The C IV ionisation equilibrium	200
5.8	Emission measure distributions for several ions observed during a flare on AD Leo, together with a theoretical DEM curve	200
5.9	Light curves of solar-type Pleiades dwarfs showing photometric modulation typical of starspots rotating across the stellar surface	202
5.10	Light curves of the Ca II K index in G and K stars, from the Mount Wilson survey	203
5.11	Light curves off AU Mic, spanning a period of 10 years	204
5.12	Soft X-ray light curve observed with EUVE during a flare on AD Leo	207
5.13	The photometric light curves in the U-band filter from two observatories, together with the soft X-ray and H γ line flux data	208
5.14	Correlations between the energy emitted in chromospheric emission lines, the white-light continuum and the thermal corona are characteristic of both solar and dMe flares	209
5.15	The first example of the Neupert effect measured in a stellar flare	212
5.16	Active stars in the field and in clusters lie systematically above and to the red of the main sequence defined by the inactive stars	214
5.17	Activity effects on the TiO bandstrengths	215
5.18	The fraction of M dwarfs which are active vs. spectral type illustrates the steep rise through type M8 attributed to the increasing longevity of activity at later types	217
5.19	The activity strength measured by $L_{\text{H}\alpha}/L_{\text{bol}}$ shows a nearly constant value through the early-mid M dwarfs, and declines at type M6 and later	218

5.20	The active fraction of M7–8 dwarfs vs. vertical distance from the Galactic Plane	219
5.21	Colour–magnitude diagram for the open cluster NGC 2516, illustrating the concept of the H α -limit mass	220
5.22	Age–activity relationships for M dwarfs using H α limit measurements in clusters	221
5.23	The ratios of the H α , X-ray and radio luminosities to the bolometric luminosity are compared for early M through mid L dwarfs	222
5.24	The activity strength is nearly uncorrelated with rotation, showing evidence only for a possible threshold value below 2 km s ^{−1}	225
5.25	Flare spectrum of a dM9.5e 2MASS dwarf, together with a spectrum of the interesting dM9.5e dwarf PC0025	227
6.1	The lithium depletion line superimposed on evolutionary tracks for low-mass stars and brown dwarfs predicted using the Tucson models	238
6.2	The flux distribution of G 29-38 – the narrow features at optical wavelengths are the Balmer absorption lines	242
6.3	The optical and infrared spectrum of GD 165B, the first known L dwarf	247
6.4	The radial velocity curve measured for HD 114762	249
6.5	Ground-based and HST images of Gl 229B	251
6.6	The spectral energy distribution of Gl 229B	251
6.7	Selection criteria for the initial 2MASS L dwarf sample	255
6.8	ε Indi Bab, the binary T dwarf companion to the nearby K5 dwarf, ε Indi A	258
6.9	Far-red and near-infrared colours for late-type dwarfs as a function of spectral type	260
6.10	M _J and M _K as a function of spectral type for L and T dwarfs with accurate trigonometric parallax measurements	261
6.11	3–5 μm colours of M, L and T dwarfs	262
6.12	Bolometric corrections and effective temperatures of ultracool dwarfs as a function of spectral type	263
6.13	Optical blue/green spectra of late-type M and L dwarfs	265
6.14	The empirical diagram for M, L and T dwarfs compared to the predicted evolution of brown dwarfs with dust-free and dusty atmospheres	266
6.15	The Q branch of the fundamental ν ₃ methane band at 3.3 μm in L dwarfs	267
6.16	Far-red optical spectrum of the T5 dwarf, 2MASS 0559-1404	269
6.17	Gravity sensitive features in late-type dwarfs	271
6.18	Gravity sensitive features in late-type dwarfs	272
6.19	H α emission in the T6.5 dwarf, 2MASS1237 + 6526	273
6.20	Activity in ultracool dwarfs	274
6.21	Changes in H ₂ O bandstrength in 2MASS1523 + 3014, Gl 584C	275
6.22	The distribution of component magnitude difference as a function of angular separation for ultracool dwarfs with HST observations	280
6.23	Model spectra of a 5 M _J brown dwarf at ages of 0.1, 0.5, 1 and 5 Gyr	282
6.24	A far-red optical spectrum of an ultracool dwarf	288
7.1	Strömgren asymmetric drift	291
7.2	The edge-on spiral galaxy NGC 4565	295
7.3	The globular cluster M92, from a plate taken with the 48-inch Oschin Schmidt telescope	296

7.4	Colour-magnitude diagrams for the globular clusters M5 and M68. (M68 courtesy of A. Walker.)	297
7.5	(M_V , ($B-V$)) data for subdwarfs with abundance determinations, compared with 12-Gyr isochrones for $[\text{Fe}/\text{H}] = -0.7, -1.0, -1.5$ and -2.0	298
7.6	The present-day distribution of distance above the Galactic Plane of globular clusters, plotted as a function of abundance.	299
7.7	The variation in the α -element/iron abundance ratio as a function of iron abundance	303
7.8	The UBV and $uvby$ photometric passbands matched against the spectral energy distribution of the F-type subdwarf, HD 19445	304
7.9	The $((U-B), (B-V))$ diagram	305
7.10	The passbands of the DDO photometric system matched against the spectrum of a K0 disk giant	305
7.11	Spectra of two sets of G-type dwarf with identical ($B-V$) colours and abundances of $[\text{Fe}/\text{H}] = 0, -1$ and -2	306
7.12	The halo abundance distribution.	308
7.13	Theoretical isochrones for metal-poor globular clusters.	310
7.14	Age calibration for globular clusters	311
7.15	The correlation between rotation and abundance	314
7.16	The (V, U) and (V, W) velocity distributions of the nearby M dwarf sample and the probability plots for each component	318
7.17	The radial surface brightness profile of the disk in the face-on spiral galaxy UGC 438 (NGC 214)	320
7.18	The stellar density distribution perpendicular to the Galactic Plane	322
7.19	The metallicity distribution of an unbiased sample of long-lived stars in the Solar Neighbourhood	324
7.20	The $[\text{Mg}/\text{Fe}]$ vs. $[\text{Fe}/\text{H}]$ for a sample of FG stars in the Solar Neighbourhood and the $[\text{Fe}/\text{H}]$ distribution for the $[\text{Mg}/\text{Fe}] > 0.2$ ‘thick disk’ stars	325
7.21	The age–metallicity relationship for stars in the Galactic Disk.	328
7.22	An outline of the main spiral arm complexes in the vicinity of the Sun	330
7.23	A schematic outline of the 800-parsec square region centred on the Sun.	331
8.1	The changing form of the luminosity function	343
8.2	Luminosity functions derived using the method of mean absolute magnitude	346
8.3	The relationship between reduced proper motion and absolute magnitude defined by modern observations of stars in the nearby-star catalogue with trigonometric parallaxes accurate to better than 10%	347
8.4	The three luminosity functions derived by Kuiper from the analysis of nearby stars.	349
8.5	A comparison of Luyten’s 1939 luminosity function and Sanduleak’s analysis of the north polar objective prism survey and a comparison of Luyten’s 1968 results and Weistroop’s luminosity function.	351
8.6	A comparison of the luminosity function derived by Smethells, Reid and Luyten Lutz–Kelker corrections.	353
8.7	The V -band luminosity function, derived from a nearby star.	356
8.8	V -band luminosity functions derived from the photometric parallax surveys identified in the figure legend	364
8.9	367	

8.10	The ($M_V, (V-I)$) and ($M_I, (R-I)$) colour–magnitude diagrams for stars in the 8-parsec sample compared with the mean colour-magnitude relations used to infer photometric parallaxes	368
8.11	Systematic biases introduced in $\Phi(M_V)$ by using a linear ($M_V, (V-I)$) relationship	369
8.12	Finding nearby M dwarfs in the NLTT	371
8.13	Finding nearby M dwarfs in the SuperCOSMOS proper motion survey	372
8.14	The (M_J , spectral type) relation for ultracool dwarfs; the solid line is the best fit relation listed in the text	373
8.15	$\Phi(M_J)$, the J -band luminosity function	374
8.16	The spectral energy distribution of binary systems including a white dwarf and a later-type M dwarf	377
8.17	The proper motion distribution in (μ_u, μ_t) for stars with $16 < V < 16.5$ in a 50 square degree region near the Hyades	381
8.18	The ($M_V, (V-I)$) colour–magnitude diagram for the Hyades proper-motion candidate members	382
8.19	Mass segregation in older open clusters	385
8.20	$\Phi(M_V)$ for the Hyades and $\Phi(M_I)$ including fainter stars	387
8.21	The solid histogram plots the Pleiades luminosity function, while the shaded histogram gives the results derived by a deeper survey	390
8.22	The lithium depletion boundary in the Pleiades	392
8.23	The current best estimates of the V -band and J -band luminosity functions for stars and brown dwarfs in the Solar Neighbourhood	395
9.1	The Salpeter mass function	405
9.2	The relative orbit of the G/K dwarf binary Σ248	410
9.3	Radial velocity curves for the double-lined spectroscopic binary CM Draconis	412
9.4	Lensing geometry and flux amplification	414
9.5	The mass–luminosity relation at visual wavelengths	417
9.6	The K-band mass–luminosity relation	420
9.7	A comparison of the derivatives of (mass, M_V) relations ($d(\text{mass})/dM_V$) and $\Phi(M_V)$ for the nearby stars	422
9.8	A comparison of the location of the four components of the pre-main sequence quadruple GG Tauri against theoretical isochrones computed by Baraffe <i>et al.</i>	424
9.9	The nearby-star mass function	427
9.10	Predicted age and mass distributions for ultracool dwarfs as a function of spectral type	429
9.11	A comparison between the observed J -band luminosity function for ultracool dwarfs and theoretical predictions based on a constant birth rate and a power-law mass function	430
9.12	The initial mass function in the field	431
9.13	Theoretical (mass, M_V) relations for solar-abundance dwarfs with ages 0.1, 1 and 10 Gyr	434
9.14	The Pleiades mass function	435
9.15	An R -band image of the cluster IC 348	439
9.16	K -band source counts towards IC 348 and the K -band luminosity function matched against predictions based on a log-normal mass function and a truncated function	440
9.17	The H–R diagram for lower-luminosity stars in the Orion Nebula Cluster	442

9.18	Mass functions for a representative sample of young clusters and associations	446
10.1	The ($M_V, (B-V)$) and ($M_V, (V-I)$) H–R diagrams for late-type subdwarfs	459
10.2	The expected distance distribution of a sample of stars selected by proper motion	460
10.3	The ($H_V, (V-I)$) reduced proper motion diagram for stars in <i>Luyten's LHS Catalogue</i>	462
10.4	The narrowband CaH/TiO classification system	464
10.5	The sdM spectroscopic sequence	465
10.6	The esdM spectroscopic sequence	466
10.7	A comparison between spectra of 2M0532 + 82 and the L7 dwarfs, Denis 0205–1159	467
10.8	Infrared spectra of late-type sdM and L-type subdwarfs	469
10.9	Tangential velocity distributions predicted by Monte Carlo simulations for kinematic models based on either Layden <i>et al.</i> RR Lyrae analysis or Norris' observations of local subdwarfs	470
10.10	A comparison between Schmidt's halo luminosity function and the nearby-star luminosity function scaled by a factor of 1/400	471
10.11	The halo luminosity function in the Solar Neighbourhood	473
10.12	The halo subdwarf luminosity function for the Solar Neighbourhood	475
10.13	Luminosity functions derived for the nearby cluster NGC 6397, from ground-based observations and using HST data	477
10.14	The main sequence of NGC 6397, from HST observations, compared with theoretical isochrones for metal-poor stars	478
10.15	$\Phi(M_V)$ (and $\Phi(M_I)$) for the clusters M15, M30, M92, NGC 6397 and M4	479
10.16	Luminosity functions for the metal-rich globular cluster 47 Tuc and the intermediate abundance clusters Ω Cen, M3 and NGC 6752	480
10.17	(mass– M_V) relationships predicted by intermediate-abundance models	482
10.18	The (mass– M_V) and (mass, M_K) relationships predicted by the Lyon models	483
10.19	The halo mass function derived from field subdwarfs adopting the Lyon mass-luminosity relationship	484
10.20	Present-day mass functions for the globular clusters NGC 6397, M3, M4 and M30	485
10.21	Main-sequence fitting, using M subdwarfs	488
11.1	The astrometric displacement of the Sun produced by the Solar System planets, as seen from a distance of 10 parsecs	499
11.2	HST observations of the planetary transit on HD 209458	504
11.3	A microlensing planet	505
11.4	Near-infrared observations by the NICMOS camera on HST of dust disks around two nearby stars	507
11.5	The debris disk in β Pic	colour
11.6	Hubble Space Telescope coronagraphic observations of the edge-on disk around AU Mic and the face-on disk around HD 107146	colour
11.7	H–R diagram for the nearby stars	517
11.8	The mass/semi-major axis distribution for the known planetary systems of a solar-mass primary star	518
11.9	The brown dwarf desert	520

11.10	The mass function of low-mass companions to solar-like stars.	521
11.11	The distribution of eccentricity as a function of mass and period for extrasolar planetary systems	523
11.12	The metallicity distribution of extrasolar planet hosts and the frequency of planetary systems as a function of metallicity.	525
11.13	Schematic diagram of relative astrometry with a ground-based, long-baseline interferometer	530
11.14	The predicted astrometric accuracy ε_θ in arcseconds, as a function of angular separation for interferometric observations made with baselines of 1 m, 10 m and 100 m.	531

Tables

1.1	Broadband filter characteristics	20
2.1	Radii of low-mass stars	69
2.2	The fraction of dwarfs with H α emission as a function of spectral type	81
2.3	Distance distribution of nearby stars	93
4.1	Fundamental properties of M dwarfs	169
4.2	Temperatures of M subdwarfs and extreme subdwarfs	177
4.3	Temperatures of L dwarfs	180
5.1	Active dMe dwarfs	191
5.2	Magnetic field measurements of dMe stars	192
5.3	Magnetic activity properties of low-mass dwarfs	228
6.1	Benchmark ultracool dwarfs	245
6.2	L/T binaries with main sequence or post main sequence primaries	278
6.3	Ultracool binary systems	279
7.1	The Vatican stellar population scheme	293
7.2	The kinematics of the stellar halo	310
7.3	Velocity dispersions of disk stars	317
8.1	The declination distribution of CNS3 stars	362
8.2	Completeness limits for nearby-star analyses	362
8.3	Visual luminosity function	394
8.4	J-band luminosity function	394

xxxii **Tables**

9.1	Low-mass binary systems	418
10.1	Molecular bandstrength indicators	464
11.1	The pulsar planets	509
11.2	Radial velocity searches for planets	512

Acknowledgements

Most of the writing of this book was completed while Reid was at the California Institute of Technology, Pasadena, and Hawley was at the Space Sciences Laboratory, University of California, Berkeley, and at the University of Washington, Seattle, and we acknowledge access to the facilities at those establishments. We would also like to thank the staff at the Palomar and Keck Observatories for their able assistance in acquiring many of the observations discussed here. Particular credit goes to the Palomar Sky Survey team, responsible for completing the Second Palomar Sky Survey (POSS II). One of the authors (Reid) has been fortunate enough to be a participant in a Core Science project associated with the 2MASS sky-survey, undertaken by the University of Massachusetts and the Infrared Processing and Analysis Center (IPAC) at Caltech. We acknowledge the efforts of all associated with this very successful project, which (as Chapter 6 illustrates) has had a very substantial impact on our understanding of both low-mass stars and brown dwarfs.

During the course of this undertaking we have profited from discussions with many of our colleagues. Amongst those whom we would like to particularly thank are France Allard and Peter Hauschildt, for predictions, diagrams and discussions about their recent grids of model atmospheres; John Gizis, for comments, criticism and many figures, particularly in Chapter 10; Davy Kirkpatrick, Jim Liebert, Dave Monet, Adam Burgasser and Conard Dahn of the 2MASS ROT (Rare Objects Team) for stimulating discussions and new discoveries in the low-mass star/brown dwarf field; David Koerner, for reviewing Chapter 11; Lynne Hillenbrand, for figures and comments on Chapters 6, 7 and 8; Manuela Zoccali, for discussions on globular cluster luminosity functions; Bill Abbott, George Fisher, Jeff Valenti and, especially, Chris Johns-Krull, who also provided several figures, for consultations on stellar activity; Adam Burrows, for supplying data from theoretical models and for related discussions; Jeremy Mould, for stimulating interest in M dwarfs, and general joviality; Susana Deustua and Jamie Schlessman for comments and helpful

xxxiv Acknowledgements

criticism. We would also like to thank the following scientists for providing figures and pre-publication results: Sallie Baliunas, Tom Fleming, Hugh Jones, Anita Krishnamuthi, Greg Laughlin, Jenny Patience, Don Terndrup and Chris Tinney. As a technical credit, NASA's Astrophysics Data System proved an invaluable resource for identifying and locating relevant literature. Needless to say, remaining errors and omissions are the responsibility of the authors.

We are grateful to Clive Horwood for his patience during the three-year germination of this manuscript, and to Bob Marriott for ably editing the text. Finally, we would like to thank our respective families: Helen and Dwight Hawley, Jim, Rosa and Tom Hughes, on the one side; and Elizabeth Reid, Jamie and the four-legged ones, Ruaraidh, Gwen and Daisy, on the other. They provided tolerance and support during the completion of this manuscript.

1

Astronomical concepts

1.1 INTRODUCTION

M dwarf stars rank among the least spectacular constituents of the local stellar population, yet they represent almost half of the total mass locked up in stars within the Milky Way galaxy. As a result, apart from being interesting in their own right, the global characteristics of these stars have wider implications for studies of galactic structure and star formation, both within our own Galaxy and elsewhere. The main aim of this book is to summarise our current understanding of the physical structure and properties of these stars. That knowledge has been garnered through applying a variety of commonly-used astronomical techniques. This introductory chapter provides an outline of the basic concepts involved, placing a particular emphasis on practical observation.

1.2 THE ELECTROMAGNETIC SPECTRUM

Most astronomical observations are made through the Earth's atmosphere, which is not perfectly transparent. Radiation is absorbed and scattered with various degrees of efficiency at different wavelengths (Figure 1.1). The relevant physical process depends on the energy of the incoming radiation. Wavelength, frequency and energy are related by the equation

$$E = h \frac{c}{\lambda} = h\nu \quad (1.1)$$

where ν is the frequency, λ is the wavelength, c is the speed of light, and h is Planck's constant.

At the longest wavelengths, $\lambda > 100$ m, radiation incident on the Earth's atmosphere is completely reflected by the ionosphere. Reflection from the underside of the

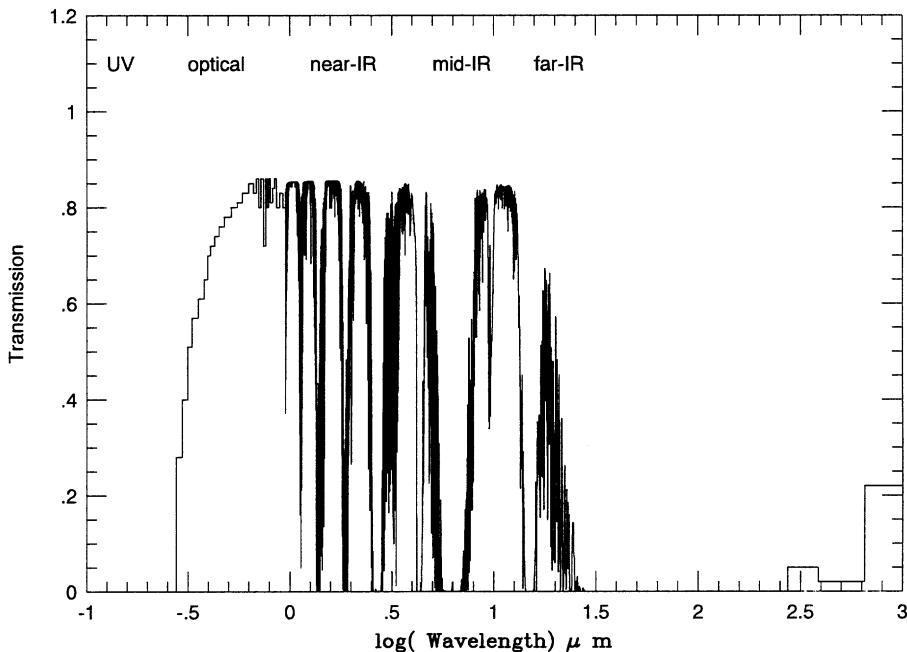


Figure 1.1. Atmospheric transmission at an altitude of 4 km, equivalent to the altitude of Mauna Kea Observatory. (Long-wavelength data courtesy of Peter Stockman.)

same layer allows long-wavelength radio stations to broadcast beyond their local horizon. The atmosphere is relatively transparent to higher-frequency radio waves, spanning wavelengths between 100 m and 1 cm. At wavelengths shorter than 1 cm, electromagnetic radiation has sufficient energy to excite vibrational motion in molecules such as CO, OH and, particularly, H₂O. Those molecules absorb a substantial fraction of the incident millimetre and infrared (1 cm to 1 μm) radiation, although there are regions of moderate transparency, notably at ∼5, 3.4, 2.2, 1.6 and 1.25 μm.

Shortward of 1 μm, the OH radical contributes a number of absorption bands (as well as significant emission) and there are two strong O₂ bands at $\lambda \sim 7,500 \text{ \AA}$ (the A-band) and $\lambda \sim 6,800 \text{ \AA}$ (the B-band) (see [R2]). However, by $\lambda \sim 7,000 \text{ \AA}$ ($7 \times 10^{-7} \text{ m}$), the photon energy exceeds that of almost all molecular vibrational transitions. As a result, there is a second highly-transparent window between that wavelength and ∼3,250 Å – the visible region of the spectrum. This matches the peak in the solar energy distribution, so it is no coincidence that our own (and most other terrestrial) eyes are most sensitive to light at those wavelengths.

The optical window is not completely transparent, since the radiation has a wavelength comparable to the size of aerosol particles in the atmosphere. As a result, scattering occurs, particularly at blue wavelengths, since Rayleigh scattering has a λ^{-4} dependence. This wavelength dependence accounts for the fact that the sky is blue. The proportion of the incident radiation that is scattered depends on the

total path-length through the atmosphere. Thus, an object appears both fainter and redder (since blue light is scattered preferentially) when near the horizon.

At shorter wavelengths, the high-energy radiation can ionise atoms, and is absorbed completely. As a result, the atmosphere is totally opaque to ultraviolet radiation, $\lambda \sim 3,250$ to 120 \AA (100 eV) and soft X-rays ($\lambda > 0.25\text{ \AA}$ or $E < 50\text{ keV}$). At even higher energies, Compton scattering prevents the radiation from penetrating the atmosphere, although γ -rays ($E > 1\text{ MeV}$) can be detected from high-energy particle showers as they mutually annihilate through pair-production in the upper atmosphere.

1.3 POSITIONAL ASTRONOMY

A well-defined co-ordinate system is an essential requirement if individual objects are to be identified in a unique manner. In astronomy, objects are observed on the celestial sphere with positions defined in spherical co-ordinate systems, measured in angular units (degrees, radians). These are directly analogous to the system of longitude and latitude used to determine positions on the surface of the Earth; indeed, the primary celestial system is a direct projection of the geographic system.

There are four main celestial systems, each defined with reference to a fundamental reference plane which passes through the centre of the sphere (the observer). The circle defined by where this plane intersects the celestial sphere (AFBC in Figure 1.2) is a great circle: defining the diameter of the sphere as $r = 1$ unit, the length of this circle is 2π , the maximum possible. Any circle defined by a plane which does not pass through the centre of the sphere (such as DSE, parallel to ABC, in Figure 1.2) has a smaller circumference and is known as a small circle. Each great circle has two poles – the two diametrically-opposed points on the celestial sphere which lie 90° from every point on the great circle. In Figure 1.2, P and Q identify the two poles of the great circle ABC.¹

Any circle drawn through both poles is a great circle, intersecting the reference circle ABC at an angle of 90° . Choosing one 180° segment of a polar circle (for example, PBQ in Figure 1.2) as a second reference then defines an orthogonal co-ordinate system. Consider the point S in Figure 1.2. This point is defined uniquely by two angles: the angle between the reference polar circle, PBQ, and the polar circle passing through S (the angle BPS, which is also the angular length of the arc BF); and by the angular distance PS. In most systems, the latter angle is defined with respect to the fundamental circle, so the angle becomes FS = $90^\circ - PS$. In the case of the terrestrial system of latitude and longitude, the fundamental circle is the equator, while the Greenwich meridian defines the reference polar circle. The angle BPS then defines the longitude of a given location on the Earth's surface, while FS gives the latitude.

¹ Since P and Q are the poles of only one great circle, ABC, it is clear that a given co-ordinate system can be defined unambiguously either by the position of the reference great circle or by the positions of the poles of that great circle.

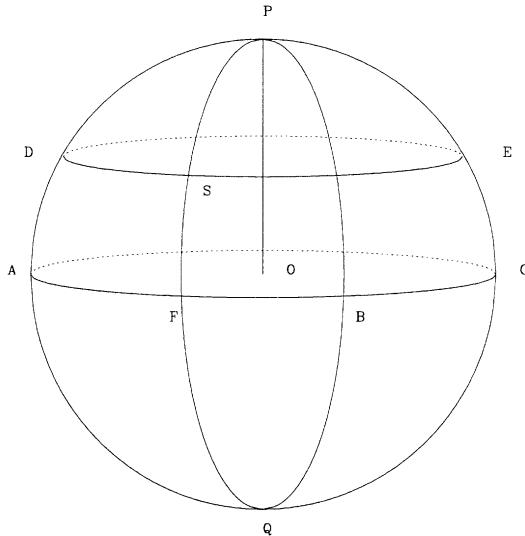


Figure 1.2. Great circles and small circles.

Two other properties of spherical geometry are worth bearing in mind. First, there is only one great circle that passes through any two points on a sphere, unless the two points form the poles of a great circle. Second, just as a straight line is the shortest distance between two points on a flat surface, the shortest distance between two points on a sphere is measured along the great circle joining those two points.

1.3.1 Co-ordinate systems

The first three systems

The alt-azimuth system The first astronomical co-ordinate system is the altitude/azimuth system. The fundamental plane, the horizon, is defined by the poles: the zenith, the point directly overhead, and the nadir, the diametrically-opposed pole, directly underneath the observer. Figure 1.3 indicates the fundamental plane (NS) and the zenith (Z). Semi-great circles passing through the zenith are known as vertical circles, and the vertical circle passing through the horizon due north of the observer is the principal vertical circle, which defines the zero point of the azimuth scale. The prime verticals are the vertical circles due east (azimuth = 90°) and due west (azimuth = 270°), and the altitude of a given object is the angular distance above the horizon, measured along the appropriate vertical.

The zenith distance, z , is the complement of the altitude. This angular distance provides a measure of the path-length through the Earth's atmosphere, a quantity important both for determining accurate fluxes (see Section 1.5.2) and the degree of refraction. The latter is approximately represented by

$$\zeta = R \tan(z) \quad (1.2)$$

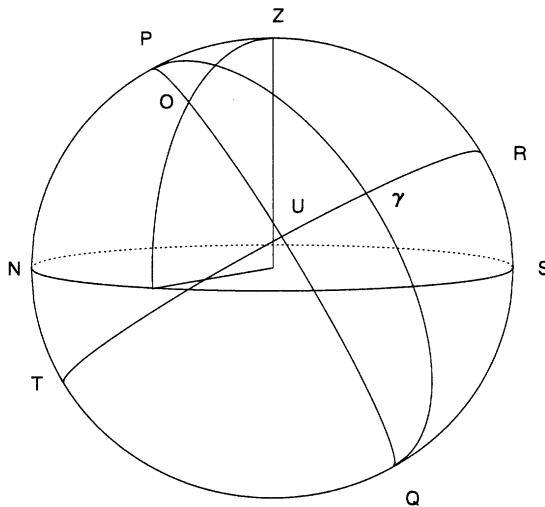


Figure 1.3. Alt-azimuth and equatorial co-ordinate systems.

where ζ is the angle of refraction and R is the constant of refraction, usually expressed in arcseconds. Since blue light is refracted to a larger extent than red light, R is wavelength dependent, with $R \sim 59.^{\circ}4$ at 3,500 Å, $R \sim 58.^{\circ}2$ at 5,000 Å, $R \sim 57.^{\circ}7$ at 6,500 Å and $R \sim 57.^{\circ}5$ at 9,000 Å [F1]. This variation leads to differential chromatic refraction, with a point source dispersed into a mini-spectrum at low altitudes.²

The equatorial system All celestial objects rise and set, and the trajectory of their motions depends on the (terrestrial) latitude of the observer. Thus, altitude and azimuth are both time- and location-dependent, and while the alt-azimuth system is used operationally by most large, modern telescopes, it cannot provide a universal reference system. The equatorial system fulfils that purpose by taking the reference frame as the projection of the Earth's equator onto the celestial sphere. The poles of the latter system are therefore the projection of the north and south terrestrial poles.

If we consider observations made from a given terrestrial location, latitude ϕ , then the zenith distance of the pole, PZ in Figure 1.3, is $(90^\circ - \phi)$. As the Earth rotates, an object, O, describes a circle at constant angular distance, PO, from the pole P. The complement of this angle, the angular distance of the star from the celestial equator as measured along a great circle (angle OU in Figure 1.3), defines the declination, the celestial equivalent of latitude.

A second reference circle is required to specify the position of O on the celestial sphere. This is provided either by the observer's meridian, the semi-great circle

² This effect was actually used to identify quasars on Palomar Sky Survey blue plates taken at moderate to large zenith distances: the strong ultraviolet continuum in low-redshift QSOs led to their having more extended images than stars.

PZRSQ (where S is the southern point on the observer's horizon), or by the semi-great circle $P\gamma Q$. In the former case, the east/west reference line is fixed relative to the observer, and angle ZPO defines the hour angle (HA), measured east or west of the meridian in either degrees or, more usually, hours, minutes and seconds of time. West of the meridian, the HA is the time since the object crossed the meridian (transit); east of the meridian, the HA is the time until transit. Thus, the hour angle of an object is changing continuously.

In contrast to the circle PSQ, $P\gamma Q$ is a fixed reference circle on the celestial sphere. γ is a reference point on the celestial equator, the first point of Aries (see below). As a result, the right ascension (RA) of O, the angle γPO , is constant for the equatorial system $R\gamma UT$, and defines an unambiguous position. As with hour angle, right ascension is usually measured in units of time, increasing as one moves east of γ . Right ascension and declination are notated as α and δ , respectively.

The ecliptic system The third reference system is defined by the plane of the Earth's orbit, traced out by the ecliptic, the apparent motion of the Sun on the celestial sphere. The angular separation of the ecliptic and equatorial poles is $23^\circ 27'$, and ecliptic co-ordinates are specified in latitude (β) and longitude (λ). The ecliptic intersects the celestial equator at two points (nodes) 180 degrees apart: the vernal equinox where the Sun moves from south to north of the equator, and the autumnal equinox where the Sun returns to the southern celestial hemisphere. The former node is the first point of Aries, denoted γ as mentioned above (Figure 1.3). This point serves as the reference for both ecliptic longitude and right ascension.

Precession The equatorial and ecliptic systems are not invariant in time. Since the Earth's rotational axis is inclined with respect to both the poles of the ecliptic and the poles of the lunar orbit, and since the Earth is not a perfect sphere, an imbalance of gravitational forces due to both the Moon and the Sun (an applied torque) leads to precession. The main result of luni-solar precession is that the Earth's pole describes a small circle, radius $\sim 23^\circ$, about the ecliptic pole, with a period of $\sim 26,400$ years.³ The vernal equinox moves backwards (towards negative longitude) along the ecliptic by $\sim 50'' \text{ yr}^{-1}$. Thus, when giving a position in either equatorial or ecliptic co-ordinates, the reference equinox (by year) must be specified. Most catalogues list positions for equinox 1900, 1950 or 2000.

Confusion is often encountered in the astronomical literature over the meaning of the terms equinox and epoch. Equatorial (and ecliptic) co-ordinates are specified with reference to the position of the vernal equinox on a given date; hence, they are cited as, for example, equinox 1950 co-ordinates. However, stars in the immediate Solar Neighbourhood can possess appreciable angular motion (proper motion) due to their velocity relative to the Sun. One can allow for those motions and correct the position to a given date (epoch), but still maintain the same co-ordinate system

³ There are additional shorter-term effects, notably nutation, which has a period of ~ 18.6 years. Full details are given in [M1] and [S1].

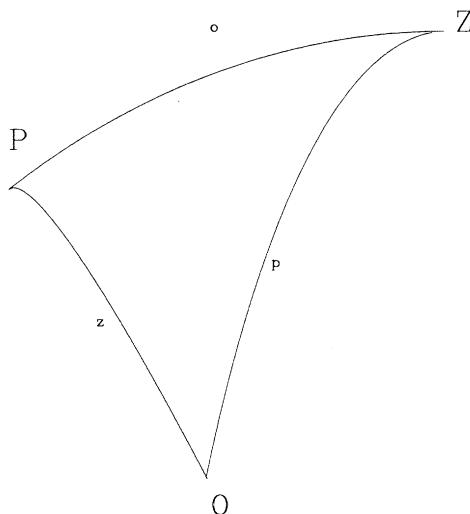


Figure 1.4. Spherical triangles.

(equinox). In other words, the equinox listed for a given object specifies the position of the reference system and the epoch specifies the date when the object was at the stated co-ordinates.

Spherical trigonometry

Before defining the fourth co-ordinate system, it is useful to outline the basic formulae of spherical trigonometry which allow one to manipulate co-ordinates and transform between different systems. The full derivation of these formulae is given in [S2], and we simply state the results here. Consider the spherical triangle PZO (Figure 1.4). We shall denote the inner angles as Z , P and O , while the angular lengths of the three sides are designated z , p and o . The two most useful formulae are the sine rule

$$\frac{\sin O}{\sin o} = \frac{\sin P}{\sin p} = \frac{\sin Z}{\sin z} \quad (1.3)$$

and the cosine rule

$$\cos o = \cos z \cos p + \sin z \sin p \cos O \quad (1.4)$$

There are comparable expressions for $\cos z$ and $\cos p$. Also occasionally useful are the mixed formula

$$\sin z \cos O = \cos o \sin p - \sin o \cos p \cos Z \quad (1.5)$$

and the four-parts formula

$$\cos z \cos O = \sin z \cot p - \sin O \cot P \quad (1.6)$$

Thus, if we know the hour angle and declination of a given star, we can determine

the altitude using the cosine rule, since $z = 90^\circ - \delta$; $o = 90^\circ - \phi$; P is the hour angle; and $p = 90^\circ - \text{altitude}$. Given p , the azimuth (Z) can be determined by applying the sine rule. Angular distances between sources can also be determined using these formulae.

Galactic co-ordinates

Our Galaxy is a disk system (Chapter 7). Fortified by trigonometry, we can now define the fourth reference system. Galactic longitude and latitude (l, b) are measured with respect to the plane of the Milky Way. The present system (l^H, b^H) dates from 1958, and replaces the (l^I, b^I) system defined by Ohlsson [O1]. Under the new system, the north Galactic pole is defined (by the International Astronomical Union) to be at position $\alpha = 12h\,49m$, $\delta = +27^\circ\,24'$ (1950 equinox), while the reference great circle at $l^H = 0^\circ$ is defined to lie at a position angle of 123° from the great circle through the north Galactic and north celestial poles. These criteria place the origin of the (l^H, b^H) co-ordinate system at $\alpha = 17h\,42m\,26s.603$, $\delta = -28^\circ\,55'0''445$ for equinox 1950 [L4].⁴ The Galactic plane intersects the celestial equator at $l = 33^\circ, 23'$. Transforming (α, δ) to (l, b) is more complicated than moving between equatorial and ecliptic co-ordinates, since the two systems do not share the same zero-point. However, the following formulae can be applied:

$$\cos(b) \cos(l - 33^\circ) = \cos(\delta) \cos(\alpha - 282^\circ.25)$$

$$\cos(b) \sin(l - 33^\circ) = \cos(\delta) \sin(\alpha - 282^\circ.25) \cos(62^\circ.6) + \sin(\delta) \sin(62^\circ.6)$$

$$\sin(b) = \sin(\delta) \cos(62^\circ.6) - \cos(\delta) \sin(\alpha - 282^\circ.25) \sin(62^\circ.6)$$

$$\cos(\delta) \sin(\alpha - 282^\circ.25) = \cos(b) \sin(l - 33^\circ) \cos(62^\circ.6) - \sin(b) \sin(62^\circ.6)$$

$$\sin(\delta) = \cos(b) \sin(l - 33^\circ) \sin(62^\circ.6) + \sin(b) \cos(62^\circ.6) \quad (1.7)$$

These equations allow co-ordinates to be transformed between the equatorial and Galactic systems.

1.3.2 Stellar astrometry

Measuring accurate positions of celestial objects is the astronomical subdiscipline with the longest pedigree. Indeed, astrometry *was* astronomy until the latter half of the nineteenth century. From the perspective of studying low-mass dwarfs, two astrometric parameters are of prime importance: parallax and proper motion.

Stellar parallax is the apparent motion of a star due to our changing perspective as the Earth orbits the Sun. Measured against a reference frame of more distant objects, the target star describes an ellipse, the semi-major axis of which is the

⁴ Sagittarius A*, the radio source associated with the Galactic centre, is actually at $\alpha = 17h\,42m\,02s$, $\delta = -28^\circ\,47'6.0''$ (equinox 1950), or $l = 0^\circ.06$, $b = +0^\circ.15$.

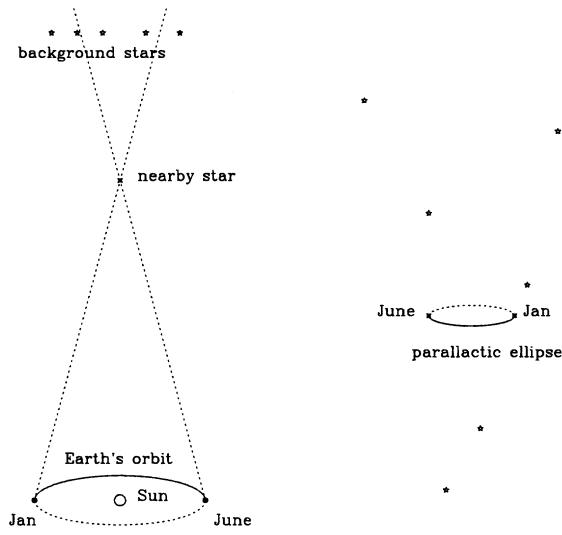


Figure 1.5. Stellar parallax. (*left*) The change in perspective of an Earth-bound observer; (*right*) the parallactic ellipse traced on the sky by an idealised nearby star (four-point star) against the reference frame of (five-point) background stars. In reality, the motion would include a linear component due to proper motion.

parallax angle, π , and the semi-minor axis $\pi \cos(\beta)$, where β is ecliptic latitude. The ellipse is the projection of the Earth's orbit (Figure 1.5). Thus, a star at the ecliptic pole describes a near-circular path, while a star in the ecliptic (the plane of the Earth's orbit) exhibits linear motion.

The semi-major axis of the Earth's orbit is 149,597,870 km, defined as one astronomical unit (AU) [L5]. Since this distance, a_E , is much smaller than the distance to even the nearest star, the parallax of a star at distance r can be written using the small angle approximation as

$$\pi = \frac{a_E}{r} \quad (1.8)$$

The units used in astronomical measurement follow from this equation: a star with $\pi = 1$ arcsec is defined to be at a distance of 1 parsec, equivalent to 3.2616 light years or 206,265 AU. The first successful measurements of stellar parallaxes were by Bessel (61 Cygni, 1838), Henderson (α Centauri, 1839) and Struve (Vega, 1840). The nearest known star to the Sun, Proxima Centauri, an M dwarf companion of α Centauri, has a parallax of only $0.^{\circ}772$. The current limit of precision for parallax measurement is 0.5–1 milliarcseconds (mas).

An important point to note is that *all* objects in a given direction have shared parallactic motion, but with amplitudes which depend on the reciprocal of their distance. Parallax measurements are generally made from time-series of direct images, taken either by photographic plates or, more recently, by digital detectors such as CCDs (see Section 1.4.2 and [M1]). Reference stars are limited to separations

of less than a degree from the target. Their parallactic motion, in phase with the target, leads to the measured apparent parallax of the target, π , being an underestimate of the true parallax, π_o . The offset can be corrected if other techniques are used to estimate distances to the reference stars, which typically lie at distances of 500–3,000 parsecs. The *Hipparcos* astrometric space mission [E2] circumvented this problem by using a compound telescope to project two 1 square degree areas of sky, separated by an angle of $\sim 58^\circ$, onto the same focal plane. This technique – rendered impossible in ground-based observations by atmospheric distortions – allows measurement of the angular separations of stars with very different parallactic motions, so that absolute astrometric parameters can be determined [V1], [K3].

Parallax is a cyclical motion. Proper motion, in contrast, is secular and cumulative. Stellar proper motions reflect the changing relative positions of the Sun and the target star due to their changing position in the Galaxy. Writing the annual proper motion as μ , then

$$\mu = \frac{V_T}{\kappa r} \text{ arcsec yr}^{-1} \quad (1.9)$$

where V_T is the heliocentric transverse velocity of the target in km s^{-1} , r is the distance in parsecs and $\kappa = 4.74$. As with parallax, proper motions are generally measured with respect to a reference grid of more distant stars which are typically members of the Galactic disk, and a small correction is usually required to transform the target measurement to an absolute reference frame. However, unlike parallax measurements, faint (15th–18th magnitude) galaxies can be used to define the reference frame for measurements, and absolute motions determined directly [V2].

Many of the nearest stars have proper motions of 1 arcsec or more per year. Figure 1.6 shows an example, the low-luminosity M dwarf LHS 2924, photographed on 20 May 1950 by the Oschin Schmidt telescope as part of the original Palomar Observatory Sky Survey, POSS I, and again on 2 May 1992 in the course of the second Palomar Sky Survey, POSS II. With $\mu = 0.^{\circ}802 \text{ yr}^{-1}$, the displacement is obvious. Proper-motion surveys based on this type of plate material have been responsible for identifying many of the lowest-luminosity dwarfs currently known to lie in the vicinity of the Sun.

1.3.3 Stellar kinematics

The average space velocity and the velocity dispersion of a group of stars provide a measure of, respectively, the mean Galactic orbital velocity and the fraction of kinetic energy resident in random, rather than ordered, motion. Proper motions measure displacement tangential to the line of sight, while radial velocities supply the third co-ordinate. The radial velocity observed at any epoch includes the projected contribution from both the orbital motion of the Earth around the Sun (or, rather, the combined motion of the Earth around the Earth–Moon barycentre and the motion of the latter around the barycentre of the Solar System) and the Earth’s diurnal rotation. In round numbers, the former amounts to $\pm 18.5 \text{ km s}^{-1}$, while the latter is no more than $\pm 450 \text{ m s}^{-1}$. Both effects must be taken into account

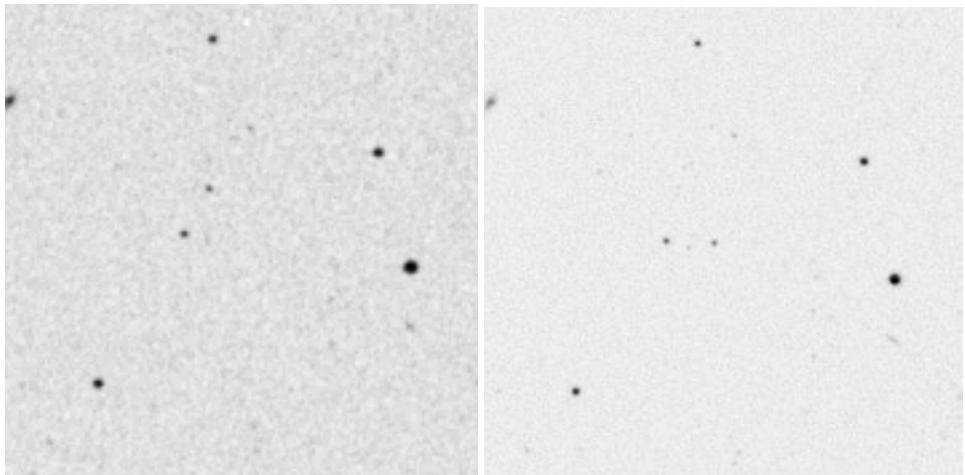


Figure 1.6. Images of a 5×5 arcmin region from the STScI digitisation of POSS I (*left*, 1950) and POSS II (*right*, 1992) photographic plates. The low-luminosity proper motion star LHS 2924 is centred in the POSS II image.

in precision velocity work, such as searches for planetary-mass companions (see Chapter 11).

Proper motions are usually measured in the (α, δ) equatorial system. However, it is more useful for statistical purposes to transform these to the Galactic (l, b) system

$$\begin{aligned}\mu_l &= \mu_\alpha \cos(\phi) + \mu_\delta \sin(\phi), \\ \mu_b &= -\mu_\alpha \sin(\phi) + \mu_\delta \cos(\phi)\end{aligned}\quad (1.10)$$

where ϕ is the angle between the direction towards the north celestial pole and the north Galactic pole. Simple application of spherical trigonometry shows that

$$\sin \phi = \frac{\sin(90 - \delta) \sin(\alpha - \alpha_0)}{\sin(90 - b)} \quad (1.11a)$$

and

$$\cos \phi = \frac{\cos(90 - \delta_0) \sin(90 - \delta) - \sin(90 - \delta_0) \cos(90 - \delta) \cos(\alpha - \alpha_0)}{\sin(90 - b)} \quad (1.11b)$$

where (α, δ) are the equatorial co-ordinates of the star in question, and (α_0, δ_0) are the equatorial co-ordinates of the north Galactic pole.

If the distance of the programme object is known, then proper motions can be transformed to velocities, giving the heliocentric velocity triad (V_l, V_b, V_r) . For Galactic structure studies it is more useful to transform these velocities to a (right-handed) co-ordinate system with orthogonal axes directed towards the Galactic centre (defined as the U velocity), in the direction of Galactic rotation

($l = 90^\circ$, $b = 0^\circ$, V velocity) and perpendicular to the Galactic plane (W velocity). The choice of axes is dictated by the fact that the Galactic disk is a flattened, nearly axisymmetric system, with ordered rotational motion (see Chapter 7). The observed (V_l, V_b, V_r) are transformed to (U, V, W) using

$$\begin{aligned} U &= V_r \cos(l) \cos(b) - V_b \cos(l) \sin(b) - V_l \sin(l) \\ V &= V_r \sin(l) \cos(b) - V_b \sin(l) \sin(b) + V_l \cos(l) \\ W &= V_r \sin(b) + V_b \cos(b) \end{aligned} \quad (1.12)$$

In determining stellar kinematics, stars are usually grouped together based on similarity in properties such as broadband colours or spectral type. The mean heliocentric velocities in U, V and W measure the solar motion, while the velocity distribution is usually characterised as orthogonal Gaussian dispersions (the Schwarzschild velocity ellipsoid). Traditionally, velocity dispersions have been measured for the (U, V, W) axes as $(\sigma_U, \sigma_V, \sigma_W)$. However, in some cases, notably for young stars, the best-fit velocity ellipsoid is misaligned with the (U, V, W) axes. The offset from the radial (U) axis is described as the ‘vertex deviation’. Chapter 7 presents recent observational analyses of Galactic stellar kinematics.

1.4 TELESCOPES AND DETECTORS

1.4.1 Telescopes

The two main properties of a telescope are the diameter of the primary optical element (D), which defines both the light grasp and the limiting angular resolution, and the focal length, f , which determines the plate-scale, the number of arcsec per millimetre in the focal plane. The angular resolution, Θ , is set by the diffraction limit:

$$\Theta = \frac{\lambda}{D} \times 206,265 \text{ arcsec} \quad (1.13)$$

where λ is the wavelength of the observations, and 206,265 is the number of arcseconds in a radian. Thus, the human eye, with an effective aperture of 5 mm, has an angular resolution of ~ 20 arcsec; an 8-inch (20-cm) telescope can resolve double stars separated by 1 arcsec; and the 200-inch (5-metre) Hale telescope at Mount Palomar has a diffraction limit of 0.02 arcsec at visual wavelengths, although atmospheric turbulence (seeing) prevents this resolution being attained in direct observations.

The focal length, f , of a telescope is the distance between the primary optical element and the principal focus. The plate-scale at that focus is given by

$$p = \frac{206,265}{f} \text{ arcsec mm}^{-1} \quad (1.14)$$

The focal ratio of the system is the ratio of the focal length and the diameter of the

primary, $R_f = f/D$. Thus, the 200-inch Hale telescope has a primary mirror with a focal ratio of 3.3 (written f/3.3), corresponding to a focal length of 55 feet (16.76 metres) and a plate-scale of $12.3 \text{ arcsec mm}^{-1}$.

The majority of large telescopes built for astronomical research (and, indeed, for amateur work) are reflecting telescopes. The main optical element is a mirror made of low-expansion glass and coated with a thin, highly-reflective layer of aluminium or gold. The primary advantages of a reflector are ease of construction, and cost. The objective lens in a refracting telescope must be edge-supported, and therefore must be sufficiently rigid to minimise deformation as the telescope is slewed. The structural strength of glass sets an upper limit of $\sim 1 \text{ m}$ to the diameter of refractors. In contrast, with reflecting optics the primary mirror is back-supported, and it is therefore possible to make larger-aperture telescopes – up to 8 m for monolithic mirrors (such as the Gemini telescopes and the ESO Very Large Telescopes) and 10 m for segmented-mirror telescopes (the Keck telescopes on Mauna Kea). Plans are being made for 30 m+ segmented mirrors (e.g., T2).

Not only is it substantially more expensive to fabricate high-quality transmitting optics than to make reflecting optics of the same aperture, but mirrors can be ground to faster optical ratios, leading to shorter focal-length telescopes, and hence smaller enclosures (domes). Each of the Keck 10-m telescopes, for example, has a primary mirror with a focal ratio f/2, producing a focal length of only 20 m, and allowing the telescope to be accommodated within a dome little larger than that of the Palomar 200-inch. On the other hand, the largest refractor constructed, the 40-inch Yerkes telescope at Williams Bay, Wisconsin, has a focal ratio of f/19 and a focal length of 63 feet – 8 feet longer than the 200-inch telescope. Very few large refracting telescopes have been constructed since the end of the nineteenth century.

Until recently, most telescopes were mounted equatorially: that is, a rotation axis is aligned with the Earth's polar axis, so motion due to diurnal rotation can be eliminated by tracking in only one co-ordinate (RA). However, it is simpler mechanically to construct an alt-azimuth telescope, where a rotational axis is aligned with the local gravitational field. Advances in computer technology now permit accurate tracking in two co-ordinates, taking full account of effects such as telescope flexure and refraction, and most telescopes built in the last 15 years have alt-azimuth mounts.

The standard telescope design is based on the Cassegrain system: a secondary mirror (B) brings the light to a focus (C) at a point below the primary (A) (Figure 1.7). This provides a stable mount for moderate-sized instruments. In conventional telescopes, the distance AB is close to the focal length of the primary mirror, principally to minimise the diameter of the secondary mirror, and hence the size of the obstruction of the primary beam. This positioning requires that the secondary be suitably figured, which changes the effective focal ratio. Typical values are f/8, although ratios as high as f/400 are possible, with a consequent increase in the plate-scale. Alternatively, in many telescopes the secondary can be removed, and observations undertaken directly at the prime focus, P, at the primary focal ratio of f/2.5 or f/3. The latter focal station gives the largest plate-scale (most arcsec/mm), and thus provides easy coverage of large solid angles, but only limited space for

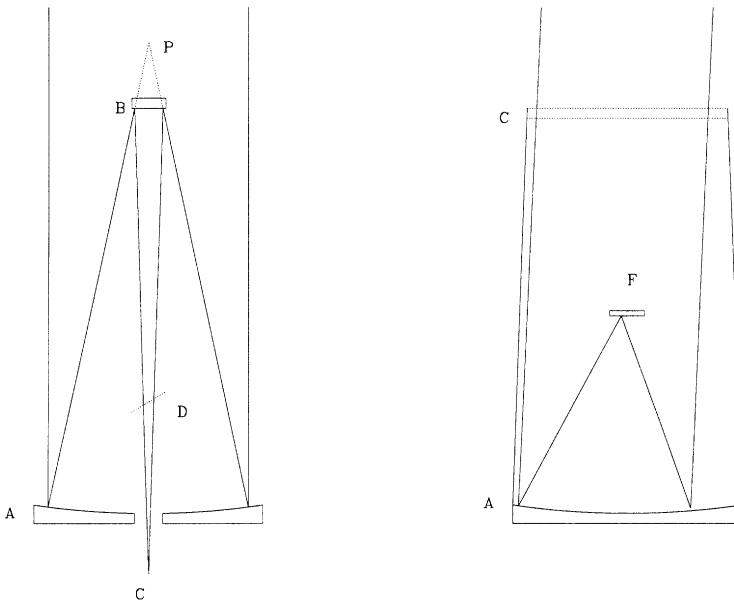


Figure 1.7. Schematics of optical telescopes: (*left*), a conventional Cassegrain reflector; (*right*), a Schmidt telescope.

mounting instruments. Large instruments obstruct the primary mirror and require a mechanically rigid telescope mount.

The classical Cassegrain design has a parabolic primary mirror. This has the disadvantage of producing off-axis images with significant optical aberrations: coma and spherical aberration [J1], [W2]. These can be corrected to some extent using a lens (corrector plate), but for only a limited field of view ($\sim 0^\circ.3$ in diameter). The Ritchey–Chrétien modified Cassegrain design has a hyperbolic primary which, combined with the appropriate secondary, results in less distortion and a larger useable field of view (up to 1 square degree). Most telescopes built since 1950 have Ritchey–Chrétien primary mirrors (see [W2] for an historical review and thorough discussion of contemporary designs).

Two other foci deserve mention. Both require the introduction of a third mirror at D (usually with additional optics) and direct the beam to larger instruments placed at stationary locations. In equatorial telescopes, the tertiary mirror sends the beam down the polar axis to the coudé room. With a typical focal ratio of $f/45$, the plate-scale is usually $\sim 1''/\text{mm}$, enabling very high spectral resolution (see Section 1.7). In alt-azimuth telescopes, the equivalent focal positions are the Nasmyth platforms, mounted alongside the altitude rotational axis.

Most conventional telescopes, even with Ritchey–Chrétien primaries, are limited to relatively small fields of view, covering no more than ~ 1 square degree (or 0.0025% of the sky) at the prime focus. For most of the twentieth century, those telescopes were ill-suited to surveying large fractions of the celestial sphere – the

survey would simply take too long using the only available detectors, photographic plates. (Circumstances have changed recently, as outlined further in Section 1.6.3.) Specialised wide-field telescopes were designed to deal with this issue. The most successful of these survey telescopes is the Schmidt telescope, named after Bernhard Schmidt [O3]. The Schmidt primary (A in Figure 1.7, right) has a spherical surface, which would normally introduce spherical aberration. However, a corrector lens⁵ (C) placed at the centre of curvature of the primary, can ‘pre-correct’ the beam, leading to high-quality images over a field of view several degrees in diameter. The focal plane, F, is a spherically curved surface, and detectors (principally photographic plates) must be shaped to conform with that surface.

Finally, with the advent of satellite observatories in the early 1960s, it has become possible to survey the sky at wavelengths which are absorbed or scattered by the Earth’s atmosphere. Longer wavelength (infrared, far-infrared) and ultra-violet telescopes (such as IRAS, ISO and IUE) are similar in design to their ground-based optical counterparts. X-rays, however, are not reflected when they strike a mirror at normal incidence (perpendicular to the plane of the mirror) but *are* reflected when they strike at grazing incidence (that is, angles of only 1 or 2 degrees to the surface). If the mirrors have suitable parabolic and/or hyperbolic figures, then the radiation can be collected and focused as in an optical telescope [C2], [T1]. Unlike optical telescopes, these mirrors are often made from (or at least coated with) high atomic-number metals. Thus, the primary reflector in an X-ray telescope consists of a metal sleeve at the top-end of the telescope (Figure 1.8), rather than a conventional mirror. Recent X-ray missions include Einstein, ROSAT and Chandra.

1.4.2 Detectors

Astronomical observations require the detection of light at extremely low intensity levels. This in turn demands detectors which have both high quantum efficiency (QE)⁶ and low noise characteristics, so that a high fraction of the photons striking the detector are recorded, and the detector does not degrade the signal significantly in the act of recording. Photography was a mainstay of astronomical imaging and spectroscopy for more than 150 years after its invention. Photographic emulsions are sensitive primarily to radiation in the wavelength régime 2,000–9,000 Å, while photographic plates can be manufactured at a size large enough to take full advantage of the wide field of view offered by Schmidt survey telescopes. The overall QE, however, is only \sim 2–5%, and over the last two decades other more sensitive detectors have superseded photography in essentially all other roles.

Most modern optical and infrared instruments use semiconductors as detectors. Those devices rely on the photoelectric effect to detect radiation. As mentioned in

⁵ The aperture of a Schmidt telescope is the diameter of the corrector plate, not the primary mirror.

⁶ The QE of a detector is the fraction of incident photons which are detected as photoelectrons. Thus, if 100 photons fall on a detector and only five are detected, the QE is 5%.

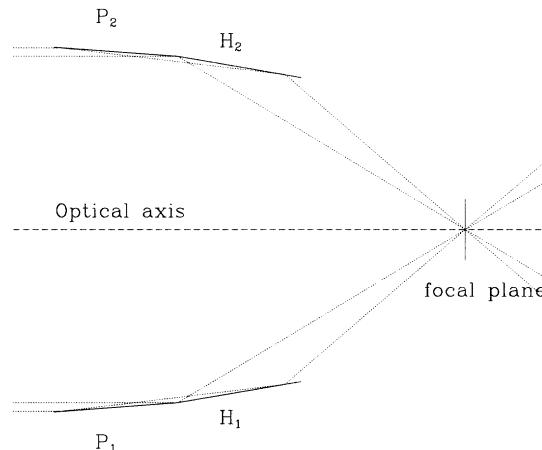


Figure 1.8. A schematic of an X-ray telescope. P_1 and P_2 are grazing-incidence paraboloidal mirrors, and H_1 and H_2 are hyperboloids.

Section 1.2, a photon of frequency ν has energy $h\nu$. If this energy exceeds the binding energy of an electron of an atom in the semiconductor, then the electron is released. Counting the number of free electrons (photoelectrons) emitted within a given time provides a measure of the number of photons striking the photosensitive material; that is, the brightness of the light source. In the original photometers, a voltage was placed across the detector, and the signal detected as a current. Even the brightest sources generate very low currents, however. At optical wavelengths, the signal can be amplified using a photomultiplier tube. A photosensitive material, such as Cs_3Sb or $GaAs$, held at a high negative potential, serves as the detector, the photocathode. The potential gradient (typically $\sim 1\text{ kV}$) leads to photoelectrons striking a series of electrodes (dynodes) in a vacuum tube (Figure 1.9), releasing many secondary electrons at each strike and producing a pulse at the anode. This amplification results in a detectable current spike.

Proportional counters operate in a similar fashion in detecting X-rays, although the primary detector is a high atomic-number gas rather than a photocathode. Again, a high-energy photon interacts with an atom, leading to the emission of a photoelectron, which is accelerated in a potential field, colliding with other atoms to produce a cascade of secondary electrons. Unlike the optical photomultiplier, proportional counters can provide some information on the frequency of the photons detected. The higher the frequency, the greater the energy and the larger the number of photoelectrons produced. Comparable detectors, with limited spectral resolution, are being developed for work at optical wavelengths.

The first photon-counting detectors provided no positional information, only the total photon rate within an area of sky defined by either the field of view of the telescope or by an aperture mask in the focal plane. Two-dimensional imaging detectors were developed in the 1970s: notably, the Imaging Photon Counting System [B5], but also reticon arrays and multichannel-plate detectors [E1]. These

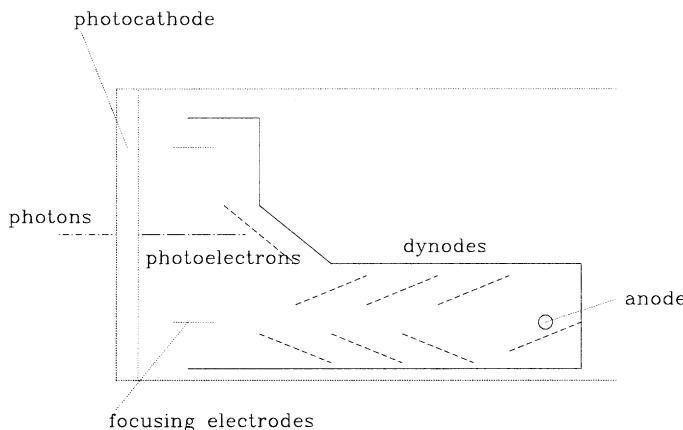


Figure 1.9. A schematic of a photon-counting detector.

provided spatial information either by accurate timing or by subdividing the photocathode into separate regions, each with its own photomultiplier chain. As with single-channel devices, each photon is detected separately, and these devices are very efficient in working at very low light-levels, particularly at blue wavelengths where the photocathodes are most efficient (20–30%). However, detecting each pulse of secondary electrons takes a finite time (typically a few microseconds), and the system cannot identify separate pulses due to other (coincident) photons arriving within that deadtime. Thus, the photon count-rate for bright sources is underestimated; indeed, if the flux level is too high, the resultant high current can damage the detector. This coincidence problem means that photon-counting devices are ill-suited for broadband, wide-field imaging, where multiple sources inevitably lead to high photon-rates, while the low efficiency of photocathodes at wavelengths longer than $\sim 6,000 \text{ \AA}$ limits their utility at red spectral regions. Given these limitations, recent instrumental development has centred on charge-coupled devices (CCDs) – array detectors which possess neither of these shortcomings, and which have become the standard detector in optical and infrared astronomical equipment.

A CCD consists of an insulating material bonded onto a doped silicon semiconductor (Figure 1.10). A grid of electrodes is embedded within the insulator, with each electrode held at a positive potential. When the device is exposed to light, freely-moving photoelectrons are produced in the doped semiconductor, and these (or the positive ‘holes’) migrate towards the local electrode. The efficiency is increased, and the noise contribution from random motion of electrons (dark current) reduced, if the device is cooled to temperatures of $\sim 100\text{--}170 \text{ K}$, usually using liquid nitrogen or liquid helium as a refrigerant. The total charge collected by each electrode is a measure of the photon flux incident on that ‘pixel’ (picture element) of the detector. By manipulating the voltages appropriately, the charge collected in each pixel can be read out at the end of an exposure, and hence the image is reconstructed. Reading out the CCD chip contributes additional noise to the counts detected in each pixel (readout noise), but this is a small

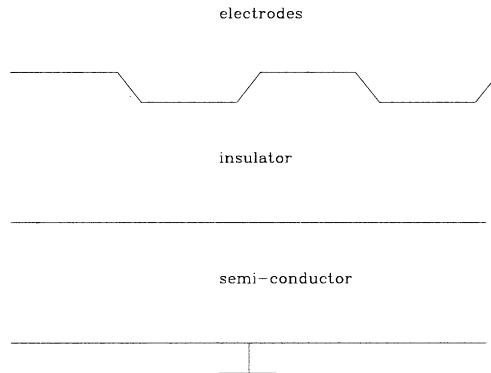


Figure 1.10. A schematic of a CCD array.

price to pay for quantum efficiencies of 30–40% at blue wavelengths and 80–90% at 6,000–9,000 Å.

The largest optical CCDs currently manufactured consist of arrays of $4,096 \times 4,096$ pixels, with individual pixels 12–15 µm in size; 2,048-square optical devices with 12–25 µm pixels are also in use on many telescopes, while the largest infrared arrays have 1,024 pixels to a side. None of the individual chips is much larger than $\sim 5.5 \times 5.5$ cm, so a single-CCD camera cannot provide the same areal coverage as a single photographic plate. However, the solid angle surveyed can be increased by constructing arrays of many CCDs, as has been done at the Japanese Kiso Schmidt [S1], and in the main camera used in the Sloan Digital Sky Survey [G1]. The latter camera includes 30 2,048-square CCDs, which sparsely sample the sky in a 6×5 grid on the focal plane, scanning a strip of sky $\sim 2^\circ 5$ wide. With improvements of nearly a factor of ten in quantum efficiency over photographic plate material, these arrays can detect objects substantially fainter than the limiting magnitude of the large-scale photographic sky surveys.

The response of optical CCDs declines significantly beyond 9,000 Å, but specialised infrared-array detectors can be used to cover the 1–5 µm region of the spectrum. These devices are fabricated from semiconductors such as InSb and HgCdTl (mercadtelluride), and the technology is less advanced than is the case for the visual-wavelength detectors. However, increasing numbers of 1,024-square arrays are becoming available at observatories throughout the world. PbS and SiAs detectors are used at wavelengths beyond 5 µm, where low atmospheric transmission and the high background due to thermal emission make observations extremely difficult. For highest efficiency, the detectors have to be cooled to substantially lower temperatures than optical devices, usually requiring liquid helium as a coolant.

1.5 STELLAR PHOTOMETRY

The telescopes and detectors discussed in the previous section are used astronomically for two main purposes: first, direct imaging, usually for photometry (measuring

the brightness of individual objects); and second, spectroscopy, determining the energy distribution as a function of wavelength. The following sections outline the fundamental principles involved in these observational measurements.

1.5.1 The magnitude system

The apparent brightness of astronomical objects is usually measured in units of *magnitude*. The system originated with Hipparchus' division of the naked eye stars into six subgroups, with the brightest stars grouped together in the 'first magnitude' and the faintest stars visible to the naked eye described as being of the 'sixth magnitude'. The human brain/eye combination tends to judge brightness differences as ratios, rather than linear differences. If there are three light sources, A, B and C, where B is twice as bright as A and C twice as bright as B, a visual observer will estimate the difference between A and B as the same as that between B and C, although in linear terms, the relative brightnesses are 1, 2 and 4, respectively. The result is that the magnitude scale is logarithmic, rather than linear, and a given difference in magnitude corresponds to a particular brightness ratio.

Pogson [P2] quantified Hipparchus' original qualitative scale into a system where a difference of five magnitudes is equivalent to a factor of 100 in apparent magnitude, retaining the convention of numerically-increasing magnitudes with decreasing intensity. Hence, magnitude is defined as

$$m = -2.5 \times \log_{10}(f) + \text{constant} \quad (1.15)$$

where f is the apparent flux (in $\text{Watts m}^{-2} \text{Hz}^{-1}$, $\text{erg sec}^{-1} \text{cm}^{-2} \text{\AA}^{-1}$ or equivalent units). One of the striking advantages of this convention is that the enormous brightness difference of 10^{21} between the apparent magnitude of the Sun (magnitude –26) and the faintest object detectable by the Hubble Space Telescope (magnitude 30) spans only 56 magnitudes. Thus, the magnitude system expresses large brightness differences in a compact, and widely understood, form. It is primarily for this reason that, despite the rumblings of some astrophysicists (for example, [L9]), the system remains in common use today.

The magnitude scale is defined as

$$m_p = -2.5 \log_{10} \frac{f_p}{F_0} \quad (1.16)$$

where f_p is the measured flux emitted by the source within a particular wavelength region (passband), usually defined by optical filters, and F_0 is the flux density produced by a star which has magnitude 0 in that passband. The latter zero-point is arbitrary, but is usually set to give an A0 star (comparable to Vega) equal magnitudes at all wavelengths. The main exception to this convention is the Gunn ugriz system, used with the Sloan Digital Sky Survey (Section 1.6.3), which adopts a uniform zero-point in all passbands. Magnitudes in the Gunn system are defined as

$$m = -2.5 \log_{10} f_\nu - 48.60 \quad (1.17)$$

where f_ν is the flux in $\text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}$. These are known as AB magnitudes.

Table 1.1. Broadband filter characteristics.

Band ¹	$\Delta\lambda$ μm	λ_{eff} A0	λ_{eff} M0	λ_{eff} M2	λ_{eff} M6.5	F ₀ Janskys	Reference ²	m ³ _{sky}
U	0.325–0.395	0.366	0.366			1,181	J2/B2	22.05
B	0.39–0.49	0.44	0.44			4,520	J2/B2	22.4
V	0.50–0.59	0.542	0.551			3,711	B3/L7	21.3
R	0.565–0.725	0.638	0.722			3,180	B3/L7	20.6
I	0.73–0.88	0.787	0.800	0.809	0.818	2,460	L7	18.8
J _{CIT}	1.16–1.35	1.22		1.257	1.256	1,568	L7	15.5
H _{CIT}	1.49–1.80	1.63		1.635	1.633	1,076	L7	14.5
K _S	2.00–2.30	2.15		2.15		650	P1	13.3
K _{CIT}	2.02–2.43	2.19		2.211	2.209	674	L7	13.0
L _{CIT}	3.24–3.73	3.45			3.472	281	B4	8.0
L'	3.52–4.12	3.80			3.80	235	B4	
M	4.5–5.05	4.75	4.75			154	B4	
u	463	3,540				3,631	F3, S5	22.7
g	988	4,750				3,631	F3, S5	22.0
r	955	6,222				3,631	F3, S5	20.9
i	1,064	7,632				3,631	F3, S5	20.1
z	1,248	9,049				3,631	F3, S5	18.8

¹ These passbands are the Cousins UBVRI system as characterised by Bessell ([B2], [B3]), the Caltech (CIT) version of the Johnson near-infrared JHKL system and the Gunn ugriz system [F3]. The K_S (K-short) passband is truncated at long wavelengths to minimise thermal background radiation. This passband is used in the DENIS and 2MASS near-infrared sky surveys (Chapter 6). Similarly, the L' passband avoids the worst atmospheric absorption in the CIT L-band.

² References are for flux zero-points in Janskys – units of 10⁻²⁶ Watts m⁻² Hz⁻¹.

³ m_{sky} is the sky brightness in magnitudes arcsec⁻¹ at a dark-sky site, such as Mauna Kea. Sky brightness longward of ∼2.3 μm is dominated by thermal radiation from the atmosphere, and is therefore highly variable on short time-scales. The sky brightness measurements for the Gunn filters were computed by A. West, J. Tucker and J. Gunn (priv. commun., 2004).

Broadband photometric systems, with filter-defined passbands of full-width at least 500 Å, are used to map the overall spectral energy distribution of celestial objects; narrowband filters are designed generally to examine features in the energy distributions of particular types of star. The most frequently used broadband photometric system is Johnson/Cousins UBVRIJHKLM [J2], [C1], spanning the wavelength range 3,000 Å to 5 μm with each passband having a width of ∼1,000 Å (Table 1.1). Within each passband, the measured flux density corresponds to the stellar flux at the effective wavelength of the filter. The latter quantity is found by convolving the spectral energy distribution of the star ($S(\lambda)$) with the shape of the filter bandpass ($B(\lambda)$):

$$\lambda_{eff} = \frac{\int \lambda S(\lambda) B(\lambda) d\lambda}{\int S(\lambda) B(\lambda) d\lambda} \quad (1.18)$$

The effective wavelength can vary depending on the spectral energy distribution

of the target (the spectral type of the star observed). For example, decreasing temperature moves the peak in the emergent energy distribution towards longer wavelengths, steepening the spectral slope at optical wavelengths and moving the effective wavelength to the red. This effect is particularly important in the broad Cousins R-band, where λ_{eff} changes from 6,380 Å for an A0 star to 7,220 Å at spectral type M0 [B2]. Other passbands are less affected, with typically 200 Å differences in λ_{eff} between spectral types A0 and M0.

Photometric colours are defined as a magnitude difference; for example, $(B - V)$. From the definition of magnitude (equation 1.15), a ‘colour’ therefore measures the flux ratio in the two passbands:

$$\text{mag}_1 - \text{mag}_2 \equiv \frac{\text{flux}_1}{\text{flux}_2} \quad (1.19)$$

Traditionally, colours are expressed as the shorter wavelength magnitude *minus* the longer wavelength, so a negative (blue) colour implies $f_{\text{short}}/f_{\text{long}}$ is high, and a positive (red) colour indicates that $f_{\text{short}}/f_{\text{long}}$ is low. As discussed further in Chapter 2, colours provide a means of estimating stellar temperatures.

1.5.2 Measuring magnitudes

Observationally, the standard photometric systems were defined using photoelectric photometers and aperture photometry techniques, but most current measurements are made with array detectors. In the former case, a photomultiplier tube was used to measure the brightness of a given source through a circular aperture (usually 20–30 arcsec diameter); in the latter case, software techniques are used to measure the flux within a given radius centred on the star. In either case, since the aim is to measure a large fraction of the stellar flux, the effective aperture size chosen depends on the prevailing atmospheric seeing and guiding accuracy. However, the underlying night sky also contributes to this measurement and this contribution must be subtracted, either by obtaining a separate offset measurement in aperture photometry, or from the ‘sky’ pixels immediately adjacent to the object in array photometry.

Typical surface brightness values for a dark-sky site are listed in Table 1.1. Airglow (primarily emission from OH and OI; see [R2]) is a strong contributor shortward of 2 μm, and the sky brightness at these wavelengths is well-correlated with the solar cycle, being higher when the Sun is more active [L6]. Volcanic eruptions can also affect m_{sky} if dust is introduced into the upper atmosphere, where it gradually diffuses around the Earth [L8]. The eruptions of the Mexican volcano El Chichón (1982) and the Philippine volcano Pinatubo (1990) not only produced spectacular sunsets, visible for several months afterwards in North America and Europe, but also raised the night sky brightness (and the optical extinction) for more than two years. At longer wavelengths, thermal radiation from the atmosphere and telescope dominates the background.

The signal-to-noise of an observation is given (following [L9]) by

$$\frac{S}{N} = \frac{A_{\text{eff}} N_\nu \Delta\nu t}{[A_{\text{eff}} N_\nu \Delta\nu t + \Omega A_{\text{eff}} S_\nu \Delta\nu t + Dt + R^2]^{0.5}} \quad (1.20)$$

where t is the integration time; $\Delta\nu$, the bandwidth (Hertz); A_{eff} , effective area of telescope, in m^{-2} ; N_ν , the source flux density, in photons, $\text{m}^{-2} \text{s}^{-1} \text{Hz}^{-1}$; S_ν , sky brightness, in photons, $\text{m}^{-2} \text{s}^{-1} \text{Hz}^{-1} \text{arcsec}^{-1}$; Ω , solid angle of effective aperture (either the physical diameter of the aperture in the photoelectric photometer, or the circle of integration used in analysing the array photometry); D , dark current; and R , readout noise (zero for aperture photometry).

Modern detectors have both low dark-current and low read-noise (3–7 electrons is typical for a CCD working at liquid nitrogen temperatures). Consequently, photon statistics in the sky level constitute the dominant source of noise in photometry of faint objects. In aperture photometry, with a photomultiplier tube, the sky measurements are made separately from the object (+sky) observations. One of the major advantages of array photometry is that the sky level can be determined from the same exposure used to measure the source. Moreover, a 2,048-square optical CCD array covers a typical solid angle of at least 150 square arcmin, encompassing many stellar (and non-stellar) objects in a deep exposure on even a moderate-sized telescope. Since these objects are all observed simultaneously, high accuracy relative photometry is possible even during inclement conditions. A caveat is that individual CCD pixels can have slightly different sensitivities, so the intensity levels in each frame must be normalised using a flat-field exposure – an image made by illuminating the CCD with a diffuse, uniform light-source.

Once CCD images have been normalised, sophisticated profile-fitting techniques can be used to determine the relative flux of each object, minimising the contribution from sky-noise. Clearly, the more concentrated the stellar profile (the better the seeing), the smaller the solid-angle for profile-fitting, the lower the contribution from sky-noise and the fainter the limiting magnitude attained in a given exposure time. The instrumental flux measurements themselves are calibrated through observations of standard stars with well-determined magnitudes on particular photometric systems. Extensive lists are provided by Landolt [L2], [L3] for the frequently-used Johnson/Cousins optical system (Table 1.1), and Persson *et al.* [P1] provide standard-star lists for the near-infrared.

In general, an instrumental magnitude is measured for each source, defined by (taking the V-band as an example)

$$v = -2.5 \log \left(\frac{N_V}{t} \right) \quad (1.21)$$

where N_V is the total number of counts measured for the source in an integration time of t seconds. Observations of standard stars are used to solve for the constants in an equation of the form:

$$V = v + k_v \times \sec(z) + C_v \times (V - I) + Z_v \quad (1.22)$$

where V is the magnitude on the standard system; v , the instrumental magnitude; z , the angular distance from the zenith; k_v , the extinction coefficient; C_v , the colour term; and Z_v , the zero-point. The extinction term corrects for absorption through the Earth's atmosphere ($\sec(z)$ is known as the ‘airmass’ of an observation) using a plane-parallel approximation for atmospheric depth along the line of sight. The colour term (which could also be $(B-V)$ or $(V-R)$ in this case) allows for a potential mismatch between the effective wavelength of the reference system and that used in the observations. Defining these terms accurately requires repeated observations of a reasonable number (15–20) of standard stars, well distributed in both colour and airmass. A more thorough discussion of photometric techniques is given by Henden and Kaitchuck [H2].

The observed magnitude of a star is termed the apparent magnitude (m). The absolute magnitude (M) is defined as the apparent magnitude that a source has at a distance of 10 parsecs. Hence, since brightness decreases with the square of the distance:

$$m = M + 5 \times \log_{10}(r) - 5 \quad (1.23)$$

The quantity $(m - M)$ is known as the distance modulus, a term often used in citing the distances to star clusters.

1.5.3 Bolometric magnitudes and effective temperatures

Summing the total energy emitted at all wavelengths for a star determines its *bolometric* magnitude,

$$m_{bol} = -2.5 \log_{10}(f_{tot}/F_0) \quad (1.24)$$

Given a known distance, this can be converted to the absolute bolometric magnitude which, in turn, can be expressed as a luminosity, usually in solar units. Based on absolute measurements of the energy distribution, primarily from satellite data, the Sun has an absolute visual magnitude of $M_V \simeq 4.79$ [L5]. The bolometric correction is approximately -0.12 magnitudes [B7], so $M_{bol}(\odot) \simeq 4.67$, corresponding to $L_\odot \simeq 3.83 \times 10^{26}$ Watts. Stellar luminosities are then given by

$$-2.5 \log_{10}\left(\frac{L}{L_\odot}\right) = M_{bol}(\text{star}) - 4.67 \quad (1.25)$$

The luminosity of a star is also used to define the quantity known as the effective temperature. An ideal radiator of temperature T produces a pure continuum spectrum, with neither absorption nor emission features, and a spectral energy distribution that is described by the Planck formula. The flux from the black-body distribution can be written as a function of frequency:

$$F_\nu \delta\nu = \frac{2\pi h}{c^2} \frac{\nu^3 \delta\nu}{e^{\frac{h\nu}{kT}} - 1} \quad (1.26)$$

where h is the Planck constant, k is the Boltzmann constant, c is the velocity of light, T is the surface temperature and ν is the frequency. In this case the usual units for

the flux, F_ν , are Janskys, where 1 Jansky is 10^{-26} Watts $\text{m}^{-2}\text{Hz}^{-1}$ (MKS units). Alternatively, the Planck curve can be written in wavelength units:

$$F_\lambda \delta\lambda = \frac{2\pi hc^2}{\lambda^5} \frac{\delta\lambda}{e^{\frac{hc}{\lambda kT}} - 1} \quad (1.27)$$

where the usual units are $\text{erg cm}^{-2}\text{\AA}^{-1}$ or $\text{erg cm}^{-2}\mu\text{m}^{-1}$ (c.g.s. units). Integrating this distribution gives Stefan's law, which states that the total energy emitted is proportional to the product of the surface area and the fourth power of the temperature. This leads to the definition of the effective temperature of a star:

$$L = 4\pi R^2 \sigma T_{\text{eff}}^4 \quad (1.28)$$

where σ is Stefan's constant, $\frac{2\pi^5 k^4}{15c^2 h^3}$. The effective temperature, T_{eff} , is defined by equation (1.28) as the temperature of a black-body of the same radius as the star which radiates the same total energy.

The black-body distribution peaks at a wavelength, λ_{max} , whose value varies inversely with the temperature. The result is Wien's law, which can be derived by determining when the derivative of equation (1.27) is zero. The solution is

$$\lambda_{\text{max}} = \frac{2,898}{T} \mu\text{m} \quad (1.29)$$

Hence, the Sun, with $T_{\text{eff}} = 5,777\text{ K}$ [L4], has an emergent spectrum which peaks at $0.502\mu\text{m}$. In contrast, emission from the Earth's atmosphere at $T \sim 300\text{ K}$ peaks at close to $10\mu\text{m}$.

At wavelengths well longward of λ_{max} , the quantity $hc/\lambda kT$ is small, and $e^{\frac{hc}{\lambda kT}} \approx 1 + hc/\lambda kT$. Thus, equation (1.27) becomes

$$F_\lambda \propto T \lambda^{-4} \quad (1.30)$$

This is the Rayleigh–Jeans approximation, and the black-body spectrum at $\lambda \gg \lambda_{\text{max}}$ is referred to as the Rayleigh–Jeans tail.

1.5.4 Interstellar absorption

Dust particles are found in higher-density regions of the interstellar medium, which are largely confined to latitudes within 5–10 degrees of the Galactic plane. These particles modify the spectral energy distribution of the background stars: large particles ($>100\mu\text{m}$) produce Mie scattering, which has a relatively weak wavelength dependence (λ^{-1}); the smallest particles produce Rayleigh scattering, as in the Earth's atmosphere, which has λ^{-4} dependence and preferentially scatters shorter wavelength radiation. Both effects produce extinction, a net dimming of the source, and ‘redden’ the overall spectral energy distribution. This interstellar reddening is usually quantified by estimating the ‘colour excess’ of a star – the difference between the observed colour, usually ($B-V$), and the intrinsic colours, estimated from the

spectral type (see Chapter 2). Identifying the latter with the subscript 0, the colour excess is

$$E_{B-V} = (B - V) - (B - V)_0 \quad (1.31)$$

Dust grains in most regions of the interstellar medium have similar properties, so the colour excess at other wavelengths can be scaled to the $(B - V)$ colour excess [B4]. Thus,

$$E_{V-I} = 1.25 \times E_{B-V}, \quad E_{V-K} = 2.78 \times E_{B-V} \quad (1.32)$$

and the total absorption is given by

$$A_V = V - V_0 = 3.12 \times E_{B-V}, \quad (1.33)$$

$$A_K = K - K_0 = 0.34 \times E_{B-V}, \quad (1.34)$$

The quantity $\frac{A_V}{E_{B-V}}$ is known as the ratio of total to selective absorption, and this parameter can increase in high-density regions, probably due to the grains having a different composition and different size distribution [W1]. The above equations clearly demonstrate the advantages of using infrared observations to probe dusty regions near the Plane of the Galaxy.

1.6 SKY SURVEYS

Large-scale surveys, providing imaging and photometric catalogues for substantial fractions of the celestial sphere, are now available at a wide range of wavelengths. Over the last few years, these surveys have become invaluable research tools for investigating the properties of intrinsically rare objects. This section outlines the main characteristics of the surveys that have proven most useful in searches for low-mass stars and brown dwarfs.

1.6.1 Photographic surveys

The first large-scale imaging surveys were based on photographic plates. They have supplied the foundation for many large-scale astronomical projects over the last fifty years or more. The most widely used are the northern and southern sky surveys, based on plates taken, respectively, by the 48-inch Palomar (now Oschin) and UK (now Anglo–Australian Observatory) Schmidt telescopes. The first of these surveys, the Palomar Observatory/National Geographic Sky Survey (POSS I), was taken between 1949 and 1957, and was designed specifically to provide faint targets for the newly completed Palomar 200-inch telescope. The original survey used blue and red-sensitive emulsions (Kodak 103aO and 103aE, respectively) to cover the sky to $\delta > -33^\circ$ [M4]. Each 14-inch square plate covers 6.5×6.5 degrees to a limiting magnitude of 20 to 20.5. That survey was extended in the red to -48° (or 15° altitude) in the early 1960s.

The UK Schmidt telescope is essentially a copy of the Palomar Schmidt, constructed in the 1970s to complement the Anglo–Australian 3.9-m telescope at Coonabarabran, New South Wales. Initially, the UK Schmidt surveyed the southern skies using the newly developed blue–green IIIaJ emulsion, while the 40-inch ESO Schmidt (at La Silla Observatory, Chile) provided an *R*-band survey using Kodak 098 emulsion; the UK Schmidt later added an *R*-band survey using IIIaF emulsion. The type IIIa emulsions are significantly more sensitive than the POSS I plate material, reaching limiting magnitudes of 21–22. Moreover, the development of IVN emulsion provided the first opportunity for wide-field imaging in the I-band, marking a major advance in surveys for low-mass stars, as discussed further in Chapter 8. With the increased sensitivity and wavelength coverage, the northern survey was repeated in the late 1980s and 1990s as POSS II [R3], while second epoch southern surveys with blue (IIIaJ) and red (IIIaF) plates were taken at the UK/AAO Schmidt in the late 1990s [M5].

All of these photographic surveys have been copied and widely distributed on both glass and film. In addition, the original plates have been scanned and digitised by several research groups with access to automated measuring engines. In particular, the plate scans made by the research group at the Space Telescope Science Institute, led by the late Barry Lasker, are freely available as the Digitised Sky Survey (L11, L12). This database has played a crucial role in a multitude of astronomical projects, ranging from brown dwarf searches through galaxy cluster studies to analysis of gamma-ray bursts. Over the last few years, the Palomar Schmidt has been converted to a wide-field CCD imager, while the AAO Schmidt now spends a substantial fraction of its time taking spectra, using optical fibres placed on the focal plane to feed a bench-mounted spectrograph. POSS II and the second-epoch southern sky surveys represent the final era of photographic sky surveys.

1.6.2 Infrared surveys

The infrared is a particularly crucial wavelength regime for finding and studying low-mass stars and brown dwarfs, but, until recently, technology placed severe constraints on the sensitivity and image quality of observations at these wavelengths. The first infrared detectors were developed in the early 1960s, but were limited to single-element detectors. Under those circumstances, aperture photometry is the only technique available for detecting sources and measuring magnitudes. As outlined in Section 1.5.2, the brightness of an object is determined by measuring the total flux with the source centred in the aperture, then offsetting to a blank sky position to measure the contribution from the sky background. The near-infrared sky is bright and rapidly variable, particularly longward of $2\,\mu\text{m}$ (Table 1.1), where thermal emission from the telescope also contributes. As a result, observers have to cycle rapidly between observations on-source and off-source; this is usually achieved by using a moving mirror to ‘chop’ at 5–10 Hz between two positions separated by 30–40 arcseconds on the sky, Beam A (source) and Beam B (sky). In addition, the telescope is ‘nodded’ every few seconds to move the source to Beam B and sample a second sky region.

Aperture photometry is effective, but slow. To give a specific example, the sky brightness is $K = 8.2$ magnitudes within a 10 arcsecond aperture, a typical size for photometry. If we assume that the overall observing efficiency (time integrating on the target) is 30% and make due allowance for optical light losses, then a 1-m telescope detects $\sim 300,000$ photons s $^{-1}$ from the sky alone. Applying the signal-to-noise formula (1.20), we find that a 10 second exposure of a $K = 13.5$ source should give photometry of 10% accuracy. This is a respectably faint magnitude. However, with this setup we would require approximately 7×10^9 observations (and $\sim 2,200$ years integration time) to survey the sky. This underlines why deep, wide-field surveys had to await the development of infrared arrays.

Surveys at near-infrared wavelengths

The infrared is subdivided based on detector technology. Thus, the near infrared is generally defined as covering the JHK_L bands (1 to ~ 4 μm), where observations can be made using PbS, InSb and HgCdTd detectors. The first sky survey at these wavelengths was the Two Micron Sky Survey (TMSS)[N1]. The observations were made using a custom-built 62-inch f/1 aluminised plastic mirror to feed two detectors: a set of 8 PbS cells, mounted in four pairs along a north–south axis and equipped with 2.0–2.5- μm filters; and an adjacent Si detector with an I-band filter. Each PbS detector covered 10×10 arcminutes, so the array of detectors spanned 40 arcminutes in declination, while the Si detector had a 20-arcminute field of view. Observations were made by scanning the telescope in RA at 15 or 30 times the sidereal rate, chopping back and forth between the two sets of PbS detectors using the primary mirror. The limiting magnitudes for detecting sources were $I \sim 9$ and $K \sim 5$, but the final catalogue was limited to 3rd magnitude at K . In total, the TMSS has only 5,612 catalogued sources – comparable to the number of stars visible to the naked eye. The TMSS catalogue, however, includes some of the most unusual infrared sources in the Galaxy. In particular, this survey provided the first indication of the dramatic importance of mass loss and circumstellar dust shells in the later stages of evolution of red giant stars.

The TMSS stood alone for over three decades, until the development of large-format infrared arrays. Over the last decade, two separate, but similar, projects have exploited these advances, and surveyed the sky to flux levels more than 10,000 times fainter than reached by the TMSS. The source catalogues generated by the Deep Near Infrared Survey (DENIS) and the 2-Micron All Sky Survey (2MASS) include more than 200 million objects.

The DENIS project is a European/Brazilian collaboration, headed by a group from Observatoire de Paris [E3], [C5]. The survey covers wavelengths from 0.8–2.5 μm , using a 1,024-square CCD detector to obtain I -band data, while a pair of 256-square HgCdTd detectors take data in the J and K_S passbands. These cameras are mounted on a 1.0-m telescope at the European Southern Observatory, La Silla, Chile, and the survey is limited to the southern skies, $-88^\circ < \delta < +2^\circ$. The survey, which reaches limiting magnitudes of $I \sim 17.5$, $J \sim 16$ and $K_S \sim 13$, started in

January 1996 and was completed in early 2001, with the final source catalogue scheduled to be released in 2004/2005.

In contrast, 2MASS, as its name implies, is an all-sky survey, combining observations from two 1.3-m telescopes (at Mt. Hamilton, Arizona, and Cerro Tololo, Chile) to cover both northern and southern hemispheres [S4]. Rather than combining optical and near-infrared, 2MASS used three HgCdTd 256-square arrays to obtain data simultaneously in the J , H and K_S passbands. With larger telescopes and a slightly longer effective integration time, 2MASS is more sensitive than DENIS, reaching limiting magnitudes of $J \sim 16.5$, $H \sim 15.8$ and $K_S \sim 14.5$. First light for the survey was in June 1997 in the North and March 1998 in the South. Observations were completed in 2001, and the final catalogue was released in January 2003. As discussed further in Chapter 6, both DENIS and 2MASS have had an enormous impact on our understanding of low-mass stars and brown dwarfs.

Surveys in the mid-infrared

The mid-infrared runs from $\sim 5\text{ }\mu\text{m}$ (M -band) to $\sim 30\text{ }\mu\text{m}$. Observations at those wavelengths are hampered not only by a brighter background, due to thermal emission, but also by numerous, deep terrestrial absorption bands (Figure 1.1). As a result, mid-infrared surveys are generally conducted from above the atmosphere.

The mid-infrared equivalent of the TMSS is the Air Force Geophysics Laboratory (AFGL) four-colour survey, covering 90% of the sky at wavelengths of 4.2, 11.0, 19.8 and $27.4\text{ }\mu\text{m}$ [P3]. The observations were made using 16.5-cm telescopes, lifted to the upper boundaries of the atmosphere in a series of nine rocket launches. A total of 2,363 sources were detected to a flux level of $\sim 50\text{ Jy}$ at $4.2\text{ }\mu\text{m}$. As with the TMSS, these objects are primarily of high intrinsic luminosity, and include evolved stars and many of the nearer star-forming regions.

The successor to the AFGL survey was IRAS, the InfraRed Astronomical Satellite, launched in January 1983 as a joint venture by the USA, UK and the Netherlands. The main mission project was an all-sky survey in four passbands, centred at 12, 25, 60 and $100\text{ }\mu\text{m}$. The satellite used 62 Ge:Ga detectors to cover a 30-arcminute field. Individual sources were identified and extracted from continuous scans, and positions, flux densities and crude morphological parameters determined. Scanning was completed in November 1983, and the initial source catalogues were issued the following year. The Point Source Catalogue includes 250,000 sources with flux densities exceeding 0.5 Jy at 12, 25 and/or $60\text{ }\mu\text{m}$, one hundred times fainter than the AFGL survey. Using more sophisticated data analysis, that limit was extended to 0.17 Jy for the Faint Source Catalogue, which included 173,000 sources with galactic latitudes $|b| > 10^\circ$.

The next step in mid-infrared surveys was taken very recently with the launch, on 25 August 2003, of the Spitzer space telescope (formerly known as SIRTF, the Space InfraRed Telescope Facility). Unlike IRAS, which was placed in a conventional 90-minute orbit, Spitzer is in a solar orbit, radius 1 AU, trailing the Earth at a distance that will gradually increase to 0.32 AU by early 2006. Spitzer has an 85-cm diameter telescope and three instruments: the InfraRed Array Camera (IRAC),

providing imaging at 3.5, 4.5, 6.3 and 8.0 μm ; the Multiband Imaging Photometer Spectrometer (MIPS), covering 20–160 μm with imaging and low-resolution spectroscopy; and the InfraRed Spectrograph (IRS), providing low to moderate resolution in the 5–40- μm wavelength regime. Observations have been underway since late 2003, reaching limiting sensitivities close to the predicted values of $\sim 1 \mu\text{Jy}$ (IRAC), 0.01 to 0.1 mJy (MIPS imaging) and 0.3 mJy (IRS). Spitzer has a projected lifetime of 2.5 to 5 years. Within that period, it is scheduled to complete several projects that will provide multi-waveband surveys covering 100–1,000 square degrees.

1.6.3 Deep optical imaging surveys

CCDs offer much higher quantum efficiencies at optical wavelengths than photographic emulsions, but it is only within the last decade that technology has progressed to the point that true wide-field surveys can be undertaken. As noted above, the Oschin Schmidt on Palomar Mountain, like the Kiso Schmidt telescope [S1], is now equipped with an array of CCD detectors mounted on the curved focal plane, and is being used for a variety of programs, including searching for near-Earth asteroids and a general-purpose imaging survey (Palomar–Quest). In addition, several 4-m telescopes, including the Mayall telescope at Kitt Peak National Observatory, the Blanco telescope at Cerro Tololo and the Canada–France Hawaii telescope on Mauna Kea, are equipped with multi-CCD imaging cameras that cover ~ 1 square degree and reach 23rd magnitude in a matter of seconds.

The Sloan Digital Sky Survey (SDSS) is the principal large-scale, CCD-based survey currently under way. Covering approximately π steradians (1/4 of the sky) at high galactic latitudes, SDSS is a combined imaging and spectroscopic survey, aimed primarily at studying the large-scale distribution of galaxies. The direct images are being used to identify $\sim 1,000,000$ field galaxies, which are then targeted for multi-object spectroscopy using optical fibres. However, the multi-waveband imaging data are also extremely useful for a wide variety of research projects in many other areas, including low-mass stars [W4] (see Chapter 5).

SDSS uses thirty 2,048-square CCDs mounted at the focal plane of a 2.5-m telescope at Apache Point Observatory, New Mexico, USA. These are used to scan the sky in five passbands (ugriz, Table 1.1) to limiting magnitudes of ~ 23.1 in r, 22.5 in i and 20.8 in z. The survey started in June 1998 and completion is expected in mid-2005, although interim data sets covering part of the survey area have already been released. As discussed further in Chapter 6, SDSS has already proven an invaluable asset in searching for T dwarfs, the coolest brown dwarfs currently known.

1.6.4 X-ray surveys

M dwarfs prove to be surprisingly luminous at X-ray wavelengths (see Chapter 5). Thus, X-ray surveys are also pertinent to analyses of the properties of late-type dwarfs. The Earth's atmosphere is completely opaque to X-ray wavelengths, a

fortunate circumstance for terrestrial lifeforms. However, as a result, X-ray observations can only be made using telescopes that have been taken above the bulk of the atmosphere using balloons, sounding rockets or orbital satellites. The two satellite observatories that have been most useful for M dwarf studies are Einstein and ROSAT.

The Einstein Observatory was launched, as High Energy Astrophysics Observatory-2 (HEAO-2), in November 1978 and operated until April 1981 [G3]. It was the first satellite to use the glancing-incidence mirror technology illustrated in Figure 1.8, and therefore the first satellite to give both accurate positions (\sim 2–4 arcseconds within the 30-arcminute field of the High Resolution Imager) and high sensitivities (10–100 μ Jy) for X-ray sources. The Einstein instruments were capable of detecting photons with energies between 0.16 and 3.5 keV (corresponding to wavelengths between 75.5 and 3.5 Angstroms). Einstein did not conduct an all-sky survey, but, with the large field of view encompassed by the primary imaging instruments, the 5,000 pointed observations, which included observations of the Hyades and Pleiades clusters, provided the first opportunity to measure the X-ray properties of average stars.

Einstein's principal successor was the Roentgen-Satellit, ROSAT, a joint venture between the US, the UK and Germany. Launched on 1 June 1990, ROSAT used its first 6 months of observations to produce the first imaging survey of the entire sky at X-ray wavelengths, the ROSAT All-Sky Survey. The Bright Source Catalogue (RASS-BSC) includes 18,811 sources with fluxes brighter than \sim 60 mJy (the formal limit is >0.05 counts/second) between 0.1–2.4 keV (5.1–124 Angstroms) [V4]; the Faint Source Catalogue (RASS-FSC) includes almost 106,000 sources, each with at least 6 detected photons. In comparison, the most extensive previous all-sky catalogue, by HEAO-1, included only 841 sources. ROSAT remained operational until 12 February 1999, primarily undertaking targeted observing programs of specific objects for individual Guest Observers.

1.6.5 Digitised surveys and virtual observatories

All of the catalogues and surveys described in this section are, or will shortly be, available to the astronomical community through the worldwide web. This marks a distinct new trend in astronomical research. Extensive resources in both the USA and Europe are being devoted to devising more efficient ways of linking and interrogating these and other data sets, providing a ‘virtual observatory’ that, like telescopes at actual observatories, can be used to identify and collect ‘observations’ that are designed to answer specific scientific questions.

The goal is to create a whole that really is greater than the sum of its parts – a network combining data from X-ray to radio, that not only permits the identification of intrinsically rare sources, with unusual energy distributions, and an accurate characterisation of the properties of ‘average’ stars and galaxies, but also taps into the astronomical literature to provide background information on specific objects. A true virtual observatory is still some years distant, but the growing trend toward freely accessible astronomical databases is already widening the research community.

Those data sets are being used not only by professional astronomers, but also by many amateur astronomers. And with the increasing capabilities of both personal computers and the instrumentation available for backyard telescopes, those amateurs are already making interesting discoveries; in the near future, they can be expected to make a growing contribution to research in many areas.

1.7 SPECTROGRAPHS AND SPECTROSCOPY

Photometry allows estimation of stellar luminosity (given a known distance) and surface temperatures; spectroscopy permits measurement of individual spectral lines, which can be used to determine composition, temperature and motion toward or away from the observer (the radial velocity). Astronomical spectroscopy has its origins in Newton's optical experiments in the 1660s. Placing a glass prism in a beam of sunlight passing through a hole in the window shutters of an otherwise darkened room, Newton discovered that 'white' sunlight was actually compounded of a mixture of different colours. Relatively little use was made of this discovery until the early nineteenth century, when first Wollaston (1802) and then Fraunhofer (1815) discovered that the solar spectrum was not a continuous band of colours, but exhibited a number of dark bands. The nature of those bands, whether intrinsic to the Sun or terrestrial in origin, remained controversial until 1861, when Kirchhoff [K2] identified the Fraunhofer D-line with the element sodium. Once it became clear that the individual spectral features provided signatures of the presence of different atomic and molecular species within the solar and stellar atmospheres, it became possible to devise stellar classification schemes (see [H1] for a thorough historical review). Spectral classification led to the recognition of underlying patterns of behaviour, eventually formalised as the Hertzsprung–Russell diagram (discussed in more detail in Chapter 2).

The first spectrographs used prisms to disperse the incident starlight, but modern instruments generally use ruled gratings as the principal dispersive optical element. Instruments designed to work at ultraviolet, visual and near-infrared wavelengths – the wavelengths of most interest for M dwarf studies – follow the same basic optical design, although the optical elements in infrared spectrographs are usually cooled to minimise their contribution to the thermal background. The general principles of spectrographic design are outlined by Bowen ([B6], and see also [J1]). Figure 1.11 shows the schematic design of a typical optical spectrograph, consisting of a slit, placed in the focal plane of the telescope, collimator, diffraction grating, camera and detector.

The spectrograph slit isolates the light from a particular astronomical target and minimises the contribution from the sky background. The standard single-object, long-slit assembly is usually both tilted to the normal of the optical axis of the telescope and polished, so that the field can be viewed with an acquisition TV camera (Figure 1.11). The collimating lens has the same focal ratio as the telescope. Placing that lens at the appropriate place behind the slit transforms the diverging beam into a parallel beam, which is directed towards the diffraction grating

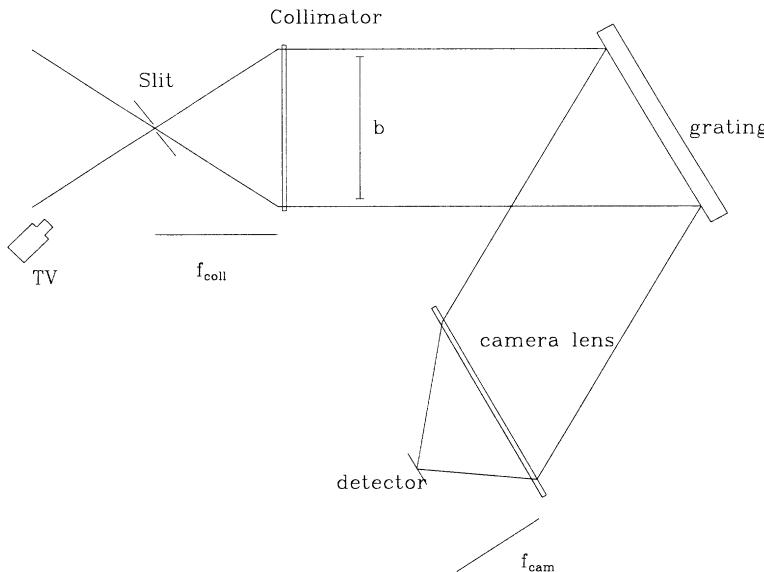


Figure 1.11. A schematic of a diffraction-grating spectrograph.

or prism. After dispersion, the beam is imaged onto a detector by the camera lens. Without the dispersing element, this is simply a re-imaging system, with the magnification given by the ratio between the focal length of the camera (f_{cam}) and the focal length of the collimator (f_{coll}). The inverse of this ratio is known as the scale factor. A length, δl , at the focal plane projects to a size δx at the detector, given by

$$\delta x = \frac{f_{cam}}{f_{coll}} \times \delta l \quad (1.35)$$

Generally, the magnification f_{cam}/f_{coll} is chosen so that a point source viewed under typical seeing conditions (see Section 1.8) projects to ~ 2 pixels at the detector. Thus, for the Hale 200-inch telescope, the plate-scale at Cassegrain focus ($f/15.7$) is $2.59 \text{ arcsec mm}^{-1}$. Given 0.8 arcsec (0.31 mm) as an estimate of the best likely seeing, and a CCD pixel size of $\sim 15 \mu\text{m}$, scale factors of at least 10 are required to avoid image degradation.

Prisms act as dispersing elements because the angle of refraction decreases as a function of wavelength – the same phenomenon responsible for differential atmospheric refraction. The dispersion formula is written as

$$\frac{d\theta}{d\lambda} = \frac{B}{b} \frac{dn}{d\lambda} \quad (1.36)$$

where b is the width of the beam (Figure 1.11), B is the base-width of the prism, and n is the refractive index of the prismatic material. The wavelength dependence of n

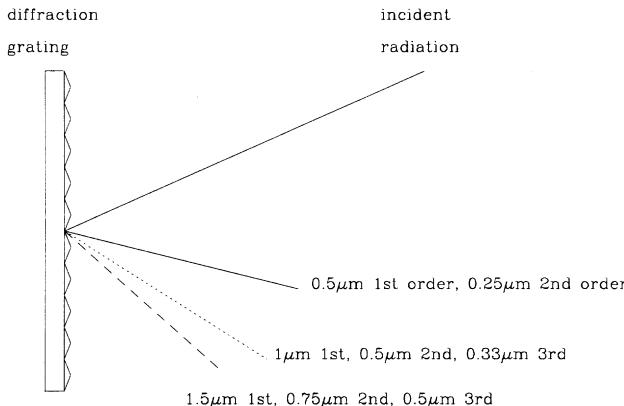


Figure 1.12. Spectral orders produced by a diffraction grating.

follows Cauchy's equation:

$$\frac{dn}{d\lambda} \propto \lambda^{-3} \quad (1.37)$$

Thus, prisms have the significant disadvantage that the dispersion, and hence the spectral resolution, decreases rapidly towards longer wavelengths. Moreover, a high dispersion requires that the width of the prism is large compared with that of the beam, which in turn demands a large, expensive and difficult to fabricate optical element. Several prisms can be combined to increase the overall dispersion, but at the potential expense of additional light loss at the air/glass surfaces. As a result of these constraints, diffraction gratings, rather than prisms, are preferred as the primary dispersive element in modern spectrographs.

A diffraction grating consists of a block of optical glass with a series of grooves ruled at regular intervals. The grating can be either reflecting or transmitting. The general properties were originally discovered in 1819 by Fraunhofer (hence Fraunhofer diffraction), who used a series of finely-spaced metal wires to form a grating. Technical difficulties limited their initial usefulness for astronomical purposes, but with the development of more precise ruling techniques it became possible to manufacture large, accurate reflection gratings.

Figure 1.12 illustrates the main properties of a diffraction grating. If we consider a parallel beam incident on a series of apertures, constructive interference between the individual diffraction patterns at each wavelength leads to a series of intensity maxima (orders) in the re-imaged beam. If the spacing between grooves is d , and the angle of incidence is i , then the grating equation

$$d(\sin(i) + \sin(\theta)) = m\lambda \quad (1.38)$$

gives the angle of diffraction, θ , for order m . In most spectrographs, the angle between the optical axes of the collimator and camera, δ_{CC} , is fixed. θ is measured in the same sense as i , so from Figure 1.11 it is clear that

$$\delta_{CC} = i - \theta \quad (1.39)$$

Thus, the central wavelength on the detector can be set by changing the tilt of the grating with respect to the optical axis of the collimator. The grating tilt is measured with respect to the normal to that optical axis, and is therefore equal to i in equations (1.38) and (1.39). As an example, consider the red camera on the Hale telescope double spectrograph [O2]. In this case, $\delta_{CC} = 35^\circ$, so if a 600 line/mm grating ($d = 1.67 \mu\text{m}$) is used, and the desired central wavelength is $\lambda = 6,000 \text{ \AA}$ in first order, then the required grating tilt is $\sim 28.25^\circ$. At that angle, second-order radiation of wavelength $3,000 \text{ \AA}$ (and third order at $2,000 \text{ \AA}$) is centred on the detector. Since the atmosphere is opaque to radiation shortward of $3,250 \text{ \AA}$ (Section 1.2), this is not a problem for this particular set-up. Indeed, a dichroic is usually employed in the double spectrograph to divert blue light to a separate camera. However, there are situations where a blocking filter is required to isolate the particular wavelength region of interest.

The angular dispersion for a given order can be derived by differentiating the grating equation to produce

$$\frac{\Delta\theta}{\Delta\lambda} = \frac{m}{d \cos \theta} \quad (1.40)$$

This can be expressed as the reciprocal linear dispersion, in \AA mm^{-1} at the detector, by inverting the above equation and dividing by the focal length of the camera:

$$\frac{\Delta\lambda}{\Delta l} = \frac{d \cos \theta}{mf_{cam}} \quad (1.41)$$

A diffraction grating re-distributes the incident light amongst several orders, with the intensity distribution dependent on the shape of the grooves. For a simple, square-wave grating, most of the energy resides in the undispersed, zero-order image. However, the shape of the grooves can be adjusted to concentrate up to 90% of the flux within a relatively narrow range of θ . Following the grating equation, for any given wavelength this places most of the intensity within a specific order. These gratings are usually described as being *blazed* for the appropriate wavelength in the first order spectrum.

The spectral resolution of a particular grating is given by

$$R = \frac{\lambda}{\Delta\lambda} = mN \quad (1.42)$$

where N is the total number of grooves within the optical beam. However, in practice the resolution is usually set by either the pixel size of the detector or by the projected width of the spectrograph slit. For example, the red camera on the original Hale double spectrograph ($f_{cam} = 152 \text{ mm}$) has a beam size of 146 mm , so the formal spectral resolution with a 600 line/mm grating is therefore $R = 87,600$. This corresponds to a resolution of 0.069 \AA at $6,000 \text{ \AA}$, or, in linear units at the detector, $0.6 \mu\text{m}$. In comparison, a 1-arcsec slit projects to a linear size of $25.6 \mu\text{m}$, and the original TI CCDs used had a pixel size of $15 \mu\text{m}$. Actual observations usually represent a trade-off between high signal-to-noise (as wide a slit-size as possible, to collect as much light as possible) and high resolution (as narrow a slit-size as feasible). A typical

compromise is to match the slit-width to two detector elements (Nyquist sampling); hence, observations with the Hale double spectrograph would be taken with a 1.2-arcsec slit, giving a resolution of $\sim 3.3 \text{ \AA}$ with the 600 line/mm grating at $6,000 \text{ \AA}$ ($R \sim 1,800$).

1.8 IMPROVING IMAGE QUALITY

The theoretical resolution of a telescope (the diffraction limit) is defined as

$$L = \frac{206,265}{10^6} \left(\frac{\lambda}{D} \right) \quad (1.43)$$

where λ is the wavelength in microns, D the telescope diameter in metres, and L the resolution in arcseconds. Astronomical seeing, however, introduces distortions, with the result that the best optical images obtained from ground-based telescopes seldom have profiles with full-width half-maximum less than ~ 0.6 arcsec at optical wavelengths.

'Seeing' is the consequence of the passage of starlight through a variable-density medium – the Earth's atmosphere. Operationally, atmospheric seeing can be regarded as including two components: image motion, due to the light passing through severely turbulent regions; and a general broadening of the stellar profile, originating as the wavefront passes through regions of different densities (atmospheric cells). The latter have diameters of approximately 15–20 cm, so a large telescope is combining light from many cells at any point during an observation. Significant distortions can be introduced by poor air-flow within the dome itself. The final observed stellar point-spread function (PSF) is defined by both these refractive effects and by scattered light within the optics of the telescope and camera. The former effects generally dominate in the core of the image, while the latter are more important at larger radii (such as stellar diffraction spikes, due to light diffracted by the secondary mirror supports). Innovative designs have been proposed for eliminating many of these sources of scattered light. For example, in telescopes with off-axis foci, the secondary mirror lies outside the primary beam and therefore introduces no obstruction, with a consequent reduction in scattered light [K4]. The majority of telescopes, however, are still constructed along more conventional lines.

The most extensive analysis of the light profile of a star was carried out by King [K1], who used photographic techniques to determine the surface brightness at radii of more than 10,000 times the half-width at half peak intensity (HW) of the image – a dynamic range of almost 28 magnitudes (Figure 1.13). More recent observations using electronic detectors in the optical (*B*-band – matching the photographic data) and near-infrared (*H*-band) have confirmed the general form of the profile and its invariance, once scaled to the appropriate HW, with wavelength [R1]. Racine has shown that, once diffraction spikes are eliminated, the observations within 10 HW are an excellent match to the profile predicted by Kolmogorov turbulence theory, suggesting that atmospheric effects dominate in the inner part of the profile. However, the extended aureole at larger radius is almost certainly due to

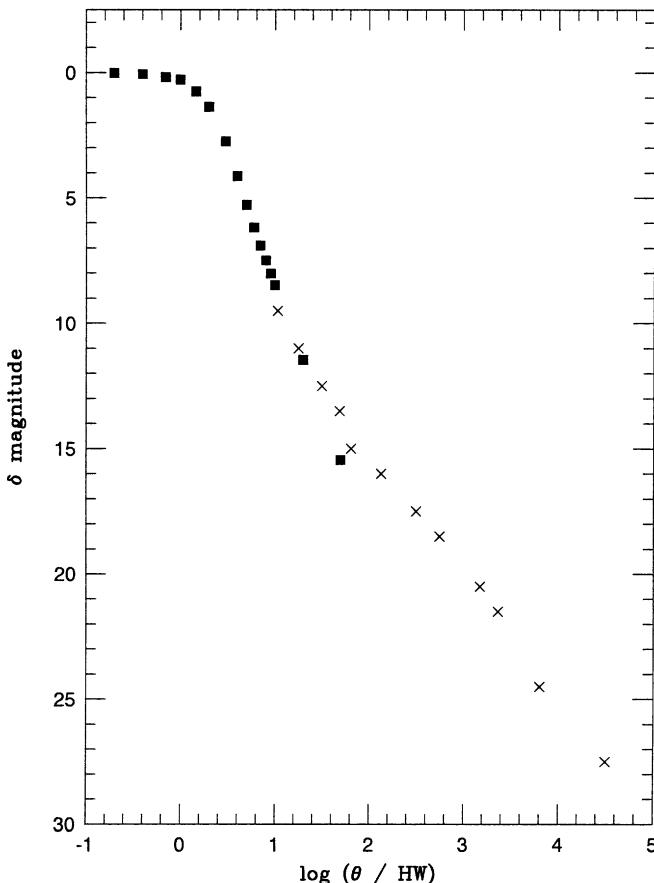


Figure 1.13. The stellar point-spread function, adapted from the measurements by King (crosses) and Racine (solid squares).

instrumental effects. In the case of King's analysis, Racine suggests that a significant component is due to scattered light from internal reflections in the 1-mm thick Schmidt plates used for the observations.

Improving seeing is currently a major priority at all large optical observatories, both through technological modifications (better temperature control, free air flow) and through the application of new techniques. Not only do such improvements permit higher spatial-resolution observations (resolving close binaries or structure in nearby galaxies), but concentrating the light from a point source also produces greater contrast with the sky background, and allows detection (or spectroscopy) at fainter magnitudes. Thus, a greater light-grasp can be achieved not only through the construction of new, larger mirrors, but also through improving the performance of existing telescopes. The present section provides an introduction to two techniques which have proven of considerable importance in studying low-mass dwarfs.

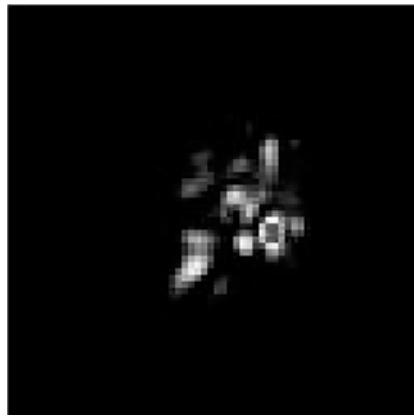


Figure 1.14. A speckle pattern.

1.8.1 Speckle interferometry

Labeyrie [L1] showed that when radiation from a star passes through the Earth's atmosphere, the wavefront passes through regions of different densities (atmosphere cells) which introduce phase fluctuations. Integrating these fluctuations over time leads to the broad seeing disk illustrated in Figure 1.13. If the exposure time is limited to extremely short timescales ($< 100\text{ msec}$) then some phase coherence remains. A single snapshot reveals a series of bright speckles which contain information on the image structure at a resolution matching the diffraction limit (Figure 1.14). Each individual observation includes only a few photons, and cannot provide sufficient signal to allow reconstruction of the image structure. However, many (hundreds) of such frames can be accumulated and combined, and can be analysed using Fourier techniques to reconstruct a diffraction-limited image.

Angular resolution decreases towards longer wavelengths – a circumstance that would seem to favour speckle observations at optical wavelengths. However, other factors favour observations in the near-infrared. First, the time-scale over which the atmosphere remains coherent, the correlation time, increases with increasing wavelength ($\propto \lambda^{6/5}$), permitting longer integration times, and a higher signal in each speckle frame in infrared observations. Second, the magnitude difference between any two main sequence stars is smallest at infrared wavelengths. Even in Fourier space, binary companion searches are limited by the stellar profile, which sets the background level. As a result, the magnitude limit for detecting a secondary star depends on both the brightness of the primary star and the angular separation between the components. Typical observations can reach flux ratios of 60–100 (4–5 magnitudes) at separations of more than twice the diffraction limit. Working at near-infrared wavelengths therefore minimises the magnitude difference and maximises the chances for detection of very low-mass companions.

Infrared speckle observations can achieve diffraction-limited resolution – often a factor of 10 improvement over direct imaging – and are capable of detecting

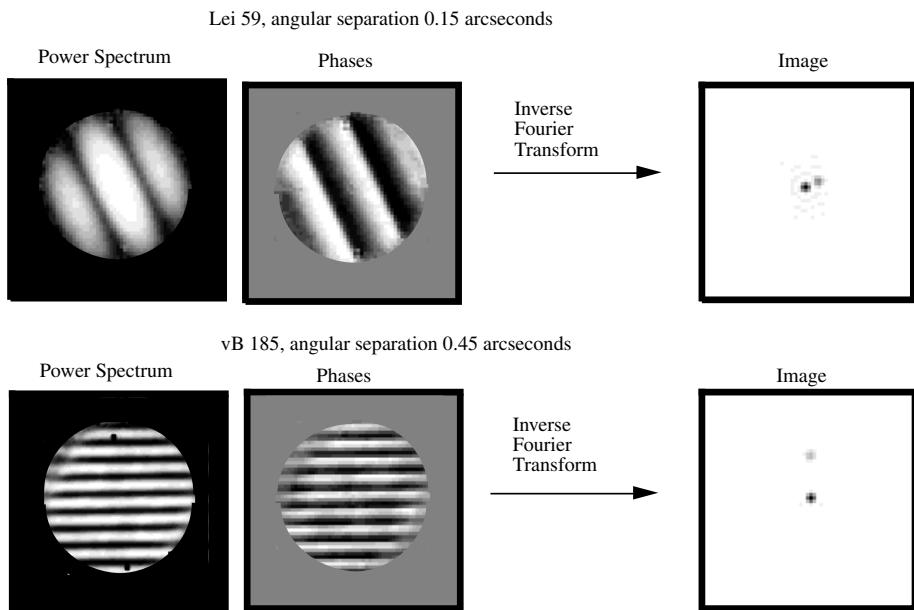


Figure 1.15. Speckle imaging analysis of two binaries in the Hyades. In each case, the left panels plot the Fourier amplitude and phase distribution, which are combined to give the diffraction-limited image on the right. Note that the orientation of the binary system is perpendicular to both amplitude and phase fringe patterns, while the spacing between the fringes increases, and the number of fringes decrease, as the separation decreases. (Illustration courtesy of J. Patience.)

companions 100 times fainter than the primary at separations of $\sim 0''.25$. This technique has played a crucial role in allowing measurement of the orbits of binary stars and (as described further in Chapter 9), the estimation of stellar masses. Figure 1.15 shows the results of Fourier analysis of speckle data for two representative binaries, two-dimensional ‘images’ of the derived phases and amplitudes. The cyclical fringe pattern is essentially an interferogram, with the period dependent on the separation of the two stars (wide separation, many fringes; narrow separation, few fringes) and the amplitude of the fringes dependent on the relative luminosity of the two sources. The orientation of the binary is perpendicular to the phase pattern. In principle, a binary of separation $\lambda/2D$ can be detected, since this produces one-half cycle of variation in the phase diagram – although this variation is detectable only for components of nearly-equal luminosity. This type of analysis has led to a more than threefold increase in the number of M dwarfs with masses determined to an accuracy of 20% or better.

1.8.2 Adaptive optics and interferometry

Speckle imaging ‘freezes’ the atmosphere in short exposures and reconstructs the image in Fourier space; adaptive optics techniques correct atmospheric wavefront

distortions above the image plane of the telescope. In the simplest case, a tip–tilt mirror takes out image motion introduced by seeing and telescope vibrations; in more complex instrumentation, multi-element ‘rubber’ mirrors are used to eliminate higher-order distortions (see [B1] for an excellent introductory review). The efficiency of these systems is measured by the Strehl ratio – the ratio between the peak flux in the core of a point source and the value expected for diffraction-limited imaging.

The advantage of AO techniques is that they provide sharpened images of individual objects which can be fed directly into CCD cameras or onto the slit of a spectrograph. As a result, they reveal complex structures, which would defy identification in Fourier space. AO systems are driven by feedback loops, using relatively bright stars (12th–14th magnitude for 4–8 metre class telescopes) as reference sources. The main limitation is that wavefront disturbances can be considered as uniform (and therefore correctable) over only a small angle, θ_0 , the isoplanatic angle. This angle depends on the coherence length of the atmosphere, r_0 , and inversely on the average distance of the atmospheric layer responsible for seeing distortions. The latter parameters are observatory dependent, but θ_0 is typically 2–3 arcsec at optical wavelengths, rising to 10–20 arcsec at 2.2–3.3 μm ($\theta_0 \propto r_0 \propto \lambda^{6/5}$). Since a bright guide-star within the isoplanatic patch is required, only a small fraction of the sky is accessible (\sim 1–2% at K-band, <0.01% at I) with natural guide-stars.

Adaptive optics systems using natural guide stars are now operational at several observatories, including Palomar [H3], Keck [V3], Gemini [G2], the MMT [W3] and the Very Large Telescope (VLT) of the European Southern Observatory (L10). The last system is particularly interesting, since it uses infrared light for wavefront correction, allowing access to red, optically-faint sources. Most systems produce Strehl ratios of 20–50% at wavelengths of 1–2.5 μm , and can achieve ratios exceeding 95% (effectively diffraction limited) at mid-infrared wavelengths (5–12 μm). The typical resolution at near-IR wavelengths is 0.1 arcseconds, and the AO systems are coupled with either imaging cameras or narrow-slit spectrographs, designed to take full advantage of the lower background offered by AO imaging. This instrumentation has already permitted the discovery and characterisation of numerous low-mass companions in close binary systems in both clusters (e.g., [C3]) and the field (e.g., [C4, F2, M2, M3, S3]).

In searches for low-luminosity companions, the potential primary often serves as its own guide star. AO technology will become more widely applicable as laser guide star systems are developed. A powerful (75 kW) pulsed laser excites sodium emission in the mesospheric layer at an altitude of 90 km. That bright patch acts as a guide star. Laser systems are currently under development at most of the new generation of 8–10-m telescopes.

Even the availability of laser guide stars, however, will not allow ground-based AO to match the performance of orbital telescopes at all wavelengths. The moderate Strehl ratios lead to higher effective sky brightness, while the point-spread function degrades significantly as one moves towards the edge of the isoplanatic patch. Multi-conjugate AO systems (e.g., [D1]) may overcome some of these problems eventually,

but for the next few years space-based systems, specifically the Hubble Space Telescope, will still provide the most sensitive high-resolution imaging at optical and near-infrared wavelengths.

Finally, interferometry has a long history of use in observations at radio wavelengths: data are combined from individual telescopes of modest aperture (5–10 m) separated by a much longer baseline (~ 100 m or more), giving a spatial resolution compatible with the latter. These techniques are being extended to near-infrared and optical wavelengths, both in ground-based observations (notably with the Keck and VLT telescopes) and from space.

1.9 SUMMARY

In the preceding sections, we have outlined some of the more important techniques in observational astronomy, particularly those relevant to the studies of low-mass dwarfs described in the following chapters:

- the definition of co-ordinate systems and the measurement of proper motion and parallax in stellar astrometry;
- an outline of telescope and detector systems;
- a summary of the main aspects of broadband stellar photometry;
- an introduction to stellar spectroscopy and spectrographs;
- and a brief review of image-sharpening techniques.

The aim has been to provide a broad overview rather than a rigorous development, emphasising aspects which are of most practical use to the astronomical observer. More extensive discussions of each of these topics are included amongst the literature cited in the appended bibliography.

1.10 REFERENCES

- B1 Beckers, J. M., 1993, *ARA&A*, **31**, 13.
B2 Bessell, M. S., 1979, *PASP*, **91**, 589.
B3 Bessell, M. S., 1990, *PASP*, **102**, 1181.
B4 Bessell, M. S., Brett, J. M., 1988, *PASP*, **100**, 1134.
B5 Boksenberg, A., 1972, in *Auxiliary Instrumentation for Large Telescopes: Proceedings of the ESO–CERN Conference*, eds. S. Lausten and A. Retz, p. 295.
B6 Bowen, I. S., 1962, in *Astronomical Techniques*, ed. W. A. Hiltner, University of Chicago Press.
B7 Buser, R., Kurucz, R. L., 1978, *A&A*, **70**, 555.
C1 Cousins, A. W. J., 1982, *SAAO Circulars*, No. 1.
C2 Culhane, J. L., Sanford, P. W., 1981, *X-ray Astronomy*, Charles Scribner's Sons, New York.
C3 Close, L. M., Widi, F., Lloyd-Hart, M. *et al.*, 2003, *ApJ*, **599**, 537.
C4 Close, L. M., Siegler, N., Freed, M., Biller, B., 2002, *ApJ*, **587**, 407.
C5 Copet, E. *et al.*, 1997, *Ap. Sp. Sci.*, **217**, 25.

- D1 Dekany, R. G. *et al.*, 2003, *Proc. S.P.I.E.*, **4839**, 1165.
- E1 Eccles, M. J., Sim, M. E., Tritton, K. P., 1983, *Low Light Level Detectors in Astronomy*, Cambridge University Press.
- E2 ESA, 1997, The Hipparcos catalogue, ESA SP-1200.
- E3 Epchtein, N. *et al.*, 1994, in *Science with Astronomical Near-Infrared Survey*, *Ap. Sp. Sci.*, **217**, 3.
- F1 Filippenko, A., 1982, *PASP*, **94**, 715.
- F2 Freed, M., Close, L. M., Siegler, N. 2003, *ApJ*, **584**, 253.
- F3 Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., Schneider, D. P., 1996, *AJ*, **111**, 1748.
- G1 Gunn, J. E., Carr, M., Rockosi, C., Sekiguchi, M. *et al.*, 1998, *AJ*, **116**, 3040.
- G2 Graves, J. E., Northcott, M. J., Roddier, F. J., Roddier, C. A., Close, L. M., 1998, *Proc. S.P.I.E.*, **3353**, 34.
- G3 Giacconi, R. *et al.*, 1979, *ApJ*, **230**, 540.
- H1 Hearnshaw, J. B., 1986, *The Analysis of Starlight*, Cambridge University Press.
- H2 Henden, A., Kaitchuck, R. H., 1982, *Astronomical Photometry*, Van Nostrand Reinhold, New York.
- H3 Hayward, T. L., Brandl, B., Pirger, B., Blacken, C., Gull, G. E., Schoenwald, J., Houck, J. R., 2001, *PASP*, **113**, 105.
- J1 Jenkins, F. A., White, H. E., 1976, *Fundamentals of Optics*, McGraw-Hill Inc.
- J2 Johnson, H. L., 1966, *ARA&A*, **4**, 193.
- K1 King, I. R., 1971, *PASP*, **83**, 199.
- K2 Kirchhoff, G., Part I: *Abhandl. d. Berliner. Akad.*, p. 63, 1861; p. 227, 1862. Part II: *ibid.*, p. 225, 1863.
- K3 Kovalevsky, J., 1998, *ARA&A*, **36**, 99.
- K4 Kuhn, J., Hawley, S. L., 1999, *PASP*, **111**, 601.
- L1 Labeyrie, A., 1970, *A&A*, **6**, 86.
- L2 Landolt, A. U., 1983, *AJ*, **88**, 439.
- L3 Landolt, A. U., 1992, *AJ*, **104**, 340.
- L4 Lane, A. P., 1979, *PASP*, **91**, 405.
- L5 Lang, K. R., 1992, *Astrophysical data: Planets and stars*, Springer-Verlag, New York.
- L6 Leinert, C., Richter, I., Planck, B., 1982, *A&A*, **110**, 111.
- L7 Leggett, S. K., Allard, F., Berriman, G., Dahn, C. C., Hauschildt, P. H., 1996, *ApJS*, **104**, 117.
- L8 Levizzani, V., Prodi, F., 1988, *J. Geophys. Res.*, **93**, 5277.
- L9 Longair, M. S., 1981, *High-Energy Astrophysics*, Cambridge University Press.
- L10 Lagrange, A.-M., Chauvin, G., Fusco, T. *et al.*, 2003, *Proc. S.P.I.E.*, **4841**, 860.
- L11 Lasker, B. M., 1995, *PASP*, **107**, 763.
- L12 Lasker, B. M., Greene, G. R., Lattanzi, M. J., McLean, B. J., Volpicelli, A. 1998, in *Astrophysics and Algorithms*.
- M1 Monet, D. G., 1988, *ARA&A*, **26**, 413.
- M2 Macintosh, B. A., Becklin, E. E., Kaisler, D., Konopacky, Q., Zuckerman, B., 2003, *ApJ*, **594**, 538.
- M3 McCaughrean, M. J., Close, L. M., Scholz, R.-D., Lenzen, R., Biller, B., Brandner, W., Hartung, M., Lodieu, N., 2004, *A&A*, **413**, 1029.
- M4 Minkowski, R., Abell, G. O. 1963, in *Stars and Stellar System*, Vol. 3., Basic Astronomical Data, ed. K. Aa. Strand (Univ. of Chicago Press, Chicago), p. 481.
- M5 Morgan, D. H. 1995, in *The future utilisation of Schmidt telescopes*, *ASP Conf. Ser.*, **84**, 137.

- N1 Neugebauer, G., Leighton, R. B., 1969, *The Two Micron Sky Survey*, NASA SP-3047.
- O1 Ohlsson, J., 1932, *Ann. Lund. Obs.*, **3**.
- O2 Oke, J. B., Gunn, J. E., 1982, *PASP*, **94**, 586.
- O3 Osterbrock, D. E., 1997, *J. Hist. Astr.*, **25**, 1.
- P1 Persson, S. E., Murphy, D. C., Krzeminski, W., Roth, M., Rieke, M. J., 1998, *AJ*, **116**, 2475.
- P2 Pogson, N., 1856, *MNRAS*, **17**, 12.
- P3 Price, S. D., Walker, R. G., 1976, *The AFGL Four Color Infrared Sky Survey*, AFGL-CR-83-0161, Air Force Systems Command, USAF.
- R1 Racine, R., 1996, *PASP*, **108**, 699.
- R2 Roach, F. E., Gordon, J. L., 1973, *The Light of the Night Sky*, Geophysics and Astrophysics Monographs, V. 4.
- R3 Reid, I. N. *et al.*, 1991, *PASP*, **103**, 661.
- S1 Sekiguchi, M., Iwashita, H., Doi, M., Kashikawa, N., Okamura, S., 1992, *PASP*, **104**, 744.
- S2 Smart, W. M., 1971, *Spherical Astronomy*, 5th edition, Cambridge University Press.
- S3 Siegler, N., Close, L. M., Mamajek, E. E., Freed, M., 2003, *ApJ*, **598**, 1265.
- S4 Skrutskie, M. F. *et al.*, 1997, in *The Impact of Large-Scale Near-IR Sky Surveys*, ed. F. Garzon *et al.*, Kluwer, The Netherlands, p. 25.
- S5 Smith, J. A. *et al.*, 2002, *AJ*, **123**, 2121.
- T1 Tucker, W., Giacconi, R., 1985, *The X-Ray Universe*, Harvard University Press, Cambridge, Massachusetts.
- T2 Taylor, K. T., McLean, I. S., 2002, *Proc. S.P.I.E.*, **4841**, 55.
- V1 van Leeuwen, F., 1997, *Sp. Sci. Rev.*, **81**, 201.
- V2 Vasilevskis, S., 1967, *AJ*, **72**, 583.
- V3 van Dam, V. A., Macintosh, B. A., 2003, *Proc. S.P.I.E.*, **5169**, 1.
- V4 Voges, W. *et al.*, 1999, *A&A*, **349**, 389.
- W1 Whittet, D. C. B., 1992, *Dust in the Galactic Environment*, the graduate series in astronomy, Institute of Physics Publishing, Bristol, Philadelphia, New York.
- W2 Wilson, R., 1996, *Reflecting Telescope Optics I: Basic Design Theory and its Historical Development*, Springer-Verlag, Berlin.
- W3 Widi, F. P., Brusa, G., Lloyd-Hart, M., Close, L. M., Riccardi, A., 2003, *Proc. S.P.I.E.*, **5169**, 17.
- W4 West, A. *et al.*, 2004, *AJ*, **128**, 426.

1.11 HOMEWORK PROBLEMS

Problem 1

Arcturus, the brightest star in the constellation Bootes, has co-ordinates ($\text{RA} = 14\text{ h }15\text{ m }39.7\text{ s}$, $\text{Dec} = +19^\circ 10' 57''$; equinox 2,000). Calculate the expected rise and set times (LST and local time) for this star as viewed from Sydney, Australia, on 15 June 2000. Suppose that the effective western horizon has a zenith distance of 88° – estimate the actual time that Arcturus disappears.

Problem 2

Arcturus is one of the nearest stars, with a measured trigonometric parallax of 88.9 milliarcseconds. The proper motions are $-0.0776 \text{ s year}^{-1}$ in RA and $-1.999''/\text{year}^{-1}$ in Dec, while the radial velocity is -5.2 km sec^{-1} . Calculate the position of Arcturus in Galactic co-ordinates and compute its space motion relative to the Local Standard of Rest.

Problem 3

A spectrum taken of an M star with $V = 15$ using a 4-m class telescope and a single-slit spectrograph has $S/N = 20$ per pixel at a continuum wavelength near $H\alpha$ in a one-minute exposure. The read noise from the detector is 6 electrons and the sky contributes $0.01 \text{ photons sec}^{-1} \text{ pixel}^{-1}$ at this wavelength.

- (a) What supporting observations must be made to properly reduce and calibrate the spectrum?
- (b) What is the total number of star counts observed in the spectrum at this wavelength? Is the spectrum dominated by noise from the sky, readnoise, or Poisson statistics?
- (c) What should the exposure time be in order to reach $S/N = 40$?
- (d) Suppose an O star with $V = 15$ was observed to the same $S/N (= 20)$. What is the exposure time?
- (e) The original spectrum used a grating that gave resolution = 1,000. If a higher resolution grating, giving $R = 5,000$ was used, what is the S/N for the same one minute exposure?

Problem 4

Show that κ , the constant used in the conversion from tangential velocity to proper motion (equation 1.9), is equal to 4.74. What are the units of κ ?

Problem 5

Derive equation (1.7).

2

Observational properties of low-mass dwarfs

2.1 INTRODUCTION

'It is a capital mistake to theorise before one has data.'

Arthur Conan Doyle

Astronomy is an observational science. In disciplines such as chemistry, biology and physics, carefully controlled laboratory experiments can be set up to test the effect of altering one specific variable, such as temperature, pressure or concentration. However, the physical conditions that pervade in most celestial objects are so extreme that there can be no possibility of their being matched in terrestrial laboratories, so direct experimentation is out of the question. Consequently, physical understanding of the conditions and processes which govern the formation and evolution of stars rests on our ability to interpret information carried by electromagnetic radiation. Observations of position as a function of time yield estimates of distance and of motion perpendicular to the line of sight; summing the energy output over all wavelengths gives the total (bolometric) luminosity; surface temperature can be estimated from the distribution of energy as a function of wavelength, and from the relative ionisation of different chemical species; individual spectral line profiles of various atoms allow us to estimate the density, gravity and gas pressure in the atmosphere; chemical composition influences the relative strengths of lines and bands in the spectrum; and the exact position of spectral lines tells us the radial velocity of the star.

We cannot vary the conditions prevalent on an individual star (as one astronomer puts it, ‘we don’t have a long enough stick to poke them with’), so observations of many individual stars are combined to build up a picture of how stars of different mass and composition evolve. This is a particularly powerful technique when the stars in question are drawn from the same star cluster and are therefore effectively coeval, but can also be illuminating when applied to stars in the immediate vicinity of

the Sun (the Solar Neighbourhood). While the latter stars span a wide range of age and composition, the fact that distances can be measured with high precision makes this group a uniquely useful probe of intrinsic properties. Whether looking at individual stars or at ensembles, observations provide the yardstick for measuring the success of analytic and numerical models devised to explain the physical processes present.

With these issues in mind, this chapter reviews the observational characteristics that define M dwarfs and the recently-identified ultracool L and T dwarfs. An empirical and phenomenological approach is adopted, concentrating on qualitative inferences that can be drawn directly from observation. This lays the foundation for more detailed consideration of the underlying physics in succeeding chapters.

The next two sections provide an overview of the spectroscopic and photometric characteristics of low-mass dwarfs. It will become clear that there are substantial similarities between the latest K dwarfs and ‘classical’ M dwarfs. In particular, most of the molecular features that identify M dwarfs – TiO, H₂O, the metal-hydride bands – start to appear at spectral types earlier than M0. For that reason, in this book the term ‘M dwarf’ includes all stars with absolute visual magnitudes fainter than $M_V = 7.5$ – a definition which encompasses many stars formally classified as K5 or K7 dwarfs. Very low-mass (VLM) or ultracool dwarfs are defined as having spectral types later than M6.5, therefore including members of the new spectral classes L and T.

2.2 SPECTRAL CLASSIFICATION: WHAT ARE M, L AND T DWARFS?

2.2.1 Early observations

The advent of optical spectroscopy revolutionised astronomy in the mid-nineteenth century. For the first time, astronomers were placed in a position where they could probe the nature of stellar material. This ability was enhanced greatly when photography was used to record stellar spectra for relatively leisurely inspection, rather than relying on fleeting impressions made at dead of night under strained physiological conditions. Hearnshaw [H3] ably describes the early development of stellar spectroscopy and its transformation into a semi-exact science. From the outset it was clear that there is order in the way that stars behave: patterns amongst the various spectral features allowed certain stars to be grouped together. In 1864, Secchi, in the first serious attempt at spectral classification, placed stars under three headings: types I, II and III. Type III objects were ‘coloured’ stars with wide absorption bands in their spectrum, but it was not until 1904 that Fowler [F3] identified these spectral features as due to either the metal titanium or a compound containing titanium. Nonetheless, titanium oxide absorption bands – the defining feature of M stars – were clearly distinguished at the outset.

All of the early classification systems ordered stars from blue to red, reflecting the suspicion first voiced in 1874 by Vogel that this represented a scale of decreasing temperature, and perhaps an evolutionary sequence. A relatively small number of

researchers – including Lockyer and (at least initially) Hale – argued that changes in the relative strengths of different absorption features (line-blanketing) was the primary cause of both spectral and colour variations. It was not until 1909 that Wilsing and Scheiner [W2] demonstrated conclusively that O-type and M-type stars did indeed lie at opposite extremes in the stellar temperature scale (see [H3] for a more extensive review).

The origination of the designation ‘type M’ lies with the scheme – devised at Harvard, by Mrs W. P. Fleming – for spectral classification outlined in the *Draper Memorial Catalogue* of 1890 (with the ‘M’ stemming from the fact that Secchi’s class I and II stars were divided into eleven subclasses, A to K, omitting J). This catalogue included 10,351 of the brightest northern stars – a large, but not overwhelming number – and the Harvard system had some competitors, notably Lockyer and Vogel. However, as Hearnshaw explains, Pickering’s championing of his local system, coupled with the irresistible force of Cannon’s classification of the 225,300 stars in the Henry Draper catalogue (1918–1924), led to the Harvard spectral types becoming the *de facto* astronomical reference system by the 1920s.

Early spectral catalogues, based on visual or photographic data, were restricted to stars with bright apparent magnitudes. As a result, almost all M stars in these catalogues are giants rather than dwarfs. This became apparent only after Hertzsprung [H5] and Russell [R5] independently arrived at versions of the diagram that bears their names. Plotting luminosities against a temperature indicator (colour indices in Hertzsprung’s case [H6], and spectral types by Russell [R6]) produced the wishbone-shaped diagram with ‘giants’ and ‘dwarfs’ (named by Russell) and a ‘main sequence’ (defined by Hertzsprung from his observations of Hyades and Pleiades cluster members). Figure 2.1 shows the modern incarnation of this diagram, based on stars with parallaxes measured by the ESA *Hipparcos* satellite. The initial H–R diagrams included few late-type (cool, low-mass) dwarfs, but nevertheless demonstrated that luminosity decreased with later spectral types along the main sequence.

The Harvard system provided a ready means of comparing the general properties of many stars, but only in an approximate fashion. Its utility was limited by its dependence on internal classification criteria: that is, the relative strengths of lines and bands in the programme star ($H\beta$ against CN, for example) were measured and compared against a reference list of line ratios, and the spectral type assigned based on that comparison, rather than by matching against a set of standard stars. Moreover, the final calibration was rather coarse, with each spectral class spanning a wide range of physical properties, particularly with spectral type M. At Mount Wilson, Adams attempted to address the latter problem by adding a numerical qualifier, based on the absolute magnitude estimated from the relative strength of individual spectral lines. However, this approach had the unfortunate consequence that a change in the absolute magnitude calibration changed the entire spectral type scale.

Both problems were addressed by Morgan and Keenan in their definition of the MK spectral classification system (see [O2] for a thorough discussion of the genesis of this system). The main achievement of their system was its empirical definition of

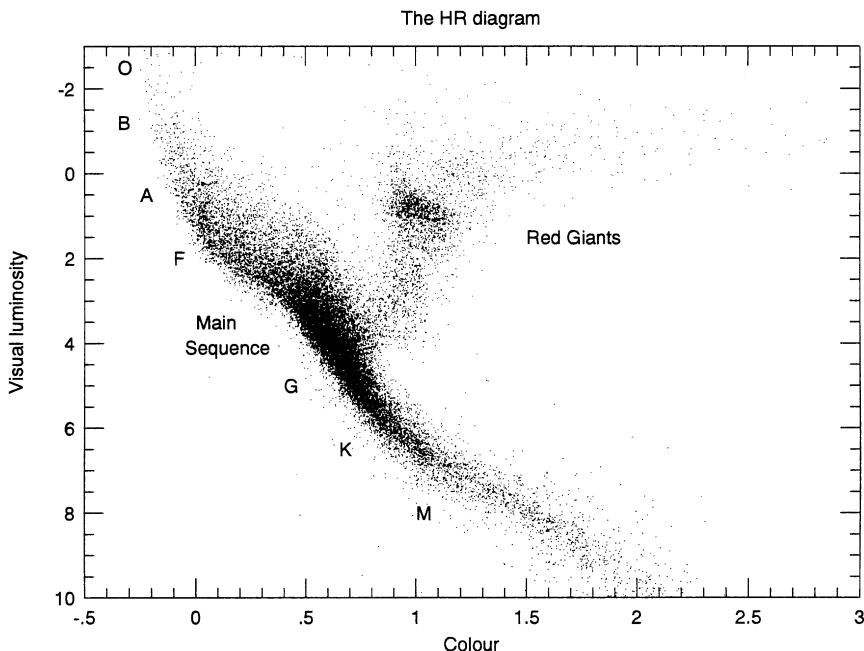


Figure 2.1. The Hertzsprung–Russell diagram, as defined by stars with parallaxes measured to an accuracy of better than 10% by the ESA *Hipparcos* satellite [E1].

spectral types with respect to a set of particular standard stars. Spectra of programme stars are classified by comparing the relative strength of specified features within a particular wavelength range (initially 3,930–4,860 Å) against observations, taken to the same signal-to-noise and at the same dispersion, of the standard stars. Moreover, the MK system added a formalised definition of luminosity class, with types I, III and V replacing the intermittently-used ‘c’ (narrow-lined), ‘g’ (giant) and ‘d’ (dwarf) qualifiers. The initial grid of standards was laid out in the MKK atlas of stellar spectra [M8]. The overall success of the system can be judged by its continued use (albeit in modified form) in current astronomical research. Many of the subsequent modifications and improvements have centred on the treatment of M dwarf stars.

2.2.2 An M dwarf classification system

Spectral classification is based purely on morphology – regularities in the appearance and disappearance of particular spectral features. The expectation is that a well-designed system, ranked by spectral variation, is also based on physical properties. Spectral type therefore provides a shorthand method of describing the overall physical characteristics of a given star, although it should be emphasised that the physics (why stars show spectral variations) post-dates, and is independent of, the definition of the classification system.

The traditional method of defining a spectral classification system is to take observations of an ensemble of stars – preferably of known absolute magnitude and luminosity class – and identify several key spectral features, prominent enough to be recognised easily, that show smooth behavioural trends when the individual spectra are suitably arranged. Thus, the hydrogen Balmer-series lines increase in strength from type O through B to type A, subsequently diminishing in strength through types F, G and K. Earlier (OB) and later (FG) stars can be distinguished by considering other features, such as the presence and strength of He I lines, the G-band (due to the CH molecule) or the H and K lines due to ionised Ca (Ca II).

The TiO bands that dominate optical spectra of M stars are an obvious choice as a primary indicator of spectral type. The next operational step in defining a workable classification system is to set up a grid of standards. Early observations were limited to stars with relatively bright apparent magnitudes, and, as a result, included few late-type M dwarfs. Moreover, most photographic observations before the 1940s were confined to the blue–green region of the spectrum, where spectral features saturate rapidly with decreasing temperature. As a result, the first MK system is of limited utility in M dwarf classification.

Recognising these limitations, the MK system was originally defined only for main sequence stars earlier than type M2. With the development of larger telescopes and more efficient spectrographs and detectors, observations of later-type dwarfs accumulated, and it became necessary to extend the classification to these cooler dwarfs. However, with no generally accepted guidelines, several mutually incompatible systems arose. The two most widely used were the Yerkes system, developed by Morgan [M7] and Kuiper [K10], and Joy’s Mount Wilson system, which took TiO band-strength in the blue spectral régime as the main indicator of spectral type. Based on different criteria at different wavelengths, these systems diverged in their classification of the later-type M dwarfs. The star Wolf 359 was classed as type M8 by Morgan, but as type dM6e by Joy. Thus it was necessary to determine what system had been used to classify a star before observations of different M dwarfs could be compared.

Part of the problem in classifying M dwarfs using visual or blue spectra is that those wavelengths lie far from the peak of the energy distribution, making it more difficult to obtain high signal-to-noise observations. While Keenan and MacNeil [K1] added to the wavelength coverage of the revised MK system, Boeshaar [B6] first included features as red as 6,800 Å. Kirkpatrick and collaborators [K2] further extended the system to both longer wavelengths and later-type stars.¹ Their calibration – designated the KHM system – is based on both the relative strengths

¹ Those observations concentrate on the nearest stars, most of which are identified either by their number in Willem Luyten’s catalogue of stars with annual proper motions exceeding 0.5 arcsec (the Luyten Half Second Catalogue, designated LHS, as in LHS 2924), or by their number from the nearby star catalogues compiled by Wilhelm Gliese and Hartmut Jahreiss (designation Gl or GJ; for example, Wolf 359 is Gl 406, the 406th star in Gliese’s 1969 catalogue [G5].

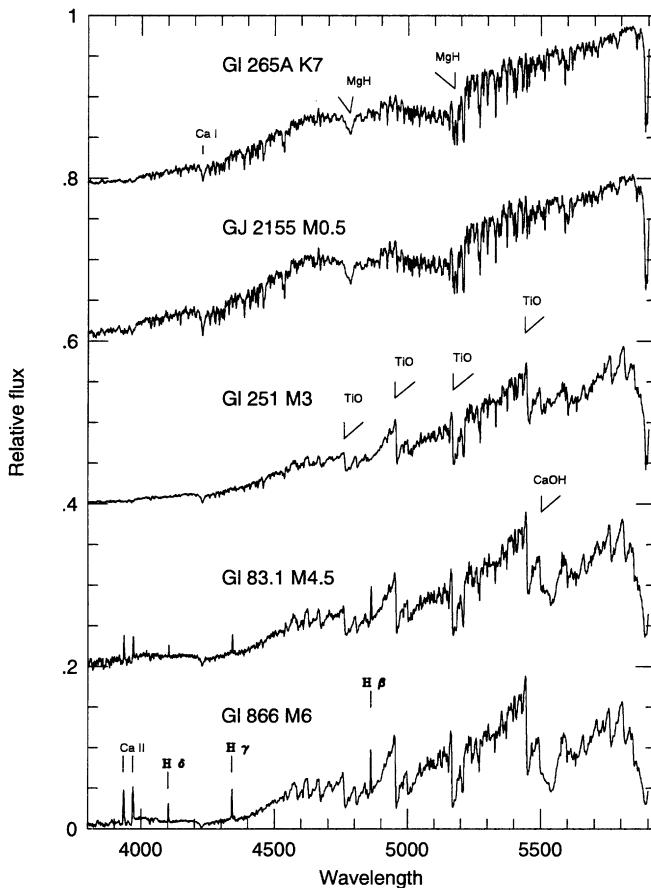


Figure 2.2. Blue–green optical spectra of M dwarfs, illustrating the main features used to calibrate spectral type. Prominent molecular bands and atomic features have been identified.

of spectral features in the range 6,300–9,000 Å and on the overall spectral slope across these wavelengths. Classification is through least-squares matching of flux-calibrated spectra against observations, made at the same resolution, of an extensive grid of standards. All spectral features contribute to some extent to the final type determination. However, most of the weight in the calibration rests with the stronger molecular bands (titanium oxide, vanadium oxide and calcium hydride), and the strength of individual features can be used to define a scale that approximately matches the KHM system. For example, measurements of the depth of the 7,050 Å TiO bandhead provide spectral types for dwarfs earlier than type M6 (where the bandhead saturates), with the observations calibrated against data for KHM standard stars [R2].

Bessell [B5] has defined an alternative spectral-type calibration for M dwarfs. His system is related to Wing’s [W3] giant-star spectral types, based on TiO band-

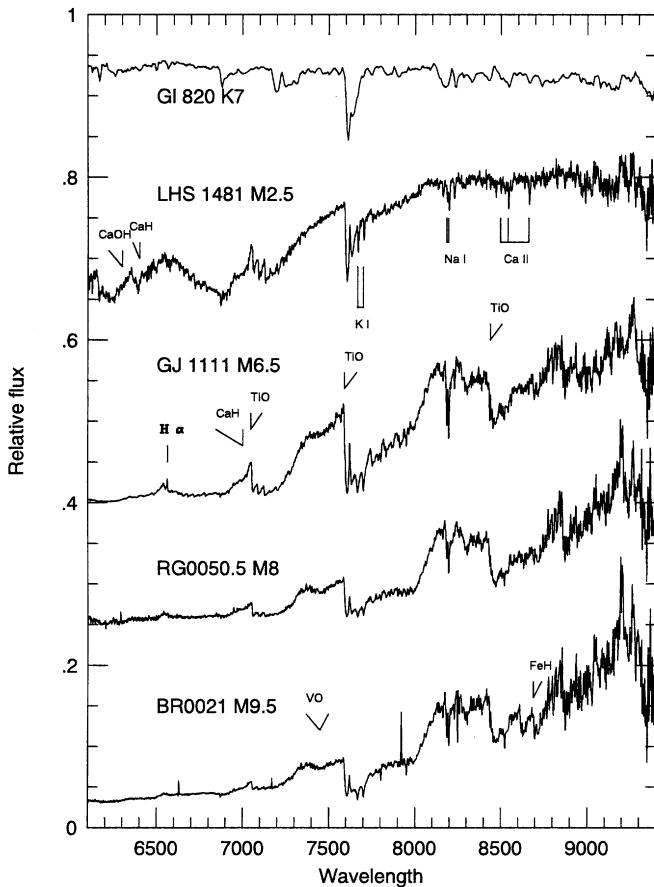


Figure 2.3. Red optical spectra of M dwarfs, illustrating the main features used to calibrate spectral types.

strength for earlier-type M dwarfs and on VO for the later-type stars where the TiO bands saturate. There is reasonable agreement (~ 0.5 spectral classes) with the KHM system in the former case, but Bessell assigns systematically earlier types to the later VLM dwarfs: for example, Gl 752 B (VB 10) is type M8 on the KHM system, but type M7 on Bessell's system. The more widely used KHM system is adopted in this book.

Figures 2.2 and 2.3 present optical spectra of representative K and M dwarfs, covering the blue-green and far-red wavelength regions, and identifying the principal spectral features. There are obvious correlations in the strength of individual lines and bands with changing spectral type. TiO is present, but weak, at type K7, and grows in strength until type M6, where most of the bandheads saturate. Bands due to several metal hydrides – MgH, FeH and, particularly, CaH – also first become detectable among K7 stars and grow in prominence with later spectral type. At

approximately the same temperature as the redder TiO bands saturate, VO becomes evident at 7,330–7,530 Å. FeH bandheads also appear in the latest spectral types at 7,786, 8,692 and 9,020 Å.

The strongest atomic lines at these wavelengths are Ca I at 4,227 Å, the Na I doublets at 5,890/5,896 Å (the D lines, at the long wavelength limit of the blue–green spectra plotted in Figure 2.2) and 8,183/8,195 Å, the K I doublet at 7,665/7,699 Å and, to a lesser extent, the Ca II ‘infrared’ triplet at 8,498, 8,542 and 8,662 Å. The last mentioned decline in strength towards later spectral types, besides being more difficult to detect as the 8,432 Å TiO bandhead increases in strength. The other atomic features show a general increase in strength, although Ca I 4,227 Å is swamped by TiO absorption at later spectral types. The KI doublet narrows in equivalent width between \sim M7 and M9.5, before broadening dramatically at later types. Detailed M dwarf spectroscopic atlases are provided by Kirkpatrick *et al.* [K2], [K3] and, for the latest M spectral types, Tinney and Reid [T1].

2.2.3 Beyond M: spectral type L...

M dwarfs were long regarded as defining the lowest extremities of the hydrogen-burning main sequence – the *Ultima Thule* of the H–R diagram. However, the marked improvement in sensitivity of wide-angle photometric surveys, particularly at near-infrared wavelengths, has resulted in the detection of increasingly fainter and cooler low-mass dwarfs. The most extreme of those objects have spectral characteristics which cannot be accommodated in class M, requiring the definition of a new spectral class – the first for almost half a century – class L.

The first L dwarf, discovered in 1988, was GD 165B, an extremely red, low-luminosity companion of a DA white dwarf [B2]. Initial spectroscopic observations [K4] suggested puzzling dissimilarities between this dwarf and well studied late-type M dwarfs, such as VB 10, LHS 2924 and LHS 2065. However, with only a 2.3-m telescope, the signal-to-noise in the available spectrum was low, while the much brighter white dwarf companion limited coverage to longward of \sim 7,500 Å. Finally, the (remote) possibility was raised of significant atmospheric contamination by metal-rich ejecta during the asymptotic giant branch or planetary nebula evolutionary phase of the degenerate companion. Thus, GD 165B was, by-and-large, considered a unique oddity.

Nonetheless, the quest for lower-temperature, lower-luminosity dwarfs continued, spearheaded by Kirkpatrick, Henry and collaborators, who compiled samples of ‘ultracool’ ($>$ M7) dwarfs from proper motion catalogues [K5]. As described in more detail in Chapter 6, the breakthrough came in 1997. First, follow-up spectroscopy of candidate late-type dwarfs in the general field, selected using near-infrared photometry obtained with the prototype camera for the 2-Micron All-Sky Survey (2MASS, [S6]), resulted in the identification of one extreme ultracool dwarf, 2MASSP J0345432+254023, initially classed as type $>$ M10 [K6]. Shortly thereafter, Ruiz *et al.* [R4] discovered a faint proper motion star, named Kelu 1 (Kelu is Mapuche for red), with an unusual optical spectrum, reminiscent of GD 165B.

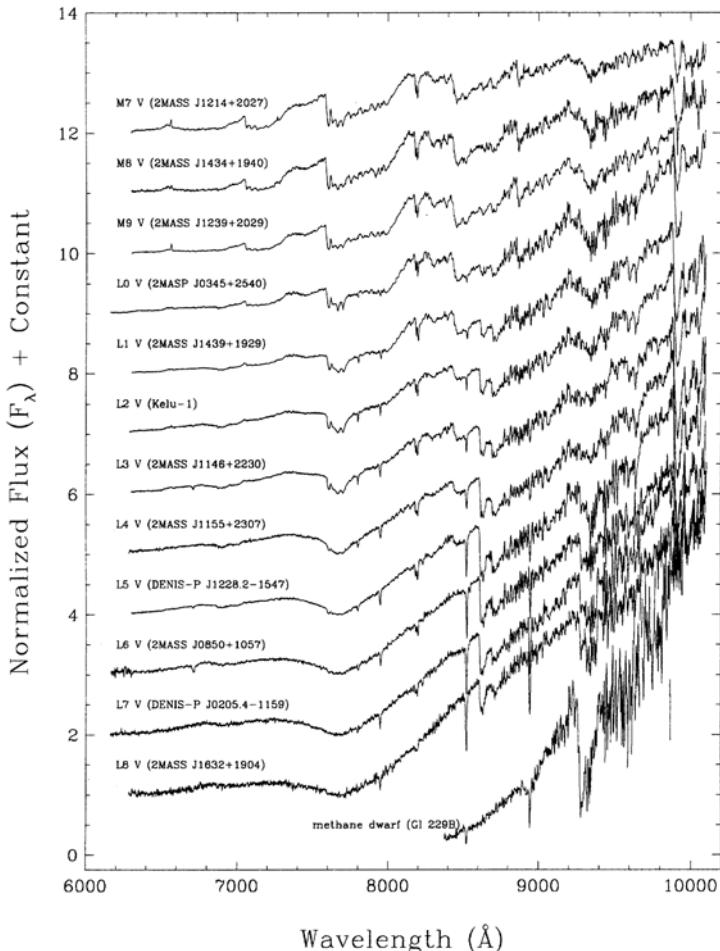


Figure 2.4. The L dwarf spectral sequence. (From [K8], courtesy of the *Astrophysical Journal*.)

The trickle turned into a flood with the first results from the wide-field near-infrared sky surveys, DENIS and 2MASS. Spectroscopy of ultracool candidates from the DENIS brown dwarf mini-survey [D1] revealed three isolated dwarfs with spectra similar to GD 165B. Similarly, follow-up observations of colour-selected candidates from the 2MASS survey resulted in the identification of more than 85 such objects between August 1997 and July 1999 [K8]. Subsequent studies, based mainly on data from either the Sloan Digital Sky Survey (SDSS – F4, S7) and 2MASS (C3, C4), have more than tripled the current sample. Many of these follow-up observations have the advantage of the unparalleled light grasp of the Low Resolution Imaging Spectrograph [O1] on the Keck 10-m telescope, but even so, many L dwarfs are too faint to observe at $\lambda < 7,000 \text{ \AA}$. With more than 300 L dwarfs now known, GD 165B has been transformed from an oddity to a prototype.

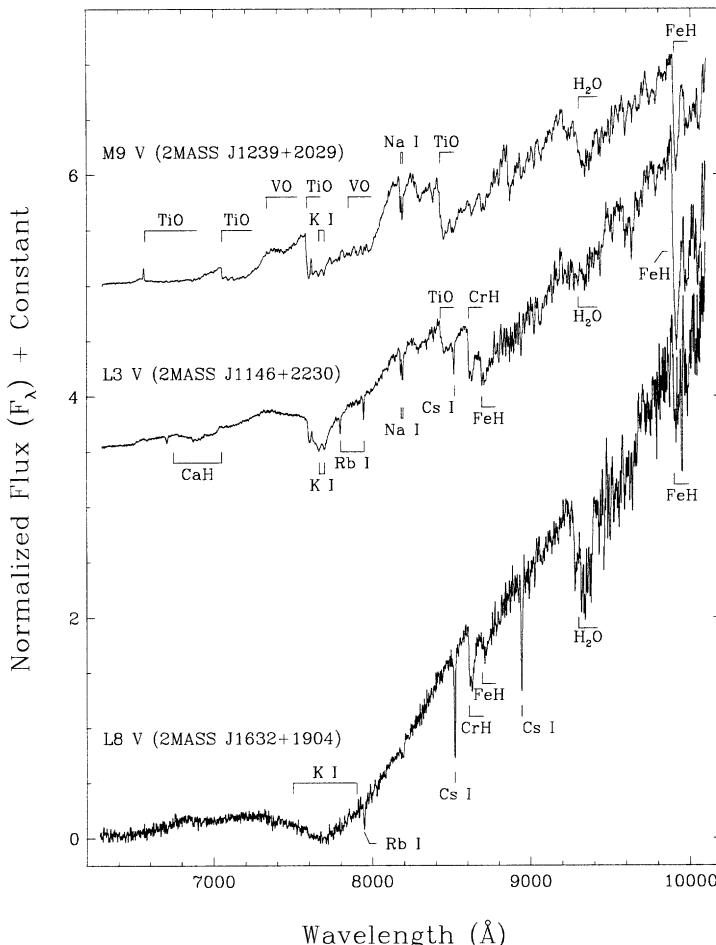


Figure 2.5. Key features defining spectral class L. (From [K8], courtesy of the *Astrophysical Journal*.)

The 2MASS dataset is sufficiently rich that it provides a well-sampled sequence for the new spectral class (Figure 2.4). As with the KHM M dwarf sequence, the primary classification is based on spectral behaviour over the wavelength range 6,000–10,000 Å. Specific features are identified in Figure 2.5. The main characteristic separating type L from type M is the diminishing strength of TiO, the primary criterion of spectral class M. VO remains prominent among the earliest L dwarfs, but is barely detectable in Kelu 1 (spectral type L2). The dominant molecular features are metal hydrides: CaH, FeH and CrH at far-red wavelengths, and MgH shortward of 6,000 Å. The Na I doublet at 8,183/8,192 Å weakens with progressively later spectral type, but the resonance lines due to K I, Cs I, Rb I and the sodium D lines all become increasingly stronger. The behaviour of the K I 7,665/7,699 Å

doublet is particularly interesting, broadening in a pronounced fashion at spectral type L4. The two components are essentially indistinguishable, with a combined equivalent width of more than 100 Å. The sodium D lines show similar behaviour in the few L dwarfs bright enough for observation at those wavelengths.

The physics underlying these variations is discussed in detail in Chapters 4 and 6. In qualitative terms, dust formation is believed to remove TiO and VO from the atmosphere, reducing the opacity. As a result the photosphere (the stellar ‘surface’) lies at relatively high depth, where pressure broadening leads to line widths of white dwarf proportions amongst the more abundant species.

Several of the spectra plotted in Figure 2.4 exhibit an absorption line at 6,708 Å (for example, 2MASS J1146+2230 and 2MASS J0850+1057). This feature is due to lithium and, as explained in Chapter 3, identifies these objects as having masses below $0.06 M_{\odot}$ – well within the brown dwarf régime. However, the theoretical models described in Chapter 3 suggest that stars with masses below $0.08\text{--}0.085 M_{\odot}$ can reach temperatures lower than the likely M dwarf/L dwarf boundary (\sim 2,000 K). Thus, early-type L dwarfs are a mixture of very low-mass hydrogen-burning stars and degenerate brown dwarfs.

2.2.4 ... and spectral type T

The discoveries made in the mid-1990s managed to exhaust even the L dwarf classification. As will be discussed further in Chapter 6, a new species of brown dwarfs was uncovered with the identification of Gl 229B, a very low luminosity companion to a M0.5 dwarf lying only 5.5 parsecs from the Sun [N1]. That object is characterised by strong methane absorption at near-infrared wavelengths, similar to spectra of the giant planets in the Solar System. This radical change in properties separates Gl 229B from L dwarfs, and marks it as the prototype of spectral class T.

All T dwarfs are substellar-mass brown dwarfs – the effective temperatures are too low to accommodate any central hydrogen fusion. These objects have extremely low luminosities ($<10^{-5} L_{\odot}$), and, as a result, the discovery of other examples has proceeded at a much slower rate than for the brighter L dwarfs. Nonetheless, the SDSS [S8, L10] and 2MASS [B8, B9] surveys have combined to catalogue more than 60 confirmed T dwarfs by late 2004 (see Section 6.6.2).

The T dwarf classification system rests mainly on the spectral morphology at near-infrared wavelengths (1–2.5 μm), where the methane bands dominate (see Section 2.2.6). At far-red optical wavelengths, the spectra continue the trends evident in the latest-type L dwarfs (Figure 2.6; see also [B10]). The K I 7,665/7,699 doublet continues to grow in strength, coupling with the Na I D lines to absorb most of the flux shortward of 8,000 Å. Chromium hydride absorption is weak, but may still be present in early-type T dwarfs, while rubidium and caesium atomic lines weaken with increasing type, but persist to \sim T7. Water absorption at 9,250/9,450 Å increases in strength in later-type T dwarfs, and the overall shape of the spectrum steepens toward the red from T2 to T8.

Two final issues deserve mention. First, why spectral types L and T? The rationale is given by Kirkpatrick [K7]: after eliminating possible confusion with

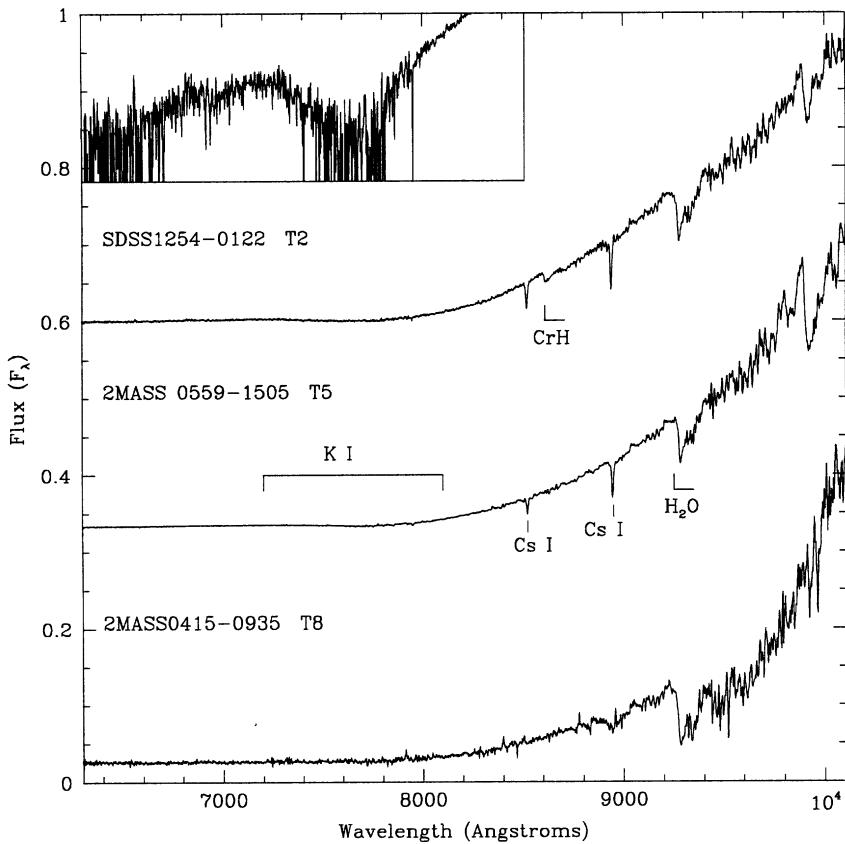


Figure 2.6. Far-red optical spectra of T dwarfs; the inset plots the R/I-band spectrum of SDSS1254 on a logarithmic flux scale, to better illustrate the broad Na I and K I absorption. (From [B10], courtesy of the *Astrophysical Journal*.)

white dwarfs, elliptical galaxies, supernovae and so on, only five letters – H, L, T, Y and Z – remain as unambiguous spectral type designations. Of those five, L lies closest to M, and T is the next in sequence; Y has already been reserved for even cooler brown dwarfs [K14]. The second issue concerns devising an appropriate mnemonic for the new spectral classification scheme, OBAFGKMLT(Y). This matter is left as an exercise for the reader.

2.2.5 ...and carbon dwarfs

The last few years have seen an unexpected addition to the spectral pantheon – carbon dwarfs, forming a parallel sequence to K and M dwarfs in the H–R diagram. Carbon rich giants were noticed first by Secchi in 1868 [S9], who commented that the stellar spectra showed, ‘a considerable analogy with the reversed spectrum of carbon’ (reversed, since the laboratory carbon spectrum is in

emission, while the stellar spectrum is in absorption). Secchi classed these as Type IV in his spectral classification scheme. Fleming included carbon giants in her investigation of ‘peculiar spectra’, and added most to the Harvard system as class N [F5]. Pickering identified a number of bluer carbon stars, which were assigned to class R. Keenan and Morgan absorbed both sets of stars into the MKK system as type C, classed as C0 to C9 based on the strength of the atomic lines, rather than molecular features [K11]. At the same time, Keenan and Morgan noted the existence of CH stars – carbon stars, ‘characterized by unusually great CH intensity’. These prove to be metal-poor halo subgiants (see Chapter 7). Finally, J-type carbon stars were added to the mix – stars where the C^{13}/C^{12} isotope ratio is unusually high, leading to strong $C^{13}N$ absorption bands [B14].

The first carbon dwarf – a proper motion star, G 77-61, at a distance of 58 parsecs [D3] – was identified over a century after Secchi first brought carbon giants to the attention of the astronomical community. Only a handful of other dwarf carbon stars were discovered over the next two decades, but follow-up observations of very red stars from the Sloan survey have uncovered increasing numbers within the last few years [M14, D4]. While relatively few have direct distance measurements, it is clear that these stars have absolute visual magnitudes matching those of late-K, early- and mid-type M dwarfs. Optical spectra of typical examples are shown in Figure 2.7.

The formation mechanism for carbon giants is well understood. Carbon is a product of the triple- α nuclear reaction (see Chapter 3), where three helium nuclei combine to form one carbon nucleus. This reaction ignites near the top of the red giant branch. During evolution up the second, or asymptotic giant branch (AGB: post Cepheid/horizontal branch evolution), the outer convection zones dip far below the surface and ‘dredge up’ nucleosynthetic products from the inner regions. Those products change the chemical composition of the outer envelope. Carbon bonds strongly to oxygen to form carbon monoxide. Once the number of carbon atoms exceeds the number of oxygen atoms in the stellar envelope ($C/O > 1$), effectively all of the oxygen is bound in CO, so none is available to bond with titanium to form TiO and produce the characteristic M-type absorption spectrum. As a result, the optical spectrum is dominated by molecular carbon (C_2), CH and CN.

The complication to this scenario is that the triple- α reaction occurs only at temperatures exceeding 10 million degrees, values attained only in the cores of intermediate-mass stars, $M > 1.5 M_\odot$, and only after those stars have evolved beyond the main sequence. Given those constraints, how could a $0.5 M_\odot$ main sequence dwarf acquire so much carbon? The answer seems to be that carbon dwarfs are not single stars, but the more luminous component of a binary system where the other, (initially) more massive star has evolved through red giant, Cepheid, AGB and planetary nebula stages to become a white dwarf. During the planetary nebula phase, the massive component ejects most of its envelope, and some of the processed material is accreted by the lower-mass companion. If those debris come from a C-type AGB star, the outer envelope of the lower-mass component can be polluted to the extent that $C/O > 1$. The net result is a carbon dwarf [G8], classified as type dC. There is also some evidence that the known dC stars are relatively metal-

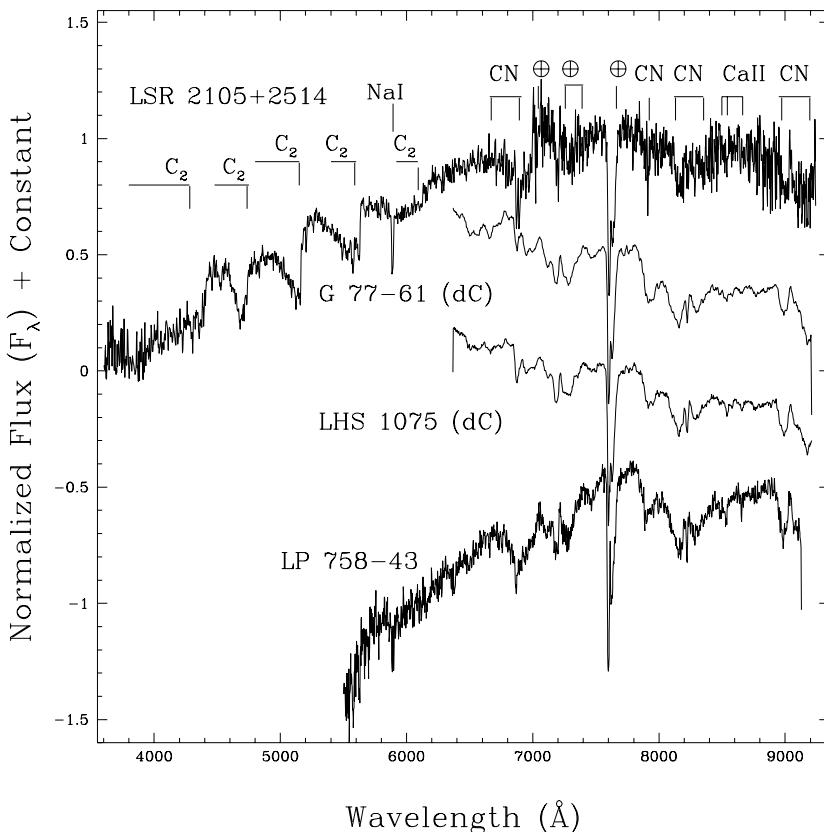


Figure 2.7. Optical spectra of representative dC dwarfs, including G77-61, the prototype. The main molecular absorption features are marked. (From [L11], courtesy of the *Astrophysical Journal*.)

poor, reducing the total mass of C required to give $C/O > 1$. CH stars are similar systems, where the lower-mass component is caught on the subgiant branch. Interestingly, the most recent estimates suggest that carbon dwarfs have a space density of $\sim 3 \times 10^{-3}$ stars pc^{-3} , making them intrinsically rare, but still over a hundred times more common than their very luminous AGB counterparts.

2.2.6 Near-infrared spectroscopy

M, L and T dwarfs emit most of their energy at wavelengths between 1 and $2.5\text{ }\mu\text{m}$. Until recently, the only observations possible were of low spectral resolution – effectively narrow-band photometry with a tuneable filter and a single-element detector [F2]. This state of affairs has changed radically with the availability of large format (up to 1,024-square) two-dimensional infrared detectors, and it is

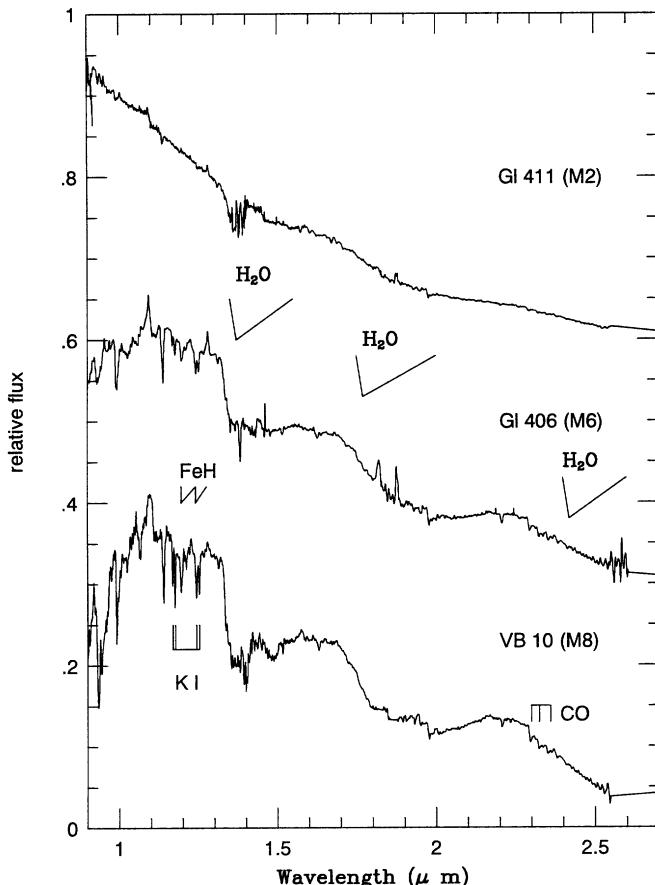


Figure 2.8. Near-infrared spectra of three late-type dwarfs: Gl 411 (M2), Gl 406 (M6) and VB 10 (M8). The most prominent features are the steam bands due to H_2O in the atmospheres of these cool stars, but CO and FeH bandheads as well as various atomic lines are also present. (Spectra courtesy of H. R. A. Jones.)

now possible to obtain near-infrared spectra that are comparable with optical data in both quality and resolution.

Figure 2.8 shows $1\text{--}3\,\mu\text{m}$ spectra of early-, mid- and late-type M dwarfs. The most prominent features – striking even in the first low-resolution observations – are the broad absorption bands at 1.4 , 1.8 and $2.4\,\mu\text{m}$ due to water. These bands are also present in the Earth’s atmosphere and, indeed, are effectively opaque at low altitudes above sea-level. Thus, only the wings of the stellar steam bands, broadened at the higher temperatures and pressures of the M dwarf atmospheres, are detectable from most astronomical observatories. However, there are a few dry, high-altitude sites – notably Mauna Kea, and parts of Antarctica – where it is possible to detect stellar radiation across at least the two shorter wavelength bands. As with TiO at optical

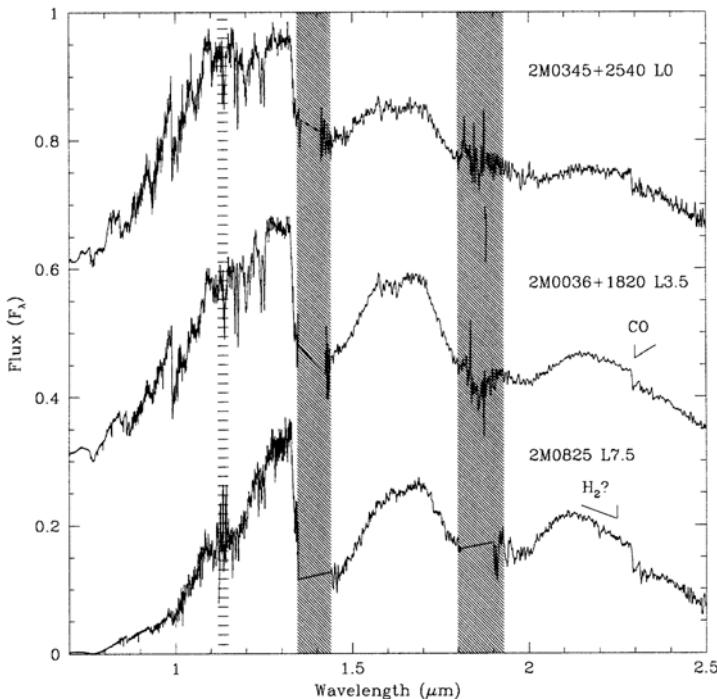


Figure 2.9. Near-infrared spectra of L dwarfs (from [R7], courtesy of the *Astronomical Journal*); the shaded columns mark regions of the spectrum affected by terrestrial water absorption. Most of the features are also present in M dwarfs, but are stronger in the L dwarf spectra.

wavelengths, the steam bands are barely perceptible at spectral types K7 and M0, but grow in strength with decreasing temperature.

Other weaker molecular bands – notably CO at $2.3\,\mu\text{m}$ – are also present, while recent higher-resolution observations have identified FeH in the $1.1\text{--}1.2\,\mu\text{m}$ region [J2]. The latter authors also identified atomic lines due to Fe I and Mg I in the same region, while strong K I lines were detected previously [K3], and Na I absorption may have been detected at $2.1\,\mu\text{m}$. These lines mirror the behaviour of the atomic lines at optical wavelengths, growing stronger in the later spectral types. An increasing number of L dwarfs have been observed at these wavelengths [L12, R7], and the overall trend continues in these cooler objects (Figure 2.9).

As in the far-red spectra, the atomic lines and hydride bands evident near $1.2\,\mu\text{m}$ weaken with increasing spectral type, while the water bands strengthen. Collisionally induced H_2 absorption appears in the K band ($2.2\,\mu\text{m}$) in the latest-type L dwarfs, flattening the flux distribution shortward of the CO $2.3\,\mu\text{m}$ bands [T4].

As the brown dwarf evolves, the atmosphere continues to cool. At some point, the temperature in the atmospheric layers that correspond to the photosphere (see

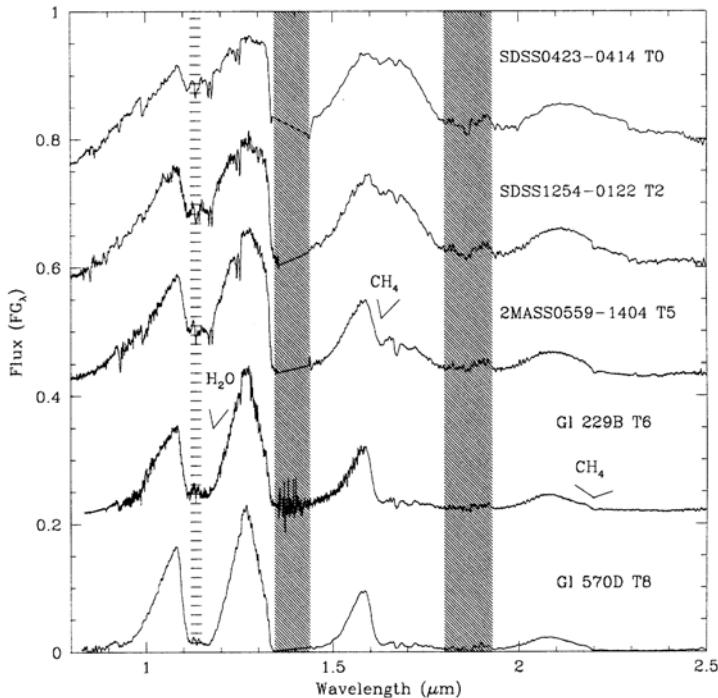


Figure 2.10. Near-infrared spectra of T dwarfs – as in Figure 2.9, the hatched areas mark regions affected by terrestrial water absorption. The strongest features are the methane bands at 1.6 and 2.2 μm and the water band at 1.1 μm .

Section 4.1) drops to sufficiently low values that extensive methane formation is allowed. At that juncture, strong overtone CH_4 absorption bands appear at near-infrared wavelengths and remove over half the flux in the H (1.6 μm) and K passbands. The presence of those bands in the 1–2.5 μm range is the primary characteristic that defines spectral class T. It is misleading to refer to T dwarfs as ‘methane dwarfs’, however, since a stronger methane band lies at 3.3 μm , and that feature is detected in mid-type L dwarfs. Moreover, recent observations suggest that a handful of late L dwarfs (L6.5 to L8) exhibit very weak CH_4 absorption at H and K (see Section 6.7.3).

Near-infrared spectral classification schemes for T dwarfs have been defined by both the SDSS [G9] and 2MASS [B11] groups. Both systems are based primarily on the strength of the water bands, particularly the depth at 1.5 μm (see also [R7]), and the methane bands. By and large, the two calibrations agree to better than 1 spectral class for individual T dwarfs. Both systems can also be used to estimate spectral types for L dwarfs, although with some caveats (as described in more detail in Section 6.7). Figure 2.10 shows representative T dwarf spectra, illustrating how the methane bands and H_2O absorption in the J-band (1.2 μm) strengthen with increasing spectral type. The underlying changes in the brown dwarf atmospheres are considered in more detail in Section 6.7.

2.2.7 Mid-infrared and longer wavelengths

Ground-based observation of late-type dwarfs at mid-infrared ($\lambda = 3$ to $100\,\mu\text{m}$) and sub-millimetre ($100\,\mu\text{m}$ to $1\,\text{mm}$) wavelengths is complicated by the absence of any clean atmospheric windows. Faint targets must be detected against a bright, variable sky background. Broadband photometry in the L ($3.5\,\mu\text{m}$) and M ($5\,\mu\text{m}$) windows has become more common in recent years, as discussed in the following section, but spectroscopic data are rare. The few available observations, such as the 8 – $12\,\mu\text{m}$ spectrophotometry of Gl 411 and Gl 406 obtained by Aitken and Roche (see [B3]), show little evidence for significant deviations from a smooth black-body distribution.

Space-based observatories have higher sensitivity, since the transparency is higher and the background significantly lower. The IRAS satellite provided the first information on M dwarfs at these wavelengths, with 12 , 25 , 60 and $100\,\mu\text{m}$ data for early- and mid-type M dwarfs ([W1], [W2], [M4]). Those observations show that, in general, the $12\,\mu\text{m}/K$ -band flux ratio rises from $\sim 5\%$ for M1 dwarfs to $\sim 10\%$ for M5 dwarfs, while the fluxes at $25\,\mu\text{m}$ are lower by approximately a factor of two. These ratios are broadly consistent with smooth, black-body distributions, with only a few stars showing any evidence for excess flux.

With only a 20-cm telescope, IRAS was unable to provide spectroscopy of faint, late-type dwarfs. Nor could its immediate successor, ISO, the European Infrared Space Observatory. However, late 2003 saw the launch and successful deployment of the Spitzer telescope (known formerly as SIRTF, the Space Infrared Telescope Facility). Spitzer will be undertaking a thorough survey of late-type dwarfs over the next few years, but even the initial few observations provide key insights to spectral variations, with the detection of ammonia absorption at 7.8 and $10\,\mu\text{m}$ in late L and T dwarfs ([R9], see Figure 2.11).

Moving to submillimetre and radio wavelengths, there is some evidence that M dwarfs are significantly brighter than expected. In one case, the young M dwarf Gl 803 (or AU Mic), the sub-millimetre excess has been tracked down to a specific cause – a dusty debris disk [K12]. We consider this interesting star in more detail in Chapter 11. MIPS observations with Spitzer should allow a more reliable determination of the characteristic properties of M dwarfs at these longer wavelengths.

A number of stars (including Gl 65AB, Gl 285 and Gl 873) have fluxes at centimetre wavelengths that are four orders of magnitude above the black-body extrapolation [C2], while a few stars may also have been detected at sub-millimetre wavelengths [M4]. Active flar stars are more prone to detection (for example, see Lim *et al.* [L3], which describes the intermittent detection of Proxima Centauri at a wavelength of $20\,\text{cm}$). Mullan *et al.* [M13] note that the infrared/radio excess can be matched by a power-law, $F_\nu \propto \nu^{0.7}$, and suggest its origin in an ionised wind, although Lim and White [L4] argue against that hypothesis.

Nonetheless, the radio emission generated by M dwarfs appears to be related to X-ray activity. Gudel *et al.* [G11] have compared $3.6\,\text{cm}$ and $6\,\text{cm}$ observations of 21 M dwarfs against X-ray data, and find that the luminosities are correlated almost

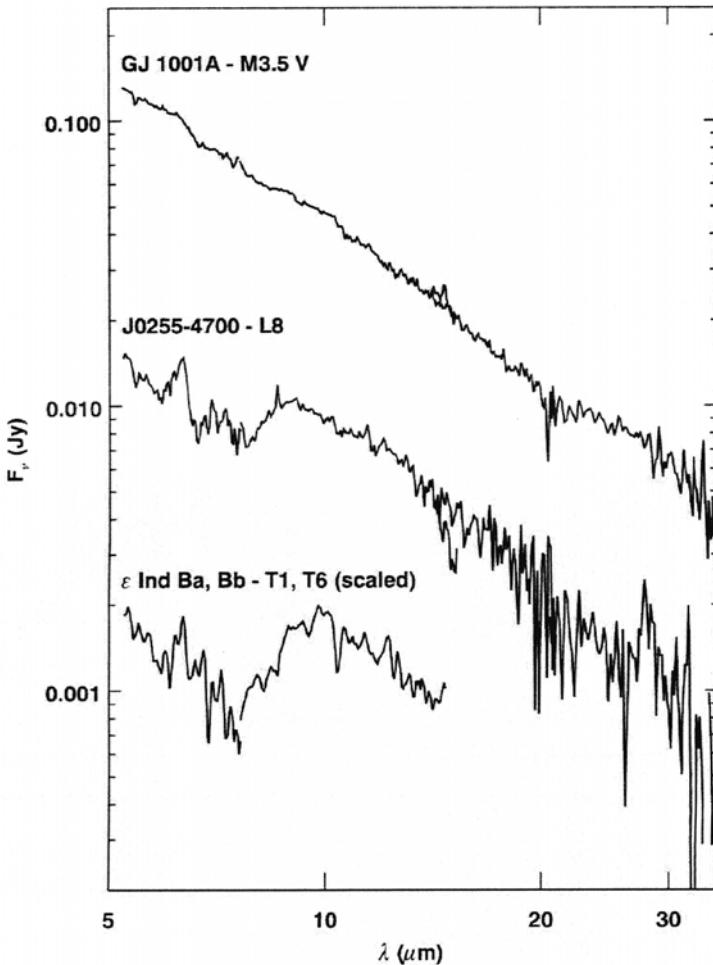


Figure 2.11. Spitzer LRS 5–15- μm spectra of an M dwarf, GJ1001A (LHS 102A), a late-type L dwarf, 2M0255-4700, and the T dwarf binary, ε Indi Bab. Broad ammonia absorption at 7.8 and 10.5 μm is evident in the latter two spectra. (From [R9], courtesy of the *Astrophysical Journal*.)

linearly. The latest type dwarf in the Gudel *et al.* sample is VB8 (spectral type M7), which is not detected at radio wavelengths. Gl 406 (M5) has a radio detection (but no X-ray detection) during a flare, and is undetected at radio wavelengths (but has an X-ray measurement) during quiescence. Gl 65AB (M5.5 + M6) has both radio and X-ray detections, but the components are unresolved at X-ray wavelengths. The latest type dwarfs with reliable data at both wavelengths are Gl 285 (YZ CMi) and Gl 729 (V1216 Sgr), both spectral type M4.5e. It is not clear whether the observed correlation can be extrapolated to spectral types later than M5. Radio

observations of active stars are described in Section 5.5.3, and brown dwarfs are discussed in Section 6.7.5.

2.2.8 Spectroscopic dwarf/giant indicators

M giants are typically 10^4 to 10^6 times more luminous than M dwarf stars of similar temperatures, implying (from equation (1.28)) that giants have radii 100–1,000 times larger than those of dwarfs. Both gravity and pressure are much lower in the more extended atmospheres, leading to spectroscopic differences with the more compact dwarfs. Atomic lines are significantly weaker in the giant stars. Indeed, the strength of Ca I 4,226 Å is one of the principal luminosity discriminants for early-type M-stars in the revised MK system [K1], while the Na D lines (5,890, 5,896 Å), the 8,184/8,192 Å Na I doublet and the Ca II triplet (8,498, 9,542, 8,662 Å) can also be used to discriminate between later-type dwarfs and giants. Balmer lines, in contrast, are stronger in giants earlier than type M3.

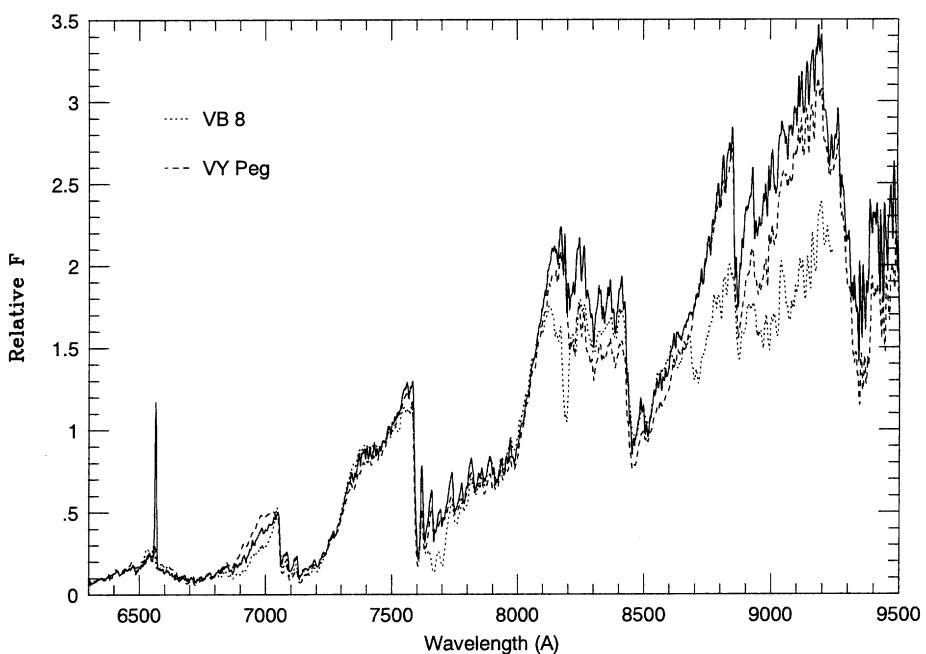


Figure 2.12. Luminosity-sensitive features of M-type stars. The solid line plots data for GG Tau Bb, a 10^6 -year-old T Tauri star in a quadruple system. This spectrum is compared with the M7 giant VY Peg and the M7 dwarf VB 8. The main sequence dwarf has the strongest atomic lines (notably K I at 7,665/7,699 Å) and CaH bands; GG Tau Bb has giant-like atomic features, but hydride bands intermediate in strength between VB 8 and VY Peg. (Courtesy of Russel White and the *Astrophysical Journal*.)

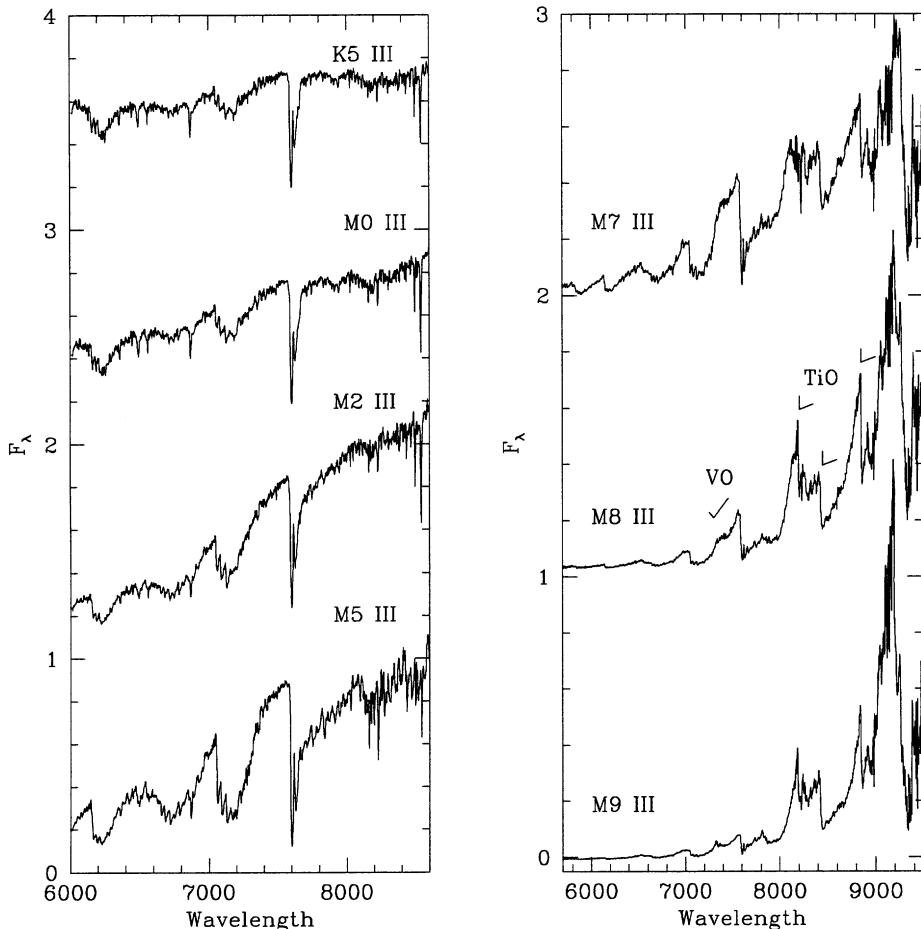


Figure 2.13. An M giant spectral sequence. Note the absence of metal hydride absorption, particularly CaH, which leads to the spectrum having a concave, rather than convex, shape between 6,800 and 7,000 Å. The later-type M giants have a radically different appearance than the dwarfs longward of 8,000 Å.

Molecular bands provide another luminosity-dependent spectral signature. Metal hydride bands are significantly stronger in dwarfs than in giants, and both MgH (in early M) and CaH (in later types) have been used to define luminosity classification. As Figure 2.12 shows, this classification can be extended to include young M-type T Tauri stars, which exhibit CaH bandstrengths intermediate between dwarfs and giants. The latter trait is consistent with the hypothesis that these stars are still contracting onto the main sequence (see Section 3.6). In contrast, CO and CN are significantly stronger in giants than in dwarfs; indeed CN is essentially absent from dwarfs.

M giant spectra become increasingly distinct from dwarf spectra at later spectral types (see Figure 2.13). Vanadium oxide is more prominent due to the absence of significant hydride absorption, and CN bands, which are essentially absent in dwarfs, dominate longward of 8,000 Å.

Narrowband photometric systems can be designed to measure the strengths of these molecular features. Either specifically designed filters are used to measure the flux within an absorption band and at a nearby pseudo-continuum point, or the equivalent data are determined from flux-calibrated spectra. Matched against a suitable temperature indicator (broadband colours, spectral type), the on-band/pseudo-continuum flux ratio measures the bandstrength. Wing [W3] devised an eight-colour system which includes bands centred on TiO (7,120 Å) and the CN band (at 10,395 Å), while Mould and McElroy [M10] used filters centred at 6,880 and 7,120 Å to compare the relative strength of the CaH 6,880 Å and TiO $\gamma(0,0)$ bands. The latter system has been developed to provide a spectral classification system for subdwarfs by Gizis ([G3], see also Chapter 10). In each case, the gravity-sensitive index is plotted against a gravity-insensitive index (such as spectral type or TiO band-strength), allowing separation of dwarfs and giants.

In the near-infrared, the 2.295 μm CO band is stronger in giants than in dwarfs of the same spectral type. The CO, H₂O narrowband system [B1] was designed specifically to measure the strength of that band relative to the temperature-sensitive 1.9 μm steam band. Molecular luminosity-class indicators are more effective at later spectral types, where the bands are intrinsically stronger.

2.3 BROADBAND PHOTOMETRY OF M DWARFS

2.3.1 Spectral energy distributions, temperatures and radii

Spectral type can be used to estimate surface temperature. If a star lies on the main sequence, its intrinsic luminosity can be deduced by matching the temperature estimate against empirical or theoretical H–R diagrams. There are, however, limitations in using spectral type for this purpose. Figure 2.14 plots absolute visual magnitude against spectral type for single stars with accurate parallax measurements. Since types are quoted to only 0.5 subclasses and have an accuracy of at best ± 0.5 , the resultant H–R diagram is coarse-grained, with an rms dispersion of $\sigma(M_V) = \pm 1.5$ magnitudes for early to mid-M dwarfs. Distances estimated from spectroscopic type alone are accurate to only a factor of two, and are therefore of limited value.

Broadband colours provide an alternative means of estimating stellar temperatures. Figure 2.15 illustrates how the Johnson/Cousins passbands sample the spectral energy distribution of late-type M dwarfs. Given the flux zero-points defined in Table 1.1, broadband photometry can be transformed to apparent flux densities, and approximate spectral-energy distributions determined. The peak in the energy distribution shifts towards longer wavelengths with increasing spectral type, as

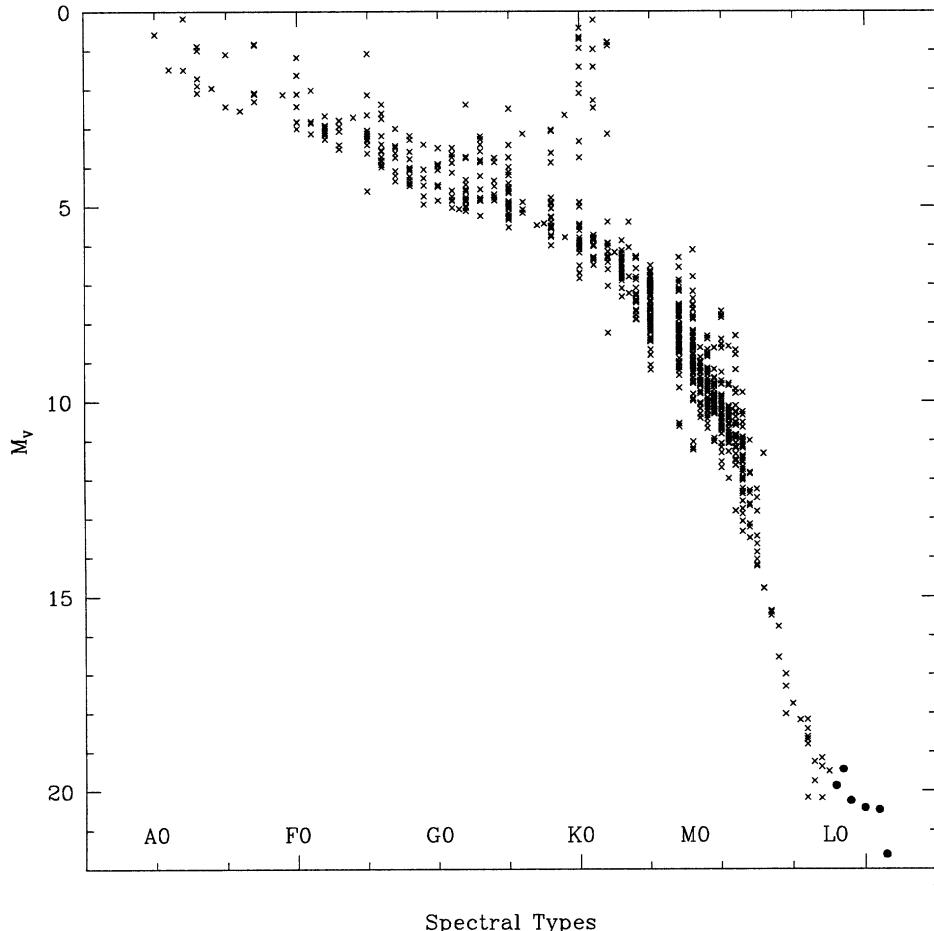


Figure 2.14. The relationship between absolute visual magnitude, M_V , and spectral type, defined by stars with accurate trigonometric parallax data.

increased molecular absorption depresses the continuum at optical wavelengths. As a consequence, the fraction of the total flux emitted shortward of $7,000\text{ \AA}$ decreases from $\sim 10\%$ at spectral type K7 to $< 0.5\%$ at M8.

Integrating the total energy beneath the flux distribution provides an estimate of bolometric magnitude, and hence luminosity. In undertaking these calculations, due allowance must be made for the fact that the near-infrared *JHKLM* passbands were chosen to avoid H_2O absorption in the terrestrial atmosphere, and therefore also avoid the infrared steam bands in late-type M dwarfs (see Figures 2.8 and 2.9). Appropriate corrections (seldom exceeding 0.15 magnitudes) can be incorporated [B7], [L6]. M dwarfs have bolometric magnitudes ranging from $M_{bol} \sim 6.5$ at type K7 to $M_{bol} \sim 13$ for the M9/M9.5 dwarfs LHS 2924 and BR 0021-0216. These

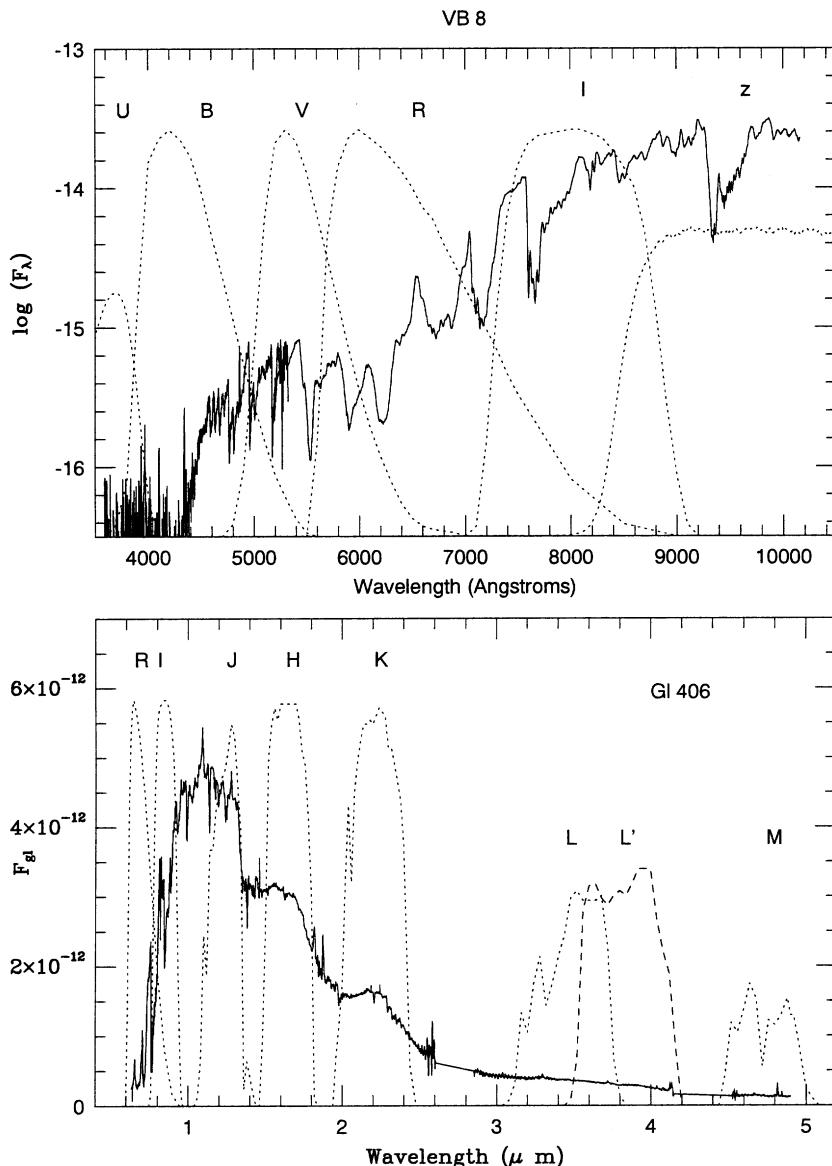


Figure 2.15. Broadband filter response curves.

correspond to luminosities of $0.2 L_\odot$ to $4.5 \times 10^{-4} L_\odot$. L dwarfs reach substantially lower luminosities (as discussed further in Chapters 4 and 6).

In principle, luminosity determinations can be combined with measurements of stellar radii to determine effective temperatures using equation (1.28). In practice, direct measurements of radii of M dwarfs are still relatively rare. Until recently, such

Table 2.1. Radii of low-mass stars.

Star	Sp. type	M _V	V-K	M _{bol}	Radius (R _{Sun})	T _{eff}	Reference
GJ 105A	K3	6.52	2.47	6.17	0.71±0.05	4,300	L13
GJ 380	K7	8.15	3.38	7.08	0.61±0.02	3,900	L13
GJ 887	M0.5	9.77	3.67	8.31	0.491±0.014	3,630	S10
GJ 191	sdK7	10.90	3.79	9.52	0.291±0.025	3,570	S10
GJ 15A	M1	10.32	4.02	8.72	0.38±0.02	3,700	L13
GJ 205	M1.5	9.17	4.08	7.66	0.702±0.063	3,520	S10
GJ 411	M1.5	10.43	4.11	8.82	0.393±0.008	3,570	L13
GJ 699	M4	13.26	5.04	10.90	0.196±0.008	3,160	L13
GJ 551	M5.5	15.49	6.73	11.95	0.145±0.011	3,040	S10
GJ 278C	M0	9.0	3.9	7.55	0.66±0.02	3,800	YY Gem, L7
GJ 278D	M0	9.0	3.9	7.55	0.58±0.02	3,740	
GJ 630.1Aa	M4.5	12.80	5.1	10.25	0.252±0.002	3,150	CM Dra, L1, M5
GJ 630.1Ab	M4.5	12.94	5.1	10.40	0.235±0.002	3,150	
GJ 2069Aa	M3.5	11.95	5.1	9.19	0.432±0.005	3,140	CU Cnc, R10
GJ 2069Ab	M3.5	12.31	5.2	9.45	0.391±0.009	3,140	

data were available only for the stars in eclipsing binary systems. However, with the development of infrared interferometers at Palomar, Keck and the ESO Very Large Telescope (VLT), the nearest M dwarfs can be resolved. Despite the fact that these stars have diameters of $\sim 0.5\text{--}1$ milliarcseconds, their radii can be measured to an accuracy of $\sim 5\%$. Table 2.1 summarises the results to date, listing interferometric data and then binary-star measurements. The eclipsing systems are YY Gem (GJ 278CD), a wide companion of the A-type binary, α Geminorum (Castor); CM Dra (GJ 630.1Aab), which has a white dwarf companion (GJ 630.1B) at a 380 AU separation; and CU Cancri (GJ 2069Aab). There are some anomalies in the interferometric data (e.g., the VLT measurements of the M1.5 dwarf, GJ 205, give a surprisingly large radius). The measured radii range from $\sim 0.6 R_{\odot}$ at spectral type K7 to $\sim 0.15 R_{\odot}$ for Proxima Centauri (GJ 551, M5.5).

These direct measurements provide only a sparse calibration. Determining temperatures using other techniques is a complex procedure, and is discussed in detail in Chapter 4. However, *representative* temperatures can be derived by matching black-body curves to broadband energy distributions. Pioneered by Greenstein *et al.* [G7], later studies have anchored black-body curves at either 2.2 μm [R1], [B3], [B4] or the 3.5 μm L-band [T2]. The resultant temperature estimates are $T \sim 3,800 \pm 200\text{ K}$ at M0 and $T \sim 2,100 \pm 200\text{ K}$ at M9, implying radii of $\sim 0.1 R_{\odot}$ for VLM dwarfs such as VB8, VB10 and LHS 2924. From equation (1.29), this places the peak of the energy distribution in F_{λ} between 0.75 and 1.5 μm . The M dwarf and L dwarf temperature scales are discussed in more detail in Chapter 4.

2.3.2 Empirical colour–magnitude diagrams

Figures 2.2–2.5 show that the slope of the spectrum at optical wavelengths increases towards later types. Broadband colours measure that slope. Plotting absolute magnitudes against those colours provides one of the fundamental tools of twentieth-century astrophysics – the colour–magnitude diagram, a proxy for the (L, T_{eff}) H–R diagram. Figure 2.16 plots the ($M_V, (B-V)$) diagram for nearby stars with accurate parallaxes and reliable photometry. Stars with known unresolved binary companions are omitted from this figure and succeeding diagrams. Undetected companions may remain amongst the stars we have plotted, but as far as possible the measurements trace the behaviour of single stars.

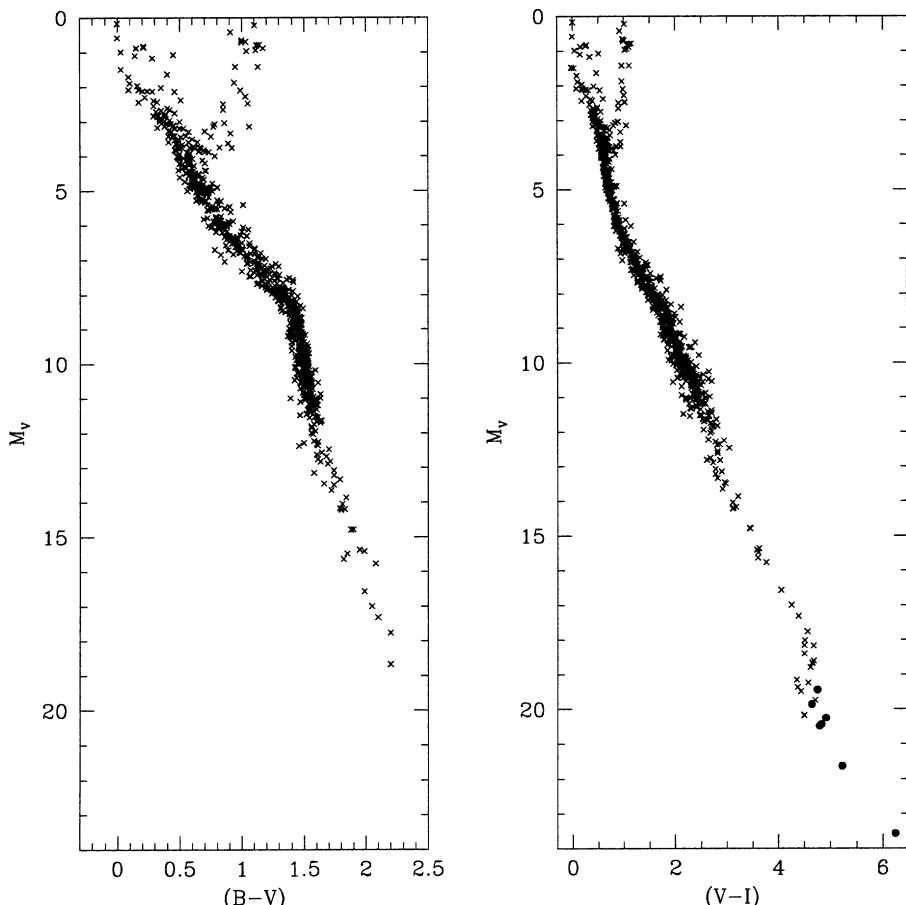


Figure 2.16. The ($M_V, (B-V)$) and ($M_V, (V-I)$) colour–magnitude diagrams for nearby single dwarfs with accurate trigonometric parallaxes and reliable photometry. L dwarfs are plotted as solid points.

The $(M_V, (B-V))$ diagram is the traditional stellar colour–magnitude diagram, a direct descendant of the ‘photographic/photovisual’ diagrams of the 1950s. However, this blue–green colour index is ill-suited for observations of M dwarfs, since the main sequence steepens sharply at $(B-V)$ colour ~ 1.4 (spectral type $\sim M1$). There are two reasons for this. First, this marks the onset of strong molecular bands at optical wavelengths (Figures 2.2 and 2.3); and second, the peak of the energy distribution moves longward of 8,000 Å, placing the B and V bands on the high-frequency tail of the energy distribution. Less than 0.5% of the total flux is emitted in the V -band by spectral type M5, while the B -band contributes a mere 0.15%. The result is that a change in the $(B-V)$ colour of only 0.15 magnitudes encompasses a decrease in M_V of ~ 4 magnitudes. Note, however, the shallower slope at $M_V > 12.5$.

One of the main applications of the colour–magnitude diagram is estimation of photometric parallaxes – using the observed colour to infer an absolute magnitude, which, coupled with the apparent magnitude, allows an estimate of distance. The steep slope of the $(M_V, (B-V))$ relationship clearly mitigates against the effective use of this diagram for late-type stars. The intrinsic dispersion on the main sequence is ~ 1.5 magnitudes in M_V at a given $(B-V)$, leading to distance uncertainties comparable to those derived from the spectral-type/ M_V relation (Figure 2.14).

Photometric parallax accuracy can be improved by using a colour with a larger dynamic range, giving a main sequence with a shallower slope. This can be achieved by choosing passbands spanning a longer baseline in wavelength, although one should avoid choosing two passbands close to the peak in the spectral energy distribution. The most effective passband combination includes one filter close to the energy maximum in the near-infrared, with the second placed at optical wavelengths.

The $(V-I)$ colour index offers many such advantages. Technologically, the I-band is the reddest well-calibrated passband measurable with photomultiplier tubes or CCD cameras, so both V and I can be measured with the same instrument (as opposed to combining optical and near-infrared observations, which require different instrumentation). The photometric system is also well defined, with an extensive grid of standards of all spectral types (see Chapter 1). Figure 2.16 demonstrates the advantages offered by $(V-I)$ for classifying early- and mid-type M dwarfs, although $(V-I)$ reaches a maximum for stars of $M_V \sim 18$ (spectral type M8) and becomes bluer for lower-luminosity stars. The latter behaviour probably reflects increased molecular absorption in the I-band by VO and FeH at these very low temperatures, as illustrated in Figure 2.3.

The main sequence changes slope at several points over the 20-magnitude range plotted in Figure 2.16, at $M_V = 7, 8.5, 12$ and 13 . The last two inflections lead to an almost vertical section at $(V-I) \sim 2.9$. Above and below this step in the main sequence, the data are well matched by the following polynomials,

$$M_V = 3.98 + 1.437(V - I) + 1.073(V - I)^2 - 0.192(V - I)^3, \\ 0.85 < (V - I) < 2.85 \quad (2.1)$$

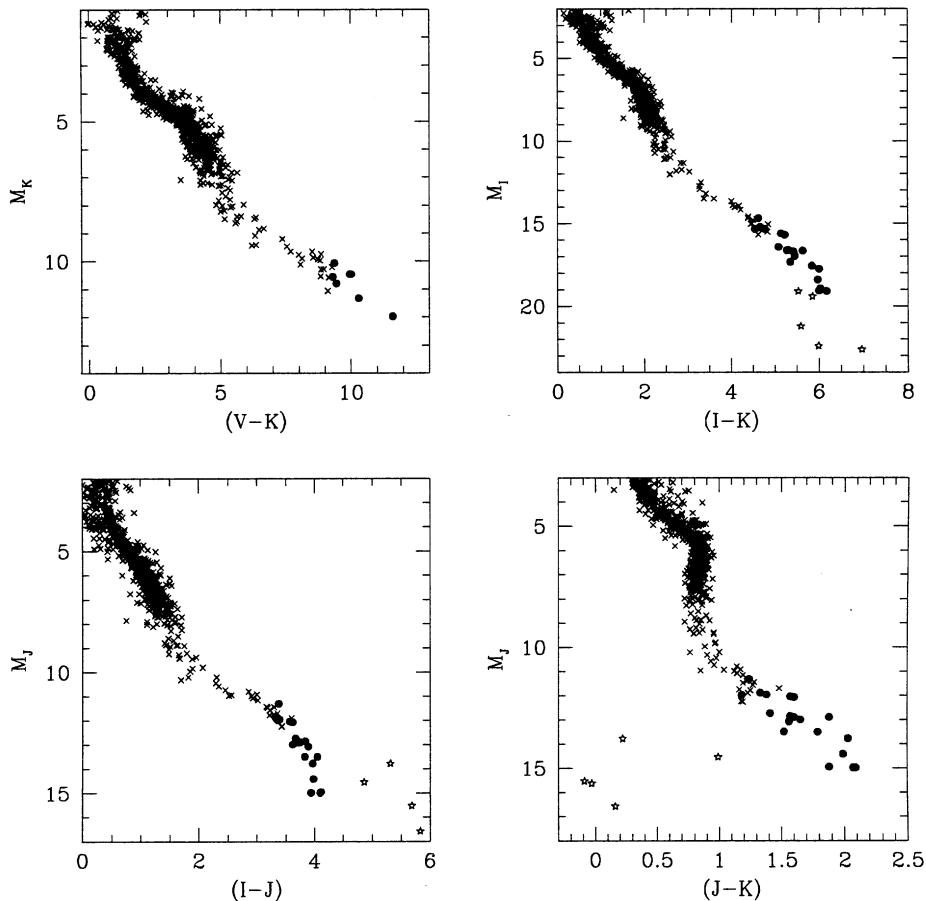


Figure 2.17. Optical/near-infrared colour–magnitude diagrams for nearby stars and brown dwarfs with accurate photometry and reliable trigonometric parallax measurements. AFGKM stars are plotted as crosses, L dwarfs as solid points, and T dwarfs as five-point stars.

and

$$M_V = 3.66 + 4.46(V - I) - 0.517(V - I)^2 + 0.0448(V - I)^3, \quad 2.96 < (V - I) < 3.45 \quad (2.2)$$

The dispersions about these relationships are only $\sigma_{M_I} \sim 0.32$ magnitudes, but at $(V-I) \sim 2.9$ (spectral type $\sim M4$) stars span a range of 1.5 magnitudes in M_V , or a factor of four in luminosity [R8]. Possible physical reasons for these features are discussed in Chapter 3.

Colours based on optical and near-infrared wavelength passbands offer advantages in studying the lowest luminosity dwarfs. Figure 2.17 plots three such colour–magnitude diagrams, $(M_J, (I-J))$ and $(M_K, (V-K))$, together with the near-infrared

$(M_K, (J-K))$ diagram. The photometry from a variety of sources [B12, L5, K13, 2MASS database]. There is significantly more dispersion in colour at a given absolute magnitude among the early and mid-M spectral types ($(I-K) < 3$) than in Figure 2.16. The reddest points in $(I-K)$ and $(I-J)$ are L dwarfs and T dwarfs with known parallax. Methane absorption at $2.2\text{ }\mu\text{m}$ leads to relatively blue $(J-K)$ colours for T dwarfs. The $(J-K)$ colour clearly provides no luminosity information for M dwarfs with $M_K < 10$ (spectral types earlier than M8).

Passbands longward of $2.2\text{ }\mu\text{m}$ lie on the Rayleigh–Jeans tail of the energy distribution, and therefore show only limited variation with decreasing temperature and luminosity. $(K-L')$ colours range from ~ 0.15 mag. at spectral type M0 to ~ 0.65 mag. at M9 [L5]. Relatively few M dwarfs have $5\text{ }\mu\text{m}$ M -band data, with colours of ~ 0.2 mag. at spectral type M6 [B3].

Broadband colours can also be used to estimate bolometric magnitudes. The bolometric correction for a given passband is defined² as

$$BC_\lambda = M_{bol} - M_\lambda \quad (2.3)$$

To first order, BC_λ depends on the effective temperature of the star, and hence the broadband colour. The empirical relationship between BC_λ and a given colour can be calibrated through moderate-resolution spectrophotometry of representative standard stars. The most thorough analysis has been undertaken by Leggett *et al.* [L6, L14, L15], who derive temperatures and bolometric magnitudes for M dwarfs, L dwarfs and M subdwarfs by combining 0.35 to $5\text{ }\mu\text{m}$ spectrophotometry and broadband measurements. Those data span dwarfs and subdwarfs with spectral types in the range K7 to M6.5. Recently, coverage has been extended to the later-type M, L and T dwarfs by Golimowski *et al.* [G10]. Most of the latter dwarfs lack accurate photometry at visual wavelengths, but K -band bolometric corrections can be derived as a function of spectral type.

Figure 2.18 shows the resulting bolometric corrections, plotting V -, I - and K -band data for the early and mid-type M dwarfs, and K -band corrections for the latest spectral types. The mean relations fitted to those data are as follows:

$$\left. \begin{aligned} BC_V &= 0.27 - 0.604(V-I) - 0.125(V-I)^2, \sigma = 0.059 \text{ mag} \\ BC_I &= 0.02 + 0.575(V-I) - 0.155(V-I)^2, \sigma = 0.061 \text{ mag} \\ BC_K &= 0.42 + 1.486(I-K) - 0.220(I-K)^2, \sigma = 0.048 \text{ mag} \\ BC_K &= 3.93 - 0.0383SpT + 5.36 + 10^{-2}SpT^2 - 2.655 \\ &\quad + 10^{-3}SpT^3 + 4.086 + 10^{-5}SpT^4, \sigma = 0.13 \text{ mag} \end{aligned} \right\} \quad (2.4)$$

The last relation is taken from [G10], where SpT represents spectral type (M1–M9 = 1–9, L0–L9 = 10–19; T0–T8 = 20–28). The substantial corrections in BC_V and their rapid increase with redder colours (decreasing temperature) emphasise the increasing separation between that band and the peak of the energy distribution.

² Bolometric corrections are sometimes quoted in the opposite sense as $M_\lambda - M_{bol}$.

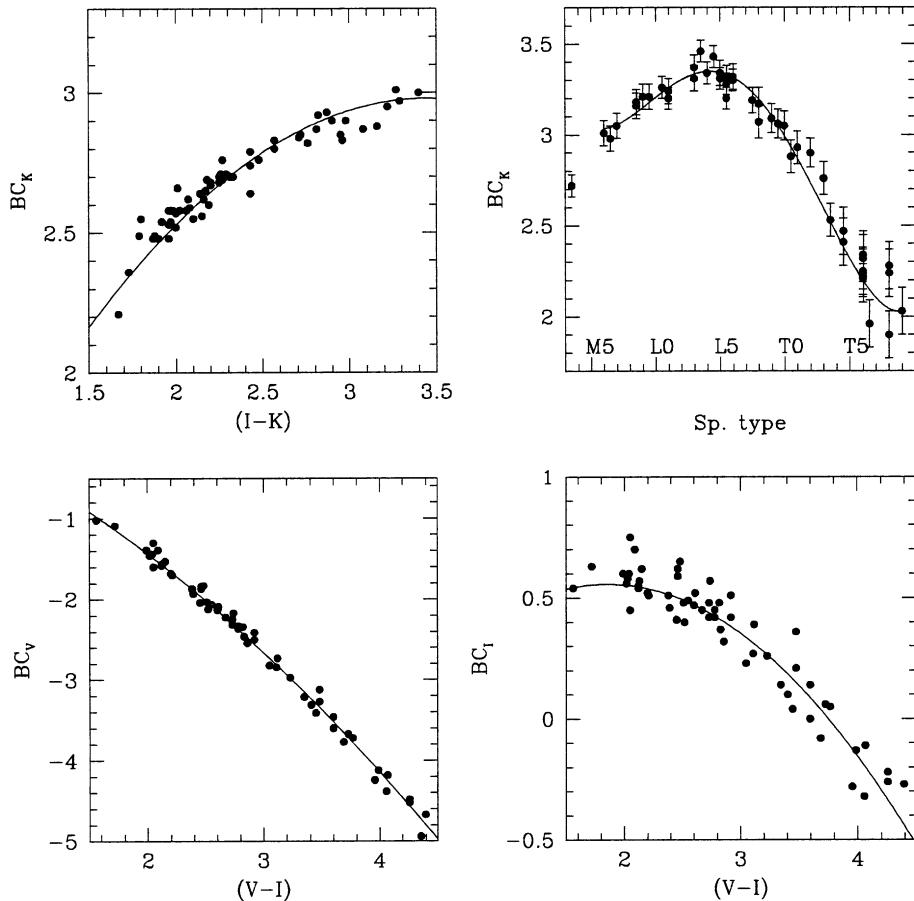


Figure 2.18. Bolometric corrections in the V, I and K bands as a function of $(V-I)$, $(I-K)$ and spectral type. The photometric data span spectral types K7 to M6.5 and are from [L6, L14, L15]; data for later-type dwarfs are from [G10].

The K -band corrections, however, vary by only 0.4 magnitudes between $(I-K) \sim 2$ (spectral type M2) and the latest-type L dwarfs.

2.4 ABUNDANCE EFFECTS ON THE H-R DIAGRAM

The Russell–Vogt theorem states that the position of a star on the H–R diagram is a function of its mass, abundance and age. The overwhelming majority of dwarfs in the vicinity of the Sun are members of the Galactic Disk and have abundances within a factor of two of that of the Sun. The discussion in the previous two sections centred on their spectroscopic and photometric properties.

This section briefly reviews the observational properties of lower metal-abundance subdwarfs in the Galactic Halo (described further in Chapters 4 and 10).

2.4.1 Spectroscopic bandstrengths, absolute magnitudes and abundance

The term ‘subdwarf’ was devised by Kuiper [K9] to describe FGK-type stars which lie significantly below the main sequence defined by the nearby stellar population in the $(M_V, (m_{pg} - m_v))$ diagram. Most of these stars also have high space motions relative to the Sun, and several early studies (notably Chamberlain and Aller [C1]), recognised their low metal abundance relative to that of the Sun. Modern

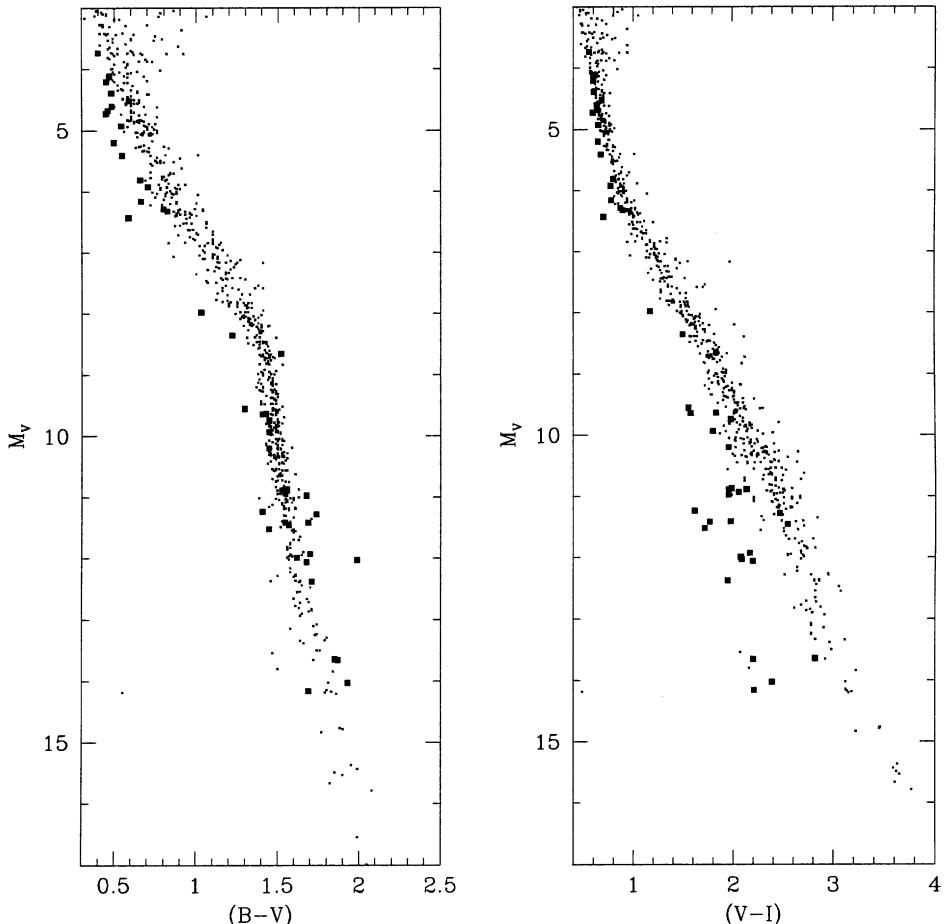


Figure 2.19. The subdwarf sequence in the $(M_V, (B-V))$ and $(M_V, (V-I))$ planes. The subdwarfs, plotted as solid points, all have accurate π_{trig} [M6].

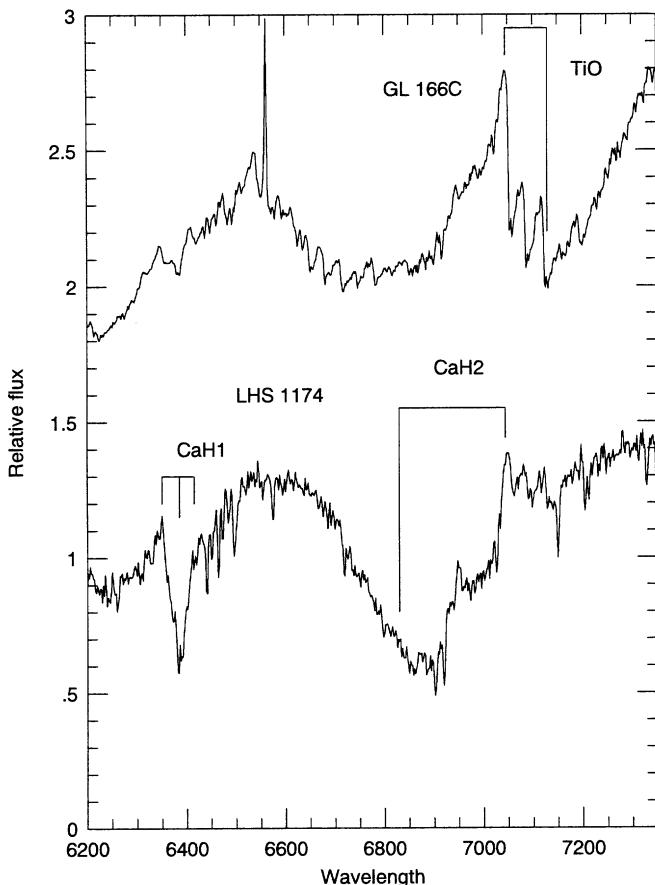


Figure 2.20. Red spectra of the near-solar abundance disk M dwarf GL 166C, $M_V = 12.81$, and of the high-velocity halo subdwarf LHS 1174, $M_V = 12.97$. Despite the close similarity in absolute visual magnitude, there are striking spectroscopic differences, principally in the relative strength of the TiO and CaH bands.

analyses show that the abundance distribution peaks at $[m/H] \sim -1.5$,³ with a low abundance tail reaching $[m/H] = -4.5$. Sandage and Eggen [S4] first demonstrated that the subdwarf sequence lies below (or, rather, at higher temperatures) than the disk dwarf sequence in the $(\log(L), T_{\text{eff}})$ H–R diagram.

The subdwarf sequence(s) extend to the lower main sequence and spectral type M, with Kapteyn's Star (Gl 191; see Appendix) the classic example. Trigonometric parallax measurements of proper-motion stars show that subdwarfs can lie as much

³ $[m/H]$ is the average metal abundance relative to the solar abundance ($Z = 2\%$ by mass) expressed on a logarithmic scale. Thus, $[m/H] = -1.0$ is equivalent to an average metal abundance one-tenth that of the Sun. See Section 4.3, equation (4.6).

as two magnitudes below the disk main sequence in the (M_V , $(V-I)$) plane, but do not become either as faint (at least at visual wavelengths) or as red as the latest type M dwarfs (Figure 2.19). However, the subdwarf sequence crosses the main sequence in (M_V , $(B-V)$), and the latest subdwarfs are redder than disk dwarfs with similar absolute magnitudes. This behaviour stems from the lower metallicity, which leads to lower opacities and higher temperatures at a given mass. As a consequence, a higher proportion of the total luminosity is emitted at optical wavelengths (see Chapters 4 and 10).

Subdwarf stars are identified spectroscopically by comparing the relative strengths of the TiO and metal hydride bands – MgH in K subdwarfs and CaH in M subdwarfs. Figure 2.20 compares red spectra of the disk dwarf Gl 166C and the halo subdwarf LHS 1174. Both stars have absolute visual magnitudes of $M_V \sim 13$, but the subdwarf has only weak TiO absorption. The differences can be quantified using narrow-band spectroscopic indices, designed to measure the flux ratio at the base of a given band against a pseudo-continuum value; that is, by determining the depth of a given spectroscopic feature. This is the same technique employed in dwarf/giant discrimination (Section 2.2.8). Figure 2.20 illustrates the greatly increased strength in CaH absorption in an extremely metal-poor halo subdwarf as compared with a solar-abundance M dwarf of the same luminosity. Chapter 10 outlines how these measurements can be used to estimate approximate chemical abundances.

2.4.2 Effects on broadband photometry

Late-type halo subdwarfs have noticeably different broadband colours than the local disk stars. Plotting the $((B-V)-(V-I))$ and $((V-R)-(R-I))$ two-colour diagrams (Figure 2.21) shows that subdwarfs lie 0.1–0.3 magnitudes above disk dwarfs in both planes: that is, subdwarfs are redder in $(B-V)$ at a given $(V-I)$ and in $(V-R)$ at a given $(R-I)$. This mainly reflects reduced TiO absorption in the I -band.

There are also discernible differences in the behaviour at near-infrared wavelengths. Disk dwarfs and giants earlier than spectral type M0 outline identical sequences in the $((J-H)$, $(H-K)$) diagram, but diverge at later types, with $(J-H) \sim 0.7$ mag. for dwarfs. This probably reflects the change in the temperature gradient with the onset of large-scale convection, reducing the flux emitted in the H^- opacity minimum at $1.6\,\mu\text{m}$ [M9], [M11]. (See also Sections 3.4 and 4.4.) Mould and Hyland [M9] first pointed out that K and M subdwarfs lie below main sequence stars in this plane (Figure 2.22). Anticipating Chapters 4 and 10, the source of this behaviour lies with the increased role of pressure-induced H_2 opacity in the metal-poor atmospheres, with the increased opacity at $1.6\,\mu\text{m}$ reducing the flux emitted in the H -band. In contrast, the reduced H_2 opacity in low-gravity giants leads to increased flux in the H -band and redder $(J-H)$ colours at a given $(H-K)$. The offset between the disk and subdwarf sequences is at most 0.1 magnitude, providing only limited dynamic range for quantitative subdwarf classification. A number of stars with disk-like kinematics (moderate to high Galactic rotational

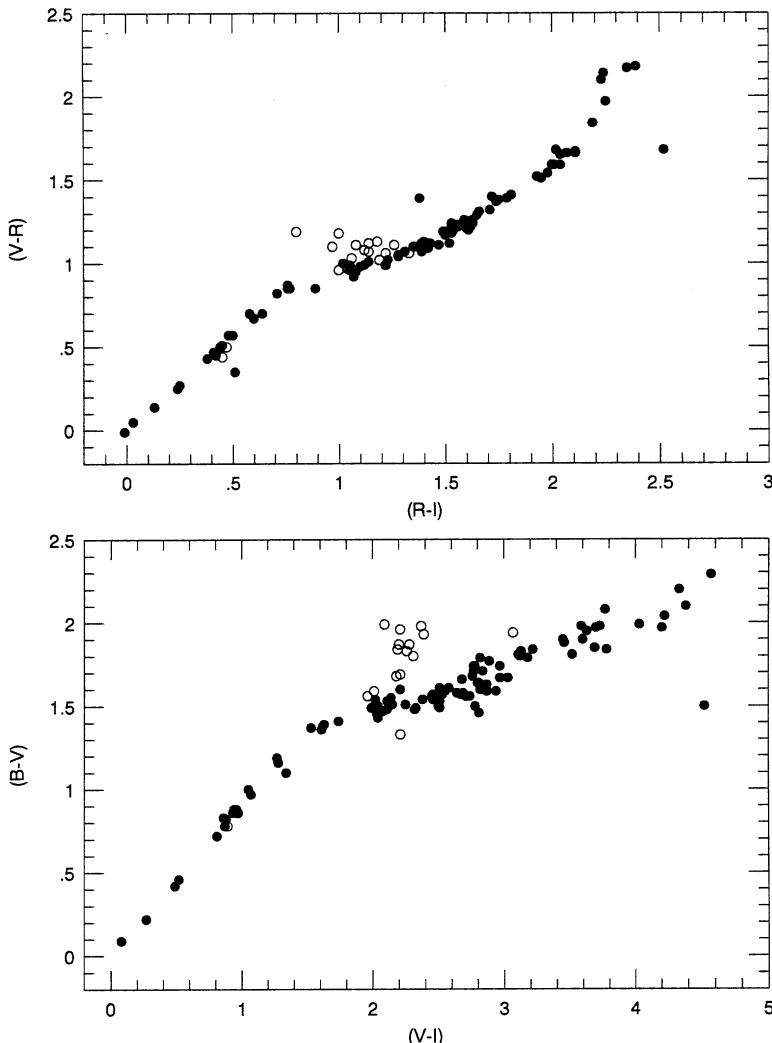


Figure 2.21. The $((B-V)/(V-I))$ and $((V-R)/(R-I))$ two-colour diagrams for main sequence stars. Disk dwarfs are plotted as solid points, while halo stars (photometry by Bessell) are plotted as open circles.

velocities; see Chapter 7) lie close to subdwarfs in this *JHK* diagram, and are identified as ‘old disk subdwarfs’ by Mould and McElroy [M10].

A consequence of the reduced opacity in subdwarfs is that the main sequence terminates at a much brighter visual absolute magnitude than is the case for solar-abundance disk dwarfs. Luminosities at the end points, however, are more similar. LHS 1742a – the lowest luminosity subdwarf in the USNO parallax sample – has an absolute visual magnitude $M_V = 14.74$, and a $(V-I)$ colour of 2.74 magnitudes. The

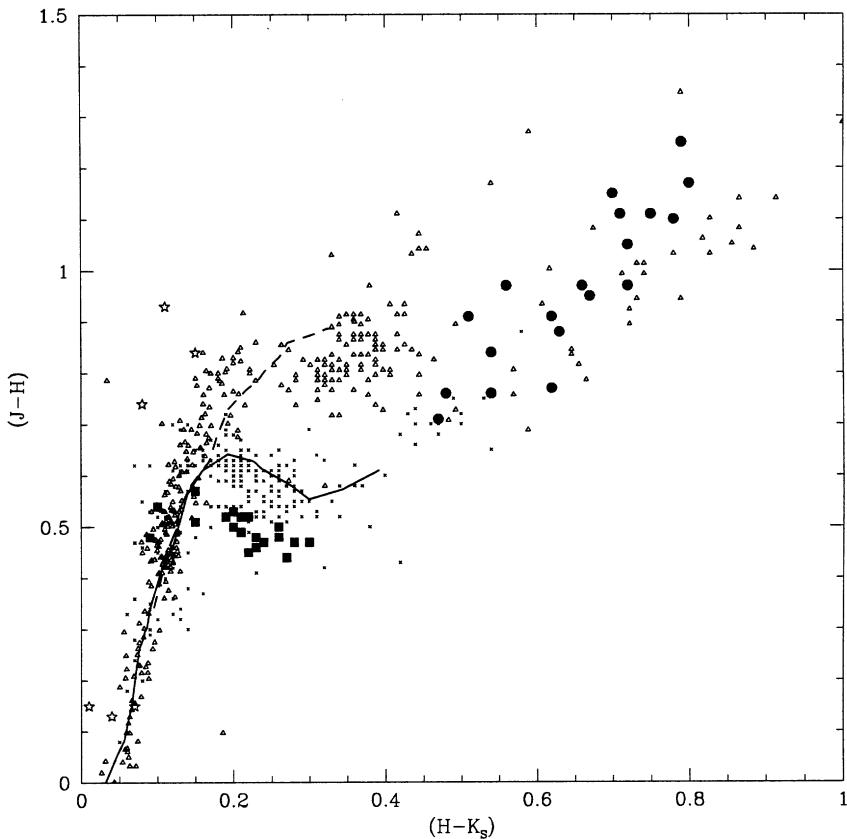


Figure 2.22. The near-infrared $(J-H)/(H-K)$ diagram – disk dwarfs are plotted as crosses, L dwarfs as solid points, T dwarfs as five-point stars, M subdwarfs as solid squares and red giants as open triangles. The solid and dashed lines outline the dwarf and giant sequences.

bolometric magnitude is $M_{bol} = 12.5$, comparable to the luminosity of the disk dwarf VB 10 ($M_{bol} = 12.9$), which has $(V-I) = 4.5$ magnitudes and $M_V = 19.5$. LHS 1742a is close to the limit of the hydrogen-burning main sequence for halo stars (see Chapter 3). Recent surveys have resulted in the identification of even cooler objects, as will be discussed further in Chapter 10.

2.5 ACTIVITY

Most M dwarfs in the Galactic disk appear to maintain nearly constant photometric properties. A subset, however, exhibit substantial, short-term luminosity variations – stellar flares. As with most variable stars, the first flare star was discovered in the course of an unrelated survey project. Hertzsprung [H7] had taken a number of

plates of the Carina region of the Milky Way in 1924. One star (subsequently designated DH Carinae) was brighter by approximately two magnitudes on one exposure, but had returned to its original magnitude in a subsequent exposure. Given that the plates for the project were taken over a relatively short period of time, it was clear that the variability was rapid (days rather than months) but sporadic, since only the one event was observed. With such sparse data and an apparently unique event, Hertzsprung even suggested that a possible cause might be a stellar/planetary collision.

It was not until 1940 that two similar variables were identified by van Maanen [M1, M2]; Lalande 21258B (Gl 412B – WX UMa) and Ross 882 (Gl 285 – YZ CMi). Again these were serendipitous discoveries, made from plates taken for a separate project, and the observations suggested that the increase in brightness was a relatively short-lived phenomenon. However, the extremely short timescale of stellar flares did not become fully apparent until Carpenter’s December 1947 observations of the high proper-motion pair L726-8 (Gl 65AB). Carpenter was working on obtaining a more accurate determination of the trigonometric parallax, taking multiple short exposures on a series of photographic plates. One of the plates showed the fainter component in the pair (UV Ceti) brightening by 2.7 magnitudes in the 20 minutes between consecutive exposures and fading by a factor of four over the succeeding 40 minutes [L8]. Later that same year, Joy and Humason’s spectroscopic observations [J5] of M dwarfs revealed particularly strong emission lines due to the Balmer series of hydrogen and to Ca II H and K during flares.

Since these initial discoveries, numerous other M dwarfs have been found to be similarly active (including VB 10, [H8]), while others, although not yet observed to flare, exhibit substantial Ca II and Balmer-line emission. Stars with Balmer emission are designated dMe stars. As on the Sun, magnetic fields are presumed to provide the mechanism for heating the outer atmosphere (chromosphere, transition region and corona) of the star. The magnetic field strengths for the handful of M dwarfs with direct measurements are $\sim 2\text{--}3 \text{ kG}$ [S1], [S2], [J1], comparable with those measured for large sunspots. The following sections present the general observational picture. A more thorough discussion of magnetic activity on M dwarfs is given in Chapter 5.

2.5.1 Chromospheric activity

The Sun is the only star for which detailed resolution of the density and temperature structure is possible. While the density drops monotonically, with increasing height above the photosphere, the temperature reverses its decline at 600 km above the visible surface and increases rapidly thereafter, reaching 10,000 K at a height of 2,000 km and 10^6 K at 5,000 km. The inner part of this structure is the chromosphere (named because of its strong reddish colour due to $\text{H}\alpha$ emission), visible during eclipses. The outer regions, where the density falls to less than $10^{16} \text{ particles m}^{-3}$, form the solar corona. There is a thin transition region, with temperatures of $10^4\text{--}10^6 \text{ K}$, between the chromosphere and corona.

Most M dwarfs exhibit core emission at the Ca II H and K lines, which are ground state resonance transitions with $E.P. = 4 \text{ eV}$. Hydrogen emission is less

Table 2.2. The fraction of dwarfs with H α emission as a function of spectral type.

Spectral type	<i>N</i> (total)	<i>N</i> (emission)	% (emission)
M0	1,072	9	0.8
M1	871	17	2.0
M2	1,115	31	2.8
M3	1,365	88	6.4
M4	669	160	24
M5	698	353	51
M6	927	527	57
M7	777	497	64
M8	175	129	74
M9	108	70	65
L0	40	23	58
L1	13	4	31
L2	6	2	33

prevalent, since the hydrogen atoms must have a significant population in the $n = 2$ state ($E.P. = 10.2$ eV) in order to produce Balmer emission lines. There is not a strong correlation between the strength of the Ca II and the Balmer emission lines, and many stars with substantial H and K emission exhibit H α in absorption. As discussed in Chapter 5, this may point to a difference in the formation conditions. Ultraviolet and X-ray observations probe the high-temperature gas in the transition region and corona.

Joy and Abt [J4] originally divided M dwarfs into dM and dMe stars, based on the presence or absence of significant (>1 Å equivalent width)⁴ H α emission. Even the initial spectroscopic surveys demonstrated that dMe stars are more common among the later spectral types. Joy [J3], for example, found that 25 of 32 stars with spectral types later than M4 had detectable emission, while only 16 of the 78 M0–M4 stars were dMe dwarfs. Indeed, every star later than M5 in the [J4] survey is listed as an emission-line star, suggesting that this is a defining characteristic of all late-type dwarfs. This proves not to be the case, however. Giampapa and Liebert [G2] carried out the first systematic survey of H α emission in VLM dwarfs and found chromospherically inactive stars even among stars as faint as $M_V = +19$: for example, LHS 2924, spectral type M9, has very weak H α emission. Since then, spectroscopic observations of nearly all of the M dwarfs in the most recent version of the nearby-star catalogue ([G6], the CNS3) have been obtained [H1, G4], and together with the extensive observations of M and early L dwarfs from the SDSS [W4], show that the proportion of dMe dwarfs appears to peak around spectral type M7–M8, with emission becoming increasingly rare at later spectral types (Table 2.2). There

⁴ See Chapter 4 for a discussion of equivalent width.

are, however, a few later L dwarfs with detectable emission, so activity is not quenched completely even in very low-mass dwarfs.

But is this apparent increase in activity among the lower luminosity stars real? As already discussed, increasing spectral type reflects decreasing temperature, decreasing total luminosity, a shift of the peak of the energy distribution towards longer wavelengths and a consequent reduction in the percentage of the total flux emitted at optical wavelengths. The net result is that the continuum flux in the vicinity of the H α line is lower in later-type stars, so a weaker emission line is more easily detectable. In other words, an equivalent width of 1 Å at type M5 corresponds to a lower line-flux in H α than in a line of the same equivalent width at type M0. The extent to which this dilution in the continuum flux affects the dMe statistics can be tested by calculating the ratio between the luminosity in the H α line and the bolometric luminosity, L_α/L_{bol} . This ratio measures the strength of the active emission relative to the total energy budget of the star, and is thus a true measure of ‘activity’, irrespective of the continuum flux in a particular band.

Figure 2.23 plots the log of the F_α/F_{bol} ratio against bolometric magnitude and spectral type. Repeated observations of dMe stars show that most, if not all, exhibit

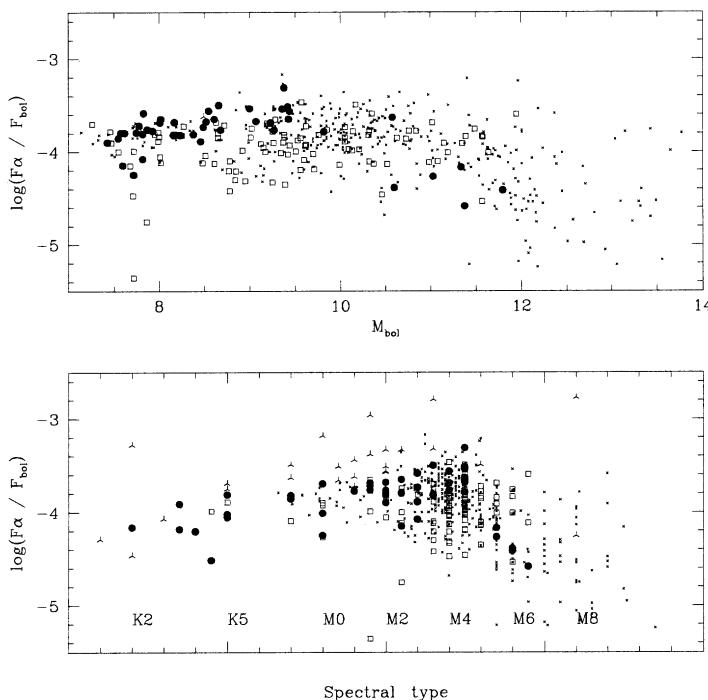


Figure 2.23. The log of the ratio between the flux emitted in the H α line, F_α , and the bolometric flux as a function of M_{bol} and spectral type for field stars in the Solar Neighbourhood (crosses), members of the Hyades cluster (open squares), Pleiades cluster (solid points) and the TW Hydrae Association (triangles).

night-to-night variability of up to a factor of two in the equivalent width (and hence the flux) in the Balmer lines. This intrinsic variability undoubtedly contributes to the scatter in the observed distribution. Nonetheless, the mean level of activity is effectively constant with luminosity to $M_{bol} \sim 12$. Mid-type dMe dwarfs are not predominantly less active than early-type dMe dwarfs, implying that the higher dMe/dM ratio indicates a larger fraction of active stars. The activity strength declines at types later than M6 [W4], as discussed further in Chapter 5.

Figure 2.23 also plots H α activity data for members of three clusters or associations in the vicinity of the Sun: the Hyades cluster (age ~ 500 Myr); the Pleiades (age ~ 100 Myr) and the TW Hydrae Association (age $\sim 10\text{--}20$ Myr). It is clear that the level of activity for stars of a given spectral type increases with decreasing age. The origin of this phenomenon will be discussed in more detail in Chapter 5.

2.5.2 The corona

The existence of the tenuous gas that forms the solar corona has been known (if not understood) since the first observations of a solar eclipse, and photographic studies of coronal structure have been a focal point of most eclipse expeditions since the 1850s. The invention of the coronagraph allowed optical observations of the brighter inner corona outside of eclipse, but detailed quantitative study of the high-temperature gas only became possible with the development of UV and X-ray instrumentation for sounding rockets and, subsequently, satellite observatories. Orbiting X-ray telescopes allowed the first searches for coronal gas around other stars, and one of the major surprises of the extensive survey of the X-ray sky by the Einstein satellite (launched in November 1978) was the discovery that M dwarfs are not only readily detectable at X-ray wavelengths, but that they are amongst the most luminous main sequence stars at high energies. With X-ray luminosities of 10^{20} to 10^{22} Watts, M dwarfs are only slightly less luminous than solar-type stars, and most are significantly more luminous (in absolute terms) than Sirius. Comparing the energy output at X-ray wavelengths against that in the visual, M dwarfs can reach a fractional output L_X/L_{vis} of 10% – a factor of 1,000 higher than the typical value for G-type stars.

Satellites with more sensitive, higher spatial – and spectral-resolution instrumentation – notably the Röntgensatellit (ROSAT) launched in 1990 – have provided direct observations of almost all of the K and M dwarfs within 7 parsecs of the Sun [S5]. While none of these stars are bright enough to allow even moderate-resolution spectroscopy, it is clear that the bulk of the emission is at relatively soft X-ray energies of <1 keV. All but a handful of the K and M dwarfs within 7 parsecs were detected by ROSAT. As with H α , coronal activity is measured by determining the fraction of the total luminosity emitted at X-ray wavelengths. Figure 2.24 plots coronal activity as a function of spectral type for nearby stars and members of the Hyades, Pleiades and TW Hydrae Association, and shows that F_X/F_{bol} ranges from 0.0001 to 0.1%, with later-type stars having relatively higher coronal emission. The absolute luminosities, however, are sufficiently low that X-rays have only been detected from a handful of dwarfs later than spectral

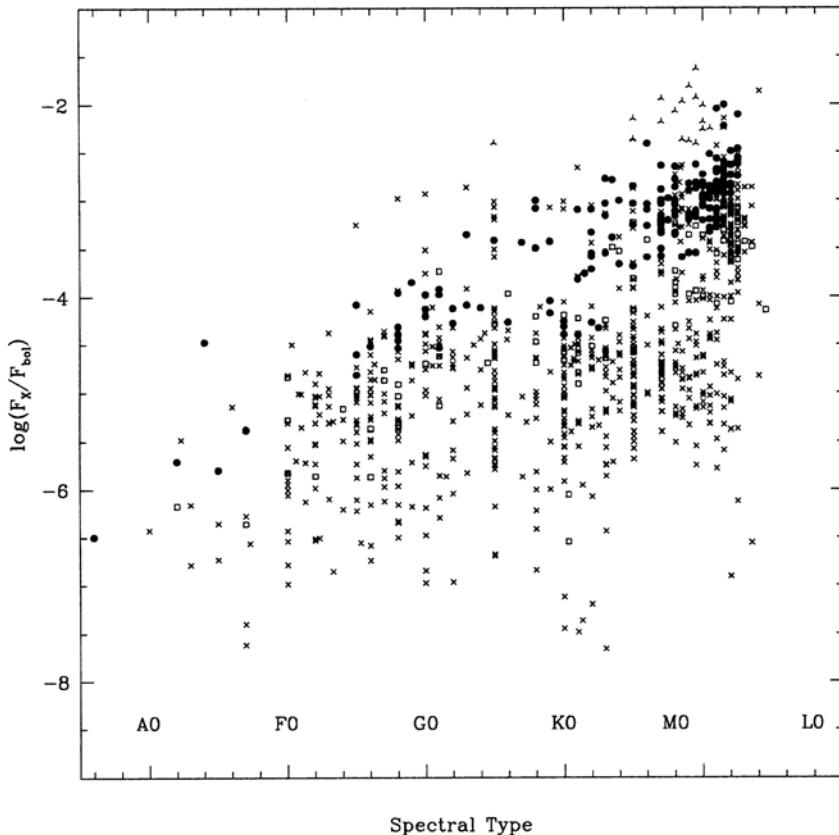


Figure 2.24. The logarithmic ratio between X-ray flux and bolometric flux plotted as a function of spectral type for nearby stars (crosses) and members of the Hyades (open squares), Pleiades (solid points) and TW Hydrae Association (triangles).

type M7. As with $H\alpha$, young stars are more active than older stars of the same spectral type.

The chromosphere and the corona can be regarded as the inner and outer layers of a stellar atmosphere, and the mechanisms responsible for their activity are almost certainly related. Coronal activity is clearly a long-lived phenomenon, as evidenced by the ROSAT detection of significant X-ray flux from most of the local dM dwarfs, some of which are likely to have ages approaching 10 Gyr. Indeed, several low-mass halo stars have been detected, albeit with extremely low X-ray luminosities.

2.5.3 Flares

M dwarf chromospheres, like the solar chromosphere, vary in their level of activity over a wide range of timescales. As already noted, $H\alpha$ linestrengths can vary by a

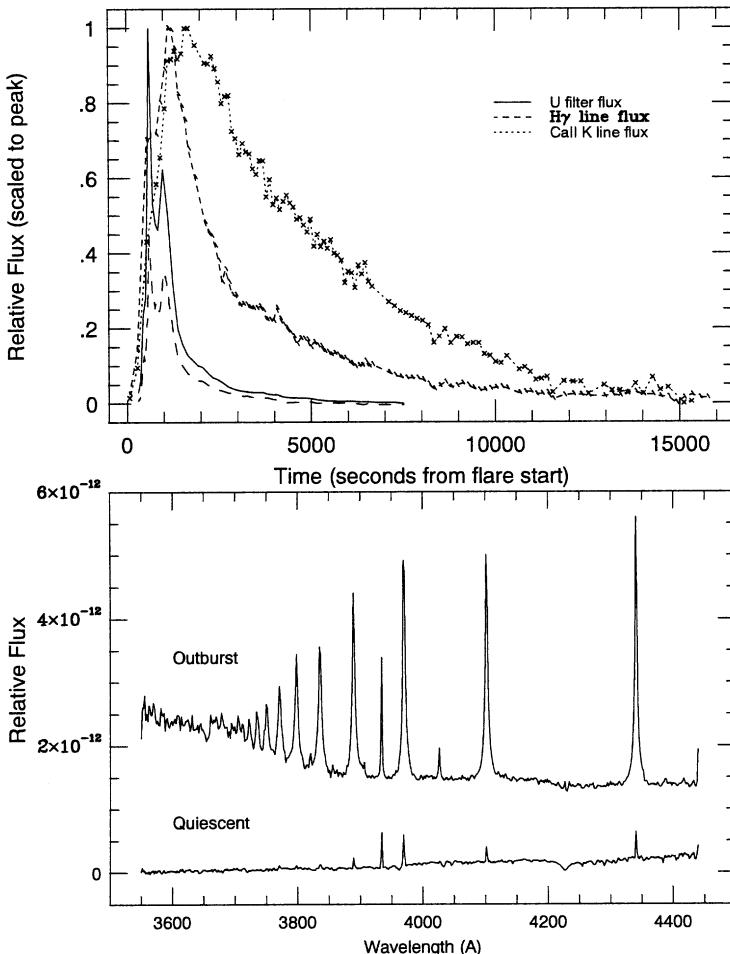


Figure 2.25. (a) Light curves of AD Leo during its outburst of 12 April 1985. The relative variations in the U-band and V-band are compared with the changing Ca II K and $H\gamma$ emission. (b) A comparison between the ultraviolet–blue spectrum of AD Leo in outburst and in quiescence. (Data are from [H2].)

factor of two from one night to the next. Some stars, however, are prone to much more substantial outbursts. The blue and ultraviolet flux can increase by a factor of 100 or more, and emission lines strengthen by an order of magnitude in a matter of a few seconds or minutes, with the star returning to its quiescent state after minutes or hours. These are M dwarfs like UV Ceti, one of the prototype flare stars.

Figure 2.25 plots line and continuum intensity variations during an outburst of the well-studied flare star Gl 388, also known as AD Leo [H2]. As pointed out originally by Bopp and Moffett [B7] in their pioneering high time-resolution observations, the continuum enhancement generally has a strong impulsive phase

and a longer-lasting gradual phase, in accord with solar flare observations. Flares are energetic phenomena with an equivalent black-body temperature of some 9,000–10,000 K; hence the photometric U-band is preferred for observations. The emission lines rise somewhat more slowly and are enhanced for longer than the continuum. The dominant optical emission lines are those of the hydrogen Balmer series and Ca II H and K, which can attain line fluxes of 10–100 times the values seen during quiescence. Emission lines of He I and II, and atomic species such as Ca I, Fe I and II are also commonly observed, while the ultraviolet reveals a rich spectrum of highly ionised emission lines such as C IV, Si IV and N V.

The statistics on flare frequency are sparse, with the seminal work being that of Lacy, Moffett and Evans [L2]. Using the excess flux in the U-band, E_U , as a measure of total flare energy, they found that intrinsically fainter (lower mass, later spectral type) flare stars had more frequent but less energetic flares. Brighter flare stars emit a higher fraction of their bolometric luminosity in flares. In the case of AD Leo, Pettersen *et al.* [P2] derive the relation

$$\log \frac{N}{T} = (15.0 \pm 2.1) - (0.62 \pm 0.09)E_U \quad (2.5)$$

for the cumulative distribution. This predicts, for example, an average time interval of $10^{6.08}$ seconds (~two weeks) between flares with energy $E_U = 10^{27}$ Joules. Observations of other flare stars suggest that the slope of this distribution changes with L_{bol} , with more frequent but less energetic flares on stars with smaller L_{bol} . There is also a statistical correlation between flare rise time and flare energy, such that faster flares are less energetic as a whole.

The most active flare stars emit ~0.1% of their bolometric luminosity in the form of flares. Flares have also been observed on dM stars, where the fractional flare luminosity may be as low as 0.0001%. According to Pettersen [P1], it is the frequency of flaring, and not the average energy per flare, that has declined in dM dwarfs. Section 5.4.2 describes analysis and modelling of stellar flares.

2.6 MASSES OF M DWARFS

Mass is the fundamental parameter that determines the position a star occupies on the main sequence. Unfortunately, the determination of accurate masses is one of the most difficult observational problems in stellar astronomy, since Kepler's laws and Newtonian orbital dynamics provide the only method of directly measuring accurate stellar masses.⁵ This limits observations to stars in binary systems with known heliocentric distances, permitting transformation of angular measurements to

⁵ Gravitational microlensing also provides mass estimates, but the typical uncertainties involved in a single measurement are substantial, while the actual lens is seldom observed directly. Thus, while this technique may provide a statistical determination of the stellar mass function, it is not yet useful for estimating masses of individual stars. See Section 9.3.3 for further discussion.

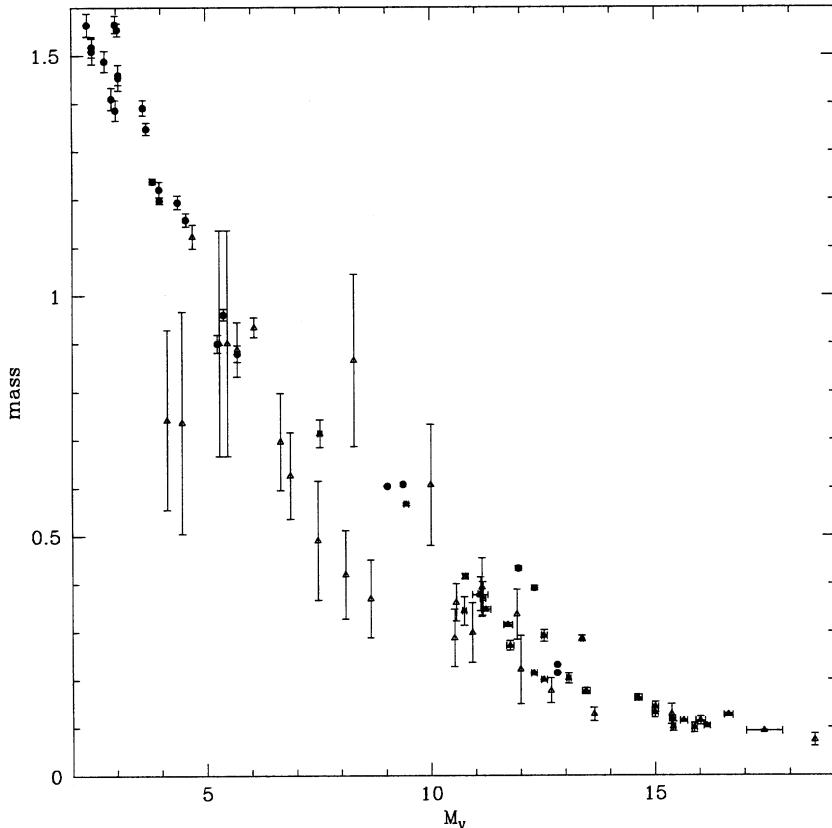


Figure 2.26. The (mass, M_V) relationship defined by stars in binaries with well-determined orbits. Solid points are eclipsing binaries (from [A2] and [D5]) and open triangles are astrometric binaries (from [H4] and [S10]).

linear separations: where the angular separation between components is sufficient to allow measurement of both, but where the period is short enough to permit an orbit determination within a human lifetime. These criteria are met only by a small subset of the stars in the immediate Solar Neighbourhood, within 10–15 parsecs of the Sun.

Masses have been determined for about 30 M dwarfs, either in eclipsing binaries or wider systems with astrometrically determined orbits. Chapter 9 provides a description of the measuring techniques used. Figure 2.26 plots the resulting mass–luminosity relationship at visual wavelengths. Defining M dwarfs as main sequence stars with $M_V > 7.5$, the inferred masses range from $\sim 0.6 M_\odot$ at spectral type M0 to less than $0.1 M_\odot$. As discussed further in Chapter 3, the lower boundary lies close to the mass limit for hydrogen fusion, M_{HBL} . Some L dwarfs, with lower luminosities, and all T dwarfs fall below that limit. Objects with masses $M < M_{HBL}$ have no long-term energy source, and cool rapidly to extremely low luminosities. These are brown dwarfs.

2.7 STELLAR STATISTICS

Previous sections consider late-type dwarfs on a star-by-star basis. However, one of the most important scientific issues centres on the determination of frequency as a function of mass – the mass function. This parameter is of fundamental importance for both star formation theory and Galactic structure. Low-mass stars and brown dwarfs are prime dark-matter candidates, and determining the number density of such objects has occupied the attention of many astronomers over the last 30 years. Chapters 8 and 9 review the extensive studies of this issue. A general point worth emphasising is that statistical analyses are only as reliable as the initial definition of the parent sample. With that in mind, this introductory section concentrates on the statistical properties of stars currently located within the immediate Solar Neighbourhood.

2.7.1 Binary and multiple star systems

An issue which must be addressed in statistical studies is the fact that stars (and brown dwarfs) are not always found in isolation. Indeed, binary and multiple star systems are relatively common. Their frequency can be quantified using two parameters: the multiple star fraction,

$$m.s.f. = \frac{N_{bin} + N_{tri} + N_{quad} + \dots}{N_{sys}} \quad (2.6)$$

where N_{bin} is the number of stellar systems that are binary, N_{tri} , the number of triples, and so on; and the companion star frequency,

$$c.s.f. = \frac{N_{bin} + 2 \times N_{tri} + 3 \times N_{quad} + \dots}{N_{sys}} \quad (2.7)$$

The m.s.f. gives the fraction of stellar systems which include more than one component; the c.s.f. gives the probability that a star, picked at random, has a more luminous companion.

There are two main techniques used to search for binary and multiple systems: direct imaging and radial velocity monitoring. The former is better suited to detecting wide companions in long period orbits, with a detection efficiency that decreases with distance from the Sun. The latter searches for reflex motion in the primary star, and is most sensitive to companions at small separations and high orbital inclinations. It has an efficiency that depends on the apparent magnitude and spectral type of the primary. Both techniques are discussed in more detail in Section 9.3.

Early surveys centred on radial-velocity monitoring of solar-type stars, and suggested that the m.s.f. was close to 100% [A1]. Those analyses, however, were based on magnitude-limited samples, and were therefore subject to several selection effects – notably, a bias towards equal-mass spectroscopic binaries. Such binary stars

have twice the luminosity of a single star of the same spectral type and, as a result, the effective distance limit for detecting these binaries is higher (by $\sqrt{2}$) in a magnitude-limited sample; the sampling volume is therefore higher by $2^{3/2}$; and the frequency of such systems is overestimated by a corresponding amount. More recent surveys are based on volume-limited samples, with the requirement being a fair sampling of the local stellar population. Of these analyses, the most influential is the survey of solar-type stars undertaken by Duquennoy and Mayor [D2]. Their dataset includes 164 primaries with spectral types between F7 and G9, and parallaxes (from the CNS1) greater than $0.^{\circ}045$, although the sample is *not* complete within the corresponding distance limit. Seventy-two (44%) of these stars are spectroscopic, visual or common proper motion binaries, and, allowing for observational selection effects (relatively few of the stars have high-resolution imaging, for example), Duquennoy and Mayor deduce that the total binary fraction may be as high as 70%, implying that over 80% of solar-type stars have companions.

Lower-luminosity stars – in particular, M dwarfs – have received less attention for the usual reason: with fainter apparent magnitudes, observations are more difficult. Limiting a survey to M dwarfs alone should lead to an underestimate of the true binary frequency, since M dwarfs with main sequence companions of earlier spectral type are excluded *a priori*. However, since M dwarfs are four times more common than all other stars combined, this selection effect is correspondingly reduced.

Initial investigations of M dwarf binarity suggest a significantly lower frequency than Duquennoy and Mayor's G-dwarf results. Fischer and Marcy [F1] combined results from a radial velocity survey of 70 M dwarfs [M3] with complementary imaging data, and deduce an overall binary fraction of only $\sim 35\%$. While these stars do not constitute a complete sample, Henry and McCarthy's survey [H4] of the northern ($\delta > -25^\circ$) M dwarfs indicates a nearly identical multiplicity function (34.5%), albeit based on only 29 stellar systems, including eight binaries and two triples. Two of the multiple systems – 40 Eridani and Stein 2051 – have components which are not M dwarfs.

The Appendix to this book lists the current census of all systems known to be within 8 parsecs of the Sun. As discussed in the following section, this sample is likely to be substantially complete for declinations north of -30° , a sample we shall refer to as the northern 8-parsec sample. While as-yet undiscovered low-luminosity companions cannot be ruled out completely, nearly all of the stars have been included in either radial-velocity monitoring programmes or high-resolution imaging surveys, and most have been scrutinised using both techniques. It is therefore likely that only a few stellar companions remain hidden among this sample.

The northern 8-parsec sample includes 143 main sequence stars, nine white dwarfs and four brown dwarfs in 111 systems (including the Sun). Those systems comprise 78 single objects (including four white dwarfs and two brown dwarfs, LP 944-20 and 2MASS1507-1621), 27 binaries (four with white dwarf companions and one, Gl 229A, with a brown dwarf companion), seven triples and the

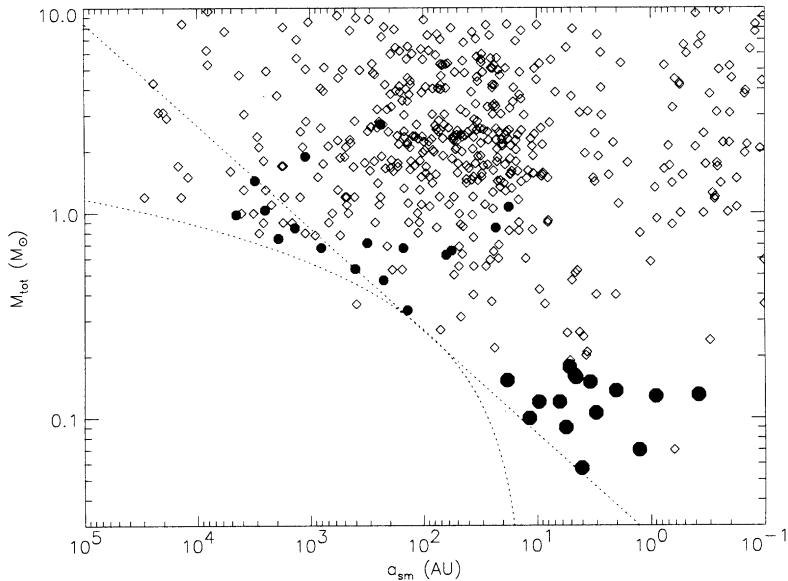


Figure 2.27. Binary component separation (in astronomical units) plotted against the total system mass for a large sample of binaries. Systems plotted as solid points include brown dwarf components. Low-mass systems are confined to small separations. The curved dotted line represents a linear relation between maximum separation and total system mass, and appears to match the envelope of the distribution at high masses; the straight line represents an M^2 dependence, and better matches the low-mass envelope. (From [B13], courtesy of the *Astrophysical Journal*.)

quadruple systems Gl 570 and Gl 644. The overall m.s.f. is only 32.4%, while the c.s.f. is 42.3%.

The 8-parsec binary fractions are similar to the [F1] M dwarf analysis, which is not surprising given the predominance of M dwarfs in the local sample. This reinforces the hypothesis that the stellar multiplicity fraction is mass dependent. Observations of young T Tauri stars suggest that the environment plays a role in defining stellar multiplicity, since the binary frequency is much higher in dispersed star-forming regions, like the Taurus clouds, than in dense associations, like Orion. Recent statistical analyses, however, have identified another important factor.

Figure 2.27 shows the distribution of binary-component separations as a function of the total system mass. The data set is an amalgam of several published binary star catalogues, and includes the most recent results from high-resolution ground and HST observations of late M and L dwarfs (see Section 6.8). There is an obvious trend toward smaller maximum separation at lower masses – Δ_{max} decreases approximately linearly with M_{tot} at high masses, with the trend steepening to M_{tot}^2 below $\sim 0.2 M_\odot$ [B13]. This runs against expected selection effects, since wide binaries are the easiest systems to detect. Clearly, if M dwarf binary systems are restricted to a smaller range of orbital separations than higher mass systems, then,

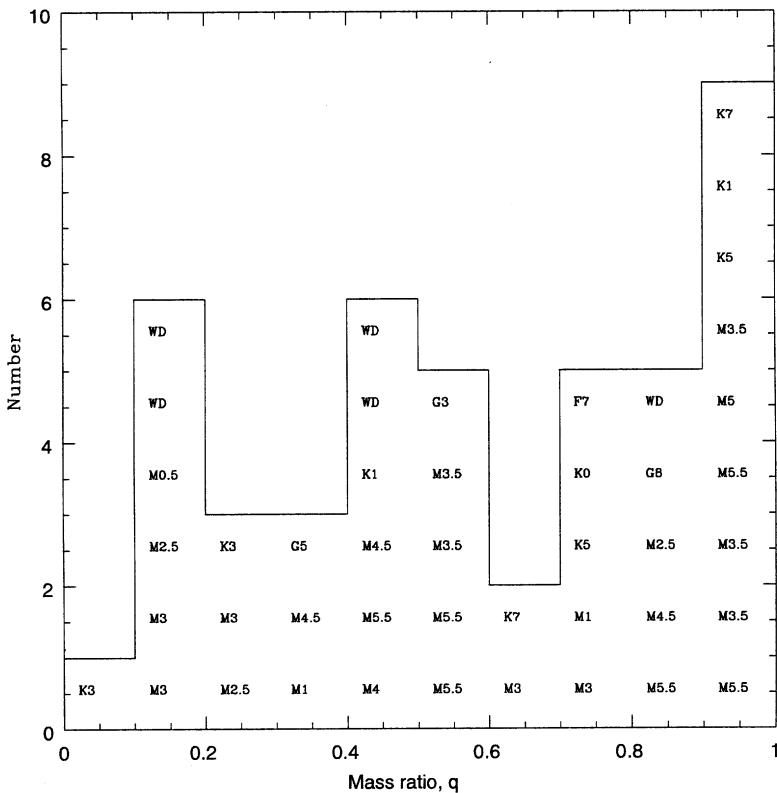


Figure 2.28. The distribution of mass ratios for binaries in the 8-parsec sample.

even if those small-separation systems form with the same frequency per unit separation as, say, G dwarf binaries, the *total* number of M dwarf binaries must be less than the total number of G dwarf binaries.

Besides deriving the multiplicity fraction, the nearby-star data probe the distribution of mass ratios, $q = M_B/M_A$, and test whether they are dependent on the separation between components. This is a thorny issue, beset by observational selection effects. Early statistical studies suggested that there was a preference for mass ratios close to 1 – at least amongst close binary systems. However, a more thorough analysis suggested that this result stemmed from a bias towards double-lined spectroscopic binaries in the observational sample – binaries which, by definition, must include stars of near-equal mass. Trimble [T3] reviewed and re-analysed prior results, and concluded that on balance, the observational data favoured a distribution, $F(q) \propto q^{-1}$; that is, a preference for low mass-ratios. This is the form of the distribution expected if stars are selected at random from a mass function which rises with decreasing mass. However, it is important to bear in mind that most stars in the various surveys analysed in [T3] are at least as massive as the Sun.

Orbits have not yet been determined for all of the binaries in the 8-parsec sample, as systems with separations of tens or hundreds of AU have periods of centuries or more. However, if the orbital inclination is random, then the projected separation can be statistically related to the semi-major axis through the simple scaling

$$a = 0.80 \times \Delta \quad (2.8)$$

where Δ is the observed separation. There is no clear correlation between q and a for stars in the 8-parsec sample. However, if the mass ratio distribution is considered (Figure 2.28) then, contrary to results from analyses of higher-mass stars, there seems to be a preference for near equal-mass systems. Clearly a larger sample and better statistics are required before firm conclusions can be drawn.

2.7.2 The stellar luminosity and mass functions

The stellar luminosity function is defined as the number of stars per unit absolute magnitude per unit volume, $\Phi(M)$. Combined with an appropriate mass–luminosity relationship, this can be used to derive the mass function, $\Psi(M)dM$ or $\xi(\log M)d\log M$, which describes the relative frequency of formation of stars of different masses: that is, how a molecular cloud fragments.⁶ Integrating the latter determines the stellar contribution to the local mass density – a parameter of significance in understanding Galactic dynamics and the nature of dark matter. The simplest method of determining $\Phi(M)$ is to conduct a stellar census within a specified volume. The crucial point lies in assessing the likely completeness of the resulting sample – an exercise which can be undertaken for the sample of stars within 8 parsecs of the Sun.

The most recent version of the *Catalogue of Nearby Stars*, supplemented by data from the *Hipparcos* astrometric satellite [E1], lists 191 stars in 139 systems with parallaxes exceeding 125 mas ($r < 8$ parsecs). Data for these stars are listed in the Appendix, where the sample is divided into northern and southern subsets, setting the boundary line at $\delta = -30^\circ$. The rationale for this split lies in the distribution of terrestrial observatories: more than 90% are in the northern hemisphere, reflecting the distribution of land mass and population. As a result, northern skies have been scrutinised in more detail, and for a longer period, than the southern skies. Most nearby stars were discovered through their having high proper motion, and the deepest such surveys are Luyten's Palomar proper-motion catalogues [L9]. Since Palomar lies at a latitude of $+30^\circ$, Luyten's observations became increasingly difficult at southern declinations, and the declination $\delta = -30^\circ$ provides a natural division of the sky. This eliminates from consideration only the southernmost 25% of the sky. In fact, comparison of Tables A.1 and A.2 show that the relative number of systems north and south of this division is close to the expected ratio of 3 : 1; although there are almost 25% fewer companions known amongst the southern sample.

⁶ M is used for magnitude, M is used for mass when discussing the luminosity and mass functions.

Table 2.3. Distance distribution of nearby stars, $\delta > -30^\circ$.

M_V	< 5 parsec	$5 < r < 8$ parsec	$8 < r < 10$ parsec
≤ 4	3	6	10
$4 < M_V \leq 7$	4	16	29
$7 < M_V \leq 10$	2	13	22
$10 < M_V \leq 12$	8	26	55
$12 < M_V \leq 14$	5	26	47
> 14	11	17	30
All	34	108	204

But how complete is the current northern 8-parsec catalogue? Since most stars were identified from proper-motion surveys, this introduces the possibility of a bias against stars with low tangential velocities relative to the Sun. Luyten's main catalogue, however, is the LHS (Luyten Half Second) catalogue, including stars with $\mu > 0.^{\circ}5 \text{ yr}^{-1}$. These stars have extensive follow-up spectroscopy, and it is unlikely that any LHS stars within 8 parsecs remain unidentified. The proper motion limit corresponds to only 18 km s^{-1} at 8 parsecs, and fewer than 20% of disk stars can be expected to have such low rates of motion. In fact, 14 of the 103 systems in Table A.1 have $\mu < 0.5 \text{ arcsec yr}^{-1}$, suggesting that at most a handful of such low-velocity systems remain to be discovered. Indeed, several such systems have been discovered in the last few years (see [R10]).

Sample completeness can also be tested by comparing the number distribution with distance. Given the small number of systems in total, the northern 8-parsec sample cannot be subdivided too finely. Table 2.3 lists the number of systems within the distance limits 0–5, 5–8 and 8–10 parsecs for several luminosity intervals. Binary and multiple star systems are binned in M_V based on the brightest star in the system. Volume sampling indicates that the observed star counts should increase by a factor of 3 between the first and second distance intervals, while there should be a factor of 2 difference between the numbers in the second and third columns. The expected ratios hold for $M_V < +14$. Taken at face value, the data for the faintest stars suggest a shortfall of a factor of 2 in numbers at $r > 5$ parsecs, although the deficit is less than 2σ given the sampling uncertainties in the 5-parsec sample. Overall, the northern 8-parsec sample is likely to be $\sim 90\%$ complete for H-burning stars.

Accepting these uncertainties, Figure 2.29 shows $\Phi(M_V)$ and $\Phi(M_J)$ derived from data for the main sequence stars in the northern 8-parsec sample. These luminosity functions are constructed on a star-by-star basis, with the dashed lines indicating the contribution from stellar companions. The visual luminosity function reaches a maximum at $M_V \sim +12$, and declines toward fainter magnitudes; $\Phi(M_J)$ peaks at $M_J \sim 8$. The extended tail towards faint absolute magnitudes in $\Phi(M_V)$ reflects decreasing temperature and increasing bolometric corrections rather than a substantial change in luminosity; $\Phi(M_J)$ provides a better representation of the distribution with bolometric luminosity.

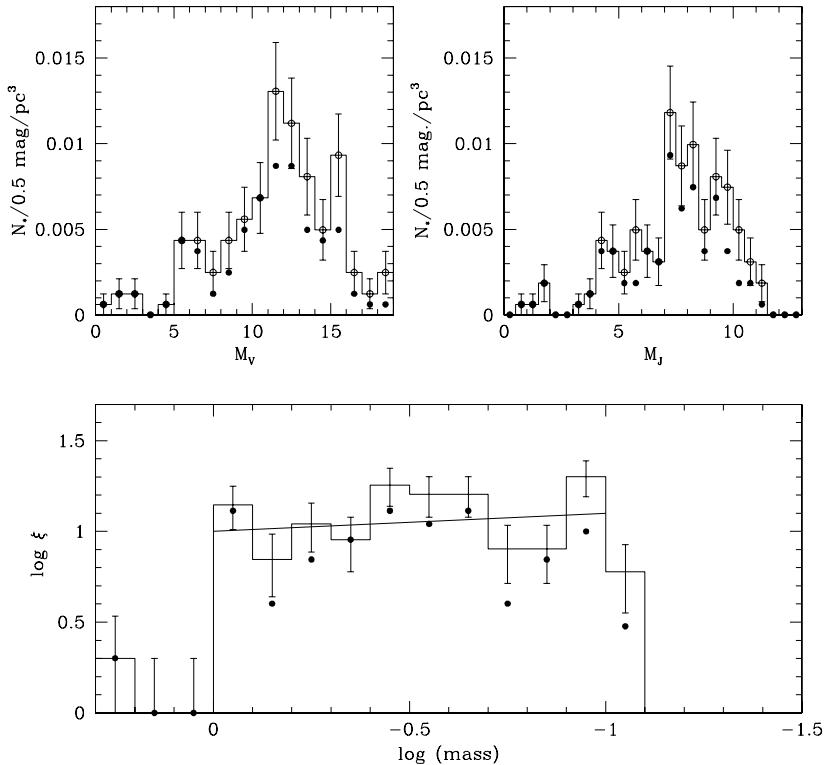


Figure 2.29. $\Phi(M_V)$ and $\Phi(M_J)$ for the northern 8-parsec sample and the corresponding mass function, $\xi(\log M) d \log M$. The solid points in each figure show the contribution from single stars and primaries in binary/multiple systems.

The stellar mass function is derived from the luminosity function data by applying the appropriate mass–luminosity (or mass–absolute magnitude) relationship. Figure 2.29 presents the results for the northern 8-parsec sample, and the systemic mass function (including only single stars and primary stars in multiple systems) is also plotted. Traditionally, the mass function is expressed as a power law, $\Psi(M) = dN/dM \propto M^{-\alpha}$ [S3]. Figure 2.29 shows that $\Psi(M) \propto M^{-1.15}$ gives a reasonable approximation of the star-by-star 8-parsec data for $1 > M/M_\odot > 0.1$; an exponent of $\alpha = 0.9$ matches the systemic function. The inferred total mass density due to main sequence stars is $\sim 0.048 M_\odot \text{ pc}^{-3}$. Chapter 9 provides more extensive discussion of these results.

2.8 SUMMARY

The main aim of this chapter has been to provide an empirical overview of the properties of cool, late-type dwarfs, and to discuss how those properties can be

used to illuminate broader issues such as star formation, nucleosynthesis, stellar magnetic fields, Galactic structure and galaxy formation. Defining M dwarfs as main sequence stars with TiO absorption, their properties can be summarised as follows:

- Their spectra are characterised by molecular absorption bands, notably TiO, VO and metal hydrides at optical wavelengths; H₂O and CO in the near-infrared.
- They have absolute visual magnitudes between $M_V = 7.5$ and $M_V = 20$.
- Their luminosities range from $0.2 L_\odot$ to less than $5 \times 10^{-4} L_\odot$.
- They have increasingly red optical and near-infrared colours with increasing spectral type: $1.5 < (V-I) < 5$ and $1 < (I-K) < 6$.
- Their effective temperatures lie between $\sim 3,800$ K (M0) and $\sim 2,300$ K (M9).
- Their radii lie between $0.6 R_\odot$ and $0.1 R_\odot$.
- They often have active chromospheres and coronae.
- Their masses lie between $0.6 M_\odot$ and $0.1 M_\odot$.

The newly recognised L dwarfs and T dwarfs represent an extension of the M dwarf sequence to cooler temperatures, lower luminosities and lower masses, below the hydrogen-burning limit. TiO and VO become progressively less prominent in late L dwarfs, probably through the formation of dust in the stellar atmosphere. Metal hydride bands and alkali lines dominate the optical spectrum.

Sub-stellar mass brown dwarfs can cool to temperatures below $\sim 1,200$ K, where methane forms in the atmosphere. This leads to substantial absorption in the *H* and *K* passbands, and blue near-infrared colours, as exemplified by the prototype T dwarf, Gl 229B.

In the succeeding chapters, the properties of these cool dwarfs are discussed in detail, giving more attention to the underlying physics required for theoretical modelling. The observations discussed in this chapter, however, represent the primary constraints on those models and the fundamental description of the nature of M dwarfs.

2.9 REFERENCES

- A1 Abt, H., A., Levy, S. G., 1976, *ApJS*, **30**, 273.
- A2 Andersen, J., 1991, *Astr. Ap. Rev.*, **3**, 91.
- A3 Arnaud, K. A., Gilmore, G., Collier Cameron, A., 1989, *MNRAS*, **237**, 495.
- B1 Baldwin, J. R., Frogel, J. A., Persson, S. E., 1973, *ApJ*, **184**, 427.
- B2 Becklin, E. E., Zuckerman, B., 1988, *Nature*, **336**, 656.
- B3 Berriman, G., Reid, I. N., 1987, *MNRAS*, **227**, 315.
- B4 Berriman, G., Reid, I. N., Leggett, S. K., 1992, *ApJ*, **392**, L31.
- B5 Bessell, M. S., 1991, *AJ*, **101**, 662.
- B6 Boeshaar, P., 1976, PhD thesis, Ohio State University.
- B7 Bopp, B. W., Moffett, T. J., 1973, *ApJ*, **185**, 239.
- B8 Burgasser, A. J. *et al.*, 1999, *ApJ*, **522**, L65.

- B9 Burgasser, A. J., McElwain, M. W., Kirkpatrick, J. D., Cruz, K. L., Tinney, C. G., Reid, I. N., 2004, *AJ*, **127**, 2856.
- B10 Burgasser, A. J., Kirkpatrick, J. D., Liebert, J., Burrows, A., 2003, *ApJ*, **594**, 510.
- B11 Burgasser, A. J. *et al.*, 2002, *ApJ*, **564**, 421.
- B12 Bessell, M. S., 1990, *A&AS*, **83**, 357.
- B13 Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., Brown, M. E., Miskey, C. L., Gizis, J. E., 2003, *ApJ*, **586**, 512.
- B14 Bouguer, R., 1954, *Ann. Ap.*, **17**, 104.
- C1 Chamberlain, J. N., Aller, L. H., 1951, *ApJ*, **114**, 52.
- C2 Cox, J. J., Gibson, D. M., 1985, in *Radio Stars*, ed. R. M. Hjellming, D. M. Gibson, Reidel, Dordrecht, p. 233.
- C3 Cruz, K. L., Reid, I. N., Liebert, J., Kirkpatrick, J. D., Lowrance, P. J., 2003, *AJ*, **126**, 2421.
- C4 Cruz, K. L., Reid, I. N., Liebert, J., Kirkpatrick, J. D., Lowrance, P. J., 2005, *AJ*, in press.
- D1 Delfosse, X., Tinney, C. G., Epcstein, N., Bertin, E., Borsenberger, J., Copet, E., de Batz, B., Fouquem, P., Kineswenger, S., Le Bertre, T., Lacombe, F., Rouan, D., Tiphene, D., 1997, *A&A*, **327**, L25.
- D2 Duquennoy, A., Mayor, M., 1991, *A&A*, **248**, 485.
- D3 Dahn, C. C., Liebert, J., Kron, R. G., Spinrad, H., Hintzen, P. M., 1977, *ApJ*, **216**, 757.
- D4 Downes, R. A. *et al.*, 2004, *AJ*, **127**, 2838.
- D5 Delfosse, X., Forveille, T., Segranson, D., Beuzit, J.-L., Udry, S., Perrier, C., Mayor, M., 2000, *A&A*, **364**, 217.
- E1 ESA, 1997, The Hipparcos catalogue, ESA SP-1200.
- F1 Fischer, D. A., Marcy, G. W., 1992, *ApJ*, **396**, 178.
- F2 Frogel, J. A., 1971, PhD thesis, California Inst. of Technology.
- F3 Fowler, A., 1904, *MNRAS*, **64**, 16.
- F4 Fan, X. *et al.*, 2000, *AJ*, **119**, 928.
- F5 Fleming, W. P., 1912, *Harvard Obs. Ann.*, **56**, 2165.
- G1 Ghez, A. M., McCarthy, D. W., Patience, J. L., Beck, T. L., 1997, *ApJ*, **481**, 378.
- G2 Giampapa, M. S., Liebert, J., 1986, *ApJ*, **305**, 784.
- G3 Gizis, J. E., 1996, *AJ*, **113**, 806.
- G4 Gizis, J. E., Monet, D., Reid, I. N., Williams, R., 1999, *AJ*, **118**.
- G5 Gliese, W., 1969, *Catalogue of Nearby Stars*, Veröff. Astr. Rechen-Inst. Heidelberg, Nr. 22.
- G6 Gliese, W., Jahreiss, H., 1991, *Third Catalogue of Nearby Stars*, preliminary version.
- G7 Greenstein, J. L., Neugebauer, G., Becklin, E. E., 1970, *ApJ*, **161**, 519.
- G8 Green, P. J., 2000, in IAU Symp. 177, *The Carbon Star Phenomenon*, ed. R. F. Wing, Dordrecht, Kluwer, p. 37.
- G9 Geballe, T. *et al.*, 2002, *ApJ*, **564**, 466.
- G10 Golimowski, D. A. *et al.*, 2004, *AJ*, **127**, 3516.
- G11 Güdel, M., Schmitt, J. H. M. M., Bookbinder, I. A., Fleming, T. A., 1993, *ApJ*, **415**, 236.
- H1 Hawley, S. L., Gizis, J. E., Reid, I. N., 1996, *AJ*, **112**, 2799.
- H2 Hawley, S. L., Pettersen, B. R., 1991, *ApJ*, **378**, 725.
- H3 Hearnshaw, J. B., 1986, *The Analysis of Starlight*, Cambridge University Press.
- H4 Henry, T. J., McCarthy, D. W., 1993, *AJ*, **106**, 773.
- H5 Hertzsprung, E., 1905, *Zeitschrift für wissenschaftliche Photographie*, **3**, 429.
- H6 Hertzsprung, E., 1907, *Zeitschrift für wissenschaftliche Photographie*, **5**, 86.

- H7 Hertzsprung, E., 1924, *BAN*, **2**, 87.
- H8 Herbig, G. H., 1956, *PASP*, **68**, 531.
- J1 Johns-Krull, C. J., Valenti, J. A., 1996, *ApJ*, **459**, L95.
- J2 Jones, H. R. A., Longmore, A. J., Allard, F., Hauschildt, P. H., 1996, *MNRAS*, **280**, 77.
- J3 Joy, A. H., 1947, *ApJ*, **105**, 96.
- J4 Joy, A. H., Abt, H. A., 1974, *ApJ*, **28**, 14.
- J5 Joy, A. H., Humason, M., 1949, *PASP*, **61**, 133.
- K1 Keenan, P., MacNeil, R. C., 1976, *An Atlas of Spectra of the Cooler Stars: Types G, K, M, S and C*, Columbus: Ohio State University Press.
- K2 Kirkpatrick, J. D., Henry, T. J., McCarthy, D. W., 1991, *ApJS*, **77**, 417.
- K3 Kirkpatrick, J. D., Kelly, D. M., Rieke, G. H., Liebert, J., Allard, F., Wehrse, R., 1993, *ApJ*, **402**, 643.
- K4 Kirkpatrick, J. D., Henry, T. J., Liebert, J., 1993, *ApJ*, **406**, 701.
- K5 Kirkpatrick, J. D., Henry, T. J., Simons, D. A., 1995, *AJ*, **109**, 797.
- K6 Kirkpatrick, J. D., Beichman, C. A., Skrutskie, M. F., 1997, *ApJ*, **476**, 311.
- K7 Kirkpatrick, J. D., 1998, in *Brown Dwarfs and Extrasolar Planets*, ed. R. Rebolo, E. Martín and M. R. Zapatero Osorio, ASP Conf. Ser. No. 134, p. 405.
- K8 Kirkpatrick, J. D., Reid, I. N., Liebert, J., Cutri, R. M., Nelson, B., Beichman, C. A., Dahn, C. C., Monet, D. G., 1999, *ApJ*, **519**, 802.
- K9 Kuiper, G. P., 1939, *ApJ*, **89**, 549.
- K10 Kuiper, G. P., 1942, *ApJ*, **95**, 201.
- K11 Keenan, P. C., Morgan, W. W., 1941, *ApJ*, **94**, 501.
- K12 Kalas, P., Liu, M. C., Matthews, B. C., 2004, *Science*, **303**, 1990.
- K13 Koen, C., Kilkenny, D., van Wyk, F., Cooper, D., Marang, F., 2002, *MNRAS*, **334**, 20.
- K14 Kirkpatrick, J. D., 2001, in *Ultracool Dwarfs*, ed. H. R. A. Jones and I. A. Steele, Springer, p. 139.
- L1 Lacy, C. H., 1977, *ApJ*, **218**, 444.
- L2 Lacy, C. H., Moffett, T. J., Evans, D. S., 1976, *ApJS*, **30**, 85.
- L3 Lim, J., White, S. M., Slee, O. B., 1996, *ApJ*, **460**, 976.
- L4 Lim, J., White, S. M., 1996, *ApJ*, **462**, L91.
- L5 Leggett, S. K., 1992, *ApJS*, **82**, 351.
- L6 Leggett, S. K., Allard, F., Berriman, G., Dahn, C. C., Hauschildt, P., 1996, *ApJS*, **104**, 117.
- L7 Leung, K.-C., Schneider, D. P., 1978, *AJ*, **83**, 618.
- L8 Luyten, W. J., 1949, *ApJ*, **109**, 532.
- L9 Luyten, W. J., 1976, *Proper Motion Survey with the 48-inch Palomar Schmidt Telescope*, University of Minnesota publications.
- L10 Leggett, S. K. *et al.*, 2000, *ApJ*, **536**, L35.
- L11 Lowrance, P. J., Kirkpatrick, J. D., Reid, I. N., Cruz, K. L., Liebert, J., 2003, *ApJ*, **584**, L98.
- L12 Leggett, S. K. *et al.*, 2002, *ApJ*, **564**, 452.
- L13 Lane, B. F., Boden, A. F., Kulkarni, S. R., 2001, *ApJ*, **551**, L81.
- L14 Leggett, S. K., Allard, F., Dahn, C. C., Hauschildt, P. H., Kerr, T. H., Rayner, J., 2000, *ApJ*, **535**, 965.
- L15 Leggett, S. K., Allard, F., Geballe, T. R., Hauschildt, P. H., Schweizer, A., 2001, *ApJ*, **548**, 908.
- M1 van Maanen, A., 1940, *ApJ*, **91**, 503.
- M2 van Maanen, A., 1940, *PASP*, **57**, 216.
- M3 Marcy, G. W., Benitz, K. J., 1989, *ApJ*, **344**, 441.

- M4 Mathioudakis, M., Doyle, J. G., 1993, *A&A*, **280**, 181.
- M5 Metcalfe, T. S., Mathieu, R. D., Latham, D. W., Torres, G., 1996, *ApJ*, **456**, 356.
- M6 Monet, D. G., Dahn, C. C., Vrba, F. J., Harris, H. C., Pier, J. R., Luginbuhl, C. B., Ables, H. D., 1992, *AJ*, **103**, 662.
- M7 Morgan, W. W., 1938, *ApJ*, **87**, 589.
- M8 Morgan, W. W., Keenan, P. C., Kellman, E., 1943, *An Atlas of Representative Stellar Spectra*, Chicago: University of Chicago Press.
- M9 Mould, J. R., Hyland, A. R., 1976, *ApJ*, **208**, 399.
- M10 Mould, J. R., McElroy, D. B., 1978, *ApJ*, **220**, 935.
- M11 Mould, J. R., 1976, *A&A*, **48**, 443.
- M12 Mullan, D. J., Stencel, R. E., Backman, D. E., 1989, *ApJ*, **343**, 400.
- M13 Mullan, D. J., Doyle, J. G., Mathioudakis, M., Redman, R. O., 1992, in *The 7th Cambridge Workshop on Cool Stars*, ed. M. S. Giampapa and J. A. Bookbinder, ASP Conf. Ser. 26, 328.
- M14 Margon, B. *et al.*, 2002, *AJ*, **124**, 1651.
- N1 Nakajima, T., Durrance, S. T., Golimowski, D. A., Kulkarni, S. R., 1994, *ApJ*, **428**, 797.
- O1 Oke, J. B., Cohen, J. G., Carr, M., Cromer, J., Dingizian, A., Harris, F. H., Labreque, S., Lucinio, R., Schaal, W., Epps, H., Miller, J., 1995, *PASP*, **107**, 375.
- O2 Osterbrock, D. E., 1993, *50 Years of the MK Process*, Astr. Soc. Pacif. Conf., **60**, 199.
- P1 Pettersen, B. R., 1987, *Vistas in Astronomy*, **30**, 41.
- P2 Pettersen, B. R., Coleman, L. A., Evans, D. S., 1984, *ApJS*, **54**, 375.
- R1 Reid, I. N., Gilmore, G. F., 1984, *MNRAS*, **206**, 19.
- R2 Reid, I. N., Hawley, S. L., Gizis, J. E., 1995, *AJ*, **110**, 1838.
- R3 Reid, I. N., Gizis, J. E., 1997, *AJ*, **113**, 2246.
- R4 Ruiz, M. T., Leggett, S. K., Allard, F., 1997, *ApJ*, **491**, L107.
- R5 Russell, H. N., 1911, Carnegie Inst. Washington Publ. No. 147.
- R6 Russell, H. N., 1913, *The Observatory*, **36**, 324.
- R7 Reid, I. N., Burgasser, A. J., Cruz, K. L., Kirkpatrick, J. D., Gizis, J. E., 2001, *AJ*, **121**, 1710.
- R8 Reid, I. N., Cruz, K. L., 2002, *AJ*, **126**, 2806.
- R9 Roeilig, T. L. *et al.*, 2004, *ApJS*, **154**, 418.
- R10 Ribas, I., 2003, *A&A*, **398**, 239.
- S1 Saar, S. H., Linsky, J. L., 1985, *ApJ*, **299**, L47.
- S2 Saar, S. H., Linsky, J. L., Beckers, J. M., 1986, *ApJ*, **302**, 777.
- S3 Salpeter, E. E., 1956, *ApJ*, **121**, 161.
- S4 Sandage, A., Eggen, O. J., 1959, *MNRAS*, **119**, 278.
- S5 Schmitt, J. H. M. M., Fleming, T. A., Giampapa, M. S., 1995, *ApJ*, **450**, 392.
- S6 Skrutskie, M. F. *et al.*, 1997, in *The Impact of Large-Scale Near-IR Sky Survey*, ed. F. Garzon *et al.*, Kluwer, Dordrecht, p. 187.
- S7 Schneider, D. P. *et al.*, 2002, *AJ*, **123**, 458.
- S8 Strauss, M. A. *et al.*, 1999, *ApJ*, **522**, L69.
- S9 Secchi, A., 1868, *Comptes Rendus*, **66**, 124.
- S10 Segranson, D., Kervella, P., Forveille, T., Queloz, D., 2003, *A&A*, **397**, L5.
- T1 Tinney, C. G., Reid, I. N., 1999, *MNRAS*, **301**, 1031.
- T2 Tinney, C. G., Mould, J. R., Reid, I. N., 1993, *AJ*, **105**, 1045.
- T3 Trimble, V., 1990, *MNRAS*, **242**, 79.
- T4 Tokunaga, A. T., Kobayashi, N., 1999, *AJ*, **117**, 1010.
- V1 Vaiana, G. S. *et al.*, 1981, *ApJ*, **244**, 163.

- V2 Vyssotsky, A. N., Janssen, E. M., 1951, *AJ*, **56**, 58.
 W1 Waters, L. B. F. M., Coté, J., Aumann, H. H., 1987, *A&A*, **172**, 225.
 W2 Wilting, J., Scheiner, J., 1909, *Publ. Astrophys. Obs. Potsdam*, **19**, 1.
 W3 Wing, R. F., 1973, *Spectral Classification and Multicolour Photometry*, IAU Symposium No. 50, ed. Ch. Fehrenbach and B. E. Westerlund, Reidel, Dordrecht, p. 209.
 W4 West, A. A. *et al.*, 2004, *AJ*, in press.

2.10 HOMEWORK PROBLEMS

Problem 1

For spectral types M0, M3, M6 and M9 estimate the following colours and magnitudes, with uncertainties: $(B-V)$, $(V-I)$, $(V-K)$, $(I-J)$, $(J-K)$, M_V , M_I , M_J , M_K .

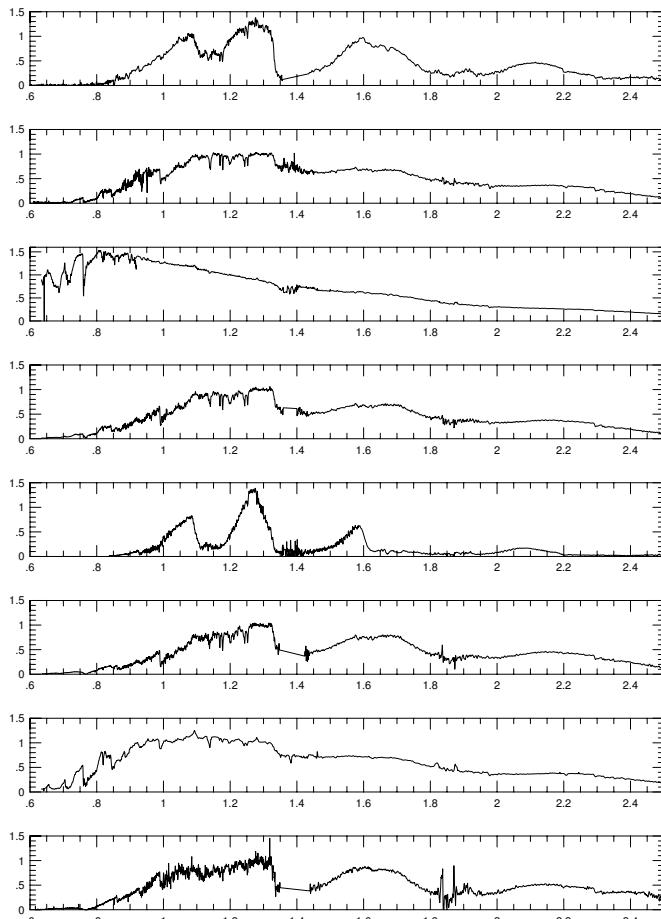


Figure 2.30. Far-red optical and near-infrared spectra of late-type dwarfs: see Problem 3.

Explain which colours are most useful for identifying early-, mid- and late-M spectral types, respectively.

Problem 2

Consider a binary system consisting of a G2 dwarf like the Sun and a T6.5 dwarf like Gliese 229B.

- (a) What are the absolute magnitudes of each component in V and in K ?
- (b) Estimate the minimum separation (in arcsec) and the maximum distance (in parsecs) for which the individual components could be detected using (1) AO on the Keck 10-m telescope; (2) speckle interferometry using the Palomar 200-inch; (3) the Advanced Camera for Surveys (ACS) and the Near-Infrared Camera and MultiObject Spectrograph (NICMOS) on HST; (4) an optical CCD camera and a near-infrared array on a typical ground-based 4-m telescope. In each case, consider observations in the V band and the K band.

Problem 3

List two defining spectral characteristics, and give approximate spectral types (e.g., mid-M, early-L) for each object shown in Figure 2.30.

Problem 4

Write a program to read the 8-pc sample data given in the Appendix, and available on the web at <http://www.stsci.edu/~inr/nldsdat.html>. Use the data to produce figures similar to Figures 2.16 and 2.17. The 8-pc sample will be used throughout the book to illustrate representative examples, so this program will be needed for future exercises.

3

The structure, formation and evolution of low-mass stars and brown dwarfs

3.1 INTRODUCTION

The development of a self-consistent theoretical description of the internal structure of stars, and the consequent construction of models which trace evolutionary behaviour, represents one of the major achievements of twentieth-century astrophysics. Most studies have centred on intermediate- and high-mass stars, with little consideration of M dwarfs with masses below $0.6 M_{\odot}$. In part, this reflects the availability of more precise observational constraints, and also greater analytic tractability in modelling higher-mass stars. Recently, however, the lower main sequence has attracted more attention, with a series of detailed models extending past the hydrogen-burning limit to the boundary between low-mass brown dwarfs and giant planets.

The aim of the present chapter is to concentrate on issues relevant to the structure of low-mass stars. The following section provides an historical overview of the development of stellar structure theory, allowing modern analyses to be placed in the proper context, while the succeeding section outlines the general principles. Thorough discussion of stellar structure theory can be found in textbooks devoted specifically to that subject, including Clayton [C4], Goldberg and Scadron [G3], Kippenhahn and Weigert [K9] and Hansen and Kawaler [H1], while the classic monographs by Chandrasekhar [C3] and Schwarzschild [S7] remain strikingly relevant in many areas.

3.2 A BRIEF HISTORY

The foundations of the modern theory of stellar structure were laid during the later stages of the nineteenth century in a series of analyses of the equilibrium configuration of gaseous spheres. The first such study – by the American

meteorologist J. Homer Lane [L2] – originated largely as a byproduct of an investigation of the surface temperature and density of the Sun. Deriving 54,000°F (30,000 K) for the former,¹ Lane hypothesised that at such temperatures the solar material was ‘torn asunder’ and could be treated as a perfect gas. Based on that assumption, he derived an analytical description of a self-gravitating sphere, and showed (implicitly) that a uniform contraction led to an increase in temperature: Lane’s law – $rT(r) = \text{constant}$. Identical results were derived independently by Ritter [R1], who extended the mathematical treatment in a series of 18 papers published in *Weidemann Annalen* between 1878 and 1889.

Underlying this analysis are two important principles: the balance between the internal pressure and the self-gravity of the Sun, the principle of hydrostatic equilibrium; and the balance between the energy produced and that radiated from the surface, the principle of thermal equilibrium. Based on the principle of energy conservation, Helmholtz [H6] had previously identified gravitational contraction (more specifically, meteoric accretion) as a possible energy source for the Sun. Lane’s analysis codified that process, while invoking Kelvin’s [K1], [K3] mechanism of ‘convective equilibrium’ as a means of energy transport. The latter had been devised in modelling the terrestrial atmosphere. Kelvin [K2] later provided his own analysis of this issue, while the last word on the subject was provided by Emden’s [E2] monograph *Gaskugeln*, which outlines the full series of equations describing gaseous spheres in hydrostatic equilibrium – the Lane–Emden equations.

An important inference which could be drawn from Lane’s law was that stars might be expected to start their lives as diffuse, low-temperature objects which increase in temperature as they evolve and contract. Should the perfect gas law become invalid at some point (as was expected to happen at high densities, for example), then Lane’s law would fail, and the temperature would decrease with time. This proposition ran contrary to the then-prevailing opinion that stars initially had high temperatures and cooled progressively with time (hence the reference to ‘early’ and ‘late’ spectral types in Pickering’s Harvard classification system) and was largely ignored (except by J. Norman Lockyer). However, the discovery of the shape of the stellar distribution with temperature, as revealed by the Hertzsprung–Russell diagram (1911–1913) appeared to provide observational support for this hypothesis, with stars on the giant branch representing the ‘ascending series’ or contraction phase, while the main sequence FGKM dwarfs were interpreted as the ‘descending’ cooling stage. It is interesting to note that this evolutionary model, although treated with some reservations, was still given serious consideration by Eddington [E1].

The gravitational contraction hypothesis also profided a means of estimating the solar lifetime by comparing the present-day luminosity with the available potential

¹ As Chandrasekhar ([C3], p. 176) has pointed out, Lane’s analysis predates the derivation of Stefan’s law (1879), and is based on a substantial extrapolation of the available observational results on emissivity as a function of temperature. Lane’s method was not at fault.

energy. This is the Kelvin–Helmholtz timescale:

$$t_{KH} = \frac{E_{PE}}{\bar{L}_\odot} \approx \frac{GM_\odot^2}{R_\odot \bar{L}_\odot} \sim 3 \times 10^7 \text{ yr} \quad (3.1)$$

where \bar{L}_\odot is the average solar luminosity. The initial applications of this analysis by Helmholtz and Kelvin implied solar lifetimes of less than 100 Myr – at odds with the ‘deep time’ required by Hutton and Lyell’s geological studies (‘...no vestige of a beginning – no prospect of an end’, [H9]). An alternative energy source was required, and Eddington suggested that such might be found in the ‘knockabout comedy of atomic physics’, where ‘subatomic energy of some kind is liberated within the star, so as to replenish the store of radiant energy’ [E1]. These suspicions were confirmed in the 1930s with the development of hydrogen fusion theory by Atkinson, von Weizsäcker, Gamow, Bethe and others.

Eddington’s own calculations, summarised in *The Internal Constitution of the Stars*, placed emphasis on radiative, rather than convective, energy transport. Originally described by Sampson [S1] and further refined by Schwarzschild [S6], radiative processes are of prime importance in stars with masses exceeding that of the Sun. Based on his models, Eddington derived a mass–luminosity relationship and a critical mass – the Eddington limit, where the energy generated in radiative pressure is sufficient to overcome self-gravity and disrupt the star.

Eddington’s models followed the standard (at that time) assumption that iron was the most common element, leading to a mean molecular weight of $\bar{\mu} \sim 2.2$, where $\bar{\mu}$ is defined as the mean mass per particle measured in units of the mass of a hydrogen nucleus (proton), m_H . Part of Eddington’s enthusiasm for this composition lay in the consequent balance between radiative pressure and gas pressure as a function of mass, allowing a ‘cloud-bound physicist’ to predict the properties of the stars ([E1], pp. 15–16). Payne [P1] had earlier demonstrated the uniformity of chemical composition from star to star, hinting at the dominance of H and He. By the end of the 1920s, Russell [R2] had shown conclusively that hydrogen was the most abundant element in the Sun, with $\log(N_H/N_{Fe}) \sim 4.3$. Strömgren [S15], [S16] subsequently developed a set of models which showed that the H–R diagram could be represented as a series of stars of varying mass in hydrostatic and thermal equilibrium, subject to radiative energy transport.

The next major advance in understanding stellar structure came with the description of the properties of white dwarf stars in terms of degeneracy theory. Adams’ [A4] spectroscopic observations of the companion of Sirius had produced the unexpected result that its spectral type was close to that of Sirius A, indicating a comparable temperature and a radius close to that of the Earth. Given a mass of $\sim 1 M_\odot$ (confirmed by gravitational redshift measurements, [A5], but see [G4]), the implied average density is close to 1 ton per cubic inch. Fowler [F1] used Fermi–Dirac statistics to provide an explanation by applying the Pauli exclusion principle to an electron gas (that is, electron degeneracy). Chandrasekhar [C2], [C3] fully developed this area of astrophysics, extending the analysis to relativistic degeneracy, famously contested by Eddington. The latter aspect of the theory was applied observationally only with the discovery of pulsars in the late 1960s.

During the 1940s and 1950s, considerable progress was made in many areas, notably the development of nucleosynthesis theory [B9] and the calculation of evolutionary, rather than static, stellar models [T1], [S7], the latter calculated painstakingly step-by-step on hand-operated Brunswegger calculators. Schwarzschild's work in particular led to what are amongst the first models of red giant stars, characterised as inhomogeneous gaseous spheres, in contrast to the homogeneous main sequence models. Böhm-Vitense [B8] devised the mixing-length approximation as a means of simulating convection in cooler main sequence stars and red giants. The continued improvement in accuracy of the input physical parameters (nuclear reaction rates, cross-sections, opacities), and the recent availability of increasingly more powerful computational power has led to the development of more complex and more realistic models (such as in [D2], [D3], [M1], [V2], [B2], [B10]).

3.3 GENERAL PRINCIPLES OF STELLAR STRUCTURE

3.3.1 The fundamental equations

The foundations of stellar structure theory rest on four fundamental equations [C3], [S7]. The first three equations are straightforward in character. The equation of continuity of mass describes the mass distribution for a spherical object:

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \quad (3.2)$$

where $M(r)$ is the mass distribution, and $\rho(r)$ is the density distribution. Second is the equation of hydrostatic equilibrium, which gives the balance between the attractive force of self-gravity and internal pressure support:

$$\frac{dP(r)}{dr} = -\frac{GM(r)}{r^2} \rho(r) = -g(r) \rho(r) \quad (3.3)$$

where $P(r)$ is the pressure distribution, and $g(r)$ is the gravitational acceleration. Third, the equation of thermal equilibrium gives the balance between energy generation and energy loss through radiation:

$$\frac{dL(r)}{dr} = 4\pi r^2 \rho(r) \varepsilon(r) \quad (3.4)$$

where $L(r)$ is the luminosity, and $\varepsilon(r)$ is the rate of energy generation.

The fourth equation describes the temperature distribution as a function of radius, and is therefore related to the mode of internal energy transport. There are three mechanisms: conduction, convection and radiation. Conduction plays no significant role in the interiors of low-mass main sequence stars and brown dwarfs, although it is important in white dwarfs and neutron stars. Radiation transport contributes significantly under some circumstances, and convection is by far the most important mechanism. The radiative and convective

transport equations are:

$$\begin{aligned}\frac{dT(r)}{dr} &= \frac{-3}{4ac} \frac{\kappa}{T^3} F(r) && \text{radiative transport} \\ &= \left(1 - \frac{1}{\gamma}\right) \frac{T}{P} \frac{dP}{dr} && \text{convective transport}\end{aligned}\quad (3.5)$$

where in the first equation, a is the radiation constant, c is the velocity of light, κ is the opacity and $F(r)$ is the flux as a function of radius, r : $F(r) = L(r)/4\pi r^2$; and in the second equation, γ is the ratio of specific heats C_p/C_v . Section 3.4 provides more details on energy transport.

In addition to these four fundamental equations, stellar modelling requires relations defining three parameters: pressure, energy generation and opacity. These are the three constitutive equations: the equation of state, which describes the pressure,

$$P = P(T, \rho, \text{composition}) \quad (3.6)$$

the equation of energy generation,

$$\varepsilon = \varepsilon(T, \rho, \text{composition}) \quad (3.7)$$

and the opacity equation,

$$\kappa = \kappa(T, \rho, \text{composition}) \quad (3.8)$$

The first two relationships are described in more detail below, and the various sources contributing to stellar opacities are outlined in Chapter 4.

Finally, these equations can only be solved given the appropriate boundary conditions. In the classical case, the four conditions are:

$$\begin{aligned}M(r) &= 0 \text{ at } r = 0 & P(r) &= 0 \text{ at } r = R \\ L(r) &= 0 \text{ at } r = 0 & T(r) &= 0 \text{ at } r = R\end{aligned}$$

where R is the radius of the star. With the development of more accurate atmosphere models (see Chapter 4), the luminosity and temperature distributions predicted by those models can be used to set the exterior boundary conditions.

3.3.2 Gas polytropes and the Lane–Emden equations

Lane's original analysis of the surface temperature of the Sun was based on the analytical approximation that pressure could be described as a function of only density, $P = P(\rho)$. This approach – generalised by Emden and codified by Chandrasekhar – still offers an insight into the internal structure of low-mass stars and brown dwarfs during evolutionary phases where the pressure distribution takes the polytropic form,

$$P = K\rho^\gamma, \gamma = 1 + \frac{1}{n} \quad (3.9)$$

where n is the polytropic index and K is a constant. The resulting solutions are models of gas polytropes.

The Lane–Emden equations are dimensionless forms of the equation of hydrostatic equilibrium. Eliminating $M(r)$, equation (3.3) can be re-written as

$$\frac{1}{r^2} \frac{d}{dr} \left(\frac{r^2}{\rho(r)} \frac{dP(r)}{dr} \right) = -4\pi G \rho(r) \quad (3.10)$$

Substituting for $P(r)$ using equation (3.9) gives

$$\frac{1}{r^2} \frac{d}{dr} \left(\frac{r^2}{\rho} \gamma K \rho^{\gamma-1} \frac{d\rho}{dr} \right) = -4\pi G \rho \quad (3.11)$$

or, defining

$$\phi = \frac{\gamma}{\gamma-1} K \rho^{\gamma-1} \quad (3.12)$$

we have

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\phi}{dr} \right) = -4\pi G \rho \quad (3.13)$$

ϕ is related to the gravitational potential, $\Phi(r)$:

$$\Phi(r) = \frac{GM(r)}{r} \quad (3.14)$$

as follows. Substituting for $M(r)$ in equation (3.14) gives Poisson's equation in spherically symmetric co-ordinates:

$$\nabla^2 \Phi(r) = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\Phi}{dr} \right) = -4\pi G \rho \quad (3.15)$$

Comparison of equations (3.13) and (3.15) shows that

$$\phi(r) = -[\Phi(r) - \Phi(R)] \quad (3.16)$$

since ρ , P and ϕ all vanish at the stellar surface, $r = R$.

Equation (3.12) can be rewritten as

$$\rho = \left(\frac{\gamma-1}{\gamma} \frac{\phi}{K} \right)^{\frac{1}{\gamma-1}} = \left[\frac{\phi}{(n+1)K} \right]^n \quad (3.17)$$

so substituting for ρ in equation (3.10) gives

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\phi}{dr} \right) + 4\pi G \left[\frac{\phi}{(n+1)K} \right]^n = 0 \quad (3.18)$$

Defining ϕ_0 as ϕ at $r = 0$, we introduce the scaled variable

$$\theta = \frac{\phi}{\phi_0} \quad (3.19)$$

and the scaled radius

$$\xi = \frac{r}{a} \text{ where } a = \left[\frac{4\pi G \phi_0^{n-1}}{((n+1)K)^n} \right]^{-0.5} \quad (3.20)$$

Rewriting equation (3.18) using these variables gives the Lane–Emden equations,

$$\frac{1}{\xi^2} \frac{d}{d\xi} \left(\xi^2 \frac{d\theta}{d\theta} \right) + \theta^n(\xi) = 0 \quad (3.21)$$

where the boundary conditions are

$$\theta(0) = 1 \quad \text{and} \quad \left(\frac{d\theta}{d\xi} \right)_{\xi=0} = 0 \quad (3.22)$$

Cast in this manner, the density, pressure and temperature distributions all scale with θ . Thus, from equation (3.17)

$$\rho = \left[\frac{\phi_0}{(n+1)K} \right]^n \theta^n = \rho_0 \theta^n \quad (3.23)$$

where ρ_0 is the central density. Similarly, from equation (3.9)

$$P = K \rho^{\frac{n+1}{n}} = \frac{1}{K^n} \left(\frac{\phi_0}{(n+1)} \right)^{\frac{n+1}{n}} \theta^{n+1} = P_0 \theta^{n+1} \quad (3.24)$$

Finally, given the perfect gas equation of state, $P_g = k\rho T/\bar{\mu}$, where k is Boltzmann's constant and $\bar{\mu}$ is the mean molecular weight, then

$$T = \left(\frac{\beta \bar{\mu}}{k} \right) \frac{P}{\rho} = \left(\frac{\beta \bar{\mu}}{k} \right) \left(\frac{\phi_0}{n+1} \right) \theta = T_0 \theta \quad (3.25)$$

with $\beta \equiv P_g/P$ and T_0 is the central temperature.

The location of the first zero in the Lane–Emden function, $\theta(\xi_1) = 0$, corresponds to the surface of the star, so ξ_1 is the scaled stellar radius

$$R = a \xi_1 \quad (3.26)$$

The total mass is given by

$$M = 4\pi a^3 \rho_0 \left[-\xi^2 \frac{d\theta}{d\xi} \right]_{\xi=\xi_1} \quad (3.27)$$

and the gravitational energy of a polytrope is

$$\Omega = - \frac{3}{5-n} \frac{GM^2}{R} \quad (3.28)$$

The polytropic approximation can be used to derive analytical models for low-mass dwarfs, as demonstrated by Burrows and Liebert [B11]. For an $n = 1.5$ polytrope (full convection or non-relativistic degeneracy), they derive

$$\rho_c \sim 5.99 \left[\frac{3M}{4\pi r^2} \right] \propto M^2 \quad (3.29)$$

$$P_c \sim 0.77 \frac{GM^2}{R} \propto M^{10/3} \quad (3.30)$$

$$R \sim 2.357 \left[\frac{K}{G} \right] M^{-1/3} \quad (3.31)$$

where ρ_c and P_c are the central density and pressure, and R is the radius. Zapolsky and Salpeter [Z1] provide the following analytical representation of the mass–radius relationship for very low-mass dwarfs:

$$R = 0.016 \left[\frac{M_\odot}{M} \right]^{\frac{1}{3}} \Big/ \left[1 + \left(\frac{M}{0.0032 M_\odot} \right)^{-0.5} \right]^{\frac{4}{3}} R_\odot \quad (3.32)$$

Extending these calculations to include evolutionary effects, [B11] gives the following relationships:

$$T_{\text{eff}} \approx 27,700 \tau^{-0.324} M^{0.827} \kappa_R^{0.088} \text{ Kelvin} \quad (3.33)$$

$$L \approx 1.74 \times 10^{-6} \tau^{-1.297} M^{2.04} \kappa_R^{0.35} L_\odot \quad (3.34)$$

where τ is the age in Gyr, M is the mass in M_\odot , and κ_R is the Rosseland mean opacity (see Chapter 4). These relationships make no allowance for complications such as deuterium burning or the strong wavelength dependence of atmospheric opacities in low-temperature atmospheres, but provide order of magnitude estimates for fully convective low-mass dwarfs ($M < 0.25 M_\odot$).

3.3.3 Energy generation

Stars and brown dwarfs form through gravitationally-induced collapse of dense gas within molecular cloud cores – a process described in more detail towards the end of this chapter. During the formative phase and the initial stages of evolution, radiant energy is generated through the transformation of potential energy into heat. The Kelvin–Helmholtz contraction timescale (equation (3.1)) for an M dwarf is substantially longer than for the Sun: for example, CM Draconis A, the more massive component of the well-known eclipsing binary, has $M = 0.237 M_\odot$, $R = 0.252 R_\odot$ and $L = 0.005 L_\odot$, giving $\tau_{KH} \sim 1.3$ Gyr. However, these longer lifetimes are still only a small fraction of the Hubble time, the age of the Universe, estimated as 12–14 billion years (Chapter 7).

The primary energy source for M dwarfs, as with other main sequence stars, is nuclear fusion. As the star contracts, the central temperature and density increase and, when the core temperature exceeds $T_c \sim 3 \times 10^6$ K (the exact value is density dependent), hydrogen fusion begins. Not all objects are sufficiently massive (have sufficient potential energy of collapse) to raise T_c above this threshold, and for those lower-mass objects – brown dwarfs – τ_{KH} is the appropriate luminous timescale. While this fact is mentioned in passing in early structural analyses of low-mass stars (see [L6], for example), the full consequences were emphasised in Kumar’s [K7] explicit calculation of the hydrogen-burning limit (see Section 3.4.2).

The net result of hydrogen fusion is the conversion of four ${}^1\text{H}$ atoms to one ${}^4\text{He}$ atom. This conversion occurs through two series of reactions: the proton–proton chain (P–P) and the CN (or CNO) cycle. The total energy, ε , generated by both processes is a function of core temperature (Figure 3.1). Interacting particles must overcome their respective Coulomb barriers, hence the rate of interaction increases

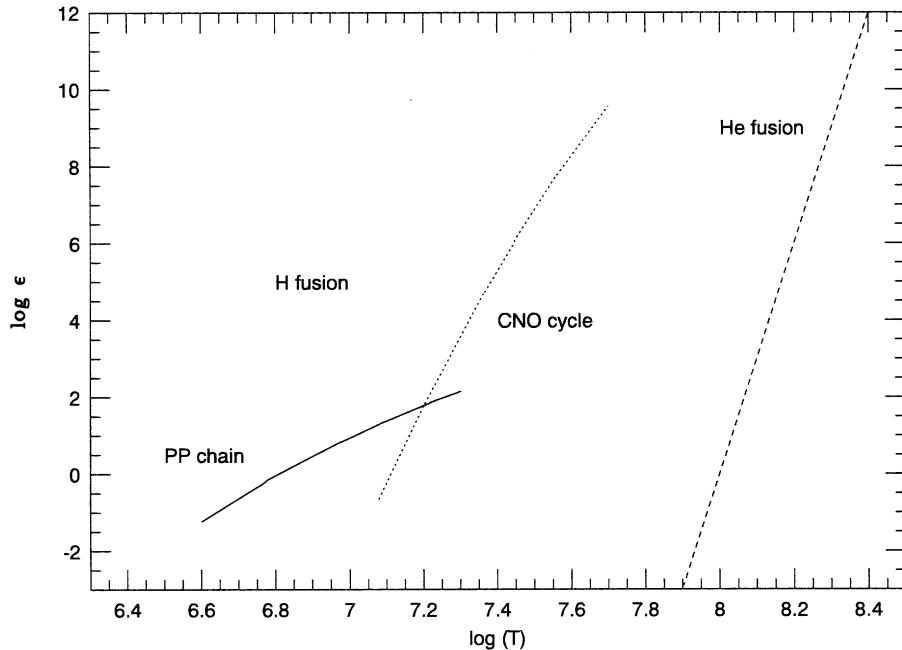
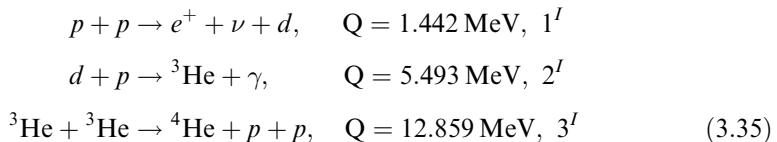


Figure 3.1. The relative energy-generation rates of the P–P chain, the CNO cycle and triple- α helium fusion. The calibrations are from [S7].

with increasing kinetic energy (kT). In the case of the Sun, both reaction series are present, although the CN cycle contributes less than 10% of the total energy.

Thorough discussions of the physics of stellar nucleosynthesis are given by Clayton [C4] and Bahcall [B1]. Here we summarise only those factors relevant to low-mass dwarfs. Amongst those stars, the P–P chain is the overwhelmingly dominant source of energy. The reactions in the basic chain are as follows:



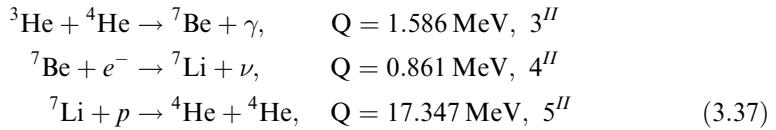
where p denotes a proton (hydrogen nucleus); d , a deuterium nucleus; e^+ , a positron; ν , a neutrino; and γ , a photon. Q is the energy released by each reaction (the neutrino in reaction 1^I carries ≈ 0.262 MeV). An alternative to the first step is the *pep* reaction



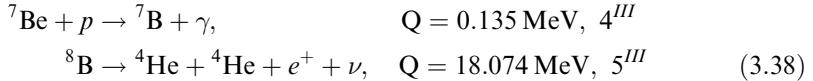
where e^- is an electron. This reaction is of secondary importance.

The P–P chain can be extended to reactions involving heavier elements if the temperature is sufficiently high (that is, in higher-mass M dwarfs). There are two subsequent reaction series: the PP II chain, starting with the products of

reaction 2^I



and the PP III chain, which follows from reaction 3^{II}



For these reactions to occur, particles must overcome the Coulomb potential barrier surrounding each nucleus, given by ([C4], Section 1.5)

$$V = \frac{Z_1 Z_2 e^2}{R} = 1.44 \frac{Z_1 Z_2}{R} \text{ MeV} \quad (3.39)$$

where Z_1 and Z_2 are the atomic numbers of the two nuclei; e is the charge on an electron, and R is the separation in femtometres (10^{-15} m). Interactions occur only when $R \lesssim 2\text{--}3$ fm, requiring energies of $E > 0.5$ MeV. Particles in the stellar core follow a Maxwell–Boltzmann velocity distribution. The most probable energy of a particle is $kT/2$, and the average energy per particle is $\frac{3}{2}kT$, where

$$kT \sim 0.862 \times 10^{-7} T \text{ keV} \quad (3.40)$$

Since the central temperature of M dwarfs never exceeds 10^7 K, it is clear that quantum tunnelling effects are vital in overcoming Coulomb repulsion and initiating fusion. The efficiency of that process varies as $e^{-b/\sqrt{E}}$, where E is the energy of the particle and b is a constant.² Nuclear reaction rates are determined by the convolution of the number of particles in the high energy tail of the Maxwell–Boltzmann distribution, $N \propto e^{-E/kT}$, and the tunnelling efficiency. The convolved function peaks sharply (the Gamow peak) at

$$E_0 = 1.22(Z_1^2 Z_2^2 A T_6^2)^{\frac{1}{3}} \text{ keV} \quad (3.41)$$

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

Typical values for E_0 range from 10 to 30 keV for reactions in the P–P chain. As T_c decreases with decreasing mass, the reaction rates for the PP II and PP III reactions also decrease and fail to achieve equilibrium. Indeed, for central temperatures below 8×10^6 K, reaction 3^I requires more than 10^9 years to reach equilibrium. At masses below $\sim 0.25 M_\odot$, temperatures are too low to allow significant production of ${}^4\text{He}$, and the P–P chain effectively terminates at reaction 2^I [B11].

When an M dwarf is on the main sequence, deuterium is produced through reaction 1^I and subsequently burned in reaction 2^I, while lithium is produced through reaction 4^{II} and burned in reaction 5^{II}. However, deuterium and lithium are both present at their interstellar abundances in a newly-formed star. Deuterium,

² $b = 31.28 Z_1 Z_2 \sqrt{A}$, where A is the reduced atomic weight, $A = A_1 A_2 / (A_1 + A_2)$, [C4].

with a primordial ratio $D/H \sim 2 \times 10^{-5}$ [T2] can be transformed to ^3He through reaction 2 $'$ when the central temperature exceeds $\approx 6 \times 10^5$ K. This process, originally highlighted by Grossman [G5], leads to a pause in the luminosity evolution of dwarfs with masses exceeding $\sim 0.012 M_\odot$, with the luminosity remaining approximately constant for between a few million years and 10^5 years due to the additional energy source. This stage has been referred to as the deuterium-burning main sequence.

Primordial lithium is destroyed through reaction 5 $''$. However, the threshold temperature for initiating that reaction is relatively high: $\sim 2 \times 10^6$ K. As a result, objects with masses below $\sim 0.06 M_\odot$ are predicted to retain an undepleted lithium fraction. This has become an important observational technique for verifying the nature of low-temperature brown dwarf candidates ([M2]; see Chapter 6).

3.3.4 Energy transport

Energy generation occurs within the stellar core, and this energy must be transported to the surface to be released as radiation. As outlined in Section 3.1, convection is the dominant process in defining the internal structure of M dwarfs, with radiation playing a lesser, but still important, role. The criterion for stability against convection was stated originally by K. Schwarzschild: if a volume element is displaced in the presence of a temperature gradient with no change in the heat content (adiabatically), then the element expands or contracts to match the change in pressure. For an adiabatic expansion,

$$P = K\rho^\gamma \quad (3.42)$$

where $\gamma = C_P/C_V$, the ratio of specific heats at constant pressure and constant volume. The condition for stability (a restoring force on the displaced volume element) is

$$\left(\frac{P + dP}{K}\right)^{\frac{1}{\gamma}} > \rho + d\rho \quad (3.43)$$

or

$$\frac{1}{\gamma} \frac{1}{P} \frac{dP}{dr} > \frac{1}{\rho} \frac{d\rho}{dr} \quad (3.44)$$

If we assume that the perfect gas law is valid,

$$P = NkT = \frac{k\rho T}{\bar{\mu}} \quad (3.45)$$

with N the number of particles per unit volume, the stability criterion can then be rewritten as

$$\frac{1}{\gamma} \frac{1}{P} \frac{dP}{dr} > \frac{1}{P} \frac{dT}{dr} - \frac{1}{T} \frac{dT}{dr} \quad (3.46)$$

where dT/dr is the temperature gradient. Hence we can define the

superadiabatic temperature gradient:

$$\Delta \nabla T = \left(1 - \frac{1}{\gamma}\right) \frac{T}{P} \frac{dP}{dr} - \frac{dT}{dr} \quad (3.47)$$

The system is stable against convection for $\Delta \nabla T < 0$. Finally, since

$$C_P - C_V = \frac{Ak}{\bar{\mu}} \quad \text{and} \quad \frac{dP}{dr} = -g(r)\rho \quad (3.48)$$

the convective inequality can be rewritten as

$$\left| \frac{dT}{dr} \right| \propto F\kappa < \frac{g}{C_P} \text{ for stability} \quad (3.49)$$

where F is the flux and κ is the opacity. Thus, a system is unstable against convection if the flux level is high (as in the central regions of intermediate- and high-mass stars); if gravity is low (as in red giant envelopes); if the opacity is high; and if C_P is high, which is the case where abundant elements are undergoing ionisation. The last two circumstances hold to varying extent for M dwarfs and brown dwarfs.

Early-type M dwarfs have convective envelopes and radiative central regions, often referred to as radiative cores.³ Approximately 90% of the total mass resides in the radiative core of a $0.55 M_{\odot}$ M0 star, with the fraction dropping to $\sim 70\%$ by $0.4 M_{\odot}$, or spectral type M2/M3 [D5]. At $\sim 0.25 M_{\odot}$, spectral type \approx M4, the star becomes fully convective, and analyses suggest that the same circumstances prevail for brown dwarfs down to temperatures of 1,500 K [B13].

Convection is generally modelled using the mixing-length approximation [B8], where each mass element participating in the convective motion is considered to rise or fall adiabatically over a distance l before achieving thermal equilibrium. The mixing length, l , is expressed in terms of the pressure scale height, H , with $\alpha = l/H$ typically taken as between 0.5 and 2. Describing this procedure as a theory of convection is something of an exaggeration, but the mixing-length approximation has been used almost exclusively in stellar modelling over the last 40 years. Canuto *et al.* [C1] have recently introduced a different approach, modelling convection using turbulent diffusion, but the technique has yet to be applied to M-dwarfs.

Finally, while convection is responsible for the bulk of energy transport in low-mass dwarfs, it is important to remember that every object has superimposed a thin radiative envelope – the atmosphere. The position of the boundary between convective and radiative transport is strongly dependent on wavelength-dependent effects tied to the complex, mainly molecular, opacities present in the atmosphere. The atmosphere determines how efficiently energy is released by the star, and therefore acts as a feedback mechanism, influencing the internal structure. Thus, uncertainties in atmospheric physics permeate inwards to produce corresponding uncertainties in the internal structure of these late-type dwarfs. These issues are described in more detail in Chapter 4.

³ The radiative core should not be confused with the nuclear core, where fusion reactions are underway. The latter occupies only the central few percent of the radiative core.

3.3.5 The equation of state

The equation of state (EOS) describes the relationship between pressure, temperature and density within the stellar model, and therefore plays a pivotal role in determining both the internal structure and the consequent predicted surface properties. The Lane–Emden equations allow the full structure to be derived if the system can be modelled as a gas polytrope. The polytropic form for the EOS is valid under certain circumstances: in convection zones, where the temperature gradient is adiabatic ($n = 1.5$); under conditions of full degeneracy ($n = 1.0$ or 1.5); and in an isothermal gas ($n = \infty$).

Schwarzschild [S7] demonstrated that the $(\log T, \log \rho)$ plane can be separated into a number of régimes depending on the dominant contributor to the EOS (Figure 3.2). At high temperatures and low densities, radiation makes a substantial contribution to the total pressure. Under these circumstances, the gas pressure contribution can be modelled using the perfect gas law, so

$$P_{tot} = P_{rad} + P_g = \frac{1}{3}aT^4 + NkT \quad (3.50)$$

Electrons have half-integral spin, and are therefore governed by Fermi–Dirac statistics. As the density increases at a given temperature, a larger fraction of the available electron states are occupied, and the Pauli exclusion principle comes into play. Once all electronic states are occupied, the gas is fully degenerate. At non-relativistic velocities the electron pressure is defined by

$$P_e = \frac{8\pi}{15mh^3} p_0^5 \quad (3.51)$$

and by

$$P_e = \frac{2\pi c}{3h^3} p_0^4 \quad (3.52)$$

at relativistic velocities, where h is Planck's constant and p_0 is the maximum momentum. In the non-relativistic case, this corresponds to

$$P_e = K_1 \left(\frac{\rho}{\bar{\mu}_e} \right)^{\frac{5}{3}} \quad (3.53)$$

while the relativistic case gives

$$P_e = K_2 \left(\frac{\rho}{\bar{\mu}_e} \right)^{\frac{4}{3}} \quad (3.54)$$

where K_1 and K_2 are constants. Thus, the EOS for non-relativistic degeneracy is equivalent to an $n = 1.5$ polytrope, while a relativistically degenerate star can be modelled as an $n = 1.0$ polytrope.

The divisions in the $(\log T, \log \rho)$ plane between regions where the dominant contribution to the EOS stems from radiation pressure, gas pressure and degeneracy, respectively, are outlined in Figure 3.2. Superimposed on the diagram are radial

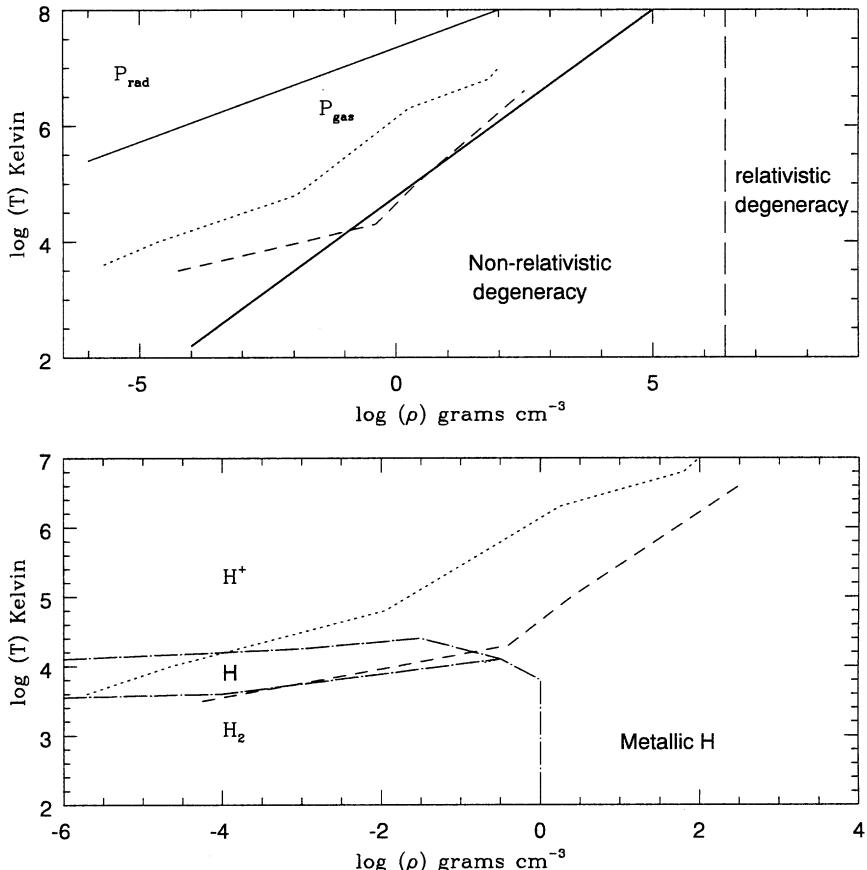


Figure 3.2. The (density, temperature) plane. The solid lines separate the régimes where radiation pressure and degeneracy are dominant; the lower diagram identifies regions where hydrogen is ionised, neutral, molecular or metallic. Both diagrams show the predicted radial density/temperature relationship for 0.5 and $0.085 M_\odot$ stars (dotted and dashed lines respectively, models from [G8]).

distributions predicted for 0.5 and $0.085 M_\odot$ dwarfs; that is M0 and M8/M9 stars. These models show that neither radiation pressure nor relativistic degeneracy are important in determining the structure of M dwarfs. However, non-relativistic electron degeneracy is significant in defining the equation of state in lower-mass objects, influencing in particular the location of the hydrogen-burning mass limit and the mass-radius relation (see Sections 3.4.2 and 3.5.3). Degeneracy becomes increasingly important with decreasing mass.

Figure 3.2 shows that fully-convective M dwarfs can be modelled with reasonable accuracy using simple polytropic approximations to the EOS and, as described in the following section, $n = 1.5$ polytropes have been applied to this purpose. However, over most of the mass range spanned by these objects, corrections must

be introduced to allow for a variety of complicating factors. Partial degeneracy affects the innermost radii at low masses well before the star becomes fully degenerate; Coulomb interactions occur due to the presence of free charges (ions and electrons) throughout the star; there are perturbations due to the close approach of outer electron shells in ions, atoms and molecules; pressure ionisation occurs – a non-ideal gas ionises at temperatures lower than expected for an ideal gas; and the surface boundary conditions (notably the atmospheric opacities) affect the interior temperature and density distributions. All of these effects must be taken into account in computing the EOS in models of low-mass dwarfs. A brief history of the theoretical treatment of these issues is given by Saumon *et al.* [S5], who also present a thorough discussion of the physics underlying the most recent developments. The equations of state for hydrogen and helium derived in that paper are used in the recent models of low-mass stars, brown dwarfs and giant planets. The following section reviews the attributes of those models, concentrating on the predicted observable properties – notably, luminosities and temperatures.

3.4 MODELS OF LOW-MASS DWARFS

3.4.1 The development of low-mass models

The complex physical nature of M dwarfs coupled with their low intrinsic luminosities, and the consequent scarcity of high-quality observations, limited the extent of analysis as the underlying theory of stellar structure was being developed. Until the mid-1970s, observational constraints in the (M_{bol} , T_{eff} , mass) planes were set by a mere handful of binaries: YY Gem (Gl 278C, or Castor C), Kruger 60AB (Gl 860AB), L726-8AB (Gl 65AB) and, to a lesser extent, VB 10 (Gl 752B). CM Draconis (Gl 630.1A) was added to the mix only in the mid-1970s [L1]. Nonetheless, there were a small number of far-reaching studies undertaken during the middle years of the twentieth century.

Early structural analyses, such as [O2], modelled low-mass stars as a radiative core surrounded by a convective envelope, but that treatment led to significant discrepancies between theoretical predictions and observations of late-type ($>M4$) dwarfs. Limber [L6] was the first to demonstrate that these discrepancies are resolved if the latter stars are fully convective: the shallower adiabatic temperature gradients lead to lower inferred T_c and M_{bol} for a given observed T_{eff} . In addition, Limber drew attention to two other important issues: the onset of electron degeneracy in the core, and the effects of H₂ formation and dissociation on the radial temperature gradient. As discussed further in the following section, the former process was first elaborated by Kumar [K7], [K8], who demonstrated that a minimum mass is required to achieve central temperatures sufficiently high for sustained hydrogen burning. Copeland *et al.* [C6] produced the first set of models which explicitly show the effects of the H₂ dissociation layer (Figure 3.3). The adiabatic gradient is reduced, leading to higher pressures and a reduced temperature gradient. Consequently, both luminosity and effective temperature are increased (although

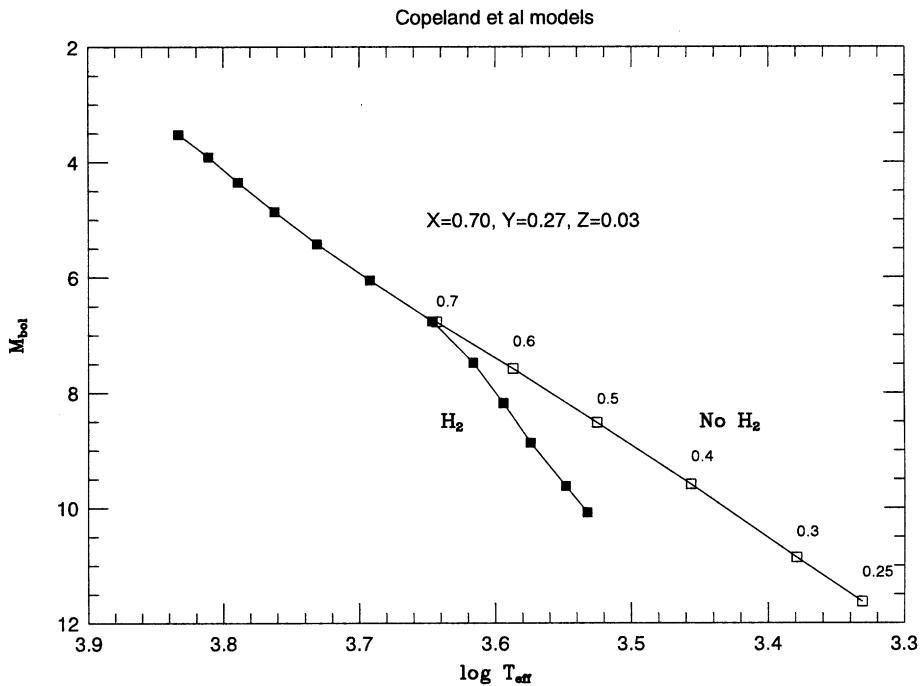


Figure 3.3. The effect of dissociation of the H_2 molecule on the form of the lower main sequence. The numbers 0.7–0.25 refer to the mass of the model. The models are from Copeland *et al.* [C6].

the radius is unchanged) as compared with models where H_2 is omitted. This results in a change of slope in the main sequence in the H–R diagram and in the mass–luminosity relationship. Both effects are observed – for example, see the steepening in slope at $M_V \sim 8.5$ in Figures 2.14 and 2.16.

The models of Copeland *et al.* were limited to masses exceeding $0.25 M_\odot$. More extensive calculations were undertaken by Grossman, Grabske and collaborators, aimed explicitly at modelling the structure of VLM dwarfs above and below the hydrogen-burning limit [G5], [G6], [G7]. The final version of these models [G8] incorporated a number of significant innovations, including allowing for fusion of primordial deuterium as an energy source, using non-grey atmospheres to define the surface boundary conditions, extensive consideration of the effects of partial degeneracy in the stellar core, and incorporating non-ideal gas properties due to Coulomb interactions. Primordial deuterium burning occurs through reaction 2' of the P–P chain when T_c exceeds $\sim 7 \times 10^5$ K, and leads to a short-lived standstill in the rate of decline of L , T_{eff} for dwarfs with $M \geq 0.012 M_\odot$. The initial calculations of Grossmann *et al.* were based on a primordial abundance $N_D/N_H = 2 \times 10^{-4}$ and suggested that low-mass ($< 0.25 M_\odot$) dwarfs would spend 10^6 – 10^8 years (τ_D increasing with decreasing mass) on a deuterium-burning main sequence, running almost parallel to the hydrogen-burning sequence, but at higher luminosities and tempera-

tures. Current estimates place N_D/N_H closer to 2×10^{-5} , with a corresponding decrease in τ_D .

Adapting the boundary conditions to allow for non-grey opacities might seem to be a minor change, since the radiative layers constituting the stellar atmosphere include less than 1% of the total mass. However, these layers have a disproportionate effect in defining the internal structure in low-mass dwarfs, since they effectively define the entropy of the system. That entropy remains constant throughout the convective regions, the full interior in dwarfs with masses below $0.3 M_\odot$. The only atmosphere calculations available for use by Grossman *et al.* (from [A8]) were limited in scope, but the data were sufficient to indicate the importance of allowing for non-grey opacities. Incorporating the latter led to both lower luminosities and lower temperatures for a given mass and age, reducing the mass-limit for hydrogen burning. The $(\log T, \log \rho)$ radial profiles of the [G8] 0.5 and $0.085 M_\odot$ models are plotted in Figure 3.2. The former has a radiative core, and as a result $\rho \propto T^3$ (an $n = 3$ polytrope) in the central regions, while $\rho \propto T^{1.5}$ (an $n = 1.5$ polytrope) in the outer convective region. The $0.085 M_\odot$ model is both fully convective and partially degenerate near the core, with $\rho \propto T^{1.5}$ at all radii except within the outermost radiative envelope.

Taken as a whole, the [G8] models reproduce qualitatively the change in slope in the main sequence ascribed by [C6] to H₂ dissociation near the lower boundary of the atmosphere, although [G8] place the inflection at $0.6 M_\odot$ rather than $0.7 M_\odot$. [G8] also predict a change in slope in the mass–luminosity relationship near that mass, partly due to non-equilibrium in the ³He/H ratio as reaction 3' decreases in efficiency (with a consequent change in energy production), and partly due to increased molecular opacity sources. The net effect is a more rapid decrease in mass with decreasing luminosity for masses below $\sim 0.5 M_\odot$ – a flattening of the mass–luminosity relationship, dM/dL , evident in Figure 3.4.

Subsequent model calculations by D’Antona and Mazzitelli [D1], [D2] revealed evidence for further structure in the mass–luminosity relationship at masses close to the hydrogen-burning limit. Besides the change in slope at $\sim 0.5 M_\odot$, D’Antona and Mazzitelli identified a pronounced steepening in dM/dL at the lowest masses (Figure 3.4) – that is, a small change in mass corresponds to a substantial change in luminosity. This result, confirmed by Sienkiewicz [S8], Vandenberg *et al.* [V1] and more recent studies, has significant implications for the interpretation of the stellar luminosity function, $\Phi(M_V)$: specifically a flat or declining luminosity function (number of stars per unit magnitude) may well be consistent with a rising mass function (number of stars per unit mass). This matter is discussed in Chapter 9.

Further developments made during the 1980s in modelling low-mass stars are well summarised in the reviews by Dorman *et al.* [D5] and Burrows and Liebert [B11]. Many of those models still placed heavy reliance on the polytropic approximation and most used grey atmospheres. The significant breakthroughs of recent years have been stimulated by, first, the development of the appropriate equation of state for metallic hydrogen [S5]; second, the inclusion of grain opacities in evolutionary models [B12], [B13]; and, finally, the marriage between interior

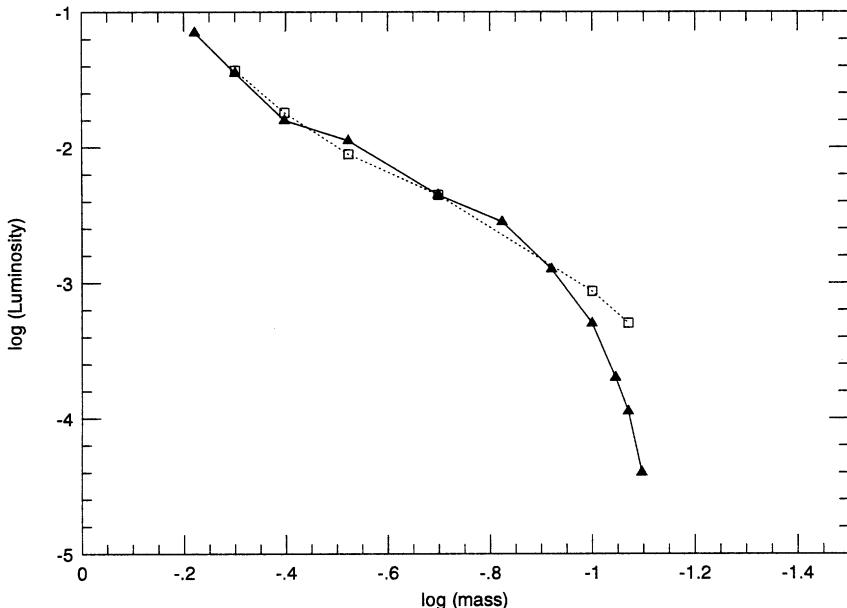


Figure 3.4. The mass–luminosity relationship derived by Grossman *et al.* ([G8], open squares) and D’Antona and Mazzitelli ([D2], solid triangles). Both mass and luminosity are given in solar units.

models [B2], [B3] and the complex low-temperature atmosphere codes devised by Allard and Hauschildt [A6] and collaborators (see Chapter 4). The properties of these recent low-mass models are summarised in Section 3.4.3.

3.4.2 The minimum mass for hydrogen burning

The existence of a lower mass limit to the hydrogen-burning main sequence for stars of a given abundance is tied to the onset of degeneracy in the stellar core during pre-main sequence evolution. As the protostar collapses, the central density rises and the core temperature, T_c , increases as potential energy is transformed to thermal energy. For solar-type stars, T_c is driven above $\sim 3 \times 10^6$ K and hydrogen fusion is initiated, providing a steady energy source which supports the star against further collapse.

Low-mass stars are required to collapse to higher densities than solar-type stars before achieving the critical T_c for fusion ignition. Since $GM^2/R \sim NkT$ in the perfect gas régime, and $M/R \sim \text{constant}$ for a given T_c , $\langle \rho \rangle \propto M^{-2}$. As the density increases, the core becomes partially degenerate, measured by the degeneracy parameter

$$\alpha_E = \frac{N_e h^3}{2(2\pi M - E k T)^{\frac{3}{2}}} \sim \frac{\text{electron chemical energy}}{kT} \quad (3.55)$$

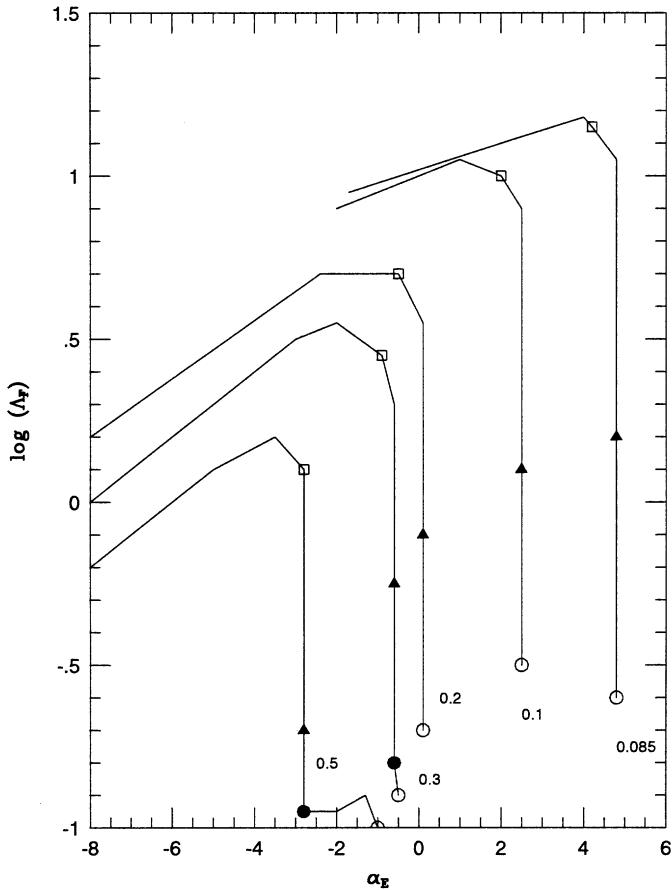


Figure 3.5. The Coulomb force, Λ_F , as a function of the radial variation in the degeneracy parameter, α_E for the 0.5 , 0.3 , 0.2 , 0.1 and $0.085 M_\odot$ models computed by [G8]. The open circle marks the centre of each model; the solid circle marks the lower boundary of the convective zone – the models are fully convective for $M < 0.25 M_\odot$. The solid triangle is the core boundary (including 97% of the total mass), while the open square indicates where degeneracy becomes important.

If $\alpha_E < -4$, the material can be treated as a perfect gas, and Maxwell–Boltzmann statistics apply, while if $\alpha_E > 20$ the material is fully electron degenerate, and Fermi–Dirac statistics are appropriate. Figure 3.5 (adapted from [G8]) plots the predicted radial variation in α_E for a range of masses. Λ_F is the plasma parameter, the ratio between the Coulomb energy and kT . A plasma parameter $\Lambda_F < 0.1$ indicates weak screening of the nucleus, while $\Lambda_F > 10$ implies strong screening. The figure illustrates the growing degeneracy with decreasing mass, with α_E constant within the convective envelope.

As degeneracy becomes more significant in a contracting protostar, an increasing fraction of the liberated potential energy is absorbed in reducing the separation

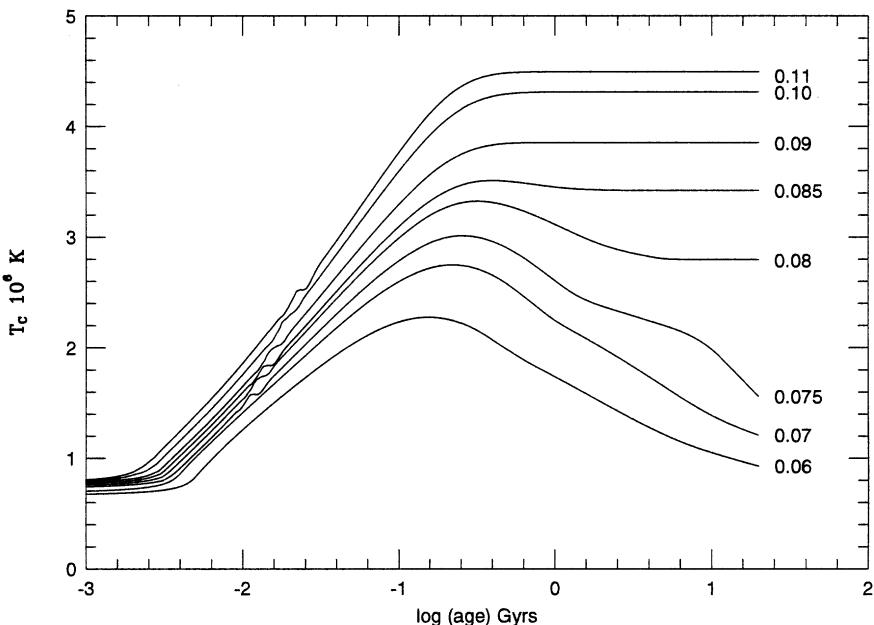


Figure 3.6. The evolution of core temperature with time predicted by Burrows *et al.* [B12], [B13] for dwarfs bridging the transition from main sequence, core-hydrogen-burning stars to brown dwarfs.

between degenerate electrons, rather than transformed to thermal energy. Degeneracy prevents the stellar radius from decreasing below $\sim 0.1 R_\odot$. The consequence, originally pointed out by Kumar [K9], is that T_c reaches a maximum value, T_{max} . Subsequent evolution depends on the total mass of the dwarf, and on the degree to which nuclear fusion can continue to provide a stable energy source. There is not, in fact, a simple, clean division between ‘stars’ – where fusion makes a significant contribution to L_{tot} – and ‘brown dwarfs’ – which, to quote Burrows and Liebert, ‘cool like a rock’.

A low-mass dwarf faces four possible futures after the onset of significant core degeneracy. These are illustrated in Figure 3.6 using models computed by Burrows *et al.* [B12], [B13]. In order of decreasing mass, they are:

- Hydrogen fusion continues at a sufficient rate to maintain $T_c \sim T_{max}$; the star remains on the main sequence, and the luminosity and effective temperature are constant for many Hubble times ($M \geq 0.09 M_\odot$ in Figure 3.6).
- Degeneracy reduces T_c below T_{max} , but the temperature remains sufficiently high to permit continuing fusion, so the dwarf achieves a main sequence configuration ($0.08\text{--}0.085 M_\odot$ in Figure 3.6).
- Fusion has been initiated, but degeneracy (eventually) reduces T_c below the critical level for continuing hydrogen-burning, so the ‘ex-star’ cools as a brown dwarf ($0.075 M_\odot$ in Figure 3.6).

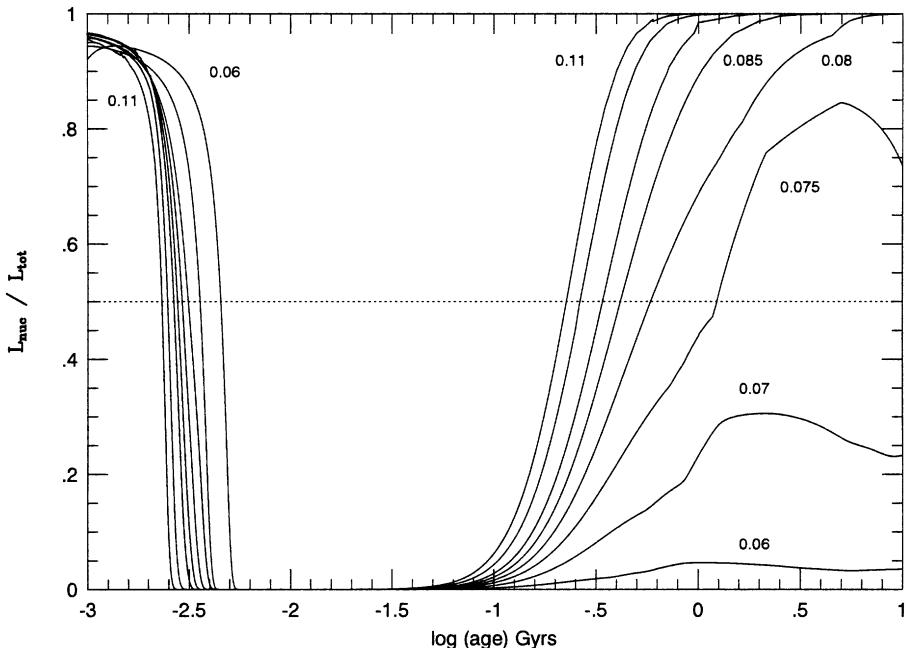


Figure 3.7. The fractional luminosity contributed by nuclear fusion reactions as a function of time for the [B12], [B13] models plotted in Figure 3.6. The initial burst of fusion marks deuterium burning which persists longest in the lowest-mass model.

- T_c never reaches sufficiently high levels to allow fusion to become a significant energy source; the object becomes a degenerate brown dwarf and cools to oblivion ($M \leq 0.07 M_\odot$ in Figure 3.6).

The first and second scenarios result in stable low-mass stars; scenario (4) produces a brown dwarf, and the hydrogen-burning limit, M_{HBL} , lies somewhere between scenario (3) (which produces what has become known as a ‘transition object’) and scenario (4). The dividing line is usually placed at a mass where the fractional contribution to the total luminosity due to hydrogen fusion, L_{nuc}/L_{tot} , never exceeds 50% – an arbitrary definition.

Kumar’s original estimate placed M_{HBL} at $\sim 0.07 M_\odot$. Figure 3.7 plots L_{nuc}/L_{tot} for the Burrows *et al.* models illustrated in Figure 3.6. These models predict that a $0.08 M_\odot$ solar-abundance dwarf achieves a stable luminosity and temperature after $\sim 2 \times 10^8$ years; that a $0.075 M_\odot$ dwarf is a transition object, able to sustain fusion for nearly 10^{10} years, but eventually fated to cool degenerately; and that a $0.07 M_\odot$ dwarf, while partially supported by fusion at ages of 1–2 Gyr, evolves as a brown dwarf. The formal hydrogen-burning limit in these models is $M_{HBL} = 0.073 M_\odot$.

The calculations plotted in Figures 3.6 and 3.7 are for solar-abundance (disk) dwarfs. Decreasing the metal content leads to lower opacities and both higher effective temperatures and higher luminosities at a given mass. As discussed in

Chapters 2 and 10, these effects lead to metal-poor stars defining separate main sequence(s), lying blueward of disk dwarfs in the H–R diagram. The higher surface temperatures in turn require higher core temperatures, with the result that M_{HBL} is driven to higher masses at the lower metallicities found in the Galactic Halo. Both D’Antona and Mazzitelli [D4] and Baraffe *et al.* [B4] have computed low-mass models for a range of abundance. The former models predict that M_{HBL} rises from $0.090 M_\odot$ at a metallicity of $[Fe/H] = -0.7$ to $0.097 M_\odot$ at $[Fe/H] = -2.3$. The latter value is almost $0.02 M_\odot$ higher than the solar-abundance M_{HBL} computed by the same authors. Finally, lower opacities and higher transparency lead to more rapid cooling of substellar-mass brown dwarfs, and a correspondingly more rapid decrease in luminosity with time. As a result, direct observations of brown dwarf members of the Galactic Halo are likely to task even the James Webb Space Telescope.

3.4.3 Properties of current models of low-mass dwarfs

The 1990s saw substantial advances in modelling both the atmospheres and the internal structure of low-mass dwarfs. Similar advances are likely to occur over the succeeding decade or more. Nonetheless, the predictions made by the current generation of stellar models provide a snapshot of our present understanding in the field. We consider the results derived from two sets of models: by the Tucson group – Burrows, Hubbard, Lunine, Saumon and collaborators; and by the Lyon group – Baraffe, Chabrier and collaborators.

For present purposes we concentrate on the Tucson calculations described by Burrows *et al.* ([B12] – the ‘X’ models) and the extension to lower masses in Burrows *et al.* [B13], while the Lyon models are from Baraffe *et al.* [B4], [B5]. The Tucson dataset is limited to low-mass dwarfs, spanning the mass range $0.009\text{--}0.20 M_\odot$, while Baraffe and Chabrier extend their calculations to $1 M_\odot$, but do not sample the substellar régime at the same resolution in mass. Both sets of models adopt similar physics for the internal structure, with the equation of state based on Saumon and Chabrier’s [S4] calculation. The main differences lie in the opacities and the boundary conditions: the [B12] Tucson models use grey atmospheres, although the later lower-mass calculations adopt non-grey boundary conditions, and both sets of models allow for dust formation and grain opacities; the Lyon calculations, on the other hand, use Allard and Hauschildt’s model atmospheres (Chapter 4) to define the surface boundary conditions, but (until recently) did not allow for extensive grain formation. This difference in formulation leads to a difference in the presentation of the model predictions: Burrows *et al.* limit their predicted properties to data in the theoretical plane (L , T_{eff}) and near-infrared (JHK) colours; Baraffe and Chabrier, with the advantage of more realistic atmospheres, provide colours and magnitudes ($BVRIJHK$) in addition to (L , T_{eff}). Taken together, these two sets of models provide a measure of the reliability of our current understanding of the structure of low-mass dwarfs.

The basic characteristics of low-mass stars and brown dwarfs are best illustrated by considering their evolution in luminosity and effective temperature as a function

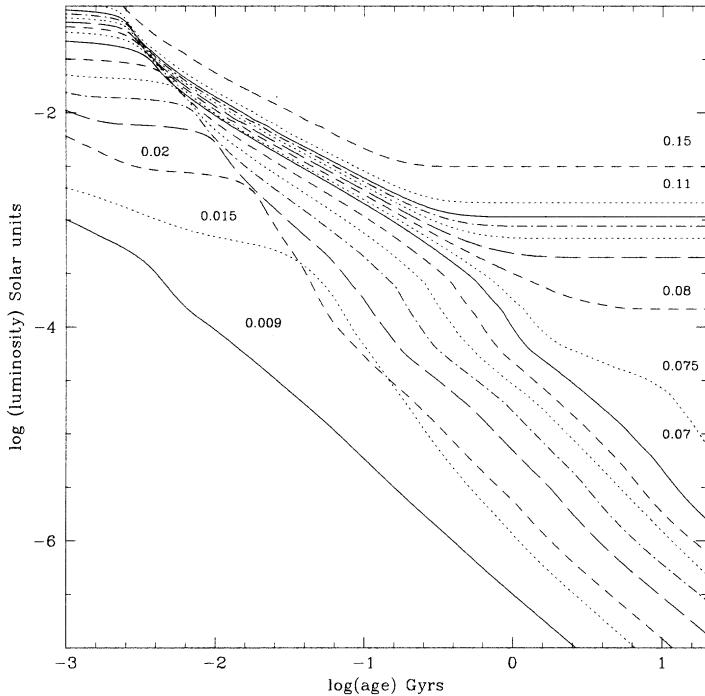


Figure 3.8. Luminosity evolution of low-mass dwarfs as computed by the Tucson group. The models range in mass from 0.009 to $0.15 M_{\odot}$. The divergence of brown dwarfs from main sequence stars, and the equivocal behaviour of transition-mass objects, is evident at ages exceeding ~ 1 Gyr.

of mass. Figures 3.8 and 3.9 plot those predictions for the Tucson models. Both diagrams show a clear bifurcation in the behaviour at $\tau > 2\text{--}3$ Gyr, with hydrogen-burning stars maintaining nearly constant properties after settling into their main sequence configurations, while brown dwarfs show a continued decline in both L and T_{eff} . The $\sim 0.075 M_{\odot}$ transition objects exhibit a standstill in that descent for up to ~ 10 Gyr (close to the age of the Galactic Disk), before fusion is extinguished and cooling takes over. The transition between M dwarfs and L dwarfs is generally placed at $\sim 2,100$ K. Thus, these models predict that early-type L dwarfs encompass a mix of both *bona fide* old, hydrogen-burning stars and younger (< 2 Gyr-old) brown dwarfs.

The effective temperature of almost all objects – save those of the lowest mass – is predicted to remain nearly constant at ages of less than $\tau \sim 10^7$ years. At the same time, the luminosity decreases by more than an order of magnitude. This obviously reflects the rapidly-decreasing radius as the dwarf contracts either onto the main sequence or to a degenerate configuration. Figure 3.10 shows the behaviour predicted by the Tucson models. Fusion of primordial deuterium provides an additional energy source during this same time period for dwarfs with masses exceeding $\sim 0.012 M_{\odot}$, as evidenced by a plateau in the evolution of the luminosity, effective

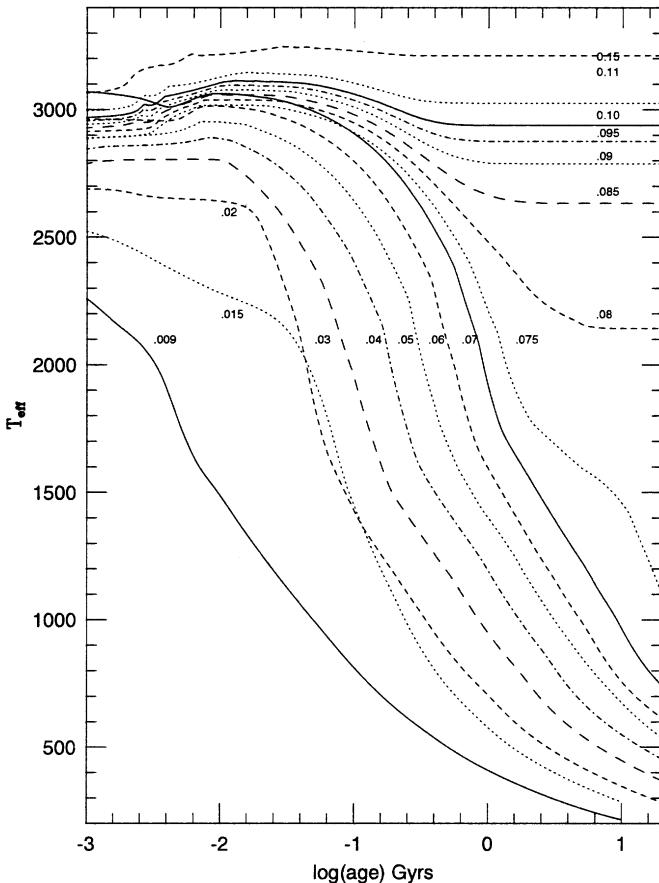


Figure 3.9. The evolution of effective (surface) temperature with time, for the same models illustrated in Figure 3.8.

temperature and radii. This behaviour must be taken into account when determining statistical parameters, such as the mass function, for young star clusters.

The same dichotomous behaviour is evident in the evolution of the mass-luminosity relationship with time (Figure 3.11). For ages up to $\tau \sim 3 \times 10^8$ years, the luminosity decreases over the full mass range with increasing age, with deuterium burning slowing the rate of decline at $\tau \leq 10^7$ years, $M \gtrsim 0.012 M_\odot$. At ages exceeding 3×10^8 years, the lowest mass stars have achieved thermal and hydrostatic equilibrium, but L continues to decline at substellar masses. This divergence was first emphasised by D'Antona and Mazzitelli, and clearly eliminates the possibility of determining masses for brown dwarfs unless age is a known quantity. The latter requirement underlines the utility of open star clusters in mass function analyses.

Finally, Figure 3.12 shows evolutionary tracks in the $(\log L, T_{\text{eff}})$ H-R diagram. The Tucson models predict $T_{\text{eff}} \sim 1,900$ K and $\log(L) \sim 10^{-4} L_\odot$ as the lowest value attainable by a hydrogen-burning star; the Lyon models place the limit at somewhat

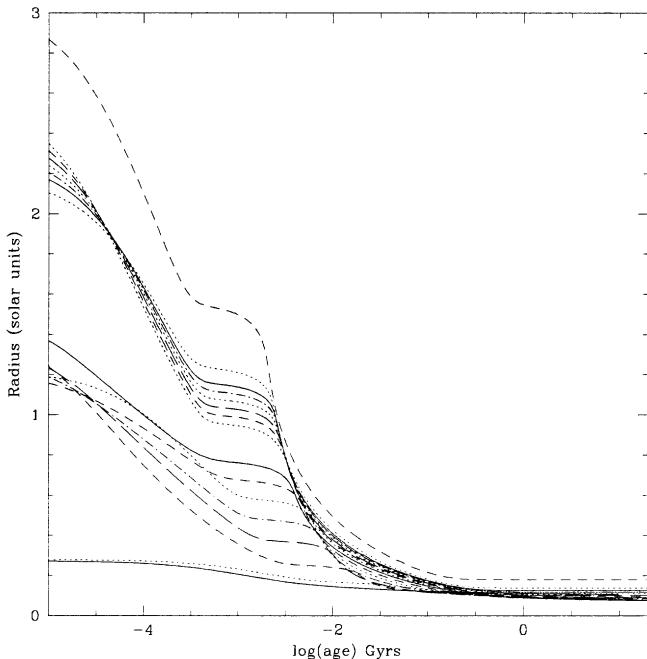


Figure 3.10. The variation with time of radii of low-mass dwarfs, as predicted by the Tuscon models. As in the $L(\tau)$ and $T_{\text{eff}}(\tau)$ diagrams, the initial plateau reflects energy generated through deuterium burning. The tracks correspond to the same masses as in Figures 3.8 and 3.9.

higher temperatures and luminosities: $\sim 2,000$ K and $\sim 10^{-3.9} L_{\odot}$. Lower-mass objects evolve through the full temperature range between $\sim 3,000$ K and < 500 K, but at different rates depending on the mass. Note that lower-mass brown dwarfs are predicted to have higher luminosities for a given T_{eff} , reflecting their larger radii. The main consequence of this temperature evolution is that the observed spectral type of substellar-mass objects changes with time, complicating the interpretation of observations of L and T dwarfs.

3.5 MATCHING MODELS AND OBSERVATIONS

The crucial test of these theoretical models is the comparison between their predicted properties and the observed characteristics of low-mass dwarfs. These comparisons can be undertaken in either the observational or theoretical planes: for example, the H–R diagram comparison can be made in either the $(\log L, T_{\text{eff}})$ plane or the (absolute magnitude, colour) plane. In both cases, uncertainties in the transformation between observed and theoretical quantities contribute significantly to the uncertainties in the final comparison. Theoretical models of low-mass stars and

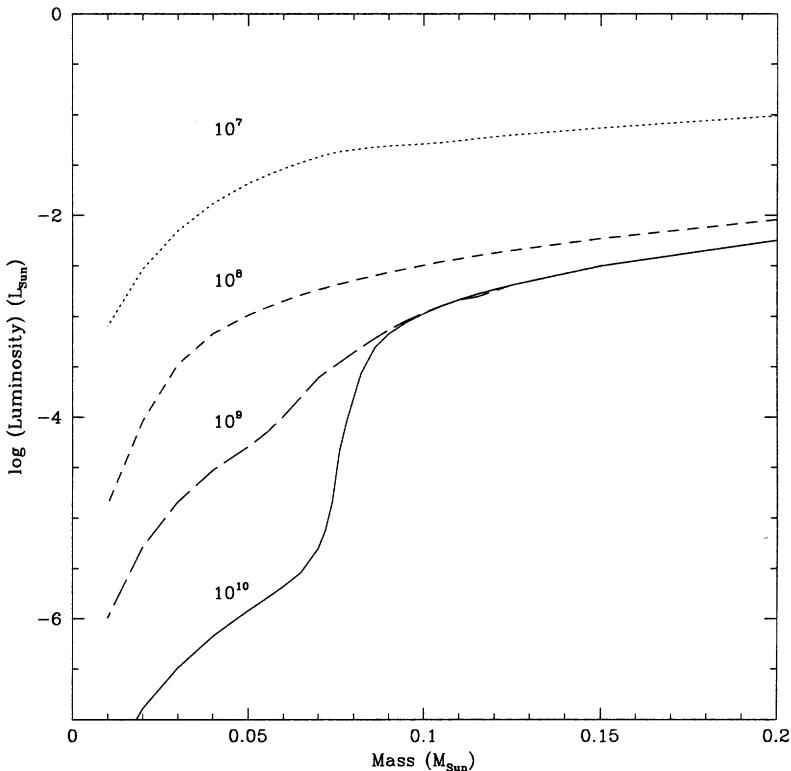


Figure 3.11. The mass–luminosity relationship at various ages, as predicted by the Tucson low-mass dwarf models. The substantial evolution in luminosity at low masses originally highlighted by D’Antona and Mazzitelli is clearly evident.

brown dwarfs showed marked improvement throughout the 1990s. For present purposes, the Tucson and Lyon datasets are taken as representative of current models of disk dwarfs, supplemented by D’Antona and Mazzitelli’s [D4] models of metal-poor halo stars. (The reader should check contemporary literature for the latest improvements.)

3.5.1 The Hertzsprung–Russell diagram

Figure 3.13(a) compares the observed and predicted locations of the lower main sequence on the H–R diagram. The empirical temperatures and luminosities are those of the mean M-dwarf scale described in more detail in Chapter 4 (see Table 4.1), supplemented by data for the few L dwarfs with measured trigonometric parallaxes. The theoretical relationships plotted are for 10 Gyr-age Tucson and Lyon models. The models are identical with those plotted in Figure 3.12.

The two sets of theoretical calculations are in good agreement. The Lyon models are $\sim 50\text{--}100\text{ K}$ cooler and $\sim 0.05\text{--}0.1$ dex less luminous in $\log(L/L_\odot)$ for masses

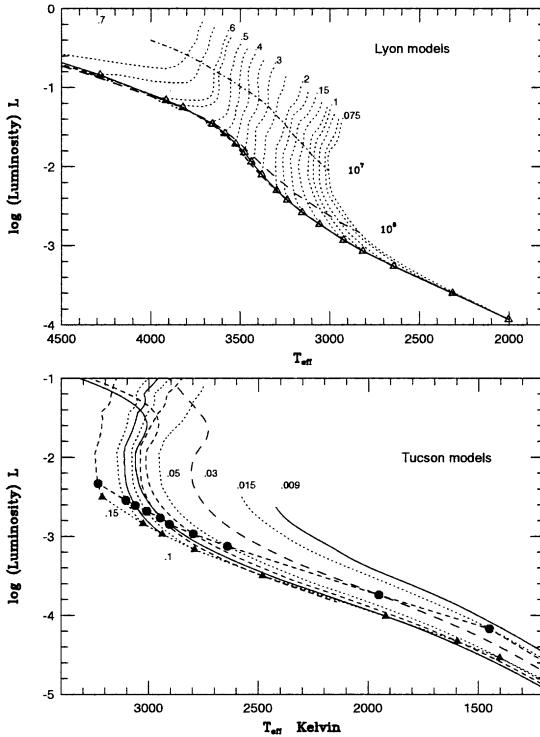


Figure 3.12. Evolutionary tracks of low-mass dwarfs. The lower diagram plots results from the Tucson group for masses between 0.15 and $0.009 M_{\odot}$ (coded as in Figure 3.9). Solid dots mark ages of 10^8 years, and solid triangles 10^9 years. The upper diagram plots predictions from the [B5] Lyon models. Representative tracks are identified by their mass and isochrones for ages 10^7 (dash-dot line), 10^8 (dashed line), 10^9 (short-dashed line) and 10^{10} years (solid line), with the open triangles marking the main-sequence location of each model. Luminosity is given in solar units in both diagrams.

above $0.08 M_{\odot}$, although differences near M_{HBL} lead to the $0.075 M_{\odot}$ Lyon model being ~ 400 K hotter and ~ 0.5 dex brighter than the Tucson counterpart. Neither set of models, however, provides a perfect match to the empirical sequence. As Figure 3.13(b) shows, this is not a new problem: theoretical models have been consistently 200–300 K hotter than the observationally-defined M dwarf sequence for more than a quarter of a century. The [G8] models provide a partial exception, matching the observed location of early M dwarfs, but overestimating both luminosity and temperature to an increasing extent with decreasing mass.

Since the Lyon [B5] models incorporate the Allard and Hauschildt model atmosphere calculations, they also predict colours and absolute magnitudes in the observational plane. Figure 3.14 compares those predictions against the observed ($M_V, (V-I)$) and ($M_I, (I-K)$) distributions described by single disk dwarfs with accurate ($\sigma_{\pi}/\pi < 10\%$) trigonometric parallax determinations. Besides the solar-

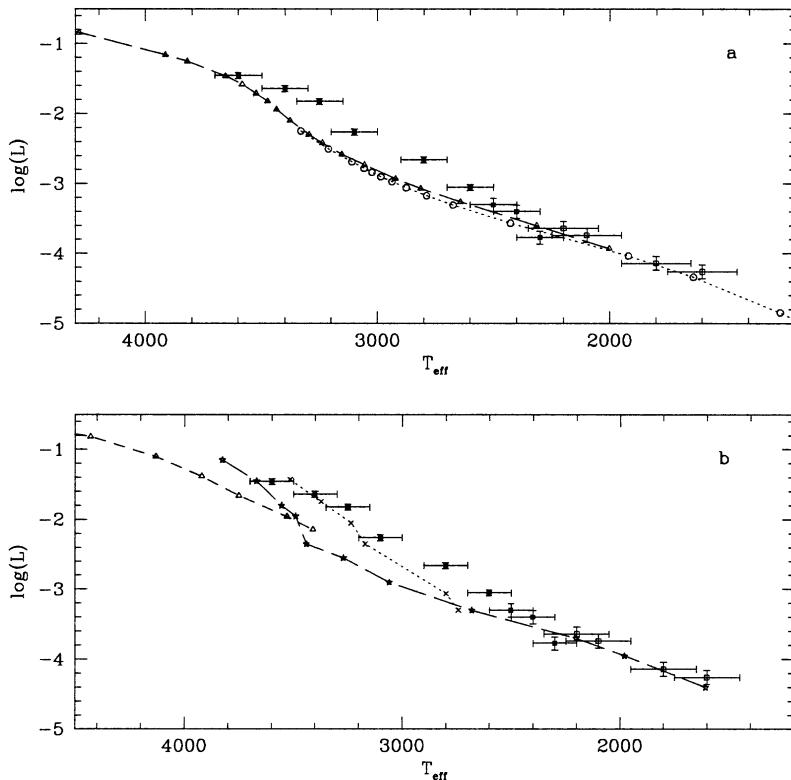


Figure 3.13. Comparison of the theoretical and observational H–R diagrams. (a) compares the Tucson (open circles) and Lyon (open triangles) solar abundance models for age 10 Gyr against the empirical M-dwarf sequence (solid squares) and data for L dwarfs with measured trigonometric parallaxes (open squares). (b) matches the same observational data against the theoretical main sequences from [C6] (triangles), [G8] (crosses), and [D2] (stars). All of the models produce significantly higher temperatures than the observed sequence. Luminosity is given in solar units in both diagrams.

abundance models, the [B5] predictions for $[\text{Fe}/\text{H}] = -0.25$ dex (a closer match to the mean abundance of the Galactic Disk) are also shown. Both models are for ages of 10 Gyr, and are limited to $M > 0.075 M_{\odot}$; that is, hydrogen-burning dwarfs in their main sequence configurations.

These models represent a significant advance over all previous analyses, but Figure 3.14 shows that there is still room for improvement. The solar abundance models are significantly bluer than the observations for $M_V > 10$, as might be expected given the higher temperatures evident in Figure 3.13. The inflection in the main sequence at $M_V \sim 9.5$ ($M \sim 0.5 M_{\odot}$, $(V-I) \sim 2.1$) in the $[\text{Fe}/\text{H}] = 0$ models is due to H_2 formation and dissociation near the photosphere, while the subsequent flattening at $M_V \sim 13$ marks the increasing importance of degeneracy. Neither the ‘step’ in the empirical main sequence at $M_V \sim 12.5$, $(V-I) \sim 3$ nor the

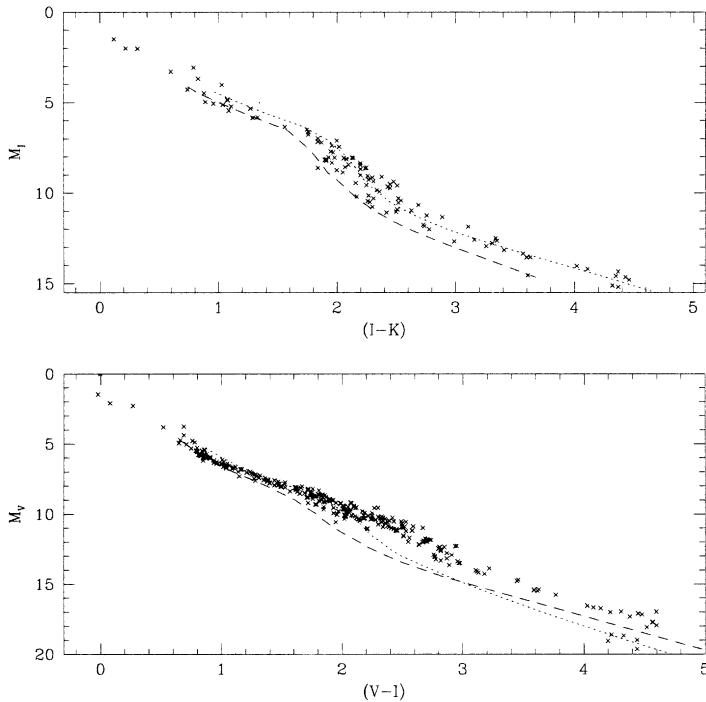


Figure 3.14. The [B5] (M_V , $(V-I)$) and (M_I , $(I-K)$) theoretical sequences compared with observations of nearby stars with accurate parallax measurements. The dotted line is the theoretical solar-abundance sequence; the dashed line marks the $[Fe/H] = -0.25$ dex predictions.

'hook' at $M_V > 17.5$ are reproduced by these (or any other) models. Moreover, the Lyon $[Fe/H] = -0.25$ models are predicted to be redder than the solar abundance models at $M_V > 15$. All of these effects may well reflect incomplete treatment of opacities in the model atmospheres.

The Lyon group have developed a new set of models for very low mass stars and brown dwarfs, incorporating dust formation. We will discuss these and other related studies in more detail in Chapter 4.

Theory and observations are in better agreement in the (M_I , $(I-K)$) plane, with the solar abundance isochrone providing a good match to the observed sequence. This may be due partly to the lesser importance of molecular opacities at 0.8 and $2.2\mu m$ – an hypothesis supported by the theoretical/empirical mass–luminosity comparisons discussed in the following section. However, the good agreement between theory and observation at near-infrared wavelengths does not extend to the cooler L dwarfs: both the Tucson and Lyon models predict a maximum ($J-K$) colour of ~ 1.0 , while the reddest L dwarfs have ($J-K$) colours exceeding two magnitudes. As discussed further in Chapters 4 and 6, the redder colours are probably due to dust formation within these cool atmospheres, requiring even more complex theoretical models.

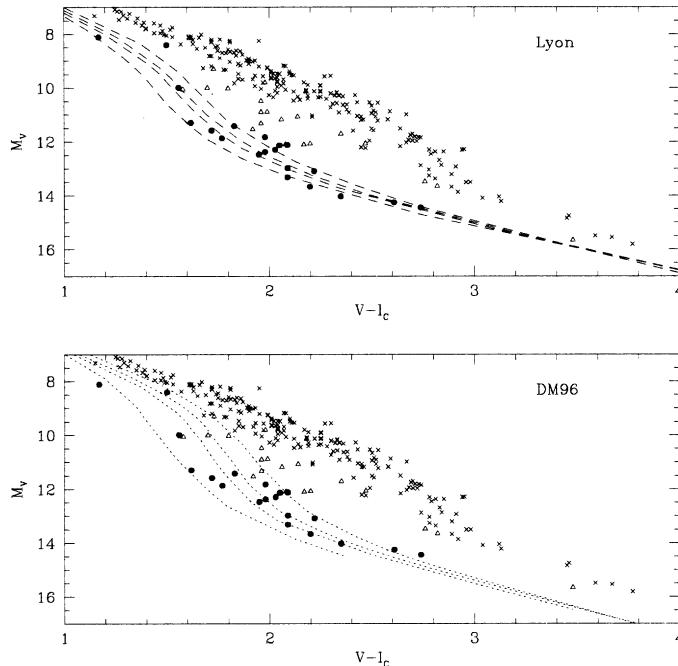


Figure 3.15. Theoretical main sequence relationships for subdwarfs, as predicted by D’Antona and Mazzitelli [D4] (DM96, dotted lines) and Baraffe *et al.* [B4] (Lyon, dashed lines). The former models are for $[Fe/H] = -2.3, -1.3, -1.0$ and -0.7 ; the latter for abundances of $-2.0, -1.5, -1.3$ and -1.0 . The disk main sequence (crosses) is defined by the nearby stars plotted in Figure 3.14; data for the halo subdwarfs are taken from [M5] and [G2]. Solid points mark extreme subdwarfs; open triangles mark intermediate subdwarfs (see Chapter 10).

As discussed in the previous section, low-mass subdwarfs follow different colour–magnitude relationships than disk dwarfs. Both D’Antona and Mazzitelli and the Lyon group have predicted $(M_V, (V-I))$ relationships for a range of abundances. It should be noted that these models adopt different helium abundances: [D4] assume that the helium mass fraction increases from 23% at $[Fe/H] = -2.3$ to 24% at $[Fe/H] = -0.7$, as would be expected given nucleosynthesis in massive stars and recycling; [B4] adopt $Y = 0.25$ for all models. Both sets of isochrones are predicted to have a characteristic S-shape due to H_2 dissociation at $\sim 0.5 M_\odot$, and the dominance of degeneracy at masses below $\sim 0.12 M_\odot$. The qualitative accuracy of these predictions has been confirmed by deep colour–magnitude data for globular clusters (Section 10.5).

Figure 3.15 compares both sets of predictions against data for nearby subdwarfs with accurately determined trigonometric parallaxes. The observational sample has been divided into intermediate and extreme subdwarfs (a classification discussed further in Chapters 4 and 10). The former stars are likely to have abundances $[m/H] > -1.5$, while the latter are more metal-poor. The models match the

extreme subdwarfs, but not the intermediate abundance stars. Uncertainties in the upper boundary conditions (the atmospheres) are likely to make a significant contribution to the discrepancy.

3.5.2 The mass–luminosity relationship

Mass is the fundamental stellar parameter, but the determination of stellar masses is extremely difficult observationally. (Chapter 9 includes extensive discussion of the methods currently used for measuring masses.) At present, masses have been measured for only a small number of disk stars in binary systems, and no empirical data are available for metal-poor subdwarfs. Thus, theoretical/empirical comparisons are limited to near-solar abundances. The mass–luminosity relationships predicted by both the Tucson and Lyon solar-abundance models prove to be in reasonable agreement with the available observational calibrators. Figure 3.16 compares these relationships (for ages of 0.6 and 10 Gyr and 1 and 10 Gyr respectively) against empirical measurements of binary star components (see Table 9.1). The agreement is well within the observational uncertainties. In similar fashion, the Lyon models provide a reasonable match to the observed distribution in the (M_K , mass), (M_V , mass) and (M_I , mass) planes. On the other hand, the scatter in the

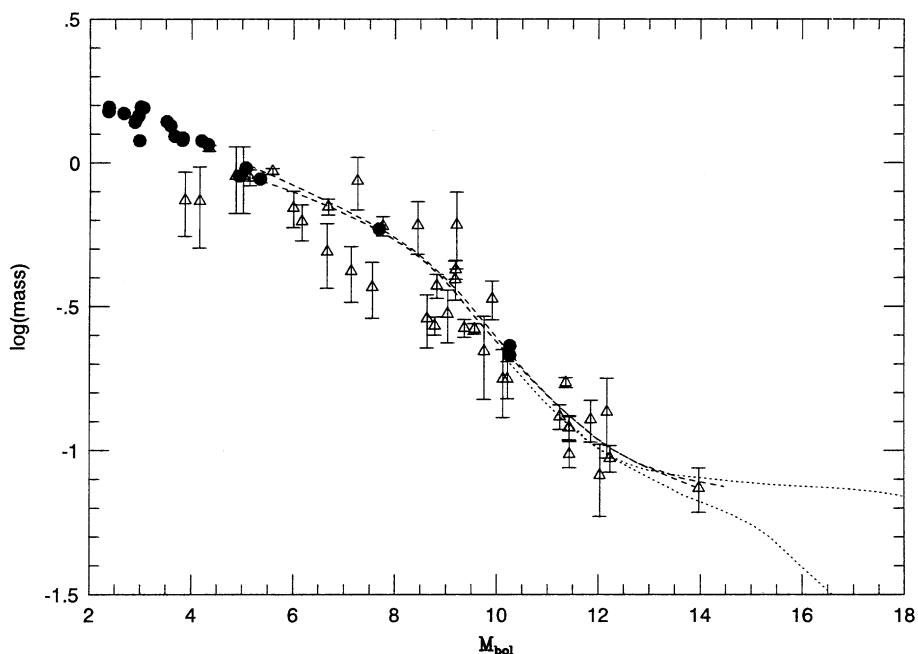


Figure 3.16. The mass–luminosity relationship for low-mass stars. Solid points denote data for eclipsing binaries (from [A7]); triangles mark astrometric binaries (from [H7], [H8]). The dashed lines are 1 and 10 Gyr isochrones from the Lyon models; the dotted lines (limited to masses below $0.2 M_\odot$) are the 0.6 and 10 Gyr Tucson predictions. Mass is given in solar units.

empirical data is sufficiently high that it is not possible to determine whether the mismatch between theory and observation in the colour–magnitude plane (Figure 3.14) stems from calibration problems in one or several passbands. This is not surprising given the shallow slope of the mass–absolute magnitude relationships below $\sim 0.2 M_\odot$, and lack of data at very low masses.

3.5.3 The mass–radius relationship

Since theoretical models disagree with the empirically-defined temperature scale but are in reasonable agreement with observed luminosities, disagreement between theoretically and observationally inferred stellar radii might be expected – and this proves to be the case. Figure 3.17(a) presents the expected mass–radius relationship for low-mass main sequence stars and 0.6 and 10-Gyr-old brown dwarfs; the Lyon and Tucson models are in good agreement in the range of overlap in mass. As outlined above, the radius is expected to decrease almost linearly with decreasing mass until close to $0.1 M_\odot$, where degeneracy becomes the dominant source of pressure support, preventing the radius decreasing below $\sim 0.08 R_\odot$. The radius increases with decreasing mass as degeneracy becomes less important: Jupiter, at $0.009 M_\odot$, is essentially non-degenerate.

Directly determined radii and masses are available only for the components of eclipsing binaries (Table 2.1). However, given a temperature scale, it is possible to calibrate a photometric index as a temperature indicator, and (as for Table 4.1) derive photometric radii from

$$R = \left(\frac{L}{4\pi\sigma} \right)^{0.5} T^{-2} \quad (3.56)$$

Luminosity calibration is straightforward given an accurate parallax and multi-wavelength photometry.

Clemens *et al.* [C5] have derived temperature calibrations for both the ($V-I$) and ($I-K$) colour indices, calibrating these relationships using well-studied stars from the photometric data compiled by Leggett *et al.* [L5]. They are consistent with the mean spectral-type calibration given in Chapter 4 – unsurprisingly, since the latter relation is also based on the [L5] calibrators. Masses are estimated using the empirical (M_K , mass) relation derived by Henry and McCarthy [H7] (see Figure 3.16). Figure 3.17(b) plots the mass–radius relationship outlined by single stars in the 8-parsec sample with the requisite photometry, and compares it with the theoretical calibrations. Data for astrometric binaries are from Popper [P1], in which photometric techniques were employed to estimate radii. The directly measured radii of M dwarf eclipsing binaries are also plotted. There is an obvious offset between theory and observation in the expected sense: the theoretical temperatures are higher, requiring smaller theoretical radii if the observed and predicted luminosities are to agree.

As yet, there is no unambiguous means of choosing between theoretical and empirical temperature (and hence radius) calibrations. However, since the theoretical relation is predicted to vary smoothly with mass, similar behaviour can be expected

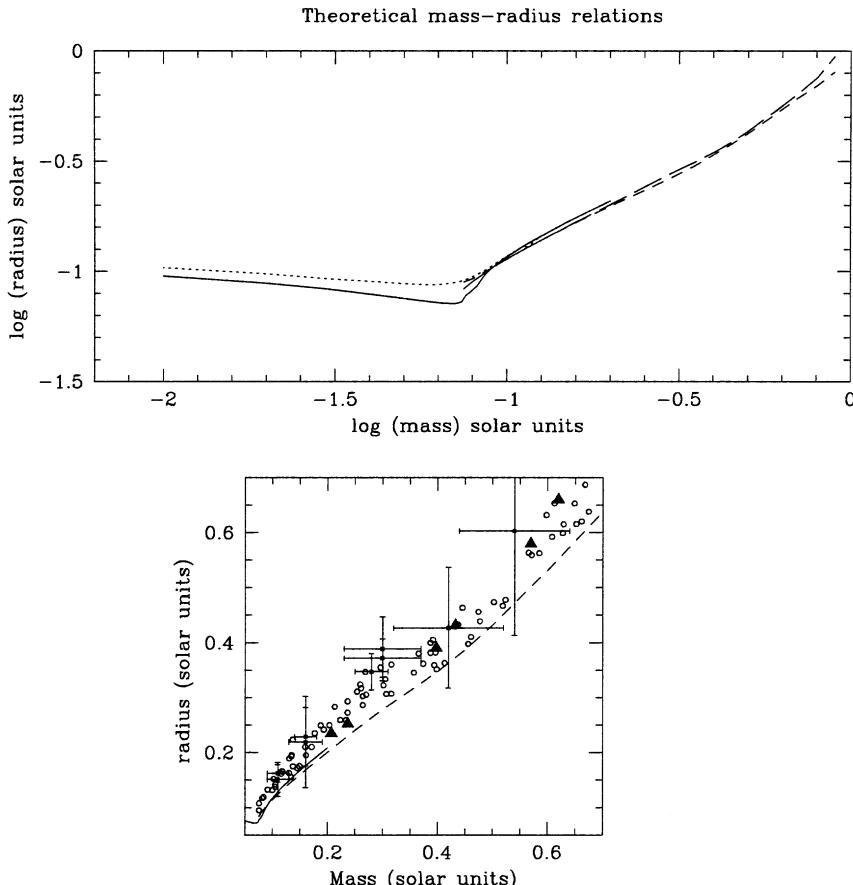


Figure 3.17. The mass–radius relationship. (a) plots the predicted behaviour of the 0.6 (dotted line) and 10 Gyr (solid line) Tucson models and the 1 and 10 Gyr Lyon models (dashed lines). (b) compares the observed and predicted relationship within the mass range spanned by M dwarfs: open circles are photometrically calibrated radii; solid squares with error bars are data from Popper's [P2] analysis of astrometric binaries; solid triangles mark the locations of the CM Dra, YY Gem and GJ 2069A components. (From [C5], courtesy of the *Astrophysical Journal*.)

in the colour–magnitude diagram, modified by opacity-induced effects. In fact, the (M_V , $(V-I)$) main sequence shows an extremely abrupt feature at $(V-I) \sim 3$, which is reproduced in other colour–magnitude and spectral type/absolute magnitude diagrams. The fact that this feature is present over a wide wavelength range suggests that the effect is global, rather than due to absorption affecting one particular passband. An obvious possibility is a sharp change in radius over a relatively small range in temperature, and the proximity to $0.3 M_\odot$ suggests a correlation with the onset of full convection. Alternatively, Clemens *et al.* [C5] hypothesise that the effect might be tied to changes in the interior structure – perhaps separation of iron

and hydrogen in the stellar core. Whatever the cause, this ‘step’ in the main sequence remains unexplained by current theoretical models.

3.6 THE FORMATION OF LOW-MASS STARS

Star formation is a fundamental (perhaps *the* fundamental) astrophysical process, underlying heavy element production and the overall evolution of galactic systems. As yet, we have only a partial understanding of the various mechanisms which govern the morphological rearrangement of clouds of diffuse gas into self-gravitating spheres in hydrostatic equilibrium. These uncertainties reflect the complexity of the process, particularly dealing with gas magnetohydrodynamics over scales of up to several parsecs. A full consideration of star formation theory and observations lies beyond the scope of this book. In this section we sketch the basic outline and refer the reader to the review of Shu *et al.* [S9], Hartmann’s monograph [H2] and Stahler and Palla’s textbook [S18], for more detailed treatment.

3.6.1 Molecular clouds: the initial phase

Stellar birth takes place within giant molecular clouds (GMCs). These systems have masses of 10^5 to $\sim 5 \times 10^6 M_\odot$, sizes of 10–60 parsecs, densities exceeding 10^9 particles cm^{-3} , and temperatures of 10–30 K. Relatively complex molecules can form at these densities and temperatures, including HC_3N , CH_3OH and HC_{11}N , in addition to SO_2 , CO and CS. GMCs lie close to the mid-plane of the Galactic Disk, and are evident in the optical merely as dark, absorbing patches, silhouetted against the background stellar population. With typical internal extinctions of $A_V > 25$ magnitudes, these complexes are impervious to study at even near-infrared (1–2 μm) wavelengths. However, thermally-induced vibrational and rotational transitions in the constituent molecules produce emission at millimetre wavelengths, and these emissions can be observed, and the clouds can be mapped using millimetre and sub-millimetre radio telescopes [S3].

As might be expected, molecular hydrogen is the most abundant species present in GMCs. Since H_2 lacks any strong emission lines, most observations centre on CO, adopting a standard CO/H₂ conversion factor to transform the observed line flux to an estimate of the total molecular mass.⁴ Given a CO flux measurement in units of (Kelvin km s^{-1}), the Galactic conversion factor is $\alpha_{\text{Gal}} \sim 2.3 \pm 0.3 \times 10^{20}$ molecules $\text{cm}^{-2} (\text{K km s}^{-1})^{-1}$. This factor is based mainly on observations of individual clouds, where the velocity dispersion of the gas can be used to estimate a dynamical mass that can be compared to the mass in CO inferred from the total flux. In addition, the diffuse γ -ray background can be used to estimate the conversion factor, since this background results from interactions between cosmic rays and interstellar gas [S17].

⁴ CS and ¹³CO are often used to trace the density distribution in high-density regions where emission from the more common isotope ¹²CO is saturated (optically thick).

The first millimetre observations were made using single-dish radio telescopes with diameters of up to 25 m, achieving typical angular resolution of 40–60 arcsec. Over the last decade, interferometric arrays, combining several individual dishes (usually each of diameter 6–10 m) with baselines of up to 200–300 m, have improved the resolution to 1–2 arcsec. These measurements are comparable to optical and near-infrared observations and provide a linear resolution of a few hundred AU in the nearby Taurus (140 parsecs) and Orion (500 parsecs) star-forming clouds. This is close to the scale expected for the pre-planetary solar nebula (Chapter 11), and has provided considerable insight on the early stages of planetary formation. These observations will become even more interesting with the completion of the Atacama Large Millimetre Array (ALMA), which will have at least 64 12-m telescopes at an altitude of 16,400 ft in Llano de Chajnantor, Chile, and will provide a resolution of 10 milliarcseconds, or 1–2 AU.

Single-dish observations revealed that GMCs are extremely inhomogeneous, with high-density regions only a few parsecs in size [S2], [B7]. These clumps include smaller regions of even higher density – known as cloud cores – which mark the sites where individual protostars are forming. The driving mechanism behind formation is gravitational collapse, suitably modified by the magnetic fields known to be present within the GMCs [S9]. The initial collapse phase is generally believed to be simple spherical accretion onto an incipient protostellar core. However, most clumps are expected to have some degree of ordered rotation due to gravitational interactions with other clumps. As the collapse progresses, conservation of angular momentum leads to the formation of an accretion disk around the central protostar [P3], [S9].

As the protostar becomes more massive, it also becomes hotter and more luminous, with

$$L \approx \frac{GM\dot{M}}{R_*} \quad (3.57)$$

where \dot{M} is the accretion rate, and R_* is the protostellar radius [S9]. At some point, the global infall decreases substantially, either due to radiation pressure or to the effect of winds generated by the star. A bipolar outflow develops along the axis perpendicular to the disk, with gas outflow velocities of 50–100 km s⁻¹. The outflow gives rise to the spectacular structures evident in Herbig–Haro (HH) objects (Figure 3.18, colour-plate section).

The timescales involved in these early evolutionary phases are somewhat uncertain. For a protostar to be visible as an HH object, the parent GMC must be at least partially disrupted, either by shock fronts encountered while crossing spiral arms, or by internal winds generated by massive protostars. GMCs are generally estimated to have lifetimes of $\sim 10^7$ years. In contrast, the typical age estimated for a protostar once it becomes accessible to optical/infrared observation lies in the range of a few $\times 10^5$ to $\sim 10^6$ years. These age estimates may be compatible, since a stellar core may well spend considerable time continuously accreting material before it is ‘born’ as a protostar. Given these uncertainties, the duration of star formation within a given GMC remains unclear, but probably does not exceed a few million years.

3.6.2 Brown dwarf formation mechanisms

Traditional star formation theories envisage brown dwarfs as the culmination of scaled-down versions of the same process that gives rise to hydrogen-burning stars. As discussed further in Section 9.8, classical models characterise the minimum fragmentation mass possible within a molecular cloud in terms of the Jeans mass,

$$M_J = 0.35(T/10)^2(P/10^6k)^{-1/2}M_{\odot} \quad (3.58)$$

where T is the cloud temperature in Kelvin; P , the core pressure; and k is Boltzmann's constant [E3]. Estimates of the typical pressure in molecular clouds range from $\sim 10^5 k$ in low-density regions, such as Taurus, to $\sim 10^7 k$ in the cores of GMCs. Taking 10 K as the standard gas temperature, those pressures correspond to Jeans masses of 1.1 and $0.11 M_{\odot}$, respectively. The star that emerges as the final product of the core collapse has a mass that is only a fraction of the Jeans mass, perhaps as small as 10%. Even so, this model implies that brown dwarfs, as a class, require lower than average cloud temperatures, while very low mass brown dwarfs ($<0.02 M_{\odot}$) require very cold, very high density environments that are also very scarce among known star-forming regions.

This theoretical result seems at odds with the recent discoveries of numerous brown dwarfs in the general field (see Chapters 6 and 9). One possible means of circumventing the issue is to place less reliance on the Jeans mass as a predictor of individual stellar mass (see Section 9.8). Alternatively, prompted by the apparent paradox, different models for brown dwarf (and low-mass star) formation have been devised. In particular, it has been suggested that brown dwarfs are stellar embryos, ripped untimely from the womb before they could achieve stardom [R5].

The embryo ejection hypothesis has its roots in theories that envisage stars forming through competitive accretion [C7], [B14]. In simple terms, rather than considering stars as forming in splendid isolation (one star per cloud core), these theories postulate formation in small groups of 2 to 4 objects. Each contracting protostar is competing for material from the same reservoir. Gravitation works on the principle that to those that have, more shall be given, so the net result is that more massive initial perturbations grow more rapidly. The separation between individual objects in these small groups is sufficiently small that significant dynamical interactions can occur; in many cases, the brown dwarfs-to-be form within the circumstellar disks of more massive objects [R6]. Simulations show that the most likely consequence of the dynamical interactions is that the lowest mass object is ejected from the group. Once those ejected objects are removed from the denser regions of the core, they are no longer able to acquire significantly more material through accretion. Consequently, if the mass of the ejected object is below the hydrogen burning limit, the embryo will become a brown dwarf [R5], [B15].

The ejection model makes a number of predictions:

- The velocity dispersion of brown dwarfs in star clusters should be higher than the velocity dispersion of the stellar members. The amplification must be

relatively small, however, since an increase of even 2–3 km s^{−1} would remove all brown dwarfs from the Pleiades [M6], contrary to observations (see Section 8.9.2).

- The brown dwarf binary fraction is predicted to be extremely low, <5% [B15]; very low-mass brown dwarf binaries are disrupted in the ejection process, while higher mass systems would have longer residence times, and therefore would be expected to reach stellar masses. This prediction is contradicted by observations, which find a brown dwarf binary fraction exceeding 15% (see Section 6.8.2).
- Many of the isolated brown dwarfs are predicted to have their origin in circumstellar disks [R6]; such objects are predicted to have minimal accretion disks of their own. Moreover, even ejected embryos that form on protostellar cores retain only severely truncated disks after ejection. In fact, infrared observations of isolated brown dwarfs in young clusters find that most have significant infrared excess radiation, consistent with relatively massive accretion disks [L7].

Thus, while the embryo ejection model avoids the Jeans mass problem for brown dwarfs, its predictions are scarcely in agreement with most observations. We return to this issue when discussing theoretical analyses of the initial mass function in Chapter 9.

3.6.3 Young stellar objects in the field

Giant molecular cloud complexes span a wide range of environments, with the extremes being typified by the high star density in the massive Orion star formation region centred on the Trapezium cluster, and the more dispersed, low-mass GMCs in the Taurus complex. The former includes more than 3,000 stars, and will probably evolve to a relatively high-density open cluster, perhaps comparable to the Pleiades, while the individual clouds in the latter system are likely to produce no more than a few dozen stars in total. As discussed further in Chapter 9, there are suggestions that the stellar mass function may vary depending on the environment, with a bias against very low-mass star formation in the lower-mass Taurus-like systems.

Once free of the parent GMCs, low-mass protostars are identifiable as T Tauri stars – pre-main sequence stars with late-type spectra (K or M for masses less than $\sim 1 M_{\odot}$), and emission lines, particularly due to the hydrogen Balmer series (for example, see the spectrum of GG Tau Bb, plotted in Figure 2.12). The systems are usually divided into classical T Tauri stars (cTTs) with strong emission lines due to both chromospheric and accretion disk emission and ‘weak-line’ T Tauri stars (wTTs) which exhibit only chromospheric emission. At this point in their evolution, protostars are fully convective and evolve towards the main sequence along near-vertical (constant temperature) tracks in the H–R diagram (as illustrated by the Lyon models plotted in Figure 3.12). The latter are known as Hayashi tracks, following the first extensive analysis of pre-main sequence stars by Hayashi [H3], [H4]. Since protostars have radii which are significantly larger than

main sequence stars of the same temperature (Figure 3.10), the luminosities are correspondingly higher.

While these protostars are no longer embedded in the high-density gas of the parent GMC, many (all?) T Tauri stars retain circumstellar gaseous disks until ages of several Myr. These are evident through the detection of excess radiation at infrared wavelengths (*JHKL*) over that expected from extrapolating a black-body with a temperature matching the optical observations. Initially, those excesses were interpreted as circumstellar shells of dust grains [M4]. However, the addition of longer-wavelength data showed that the energy distribution required emission from a range of temperatures, and was more consistent with a disk configuration [R3]. In particular, the mass inferred for the molecular gas in many cTTs is sufficient to obscure the central star at visible wavelengths if the gas is distributed in a circumstellar sphere⁵ [A1]. The presence of bipolar outflows in a variety of systems [S12] also pointed to an axisymmetric distribution of gas and dust. HL Tauri was the first system to be resolved, with millimetre observations revealing a 2,000-AU diameter disk-like structure [B6], while Keplerian motion was first identified conclusively in the 600-AU diameter disk surrounding GM Aurigae [K5]. More recently, HST observations have provided additional evidence for disks, both set against the background of the Orion Nebula (so-called protoplanetary disks or ‘proplyds’ – Figure 3.19, colour-plate section) and from direct near-infrared images ([S13], [S14], Figure 3.20, colour-plate section).

Improvements in the spatial resolution and sensitivity of millimetre interferometric arrays have led to a clearer picture of the evolution of circumstellar disks in T Tauri systems [K4]. The disk is believed to form during the initial stages of protostellar collapse, as gas with sufficient angular momentum to avoid falling directly onto the central mass settles into a centrifugally-supported disk, which grows outward with time. Velocity maps of the disk in HL Tau reveal infall over most of the outer regions, with ordered rotation present only within the central 100–200 AU [H5], [K6]. GM Aur, on the other hand, has Keplerian motion throughout, and the lower energies suggest that this is an older system. Bipolar outflows, both collimated jets (Herbig–Haro objects) and molecular flows, are common during these ‘earlier’ stages. The driving mechanism is believed to arise through coupling between the magnetic field (which entrains the outflow) and viscous accretion within the circumstellar disk. The details of that mechanism – notably, the source of viscosity – remain under investigation (see, for example, [A3]).

Approximately 50% of cTTs retain optically-thick circumstellar disks at ages of $\tau \sim 3$ Myr, and a few systems are known with ages of 10 Myr. Mid-infrared observations indicate that some cTTs with optically thick disks have central holes [S10], suggesting that the disk dissipates outwards from the inside, either through more efficient grain agglomeration at high densities or dynamical effects removing particles from the interior orbits. In most cases the disk becomes optically thin by $\tau = 5$ –10 Myr. This sets an upper limit for the formation time of gas-giant planets (as

⁵ The central star is obscured in systems where our view lies long the equatorial plane of the disk.

discussed further in Chapter 11). Particulate material survives for substantially longer periods as ‘debris disks’, identified first through IRAS detection of far-infrared excess radiation from a number of nearby stars, notably Vega [A9]. In a few systems, including β Pic [S11], the disk has been imaged directly through reflected light at optical and infrared wavelengths (see Chapter 11).

Most ‘low-mass’ T Tauri stars studied to date actually have masses that are close to that of the Sun. Many of these systems also prove to be binaries [G1] leading to limitations on the sizes of the circumstellar disks, and sometimes to the formation of circumbinary disks [J1]. It is only within the last few years that observations have extended to sufficiently faint magnitudes to identify a large number of pre-main sequence stars which will evolve to become either M dwarfs or substellar-mass brown dwarfs. The latter are discussed in more detail in Chapter 6.

3.7 POST-MAIN SEQUENCE EVOLUTION AND THE DEATH OF THE MILKY WAY

Low-mass stars have main sequence lifetimes that are orders of magnitude longer than the present age of the Universe (currently estimated as 12–14 Gyr). Little consideration has been given to the later phases of evolution of these objects, perhaps because the prospects of observing an evolved M dwarf are slight. Eventually, however, even $0.1 M_{\odot}$ dwarfs exhaust their central energy cores, cease hydrogen core-burning and evolve off the main sequence. The final stages of their existence – perhaps the ultimate luminous phase of the Universe – have been examined by Laughlin *et al.* [L4], the only detailed study of its kind to date.

The starting point for the [L4] analysis is the set of low-mass main sequence models calculated by Laughlin and Bodenheimer [L3]. The latter have properties close to those of the Tucson [B12] dataset outlined in the earlier sections of this chapter. Figure 3.21 shows the predicted evolution in luminosity and effective temperature of a $0.1 M_{\odot}$ M dwarf. After arriving on the main sequence via the Hayashi track, the luminosity and temperature increase, reflecting the gradual increase in the ^3He fraction and the consequent increased mean molecular weight. At an age of 1.4×10^{12} years (1.4 trillion years), the core temperature rises sufficiently to permit reaction 3^I in the PP chain, ^3He is converted to ^4He and the ^3He mass fraction decreases from its maximum value of 9.96% as equilibrium PPI burning is established (see Figure 3.21, inset).

During the initial 1.4 trillion years, the stellar core expands and reduces ρ_c , reacting to the higher-energy generation rate. However, the hydrogen content dwindles as ^4He becomes the majority constituent, and the core is forced to contract and becomes hotter to maintain ε at a sufficiently high level to preserve hydrostatic equilibrium. The surface temperature increases, reflecting the increase in T_c . At an age of 5.74 trillion years, the hydrogen fraction has been reduced to only 16%, the opacity is low enough to invalidate the convection criterion, and the star develops a radiative core. The stellar radius, which has increased by only $\sim 10\%$ over the main sequence value, contracts, leading to a sharp dip in luminosity.

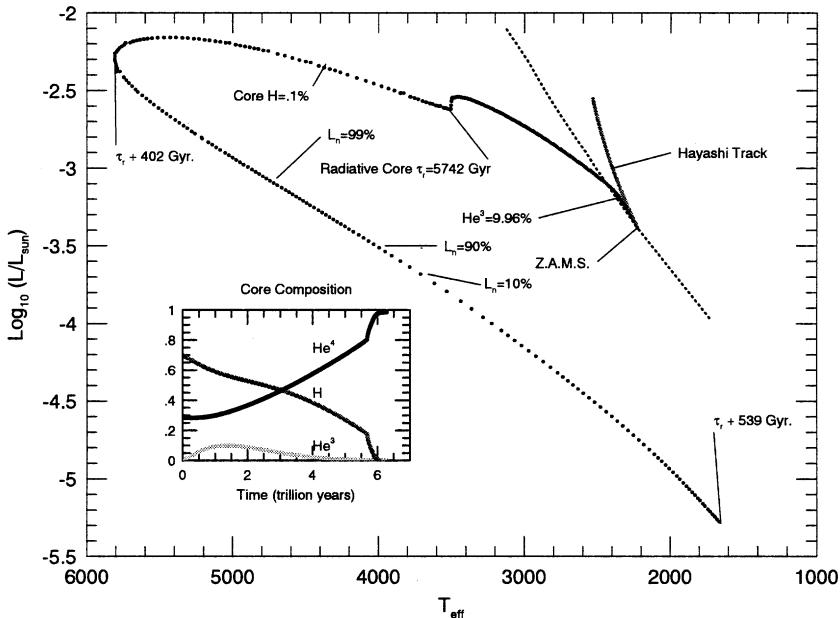


Figure 3.21. The evolution of a $0.1 M_{\odot}$ M dwarf as predicted by Laughlin *et al.* [L4]. The inset diagram shows the relative mass fractions of hydrogen, ^3He and ^4He as a function of time. (Courtesy of G. Laughlin and the *Astrophysical Journal*.)

Once the radiative core is established, the $0.1 M_{\odot}$ post-MS dwarf mimics solar-type stars in evolving through hydrogen core-burning and shell-burning phases, although in this case the initial hydrogen mass-fraction is only 16%, and timescales are measured in tens of Gyr. The surface temperature and luminosity continue to increase as the star is forced to contract even further to maintain ε despite the depleted fuel supply. T_c never comes close to the temperatures required for helium burning. Degeneracy becomes the dominant source of pressure support, conduction becomes the main form of energy transport, and the luminosity eventually starts to decline.

The $0.1 M_{\odot}$ star reaches a maximum surface temperature of $\sim 5,800$ K (comparable with the present-day Sun) at an age of nearly 6 trillion years. The hydrogen-burning shell is still active at that point, and continues to supply most of the luminosity over the succeeding ~ 80 Gyr. Figure 3.21 identifies the locations where the fractional contribution to the luminosity due to nuclear burning, L_N , reaches 99%, 90% and 10%. ε continues to decrease after passing the temperature maximum, as does the radius and hence the luminosity. All nuclear fusion eventually ceases when the dwarf cools to temperatures below 2,000 K as a helium white dwarf. Laughlin *et al.* note that the predicted final composition is 99% helium and 1% hydrogen (by mass) – an extremely efficient use of the initial fuel source.

Figure 3.22 shows comparable evolutionary tracks for a range of masses from a $0.06 M_{\odot}$ brown dwarf to a $0.25 M_{\odot}$ star with a main sequence lifetime of merely

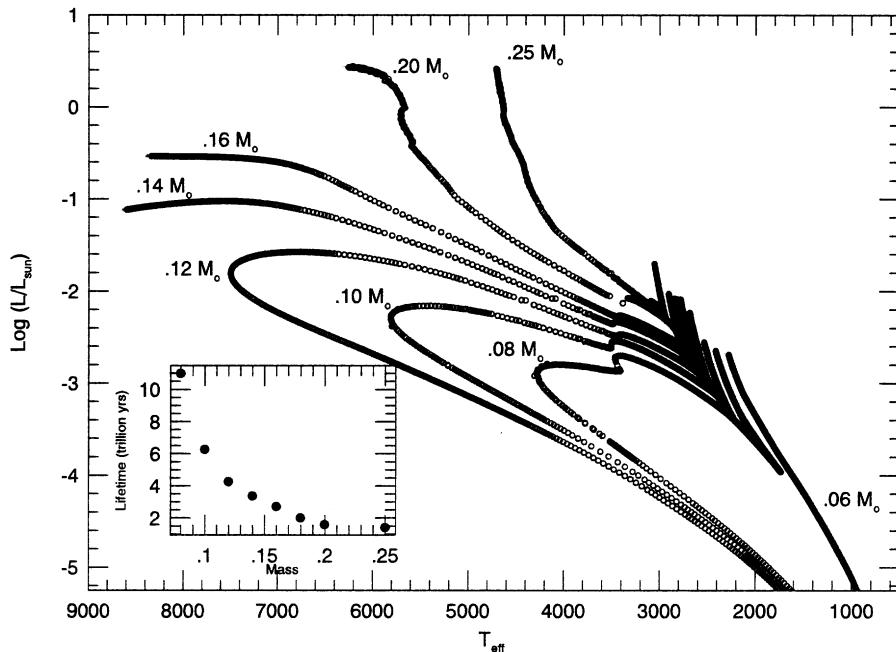


Figure 3.22. Evolutionary tracks for low-mass dwarfs. The inset diagram indicates the main sequence lifetime as a function of mass. (Courtesy of G. Laughlin and the *Astrophysical Journal*.)

~ 1 trillion years. The brown dwarf shows the continuous decline in (L, T_{eff}) that is expected given the absence of any hydrogen fusion; stars with masses $\leq 0.16 M_{\odot}$ evolve in a manner similar to that already described for the $0.1 M_{\odot}$ dwarf, although the radiative core develops at a progressively earlier stage with increasing mass, leading to a higher mass fraction of hydrogen in the envelope of the final degenerate.

The 0.2 and $0.25 M_{\odot}$ dwarfs, however, follow a somewhat different evolutionary path towards the end of their post-main sequence lifetimes. In low-mass dwarfs, T_c decreases after core degeneracy sets in; in higher-mass objects, T_c continues to increase, fuelled by more active hydrogen shell-burning. This added energy drives the radius to values exceeding $1 R_{\odot}$ – more than five times the main sequence radius. These stars ascend a giant branch, albeit as yellow(?) giants, paradoxically hotter than their main sequence progenitors.

The timescales involved in all of these calculations are extremely long: the full evolution of a $0.1 M_{\odot}$ star from protostar to white dwarf exceeds 400 Hubble times. Several interesting consequences are highlighted by [L4]. First, the evolved 0.14 and $0.16 M_{\odot}$ dwarfs spend several Gyr at luminosities of ~ 0.1 – $0.3 L_{\odot}$ and effective temperatures higher than that of the Sun. These periods of time are similar to the present age of the Solar System. On that basis, [L4] suggest that this might provide

an opportunity for Earth-like life to develop in previously frigid terrestrial planetary companions at the appropriate distance from the post-MS ‘M dwarfs’. As discussed in Chapter 11, jovian-mass planets have already been discovered around the nearby $\sim 0.25 M_{\odot}$ M dwarf, Gl 876.

The Galaxy inhabited by these hypothetical lifeforms would be very different from our own. Setting aside the possibility that the mass density in the Universe might exceed the critical value and provoke a re-collapse, most of the available baryonic matter would be locked up in low-mass stars or stellar remnants. As a result, little interstellar gas and dust would be expected, resulting in insignificant continuing star-formation. This, in turn, would lead to a dearth of intermediate- and high-mass stars. The highest-temperature stellar objects visible might well be F-type post-MS low-mass dwarfs, although the sheer number of such stars should serve to maintain the total luminosity of the Milky Way for ~ 10 trillion years. Eventually, however, the Galaxy will consist of only degenerate stellar remnants, with temperatures of ~ 60 K maintained by dark matter interactions [A2], producing a total luminosity (summed over the entire Milky Way) of only $1 L_{\odot}$.

3.8 SUMMARY

The basic principles of stellar structure were established during the middle years of the twentieth century. It is only within the last decade, however, that it has become possible to compute reliable models for very low-mass stars and brown dwarfs. We have reviewed the development of those models in this chapter. Starting with the four fundamental equations of stellar structure, the Lane–Emden equations for gas polytropes, index $n = 1.5$, allow an approximate description of the properties of fully-convective low-mass stars and brown dwarfs. Accurate analyses, however, demand more detailed calculations – notably, an equation of state which takes into account non-linear effects due to pressure ionisation, Coulomb interactions and, in particular, increasing degeneracy.

Degeneracy forces star-like objects with masses below $\sim 0.1 M_{\odot}$ to maintain a nearly constant radius of $\sim 0.1 R_{\odot}$, with a corresponding decrease in the central temperature with decreasing mass. Once that temperature drops below the critical value for initiating hydrogen fusion, the only energy sources are short-lived deuterium fusion and contraction, and the luminosity and surface temperature decrease rapidly with time. Current models place the boundary between low-mass stars and brown dwarfs at a mass of $\sim 0.075 M_{\odot}$. A comparison between the latest models and observations of low-mass stars and brown dwarfs reveals that, while there is good agreement in the predicted luminosities, discrepancies in the effective temperature scale remain. The models are typically 200 K hotter than the observed sequence for spectral types between M1 and M8, with corresponding offsets in the theoretical and empirical mass–radius relationships. Finally, we have provided an overview of current theories of the formation and evolution of low-mass dwarfs.

3.9 REFERENCES

- A1 Adams, F. C., Emerson, J. P., Fuller, G. A., 1990, *ApJ*, **357**, 606.
- A2 Adams, F. C., Laughlin, G., 1997, *Rev. Mod. Phys.*, **69**, 337.
- A3 Adams, F. C., Lin, D. N. C., 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine, Tucson, University of Arizona Press, p. 721.
- A4 Adams, W. S., 1914, *PASP*, **27**, 236.
- A5 Adams, W. S., 1925, *Proc. Nat. Acad. Sci.*, **11**, 382.
- A6 Allard, F., Hauschildt, P. H., 1995, *ApJ*, **445**, 433.
- A7 Andersen, J., 1991, *Astr. Ap. Rev.*, **3**, 91.
- A8 Auman, J. J., 1967, *APJS*, **14**, 171.
- A9 Aumann, H. H., Gillett, F. C., Beichman, C. A., de John, T., Houck, J. R., Low, F. J., Neugebauer, C., Walker, R. G., Wesselius, P. R., 1984, *ApJ*, **278**, L23.
- B1 Bahcall, J. N., 1989, *Neutrino Astrophysics*, Cambridge University Press.
- B2 Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P. H., 1995, *ApJ*, **446**, L35.
- B3 Baraffe, I., Chabrier, G., Plez, B., 1996, *ApJ*, **459**, L91.
- B4 Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P. H., 1997, *A&A*, **327**, 1054.
- B5 Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P. H., 1998, *A&A*, **337**, 403.
- B6 Beckwith, S., Sargent, A. I., Scoville, N. Z., Masson, C. R., Zuckerman, B., Phillips, T. G., 1986, *ApJ*, **309**, 755.
- B7 Blitz, L., Thaddeus, P., 1980, *ApJ*, **241**, 676.
- B8 Böhm-Vitense, E., 1958, *Zs. f. Ap.*, **46**, 108.
- B9 Burbidge, E. M., Burbidge, G. R., Fowler, W. A., Hoyle, F., 1957, *Rev. Mod. Phys.*, **29**, 547.
- B10 Burrows, A., Hubbard, W. B., Lunine, J. I., 1989, *ApJ*, **345**, 939.
- B11 Burrows, A., Liebert, J., 1993, *Rev. Mod. Phys.*, **66**, 301.
- B12 Burrows, A., Hubbard, W. B., Saumon, D., Lunine, J. I., 1993, *ApJ*, **406**, 158.
- B13 Burrows, A., Marley, M., Hubbard, W. B., Lunine, J. I., Guillot, T., Saumon, D., 1997, *ApJ*, **491**, 856.
- B14 Bonnell, I. A., Bate, M. R., Clarke, C. J., Pringle, J. E., 1997, *MNRAS*, **285**, 201.
- B15 Bate, M. R., Bonnell, I. A., Bromm, V., 2002, *MNRAS*, **332**, L65.
- C1 Canuto, V. M., Goldman, I., Mazzitelli, I., 1996, *ApJ*, **473**, 550.
- C2 Chandrasekhar, S., 1931, *ApJ*, **74**, 81.
- C3 Chandrasekhar, S., 1939, *An Introduction to the Study of Stellar Structure*, Dover Publications.
- C4 Clayton, D. D., 1983, *Principles of Stellar Evolution and Nucleosynthesis*, University of Chicago Press.
- C5 Clemens, J. C., Reid, I. N., Gizis, J. E., O'Brien, M. S., 1998, *ApJ*, **496**, 352.
- C6 Copeland, H., Jensen, J. O., Jørgensen, H. E., 1970, *A&A*, **5**, 12.
- C7 Clarke, C., 1998, *The Stellar Initial Mass Function*, ASP Conf. Ser., **142**, 189.
- D1 D'Antona, F., Mazzitelli, I., 1983, *A&A*, **127**, 149.
- D2 D'Antona, F., Mazzitelli, I., 1985, *ApJ*, **296**, 502.
- D3 D'Antona, F., Mazzitelli, I., 1994, *ApJS*, **90**, 467.
- D4 D'Antona, F., Mazzitelli, I., 1996, *ApJ*, **456**, 329.
- D5 Dorman, B., Nelson, L. A., Chau, W. Y., 1989, *ApJ*, **342**, 1003.
- E1 Eddington, A. S., 1926, *The Internal Constitution of the Stars*, Cambridge University Press.
- E2 Emden, V. R., 1907, *Gaskugeln*, Teubner.

- E3 Elmegreen, B. G., 1999, *ApJ*, **522**, 915.
- F1 Fowler, R. H., 1927, *MNRAS*, **87**, 114.
- G1 Ghez, A. M., Neugebauer, G., Matthews, K., 1993, *AJ*, **106**, 2005.
- G2 Gizis, J. E., 1997, *AJ*, **113**, 806.
- G3 Goldberg, H. S., Scadron, M. D., 1981, *Physics of Stellar Evolution and Cosmology*, Gordon & Breach Publishers.
- G4 Greenstein, J. L., Oke, J. B., Shipman, H., 1985, *Quart. J. R. astr. Soc.*, **26**, 279.
- G5 Grossman, A. S., 1970, *ApJ*, **161**, 619.
- G6 Grossman, A. S., Graboske, H. C., 1971, *ApJ*, **164**, 475.
- G7 Grossman, A. S., Graboske, H. C., 1973, *ApJ*, **180**, 194.
- G8 Grossman, A. S., Hays, D., Graboske, H. C., 1974, *A&A*, **30**, 95.
- H1 Hansen, C. J., Kawaler, S. D., 1994, *Stellar Interiors*, Springer-Verlag, New York, Berlin.
- H2 Hartmann, L., 1998, *Accretion Processes in Star Formation*, Cambridge Astrophysics Series, No. 32, Cambridge University Press.
- H3 Hayashi, C., 1961, *PASJ*, **13**, 450.
- H4 Hayashi, C., 1966, *ARA&A*, **4**, 171.
- H5 Hayashi, M., Ohashi, N., Miyama, S. M., 1993, *ApJ*, **418**, L71.
- H6 von Helmholz, H., 1854, Kant commemoration lecture, Königsberg.
- H7 Henry, T. J., McCarthy, D. W., 1993, *AJ*, **106**, 773.
- H8 Henry, T. J., Franz, O. G., Wasserman, L. H., Benedict, G. F., Shelus, P. J., Ianna, P. A., Kirkpatrick, J. D., McCarthy, D. W., 1999, *ApJ*, **512**, 864.
- H9 Hutton, J., 1788, *Theory of the Earth*, Edinburgh.
- J1 Jensen, E. L. N., Koerner, D. W., Mathieu, R. D., 1996, *AJ*, **111**, 2431.
- K1 Lord Kelvin (W. Thomson), 1862, *Collected Papers*, **3**, 255.
- K2 Lord Kelvin (W. Thomson), 1887, *Phil. Mag., ser. 5*, **23**, 287.
- K3 Lord Kelvin (W. Thomson), 1897, *Collected Papers*, **5**, 205.
- K4 Koerner, D. W., 1997, *Origins Life Ev. Bio.*, **27**, 157.
- K5 Koerner, D. W., Sargent, A. I., Beckwith, S. V. W., 1993, *Icarus*, **106**, 2.
- K6 Koerner, D. W., Chandler, C., Sargent, A. I., 1995, *ApJ*, **452**, L69.
- K7 Kumar, S. S., 1963a, *ApJ*, **137**, 1121.
- K8 Kumar, S. S., 1963b, *ApJ*, **137**, 1126.
- K9 Kippenhahn, R., Weigert, A., 1994, *Stellar Structure and Evolution*, Springer-Verlag.
- L1 Lacy, C., 1977, *ApJ*, **218**, 444.
- L2 Lane, J. H., 1869, *Amer. J. Sci.*, 2nd ser., **50**, 57.
- L3 Laughlin, G., Bodenheimer, P., 1993, *ApJ*, **403**, 303.
- L4 Laughlin, G., Bodenheimer, P., Adams, F. C., 1997, *ApJ*, **482**, 420.
- L5 Leggett, S. K., Allard, F., Berriman, G., Dahn, C. C., Hauschildt, P. H., 1996, *ApJS*, **104**, 117.
- L6 Limber, D., 1958, *ApJ*, **127**, 387.
- L7 Liu, M. C., Najita, J., Tokunaga, A. T., 2003, *ApJ*, **585**, 372.
- M1 Maeder, A., Meynet, G., 1989, *A&A*, **210**, 155.
- M2 Magazzu, A., Martin, E. L., Rebolo, R., 1992, *ApJ*, **392**, 159.
- M3 Mazzitelli, I., Moretti, M., 1980, *ApJ*, **235**, 955.
- M4 Mendoza, E. E., 1966, *ApJ*, **143**, 1080.
- M5 Monet, D. G., Dahn, C. C., Vrba, F. J., Harris, H. C., Pier, J. A., Luginbuhl, C. B., Ables, H. D., 1992, *AJ*, **103**, 638.
- M6 Moraux, E., Clarke, C., 2004, *A&A*, **429**, 895
- O1 O'Dell, C. R., Wen, Z., Hu, X. H., 1993, *ApJ*, **410**, 696.

- O2 Osterbrock, D. E., 1953, *ApJ*, **118**, 520.
- P1 Payne, C. H., 1925, Stellar Atmospheres, *Harvard Coll. Obs. Monographs*, No. 1.
- P2 Popper, D. M., 1980, *ARA&A*, **18**, 115.
- P3 Pringle, J. E., 1981, *ARA&A*, **19**, 137.
- R1 Ritter, A., 1878, *Weidemann Annalen*, **5**, 543.
- R2 Russell, H. N., Adams, W. S., Moore, C. E., 1928, *ApJ*, **68**, 1.
- R3 Russell, H. N., 1929, *ApJ*, **70**, 11.
- R4 Rydgren, A. E., Cohen, M., 1985, in *Protostars and Planets II*, ed. D. C. Black and M. S. Matthews, Tucson, University of Arizona Press, p. 371.
- R5 Reipurth, B., Clarke, C., 2001, *AJ*, **122**, 432.
- R6 Rice, W. K. M., Armitage, P. J., Bonnell, I. A., Bate, M. R., Jeffers, S. V., Vine, S. G., 2003, *ApJ*, **346**, L36.
- S1 Sampson, W., 1894, *Mem. RAS*, **51**, 123.
- S2 Sargent, A. I., 1977, *ApJ*, **218**, 736.
- S3 Sargent, A. I., Welch, W. J., 1993, *ARA&A*, **31**, 297.
- S4 Saumon, D., Chabrier, G., 1991, *Phys. Rev. A*, **44**, 5122.
- S5 Saumon, D., Chabrier, G., Van Horn, H. M., 1995, *ApJS*, **99**, 713.
- S6 Schwarzschild, K., 1906, *Gott. Nacht.*, **41**.
- S7 Schwarzschild, M., 1958, *Structure and Evolution of the Stars*, Dover Publications.
- S8 Sienkiewicz, R., 1982, *Acta Astr.*, **32**, 275.
- S9 Shu, F. H., Adams, F. C., Lizano, S., 1987, *ARA&A*, **25**, 23.
- S10 Skrutskie, M. F., Dutkiewitch, D., Strom, S. E., Edwards, S., Strom, K. M., Shure, M. A., 1990, *AJ*, **99**, 1187.
- S11 Smith, B. A., Terrile, R. J., 1984, *Science*, **226**, 1421.
- S12 Snell, R. L., Loren, R. B., Plambeck, R. L., 1980, *ApJ*, **239**, L17.
- S13 Stapelfeldt, K. R. *et al.*, 1998a, *ApJ*, **502**, L65.
- S14 Stapelfeldt, K. R. *et al.*, 1998b, *ApJ*, **508**, 736.
- S15 Strömgren, B., 1932, *Zs. f. Ap.*, **4**, 118.
- S16 Strömgren, B., 1933, *Zs. f. Ap.*, **7**, 222.
- S17 Strong, A. W. *et al.*, 1988, *A&A*, **207**, 1.
- S18 Stahler, S. W., Palla, F., 2004, *The Formation of Stars*, Wiley-VCH, Germany.
- T1 Tayler, R. J., 1954, *ApJ*, **120**, 332.
- T2 Tytler, D., Fan, X. M., Burles, S., 1996, *Nature*, **381**, 207.
- V1 Vandenberg, D. A., Hartwick, F. D. A., Dawson, P., Alexander, D. R., 1983, *ApJ*, **266**, 747.
- V2 Vandenberg, D. A., 1992, *ApJ*, **391**, 685.
- Z1 Zapolsky, H. S., Salpeter, E. E., 1969, *ApJ*, **158**, 809.

3.10 HOMEWORK PROBLEMS

Problem 1

Show that if non-relativistic degeneracy is important, the pressure and density are related by $P \propto \rho^{5/3}$. What polytropic index does this correspond to? Thus, polytropes with this index can be used to approximate the interior structure analytically, without having to use detailed numerical models.

Problem 2

Fill in the following table, using information from this chapter and references cited therein (especially [C4], pp. 88–89), for dwarfs with an age of 1 Gyr.

Spectral type	Mass	Radius	$T_{central}$	$\rho_{central}$	$\rho T^{-3/2}$
M0					
M4					
M8					
L2					
L5					
L8					

Degenerate electron pressure exceeds non-degenerate electron pressure when

$$\rho/\bar{\mu}_e > 2.4 \times 10^{-8} T_c^{3/2}$$

Which of the objects in the above table can be classified as mostly degenerate? Which particular object is the ‘most’ degenerate and why? (Note, this illustrates the same principle that applies to white dwarfs, namely that the largest mass white dwarfs have the smallest radius and hence are the ‘most degenerate’, see Section 16.1, pp. 206–210 of *Introduction to Stellar Astrophysics* (Volume 3: Stellar Structure and Evolution) by Erika Böhm-Vitense.)

Problem 3

Draw a picture and argue qualitatively that a layer is unstable to convection if

$$|dT/dr|(\text{star}) > |dT/dr|(\text{ad})$$

(see [C4], sections 3–5 for additional discussion). In practice the temperature gradient is taken to be the adiabatic gradient in convective regions, since the excess is very small. Note that convective regions have a polytropic index $n = 1.5$, and polytrope models can therefore be used to approximate the interior structure.

Now consider the discussion in Section 3.3.4, and verify that the conditions for convection depend on flux, opacity, gravity and specific heat at constant pressure as claimed. Be sure to distinguish explicitly between $|dT/dr|_{\text{star}}$ and $|dT/dr|_{\text{ad}}$. Which factors are most important for driving convection in low-mass dwarfs?

Problem 4

You observe two star clusters and find numerous M6 dwarfs in each. In cluster A, the M6 dwarfs have Li I 6,707 Å in absorption. In cluster B, the M6 dwarfs show no lithium absorption. What can you infer about the ages of the clusters and the masses of the M6 dwarfs in each cluster?

4

The photosphere

4.1 INTRODUCTION

The photosphere is the surface layer of a star that emits the radiation that we observe photometrically and spectroscopically. Due to the high gravity and pressure in the M dwarf atmosphere, the photosphere is remarkably thin, extending a mere 100–200 km into the star. Nevertheless, it performs the crucial tasks of moderating the energy flow from the nuclear reactions within, regulating the surface temperature and luminosity, and producing the line and continuum radiation that we observe. Many M dwarfs show evidence of an additional, tenuous, outer atmosphere (chromosphere, transition region and corona) located above the photosphere, probably produced by heating in regions of strong magnetic fields. Its presence may well modify the photospheric behaviour in some subtle, or even critical, fashion; however, we defer discussion of this region to Chapter 5, and concentrate here on the photosphere as a separate entity.

Analysis of the photosphere is accomplished using numerical models together with observations of the emergent spectrum. The observations often consist only of broadband colours, although a detailed, high-resolution spectrum provides much more information and consequently more insight into the atmospheric structure. With the twin tools of models and observations, it is possible to determine the temperature, gravity and metallicity of the star, and to infer such quantities as the mass, radius and luminosity.

The history of stellar atmospheres is rich with the insight of many of the great astronomers of the early twentieth century. Hearnshaw [H3] provides a detailed historical account from which we summarise the principal methods and results as background for the understanding of the modern analysis. It is our intention to provide the reader with the tools necessary to understand and evaluate current photospheric modelling efforts, and to then present the results together with observations and interpretation. We make no attempt to give a full explanation of the

process by which modern numerical models are produced. Several excellent treatises are available which treat this difficult and complex problem in the detail it deserves. See, for example, the standard text by Mihalas [M3].

4.2 HISTORICAL PERSPECTIVE

In the early 1800s, Wollaston observed dark lines superposed on the continuous spectrum of colours formed by passing sunlight through a narrow slit, and then through a prism. Fraunhofer subsequently mapped the lines, named the strongest by letter designations still in use today (for example the D lines, now known to be sodium resonance lines, and the H and K lines, now identified as resonance lines of ionised calcium), and measured the wavelengths of several hundred lines. Understanding the formation of those dark lines, and using them to interpret the physical conditions at the surface of the star, comprises the field of stellar atmospheres. This work occupied astronomers for much of the first half of the twentieth century; among the triumphs – well before the age of modern computers – were the identification of temperature as the fundamental parameter determining the spectral type of a star; the description of gravity differences between the ‘bright, narrow-lined’ (giant) and ‘faint, broad-lined’ (dwarf) stars of the same spectral type; and the realisation that the solar composition of elements was shared by almost all stars, with hydrogen being the most abundant element by several orders of magnitude.

Advances in theoretical physics in the early part of the twentieth century, including the discovery of the photoelectric effect and the development of quantum mechanics, led to the concept of quantised energy levels that were available for an electron to occupy within an atom, and paved the way for the interpretation of the dark lines as the absorption of a photon with the particular energy necessary to excite an electron to a higher energy level. Figure 4.1 illustrates a schematic energy-level diagram for the first three bound levels plus continuum of the hydrogen atom. The excitation potential, ϕ_{ij} , is the energy required to excite an electron from level i to level j , and is related to wavelength by $\phi_{ij} = hc/\lambda$, where h is the Planck constant and c is the speed of light.

Boltzmann showed that, in thermodynamic equilibrium, the relative populations in the levels i , j are given by:

$$\frac{n_i}{n_j} = \frac{g_j}{g_i} e^{-\phi_{ij}/kT} \quad (4.1)$$

where k is the Boltzmann constant. Equation (4.1) allows the calculation of the number of atoms in each state if the temperature T , the statistical weights of the levels, g_i and g_j , and the excitation potential, ϕ_{ij} , are known. Alternatively, if the number of atoms in each state can be estimated from observations of an absorption line, the temperature can be determined. Saha extended the study of level populations to include ionisation (that is, transitions from a bound level to the continuum, first discussed in detail by Eddington several years earlier), and

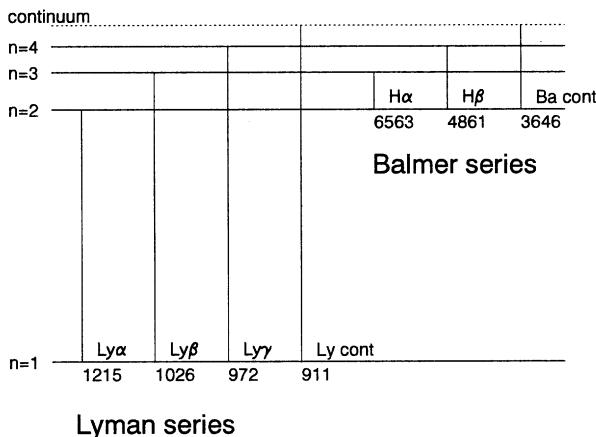


Figure 4.1. Schematic of energy levels of the hydrogen atom, illustrating bound–bound (absorption) and bound–free (ionisation) transitions in the Lyman and Balmer series. The wavelengths in Ångstroms of the transitions are indicated.

deduced the Saha equation:

$$\frac{N^+ n_e}{N} = \frac{u^+}{u} \frac{2(2\pi m)^{3/2}}{h^3} (kT)^{3/2} e^{-\chi_{ion}/kT} \quad (4.2)$$

where N^+ and N are the number densities of ions and neutrals, respectively, n_e is the electron density, u^+ and u are partition functions (similar to the statistical weights, g , in the Boltzmann equation) and χ_{ion} is the ionisation potential.

The Saha equation was instrumental in understanding the differences between the ‘bright, narrow-lined and redder’ stars of the same spectral type as the more common solar-like stars. Pannekoek reasoned that stars of the same spectral type must have nearly the same temperature, and hence the brighter stars must be much larger (giants). (He implicitly assumed that the masses were similar, which was later confirmed by binary star measurements.) A larger size implies a lower atmospheric density in the absorbing layer, and hence a higher degree of ionisation at a given temperature, since $N^+/N \propto n_e^{-1}$. Spectroscopic analysis thus gave a lower temperature for a giant of the same spectral type as a dwarf, in agreement with the observations which showed that the giants were redder. The narrower lines in the giants are a result of reduced pressure (collisional) broadening, which affects the line profile (see below).

Armed with these theoretical tools, and a ‘stellar atmosphere’ which consisted of a continuum source overlaid by an absorbing layer of constant temperature and pressure, Payne used the extensive spectroscopic data available at Harvard to show that the changes in the strength and location of stellar absorption lines for stars of different spectral type could be attributed to the changing excitation and ionisation of different atomic species as a function of temperature in the absorbing layer. The most striking example is the appearance of the hydrogen Balmer lines, which comprise transitions from the $n = 2$ excited level of hydrogen (see Figure 4.1),

and which are visible in the optical spectrum as the H α line at 6,563 Å, the H β line at 4,861 Å, and the H γ line at 4,340 Å. It was well known that these lines reached their maximum strength in stars of spectral type A (this is, in fact, the basis of the ‘A’ designation), becoming weaker both towards the hotter O stars and the cooler M stars. Payne explained this as the trade-off between having a large neutral hydrogen population in the ground state (in the cool stars), and having hydrogen being mostly ionised (in the hot stars), leading to a maximum in the number of hydrogen atoms having electrons in the first excited state at an intermediate temperature of about 10,000 K. The application of this balancing act between excitation and ionisation to a variety of atomic species that were sensitive to different temperatures, allowed her to quantitatively define a temperature scale for all spectral types (though the number of M dwarfs was small, and confined to the earliest spectral types due to the faintness of the objects).

In addition, Payne compared the line strengths of various atomic species in numerous stars of the same temperature, and found that most stars had similar values. Thus compositional differences were not common, with the majority of stars sharing the solar distribution of metallicity. Further, she found that the strength of the hydrogen lines relative to the metal lines indicated a hydrogen abundance that was several orders of magnitude larger than the metals – a result which would later prove to be accurate, though it was mistrusted at the time.

A final important piece of background physics was the description of the line profile, which reflects the probability for photon absorption as a function of wavelength. The Lorentz damping profile, shown in Figure 4.2, arises from the Heisenberg uncertainty principle. The energy levels in an atom are not infinitely

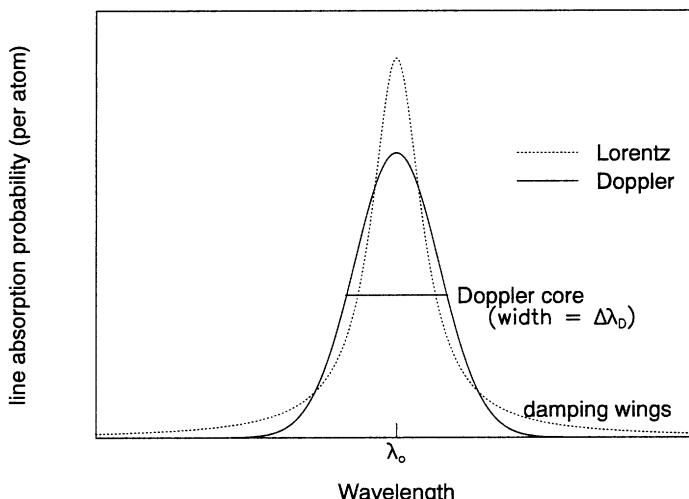


Figure 4.2. The line absorption profiles due to random thermal motions (Doppler) and finite energy width (Lorentz). Pressure broadening also has a Lorentz profile. The resulting Voigt profile is not a simple sum of the component profiles, although the properties of the Doppler core and damping wings are preserved.

sharp; each has a finite energy width ΔE and lifetime Δt such that $\Delta E \Delta t = \hbar$. Transitions with shorter lifetimes, such as resonance transitions, have large transition probabilities and large energy widths, leading to broader profiles, which are characterised by the damping constant γ ($\gamma \propto$ transition probability). The width of the Lorentz profile is typically only 10^{-4} Å, but the broad *damping wings* of the Lorentz profile can make an important contribution to the opacity at wavelengths far from the line centre. For M dwarfs, we shall see that the Lorentz profile resulting from van der Waals interactions between neutral atoms and molecules will dominate the line profile except at the line centre.

A second important effect on the line profile, which usually dominates at wavelengths near the centre, or *Doppler core*, of the line, is a result of random motion of the atoms in the atmosphere leading to a Doppler effect on the wavelength at which the atom can absorb a photon. The wavelength is modified by an amount $\Delta\lambda_D$, such that $\Delta\lambda_D/\lambda = v/c$, where the random velocity, v , depends on the mass of the atom, m , and the temperature, T ; $v = \sqrt{2kT/m}$. An M dwarf with a temperature of 3,000 K has a Doppler width (for hydrogen) $\Delta\lambda_D \sim 7 \text{ km s}^{-1}$, corresponding to 0.15 Å at H α .

The Lorentz and Doppler profiles are shown schematically in Figure 4.2, which plots the absorption coefficient, κ (the probability for photon absorption), as a function of wavelength. The combined (Voigt) profile consists of a Doppler core and Lorentz damping wings, and may be calculated analytically for any transition with known atomic properties, at a given gas temperature (see Böhm-Vitense [B4]; Mihalas [M3]).

In the late 1920s, Minnaert introduced the concept of the line equivalent width to quantitatively measure the total absorption in the line. The equivalent width is the width in Å of a perfectly dark line that subtends the same area under the stellar continuum as the actual line (see Figure 4.3). He then showed that a plot of equivalent width (EW) against the number (N) of absorbing atoms had a characteristic shape that was nearly universal. This famous diagram is known as the ‘curve of growth’, and is shown schematically in Figure 4.4. The curve of growth divides into three parts, which Schutz explained in terms of the line profile. When the line is ‘optically thin’ ($\tau < 1$; see Section 4.3), each atom added to the layer will absorb radiation, hence the EW rises linearly with the number of atoms for weak lines. In the flat part of the curve of growth, the core of the line, where the probability of absorption is high, has saturated; every available photon of that wavelength is already being absorbed, so the addition of more atoms does not change the amount of absorption. Thus the EW barely changes as N increases. Finally, when there are enough atoms present, the probability of being absorbed in the damping wings becomes significant and the EW begins to rise as the square root of N (the square root dependence comes from the explicit expression for the damping profile, see [B4], [M3]). Since the number of atoms in the absorbing layer of the star cannot be varied, the curve-of-growth method is applied by observing the equivalent widths of a distribution of weak and strong lines (corresponding to low and high transition probabilities, respectively), produced by a given atomic species such as Fe I. The curve of growth can then be calibrated in absolute units using solar measurements.

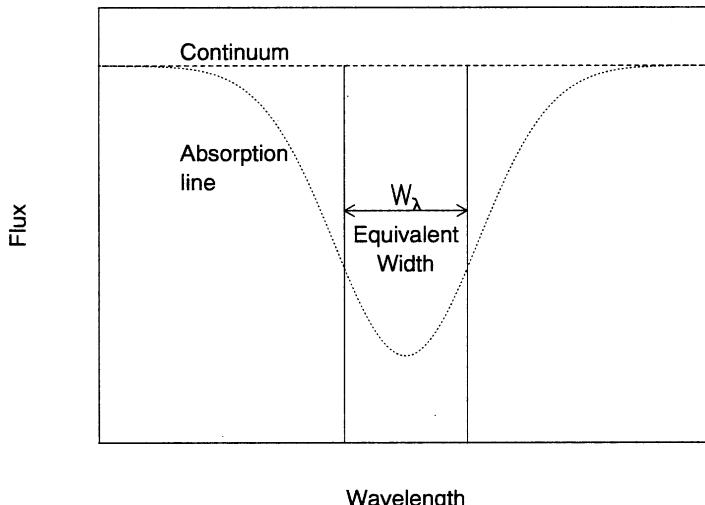


Figure 4.3. The area under the continuum of the absorption line (dotted) is the same as the area within the rectangle of width W_{λ} (the equivalent width) and height equal to the continuum flux.

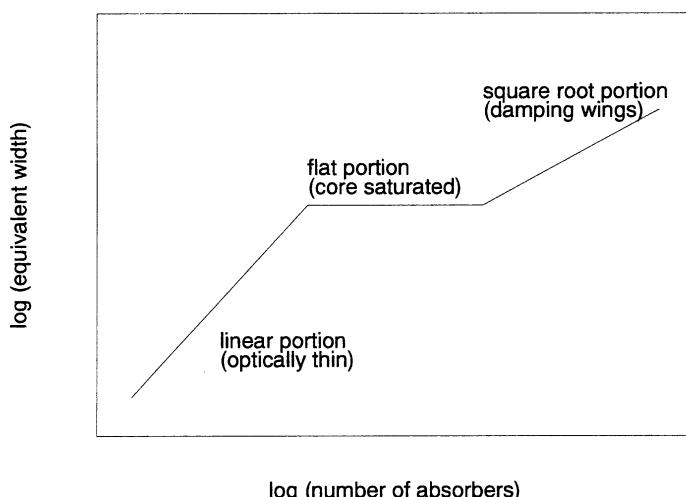


Figure 4.4. A schematic curve of growth, showing the progression in equivalent width that an absorption line would experience as the number of atoms increases. In practice, one observes many lines of the same species with varying strength to map the curve of growth (it being impossible to vary the number of atoms on a star many parsecs away!).

The curve-of-growth method remained the principal tool for abundance analysis until well into the latter half of the twentieth century.

In the late 1930s a final important puzzle was solved by Wildt, who showed that the principal source of continuous opacity in the Sun and other cool stars (including

the M dwarfs) is the H^- ion. It had been known for some time that stars did not radiate as perfect black-bodies, since the Balmer jump was easily observable in hot stars. In particular, colour temperatures determined by matching observed flux gradients to a black-body spectrum were in conflict with the ionisation temperatures found from analysis of spectral lines, with the colour temperatures being generally hotter. A source of continuous opacity was required in the optical wavelength region to bring the colour temperatures into agreement. It was thought by many that this opacity was due to a large number of metal transitions providing a ‘line blanketing’ effect. The first crude model atmospheres were being developed in the 1930s; Biermann and Unsöld had independently found that an atmosphere comprised of 2/3 metals and 1/3 hydrogen could provide the needed opacity. This result was sharply at odds with that of Payne, and later Russell, who had found hydrogen to be thousands of times more abundant than the metals.

The opacity due to the H^- ion, which consists of an electron loosely bound to a neutral hydrogen atom, resolved both the temperature and abundance discrepancies. The second electron can be ionised by photons with $E > 0.7 \text{ eV}$, or $\lambda < 1.7 \mu\text{m}$. Thus any optical or near-infrared photon can ionise an H^- ion, providing a source of continuous opacity which depends on the hydrogen abundance (rather than requiring a large metal abundance).

The period from 1940 to the mid-1960s was marked by increasing sophistication of the atmospheric models for solar and hotter stars, including particularly the first approximate description of convection and its application to hot stars by Böhm and Böhm-Vitense. Little work was done on the M dwarfs, both for lack of observational material and because of the daunting task of describing the atomic and molecular opacity sources. We shall pick up the modelling discussion again in Section 4.5, after first reviewing the methods by which models are produced, and taking a closer look at the opacity sources. The complexity and sheer volume of the modern model results is rather daunting. They should be perused while keeping in mind the cautionary words of Armin Deutsch, who in 1966 wrote in a paper entitled *Even Simpler Methods of Abundance Determination from Stellar Spectra* [D1]: ‘Of course, we must recognise that these methods produce results that are rough and not fully reliable. But this is preferable to the delusion that we can improve our results by adducing an inapplicable model, however sophisticated; or by processing irrelevant data, in quantities however vast.’

4.3 THEORETICAL CONSIDERATIONS

To understand the nature of the photosphere and how its structure is determined, we begin with some basic theoretical concepts. Additional material may be found in the introductory text of Böhm-Vitense [B4], while a more detailed and rigorous treatment is contained in Mihalas [M3]. Note that some of the equations in the following discussion have been presented in previous chapters; they are gathered here for clarity.

The photosphere may be characterised in its simplest form by three quantities: the effective temperature, T_{eff} , the surface gravity, g , and the mean metallicity, Z . The effective temperature is defined by:

$$\sigma T_{\text{eff}}^4 / \pi = F \quad (4.3)$$

where F is the total flux emerging from the photosphere, integrated over all wavelengths, and σ is Stefan's constant. The effective temperature thus represents the temperature of the star if it emitted as a perfect black-body. Ideally, the temperature determines the ionisation states and level populations of the atoms and molecules in the atmosphere through the Boltzmann and Saha equations (4.1) and (4.2), and thus the frequencies at which absorption and emission will primarily occur. The effective temperatures of M dwarfs are $\sim 2,300 < T_{\text{eff}} < 3,800$ K, while the lower mass L dwarfs reach temperatures below 1,500 K.

The surface gravity is given by:

$$g = GM/R^2 \quad (4.4)$$

where M and R are the mass and radius of the star, respectively. The gravity primarily determines the pressure and density of the atmosphere. A higher-density atmosphere has a larger number of particles that can interact with the radiation field. Surface gravity is reported as $\log g$; the surface gravity of the Sun is $\log g \sim 4.44$, and for low-mass dwarfs it ranges from $\log g \sim 4.6\text{--}5.3$.

The mean metallicity is defined as:

$$Z = M_{\text{metals}}/M \quad (4.5)$$

where M_{metals} is the total mass contained in all elements except for hydrogen and helium, and M is again the mass of the star. Another common expression used to characterise the metallicity is the metal abundance of the star:

$$[\text{m}/\text{H}] = \log 10(n_m/n_H) - \log 10(n_m/n_H)_\odot. \quad (4.6)$$

Thus, for example, a star with a metal abundance of one tenth the solar value has $Z = 0.1$ and $[\text{m}/\text{H}] = -1$. The use of 'm' to represent any metal in the star presumes that the distribution of metals in the star is the same as for the Sun.¹ As we shall see, the metallicity greatly affects the ability of the gas in the photosphere to absorb and emit radiation, since elements more massive than H and He have many more energy levels available for intercepting and producing photons.

A typical photospheric model is computed by assuming that the gas is arranged in homogeneous, plane-parallel layers of height dz , and is in a steady state. We can then apply the equation of hydrostatic equilibrium

$$dP/dz = -pg \quad (4.7)$$

which says that the pressure gradient is balanced by the force of gravity. An equation

¹ This may not always be the case; there is evidence that metals such as the α elements do not always occur in the solar ratio to the iron peak elements (see Chapter 7).

of state connects the pressure, density and temperature. For example, a perfect gas has the usual equation of state

$$P = \rho k T / \bar{\mu} \quad (4.8)$$

where $\bar{\mu}$ is the mean molecular weight of the gas. The equation of state used for M and L dwarf atmospheres is much more complicated, requiring the calculation of partial pressures for many different atomic and molecular species (and, for the lowest mass objects, dust grains) at each height in the atmosphere. The equation of state changes with height, since molecules will form at different temperatures, and hence contribute varying partial pressures at different heights.

A second equilibrium condition is the requirement of energy balance in each layer, usually expressed as flux conservation,

$$dF/dz = 0 \quad (4.9)$$

which says that all energy absorbed in a photospheric layer is exactly balanced by energy leaving the layer, so that the net flux change is zero. In contrast to earlier-type stars, the energy is carried by both convection and radiation in an M dwarf photosphere; the flux in equation (4.9) has both convective and radiative components. The treatment of convection is therefore an important aspect of the models. Modern models use a mixing length formulation for the convective flux [M3],

$$F_c = \frac{1}{2} \rho C_P \bar{v}_{conv} T \frac{l}{H} (\nabla_{atmos} - \nabla_{blob}) \quad (4.10)$$

with the mixing length parameter (the ratio of the mixing length, l , to the pressure scale height, H) set to values between 1 and 2. C_P is the specific heat at constant pressure, \bar{v}_{conv} is the average velocity of a convective blob, and ∇_{atmos} and ∇_{blob} are the temperature gradients in the atmosphere and in the blob, respectively. (See Section 3.3.4 for a discussion of the adiabatic temperature gradient which determines when convection will occur.) The convective flux depends strongly on C_P , which in turn depends on metallicity. C_P is particularly large in the atmospheres of M dwarfs, where molecules contain considerable amounts of internal energy; hence the importance of treating convection explicitly at all depths in the photosphere.

The radiative flux depends on the local density and metallicity through the optical depth, τ_ν at frequency ν ,

$$\tau_\nu = \int_z \rho \kappa_\nu dz \quad (4.11)$$

and the equation of radiative transfer,

$$\mu dI_\nu / d\tau_\nu = I_\nu - S_\nu. \quad (4.12)$$

We will refer to the local quantity $\rho \kappa_\nu$ as the ‘opacity’ of a layer, and the integrated quantity given in equation (4.11) as the optical depth, τ_ν , although many authors use the terms interchangeably. The quantity I_ν is the intensity along a particular direction θ ($\mu = \cos \theta$, not to be confused with the mean molecular weight, $\bar{\mu}$). The radiative flux is the integral of the intensity over all directions and frequencies. Equation (4.12) says that the change in intensity, dI_ν of a ray travelling through a

gas with optical depth $d\tau_\nu$, is given by the initial intensity I_ν , modified by the source function S_ν , which describes the balance between absorption and emission within the layer, at frequency ν . Complete descriptions of the derivation and usage of equation (4.12) are given in [M3].

The optical depth, τ_ν , accounts for the interaction between the photospheric matter and the radiation field through the important parameter κ_ν , the mass extinction coefficient. κ_ν is also referred to as the absorption coefficient, as in Section 4.2, but it actually contains both an absorption and a scattering component. κ_ν has units of cm^2/gm , and is essentially a cross-section per unit mass, describing how likely it is that a photon will undergo an interaction with a particle in the gas. The distance that a photon travels before it undergoes an interaction is called the photon mean free path. The opacity, $\rho\kappa_\nu$, has the units of inverse length and can be thought of as the inverse of the photon mean free path at a particular frequency. The optical depth – the integral of the opacity over a distance z – then represents the number of mean free paths a photon must travel to escape from a depth z in the atmosphere. A general rule of thumb is that radiation emitted at optical depth $\tau_\nu < 1$ (a height less than one photon mean free path) will escape the atmosphere and be observed, while radiation from larger optical depth will be absorbed or scattered before it can escape. Stars with more metals (larger Z) will in general have more absorption and scattering, and thus a larger κ and greater optical depth over smaller distances.

Observations of the photospheric spectrum at different frequencies provide a map of the atmosphere, reflecting the opacity (and hence temperature, density and metallicity) of different layers, depending on the optical depths at the frequencies that are sampled. For example, the optical depth at the line centre frequency ν_o of a strong absorption line will be very large even at a small physical distance z into the photosphere, since κ_{ν_o} is large (the atom or molecule has a high cross-section for photon interaction). Radiation from $\tau_{\nu_o} = 1$ will come from a very shallow distance in the atmosphere. Conversely, the continuum radiation emanating at a frequency ν between strong lines will experience less opacity and will reach $\tau_\nu = 1$ much deeper into the atmosphere, reflecting the conditions at those layers.

From the above equations it is clear that specifying the effective temperature, surface gravity, and metallicity of a star provides boundary conditions on the flux, pressure and density, and constraints on the interaction between radiation and matter, which lead to a self-consistent determination of the optical depth at each frequency, and finally a prediction of the radiation emitted at each frequency. (This is from the point of view of the modeller. The observer, in contrast, has data showing the radiation emitted at each frequency, and would like to use the observations to infer the temperature, gravity and metallicity of the star.)

In practice, the procedure is to postulate an initial temperature and density structure, solve the above equations in each layer, perform numerical integrations over the equilibrium equations, and determine the corrections to the initial structure that must be applied to satisfy the equilibrium and boundary conditions. The solution is then iterated until convergence, and the quantity of interest for comparison with observations (the emergent spectrum) is computed. Although this qualitative description appears straightforward, in practice it is often difficult or

impossible to carry out. Many workers have spent countless hours devising clever and innovative ways to solve the model atmospheres problem [C1].

In general, determining the interaction between the radiation field and the matter, which is hinted at by equations (4.11) and (4.12), is the most complicated and difficult part of the procedure. In particular, specifying the exact form of κ_ν at all frequencies is often an intractable problem, considering the vast number of transitions available for excitation in a typical metal atom or simple molecule.

A common assumption which greatly simplifies the problem is to assume that the atmosphere is in local thermodynamic equilibrium (LTE) – the condition that the radiation is locally in equilibrium with the matter, and that both may be characterised by a single temperature. In this case, the velocity distribution of the particles in the gas is described by a Maxwell–Boltzmann distribution of a given temperature, and the radiation field is described by a Planck function of the same temperature. The level populations of all states of the matter are then completely determined by the temperature (via the Boltzmann and Saha equations) allowing immediate calculation of κ_ν for all transitions being considered. The remaining problem, even for LTE models, is to specify sufficient transitions in the dominant atoms and molecules so that the computed opacity encompasses all of the important radiative processes.

In M and L dwarf photospheres, the effective temperature is low while the surface gravity is large, leading to high photospheric densities. Both low temperature and high density lead to increased opacity, since 1) essentially all atoms will be neutral with many transitions having low excitation energy; 2) many molecular species can form, contributing even more low-energy transitions; and 3) the number density of atoms and molecules will be large, increasing the probability of interaction with the radiation. These factors tend to move the atmosphere toward LTE, and all current photospheric models for cool dwarfs do assume LTE in the computations.

4.4 TREATMENT OF OPACITY

The description of the opacity is the single most important factor in producing an accurate photospheric model for a low-mass star. The reader may have encountered references to ‘grey’ atmospheres (computed assuming the opacity is frequency-independent), or atmospheres calculated using the Rosseland mean opacity (also frequency-independent, but weighted at each depth such that the integrated flux is correct at the surface; see [M3]). While adequate for some earlier-type stars, and particularly for calculations at deep layers in the stellar interior, these approximations are not valid for M and L dwarf photospheres. To illustrate the complexity of their spectra, Figure 4.5 shows a spectrum of the famous M8 dwarf VB 10, together with a black-body of the same effective temperature, the continuum expected from H⁻ and a model fit. The important molecular bands of TiO, VO, FeH, H₂O, and CO are indicated. It is clear that the simple black-body and H⁻ approximations to the continuum flux distribution have been greatly distorted by the (mostly molecular and

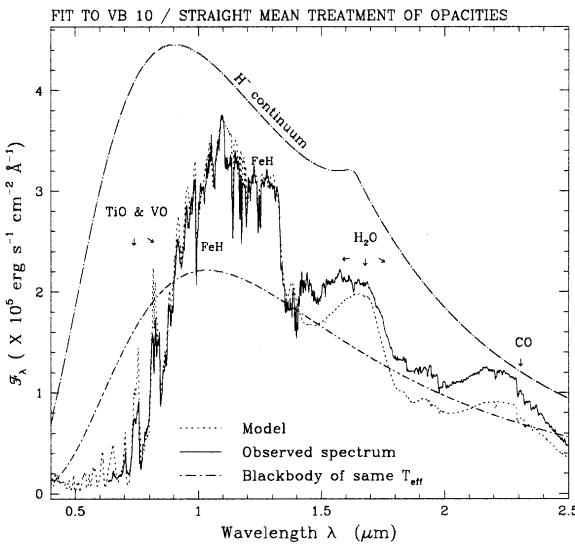


Figure 4.5. The observed spectrum of the M8 dwarf VB 10 is shown together with a model fit. The idealised black-body and H^- spectra are clearly inadequate to describe the complicated structure stemming from the significant molecular line opacity. (Courtesy of F. Allard and P. Hauschildt, and the *Astrophysical Journal* [A3].)

strongly frequency dependent) opacity, and detailed models are required to understand the spectra of these cool dwarfs.

The important opacity sources in a stellar atmosphere involve bound–bound (line), bound–free (continuum), and free–free (continuum) processes due to both absorption and scattering. In the bound–bound case, absorption occurs when a photon from the radiation field is absorbed by an atom (or molecule) in the gas and the excited state is subsequently de-excited by a collision between the atom and another particle (typically an electron, or another atom). The photon is thus destroyed, and its energy returned to the gas. Absorption therefore couples the radiation field to the matter through collisions. The collisions occur with a rate that depends on the velocity distribution of the particles, and thus reflects the local temperature of the gas. If absorption dominates the opacity, the condition of LTE will be approached, and the assumption of LTE is acceptable.

Scattering, in contrast, occurs when an excitation of an atom (or molecule) by a photon is followed by immediate radiative de-excitation (which emits a photon). No collisional processes are involved, so scattering is only weakly dependent on the local matter temperature. An important attribute of bound–bound scattering is that transitions have a finite energy width, such that a slight change in frequency may occur between the absorbed and re-emitted photons. This allows, for example, photons which are absorbed at the line centre frequency with high probability to be re-emitted at a frequency in the line wing. The probability of absorption is much smaller for a line wing photon (see Figure 4.2), so the emitted photon may travel a large distance through the atmosphere before being re-absorbed. In this way the

radiation field can be quite decoupled from the local matter temperature. When scattering dominates the opacity, the assumption of LTE is no longer valid.

The bound–bound (line) opacities are used to describe transitions between bound levels, including the extensive, closely-spaced, rotational and vibrational bands produced in molecules. In practice, for low-mass stars, TiO, CaH and other oxides and hydrides in the visible spectrum, and H₂O and CO in the infrared spectrum, dominate the line opacity under most conditions. Current references for line lists for these molecules may be found in [A2], [A4], [H1]. The huge number of transitions (literally hundreds of millions of lines for H₂O alone) makes the detailed calculation of κ_ν at individual frequencies a daunting task. As an alternative, various approximations to the opacity have been used (Carbon [C1] contains a description of many of these methods). Examples include the straight mean (SM), harmonic mean (HM), just overlapping line approximation (JOLA), and opacity distribution function (ODF) methods which approximate the opacity in a given frequency interval by some average value, smoothly varying function, or histogram of values based on the frequency of occurrence of high and low opacities in the interval. A more rigorous but computationally expensive method is the opacity sampling (OS) procedure where the true opacity is calculated on a grid of pre-specified frequency points, including all transitions that contribute opacity above a given threshold at that frequency. The OS method represents the current state-of-the-art, but the models still fail to adequately represent the observations in some important frequency intervals. There is much work yet to be done both in the compilation of better transition data for important molecules, and in the treatment of the line opacity in the models.

Bound–free and free–free continuum processes, in contrast, are straightforward to compute analytically for atomic and molecular ionisation (dissociation) and recombination (formation), and for Thomson and Rayleigh scattering. Unfortunately none of these processes is very important in the cool, high-pressure environment of the M dwarf photosphere. Figure 4.6 shows the absorption coefficients for the most important contributors to the opacity in solar metallicity and low ($[m/H] = -2.5$) metallicity models. The solar metallicity model is dominated by molecular line absorption due to H₂O in the infrared and TiO in the optical, while the low metallicity model has far less TiO opacity in the optical and is controlled by collision-induced H₂ absorption in the infrared.

Allowed transitions in the H₂ molecule occur only at electric quadrupole and higher-order moments. However, H₂–H₂ opacity may be produced as a result of photon interactions with the temporary electric dipole induced during a collision between two H₂ molecules. This process is also known as pressure-induced absorption, since the number of H₂ collisions depends strongly on the pressure. In low metallicity atmospheres, H₂ is by far the most abundant molecule and the H₂–H₂ opacity dominates at wavelengths longer than about 1 μm. Linsky [L5] carried out an initial semi-empirical investigation of the H₂–H₂ opacity, but detailed quantum mechanical calculations are required to properly model this and other collision-induced absorption (CIA) opacities; for example, those induced during H₂–H and H₂–He interactions [L4]. Borysow [B5] provides an excellent review of current modelling efforts of the CIA opacities in cool stars.

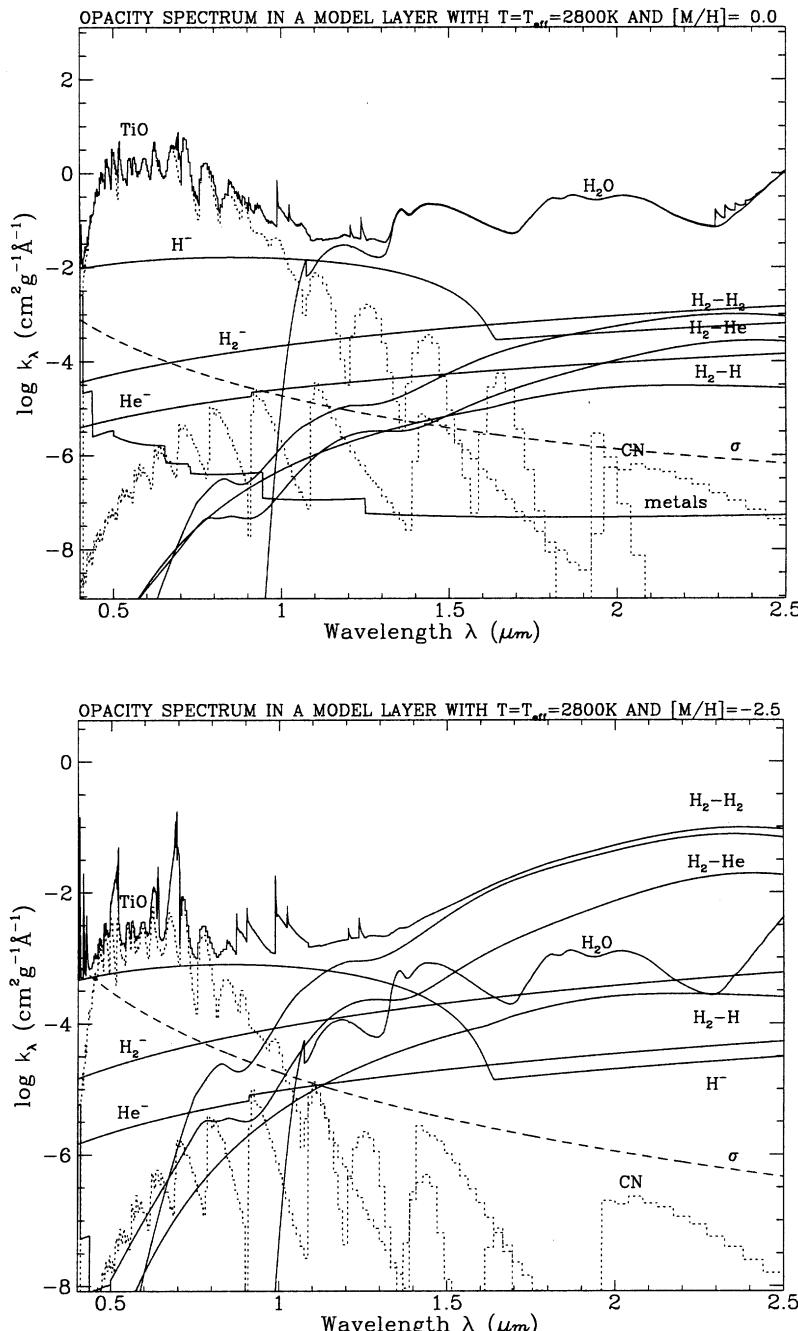


Figure 4.6. Wavelength dependence of opacity sources for solar and $[\text{M}/\text{H}] = -2.5$ models with $T = 2,800\text{ K}$ and $\log g = 5.0$. (Courtesy of F. Allard and P. Hauschildt, and the *Astrophysical Journal* [A3].)

Another important effect in the high-pressure M and L dwarf atmospheres is the presence of significant line-broadening due to van der Waals forces acting on the energy levels of the atomic and molecular species. This effect is known as ‘pressure’ or ‘collisional’ broadening, and it contributes to the Lorentz damping profile of the line with a (classical) damping constant [A4],

$$\gamma_{vdW} = 17C_6^{2/5}v^{3/5}N_p \quad (4.13)$$

where v is the relative velocity between the interacting particles, N_p is the number density of the perturbing particles, and C_6 is an interaction constant. C_6 can be determined exactly for perturbations by atomic hydrogen, and has been given various approximate forms for molecular hydrogen interactions, which are more common in M and L dwarf atmospheres [S1]. Typical widths for strong absorption lines in M dwarfs are a few Å. As discussed above, increasing the energy width of the bound–bound transition increases the importance of scattering (as opposed to true absorption) in the calculation of the opacity. This tends to move the atmosphere away from LTE.

There are several other broadening mechanisms that have a smaller effect on the line profile. The microturbulent velocity, which has a Gaussian profile, is a statistical necessity in higher-mass stars, but is generally less important for low-mass dwarfs, amounting to only 1–2 km s⁻¹. This is much less than the pressure-broadening or the width of the Doppler core due to random thermal motions. Stark broadening, which has a Lorentz profile, results from energy level perturbations due to an ambient electric field. It is generally not important in these mostly neutral photospheres, although it can play an important role in shaping the line profiles in the chromosphere. Rotation, and the effects of magnetic fields, can make important contributions to the line profiles for some stars (see Chapter 5).

4.5 PHOTOSPHERIC MODELS

There are several steps that must be undertaken in producing (and understanding) photospheric models. First, the underlying assumptions in the computation of the model must be known (for example: which atoms and molecules are included in the opacity? – what types of opacity approximations are made? – is the atmosphere in LTE? – is convection included as an energy transport mechanism?). Next, the results of the model computation must be examined, particularly noting how the structure of the model atmosphere varies with different assumptions. The latter provides insight into the sensitivity of the model calculations to the input physics, and will be discussed in this section. Finally, the atmosphere models are used with a spectrum synthesis program to produce an emergent spectrum that can be compared to observations, as described primarily in Section 4.6, although comparison with observations for some early models is briefly mentioned here. (For a comprehensive discussion of M dwarf models through 1996, see the excellent review in Allard *et al.* [A4]; a summary of newer models may be found in [A2].)

The models described in this section were computed under the assumptions of plane-parallel geometry, homogeneous layers, LTE, microturbulent velocities between $1\text{--}2\,\text{km}\,\text{s}^{-1}$, and with convection treated using a mixing length theory with mixing length parameter between 1 and 2, unless otherwise noted. Model results are typically shown as the run of temperature and density with optical depth, height, column density (mass), or gas pressure. The optical depth is reported at a standard continuum wavelength, τ_{std} ($\lambda_{\text{std}} = 1.2\,\mu\text{m}$ in the NextGen models described below), or the Rosseland mean optical depth, τ_R , is used. The height, z , is measured outward from the photosphere boundary, taken to be the location where $\tau_{\text{std}} = 1$. The column density, N , and column mass, \bar{m} , are the integrals through the atmosphere of the number and mass densities, respectively: $N = \int_r n\,dr$, and $\bar{m} = \int_r \rho\,dr$. By convention, the outermost layer considered in the model has $r = 0$, and r increases inward, thus the column density and column mass also increase inward. Note that r and z are both height variables, but are defined in the opposite sense, and have different zero-points. The gas pressure, which we use here, is related to the column mass by the hydrostatic equilibrium requirement (equation (4.7)). At a given depth, $P_g = \bar{m}g$, so if the gravity is constant (as is generally assumed), the gas pressure is proportional to the column mass.

The first attempts at modelling main sequence dwarfs with low effective temperatures were made in the mid-1960s by Gingerich and Kumar [G1], who used only continuum opacities and assumed radiative equilibrium (convection was neglected); Tsuji [T2], who included a few molecules (CaH, H₂O and CO) and attempted some models with convection; and Auman [A6], who included convection and a more detailed description of the water opacity employing an HM opacity treatment. Auman noted several results which remain true today: that convection must be included, as it is important to very small optical depth ($\tau < 0.01$) in the photosphere; that some broadband colours are not monotonic functions of T_{eff} and hence care must be taken in defining colour-temperature relationships that connect the models to observations; and that the strengths of the water bands depend critically on the detailed temperature structure and the treatment of convection. He mentioned the inadequacy of the water-band line list, but the lack of near-infrared spectra at the time did not allow a comparison with observations.

The next improvements in the models came with Mould [M4], [M5], [M6], who used a version of the Kurucz ATLAS code [K3], including convection and treating atomic and molecular line opacities with an ODF description. The line lists were much more extensive than those of the previous decade, and included the important TiO molecule which dominates the optical opacity in the M dwarfs, along with H₂O and CaH, but not CO. He found that TiO is a good temperature indicator in most models (bands become stronger with decreasing temperature), but the TiO band-strength weakens at the lowest temperatures. Also, CaH is particularly sensitive to gravity, especially in the cooler models. Comparison of the model results with the few (quite uncertain) infrared observations in the water bands that were available at that time provided the first indication that the water opacities were being overestimated in the models – a condition that still exists today.

The Mould models remained in common use for almost 20 years, as the

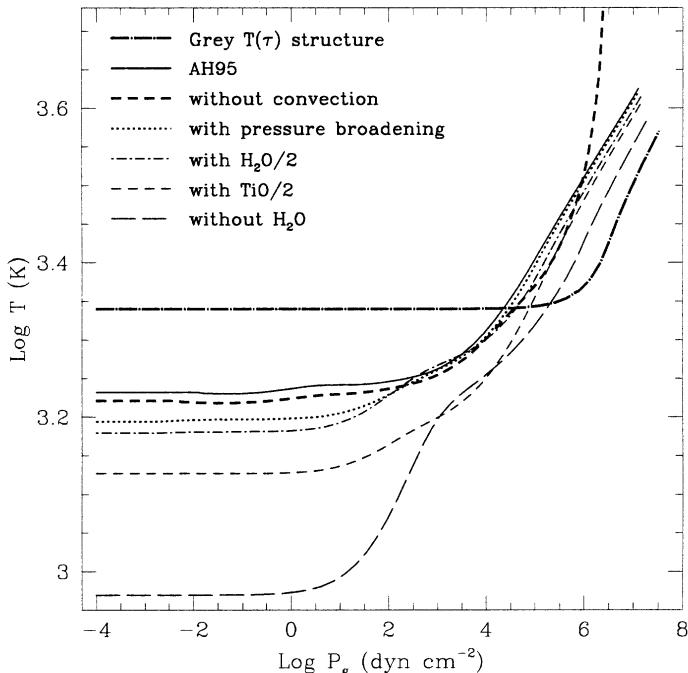


Figure 4.7. The effect of changes in model computations. The model has $T_{\text{eff}} = 2,800 \text{ K}$, $\log g = 5.0$ and solar metallicity. (Figure courtesy of F. Allard and P. Hauschildt, and the *Annual Review of Astronomy and Astrophysics* [A4].)

measurement and compilation of atomic, and particularly molecular, line data slowly progressed. The next important sets of M dwarf models were produced in the early 1990s, in the PhD theses of Allard [A1] and Ruan [R2]. In 1995, Allard and Hauschildt published improvements on the [A1] models using the PHOENIX code [A3], and Brett extended the Ruan work using the MARCS code [G3], [B1], [B7]. Both groups obtained reasonable fits to the optical spectrum, and found that the water bands from the computed models were much too strong compared to the (now much improved) infrared observations [A3], [B6].

Both Allard and Brett employed extensive water and TiO line lists: Allard used SM and JOLA opacities for these molecules and an OS method for the atomic and molecular lines, while Brett used an OS method for all the molecular opacities.² Figure 4.7 shows the effects of varying the TiO and water opacities on one of the [A3] models. The exact values of the opacities are more important for TiO than for water, although ignoring water altogether results in a dramatic cooling of the outer atmosphere by CO.

² Brett also showed that current versions of the Kurucz ATLAS models – previously used by Mould and widely used for earlier type stars – are inadequate for M dwarfs, due to the lack of many molecular (especially water) opacities [B6].

The mid–late 1990s saw the computation of several grids of models (using the PHOENIX code) by Allard, Hauschildt and co-workers. The models published in 1995 are known as a ‘base’ grid [A3]. Subsequent grids (an upgraded version of the ‘base’ grid, and the ‘extended’ grid) were also made available to the community but were not formally published. Some results are discussed in [A4]. One must beware (as we shall see in the next section) that the rapidly changing nature of the models during this time rendered the comparison with the observations quite fluid. The exact generation of models used could produce different results, for example, in the temperature scale (Figure 4.13). The situation improved in 1999 with the publication of the NextGen models, incorporating more extensive line lists and an opacity sampling treatment for all molecules, for atmospheres with effective temperatures greater than 3,000 K [H1]. These will most likely remain the standard for some years to come, and we shall concentrate on these models in our discussion of the spectra.³

Figure 4.8 is a comparison of the models produced during the years 1969–1999. There is reasonable agreement, although the Auman model is cooler, and the early Allard models (‘base’ and ‘extended’) are hotter than the others; the NextGen models are in closer agreement with the Brett and Mould models at large depth than the previous generations of Allard models. Though the models appear to be nearly the same, the spectra that are produced can be quite different in detail. This serves as a warning that the results of the computations are sensitive to the adopted physics, especially in the opacity treatment. The formation depths ($\tau_{line} \sim \text{unity}$) for several strong molecules are indicated on the figure; these are generally at $\tau_{std} \ll 1$ or $\log \tau_{std} \ll 0$. Some models (Auman, Mould, Brett) do not extend far enough into the optically thin regimes of the important molecular species to include all of the molecular opacity.

Figure 4.9 shows the run of temperature (left panels) and standard optical depth (right panels) with gas pressure, for a range of NextGen models. The top two panels indicate the effect of varying the effective temperature while holding the gravity and metallicity fixed. The outer region of the lower-temperature atmosphere has greater optical depth at a given pressure, since lower-temperature conditions favour more molecule formation and hence greater opacity at a given physical height. At larger depth, the situation reverses as the H⁻ continuum opacity becomes dominant in the higher temperature stars. Variations in gravity and metallicity (middle and bottom panels respectively) both indicate that the pressure–opacity relationship controls the temperature structure. Plots of temperature as a function of $\log \tau_{std}$ show essentially no differences between these atmospheres. In the case of varying gravity, the higher-gravity atmosphere reaches a given optical depth at a higher pressure than a lower-gravity atmosphere, due to the higher density and consequent greater abundance of molecules. The effect is even more pronounced in the low-metallicity atmospheres, where the difference in pressure at $\log \tau_{std} = 0$ is nearly two orders of magnitude

³ At temperatures below 3,000 K, many species of dust grains can condense out of the gas phase in the atmosphere, and the effects of dust on the model structures become increasingly important. Models that include dust are discussed later in this chapter.

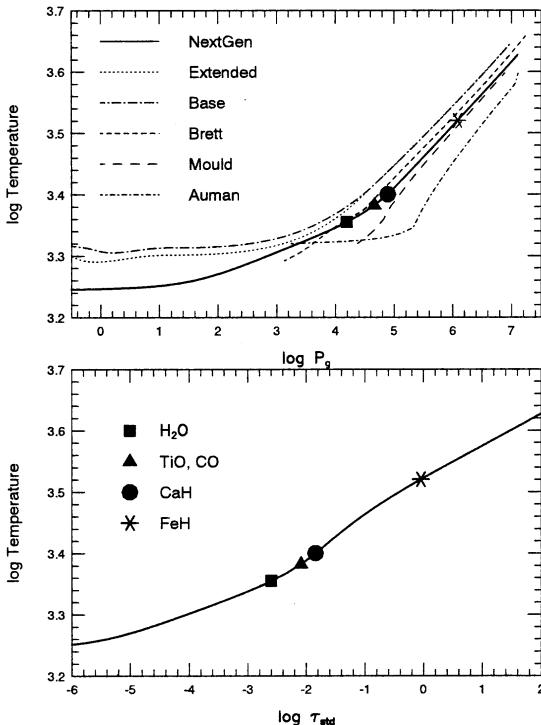


Figure 4.8. Comparison of a standard model computed by several authors, as indicated. The model has $T_{eff} = 3,000$ K, $\log g = 5.0$, and solar metallicity, except for the Mould model which has $\log g = 4.75$. The formation depths of the molecular species H_2O , TiO , CO , CaH and FeH are given on both panels; the lower panel shows that the molecules generally form at small continuum optical depth.

between the solar metallicity and $[\text{m}/\text{H}] = -2$ atmospheres. In this case, the explanation is the lack of metals and hence lower opacity in the low-metallicity atmospheres, which leads to diminished optical depth at a given pressure. It is therefore possible to look deeper into the atmosphere, to conditions of much higher pressure before attaining significant continuum optical depth. However, greater pressure and density do not always result in a higher molecular abundance for every species. When we examine the emergent spectra from these atmospheres we shall see that the complicated interplay between the metallicity and the pressure can influence molecule formation in a subtle fashion.

4.6 PHOTOSPHERIC DIAGNOSTICS

4.6.1 The comparison with observations

With the models in hand, it is necessary to produce diagnostics that can be compared directly to the observations. In other words, the modeller has an atmosphere giving temperature as a function of optical depth (for example), while the observer has a

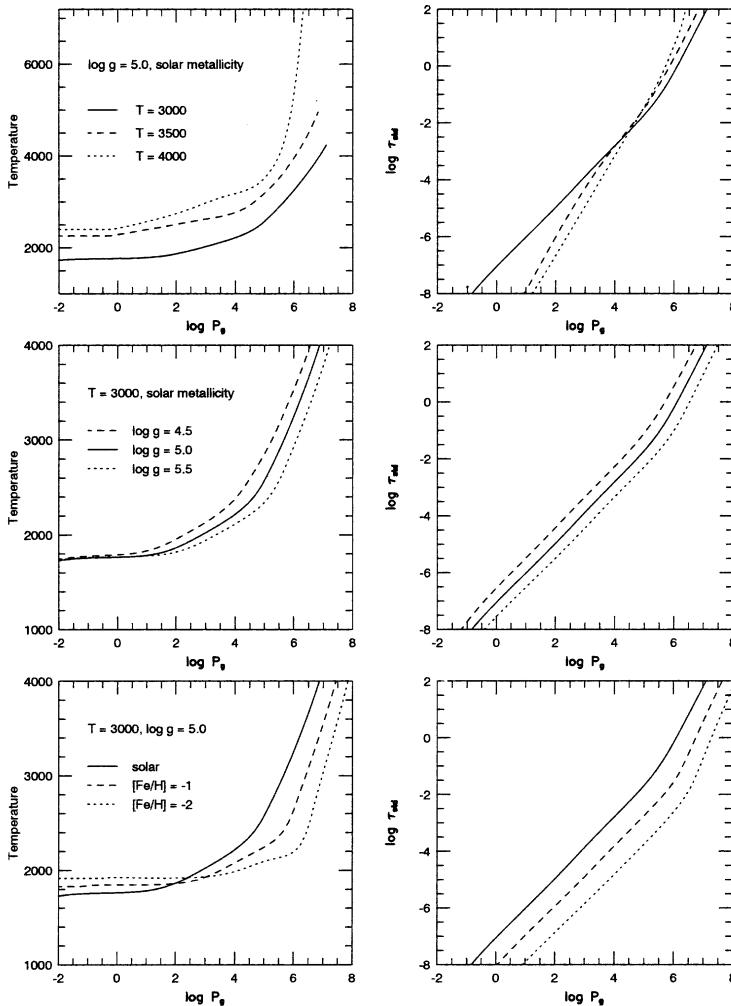


Figure 4.9. The NextGen models [H1], illustrating structural changes with temperature, gravity and metallicity. The left panels show the run of temperature with gas pressure, the right panels the change in continuum optical depth (τ_{std}). The line types apply to both panels. (Models courtesy of P. Hauschildt.)

spectrum showing various absorption lines superposed on a background continuum source. In order to determine whether the model accurately predicts the strengths and profiles of the absorption lines, and hence whether the model is an acceptable representation of the physical conditions on the star, a synthetic spectrum must be produced from the model. Most model atmosphere codes produce a spectrum as part of the model output. However, researchers who have access to the published models but not the atmosphere code, may just as easily use their own spectrum synthesis programs to produce an emergent spectrum (although care must, of course, be taken

that the assumptions used in the code to produce the models are compatible with those in the spectrum synthesis code). We shall see in Chapter 5 that modellers of the outer atmosphere adopt these photospheric models as boundary conditions, and use other spectrum synthesis codes to investigate the consequences of adding a chromosphere or flare energy [H2], or to search for evidence of photospheric magnetic fields [J1].

Note that the opacities used in the spectrum synthesis code may be much more detailed, and computed on a much finer wavelength grid, since the underlying atmosphere is already given and is assumed not to change. Variables such as temperature and density are therefore known at every atmospheric level, and, although the opacity computation is still time consuming, it does not feed back into the atmospheric structure as it does in the model computation. Thus, for example, line profiles may be produced by using the full absorption profile (Figure 4.2) in the computation, rather than approximating the opacity in the line as a single value at the central frequency, which is often the expedient method when producing the model.

Previously, we described some of the classical methods of stellar atmospheres theory and the comparison with observations, including the concept of the line equivalent width and the curve of growth. These methods are still applicable, and indeed useful, in obtaining a qualitative feel for the model output and the match to the observations. With the current models, it is now also possible to obtain a quantitative fit to the spectral data. The output of the models is a synthetic spectrum, with line profiles computed on scales of an Ångstrom or less. The synthetic spectrum is convolved with filter profiles to produce photometric colours (see Figure 2.15), or rebinned to match the instrumental resolution and compared directly to an observed spectrum. Narrow-band indices centred on important molecular features, such as the optical bands of CaH and TiO, also provide a useful measure by which to compare models and observations.

Figure 4.10 shows the emergent spectra over the optical and near-infrared wavelength régime corresponding to the models in Figure 4.9. These spectra have been convolved with a Gaussian profile ($\sigma = 5 \text{ \AA}$) to smooth out most of the spurious effects introduced by the opacity sampling method, while retaining the overall shapes of the important molecular features. Clearly the temperature has the strongest effect on the output spectra, while increasing the gravity produces a barely noticeable effect in the optical region (a slight strengthening of the CaH bands, as first noted by Mould). Reducing the metallicity is seen both in the strengthening of the hydride bands (CaH, FeH) and weakening of the oxide bands (H_2O , TiO). These effects are described in more detail below.

4.6.2 Temperature determinations

Most M dwarfs have gravity $\log g \sim 5$, and near-solar metallicity, making temperature the most important unknown parameter. Finding the temperature from a comparison of models with observations results in a spectral type–temperature calibration (using spectral data), or a colour–temperature calibration (using photometric data). Both calibrations are in common use, and provide

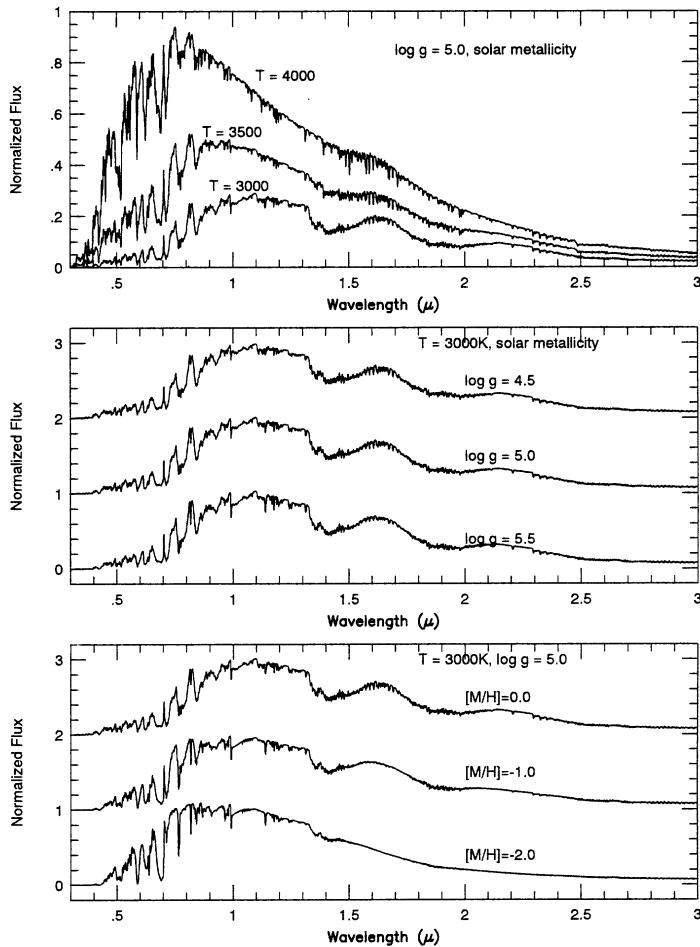


Figure 4.10. Spectra from the NextGen models shown in Figure 4.9, illustrating the changes that occur when the temperature, gravity and metallicity are varied. The temperature causes a marked change in the spectrum, shifting the wavelength peak to the blue and reducing the strength of the molecular bands in the hotter models. The changes with gravity are quite subtle, while the metallicity changes are most noticeable in their effect on the infrared water bands, which nearly disappear at low metallicity. (Model spectra courtesy of P. Hauschildt.)

methods of obtaining temperatures for large samples based on the results of a few well-studied stars.

The KHM spectral types (see Chapter 2) are usually adopted as the standard for M dwarfs. Temperature scales have been tied to these spectral types using optical and infrared spectra together with several versions of the Allard models (base grid [K1], upgraded version of base grid [L2], and extended grid [G2]), and with the NextGen grid [L6].

Photometric colour–temperature relationships work fairly well in the infrared as shown by Berriman [B1], [B2], who determined bolometric luminosities and ‘black-body equivalent temperatures’ (found by anchoring the black-body at a continuum point in the K -band) using a range of colours from U through M . Tinney [T1] used a similar procedure with black-bodies anchored in the L -band and obtained consistent results, while Leggett [L3] used infrared photometry to infer temperature, metallicity and mass for a number of late M and L dwarfs; her temperatures are given in 500 K bins, and agree within those boundaries with the other colour–temperature estimates.

The spectroscopic and photometric temperature estimates are combined in Table 4.1, to define a spectral type–($V-I$) colour–temperature relationship for solar-metallicity M dwarfs. Typical uncertainties in the temperature are ± 100 K; the spectral types are known to within 0.5 of a spectral class and the colours are measured to ~ 0.05 magnitudes. Figure 4.11 shows optical spectra for M0–M6 dwarfs, together with NextGen model spectra. The general features of the observed spectra are reproduced by the models, but there are clearly differences in fine detail. For example, the TiO bands between 6,600 and 6,800 Å, and between 7,050 and 7,250 Å, are generally too strong in the models compared to the observations, as is the CaH band between 6,950–7,050 Å. The M2 and M4 spectra are better fit than the hotter or cooler models.

Table 4.1. Fundamental properties of M dwarfs.

Spectral type ¹	$V-I$ ²	Tempera-ture ³ Kelvin	Radius ⁴ (R/R_\odot)	Mass ⁵ (M/M_\odot)	Lumi-nosity ⁶ ($10^{-2} L/L_\odot$)	Log gravity ⁷ $\text{g cm}^{-2} \text{s}^{-1}$	Prototype
M0	1.92	3,800	0.62	0.60	7.2	4.65	Gl 278C ⁸
M1	2.01	3,600	0.49	0.49	3.5	4.75	Gl 229A
M2	2.15	3,400	0.44	0.44	2.3	4.8	Gl 411
M3	2.46	3,250	0.39	0.36	1.5	4.8	Gl 725A
M4	2.78	3,100	0.36	0.20	0.55	4.9	Gl 699
M5	3.70	2,800	0.20	0.14	0.22	5.0	Gl 866AB
M6	4.06	2,600	0.15	0.10	0.09	5.1	Gl 406
M7	4.56	2,500	0.12	~ 0.09	0.05	5.2	Gl 644C (VB 8)
M8	4.66	2,400	0.11	~ 0.08	0.03	5.2	Gl 752B (VB 10)
M9	4.37	2,300	0.08	~ 0.075	0.015	5.4	LHS 2924 ⁹

¹ Spectral types on the KHM scale.

² $V-I$ colours on the Cousins system, from the Appendix, except for YY Gem [M1] and LHS 2924 [L1].

³ Temperatures compiled from several photometric and spectroscopic calibrations (see text for references).

⁴ Radii calculated from $L = 4\pi R^2 \sigma T^4$.

⁵ Mass estimates from 8-parsec sample for each spectral type, except for M0 and M9 (see notes).

⁶ L calculated from M_{bol} using $M_{bol}(\text{Sun}) = 4.62$, $L_\odot = 3.9 \times 10^{33} \text{ ergs s}^{-1}$.

⁷ $\log g$ calculated from $g = GM/R^2$.

⁸ Gl 278C (YY Gem) is a well-observed eclipsing binary consisting of two dM0e stars; we use it for the prototype, although it is not in the 8-parsec sample. The values given are the average of the two stars. The ($V-I$) colour comes from [M1].

⁹ LHS 2924 is the best studied M9 dwarf, at a distance of 10.5 parsecs.

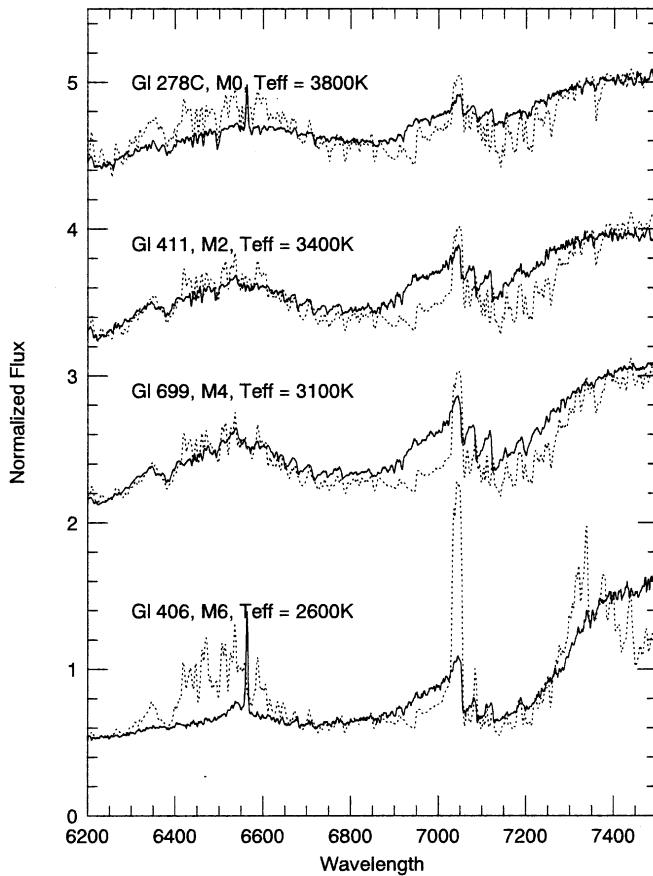


Figure 4.11. Optical spectra for the prototype stars with spectral types M0, M2, M4 and M6 in Table 4.1, together with NextGen model spectra. Models have temperatures from Table 4.1, $\log g = 5.0$, and solar metallicity, and have been convolved with a Gaussian of $\sigma = 5 \text{ \AA}$. The data and models are normalised at $\lambda = 7,400 \text{ \AA}$. (Model spectra courtesy of P. Hauschildt.)

Figure 4.12 shows the infrared spectra of the M2 and M6 dwarfs, highlighting the persistent problem with the water opacities first discussed by Auman. As explained by Allard [A4], the infrared water opacities are systematically overestimated by opacity sampling techniques due to the large overlap of the H_2O lines forming the principal bands. In addition, many of the hot (steam) bands are not included in current lists. These deficiencies tend to offset one another. Fortunately, the use of the SM opacity approximation for water in the Allard base models [A3], which underestimates the opacity, actually gave better fits to the observations than the more rigorous OS method of Brett [B6]. It appears now that the inclusion of dust opacity in the cooler M dwarf, and L dwarf, atmospheres is probably vital to the models (as discussed in Section 4.7). Preliminary models including dust [A2] provide a better fit to the spectral data as shown in the figure.

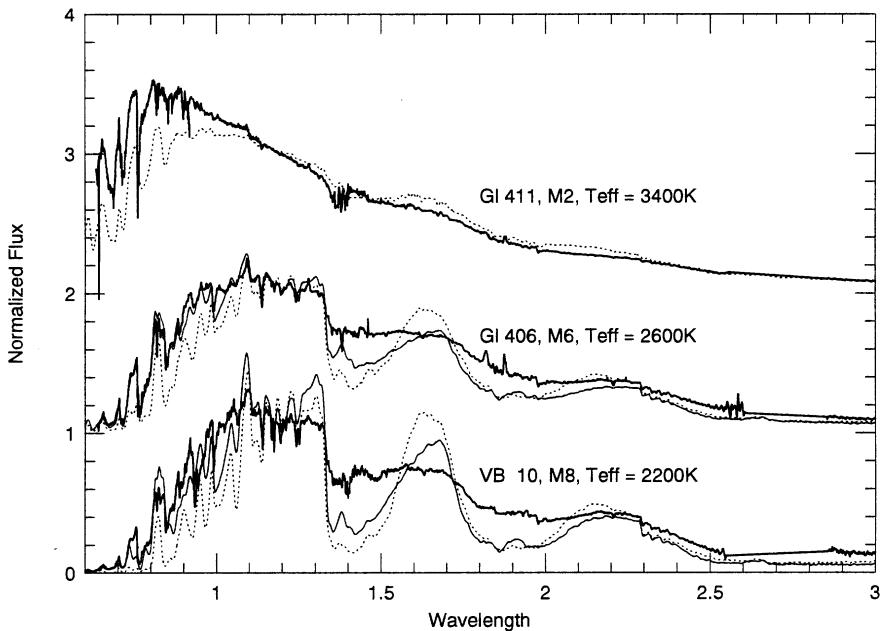


Figure 4.12. Gliese 411 (M2), Gliese 406 (M6) and VB 10 (M8) spectra from 0.6–3.0 μm (dark solid lines) are shown together with NextGen model spectra of 3,400 K, 2,600 K and 2,200 K respectively (dotted lines). The spectra are normalised at the standard model wavelength of 1.2 μm . Preliminary models including dust are shown as the thin solid lines for the 2,600 K and 2,200 K models. Table 4.1 shows a better temperature estimate for VB10 (M8) is 2,400 K. The preliminary dust models provide a much better fit in the optical, and show some improvement in the infrared water bands. (Model spectra courtesy of F. Allard and P. Hauschildt.)

The $(V-I)$ –temperature relationship from Table 4.1 is shown in Figure 4.13 (solid points), and agrees well with the one given in [A2] (solid line). It is somewhat cooler than earlier relations obtained from the Allard base and extended model grids (dotted line). Note that the $V-I$ colour saturates at spectral type M7, and becomes bluer at later types (an effect first noted by Auman; see above). This effect is also seen as the ‘hook’ in the colour–magnitude diagrams of Chapter 3; for example, the bottom panel of Figure 3.14.

Although the temperature estimates from spectra and photometry using the latest models now agree fairly well, they are in disagreement with the structural model calculations. As described in Chapter 3, the structure models tend to produce higher temperatures at a specified mass or luminosity. In other words, an observed M dwarf spectrum, when compared to an atmospheric model grid, matches a model with temperature T_o . However, the inferred mass and luminosity of that M dwarf, when compared to a structural model grid, would suggest that a different model, with $T \sim T_o + 200$ K should have provided the best fit. It is not clear whether the problem lies in the structural models – producing too high a temperature for a given mass star – or in the atmospheric models and spectral synthesis,

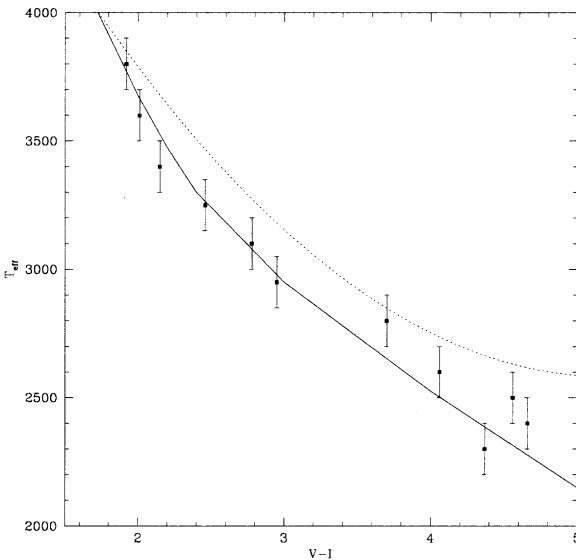


Figure 4.13. The $(V-I)$ colour–temperature relationship from Table 4.1 (points and error-bars). The solid line is the relationship for the NextGen models from [A2]; the dotted line is the relationship from [C2] which they derived from the data and model fits (to Allard upgraded base models) given in [L2]. The difference in temperature amounts to nearly 500 K for the lowest-temperature M dwarfs. The lower-temperature scale is currently favoured.

producing atomic line and molecular band strengths and continuum radiation that suggest too low a temperature for the same star. As pointed out in Chapter 3, the effective temperature of the star depends critically on the atmospheric boundary conditions that are adopted in the structural calculation, so the problem is clearly coupled. It is ultimately a result of inadequate opacity data, and the approximate treatment of the existing opacity data, and hopefully will be rectified in the near future.

4.6.3 Gravity and radius determinations

Figure 4.14 shows higher-resolution (2 \AA) model spectra in the optical wavelength region, illustrating in finer detail the spectral changes due to varying gravity and metallicity. As described in Section 4.5, the higher-gravity models reach optical-depth unity at higher pressure, resulting in increased molecular abundances and somewhat stronger bands. Figure 4.14 indicates that the CaH band at $6,900 \text{ \AA}$ is deeper in the higher-gravity model, but the TiO band at $7,050 \text{ \AA}$ is barely affected. Mould attributed the TiO behaviour to the competition for oxygen between CO, H_2O , TiO and several other oxides. At higher pressure, H_2O formation is favoured, and the abundance of the TiO molecule is essentially unchanged (at solar metallicity). In general, the dependence of the spectrum on gravity is more subtle

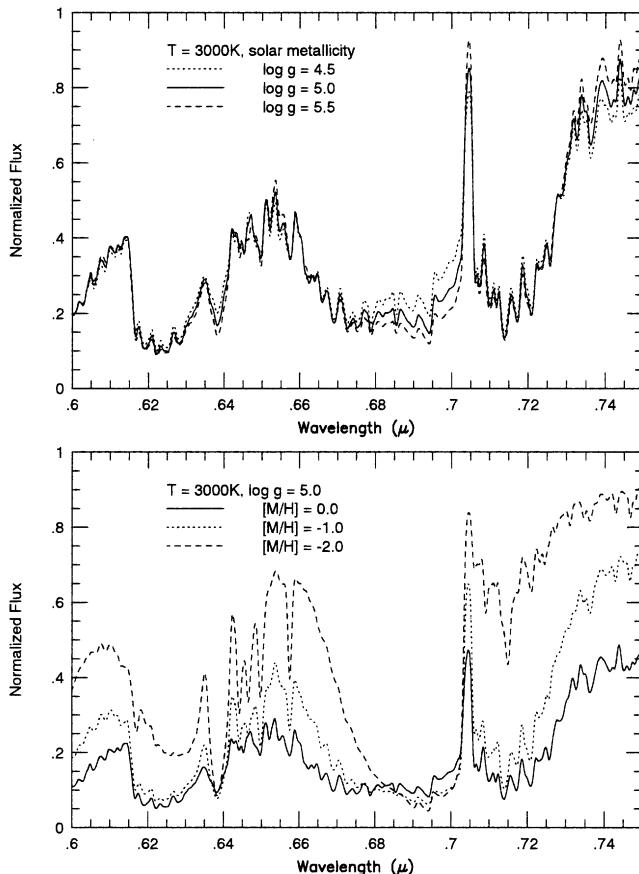


Figure 4.14. Higher resolution spectra from the NextGen models of Figure 4.9, showing the detailed changes in the optical (red) CaH and TiO bands with varying gravity and metallicity. CaH becomes stronger with higher gravity, while TiO remains about the same strength. Alternatively, with lower metallicity CaH appears stronger, while TiO becomes weaker. The continuum is also noticeably bluer in the lower-metallicity models. (Model spectra courtesy of P. Hauschildt.)

than the dependence on metallicity, and requires high-resolution observations to resolve.

The first fine analysis of such high-resolution observations was that of Valenti *et al.* [V1], who obtained $R = 120,000$ spectra of the dM3.5 star Gliese 725B. They used NextGen model atmospheres with their own spectrum synthesis program for TiO, and found $T = 3202 \pm 70$ K, in agreement with the temperature scale in Table 4.1. They also found gravity, $\log g = 4.82 \pm 0.14$, and metallicity, $[m/H] = -0.91 \pm 0.07$, which seems somewhat metal-poor for the temperature they obtained. This may be a result of not using the hydride bands in their analysis (see next section).

Using only photometric data, Leggett *et al.* [L2] employed an alternative approach to finding the gravity. They first used the observed position of a star in the H–R diagram together with an empirical mass–luminosity relationship [H4], to estimate the mass (see Figure 2.23, and discussion in Chapter 9). They compared the masses obtained in this way to theoretical predictions from the structure models of the Lyon group (see Figure 3.16) and found good agreement. The structure models give an estimate of the radius which can be combined with the mass via equation (4.4) to determine the surface gravity.

Alternatively, the luminosity of the star may be determined from the (observed) bolometric magnitude, and together with a temperature estimate the radius can be found from equation (4.3), with $F = L/4\pi R^2$. The mass is found from the same empirical mass–luminosity relationship [H4]; the mass and radius determine the gravity, as before.

The latter method for determining mass and luminosity has been applied to the 8-parsec sample [R1], [C2], and the results are given in the Appendix. Using these data we have computed radii and gravities for each star in the Appendix. Table 4.1 gives the average mass, luminosity, radius and gravity together with a ‘prototype’ star for each spectral type; the prototypes are members of the 8-parsec sample except for the M0 prototype (YY Gem, a well-observed eclipsing binary with independent mass and radius determinations for each component) and the M9 prototype (LHS 2924, the best studied M9 dwarf, located at 10.5 parsecs from the Sun).

4.6.4 Metallicity

Known or suspected low-metallicity stars – which are referred to as subdwarfs due to their position below the solar metallicity main sequence in an H–R diagram (see Chapters 2 and 10) – must have their spectral types, metallicities and temperatures determined self-consistently; simply applying the relationships from Table 4.1 will lead to an underestimate of the temperature by several hundred degrees. In general, metallicity affects the atmosphere by reducing the opacity at a given height, hence the low metallicity atmospheres are expected to reach $\tau = 1$ in the continuum at a greater physical depth in the atmosphere, where the gas pressure and temperature are higher (as illustrated in Figure 4.9). This explains why a low-metallicity star appears hotter, and has a stronger blue continuum, at the same mass and luminosity as a solar-metallicity star. The lack of metals also renders the hydrogen opacity sources more important – particularly the pressure-induced H_2 dipole opacity in the infrared. This leads to a further redistribution of flux from the infrared into the optical in metal-poor stars.

Initial attempts at the identification of metal-poor low-mass stars [M4], [M5] used the change in the infrared colours due to the increased H_2 dipole opacity in the H - and K -bands (hence bluer J – H colour) to separate subdwarfs from disk dwarfs in the J – H versus H – K two-colour diagram (see Figure 2.22). However, the change in the colours is rather small, and is difficult to measure observationally. In addition, models of varying metallicity are not yet able to reproduce the observed colours

in this diagram (see Figure 3 of [L3]), making it difficult to assign quantitative metallicities to the subdwarfs.

Analysis of the spectral changes in metal-poor stars is a more promising avenue for abundance determination. Figure 4.14 shows that changes in metallicity have a larger effect on the spectrum than changes in gravity, but that there is some overlap; for example, the CaH bands become stronger both in higher gravity and in lower-metallicity stars. To disentangle these effects requires observations of several spectral diagnostics that respond differently to changes in these parameters. In particular, the key to metallicity determination in M subdwarfs is to compare the strength of hydride (single-metal) features to species composed of two metals (for example, TiO). The hydride features actually appear to increase in strength with decreasing abundance, while the double-metal species rapidly decrease in strength. The explanation for this apparently paradoxical behaviour lies in the variation of the partial pressures of the species contributing to the line and continuum opacities. The strength of a spectral feature depends on the ratio between these opacities, and hence on the relative number densities of the relevant ions, atoms and molecules. In the optical wavelength regime, the H^- ion is the dominant source of continuum opacity (see Figure 4.6). The number density, $n(H^-)$, decreases slowly with decreasing metal abundance (since the electrons are contributed primarily by metals). The number density of TiO is reduced more rapidly – partly because TiO is a double-metal molecule, and partly because the high pressure in metal-poor atmospheres leads to most of the available oxygen forming H_2O . As a result, $n(TiO)/n(H^-)$ decreases sharply, and TiO absorption weakens at lower metallicity. In contrast, the number density of neutral hydrogen increases as the metallicity decreases, and this compensates partially for the lower absolute abundance of metals in the formation of the single-metal hydride molecules [C3], [B3]. As a result, the ratios $n(\text{hydride})/n(H^-)$ do not change substantially amongst metal-poor stars.

Mould [M6] estimated that the iron hydride (FeH) number densities, and hence bandstrengths, vary with gravity and metal abundance as

$$n(FeH)/n(H^-) \propto g^{0.5} Z^{0.25} \quad (4.14)$$

The gravity dependence in this relationship is another manifestation of the need for a high-pressure atmosphere to form molecules, and demonstrates why the hydride bands are so much weaker in red giant stars. Other hydrides (CaH, NaH, AlH, CrH) show similar behaviour which, combined with the reduced TiO absorption, leads to hydride absorption bands dominating the spectra in M subdwarfs.

Gizis [G2] and Leggett [L2], [L3] pioneered much of the study of metallicity effects in low-mass stars. Gizis empirically defined several narrowband indices centred on optical (red) CaH and TiO bands in the $\lambda\lambda 6,200\text{--}7,500\text{\AA}$ region, in order to take advantage of the differing behaviour of the two molecules with metallicity variation. Since the TiO bands are abundance-sensitive, CaH bandstrength is a better measure of spectral type (temperature) in M subdwarfs. Figure 2.20 illustrates the spectral differences, while Figure 10.4 shows the separation of the disk, subdwarf M (sdM) and extreme subdwarf M (esdM) stars using these indices. Spectral-type sequences for the sdM and esdM stars are shown in Figures 10.5 and 10.6.

To quantify the abundances for each group, representative data can be compared to atmospheric models. Figure 4.15 shows a spectrum of the sdM1.5 dwarf LHS 482 together with model fits covering a range of metallicity, temperature and gravity. The best-fitting model has $T = 3,600\text{ K}$, $\log g = 5.0$ and $[\text{m}/\text{H}] = -1.0$. An extensive set of observations and least squares fits (to the Allard extended models) define the spectral type–temperature calibration for the sdM and esdM stars given in Table 4.2 [G2]. The metallicities are typically $[\text{Fe}/\text{H}] \sim -1$ for the sdM stars, and $[\text{Fe}/\text{H}] \sim -2$ for the esdM stars.

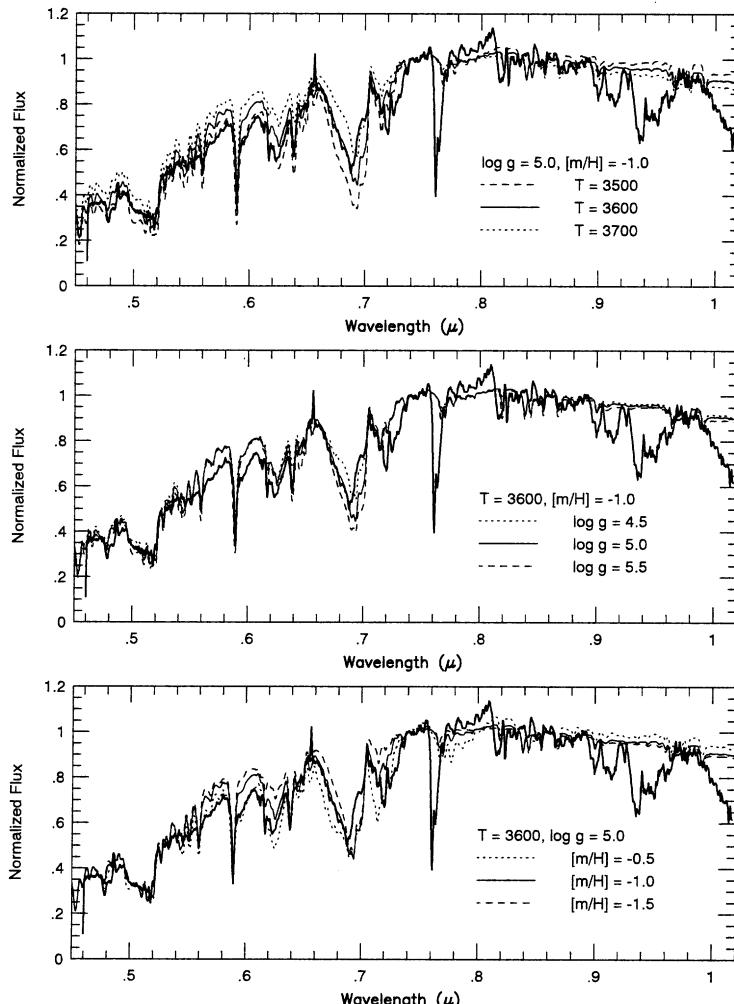


Figure 4.15. The sdM1.5 subdwarf LHS 482 spectrum (dark solid line) compared with model fits with differing temperature (*top*), gravity (*middle*) and metallicity (*bottom*). The model with $T = 3,600\text{ K}$, $\log g = 5.0$ and $[\text{m}/\text{H}] = -1.0$ provides the best fit. (Model spectra courtesy of P. Hauschildt.)

Table 4.2. Temperatures of M subdwarfs and extreme subdwarfs.

Spectral type	T_{eff}	Prototype ¹
sdM0.5	3,700	LHS 307
sdM1.5	3,600	LHS 482
sdM2.5	3,500	LHS 20
sdM3.5	3,500	LHS 2497
sdM5.0	3,400	LHS 407
sdM7.0	3,200	LHS 377
esdM0.5	3,700	LHS 3259
esdM2.0	3,600	LHS 161
esdM3.0	3,500	LHS 1174
esdM4.0	3,400	LHS 375
esdM5.5	3,300	LHS 1742a

¹ Stars are identified by their numbers in the Luyten Half-Second (LHS) catalogue.

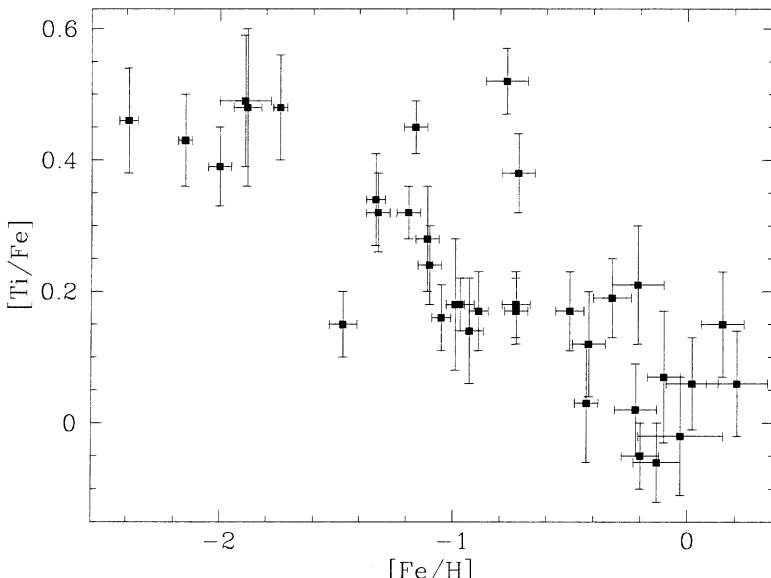


Figure 4.16. The abundance ratio $[Ti/Fe]$ is shown as a function of $[Fe/H]$ for K and M dwarfs and subdwarfs (courtesy of V. Woolf and *MNRAS*).

Recently, Woolf and Wallerstein [W1] have determined metallicities for a sample of 35 K and M dwarfs and subdwarfs. They use improved, though as yet unpublished, NextGen atmospheres to analyse a large number of Fe I and Ti I lines, measured from $R \sim 33,000$ echelle spectra, in regions of minimal molecular

absorption. Their technique provides good metallicity measurements for low-mass subdwarfs that are bright enough to be observed at such high resolution, and those estimates, in turn, can be used to calibrate the lower resolution CaH and TiO indices. Figure 4.16 shows the first determination of the [Ti/Fe] dependence on [Fe/H] for these low-mass stars. As discussed further in Chapter 7, titanium is an α -process element, and is found to be enhanced relative to iron in most low-metallicity stars of earlier spectral type. The [W1] results show that K and M subdwarfs follow a similar trend.

4.7 DUSTY ATMOSPHERES IN THE LOWEST-MASS STARS AND BROWN DWARFS

Tsuji and collaborators [T3] were the first to suggest that dust formation might be important in M dwarfs, although dust had already been identified for many years in M giants. Dust formation depletes many important metals, including (especially) titanium. The most abundant dust grains include corundum (Al_2O_3), perovskite (CaTiO_3), iron (Fe), enstatite (MgSiO_3) and forsterite (Mg_2SiO_4). As in the case of low metallicity, the hydrides increase in strength while the double-metal molecules weaken with the advent of dust formation. Thus the ‘dusty’ atmospheres are dominated by hydride bands, while TiO and VO become weak and, in the lowest-temperature stars, disappear altogether. A second important effect of adding dust to the models is the accompanying greenhouse heating which occurs when the extra dust-grain opacity is included. This enhanced continuum opacity at infrared wavelengths traps outgoing radiation, heating the atmosphere and redistributing the flux into different wavelength regions. The atmospheric heating results in increased dissociation of H_2O , so that spectra from models that include dust have weaker water bands. The dust will eventually settle gravitationally into deeper layers, leaving the atmosphere with depleted metal abundances but without the dust opacity. The time-scale over which this occurs is still open to question, and current modelling efforts include models both with and without dust opacity. Tsuji and collaborators [T3], [T4] have computed models including dust and TiO depletion, while Allard [A2] presented preliminary work on extending the NextGen models to include dust and settling. Approximately 1,000 dust species are included in the computation of the ‘NG-dusty’ models. Figure 4.12 illustrates the improved fit of the NG-dusty models, particularly in the optical spectra of dwarfs cooler than 3,000 K.

Besides the hydride bands, the most prominent features in the optical L dwarf spectra are alkali lines, including the relatively rare metals caesium, rubidium and (if $M < 0.06 M_\odot$) lithium, as shown in Figure 4.16 (from [K2]). The strength of the alkali lines likely stems from the substantially reduced opacity at optical wavelengths in these objects. Not only have TiO and VO virtually disappeared, but dust formation also removes the main electron donors, and the H^- and H_2^- opacities (see Figure 4.6) are correspondingly reduced. As a result, we see deep into the L dwarf atmosphere at optical wavelengths, and the increased column density of neutral alkali atoms leads to very strong lines.

The effect is particularly pronounced for the resonance lines of K I at 7,665 and 7,699 Å, which broaden to equivalent widths of several hundred Å at spectral type L4. The Na I resonance doublet at 5,890/5,896 Å (the D lines) shows a similar behaviour (the 8,192/8,196 Å doublet is not a resonance pair). The fact that these elements have stronger lines than the other alkali metals reflects the relative abundances of the species: on a scale where the number of hydrogen atoms is defined as $[H] = \log_{10} N(H) = 12.0$, $[Na] = 6.31$; $[K] = 5.13$; $[Li] = 3.31$; $[Rb] = 2.41$; and $[Cs] = 1.12$. The transparency of the atmosphere is such that the column densities of sodium and potassium are sufficiently high to place the resonance lines on the damping (square root) part of the curve of growth (see Section 4.2 and Figure 4.4). Figure 4.17 illustrates the extremely broad damping wings of K I at mid–late L spectral types.

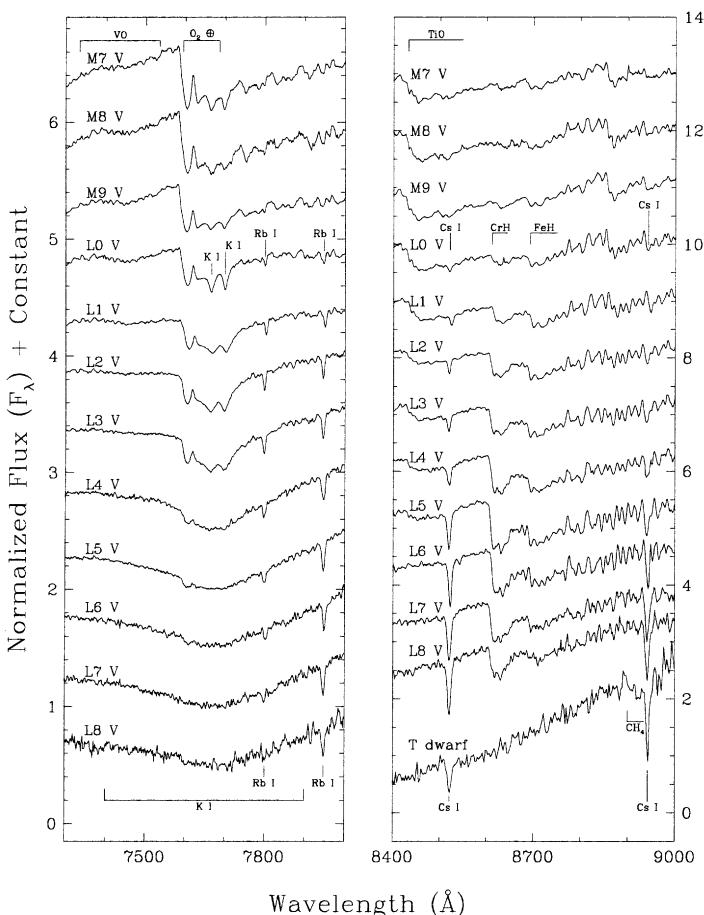


Figure 4.17. A series of late M and L dwarf spectra showing the increasingly strong alkali metal resonance lines with decreasing temperature. Hydride bands of CrH and FeH are also visible. (Figure courtesy of J. D. Kirkpatrick and the *Astrophysical Journal* [K2].)

Table 4.3. Temperatures of L dwarfs.

Spectral type	T_{eff}	Prototype ¹
L0	2,200	2MASP J0345+2540
L2	2,000	Kelu 1
L4	1,800	2MASS J1155+2307
L6	1,600	2MASS J0850+1057
L8	1,400	2MASS J1632+1904

¹ The names of the prototypes from the 2MASS survey are defined in [K2]. Kelu 1 is discussed in [R3].

Additional work on dust formation in very low temperature atmospheres has been carried out by Fegley and Lodders [F1] and Burrows and Sharp [B8], who computed atmospheres including some 120 condensates and extending over a temperature range from 2,200 K down to 100 K. Those calculations allow a first estimate of the L dwarf temperature scale [K2]. TiO is predicted to condense onto grains of perovskite (CaTiO_3) at temperatures between 2,300 and 2,000 K; in comparison, the TiO bands are observed to reach maximum strength at a spectral type of M7–8 and have disappeared (except for the 8,432 Å band) by type L2. Vanadium oxide is next to disappear, forming solid VO at temperatures between 1,900 and 1,700 K; that species is strongest at type M9.5 and disappears at type L4. CrH is expected to be entirely in the metallic phase at \sim 1,400 K, and Figure 4.17 shows that the 8,611 Å band decreases in strength beyond type L5. All of the alkali lines are predicted to form chlorides at temperatures below 1,200 K, although sodium may condense onto high albite ($\text{NaAlSi}_3\text{O}_8$) at somewhat higher temperatures; the observations show the 8,192/8,196 Å doublet as barely visible by spectral type L8. Finally, carbon is predominantly found in CO among the L dwarfs, but preferentially forms methane at a temperature between 1,500 and 1,200 K. This point marks the transition between spectral type L and type T (see Chapter 6). Combining these diagnostics with the observed change in the overall spectral energy distribution leads to the tentative temperature scale outlined in Table 4.3. The available evidence suggests that L dwarfs have radii near $0.1 R_\odot$, and luminosities ranging from $-3.5 < \log L/L_\odot < -4.7$; the luminosity depends critically on the age at young ages. Masses are less than about $0.08 M_\odot$, although few direct measurements are yet available.

As described in Chapter 6, work is proceeding to better define these quantities. In general, theoretical models for cool, methane-dominated (T-type) brown dwarfs appear on a firmer footing than studies of the hotter L dwarfs, although this may reflect the fact that very few are known. In addition to the Burrows models, Marley *et al.* [M2] and Allard *et al.* [A5] describe models of very low-temperature atmospheres. These are compared against the observations and discussed in more detail in Chapter 6. The modelling of L dwarf and T dwarf (and extrasolar giant planet) atmospheres is a very young and active field, and the reader will undoubtedly find many new references not mentioned here.

4.8 SUMMARY

Atmospheric parameters of M and L dwarfs may be found by matching synthetic spectra from model atmosphere calculations to observed data. High resolution spectral data allow quite accurate determinations, but even photometry (particularly in the infrared) can be used with some success. The models are now good enough that fine analysis of very high-resolution spectra allows metallicity, gravity and temperature to be determined with good precision. Lacking this data, gravity may be found by indirect methods using luminosity and temperature information to infer mass and radius. Metallicity may be determined by comparing the molecular band-strengths of hydrides (CaH) to double metal species (TiO). Thus, the temperature, gravity and metallicity of a low-mass star are now accessible using optical and infrared photometry and spectroscopy together with the current generation of atmospheric models. Improvements in the next few years will come with the acquisition of higher-resolution optical and infrared spectra, and from continuing work on the molecular and dust opacity data and the treatment of the opacities in the models.

4.9 REFERENCES

- A1 Allard, F., 1990, PhD thesis, University of Heidelberg.
- A2 Allard, F., 2000, in *Very Low Mass Stars and Brown Dwarfs in Stellar Clusters and Associations*, ed. R. Rebolo (Cambridge University Press), p. 144.
- A3 Allard, F., Hauschildt, P. H., 1995, *ApJ*, **445**, 433.
- A4 Allard, F., Hauschildt, P. H., Alexander, D. R. A., Starrfield, S., 1997, *ARA&A*, **35**, 137.
- A5 Allard, F., Hauschildt, P. H., Baraffe, I., Chabrier, G., 1996, *ApJ*, **465**, L123.
- A6 Auman, J., 1969, *ApJ*, **157**, 799.
- B1 Berriman, G. B., Reid, I. N., 1987, *MNRAS*, **227**, 315.
- B2 Berriman, G. B., Reid, I. N., Leggett, S. K., 1992, *ApJ*, **392**, L31.
- B3 Bessell, M. S., 1982, *Pub. Ast. Soc. Australia*, **4**, 417.
- B4 Böhm-Vitense, E., 1989, *Introduction to Stellar Astrophysics*, Volume 2, *Stellar Atmospheres*, Cambridge University Press.
- B5 Borysow, A., 1993, in *Molecules in the Stellar Environment*, ed. U. G. Jørgensen (Lecture Notes in Physics, 428), Springer, p. 209.
- B6 Brett, J. M., 1995a, *A&P*, **295**, 736.
- B7 Brett, J. M., 1995b, *A&AS*, **109**, 263.
- B8 Burrows, A., Sharp, C. M., 1999, *ApJ*, **512**, 843.
- C1 Carbon, D. F., 1979, *ARA&A*, **17**, 513.
- C2 Clemens, J. C., Reid, I. N., Gizis, J. E., O'Brien, M. S., 1998, *ApJ*, **496**, 352.
- C3 Cottrell, P. L., 1978, *ApJ*, **223**, 544.
- D1 Deutsch, A. J., 1966, in *Abundance Determinations in Stellar Spectra*, ed. H. Huben et al. (IAU Symposium No. 26), 112.
- F1 Fegley, B., Lodders, K., 1996, *ApJ*, **472**, L37.
- G1 Gingerich, O., Kumar, S. S., 1964, *AJ*, **69**, 139.
- G2 Gizis, J. E., 1997, *AJ*, **113**, 806.
- G3 Gustafsson, B., Bell, R. A., Eriksson, K., Nordlund, A., 1975, *A&A*, **42**, 407.
- H1 Hauschildt, P. H., Allard, F., Baron, E., 1999, *ApJ*, **512**, 377.

- H2 Hawley, S. L., Fisher, G. H., 1992, *ApJS*, **78**, 565.
- H3 Hearnshaw, J. B., 1986, *The Analysis of Starlight*, Cambridge University Press.
- H4 Henry, T. J., McCarthy, D. W., 1993, *AJ*, **106**, 773.
- J1 Johns-Krull, C. M., Valenti, J. A., 1996, *ApJ*, **459**, L95.
- J2 Jones, H. R. A., Tsuji, T., 1997, *ApJ*, **480**, L39.
- K1 Kirkpatrick, J. D., Kelly, D. M., Rieke, G. H., Liebert, J., Allard, F., Wehrse, R., 1993, *ApJ*, **402**, 643.
- K2 Kirkpatrick, J. D., Reid, I. N., Liebert, J., Cutri, R. M., Nelson, B., Beichman, C. A., Dahn, C. C., Monet, D. G., Gizis, J. E., Skrutskie, M. F., 1999, *ApJ*, **519**, 802.
- K3 Kurucz, R. L., 1970, *ATLAS: A Computer Program for Calculating Model Stellar Atmospheres*, Smithsonian Astrophysical Observatory Special Report No. 309, Cambridge.
- L1 Leggett, S. K., 1992, *ApJS*, **82**, 351.
- L2 Leggett, S. K., Allard, F., Berriman, G., Cahn, C. C., Hauschildt, P. H., 1996, *ApJS*, **104**, 117.
- L3 Leggett, S. K., Allard, F., Hauschildt, P. H., 1998, *ApJ*, **509**, 836.
- L4 Lenzuni, P., Chernoff, D. F., Salpeter, E. E., 1991, *ApJS*, **76**, 759.
- L5 Linsky, J. L., 1969, *ApJ*, **156**, 989.
- L6 Leggett, S. K., Allard, F., Dahn, C., Hanschildt, P. H., Kerr, T. H., Rayner, J., 2000, *ApJ*, **535**, 965.
- M1 Maceroni, C., Rucinski, S. M., 1997, *PASP*, **109**, 782.
- M2 Marley, M. S., Saumon, D., Guillot, T., Freedman, R. S., Hubbard, W. B., Burrows, A., Lunine, J. I., 1996, *Science*, **272**, 1919.
- M3 Mihalas, D., 1978, *Stellar Atmospheres*, 2nd edition, W. H. Freeman and Company.
- M4 Mould, J. R., 1976, *A&A*, **48**, 443.
- M5 Mould, J. R., Hyland, A. R., 1976, *ApJ*, **208**, 399.
- M6 Mould, J. R., Wyckoff, S., 1978, *MNRAS*, **182**, 63.
- P1 Press, W. H., Tenkolsky, S. A., Vetterling, W. T., Flannery, B. P., 1992, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition, Cambridge University Press.
- R1 Reid, I. N., Gizis, J. E., 1997, *AJ*, **113**, 2246.
- R2 Ruan, K., 1991, Ph.D. thesis, National University of Australia (also reported as Kui (1991)).
- R3 Ruiz, M. T., Leggett, S. K., Allard, F., 1997, *ApJ*, **491**, L107.
- R4 Reid, I. N., Hawley, S. L., Gizis, J. E., 1995, *AJ*, **110**, 1838.
- S1 Schweitzer, A., Hauschildt, P. H., Allard, F., Basri, G., 1996, *MNRAS*, **283**, 821.
- T1 Tinney, C. G., Mould, J. R., Reid, I. N., 1993, *AJ*, **105**, 1045.
- T2 Tsuji, T., 1966, *PASJ*, **18**, 127.
- T3 Tsuji, T., Ohnaka, K., Aoki, W., 1996a, *A&A*, **305**, L1.
- T4 Tsuji, T., Ohnaka, K., Aoki, W., Nakajima, T., 1996b, *A&A*, **308**, L29.
- V1 Valenti, J. A., Piskunov, N., Johns-Krull, C. M., 1998, *ApJ*, **498**, 851.
- W1 Woolf, V. M., Wallerstein, G., 2005, *MNRAS*, **356**, 963.

4.10 HOMEWORK PROBLEMS

Problem 1

The photospheric measurements used on low-medium resolution spectra to classify and analyse M dwarfs are often called ‘indices’. Two of the most common are the

TiO5 index, measuring the ratio of the flux at 7,120 Å to the flux at 7,050 Å (the full TiO bandhead depth) and the CaH2 index, measuring the ratio of the flux at 6,830 Å to the 7,050 Å flux (the adjacent CaH bandhead depth). These and several other indices were used originally in the PMSU survey as described in [R4]. To investigate these indices, download the M0–M9 dwarf SDSS template spectra (courtesy of J. Bochanski) available at the NLDS2 data website (<http://www.stsci.edu/~inr/nldsdat.html>) and use the definitions from [R4] to plot TiO5 and CaH2 against spectral type. Also plot against temperature and ($V-I$) colour from Table 4.1. Discuss the use of these indices as indicators of temperature, colour and spectral type for disk dwarfs.

Problem 2

Section 4.6.4 asserts that the hydride bands remain strong in metal-poor stars, while the oxide bands weaken significantly. First, summarize this argument. Then, download the subdwarf template spectra from the same website and measure TiO5 and CaH2 as in Problem 1. Plot CaH2 against TiO5 for the disk dwarfs and the subdwarfs to reproduce the metallicity effect (see also Figure 10.4).

Problem 3

Investigation of line formation and the curve of growth for a simple Lorentz profile of the form

$$I(w) = (\gamma/2\pi)/[(w - w_0)^2 + (0.5\gamma)^2]$$

where γ is the full-width half-maximum of the line profile and w is the frequency. As described in Section 4.1 (pp. 150–152), the Lorentz profile resulting from van der Waals interactions in the high-pressure M and L dwarf atmospheres often dominates the line broadening. To simulate line absorption, we will conduct a Monte-Carlo simulation where the wavelength at which the absorption occurs is chosen from a Lorentz profile probability distribution.

- (a) Show that the Lorentz profile can be written in the form

$$f(x) = C/[1 + (x - x_0)^2/a^2]$$

where C is a constant, a is related to γ and x is a dimensionless wavelength variable. x is often measured in units of the Doppler width of the line, δW_D ,

$$\delta w/w = \xi/c \rightarrow \delta W_D = \xi_0 w_0/c$$

where ξ_0 is the Doppler velocity = $\text{sqrt}(2kT/m) = 12.85(T/10^4 \text{ A})^{0.5} \text{ km/sec}$, w_0 is the line center frequency and A is the atomic weight of the absorbing atom.

- (b) A random number chosen on a uniform distribution may be converted to one chosen from a specific probability distribution (see e.g., discussion in Numerical Recipes [P1], Section 7.2). To do so, we must evaluate the inverse indefinite

integral of the Lorentz profile. Show that the form $f(x)$ given in part (a) integrates to an arctangent, so that the inverse indefinite integral we seek is given by:

$$x = a \tan(\pi U) + x_0$$

where U is a random number drawn uniformly between 0 and 1.

- (c) Write a program to choose randomly from a Lorentz distribution for an arbitrary number of times, and use the program to plot the line profile assuming that each time a random wavelength is chosen, an absorption occurs which removes one flux unit. Use an initial continuum of 1,000 flux units. Start with $a = 1$. Produce plots choosing between a few hundred and a few million times. Note that the flux cannot go below zero. To make the plotting feasible, only consider wavelengths between -100 to 100 wavelength units of the center. (Hint: in IDL, use routines ‘randomu’ and ‘histogram’.)
- (d) Now calculate the curve of growth, a plot of \log (equivalent width) against \log (number of absorptions). Recall from Figure 4.4, that there is a linear part, a flat part and a square root part on the curve of growth. Can you identify all of these parts on your graph? If not, why not? Also, what approximation are we making that eventually will make the profile inaccurate? (Hint: the equivalent width is an integral of the line under the continuum.) You don’t need an elaborate integration method. Simple trapezoidal integration assigning the flux to its value at the computed points is adequate (in IDL, use routine ‘total’).
- (e) What happens to the curve of growth for values of $a = 0.1$ and $a = 10$ (show plots)? Why? (Hint: what does a correspond to?).

5

Stellar activity

5.1 INTRODUCTION

The presence of strong magnetic fields on the stellar surface manifests itself in many ways, referred to collectively as ‘stellar activity’. The magnetic field is probably formed by the action of an interior dynamo, although its workings are not well understood. On the surface, the field may be organised in active regions, as on the Sun, leading to starspots, plages, prominences, and other solar-type phenomena. The magnetic field heats the tenuous atmospheric layers above the photosphere by non-radiative processes, and these heated layers are seen as the chromosphere, transition region and corona, giving rise to readily recognisable emission lines in the optical, ultraviolet and X-ray spectral regions. Radiation from the outer atmosphere may also feed back into the stellar photosphere, causing secondary effects. Perhaps the most spectacular manifestations of stellar activity are the occasional flares, during which the star emits an enormous amount of energy in a very short time. Some flares are accompanied by mass ejections which affect both the star and its immediate neighbourhood. Clearly, stellar activity has a profound effect on the surface layers which we observe. The magnetic field may also play an important role in the formation, interior structure and evolution of the star, although our understanding of these processes is in its infancy.

In this chapter we first provide a brief overview of current dynamo theories to provide context for understanding the observations. We then examine the evidence for magnetic fields on M dwarfs from direct observation of the photosphere and from the presence and structure of the outer atmosphere. Observations of solar-like activity, including spots and flares, are discussed and contrasted with the solar case. An extensive section on global activity properties observed in large samples of M dwarfs – both in the field and in clusters – describes the current observational situation. We finish with a brief discussion of new observational results for the

lowest mass stars and brown dwarfs, and a summary attempting to construct a coherent picture of stellar activity in low-mass dwarfs.

5.2 THE MAGNETIC FIELD

5.2.1 Dynamo theory

Galileo first observed sunspots in the early 1600s, but it was George Ellery Hale – using the spectroheliograph at Mount Wilson Observatory in the early 1900s – who identified them as regions of strong magnetic field [H1]. The modern theory for the production of the field began with the work of Parker [P1], who proposed a dynamo mechanism for generating and sustaining magnetic fields in the convection zone of the Sun. This model and its many subsequent revisions have come to be known as the $\alpha\Omega$ dynamo. The nomenclature refers to the two forces operating on the plasma. An initial poloidal field is subject to shearing forces by the interior radial differential rotation in the Sun, resulting in the generation of an internal toroidal field (the Ω effect).¹ The toroidal field suffers small-scale cyclonic motions with non-zero helicity, generating a new poloidal field by the α effect:

$$\delta B_{pol}/\delta t = \alpha B_{tor} \quad (5.1)$$

where B_{pol} and B_{tor} refer to the poloidal and toroidal magnetic fields respectively, and α is a term that depends on the kinetic helicity of the plasma (the rotational velocity field and its curl). DeLuca and Gilman [D1] and Durney *et al.* [D8] provide detailed descriptions and derivations of the magnetohydrodynamic equations describing the α effect. Note that both the Ω and α effects are a consequence of stellar rotation.

The toroidal field is twisted by the turbulent convection zone creating bundles of field which break off from the overall structure. Magnetic buoyancy causes these ‘flux tubes’ to rise to the surface and appear as loop-like structures, with opposite polarity at each of their footpoints. Individual flux tubes occur over much of the solar surface (although they cover only a small area in total). Sunspots are organised areas where many flux tubes have grouped together to make a large and noticeable magnetic region (or ‘active region’). The latitude dependence of sunspots – which are observed to appear at mid-latitudes early in the solar cycle and at low latitudes late in the cycle – is explained in $\alpha\Omega$ models by the propagation of the interior dynamo wave towards the equator as the α effect works to cancel the toroidal field and produce the new poloidal field. The observed tilt of sunspots relative to the solar equator can be reproduced with detailed modelling of the forces on flux tubes as they rise through the convection zone. Fisher *et al.* [F5] found that the Coriolis force was the major contributor to the tilt.

As described by Parker [P2], the global toroidal field must be located at the base

¹ Figure 5.1 shows the Babcock model which depends on the surface *latitudinal* differential rotation, and should not be confused with the interior *radial* differential rotation required in the $\alpha\Omega$ model.

of the convection zone, in a region of weak magnetic buoyancy. This allows the toroidal field to be maintained for a timescale of several years, corresponding to the period of the solar cycle. It may also be necessary to allow the flux tubes time to attain the large field amplitude ($\sim 10^4\text{--}10^5$ Gauss) needed in the interior for them to appear with the $\sim 1,000\text{--}2,000$ Gauss fields that are observed at the surface [D7].

The regeneration of the poloidal field with opposite polarity marks the end of the 11-year solar cycle which is observed. Obtaining the opposite polarity at the end of the cycle places constraints on both the magnitude and sign of α . Leighton [L3] showed that the Sun must have differential rotation which increases inward at a fairly high rate to meet these constraints. Recent measurements indicate that the solar convection zone has nearly constant radial differential rotation at active region latitudes [S4]. These observations have led theorists to locate the regeneration region, where the α -effect operates, at the radiative–convective boundary (the base of the convection zone) where a large differential rotation is expected but cannot yet be directly measured. Thus, the current $\alpha\Omega$ dynamo model has both the toroidal and poloidal fields being generated deep inside the Sun. The surface phenomena which we observe are merely tracers of the interior, where all the action is taking place.

An alternative, phenomenological model of the magnetic field evolution during the solar cycle was proposed by Babcock [B1], using the observed behaviour of sunspots on the solar surface. The Sun rotates faster at the equator than at the poles; this latitudinal differential rotation produces an equatorial toroidal field from an initially simple dipole (poloidal) field. The toroidal field attains maximum shear first at mid-latitudes, so flux tubes (sunspots) appear there at the beginning of the cycle. Later in the cycle, as the toroidal field becomes more concentrated toward the equator by the continued action of surface differential rotation, sunspots appear at lower latitudes, as observed. Because the toroidal field is not strictly azimuthal, but retains some poloidal component, the sunspots rise to the surface slightly tilted (as observed), and with opposite polarity in each hemisphere. As their location migrates toward the equator during the solar cycle, the toroidal components cancel and the poloidal components re-establish a poloidal field with the opposite polarity. These ideas are illustrated in Figure 5.1.

In the Babcock model, the toroidal field is produced by an Ω effect (in this case from *latitudinal* differential rotation), and must be somehow transported to the base of the convection zone where flux tube generation occurs, for the same reasons given above. On the other hand, the regeneration of the poloidal field is accomplished without an α effect, simply as a natural consequence of the surface behaviour. A combination of the Babcock and Leighton models has been described by Durney [D7]; that model also includes the observed effect of poleward meridional circulation [S7] which enhances the re-establishment of the poloidal field.

Two common features of these dynamo models are the presence of differential rotation (hence rotation), and the location of the toroidal field at the interface between the convective and radiative zones in the solar interior. For this reason, the models are often described as ‘shell’ dynamos. Although many other details

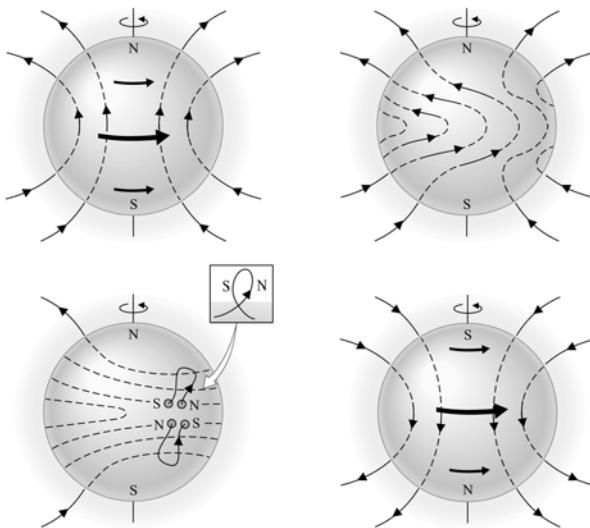


Figure 5.1. The Babcock model of the magnetic dynamo. (a) The solar magnetic field is initially a poloidal field. (b) Differential rotation drags the ‘frozen-in’ magnetic field lines around the Sun, converting the poloidal field into a toroidal field. (c) Turbulent convection twists the field lines into magnetic ropes, causing them to rise to the surface as sunspots, the polarity of the lead spots corresponding to the original polarity of the poloidal field. (d) As the cycle progresses, successive sunspot groups migrate toward the equator where magnetic field reconnection re-establishes the poloidal field, but with the original polarity reversed. (From Carroll and Ostlie, *Introduction to Modern Astrophysics*, Addison Wesley Longman, 1996.)

differ, these features are apparently required in order to produce the observed global, cyclic magnetic field on the Sun.

The cyclic behaviour is most easily observed in the sunspots, but also includes effects on the solar luminosity and the distribution of luminosity with wavelength. In particular, the Sun has a bluer colour temperature and emits more total luminosity at solar maximum. (The maximum of the solar cycle is defined as the time of strongest magnetic activity – largest spot coverage, most energetic flares, and so on). However, the Sun also possesses other, less obvious, magnetic features, including the magnetic network and the intranetwork field. The magnetic network is thought to be produced by the break-up of active regions, and is thus related to the global cyclic field. The intranetwork field, in contrast, does not appear to vary much during the solar cycle, and Durney has suggested that it is generated by a turbulent dynamo, unrelated to the large-scale global dynamo which is responsible for the cyclic activity [D8]. Rosner had earlier proposed a ‘distributed dynamo’ model that incorporated some of these ideas to explain X-ray activity in late type stars [R6]. Recent theoretical work by Bercik *et al.* shows that a turbulent dynamo may be able to explain the basal chromospheric and X-ray emission in F0–M0 stars [B14], [P9].

A turbulent dynamo produces magnetic fields by random convective motions in the convection zone, and does not require rotation (or differential rotation) or a

radiative–convective boundary layer for its operation. The field is not stored for long periods, or organised in either time or space. Field is generated, and quickly forms flux tubes which rise to the surface, appear for a short time, merge with other regions of opposite polarity and are destroyed. No cycles are expected, and the coverage of the active regions should be rather uniform over the surface, rather than concentrated at mid–low latitudes. These are all features which are observed in the solar intranetwork field. The attractive feature of the turbulent dynamo model for low-mass stars is that it can operate under conditions where the interior is completely convective (that is, in dwarfs below $\sim 0.25 M_{\odot}$; see Chapter 3). In these objects, no radiative–convective boundary layer exists for storing the global toroidal field required by the shell dynamo models. The observational predictions of the turbulent dynamo are: a weaker (or no) dependence on rotation, no evidence for cyclic behaviour; and uniform coverage of active regions. As we shall see, these are observed features of the magnetic activity on M dwarfs. There remains some question as to whether the large field strengths which have been observed can be generated by a turbulent dynamo, although recent simulations by Dobler *et al.* [D9] of magnetic field generation in fully convective rotating spheres indicate that it may be possible to generate large-scale global fields. Turbulent dynamo theory is currently a very active area of research and more results applicable to magnetic field generation in low-mass stars should be forthcoming in the near future.

5.2.2 Magnetic field observations

Direct measurements of the magnetic field are accomplished by utilising the Zeeman effect, that is, the splitting of degenerate atomic levels into separate components in the presence of a magnetic field. The normal Zeeman effect results from splitting the $(2l + 1)$ degenerate m_l levels of the electron orbital with quantum number l . The selection rule $\Delta m_l = 0, \pm 1$ means there are three spectral lines formed in transitions to this l level, as shown in Figure 5.2.

The $\Delta m_l = 0$ transition is known as the π component of the line, and retains the original wavelength (in this simple case; more complicated levels can have several π components with small wavelength shifts). The wavelength offsets of the $\Delta m_l = \pm 1$ levels, known as the σ components, are given by Johns-Krull and Valenti [J2] as

$$\Delta\lambda = \pm \frac{e}{4\pi mc} \lambda^2 g_{eff} |B| \quad (5.2)$$

To facilitate observation, the splitting should be as large as possible, and lines are chosen with large Lande g_{eff} factors (magnetically sensitive levels), and long wavelength. However, infrared ($\lambda \sim 1\text{--}5 \mu\text{m}$) spectrographs have typically been less sensitive than optical spectrographs (a situation which is gradually changing), so there is a trade-off in choosing the lines to observe.

Additional measurement uncertainty is introduced by the presence of thermal, turbulent, collisional and rotational broadening of the lines, which tend to mask the magnetic Zeeman signature. Careful analysis of the line profiles using model atmospheres and spectral synthesis (see Chapter 4) is required to accurately recover the

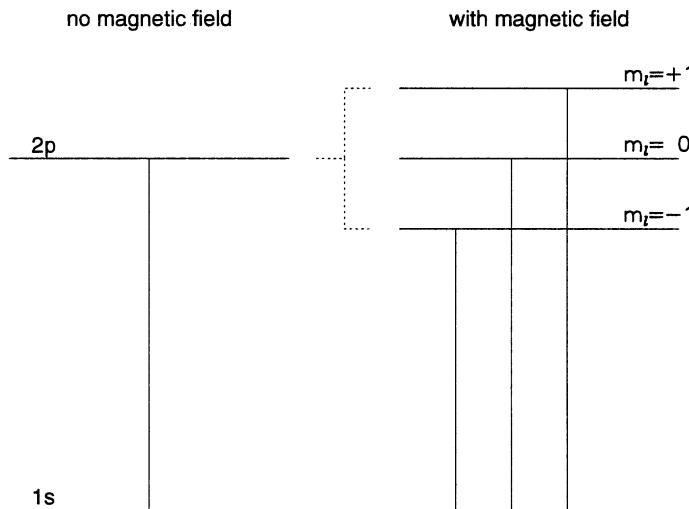


Figure 5.2. The splitting of a degenerate atomic level into its magnetic sub-levels in the presence of a magnetic field. The original transition formed only one spectral line; in the presence of a magnetic field three spectral lines are formed. The magnitude of the splitting depends on the Lande g_{eff} factor for the level, and on the strength of the magnetic field, as shown in equation (5.2).

magnetic-field information from the observed spectra. Methods of analysis fall into two categories: 1) comparison of magnetically sensitive and insensitive lines from the same star; for example, by observing two lines near to one another in wavelength, with very different g_{eff} values as described by Robinson [R4]; or 2) comparison of a magnetically sensitive line in two different stars, one active and the other inactive [J2]. The comparison star must be chosen to be a close match in spectral type, evolutionary state and other physical properties to the star of interest, so that differences in their spectra can be attributed to the effects of the magnetic field. Magnetically insensitive lines can be compared between the two stars to test the validity of this assumption. The magnetic field strength B and the filling factor f (percentage of the stellar surface covered by field) are separately determined by assuming that $f\%$ of the star is covered by field with strength B and the other $(1-f)\%$ has no field. Synthetic model spectra are fit to the data to determine the two parameters B and f .

Field strengths in excess of ~ 600 G have been measured on a number of solar-type G and K stars, using the Robinson method [M1]. M dwarfs, being much fainter, present a significant observational challenge. At present, magnetic fields have been observed on only a few of the brightest dMe dwarfs (recall from Chapter 2, that dMe dwarfs are those with H α in emission, indicating the presence of a chromosphere, and hence magnetic activity). Initial efforts by Saar and collaborators in the mid-1980s centred on Fourier transform spectroscopy, using the Robinson method to observe Ti I lines in the infrared [S2], [S3]. These resulted in detections for the dMe

Table 5.1. Active dMe dwarfs.

Gliese number (or other)	Name	Spectral type
65B	UV Cet	M6e
HD 289114	PZ Mon	K2e
278C	YY Gem	M0e
285	YZ CMi	M4.5e
GJ 2079	DK Leo	K7e
388	AD Leo	M3e
473B	FL Vir	M5.5e
494	DT Vir	M0.5e
517	EQ Vir	K5e
551	Prox Cen	M5.5e
630.1A	CM Dra	M4.5e
644C	VB 8	M7e
719	BY Dra	K7e
752B	VB 10	M8e
799A	AT Mic	M4.5e
803	AU Mic	M0e
820AB	61 Cyg AB	K5e, K7e
873	EV Lac	M3.5e
896A	EQ Peg	M3.5e

dwarf AD Leo and the K5e dwarf EQ Vir.² A more recent application using infrared echelle spectroscopy resulted in detections of DT Vir and YZ CMi [S1].

The second method has been employed with success by Johns-Krull, Valenti and collaborators [J2], [V1]. The magnetically sensitive Fe I line at 8,468.4 Å gives an observed Zeeman splitting of the σ components of $\pm 83.7 \text{ m}\text{\AA} \text{ kG}^{-1}$. Figure 5.3 illustrates the pronounced σ components of the Fe I line in the active stars compared to the inactive stars. The high-resolution coudé echelle observations shown in the figure were used to measure magnetic fields on Gl 729 and EV Lac. Observations of FeH molecular lines hold promise for measuring magnetic fields on cooler M dwarfs [V6].

Current results are summarised in Table 5.2, indicating that active M dwarfs generally have field strengths of a few thousand Gauss covering approximately half of the stellar surface. As more stars are observed, comparison of field strength and filling factor with mass, spectral type, and atmospheric diagnostics such as H α emission and X-ray emission will become possible.

The most recent observations indicate that one-component models – where a single field strength and covering factor are assumed – do not fit the data particularly well (see Figure 5.4). A range of field strengths and filling factors are required to fit the excessively broad σ components that are observed. Evidently the surface

² Active stars are commonly referred to by their variable-star names. Table 5.1 shows the Gliese numbers and variable-star names for a number of well-known dKe and dMe stars.

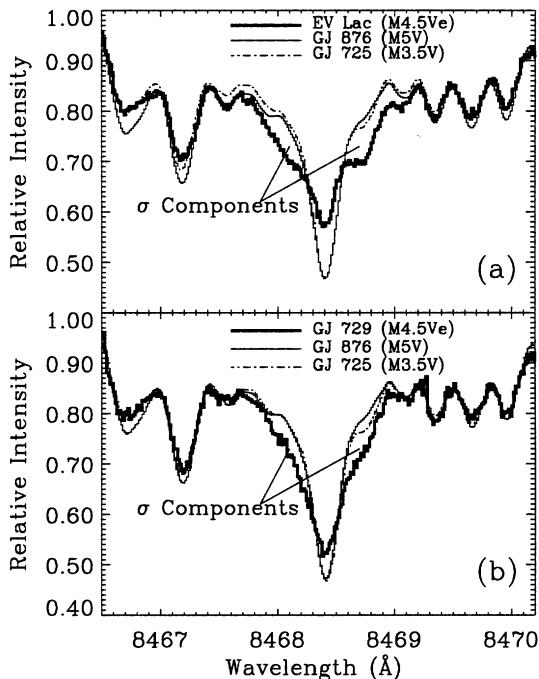


Figure 5.3. The Zeeman-broadened Fe I 8,468 Å line observed in the active stars EV Lac and Gl 729. The σ components are clearly visible when the line profile is compared with inactive stars of similar spectral type. (Figure courtesy of C. Johns-Krull and the *Astrophysical Journal*.)

Table 5.2. Magnetic field measurements of dMe stars.

Name	Spectral type	B (Gauss)	f(%)
Gl 171.2a	K2e	2,800	60
EQ Vir	K5e	2,500	80
DT Vir	M0.5e	3,000	50
AD Leo	M3e	3,800	73
		4,000	60
Gl 729	M3.5e	2,600	50
EV Lac	M3.5e	3,800	50
YZ CMi	M4.5e	4,200	67

magnetic field has a complicated spatial structure. Another, perhaps concurrent, possibility is that the field strength changes with height over the line formation region.

Additional information on the spatial (surface) structure can be obtained by using the polarisation signature of the Zeeman effect. The π component of the

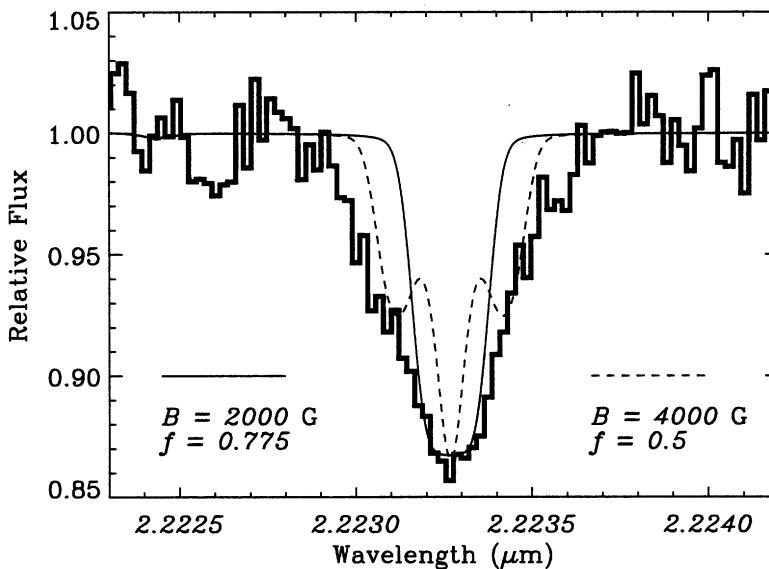


Figure 5.4. A single field strength and filling factor are insufficient to characterise the observed, Zeeman-broadened, line profile in AD Leo. (Figure courtesy of C. Johns-Krull.)

magnetically sensitive line is linearly polarised, while the σ components are circularly polarised; one component is left circularly polarised in regions of positive polarity field while the other is right circularly polarised. In regions of negative polarity field the opposite is true, so if the star is equally covered with regions of positive and negative polarity field (by uniform coverage of spots which represent both polarities), the net circular polarisation signature is zero. Thus, circular polarisation measurements are useful primarily to measure globally organised fields, such as the large dipole fields seen in Ap stars and in individual sunspots. Such measurements for M dwarfs indicate that there is no evidence for an organised global field on Gl 729 at the 1% level on daily or yearly timescales [V1]. If the field is organised into large active regions, as on the Sun, the circular polarisation signature should vary slightly as these regions rotate in and out of view. The fact that this variation is not observed at such a stringent level argues that the field may be composed of many small active regions, so that statistical fluctuations are not important.

The time-dependent studies in [V1] show that the fields on Gl 729 and AD Leo do not vary on timescales of days (similar to the rotation period) or years (similar to a solar cycle period). These magnetic field measurements are in agreement with other chromospheric and coronal indicators which show that solar-like magnetic cycles have not yet been observed on dMe stars. However, the H α emission varies considerably during the course of the observations. Apparently, the chromospheric heating is not closely coupled to changes in the surface magnetic field – at least on these timescales.

5.3 THE OUTER ATMOSPHERE

The outer atmosphere of the Sun begins at the temperature minimum region (TMR) and extends outward through the chromosphere, transition region and corona. Figure 5.5 shows the distribution of temperature with height and column mass for the solar VAL3C chromospheric model [V4]. Travelling outwards from the photosphere (height = 0 marks $\tau_{std} = 1$ in the continuum, the photospheric ‘boundary’), the temperature first decreases through the TMR where the energy balance still reflects radiative equilibrium. At the height where magnetic (non-radiative) heating becomes important in the overall energy balance, the temperature reverses its decline and begins to rise, reaching a plateau at $\sim 7,000$ K (the chromosphere) before rising abruptly through the transition region and levelling out at $\sim 10^6$ K in the corona (not shown in the figure). VAL3C is a semi-empirical model, found by postulating an atmospheric structure, computing the emergent brightness temperatures in numerous diagnostics, comparing these with observations, and changing the structure until a good match is obtained between the model and the data. The locations of formation for the principal diagnostic lines in the solar chromosphere – H α , Ca II K, and Mg II k – are indicated on the figure. Chromospheric models for AD Leo are superposed on the solar model, illustrating the compressed nature of the chromosphere in the lower mass, higher gravity M dwarfs.

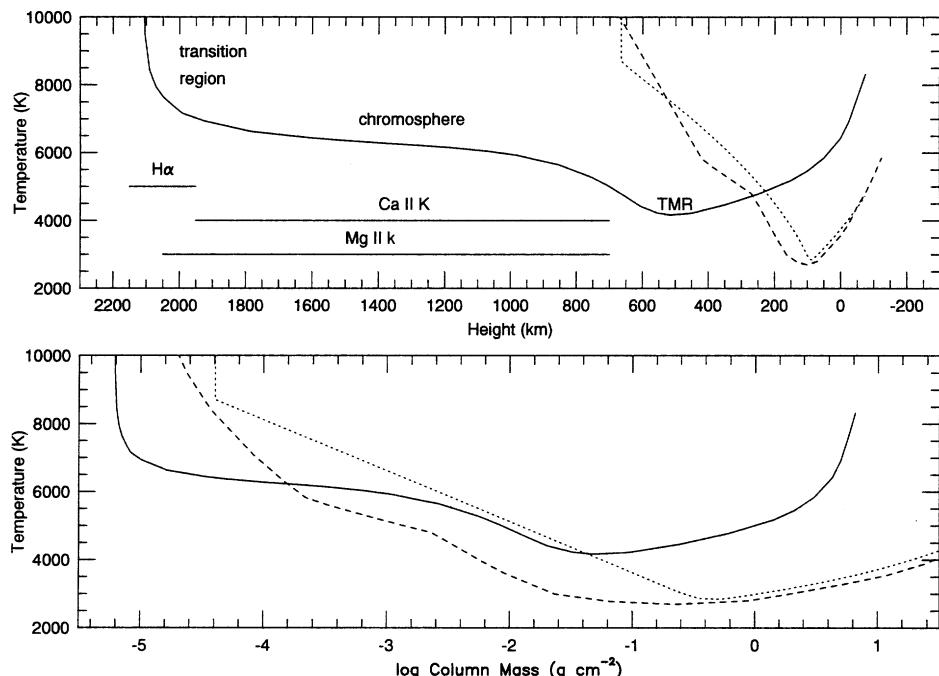


Figure 5.5. The VAL3C semi-empirical model of the solar chromosphere is the solid line. Two chromospheric models for AD Leo are shown: a semi-empirical model from Mauas and Falchi [M6] (dashed line), and a schematic model from Hawley and Fisher [H5] (dotted line).

If the solar atmosphere was in strict radiative equilibrium with the energy diffusing outward from the hot core, there would be no temperature rise outside of the photosphere, and the temperature would slowly approach interstellar values as the atmospheric density and radiative flux decreased. The thermal bifurcation model for the solar atmospheric proposed by Ayres and collaborators asserts that radiative equilibrium does occur in the quiet regions of the photosphere, and that only in the active regions (within magnetic flux tubes) is there a chromospheric temperature rise [A6], [A7]. The solar model shown in Figure 5.5 is then misleading, representing a global average over disparate regions and matching the physical properties of neither. Unfortunately, we are not yet able to resolve surface features on M dwarfs, and must use globally averaged values in our analysis.³ In the following section, we refer to the outer atmosphere as though it is uniform and ubiquitous over the surface, keeping in mind that lateral inhomogeneity of considerable complexity is almost certainly present.

5.3.1 The chromosphere

A distinctive feature of the chromosphere is the extended temperature plateau where the density falls by several orders of magnitude while the temperature remains nearly constant. The plateau is the result of a balance between the non-radiative (magnetic) heating, and radiative cooling, principally from collisionally excited emission lines of hydrogen, ionised calcium and ionised magnesium. The reason for the plateau was succinctly described by Ayres [A5], and is summarised here.

The collisional excitation which drives the cooling results primarily from collisions with electrons, and hence depends linearly on the electron density. Models with constant heating per unit mass (the simplest assumption) thus require nearly constant electron density to maintain the balance between heating and cooling that is required to keep the temperature constant. At chromospheric temperatures, the electrons are supplied by hydrogen. Constant electron density is achieved by a nearly perfect match between the increasing partial ionisation of hydrogen (freeing more electrons into the atmosphere) and the decrease in the total number of hydrogen atoms available imposed by the requirement of hydrostatic equilibrium. The chromospheric temperature plateau thus spans the hydrogen partial-ionisation region, which is quite extended (in height) in low-gravity giants, somewhat smaller in solar-type dwarfs, and very compressed in high-gravity M dwarfs. This provides a natural explanation for the observed correlation between the width of the Ca II K line and luminosity class in G and K stars, known as the Wilson–Bappu effect [W2].

The radiation that characterises the solar chromosphere comes from the same lines that provide the cooling – namely, the Lyman and Balmer series lines of hydrogen and the resonance lines of Ca II and Mg II. In low-mass stars the density is higher, the heating and cooling are larger, and the hydrogen lines are

³ Our inability to resolve surface features also means that the voluminous solar literature on spatially resolved structures such as prominences, spicules and bright points (see [Z2]) is not yet directly applicable to the study of low-mass stars.

particularly strong. The entire Balmer series is often in emission in dMe stars; the primary source of chromospheric cooling switches from the Ca II and Mg II lines in the Sun to the hydrogen Balmer lines in M dwarfs [L7]. In addition, the Ca II resonance lines near 3,900 Å are difficult to observe in these cool stars, as the continuum flux is greatly reduced due to the low effective temperatures. The combination of strong hydrogen lines and ease of detection makes the H α line at 6,563 Å the principal diagnostic of an M dwarf chromosphere, in contrast to the study of chromospheres in solar-type stars where the Ca II K line is the diagnostic of choice.

The current generation of semi-empirical chromospheric models of M dwarfs do not provide a good fit to the data. There is a serious problem in fitting several different lines (Ca II, Mg II and H Balmer and Lyman series) with the same model, which has been attributed to inhomogeneous surface structures on the star. A continuing question is the relative coverage of plage and spot regions. Some recent references are the models shown in Figure 5.5, which fit the hydrogen Balmer lines but predict Ca II as too weak [M6] and too strong [H5] respectively; Giampapa and collaborators [G4], who found that the Mg II emission is too strong when the models predict the correct Ca II flux; and a series of papers by Houdebine and collaborators, showing that a very high coronal pressure (log column mass ~ -2.9 at the beginning of the transition region temperature rise, compared to ~ -4.5 in the models shown in Figure 5.5) is required to match both the hydrogen Balmer and Lyman line fluxes. The largest set of models by the latter authors [H16] had moderate success in reproducing H α and Ca II K observations in early M dwarfs, but they noted that significant plage areas were necessary to describe the continuum variations. There is much work to be done in refining the models, which may lead to a better understanding of the distribution of surface activity on these stars.

The question of what heats the solar chromosphere is still under considerable debate, and very little work has been carried out on low-mass stars. An overview of the solar discussion is given by Kalkofen [K2].) There is agreement that Alfvén waves travelling along the magnetic field lines (in the flux tubes) must play a role, but the details of the models have yet to converge. Acoustic heating (shocks) may play some role on early type stars and the Sun (the so-called basal chromosphere, [S5], [C2]), but is unlikely to be important in the relatively denser and cooler M dwarf atmospheres. The possibility that flaring activity in the corona provides enough downward-directed energy to heat the chromosphere has been explored briefly [C6], although this begs the question of how the flaring is initially produced in the corona. The possibility that the corona itself is heated by extensive flaring activity has also been discussed for active M dwarfs [K6], [G12], [A9].

5.3.2 The transition region

When hydrogen becomes nearly completely ionised, the chromospheric energy balance is disrupted by the sudden lack of collisionally induced cooling, since there is no longer a continuing supply of electrons to drive it. The atmosphere begins to heat up, and soon a situation of thermal runaway occurs. Between about 10^4 K and 10^6 K there are many resonance transitions in abundant elements that

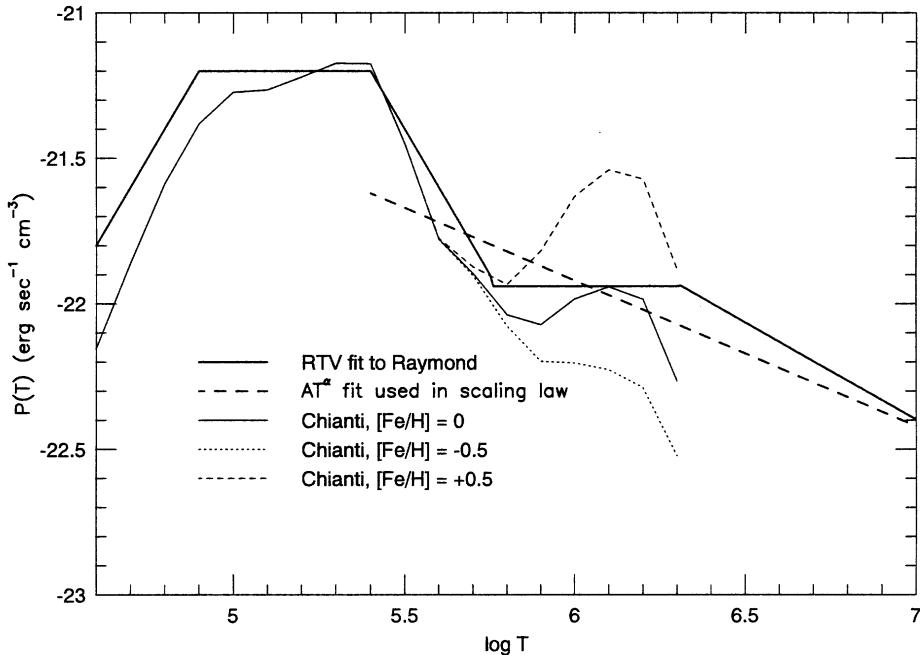


Figure 5.6. Radiative losses are shown as a function of temperature for a hot, optically thin plasma. The RTV analytical piecewise fit to the Raymond and Smith [R2] radiative loss function (solid line) and the AT^α approximation used in the RTV scaling law (long-dashed line) are shown. The other relationships are recent results from the Chianti database [D5], for several metallicities.

contribute strongly to the cooling; in fact, the radiative cooling function from an optically thin, collisionally excited gas peaks at temperatures near 10^5 K (see Figure 5.6; also Figure 5 in [C8]). The heating is not enough to maintain these temperatures, and the temperature must continue to rise (counter-intuitively) until it reaches a million degrees and the cooling is reduced, in the corona. The very thin layer between chromospheric and coronal temperatures is known as the transition region. In this region with an extremely steep temperature gradient, the energy balance is primarily between the heating due to thermal conduction and the radiative cooling:

$$\frac{dF_c}{dz} = -n_e n_H P(T) \quad (5.3)$$

where the Spitzer conductivity $F_c = -\kappa_o T^{5/2} dT/dz$ and $P(T)$ is the radiative loss function shown in Figure 5.6. Fisher and collaborators [F4] described how the temperature structure of the transition region could be obtained by numerically evaluating equation (5.3) for the Sun, and later have applied this formalism to M dwarfs [H5].

The principal emission lines that are produced in the transition region come from upper ionisation states of elements such as C, N, O, S and Si. Commonly

observed lines are in the ultraviolet; for example, C II 1,335 Å, He II 1,640 Å, C III 1,174 Å, C IV 1,550 Å, S IV 1,400 Å, and N V 1,240 Å. In the early 1980s, NASA's International Ultraviolet Explorer (IUE) satellite allowed access to the ultraviolet wavelength region for stellar exploration, and produced a wealth of information for solar-type stars. Since M dwarfs are intrinsically faint, only the nearest and brightest were detected [L7], and even those were faint except during flares [B9], [H4]. Tight correlations between the fluxes emitted in chromospheric and transition region emission lines are found for solar-type stars [A8], but the M dwarfs differ systematically, showing less Mg II flux at a given transition region flux [O1]. Oranje attributed this behaviour to the shifting of chromospheric emission from Mg II into the H Balmer lines in the lowest-mass stars.

In the 1990s, the NASA Extreme Ultraviolet Explorer (EUVE) satellite provided short wavelength ($70 \text{ \AA} < \lambda < 760 \text{ \AA}$) data which probe this temperature regime. Observations of F–M dwarfs are summarised by Mathioudakis *et al.* [M5]. They find good correlation between the strength of the EUV emission lines and other activity parameters such as Mg II and X-ray emission, with the M dwarfs again showing departures in Mg II. However, they propose that the Mg II flux is reduced because the chromosphere is ‘saturated’. As we shall see, ‘saturation’ is currently a common theme in interpreting activity observations; Sections 5.4.1 and 5.5.4 contain more discussion.

5.3.3 The corona

The corona is the outermost part of the stellar atmosphere, extending as much as a stellar radius above the photosphere, with a typical temperature of a few million degrees. Coronal emission is primarily at soft X-ray wavelengths, from collisionally-excited emission lines of high ionisation states of iron and other heavy elements. Magnetic fields confine the (ionised) coronal gas, as evidenced by solar images showing that the X-ray emitting regions are large loop-like structures whose footpoints are magnetic active regions in the photosphere. Emission from solar coronal plasma is also observed along open magnetic field lines; this directed outflow forms the solar wind.

The energy balance within a coronal loop was independently described in the classic papers of Rosner, Tucker and Vaiana (RTV) [R7], and Craig, McClymont and Underwood [C5]. In the simplest case, the heating is constant per unit mass and cooling is due to conduction and radiation. RTV provided a piecewise analytical fit to the radiative losses as a function of temperature, $P(T)$ (from [R2]). Their fit, together with a power-law approximation in temperature with index α ,

$$P(T) = AT^\alpha \quad (5.4)$$

is shown in Figure 5.6. The power-law approximation is valid from $\log T \sim 5.5\text{--}7$. More recent radiative loss calculations using the Chianti database [D5] are also shown, for solar, metal-rich and metal-poor atmospheres. The solar metallicity relationship is still quite close to the RTV fit.

With these approximations, RTV found a relationship between the loop length L , pressure P , and apex temperature T_A which has no free parameters. The ‘coronal loop scaling law’ is

$$T_A = C(PL)^{1/3} \quad (5.5)$$

where the constant, C , depends on the values of A and α in the cooling function. With $A = -18.9$ and $\alpha = -1/2$ as shown in Figure 5.6, the value of $C = 1.4 \times 10^3$ given in RTV is obtained. Solar observations indicate that this loop model reproduces the empirical values for L , P and T quite well. Similar loop models have been used to describe the coronae of M dwarfs, particularly during flares [F3], [H5], [H7], [G3], [C9].

Description of the emitting material in the transition region and corona is accomplished using the formalism of the differential emission measure (DEM). The flux emitted in an optically thin emission line is given by $F = G(T)EM$, where the emission measure EM is an integral over the electron density squared:

$$EM = \int n_e^2 dz \quad (5.6)$$

appropriate for two-body collisional excitation processes. $G(T)$ is a function incorporating the atomic physics, including the elemental abundance, the ionisation equilibrium that determines the abundance of the particular ion, and the collisional excitation rate. The latter two quantities depend sensitively on the temperature. Exact forms for $G(T)$ may be found in [H5]; calculations for a given elemental transition and temperature are straightforward, if tedious. The emission measure is thus an experimentally determined quantity describing how much plasma must be at a given temperature to produce the line fluxes that are observed. Emission measures for individual lines are typically shown as distributions versus temperature, with a characteristic U-shape reflecting the inverse dependence on the ion abundance, which is a sharply peaked function of temperature, as shown in Figure 5.7 for C IV (from [A4]). The maximum fraction of carbon that occurs in the C IV state is $\sim 35\%$, at a temperature of 10^5 K. At lower and higher temperatures, very little carbon is in the C IV state, and hence a very large amount of material would have to be at that temperature to produce the observed line flux. This leads to large inferred emission measures at temperatures far from the peak, and a minimum in the emission measure distribution at the peak formation temperature of the ion.

Observing several ions with different formation temperatures allows the mapping of the amount of emitting material as a function of temperature. Figure 5.8 shows the emission measure distributions for the C II, Si III, Si IV, C IV and N V ions observed with IUE during the strong 1985 flare on AD Leo, and for the C II, C IV and N V observations during a quiescent (non-flaring) time-period on the star. The DEM, $\xi(T)$, is defined from the models as

$$\xi(T) = \frac{n_e^2 T}{|dT/dz|} \quad (5.7)$$

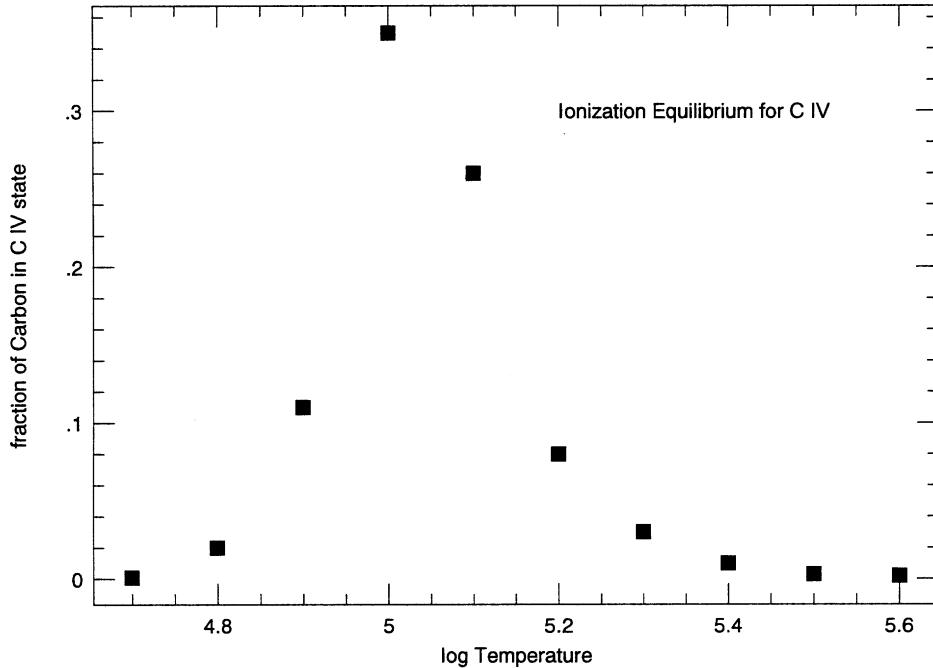


Figure 5.7. The C IV ionisation equilibrium. (From Arnaud and Rothenflug [A4].)

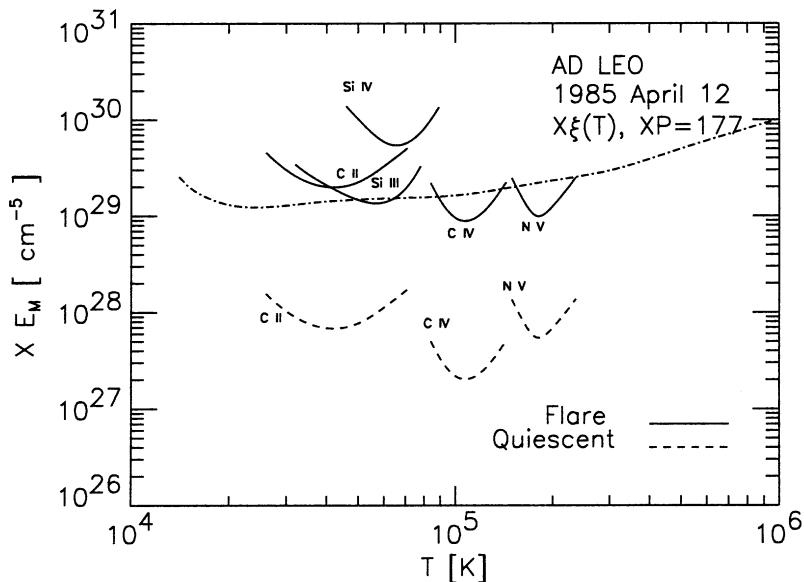


Figure 5.8. Emission measure distributions for several ions observed during a flare on AD Leo, together with a theoretical DEM curve. (Data from [H5], [H6].)

a quantity which has the same units as the emission measure. $\xi(T)$ is a differential quantity, representing the change in the amount of emitting material with the temperature gradient.⁴ The DEM may be easily computed if a model structure (temperature and density as a function of column mass) is available for the transition region and corona. In the transition region, the energy equation (5.3), together with the equation of state and a form for the radiative losses (such as in Figure 5.6) produces a structure which leads to the $\xi(T)$ shown in Figure 5.8. The figure demonstrates that this simple energy balance structure does not agree well with the AD Leo flare data. In the corona, coronal loop scaling laws (as in equation (5.5), but see also [F3] for a description of loop scaling laws applicable during flare conditions) determine the structure allowing $\xi(T)$ to be computed [C9].

The Einstein, EXOSAT and ROSAT satellites have provided extensive information on the X-ray emission from low mass stars. Surprisingly (in the 1980s), M dwarfs were found to be relatively bright X-ray sources. Nearly complete surveys out to several parsecs, together with larger surveys at greater distances, have been carried out. Survey results for early and mid-M dwarfs were described in Chapter 2 (see Fig. 2.24); new results for later M and L dwarfs are given in Section 5.6.

ROSAT observations include minimal spectral information, generally the ratio of emission in a ‘hard’ band (1–2.4 keV) to a ‘soft’ band (0.1–1 keV). The M dwarf ROSAT data are not well fit by the theoretical emission from coronal loops with a single apex temperature. A common practice is to fit the data with two temperature components, one with high $T_A(> 10^7 \text{ K})$ which contributes the harder emission, and one with low $T_A(\sim 10^6 \text{ K})$ which dominates the emission in the softer bandpass. These temperature components have been interpreted as representing two different populations of coronal loops ([G3], but see also [G8] for a cautionary viewpoint). The whole procedure of two-temperature fits is questionable, since very high ISM column densities to the source (indicating significant absorption of the softer emission by interstellar material) and low abundances (thus the emission measures must be higher in order to obtain the observed emission) are also commonly obtained from the fits, and do not agree with other observations [C9]. The difficulty probably lies in over-interpreting the available data; recent observations with Chandra and XMM-Newton allow multi-temperature fits and more reliable abundance determination [V7].

5.4 SPOTS AND FLARES

Obvious manifestations of magnetic activity on the Sun are the dark sunspots (active regions) which move across the solar disk, and the energetic flares which are typically associated with them. In this section we examine the evidence for starspots and stellar flares, and compare their properties with the solar analogues.

⁴ Volume forms of the emission measure and differential emission measure are also in common use, see [C9].

5.4.1 Spots and spot cycles

The solar dynamo gives rise to large active regions where many flux tubes are collected together to form pairs of sunspots, with opposite polarity. The evolution in the number and location of sunspots provides strong observational evidence for an 11-year cycle in the dynamo production of magnetic fields. Spots are a common feature in young, active solar-type stars; observations in open clusters show that the G and K dwarfs often exhibit strong *rotational modulation* in their light curves. Photometric rotational modulation is attributed to the presence of starspots rotating in and out of view, and is distinguished by its periodicity (the rotation period of the star) and smoothly varying, often sinusoidal, shape. Figure 5.9 shows two examples of photometric light curves for solar-type dwarfs from a survey of the Pleiades [K4]. The canonical picture says that as a solar-type star ages, its rotation slows due to angular momentum loss, and its magnetic activity therefore weakens because of the dynamo dependence on rotation. The photometric variation then becomes more difficult to observe, though there are examples of cyclic photometric variability in solar-type field stars [L9]. Magnetic activity in older solar-type dwarfs is more easily monitored in the Ca II K line, produced in plage areas surrounding the starspots. An extensive programme to observe Ca II K line variations in solar type dwarfs was begun in 1966 by Olin Wilson at Mount Wilson Observatory, and has continued through the efforts of Vaughan, Baliunas and collaborators, culminating in the large compilation in [B3]. They found that young stars generally had rapid rotation and high levels of chromospheric activity, but rarely exhibited cyclic behaviour. In contrast, older stars had slower rotation, lower levels of activity, and often exhibited smooth cycles similar to the solar cycle. The Sun falls into the latter category. Evidently the dynamo in these solar-type stars begins to produce cyclic behaviour at some stage in the star's evolution depending on its age, rotation, and perhaps other (as yet unknown) factors. They also found a small subset of stars with

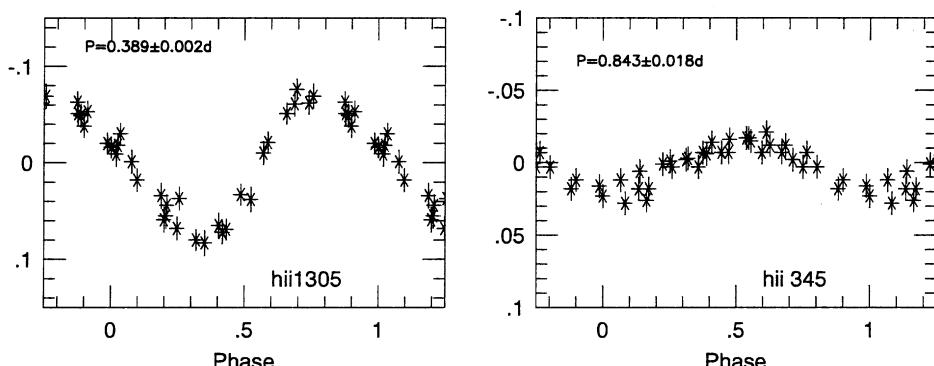


Figure 5.9. Light curves of solar-type Pleiades dwarfs showing photometric modulation typical of starspots rotating across the stellar surface. The derived periods are given on the figures. (Figure courtesy of A. Krishnamurthi [K4].)

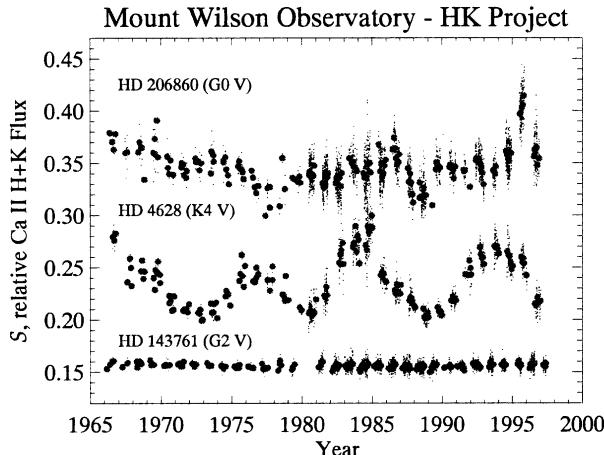


Figure 5.10. Light curves of the Ca II K index in G and K stars, from the Mount Wilson survey. Stars with less activity are more likely to show cycles (compare HD 206860 with HD 4628); a small subset of stars show very low activity and no cycles (HD 143761). (Figure courtesy of S. Baliunas.)

very low levels of activity and evidently no cyclic behaviour. Figure 5.10 plots three of their light curves, illustrating these phenomena.

Close binary stars can be extremely active, as exemplified by the RS CVn (consisting of \approx solar-mass subgiant and dwarf stars) and BY Dra systems (consisting of K-type dwarfs). These tidally-locked systems have fast rotation, which promotes shell-dynamo magnetic activity. They exhibit photometric and chromospheric rotational modulation indicative of large starspots. RS CVn systems have been subjected to extensive observation and analysis using Doppler imaging techniques to dissect the line profiles and model the spot parameters and their distribution on the surface [R3]. There is as yet no conclusive evidence for cyclic behaviour of the magnetic activity in these systems [L2], [P7].

Based on the observations of these other active stars, M dwarfs were expected to exhibit starspots and possibly spot cycles. However, considerable effort in the 1970s and 1980s failed to produce many convincing detections. Although most active stars showed variable light levels, few had the periodic light curves typical of a spot rotating in and out of view. A summary of the sparse observational material is given by Petterson [P6]. The rapidly rotating, early M dwarf Gliese 890 is the best case, with a well-defined modulation having a period of 0.43 days, in phase with the H α emission (minimum light occurs at minimum H α equivalent width) [P5], [Y1]. Spots of moderate amplitude have been observed on AU Mic [T5] and EV Lac [C4]. M dwarfs that show periodic variations in one observing run may appear non-varying in subsequent months or years [R5], [C4]. Figure 5.11 illustrates a time series of light curves for the dMe star AU Mic, a 20-Myr-old M dwarf with a dusty disk (see Section 11.3.2). A clear variation with a period of \sim 4.8 days is

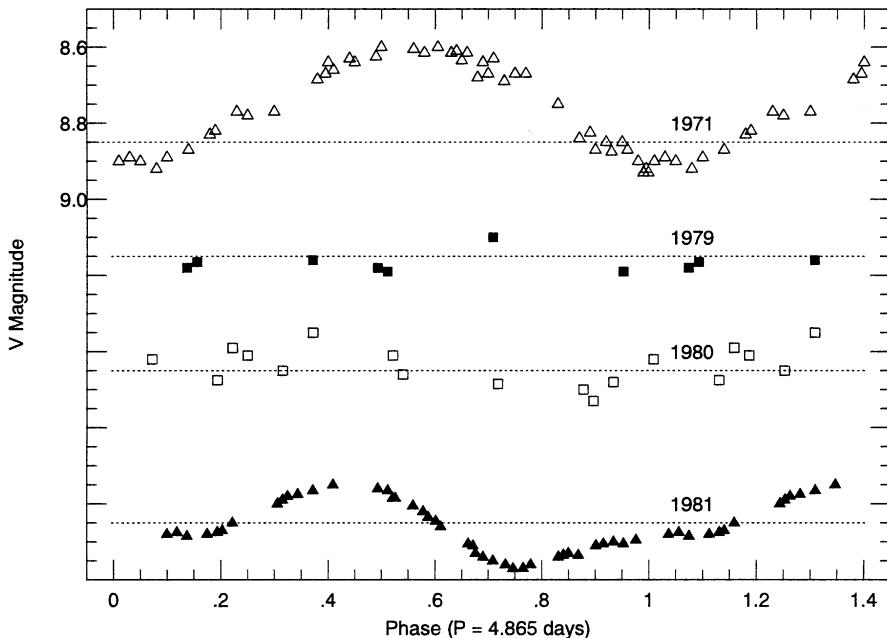


Figure 5.11. Light curves of AU Mic, spanning a period of 10 years. Clear rotational modulation with a period of 4.865 days is seen in 1971 and 1981; the intervening years show variability but no modulation. The data are plotted on the same magnitude scale as indicated in the 1971 light curve. An approximate mean magnitude $V = 8.85$ is shown by the dotted lines for each year. (Data are from [R5].)

seen for 1971 and 1981 (although the phase of light maximum appears to have shifted). Observations in 1979 and 1980 show little or no periodic modulation.

M dwarfs that have variable light levels but no sign of periodic rotational variation may have long periods similar to Proxima Centauri, whose ~ 80 day rotational modulation with an amplitude of ~ 0.01 magnitudes would surely not have been discovered were it not the subject of intense scrutiny by the HST [B6]. They might also represent cases where spots are appearing and disappearing on short time-scales, or are distributed uniformly over the surface, so that little rotational modulation is produced. The discussion in Section 5.2.1 regarding turbulent dynamos suggests that uniform surface coverage is a feature of that model; the spot observations (or lack of them) on M dwarfs support that prediction. Further support for uniform surface coverage is found in the constant polarisation measurements of the magnetic field (Section 5.2.2). An apparent wish to cling to a solar analogy (a few large, isolated spots rather than many small spots), has led many to invoke ‘saturation’ to explain the lack of convincing observations of spots. In this context, ‘saturation’ refers to a surface which is completely covered by large sunspot-like spots, such that no unspotted surface exists [V5]. While the observations do not rule this out, it is also not required, since uniform (and not necessarily

complete) coverage of small spots would produce the same unvarying effect on the light curve.

Photometric observations of young open clusters provide additional support for the view that large, isolated starspots which contrast sharply with the surrounding photosphere are not common on M dwarfs, although G and K dwarfs in the same clusters do show evidence of such behaviour. Only two M dwarfs show rotational modulation in IC 2391 [P4], while spots were found only in stars with $M > 0.5 M_{\odot}$ in the Pleiades [K4]. A large study of IC 2602 [B4] revealed low-level photometric modulation in most of the confirmed members, including several M dwarfs. Further studies of VLM cluster stars found low-amplitude modulation for one star in α Per [M4] and several stars in the Pleiades [T1], [S13]. Apparently spot activity does exist on some M dwarfs, but in most cases at a low level (amplitudes of ~ 0.02 magnitudes), which presents a challenge for detection. In very young clusters such as sigma Ori, the VLM stars and brown dwarfs show more variability, both periodic and randomly fluctuating, which could indicate that both rotational modulation and accretion processes are occurring in these pre-main sequence objects [S14].

An absence of rotational modulation is expected if spots are located only near the rotational poles, where they would be visible during the entire rotation period of the star. Both Doppler imaging and photometric studies of RS CVn stars seem to require polar spots ([H3], [B9], but see [B13] for an opposing viewpoint), primarily to explain stationary features in line profiles and changing levels of maximum brightness in long-term photometric light curves. Young *et al.* [Y1] argued that polar spots were required to model their data for Gliese 890. Theoretical support for polar spots on early-type M dwarfs was found by Buzasi [B12], who performed solar dynamo calculations on $0.4 M_{\odot}$ stars. He found that reasonable values of the rotation velocity led to flux tube emergence at much higher latitude in these stars, compared to the Sun (see also [S6] for a discussion of a similar effect in rapidly rotating solar-type stars). However, Barnes, Collier Cameron and collaborators have obtained Doppler imaging observations of a few bright, active M dwarfs and do not find evidence for polar spots. Instead their reconstructions show starspots spread over all latitudes, in agreement with predictions for a distributed (turbulent) rather than solar-like dynamo [B15], [B16].

If detection of starspot on M dwarfs is difficult, finding spot cycles poses an even greater challenge. Some tenuous evidence exists for long-term photometric variability on timescales of decades [B7], but there are as yet no convincing cases. The solitary M dwarf in the Mount Wilson Ca II K survey does not exhibit cyclic behaviour. If the dynamo operates similarly in M dwarfs as in the solar-type stars, the dMe stars could be examples of the young, active stars from the Mount Wilson study that showed strong activity but little cyclic behaviour. On the other hand, the lack of cycles could be a natural feature of a turbulent, rather than a shell, dynamo dominating the production of the magnetic field, as suggested by the Durney model [D8].

When a star that previously showed a periodic variation stops varying, it often has a *fainter* mean light level. This has been interpreted as an indication that more of

the surface is covered by spots that are darker than the surrounding photosphere – the star has become ‘more active’. Alternatively, the star may have been initially covered with bright plage areas during the period of variation, and the underlying photosphere may be revealed during non-varying periods. Pettersen *et al.* [P7], in their ‘Zebra effect’ paper on BY Dra stars, highlighted the difficulty in distinguishing between a bright star with dark spots and a dark star with bright spots. Support for the latter interpretation may be found in analyses of solar luminosity variations over the course of an activity cycle, which show that the Sun is brightest at activity maximum despite having a larger coverage of sunspots [F7], [H17]. Presumably this represents a stronger contribution from the plage areas (facular network) in the active state. However, the Mount Wilson survey [R1] provided evidence that, while relatively inactive solar-type stars (like the Sun) are brighter at the maximum of the Ca II K activity cycle, the opposite is true for active G–K stars – those stars are darker at activity maximum. Hence, the relative importance of bright plage areas compared to dark spot areas appears to vary systematically with the activity level. If late-type dwarfs mimic the behaviour of active solar-type stars, they would be expected to be darker when more active, in contrast to results for BY Dra stars [P7] and Gliese 890 [Y1], although some authors do claim to find a correlation between lower mean light level and increased flare activity [M7], [C4]. More data are required to sort out the dependence of photometric light variations on activity level in M dwarfs, and hence to clarify the interpretation of observations in terms of cyclic behaviour.

5.4.2 Flares

While spots on M dwarfs remain elusive, flares are obvious and easily observed through even a small telescope. They are distinguished by blue and ultraviolet continuum emission which can increase the brightness of the star by several magnitudes in a manner of seconds. Imagine the feeling of the photometric observer, monitoring a count rate of some few hundred counts/sec on an M dwarf flare star, and suddenly finding that the count rate has become a few hundred *thousand* counts/sec! (Experienced observers say that the adrenaline rush accompanying such observations is the only reason to tolerate the other 99% of the time when the star stubbornly refuses to flare.) The flares on M dwarfs can be remarkably energetic in comparison to the bolometric luminosity of the star, and will often completely dominate the blue, ultraviolet and X-ray emission of the quiet atmosphere. Indeed while the flux in the U band filter may increase by a factor of 100–1,000, the ultraviolet and X-ray emission may increase by up to 10^4 – 10^5 times in brightness for periods of a few minutes or more.

The observations were described briefly in Chapter 2. Flares are identified by the impulsive phase emission in the blue and ultraviolet continuum and by exceedingly strong emission lines in the optical and ultraviolet (see Figure 2.25b). The continuum emission is typically that of a black-body with a temperature of \sim 9,000–10,000 K [M8], [K1], [H5]. The optical emission lines consist of the same lines seen in the quiescent (dMe) spectrum – the hydrogen Balmer lines and Ca II H and K – with the

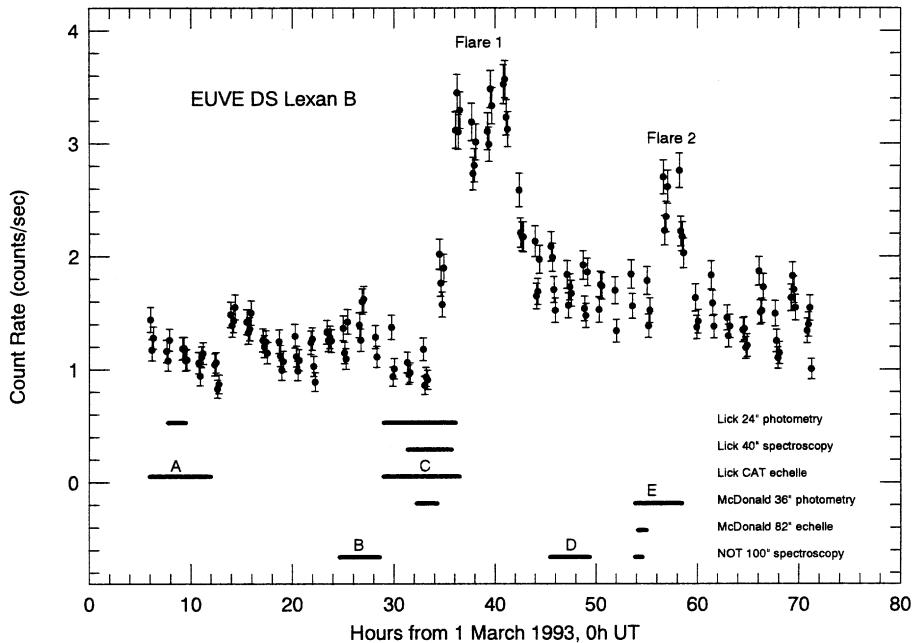


Figure 5.12. Soft X-ray light curve observed with EUVE during a flare on AD Leo. Several ground-based observatories were obtaining photometric and spectroscopic observations at the times indicated. (Data from [H7].)

addition of many short-lived species with higher excitation, such as He I and He II, the Ca II infrared triplet, and numerous singly and doubly ionised metals, all indicating a significant increase in the amount of emitting material at chromospheric, or slightly higher, temperatures. Ultraviolet emission lines such as C IV and N V become much stronger during flares, which also indicates an increase in the amount of material at transition region temperatures. The soft X-ray emission rises more slowly than the optical continuum and emission lines, peaks during the gradual phase when the optical emission has already begun to decay, and has longer rise and decay times than the impulsive phase emission. Figure 5.12 shows the soft X-ray light curve during two flares on AD Leo as observed with the NASA EUVE satellite. Optical (U-band) light curves of the impulsive and extended gradual phases of Flare 1 are illustrated together with the EUVE data and the line flux in the Balmer H γ line in Figure 5.13. Clearly the continuum brightening and optical emission lines precede the soft X-ray emission in this flare.

Impulsive phase timescales are generally seconds to a few minutes, while the gradual phase may last from minutes to hours. More energetic flares typically last longer in both phases. It is noteworthy that the Ca II H and K lines generally act more like a gradual phase indicator than the hydrogen Balmer lines (see Figure 2.25a), and have been used as a proxy for the soft X-ray emission. The difficulty in modelling the hydrogen Balmer lines and Ca II lines simultaneously

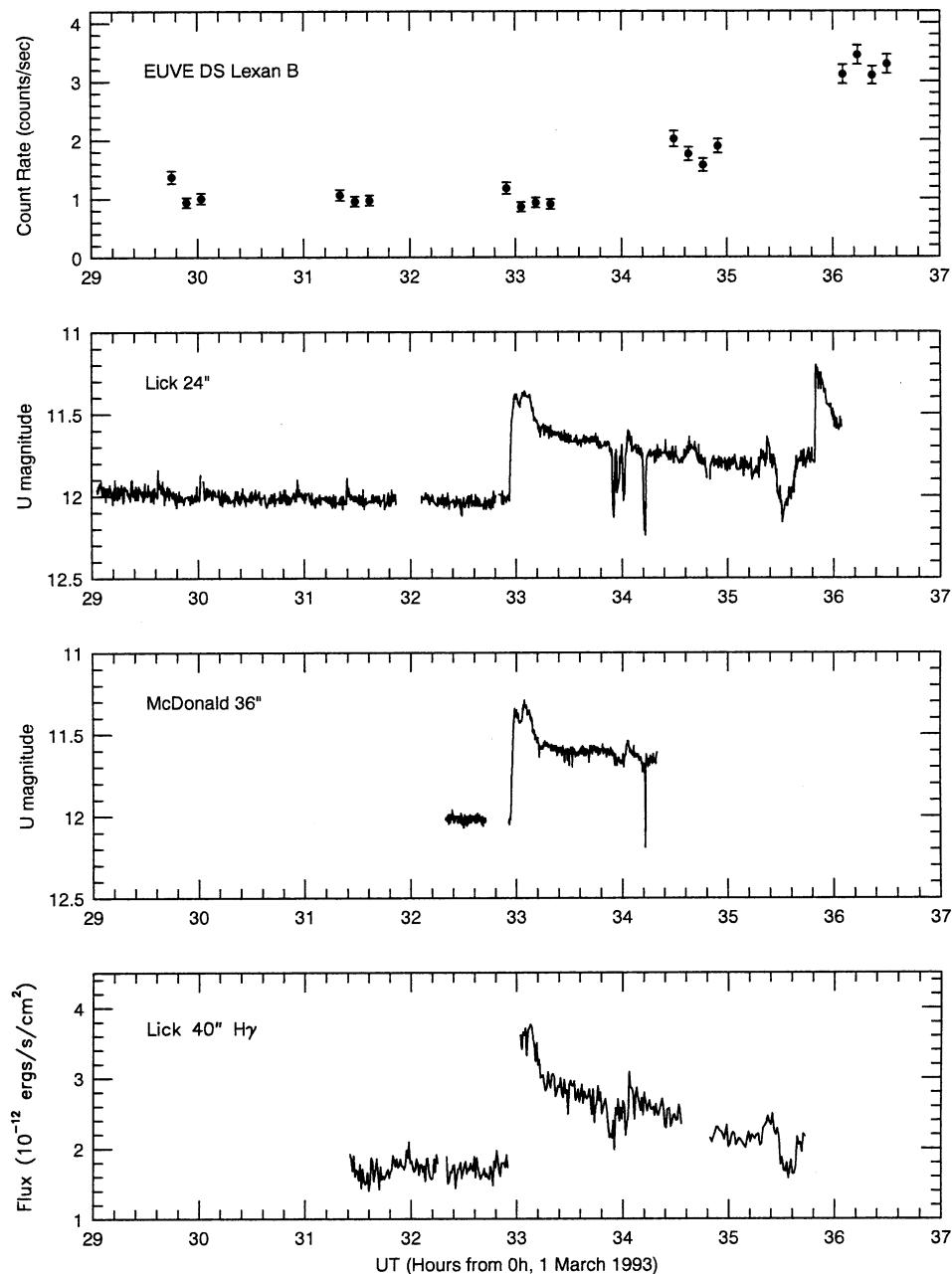


Figure 5.13. Photometric light curves of AD Leo in the U-band filter from two observatories, together with the soft X-ray and $H\gamma$ line flux data. The U-band continuum and the $H\gamma$ line flux have a similar response to the impulsive flare heating, while the X-ray emission is delayed. These observations support the chromospheric evaporation model described in the text. (Data from [H4].)

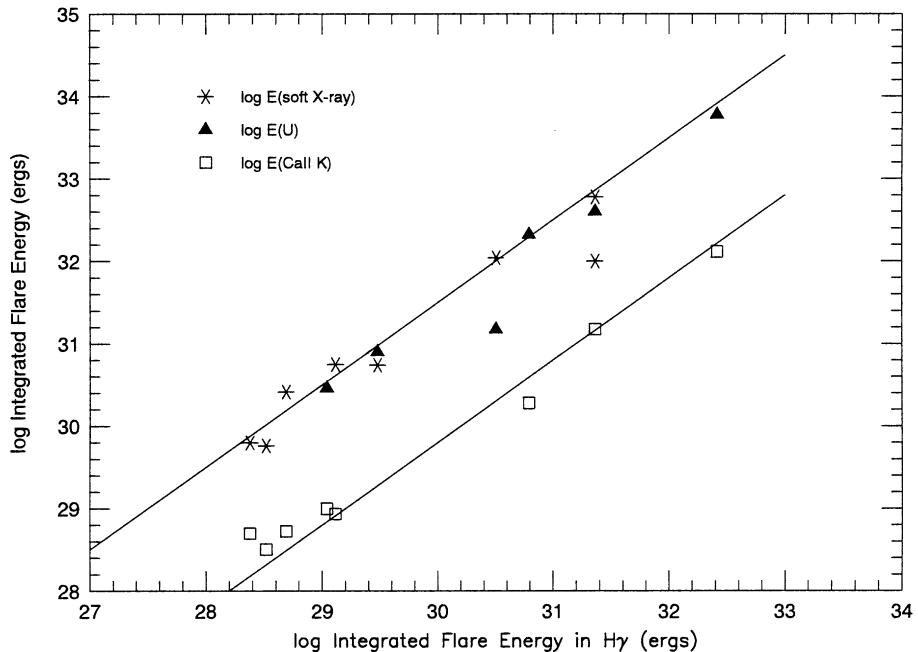


Figure 5.14. Correlations between the energy emitted in chromospheric emission lines (Ca II, $H\gamma$), the white-light continuum (U-band) and the thermal corona (soft X-ray emission) are characteristic of both solar and dMe flares. (Data from [H4].)

has already been described, and the same is true in flare models; their differing evolution during the flare may be a clue to the suspected lateral inhomogeneity in their formation on the stellar surface.

There are correlations between the total energy (summed over the flare) released in the emission lines, the continuum and the X-rays, such that $E_X \sim E_U$, $E_{H\gamma} \sim E_{CaIIK}$, and $E_{H\gamma} \sim 0.05E_U$ [B11], [H4]. Figure 5.14 illustrates these relations, which exist for flares on stars with a wide range of spectral types, and over a broad range of total flare energy ($E_U \sim 10^{26}\text{--}10^{34}$ ergs), suggesting (from an empirical standpoint) that the heating rates and mechanisms for producing the flares may be similar. Area coverage may then be the principal discriminator in determining the total energy produced by a given flare.

Note that the blue continuum, $H\alpha$ (and higher-order H Balmer lines), and Ca II K line are typically observed during M dwarf flares, since these are the most dramatically affected. $H\alpha$ often shows relatively little change during flares on early to mid-M dwarfs (although it can vary tremendously in late M dwarf flares, see Section 5.6.2). In contrast, $H\alpha$ is the principal diagnostic in solar flares, and the blue continuum radiation is rarely strong enough to be observed on the Sun, which has a much higher photospheric background. Comparison between solar and stellar flares has been hampered by the lack of similar observational material. This is not necessarily the fault of the observers; even if the underlying mechanism causing the

flares is the same, the M dwarf atmosphere responds differently to the influx of energy, making it necessary to observe the flare radiation where it chooses to appear, rather than in a more convenient, solar-type diagnostic. A step toward obtaining comparable observations has been made by Johns-Krull *et al.* [J3], who observed a solar flare using an innovative experimental set-up allowing solar observations through a stellar spectrograph. They found differences in the hydrogen Balmer line formation during solar flares compared to the stellar counterparts. More observations are needed to understand the extent to which stellar flares are analogues of solar flares, as is generally assumed.

For the Sun, an enormous body of work has led to a generally accepted model for the production of solar flares, as summarised in [L6], [D3]. The flare energy is released into the atmosphere following a catastrophic event between magnetic flux tubes (or coronal loops, as they are called when they have reached coronal height). Magnetic reconnection occurs, which produces a current sheet, and electrons are rapidly accelerated along the sheet (essentially down the magnetic field lines) towards the lower atmosphere. The electrons collide with ions in the ambient plasma and emit hard X-rays via non-thermal bremsstrahlung, from which a power-law distribution can be inferred for the electron energy spectrum. These hard X-rays have been directly observed on the Sun with satellites such as ASCA, SOHO, Yohkoh and RHESSI, but are not yet observable from stars, as the hard X-ray flux is many orders of magnitude smaller than the chromospheric and thermal (soft X-ray) emission observed later in the flare. White light (broadband optical) continuum emission is sometimes observed from small kernels at the footpoints of solar flare loops, and is temporally and spatially correlated with the hard X-ray emission [N1], implying that it is emitted at or near the sites where the accelerated electrons impact the pre-flare chromosphere and, in some cases, penetrate into the photosphere. The white-light emission is typically attributed to hydrogen recombination radiation, showing a strong Balmer jump [F1]. The spiralling motion of the energetic electrons around the magnetic field lines results in gyrosynchrotron emission at microwave frequencies, which is also observed directly in solar flares (and a few stellar flares [J1], [A1]). The non-thermal electrons lose most of their energy in the lower atmosphere, resulting in increased chromospheric emission and the heating of chromospheric plasma to coronal temperatures. The term *evaporation* is used to describe the process of heating the cool chromospheric plasma, and the subsequent flow of hot plasma up into the corona. The evaporation is accompanied by a downward moving shock front which forms a *chromospheric condensation* – a relatively cool and quite dense region that is propelled toward the photosphere, and which is the site of much of the chromospheric emission during the flare [F2]. As might be expected, the emission is strongly Doppler-shifted; redshifted H α emission is observed from the condensation, and blueshifted emission in transition region lines marks the evaporating plasma. Downflow (condensation) velocities of tens of km s^{-1} and upflow (evaporation) velocities of a few hundred km s^{-1} are typically observed [I1], [Z1].

As the corona increases in density (and hence column depth) due to chromospheric evaporation, it becomes effective at stopping energetic electrons, which now heat the corona directly while the amount of heat deposited in the lower atmosphere

diminishes. During this time the corona is dense and hot, and is emitting copious thermal soft X-rays, which in turn provide a secondary source of heating for the chromosphere. When the coronal heating ceases, the material in the loop gradually condenses back to its pre-flare configuration, and the soft X-ray flare emission dies away.

Empirical support for this model is found through observations of the ‘Neupert effect’ [N3]. Chromospheric evaporation, which is responsible for increasing the coronal emission measure and hence the soft X-ray flux, is proportional to the flare heating rate, measured by the hard X-ray flux. The instantaneous emission measure represents a time integral of the evaporation (heating) process, implying that the soft X-ray flux should be a time integral of the hard X-ray flux. Dennis and Zarro [D4] found that more than 80% of large solar flares exhibit the Neupert effect. For stellar flares, white-light continuum emission [H7] or radio gyrosynchrotron emission [G10] have been used as a proxy for the hard X-rays to estimate the flare heating rate. Both studies found evidence for the Neupert effect in M dwarfs, indicating that the general features of the solar model may be applicable on low-mass stars. Figure 5.15 illustrates the first observation of the Neupert effect during an M dwarf flare by Hawley *et al.* [H7].

The exact details of the flare emission are clearly different for M dwarfs; the white-light continuum is much stronger, and does not often show evidence for hydrogen recombination radiation as the production mechanism. Figure 2.25(b) is a spectrum from the impulsive phase of a large flare on AD Leo, with no sign of a Balmer jump [H4]. Flare models have also been largely unable to produce the white-light continuum with any realistic heating mechanisms [H5], although Houdebine [H12] showed that semi-empirical flare models with very large coronal and transition region pressure could reproduce some of the continuum properties. Models incorporating the dynamics in the chromosphere and the formation of the chromospheric condensation, together with realistic treatment of the heating and radiation, are being developed for the Sun [A2], and for M dwarfs [A10]. It is ironic that the production of the white light during M dwarf flares remains a mystery, although it is the most obvious and spectacular manifestation of these events.

The line profiles observed during flares also remain unexplained. As shown in Figure 2.25(b), extensive broadening – up to 30 Å at the base of the H Balmer lines – is a common feature during the impulsive phase of flares, but the mechanism has been variously attributed to the Stark effect, turbulence, and flows. Models incorporating the Stark effect and turbulence have not yet been successful at reproducing the observed line broadening [H5], [H12]. Since the lines are often symmetric, flows are not a promising candidate mechanism, as they would require equal amounts of material emitting at the same temperature to be flowing in both directions. When asymmetries are observed [G11], they offer a probe of the mass motions that are occurring. For example, HST/STIS observations of transition region emission lines in the ultraviolet show significant downflows during, and shortly after, the flare impulsive phase in several flares on AD Leo [H19].

The increasing availability of multi-wavelength data has shown that some stellar flares behave quite differently in various wavelength regions. A large campaign on

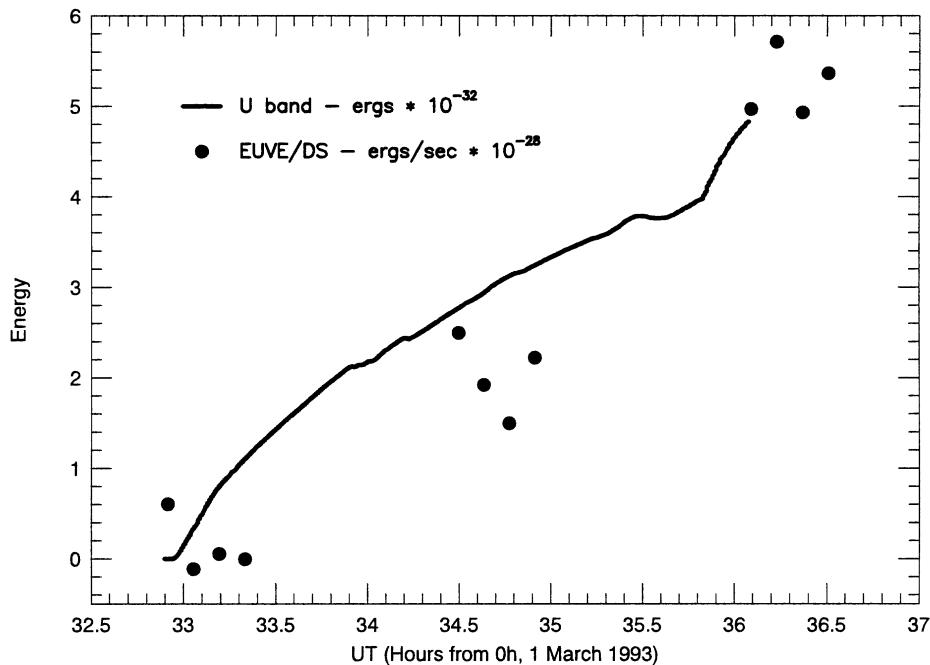


Figure 5.15. The first example of the Neupert effect measured in a stellar flare. The time integral of the U-band emission (solid line) is proportional to the instantaneous thermal soft X-ray emission observed with EUVE (filled circles) for the 1993 flare on AD Leo. The U-band (optical white light) emission is used as a proxy for the hard X-ray flux which measures the flare heating rate, while the thermal emission measures the current amount of material that has been evaporated into the corona by the flare heating. (Data from [H7].)

EV Lac employing simultaneous radio, optical, ultraviolet and X-ray observations showed several flares with strong emission in one or more regions and no emission in other regions [O2]. There have also been observations of strong radio flares in VLM dwarfs (including one L3.5 brown dwarf [B17]), at levels far beyond those expected. These observations indicate a departure from the solar paradigm, and represent a considerable challenge for stellar flare models.

A final, elusive feature of stellar flares is the presence of diminutions in the continuum radiation just before an outburst [C7], [G1], [D6], [H7]. An example is shown in Figure 5.13 at time 35.5 hours, just before the second impulsive peak, and was seen in both photometric and spectroscopic observations. Speculation about the production of these diminutions centres on enhanced opacity mechanisms [G9], or dark absorbing material from a prominence passing in front of the emitting regions. A theoretical prediction of dimming at the beginning of solar flares has been made by [A2], but the mechanism relies on the emission being due to hydrogen recombination radiation, which may not be applicable for M dwarf flares.

While the radiated flare energy has been relatively well-observed and catalogued for many flares, little work has been carried out on determining the

mechanical energy dissipated through flows or coronal mass ejections in M dwarfs. The kinetic energy and mass momentum have been computed for a flare on AD Leo [H14], and Cully *et al.* [C8] proposed a model incorporating solar-like coronal mass ejections to explain the EUVE observations of a flare observed on AU Mic. The mass loss that M dwarfs will undergo through flaring has been estimated at a few times $10^{-13} M_{\odot} \text{ yr}^{-1}$ [C3], which, when considering the large number of M dwarfs in the Galaxy, could contribute significantly to the mass and energy balance in the interstellar medium. Mullan *et al.* [M10] found that a surprisingly large value of the mass-loss rate, a few times $10^{-10} M_{\odot} \text{ yr}^{-1}$, was permitted by their radio observations, which were, however, at the detection limit of their instrumentation. This value includes both a steady wind and flare mass ejections. In response, Lim and White [L5] pointed out that such a strong wind would completely absorb the radio emission observed during some stellar flares, and found that mass-loss rates must be less than $\sim 10^{-13} M_{\odot} \text{ yr}^{-1}$. van den Oord and Doyle [V2] subsequently provided a theoretical basis for calculating properties of stellar winds in cool stars, and found that the maximum mass-loss rates were on the order of $10^{-12} M_{\odot} \text{ yr}^{-1}$. Radio and submillimetre observations of dMe stars will in future be important for obtaining direct measurements of the mass-loss rate. If mass loss from M dwarfs does play an important role in the mass balance of the ISM, there may be implications for the measured interstellar deuterium abundance as well as other light elements produced via spallation reactions in the corona during flare events.

Flare data for low-mass members of open clusters are sparse, consisting mostly of reports of detections, but with little in the way of reliable statistics. IAU Symposium No. 137 on *Flare Stars in Stellar Clusters, Associations and the Solar Vicinity*, was held in Armenia in 1989; the proceedings from that meeting together with the review by Pettersen [P8] comprise most of the older work. Newer work is mostly contained in IBVS bulletins (see also [H2], [J4]). Studies to determine the number of flare stars in clusters, and the change in flaring rate with age, would be valuable.

5.5 GLOBAL PROPERTIES OF CHROMOSPHERIC ACTIVITY IN M DWARFS

5.5.1 Activity effects on colours, magnitudes and bandstrengths

Magnetic activity, and the accompanying outer atmospheric heating, produce subtle effects on the colours, magnitudes and bandstrengths of active stars. Evidence of these effects has been known for many years; dMe stars are brighter (in absolute magnitude) than dM stars of the same colour [V3]; the V–K colours of active stars are redder than inactive stars of the same spectral type [M11]; and excess ultraviolet and blue emission is often observed in active stars [H15], [A11]. However, it was only with the availability of large, statistically complete samples – notably the PMSU spectroscopic survey of the nearby stars [H8] – that the extent of these effects became apparent.

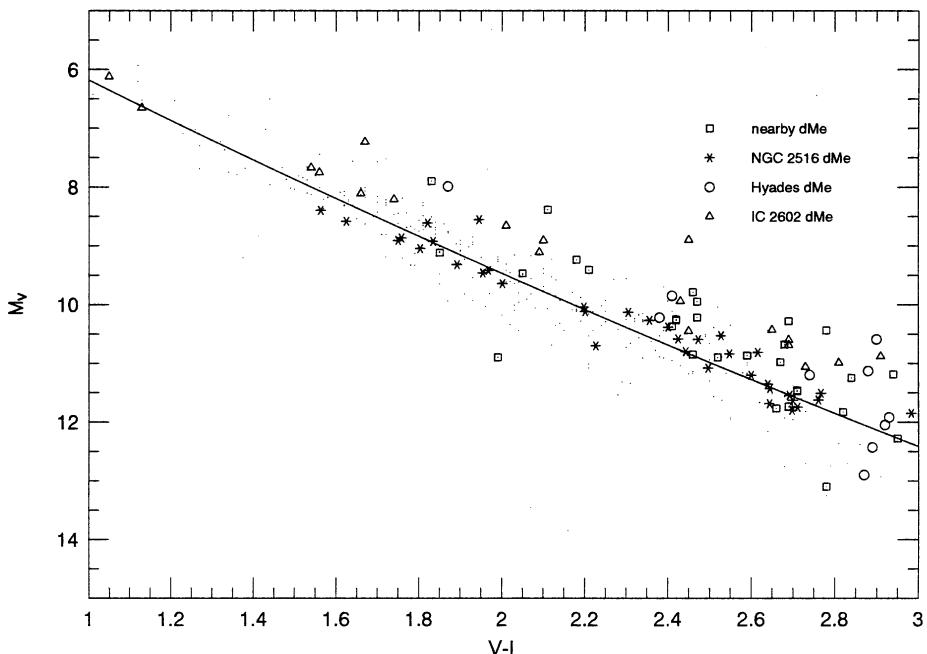


Figure 5.16. Active stars in the field and in clusters lie systematically above and to the red of the main sequence defined by the inactive stars. Inactive field M dwarfs are shown as the small dots, with the solid line being a fit.

First, field dMe stars lie systematically above and to the red of the main sequence defined by the inactive (dM) stars, particularly at early spectral type (M0–M3). The same effect is present among active cluster stars. Figure 5.16 shows data for the 8-parsec sample, other field stars with good colours and *Hipparcos* parallaxes, and the nearby clusters NGC 2516, the Hyades and IC 2602. The active stars are clearly offset from the dM main sequence. The data are not sufficient to determine whether the active stars are redder at the same luminosity, or brighter at the same colour, than the inactive stars. The recent study by West *et al.* of an SDSS spectroscopic sample of nearly 8,000 M dwarfs showed that the dMe stars were bluer than the inactive dM stars in $(u-g)$ by 0.05 ± 0.25 magnitudes and were redder in $(g-r)$ by 0.11 ± 0.16 magnitudes [W3].

Second, individual molecular bandstrengths are affected: the various sub-bands of the $\gamma(0,0)$ TiO bandhead near 7,100 Å change their behaviour, depending on the activity level of the star. Figure 5.17 shows the sub-bands plotted against one another for the active stars, with the solid lines again representing the mean relations for the inactive stars (see [H8] for a description of the TiO indices). The dMe stars lie below the mean dM relation in TiO₂, while the opposite is true in TiO₄ – in other words, the $\lambda 7,050$ Å sub-band is stronger and the $\lambda 7,120$ Å sub-band is weaker in active stars than in inactive stars. The overall bandstrength, TiO₅, is only slightly affected by the activity and therefore still serves as an indicator of spectral

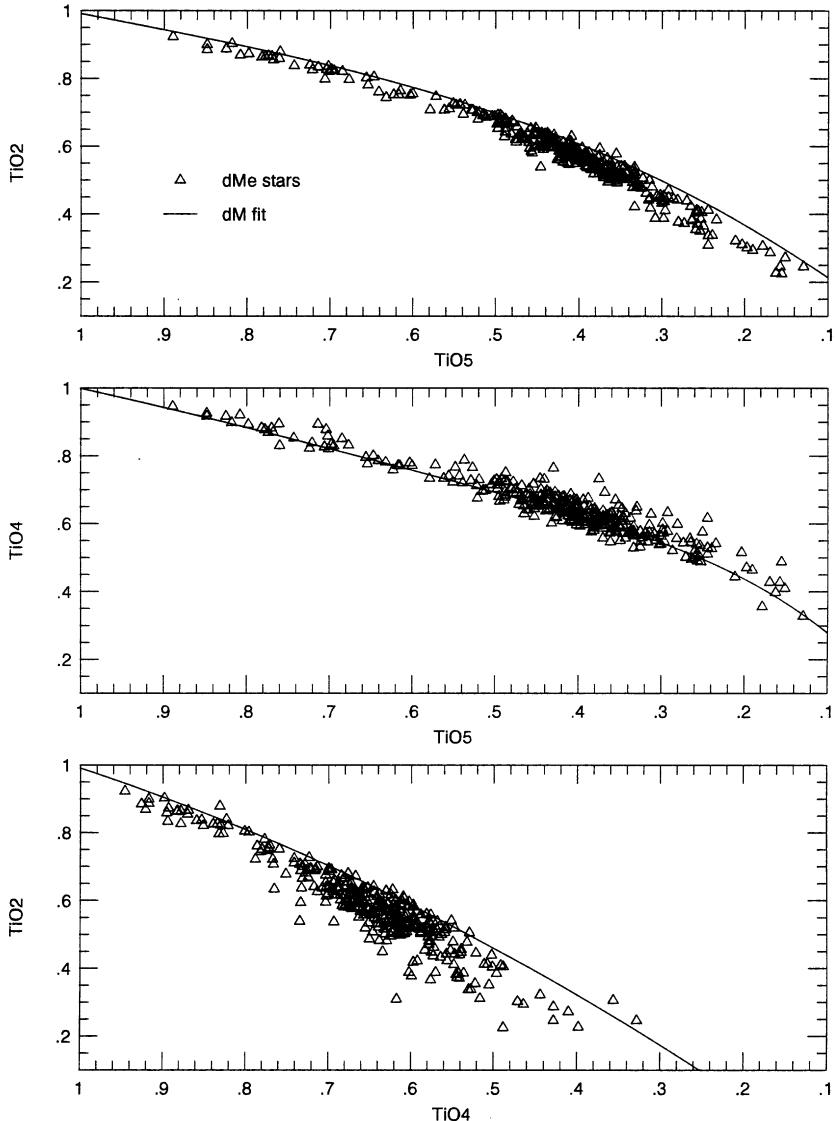


Figure 5.17. Activity effects on the TiO bandstrengths. (Data from [H8].)

type. Plotting TiO₂ against TiO₄ shows the effect most strongly, and the extent of the offset is positively correlated with the activity strength measured by $L_{H\alpha}/L_{bol}$ [H8].

The bandhead near 6,230 Å, comprised primarily of TiO in M dwarfs earlier than M4, is also affected, appearing shallower in active stars at the same $V-I$ colour as inactive ones [H9]. When the λ6,230 Å bandstrength is substituted for $V-I$ in Figure 5.16, the luminosity difference between the dM and dMe stars disappears!

Evidently the colour is most sensitive to the presence of activity. Detailed models of M dwarf atmospheres (including the chromosphere) are required to provide insight into the physical processes that lead to these activity-related effects.

5.5.2 Empirical relationships between age and activity

Observations of solar-type dwarfs in open clusters show that these stars exhibit a well-defined age–activity relationship. The emission in the Ca II K line is used to represent the activity, and this emission decays with age, such that a young solar-type star has a large Ca II K line flux, while an older star (like the Sun) has a rather small line flux. Skumanich [S8] was the first to quantify the relation as $F(Ca) = At^{-1/2}$, the now famous ‘*t* to the half’ law; Soderblom *et al.* [S9] have since shown that the relation is probably more complicated than a simple power-law fit of emission flux to age. As described further in Section 5.5.4, the underlying physical parameter is probably rotation, which also decays with age, and which the solar dynamo depends upon strongly.

Observations of M dwarfs do not obviously follow a solar-type relation (see [E1] for an opposing viewpoint). Activity in these stars appears to be a function of both age and mass, with more prolonged lifetimes in lower-mass stars.

The first clue that M dwarfs might behave differently came with the initial observation of Joy and Abt [J6] that dMe stars were much more prevalent among later type M dwarfs (e.g., type M5) than among early types. The PMSU survey results quantified the fraction of active dwarfs as a function of spectral type, showing a monotonic rise from type M0 to M6 [H8]; later follow-up of 2MASS M dwarfs indicated that the fraction decreased again at types later than M7 [G6]. The comprehensive West *et al.* SDSS survey [W3] provides the best statistics, with hundreds of stars in each spectral type bin through type M9 as shown in Figure 5.18. The rise in active fraction is ascribed to an age effect, whereby higher mass early M dwarfs lose their activity quickly, while lower mass, later-type M dwarfs retain it for much longer, perhaps nearly as long as the lifetime of the Galactic Disk. The decline in active fraction at types later than M8 may be due to atmospheric effects in the cool atmospheres of VLM dwarfs, as discussed in more detail in Section 5.6. Initial controversy over these results centred on selection effects; due to the spectroscopic resolution of most of the observations, H α emission can only be confidently detected with an equivalent width of 1 Å or greater, while a line of 1 Å EW is much more easily detected in a late-type M dwarf, where the continuum flux is greatly reduced near 6,500 Å. Selection effects are shown not to be the cause of the active fraction increase by examining the activity strength, the total luminosity emitted in the H α line compared to the bolometric luminosity. Figure 5.19 shows the most recent results for the SDSS survey – the mean $L_{H\alpha}/L_{bol}$ ratio at each spectral type is well above the detection limit (dashed line). Interestingly, the activity strength begins to decrease at type M6, while the active fraction is still rising until type M8. This difference probably provides a clue to the efficacy of magnetic field generation and/or atmospheric heating in the lower mass dwarfs, and thus to the reason for the decline in the active fraction at types later than M8.

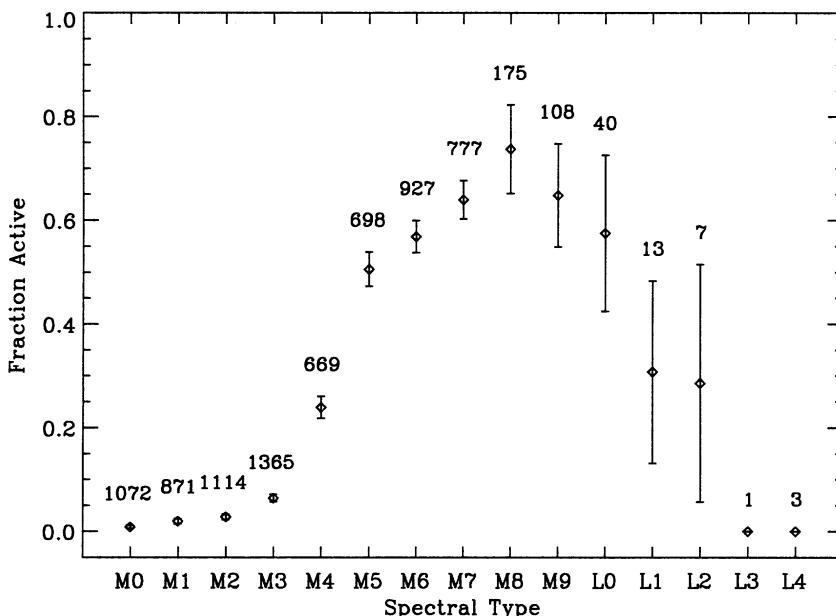


Figure 5.18. The fraction of M dwarfs which are active vs. spectral type illustrates the steep rise through type M8 attributed to the increasing longevity of activity at later types. The decline for dwarfs of spectral type later than M8 may be due to some combination of decreasing dynamo efficiency in the interior and atmospheric effects that prevent the emergence of persistent magnetic fields at the surface. (Courtesy of A. West [W3] and the *Astronomical Journal*.)

The hypothesis that activity in M dwarfs lasts longer at lower masses is bolstered by independent evidence from several other age-related indicators, including velocity dispersion, distance from the Galactic Plane, open clusters and binary systems.

Velocity dispersion

Wielen [W1] originally demonstrated that dMe stars as a group had smaller velocity dispersions and a smaller asymmetric drift than the dM stars, both indicative of a younger population. More detailed analysis of the PMSU survey confirms that the dMe dwarfs have smaller velocity dispersions than the complete dM sample [H8]. The principal axes also are not diagonal in Galactic co-ordinates (i.e., there is a significant vertex deviation) – an indication of a population that has not had time to relax. When the dM stars are subdivided into early- and late-type subsamples (at $TiO_5 = 0.5$ or $M_V \sim 11$, spectral type $\sim M3$), the later-type stars have higher velocity dispersions than the earlier-type stars, which is also attributed to an age effect: if M dwarfs of late spectral type retain their activity properties for longer than those of early spectral type, then the removal of the dMe stars from the early sample will leave some relatively young stars, while the removal of dMe stars from the late

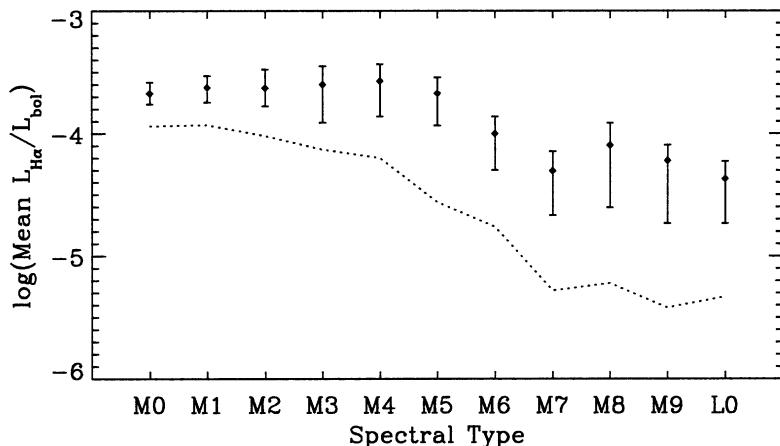


Figure 5.19. The activity strength measured by $L_{H\alpha}/L_{bol}$ shows a nearly constant value through the early-mid M dwarfs, and declines at type M6 and later. This is not a selection effect, as shown by the small scatter in this very large sample of SDSS dwarfs. The dashed line indicates the detection limit of 1 Å EW in the $H\alpha$ emission line. (Courtesy of A. West [W3] and the *Astronomical Journal*.)

sample will leave only quite old stars. These kinematic results support the idea that the later-type M dwarfs retain their activity for a longer time than the early-type M dwarfs.

Distance from the Galactic Plane

The vertical distance from the Galactic Plane is a measure of dynamical heating for Galactic Disk stars (see Chapter 7). Assuming that disk stars are formed at, or very near, the Plane, they will be heated during their lifetime by collisions with (primarily) molecular clouds, leading to an average scale height for the thin disk population of 300 pc. The dynamical heating increases with time, and thus stars found further from the Plane are expected to be an older population. Figure 5.20 shows the active fraction of M7–8 dwarfs from both the nearby 2MASS sample and the extensive SDSS sample, as a function of vertical distance from the Plane. Taking the vertical distance as a proxy for age, the active fraction clearly declines with age; with many hundreds of stars in each distance bin, this is a statistically robust result. Models of dynamical heating may eventually allow a quantitative interpretation of the average ages of stars at a given distance from the Plane.

Open Clusters

Observations of open clusters provide additional information on age effects, since the age of the cluster is independently determined; for example, from main sequence fitting (Chapter 7). M dwarfs in the Pleiades and Hyades have received the most extensive scrutiny. Stauffer and collaborators [S11] showed that cluster members exhibit strong $H\alpha$ emission at late spectral types in both clusters, and that the

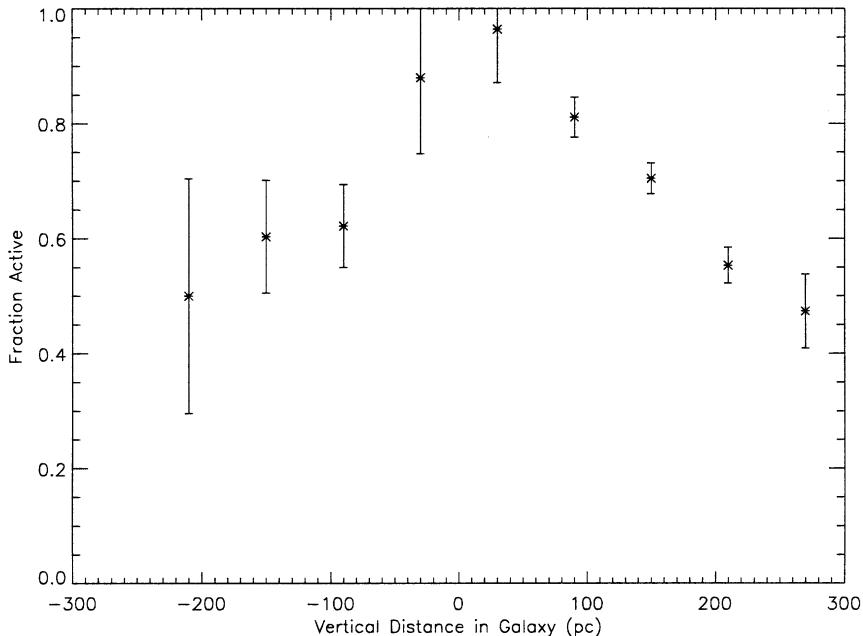


Figure 5.20. The active fraction of M7–8 dwarfs vs. vertical distance from the Galactic Plane, with positive distance being north of the Plane and negative distance south. The northern bins are better sampled, but the distribution is consistent with symmetric dynamical heating about the Plane. Taking distance as a proxy for age indicates that the active fraction decreases with age (Courtesy of A. West).

difference between them is the earliest spectral type where emission is detected. In the Pleiades the emission is prevalent among the late K stars, while in the Hyades it does not appear until the early M stars. However, the mean activity strength, measured by $L_{H\alpha}/L_{bol}$, lies at the same level in both clusters. Several clusters – including the important older cluster M67 [H10] – have now been surveyed, and all show the same effect: the spectral type (mass, colour, absolute magnitude) where the activity ‘turns on’ is what changes with age, rather than the strength of the emission at a given spectral type. This result is in accord with those above; it explains the prevalence of dMe stars at late spectral types in the field, as most of them are not old enough to have lost their activity.

The ‘ $H\alpha$ limit mass’ where the emission turns on is illustrated in Figure 5.21 for the \sim 150-Myr-old cluster NGC 2516. Using similar data for several other clusters, Hawley *et al.* [H10] constructed the age–activity relationship shown in Figure 5.22. Approximately linear relationships in log age provide an adequate fit to the available data. The age–activity relationship for M dwarfs does not indicate decay of chromospheric emission with age, as in the solar-type stars, but instead shows that activity depends on the mass of the star at different ages. This empirical relationship between age and activity provides an important (although as yet unexplained) clue concerning

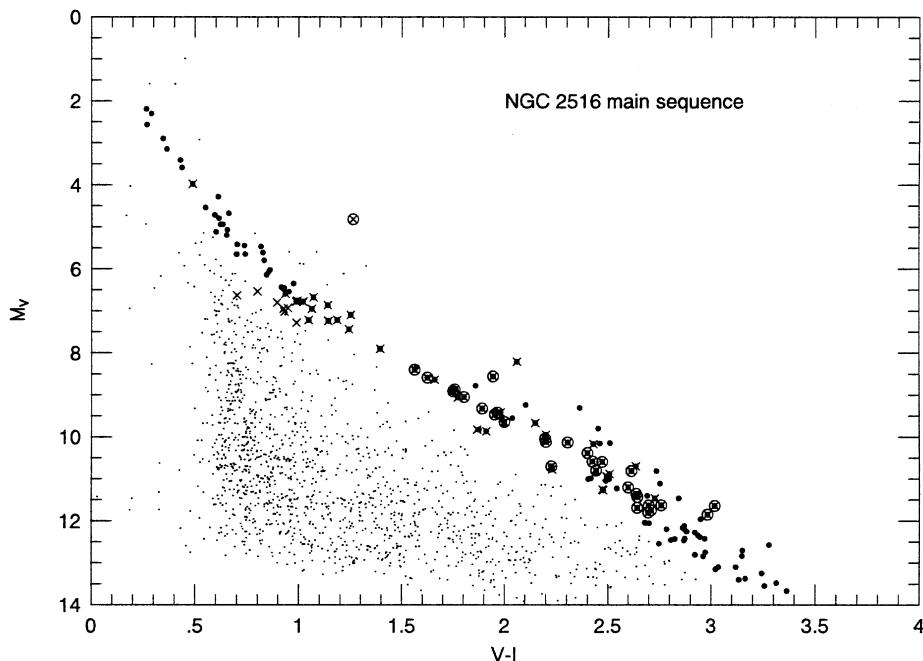


Figure 5.21. Colour–magnitude diagram for the open cluster NGC 2516, illustrating the concept of the $\text{H}\alpha$ -limit mass. Candidate cluster members are the enlarged dots, stars with spectroscopic observations are marked with crosses, and stars with $\text{H}\alpha$ emission are circled. The activity turns on at $(V-I) \sim 1.5$ in this cluster. The object at $V-I \sim 1$, $M_V \sim 5$ is a foreground dMe star, not a cluster member. (Data from [H9].)

the physical processes governing the production of the magnetic fields underlying the chromospheric heating, and perhaps the heating mechanisms themselves.

Binary systems

Binary systems consisting of an M dwarf and a white dwarf offer yet another possibility for investigating age dependence, as ages of white dwarfs can be determined through models of their cooling times (see Chapter 7). Silvestri [S15] has pioneered work in this area. The results show that activity does last longer in later type M dwarfs in binaries, but the relationship between age and activity is more complicated than the simple linear fits shown in Figure 5.22. Observations of additional binary systems with old white dwarfs, and with later type M dwarfs, are needed to fully exploit this method.

5.5.3 Emission in other wavelength regions

It is instructive to compare the diagnostics of magnetic activity across the electromagnetic spectrum; these diagnostics correspond to formation in different parts of

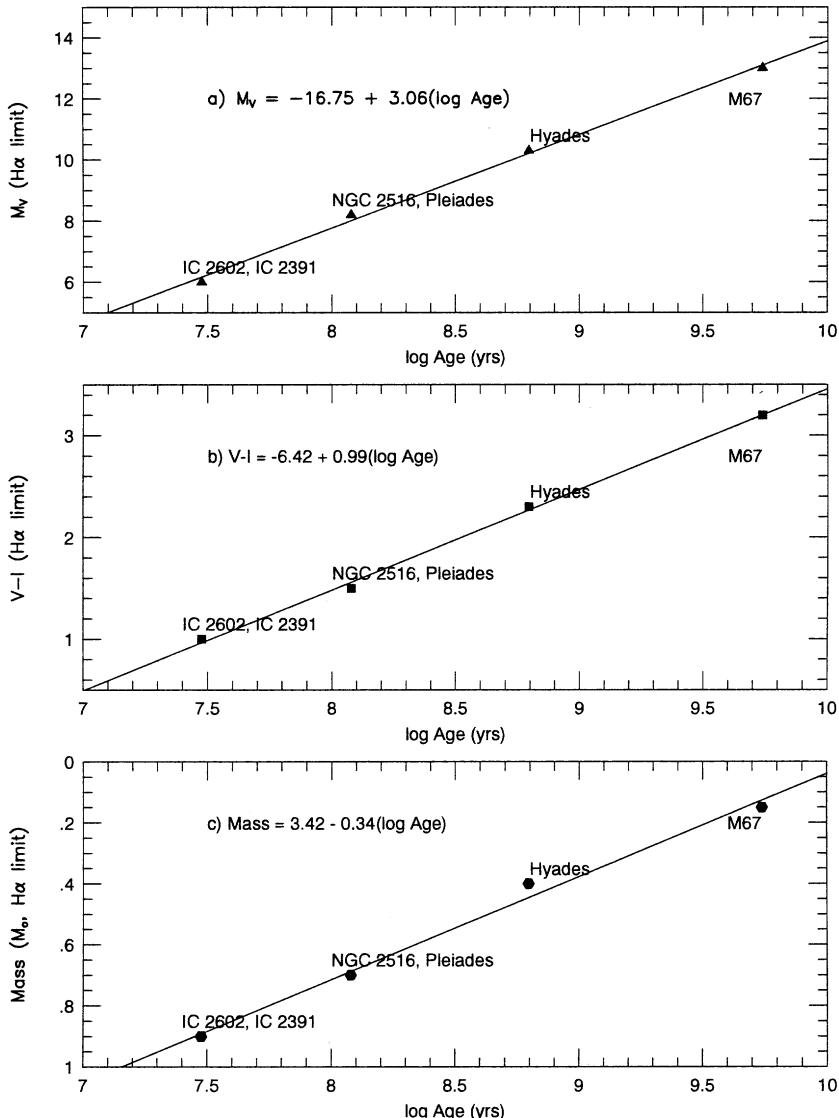


Figure 5.22. Age–activity relationships for M dwarfs using $H\alpha$ limit measurements in clusters. (a) M_V against log age; (b) $V-I$ against log age; (c) mass (M_\odot) against log age. (Data from [H10].)

the atmosphere, and possibly by different heating mechanisms, although the magnetic field is assumed to be the ultimate source of energy driving the emission processes in all cases.

Figure 5.23 compares the activity in the optical ($H\alpha$), soft X-ray (0.1–2.4 keV for ROSAT observations; 0.2–3 keV for Chandra observations) and radio bands

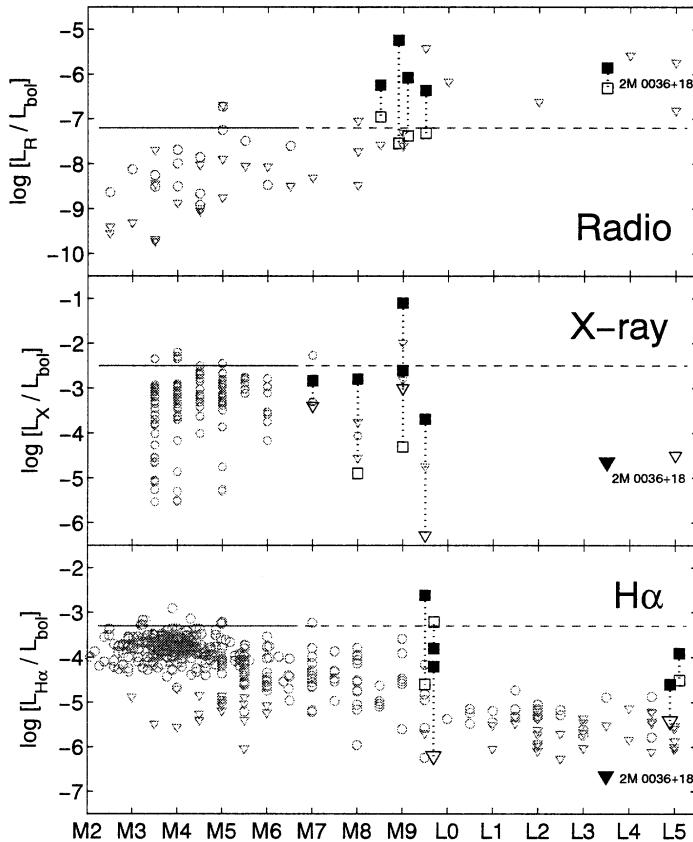


Figure 5.23. The ratios of the H α , X-ray and radio luminosities to the bolometric luminosity are compared for early M through mid L dwarfs. As shown in Figure 5.19, the $L_{H\alpha}/L_{bol}$ ratio declines past type M6, while the sparse X-ray data at late M types also shows a probable decline. The radio behaviour is in the opposite sense, with L_R/L_{bol} increasing toward later types. (Courtesy of E. Berger [B18] and the *Astrophysical Journal*.)

(average of 3.6-cm and 6-cm VLA observations, corresponding to 8.5 GHz and 4.9 GHz, respectively) [B18]. As described in Section 5.3, the optical emission is produced in the chromosphere while the soft X-rays are the result of thermal emission from the hot corona. The radio emission is probably the result of gyrosynchrotron radiation from a non-thermal distribution of electrons – this is essentially a flare phenomenon, although it appears to be persistent (i.e., present at all times) in many of the observed dwarfs. The ratios of the luminosity produced in H α , soft X-rays and radio are nearly constant for early to mid M dwarfs, with $L_X \sim 3 L_{H\alpha}$ [H20] and $L_X \sim 3 \times 10^5 L_R$ [G13]. Recent HST/STIS observations of late M dwarfs indicate that the transition region emission as measured by the ultraviolet C IV line flux also scales with $L_{H\alpha}$ throughout the M dwarf sequence, with $L_{CIV} \sim 0.1 L_{H\alpha}$ [H20]. The optical, ultraviolet and X-ray emission appear to decrease together

in the late M and L dwarfs, though more UV and X-ray observations are needed to quantify whether the emission ratios remain strictly constant. However, the radio emission exhibits a significant departure from the others, with L_R increasing by several orders of magnitude relative to L_X and $L_{H\alpha}$ for these VLM dwarfs, as shown in Figure 5.23. A possible explanation is that the heating mechanism that produces the high-temperature outer atmosphere becomes less efficient in the VLM dwarfs, while a separate heating mechanism that drives the non-thermal electrons producing the radio emission becomes more efficient. However, the theoretical underpinnings of these heating mechanisms remain elusive.

5.5.4 The connection between rotation and activity

The shell dynamo mechanism which produces the magnetic field in the Sun and solar-type stars depends critically on the stellar rotation. Empirically, studies of solar-type stars show good correlation between rotation velocity and activity indicators such as Ca II K emission, and an even stronger correlation between activity and the Rossby number, the ratio of the rotation period to the convective turnover time [N4]. The dynamo number N_D , which characterises the strength of an $\alpha\Omega$ dynamo, is proportional to the inverse square of the Rossby number, thus providing a physical explanation for the rotation–activity (and hence age–activity) correlation in solar-type stars.

Rotation velocities may be measured by modelling the effect of rotation on the spectral line profiles. All other line-broadening mechanisms – such as pressure broadening, turbulence and magnetic field effects – must also be modelled, and allowance must be made for the limb-darkening [G7]. The result of the analysis is the line-of-sight component of the rotation, $v \sin i$, where i is the inclination of the rotational polar axis. An alternative method is to measure the photometric rotation period; for example, by observing rotational modulation due to starspots (see Section 5.4.1). The period, together with an estimate of the stellar radius, allows a determination of the rotation velocity at the spot latitude (equatorial spots are generally assumed). If both spectroscopic and photometric rotation estimates are available, the inclination can be determined.

Early efforts to measure the spectroscopic rotation of low-mass stars were hampered by instrumental limitations. The stars were too faint to observe at high resolution in order to measure $v \sin i$ from the line-broadening; in addition, there was a growing realisation that rotation velocities in M dwarfs were nearly always very low compared to earlier-type stars. A notable exception is the dM0e star Gliese 890 [Y1], [P5], with $v \sin i = 70 \text{ km s}^{-1}$ and a measured rotation period of just 0.43 days, but it is the exception rather than the rule; the surveys by [S10] and [M2] found few M dwarfs above their detection limits of 10 km s^{-1} and 3 km s^{-1} respectively.

The largest modern survey was carried out by Delfosse and collaborators [D2], who detected rotation above 2 km s^{-1} in 24 out of 99 nearby stars. They found that the later-type M dwarfs (the study included stars as late as spectral type M6.5) were more likely to have measurable rotation, and that the early-type stars with measurable rotation were nearly all members of a young kinematic population,

as determined by their space velocities. A picture thus emerges in which rotation lasts longer in lower-mass stars, such that in a fairly old field sample only a few early-type stars have rotation (the young ones), while most of the later-type stars have rotation (because it lasts longer in those stars). These are characteristics very much like those described for the age–activity relationship above, and would seem, at face value, to indicate that the longer timescale for the rotation velocity to decay in later-type stars is responsible for the continued activity to late types in those stars. However, Figure 5.24 plots activity strength ($\log L_{H\alpha}/L_{bol}$) against rotation velocity, showing that while it is true that the later-type stars with measureable rotation velocity do have activity, there is no strong correlation between the magnitude of the velocity and the activity strength; stars can be equally active with velocities ranging from 0.2 km s^{-1} to $>20 \text{ km s}^{-1}$.⁵ The data indicate a rather low threshold for rotation velocity necessary to maintain magnetic activity in M dwarfs ($<2 \text{ km s}^{-1}$). The threshold effect has also been termed ‘saturation’, but note that the mean activity level shows large scatter, and by no means all stars achieve the highest observed activity strengths. The latest M and L dwarfs show very little activity, except for the interesting object PC0025+0447 (at $\log L_{H\alpha}/L_{bol} = -3.4$), and the two active late M dwarfs LHS 2397a and LHS 2924. The late-type objects at high velocity are nearly all brown dwarfs, as evidenced by the presence of lithium in their spectra. These are discussed in more detail in Section 6.7.5.

Bopp and Fekel [B8] originally noted that active close binaries (BY Dra variables) with synchronous rotation attained a rotational velocity above a threshold value of $\sim 5 \text{ km s}^{-1}$, and postulated that active single stars rotated at least this fast. Several recent investigations also show that binarity can enhance activity in late-type subdwarfs [G5], [P3]. Rotation therefore plays some role in the production of magnetic fields and magnetic activity – at least in early M dwarfs.

The lowest mass dwarfs, with spectral types later than about M7, begin to deviate significantly from even the ‘threshold’ rotation–activity relationship in the earlier M dwarfs. The first VLM dwarf to be found with very rapid rotation was LP 655-4 (BRI 0021-0214). [B5], using the HIRES echelle spectrograph on the Keck telescope, found $v \sin i = 40 \text{ km s}^{-1}$, but no measureable H α emission (H α was later measured at a low level by [T2] and at a substantial level by [R9], probably during a flare). Subsequently, several more VLM objects with similarly high rotation and little or no activity have been found. These objects are discussed further in Section 6.7.5.

5.6 ACTIVITY PROPERTIES OF THE LOWEST-MASS STARS AND BROWN DWARFS

The extreme faintness of the very low-mass (VLM) stars and brown dwarfs, coupled with the small size of the available samples, has hampered progress in the understanding of their magnetic activity. Even the choice of magnetic diagnostic

⁵ It should be noted that these are $v \sin i$ measurements for the most part; we assume that inclination effects are statistically insignificant when using a large, unbiased sample.

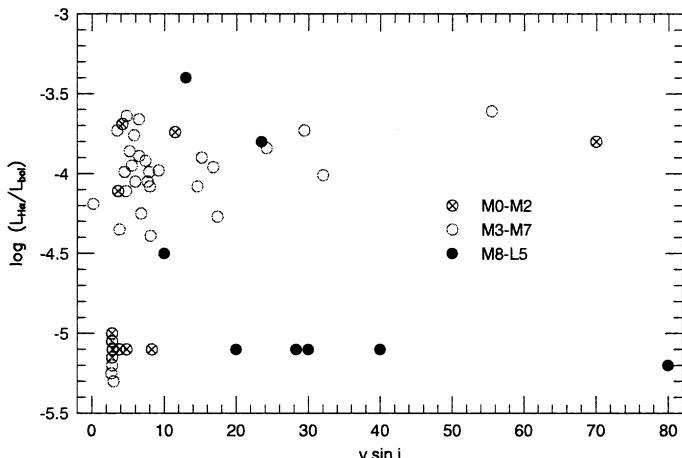


Figure 5.24. The activity strength is nearly uncorrelated with rotation, showing evidence only for a possible threshold value below 2 km s^{-1} . Very low or non-detections are shown at $\log L_{H\alpha}/L_{bol} \leq -5$; the early-type stars are arbitrarily offset to show them individually at velocities near 2 km s^{-1} .

may change – as between the G–K dwarfs and the early M dwarfs, where the chromospheric properties switch from favouring Ca II K to favouring H α . A few very young (~ 1 Myr) objects have been observed in X-rays, but most of the well-known low-mass dwarfs were not found in a systematic search of the ROSAT database [N2]. The expected X-ray emission is near or below the detection limits in most cases. Figures 2.24 and 5.23 show that even very few late M dwarfs have been detected (only the closest and most active ones). The surprising behaviour of the radio emission was described in the previous section. Here we discuss the H α emission, as well as evidence for spots and flares.

5.6.1 H α emission in VLM dwarfs

Figure 6.20 shows the decline in activity strength in the late M, L and T dwarfs. Stauffer *et al.* [S12] were the first to question whether the H α emission properties changed in the lowest mass dwarfs. Their data for VLM Pleiads showed a decrease in activity level at the latest types measured. A second indication of anomalous behaviour was the discovery of several VLM field dwarfs with no (or very low) activity, but rapid rotation [B5], [M3], [T3]. The [T3] study found that the VLM population as a group had similar kinematics to the M0–M6 nearby dwarfs in the PMSU survey, and thus there was no kinematic signature of a younger (or older) age. Instead, they observed a smooth decrease in activity with decreasing mass in objects across the stellar and substellar regimes, and postulated that the onset of dust formation (see Chapter 4) could be affecting the production of magnetic activity in the lower mass dwarfs. A notable exception to the trend of decreasing activity in the latest type M dwarfs is the remarkable dM9.5e dwarf PC0025+0447 [M9]. This

object has a Balmer emission spectrum that is very similar in activity strength to earlier-type M dwarfs, and hence exhibits enormous equivalent widths, since the stellar continuum is so depressed at this late spectral type. Martín and collaborators have suggested that PC0025 is a brown dwarf based on their detection of lithium. However, Liebert has proposed an alternative scenario, where the M dwarf is the brighter member of a VLM-dwarf/brown dwarf contact binary, with material accreted in Roche lobe overflow supplying both the high level of H α emission and intermittent lithium absorption.

Theoretical interpretation of the decline in magnetic activity at late M spectral types has centred on the increasing neutrality of the atmosphere as the effective temperature decreases, as proposed by Fleming *et al.* [F6]. Mohanty *et al.* [M12] have investigated this effect in detail, and find that the combination of low-ionisation fraction and high density lead to large resistivity and efficient diffusion of magnetic fields, making it difficult for the atmosphere to maintain a significant field strength at the surface, even if one is initially generated. It seems clear that this effect is tied to the temperature and density of the atmosphere, and not to the interior conditions – hence there should not be a difference in the activity of brown dwarfs and low-mass stars of the same spectral type. Note also that Figure 6.20 shows that there are some objects with measurable persistent emission even at very late spectral types, including an L5 and a few T dwarfs. Other parameters in addition to the low-ionisation fraction may therefore be important in discriminating between those objects with and without activity.

5.6.2 Spots and flares

As described in Section 5.4.1, spots have been observed on a few VLM objects, including brown dwarfs in the Pleiades and α Per at spectral types M6–M7. However, field objects at later spectral types show little sign of spot modulation [T1], although recent results indicate a possible detection in one object [B2]. There is a complication with ‘weather’ in dwarfs that have dusty atmospheres. Large weather systems such as the Great Red Spot on Jupiter may produce a rotational modulation signal similar to that expected from magnetic active regions, and thereby introduce ambiguity into the interpretation of the results [T4]. Further discussion of weather in brown dwarfs is found in Section 6.7.5.

Observations of flares provide additional evidence of magnetic activity in VLM stars. The well-known dM8e star VB 10 has been observed to flare in H α [H11], the ultraviolet [L8] and the X-ray [F6]. Both the UV and X-ray flares were quite energetic compared to the (undetected) quiescent emission, lying at the upper end of the flare-energy/quiescent luminosity relationship given in [L1]. A flare on a dM9.5e object discovered with 2MASS showed noticeable enhancements in H α and other optical emission lines, and in optical continuum radiation [L4]. Figure 5.25 depicts the flare spectrum, which appears quite similar to those of earlier type dMe flares. (The PC0025 object is shown for comparison; it clearly does not exhibit a flare signature.) The flare attained a very high luminosity – perhaps nearly equal to L_{bol} – during the impulsive phase. An X-ray flare on the

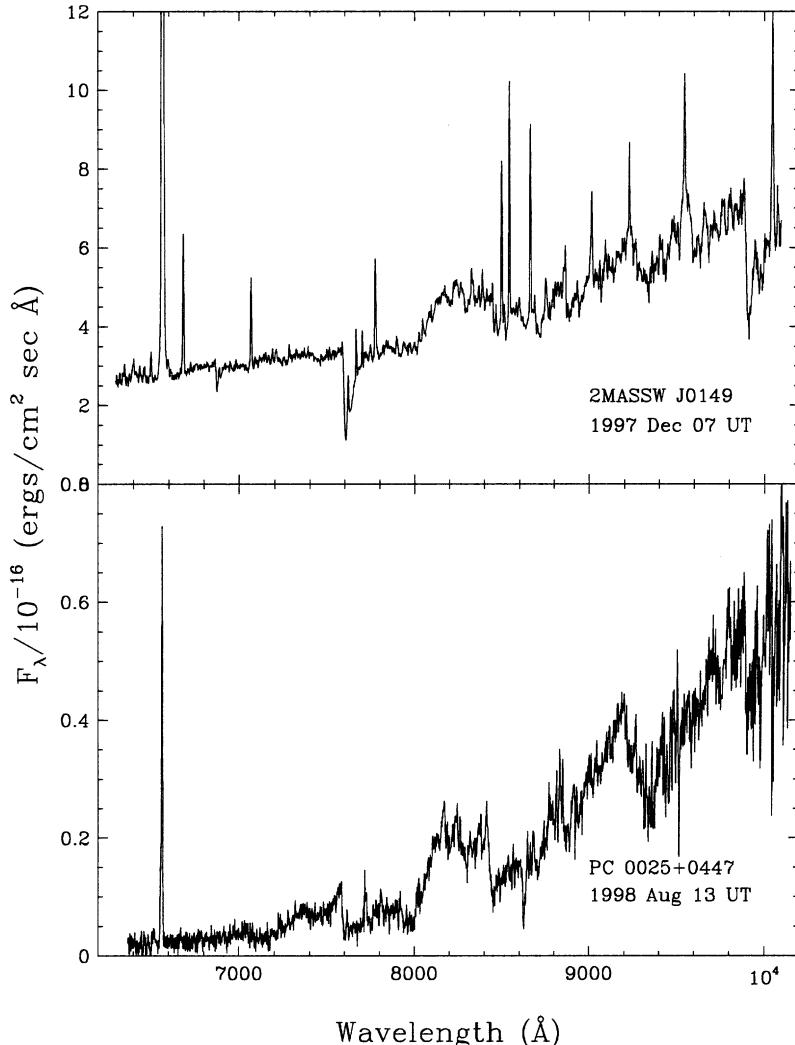


Figure 5.25. Flare spectrum of a dM9.5e 2MASS dwarf, together with a spectrum of the interesting dM9.5e dwarf PC0025. The greatly enhanced emission lines and continuum in the flare spectrum are typical of flares on earlier-type M dwarfs, while the PC0025 spectrum is clearly not of flare origin. (Figure courtesy of J. Liebert [L4] and the *Astrophysical Journal*.)

M9 brown dwarf LP 944-20 also produced a significant X-ray luminosity [R8], while radio flares on late M and L dwarfs greatly exceed the expected radio luminosities based on observations of earlier M dwarfs [B17], [B18]. The frequency of flaring in these objects is not yet documented, but will be essential in deciding whether a significant fraction of their magnetic energy is emitted during flares, as suggested by these studies.

5.7 COMPILED AND SYNTHESIS OF ACTIVITY RESULTS

Table 5.3 provides a compilation of the various results described throughout this chapter. The activity features are organised into three categories: solar-type dwarfs, early to mid M dwarfs, and late M–L dwarfs. The essential points are:

- The $\alpha\Omega$ (shell) dynamo operates in solar-type stars, and explains the activity dependence on rotation and youth. A turbulent component may be partially responsible for early chaotic behaviour, before the shell dynamo begins to dominate and produce cyclic behaviour as seen in the Sun.
- In the early M dwarfs (M0–M3), before the onset of complete convection, the shell dynamo probably still plays a role, with the turbulent dynamo becoming increasingly important at later spectral types. Moderate dependence on rotation (at least requiring a threshold level) and age, as in the solar-type stars, is expected. The lack of cycles and infrequent spot observations may be explained by the increased role of the turbulent dynamo.
- The fully convective M dwarfs (M3–M8) have the activity characteristics proposed for a turbulent dynamo: few spots, chaotic coverage with no global properties such as cycles or organised dipole fields, and little or no dependence on rotation. While the active fraction increases through type M8, the activity strength begins to decline at M6. The longevity of the activity, which increases in stars of lower mass, must be related to the efficiency of the dynamo production and/or heating mechanisms, but no theory yet exists to explain the observations in detail.

Table 5.3. Magnetic activity properties of low-mass dwarfs.

Feature	Solar type	Early–mid M	Late M–L
Magnetic fields	Yes	Yes	Yes?
B, filling factor	1,000 G, 1%	3,000 G, 50%	?
Spots	Yes	Yes (some, low amp)	Yes (some, low amp)
Cycles	Yes	No	?
Flares	Yes	Yes	Yes
Chromospheric radiation	Ca II	H α	H α ?
Fraction that are active	Small	Increasing → all	Decreasing → none
Activity strength	Weak	Strong, large scatter	Weakening → none
Rotation	Correlated	Weak/none (threshold?)	None/anti-correlation? fast rotation inhibits?
X-rays	Yes	Yes	?
Radio	Yes	Yes	Yes (strong!)
Age behaviour	Decays with time	Lasts longer at lower masses	?
Dynamo	Shell	Turbulent	Turbulent → none (primordial?)

- The latest-type dwarfs (M9–L) show a significant decline in activity. This could be attributed to a loss of efficiency in the turbulent dynamo at low mass, perhaps because the convection becomes less vigorous in the cooler objects. Other contributing factors may be the formation of radiative zones in atmospheres with significant dust opacity, perhaps impeding the emergence of magnetic flux in the surface layers; and the low ionisation fraction in the very cool atmospheres making it difficult to maintain significant fields. A further caveat is that it has not yet been established that $H\alpha$ is the best diagnostic of the magnetically heated outer atmospheres in these cool dwarfs.
- The role of the primordial field – particularly in very young brown dwarfs – has yet to be explored. The decay of the primordial field may contribute to the differences in activity observed between brown dwarfs in young star-formation regions (strong in $H\alpha$ and X-ray emission) and brown dwarfs in older clusters such as the Pleiades and in the field (weak or no activity). The activity dependence on the surface temperature, regardless of whether the object is a star or a brown dwarf, also contributes to these differences

5.8 SUMMARY

We have discussed the dynamo production of magnetic fields, and direct detection techniques in solar-type and lower-mass dwarfs. A few M dwarfs have strong magnetic fields with larger filling factors than are observed on the Sun. Properties of solar and M dwarf outer atmospheres have been reviewed, together with general modelling techniques. The evidence for spots and flares on M dwarfs has been described: spots are less prevalent than in active solar-type dwarfs, but exist at low amplitude in some objects. Flares are much more energetic compared with their bolometric luminosity in M dwarfs than in the Sun, but a model for their production that follows a solar analogy matches most of the observable phenomena. The mechanisms for producing the blue continuum emission and very broad lines remain elusive. Flare effects on the local stellar neighbourhood through mass loss and energetics may be important.

Global properties of active M dwarfs have been described: their colours, molecular bandstrengths, luminosities, and emission behaviour across the electromagnetic spectrum. The fraction of active M dwarfs rises monotonically from types M0–M8, peaking near 70% at type M8. The magnetic activity strength, measured by $L_{H\alpha}/L_{bol}$ remains at a nearly constant level from M0 through M5, declining toward later types. There is no obvious correlation with rotation, as in solar-type stars, other than a possible requirement of a ‘threshold’ value of $< 2 \text{ km s}^{-1}$. In the latest type M and L dwarfs, it is common to find rapid rotation and little or no activity. These latest objects may lose their activity because of changes in the dynamo production induced at low mass (such as reduced convection), or because the atmospheric properties preclude magnetic flux emergence in regions of significant dust formation and/or low ionisation fraction. The role of the primordial field is as yet unexplored. Understanding the properties of magnetic activity in the mid-M dwarfs,

and the decline in magnetic activity in the lowest mass stars and brown dwarfs, will be a rich topic in the coming years. In particular, we do not yet know if the effects of surface magnetic activity on the internal structure and evolution of these objects is inconsequential, or profound.

5.9 REFERENCES

- A1 Abada-Simon, M., Aubier, M., 1997, *A&AS*, **125**, 511.
- A2 Abbott, W. P., Hawley, S. L., 1999, *ApJ*, **521**, 906.
- A3 Allard, F., Hauschildt, P. H., 1995, *ApJ*, **445**, 433.
- A4 Arnaud, M., Rothenflug, R., 1985, *A&AS*, **60**, 425.
- A5 Ayres, T. R., 1979, *ApJ*, **228**, 509.
- A6 Ayres, T. R., 1981, *ApJ*, **244**, 1064.
- A7 Ayres, T. R., Rabin, D., 1996, *ApJ*, **460**, 1042.
- A8 Ayres, T. R., Marstad, N. C., Linsky, J. L., 1981, *ApJ*, **247**, 545.
- A9 Arzner, K., Guedel, M., 2004, *ApJ*, **602**, 363.
- A10 Allred, J. C., Hawley, S. L., Abbott, W. P., 2004, *AAS*, 204.0305
- A11 Amado, P. J., Byrne, P. B., 1997, *A&A*, **319**, 967.
- B1 Babcock, H. W., 1961, *ApJ*, **133**, 572.
- B2 Bailer-Jones, C. A. L., Mundt, R., 1999, *A&A*, **348**, 800.
- B3 Baliunas, S. L., *et al.*, 1995, *ApJ*, **438**, 269.
- B4 Barnes, S. A., Sofia, S., Prosser, C. F., Stauffer, J. R., 1999, *ApJ*, **516**, 263.
- B5 Basri, G., Marcy, G. W., 1995, *AJ*, **109**, 762.
- B6 Benedict, G. F., *et al.*, 1998, *AJ*, **116**, 429.
- B7 Bondar, N. I., 1995, *A&AS*, **111**, 259.
- B8 Bopp, B. W., Fekel, F., 1977, *AJ*, **82**, 490.
- B9 Bromage, G. E., Phillips, K. J. H., Dufton, P. L., Kingston, A. E., 1986, *MNRAS*, **220**, 1021.
- B10 Bruls, J. H. M. J., Solanki, S. K., Schuessler, M., 1998, *A&A*, **336**, 231.
- B11 Butler, C. J., Rodono, M., Foing, B. H., 1988, *A&A*, **206**, L1.
- B12 Buzasi, D. L., 1997, *ApJ*, **484**, 855.
- B13 Byrne, P. B., 1996, in ‘Stellar surface structure’, *Proceedings of IAU Symposium 176*, ed. Strassmeier, K. G. and Linsky, J. L., Kluwer, Dordrecht, p. 299.
- B14 Bercik, D. J., Fisher, G. H., Johns-Krull, C. M., Abbott, W. P., 2005, *ApJ*, in press.
- B15 Barnes, J. R., Collier Cameron, A., 2001, *MNRAS*, **326**, 950.
- B16 Barnes, J. R., James, D. J., Cameron, A. C., 2002, *AN*, **323**, 333.
- B17 Berger, E., 2002, *ApJ*, **572**, 503.
- B18 Berger, E. *et al.*, 2005, *ApJ*, in press.
- C1 Canfield, R. C. *et al.*, 1986, in *Energetic Phenomena on the Sun*, ed. M. Kundu and B. Woodgate, NASA Conf. Publ. 2439, Chapter 3.
- C2 Carlsson, M., Stein, R. F., 1997, *ApJ*, **481**, 500.
- C3 Coleman, G. D., Worden, S. P., 1976, *ApJ*, **205**, 475.
- C4 Contadakis, M. E., 1995, *A&A*, **300**, 819.
- C5 Craig, I. J. D., McClymont, A. N., Underwood, J. H., 1978, *A&A*, **70**, 1.
- C6 Cram, L. E., 1982, *ApJ*, **253**, 768.
- C7 Cristaldi, S., Gershberg, R. E., Rodono, M., 1980, *A&A*, **89**, 123.
- C8 Cully, S. L., Fisher, G. H., Abbott, M. J., Siegmund, O. H. W., 1994, *ApJ*, **435**, 449.

- C9 Cully, S. L., Fisher, G. H., Hawley, S. L., Simon, T., 1997, *ApJ*, **491**, 910.
- D1 DeLuca, E. E., Gilman, P. A., 1991, 'The solar dynamo', in *The Solar Interior and Atmosphere*, University of Arizona Press, p. 275.
- D2 Delfosse, X., Forveille, T., Perrier, C., Mayor, M., 1998, *A&A*, **331**, 581.
- D3 Dennis, B. R., Schwartz, R. A., 1989, *Solar Phys.*, **121**, 75.
- D4 Dennis, B. R., Zarro, D. M., 1993, *Solar Phys.*, **146**, 177.
- D5 Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., Young, P. R., 1997, *A&AS*, **125**, 149.
- D6 Doyle, J. G., Butler, C. J., Byrne, P. B., van den Oord, G. H. J., 1988, *A&A*, **193**, 229.
- D7 Durney, B. R., 1997, *ApJ*, **486**, 1065.
- D8 Durney, B. R., De Young, D. S., Roxburgh, I. W., 1993, *Solar Phys.*, **145**, 207.
- D9 Dobler, W., Stix, M., Brandenburg, A., 2005, *ApJ*, in press.
- E1 Eggen, O. J., 1990, *PASP*, **102**, 166.
- F1 Fang, C., Ding, M. D., 1995, *A&AS*, **110**, 99.
- F2 Fisher, G. H., 1989, *ApJ*, **346**, 1019.
- F3 Fisher, G. H., Hawley, S. L., 1990, *ApJ*, **357**, 243.
- F4 Fisher, G. H., Canfield, R. C., McClymont, A. N., 1985, *ApJ*, **289**, 425.
- F5 Fisher, G. H., Fan, Y., Howard, R. F., 1995, *ApJ*, **438**, 463.
- F6 Fleming, T. A., Giampapa, M. S., Schmitt, J. H. M. M., 2000, *ApJ*, **533**, 372.
- F7 Foukal, P., Lean, J., 1988, *ApJ*, **328**, 347.
- G1 Giampapa, M. S., Africano, J. L., Klimke, A., Parks, J., Quigley, R. J., Robinson, R. D., Worden, S. P., 1982, *ApJ*, **252**, 39.
- G2 Giampapa, M. S., Africano, J. L., Klimke, A., Parks, J., Quigley, R. J., Robinson, R. D., Worden, S. P., 1982, *ApJ*, **252**, L39.
- G3 Giampapa, M. S., Rosner, R., Kashyap, V., Fleming, T. A., Schmitt, J. H. M., Bookbinder, J. A., 1996, *ApJ*, **463**, 707.
- G4 Giampapa, M. S., Worden, S. P., Linsky, J. L., 1982, *ApJ*, **258**, 740.
- G5 Gizis, J. E., 1998, *AJ*, **115**, 2053.
- G6 Gizis, J. E., Monet, D. G., Reid, I. N., Kirkpatrick, J. D., Williams, R. J., 2000, *AJ*, **120**, 1085.
- G7 Gray, D. F., 1992, *The Observation and Analysis of Stellar Photospheres*, 2nd edition, Cambridge University Press, p. 375.
- G8 Griffiths, N. W., Jordan, C., 1996, in *The Ninth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, ed. R. Pallavicini & A. Dupree, ASP Conf. Ser. 109, p. 647.
- G9 Grinin, V., 1983, in *Activity in Red Dwarf Stars*, ed. P. B. Byrne and M. Rodono, Reidel, Dordrecht, p. 223.
- G10 Guedel, M., Benz, A. O., Schmitt, J. H. M. M., Skinner, S. L., 1996, *ApJ*, **471**, 1002.
- G11 Gunn, A. G., Doyle, J. G., Mathioudakis, M., Houdebine, E. R., Avgoloupis, S., 1994, *A&A*, **285**, 489.
- G12 Guedel, M., Audard, M., Kashyap, V. L., Drake, J. J., Guinan, E. F., 2003, *ApJ*, **582**, 423.
- G13 Guedel, M., Schmitt, J. H. M. M., Bookbinder, J. A., Fleming, T. A., 1993, *ApJ*, **415**, 236.
- H1 Hale, G. E., 1908, *ApJ*, **28**, 315.
- H2 Hambaryan, V. V., 1998, in *The Tenth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, ed. R. Donahue and J. Bookbinder, ASP Conf. Ser. 154, p. 1492.
- H3 Hatzes, A. P., Vogt, S. S., Ramseyer, T. F., Misch, A., 1996, *ApJ*, **469**, 808.
- H4 Hawley, S. L., Pettersen, B. R., 1991, *ApJ*, **378**, 725.
- H5 Hawley, S. L., Fisher, G. H., 1992a, *ApJS*, **78**, 565.

- H6 Hawley, S. L., Fisher, G. H., 1992b, *ApJS*, **81**, 885.
 H7 Hawley, S. L., *et al.* (10 others), 1995, *ApJ*, **453**, 464.
 H8 Hawley, S. L., Gizis, J. E., Reid, I. N., 1996, *AJ*, **112**, 2799.
 H9 Hawley, S. L., Tourtellot, J. G., Reid, I. N., 1999, *AJ*, **117**, 1341.
 H10 Hawley, S. L., Reid, I. N., Tourtellot, J. G., 2000, in *Very Low Mass Stars and Brown Dwarfs*, ed. R. Rebolo, Cambridge University Press, p. 109.
 H11 Herbig, G., 1956, *PASP*, **68**, 531.
 H12 Houdebine, E. R., 1992, *IrAJ*, **20**, 213.
 H13 Houdebine, E. R., Doyle, J. G., 1994, *A&A*, **289**, 169.
 H14 Houdebine, E. R., Foing, B. H., Doyle, J. G., Rodono, M., 1993, *A&A*, **278**, 109.
 H15 Houdebine, E. R., Mathioudakis, M., Doyle, J. G., Foing, B. H., 1996, *A&A*, **305**, 209.
 H16 Houdebine, E. R., Stempels, H. C., 1997, *A&A*, **326**, 1143.
 H17 Hudson, H. S., 1988, *ARA&A*, **26**, 473.
 H18 Hudson, H. S., Acton, L. W., Hirayama, T., Uchida, Y., 1992, *PASJ*, **44**, L77.
 H19 Hawley, S. L., *et al.*, 2003, *ApJ*, **597**, 535.
 H20 Hawley, S. L., Johns-Krull, C. M., 2003, *ApJ*, **588**, L109.
 I1 Ichimoto, K., Kurokawa, H., 1984, *Solar Phys.*, **93**, 105.
 J2 Jackson, P. D., Kundu, M. R., White, S. M., 1989, *A&A*, **210**, 284.
 J3 Johns-Krull, C. M., Valenti, J. A., 1996, *ApJ*, **459**, L95.
 J4 Johns-Krull, C. M., Hawley, S. L., Basri, G., Valenti, J. A., 1997, *ApJS*, **112**, 221.
 J5 Jones, K. L., Page, A. A., 1991, *PASAu*, **9**, 277.
 J6 Joy, A. H., Abt, H. A., 1974, *ApJ*, **28**, 14.
 K1 Kahler, S. *et al.*, 1982, *ApJ*, **252**, 239.
 K2 Kalkofen, W., 1991, 'The heating of the solar chromosphere', in *The Solar Interior and Atmosphere*, University of Arizona Press, p. 911.
 K3 Katsova, M. M., Boikko, A. Ya., Livshits, M. A., 1997, *A&A*, **321**, 549.
 K4 Krishnamurthi, A. *et al.*, 1998, *ApJ*, **493**, 914.
 K5 Krishnamurthi, A., Leto, G., Linsky, J. L., 1999, *AJ*, **118**, 1369.
 K6 Kashyap, V. L., Drake, J. J., Guedel, M., Audard, M., 2002, *ApJ*, **582**, 423.
 L1 Lacy, C. H., Moffett, T. J., Evans, D. S., 1976, *ApJS*, **30**, 85.
 L2 Lanza, A. F., Catalano, S., Cutispoto, G., Pagano, I., Rodono, M., 1998, *A&A*, **332**, 541.
 L3 Leighton, R. B., 1969, *ApJ*, **156**, 1.
 L4 Liebert, J., Kirkpatrick, J. D., Reid, I. N., Fisher, M. D., 1999, *ApJ*, **519**, 345.
 L5 Lim, J., White, S. M., 1996, *ApJ*, **462**, L91.
 L6 Lin, R. P., Hudson, H. S., 1976, *Solar Phys.*, **50**, 153.
 L7 Linsky, J. L., Bornmann, P. L., Carpenter, K. G., Hege, E. K., Wing, R. F., Giampapa, M. S., Worden, S. P., 1992, *ApJ*, **260**, 670.
 L8 Linsky, J. L., Wood, B. E., Brown, A., Giampapa, M. S., Ambruster, C., 1995, *ApJ*, **455**, 670.
 L9 Lockwood, G. W., Skiff, B. A., Radick, R. R., 1997, *ApJ*, **485**, 789.
 M1 Marcy, G. W., 1984, *ApJ*, **276**, 286.
 M2 Marcy, G. W., Chen, G. H., 1992, *ApJ*, **390**, 550.
 M3 Martin, E. L., Basri, G., Delfosse, X., Forveille, T., 1997, *A&A*, **327**, L29.
 M4 Martin, E. L., Zapatero-Osorio, M. R., 1997, *MNRAS*, **286**, L17.
 M5 Mathioudakis, M., Fruscione, A., Drake, J. J., McDonald, K., Bowyer, S., Malina, R. F., 1995, *A&A*, **300**, 775.
 M6 Mauas, P. J. D., Falchi, A., 1994, *A&A*, **281**, 129.
 M7 Mavridis, L. N., Avgoloupis, S., 1993, *A&A*, **280**, L5.

- M8 Mochnacki, S., Zirin, H., 1980, *ApJ*, **239**, L27.
- M9 Mould, J., Cohen, J., Oke, J. B., Reid, I. N., 1994, *AJ*, **107**, 2222.
- M10 Mullan, D. J., Doyle, J. G., Redman, R. O., Mathioudakis, M., 1992, *ApJ*, **397**, 225.
- M11 Mullan, D. J., Stencel, R. E., Backman, D. E., 1989, *ApJ*, **343**, 400.
- M12 Mohanty, S., Basri, G., Shu, F., Allard, F., Chabrier, G., 2002, *ApJ*, **571**, 469.
- N1 Neidig, D. F., Kiplinger, A. L., Cohl, H. S., Wiborg, P. H., 1993, *ApJ*, **406**, 306.
- N2 Neuhauser, R. *et al.* (9), 1999, *A&A*, **343**, 883.
- N3 Neupert, W. M., 1968, *ApJ*, **153**, L59.
- N4 Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., Vaughan, A. H., 1984, *ApJ*, **279**, 763.
- O1 Oranje, B. J., 1986, *A&A*, **154**, 185.
- O2 Osten, R. A. *et al.*, 2005, *ApJ*, **621**, 398.
- P1 Parker, E. N., 1955, *ApJ*, **122**, 293.
- P2 Parker, E. N., 1975, *ApJ*, **198**, 205.
- P3 Pasquini, L., Lindgren, H., 1994, *A&A*, **283**, 179.
- P4 Patten, B. M., Simon, T., 1996, *ApJS*, **106**, 489.
- P5 Pettersen, B. R., Lambert, D. L., Tomkin, J., Sandmann, W. H., Lin, H., 1987, *A&A*, **183**, 66.
- P6 Pettersen, B. R., 1983, in *IAU Colloquium 71, Activity in Red-Dwarf Stars*, Reidel, Dordrecht, p. 17.
- P7 Pettersen, B. R., Hawley, S. L., Fisher, G. H., 1992, *Solar Phys.*, **142**, 197.
- P8 Pettersen, B. R., 1989, *Solar Phys.*, **121**, 299.
- P9 Pevtsov, A. A. *et al.*, 2003, *ApJ*, **598**, 1387.
- R1 Radick, R. R., 1991, in *The Sun in Time*, University of Arizona Press, p. 787.
- R2 Raymond, J. C., Smith, B. W., 1977, *ApJS*, **35**, 419.
- R3 Rice, J. B., 1996, in ‘Stellar surface structure’, *Proceedings of IAU Symposium 176*, ed. Strassmeier, K. G. and Linsky, J. L., Kluwer, Dordrecht, p. 19.
- R4 Robinson, R. D., 1980, *ApJ*, **239**, 961.
- R5 Rodono, M. *et al.* (21), 1986, *A&A*, **165**, 135.
- R6 Rosner, R., 1980, in *The First Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, p. 79.
- R7 Rosner, R., Tucker, W. H., Vaiana, G. S., 1978, *ApJ*, **220**, 643.
- R8 Rutledge, R. E., Basri, G., Martin, E. L., Bildsten, L., 2000, *ApJ*, **538**, L141.
- R9 Reid, I. N., Kirkpatrick, J. D., Gizis, J. E., Liebert, J., 1999, *ApJ*, **527**, L105.
- S1 Saar, S. H., 1996, in *IAU Colloquium 153, Magnetodynamic Phenomena in the Solar Atmosphere – Prototypes of Stellar Magnetic Activity*, eds. Y. Uchida, T. Kosugi, H. S. Hudson, Kluwer, Dordrecht, p. 367.
- S2 Saar, S. H., Linsky, J. L., 1985, *ApJ*, **299**, L47.
- S3 Saar, S. H., Linsky, J. L., Beckers, J. M., 1986, *ApJ*, **302**, 77.
- S4 Schou, J. *et al.*, 1998, *ApJ*, **505**, 390.
- S5 Schrijver, C. J., 1987, in *The 5th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, p. 89.
- S6 Schuessler, M., Solanki, S. K., 1992, *A&A*, **264**, L13.
- S7 Sheeley, N. R. Jr., Nash, A. G., Wang, Y.-M., 1987, *ApJ*, **319**, 481.
- S8 Skumanich, A., 1972, *ApJ*, **171**, 565.
- S9 Soderblom, D. R., Duncan, D. K., Johnson, D. R. H., 1991, *ApJ*, **375**, 722.
- S10 Stauffer, J. R., Hartmann, L., 1986, *ApJS*, **61**, 531.
- S11 Stauffer, J. R., Giampapa, M. S., Herbst, W., Vincent, J. M., Hartmann, L. W., Stern, R. A., 1991, *ApJ*, **374**, 142.

- S12 Stauffer, J. R., Liebert, J., Giampapa, M., Macintosh, B., Reid, I. N., Hamilton, D., 1994, *AJ*, **108**, 160.
- S13 Scholz, A., Eisloffel, J., 2004, *A&A*, **421**, 259.
- S14 Scholz, A., Eisloffel, J., 2004, *A&A*, **419**, 249.
- S15 Silvestri, N. M., Hawley, S. L., Oswalt, T. D., 2005, *AJ*, in press.
- T1 Terndrup, D. M., Krishnamurthi, A., Pinsonneault, M. H., Stauffer, J. R., 1999, *AJ*, **118**, 1814.
- T2 Tinney, C. G., Delfosse, X., Forveille, T., 1997, *ApJ*, **490**, L95.
- T3 Tinney, C. G., Reid, I. N., 1998, *MNRAS*, **301**, 1031.
- T4 Tinney, C. G., Tolley, A. J., 1999, *MNRAS*, **304**, 119.
- T5 Torres, C. A. O., Ferraz-Mello, S., 1973, *A&A*, **27**, 231.
- V1 Valenti, J. A., Johns-Krull, C. M., Piskunov, N., 1998, in *The Tenth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, ASP Conf. Ser. 154, p. 1357.
- V2 Van den Oord, G. H. J., Doyle, J. G., 1997, *A&A*, **319**, 578.
- V3 Veeder, G. J., 1974, *AJ*, **79**, 1056.
- V4 Vernazza, J. E., Avrett, E. H., Loeser, R., 1981, *ApJS*, **45**, 635.
- V5 Vilhu, O., 1984, *A&A*, **133**, 117.
- V6 Valenti, J. A., Johns-Krull, C. M., 2003, *Stars as Suns: Activity, Evolution and Planets*, IAU Symposium 219, p. 235.
- V7 van den Besselaar, E. J. M. *et al.*, 2003, *A&A*, **411**, 587.
- W1 Wielen, R., 1974, in *Highlights of Astronomy*, Vol. 3, ed. Contopoulos, G., Reidel, Dordrecht, p. 395.
- W2 Wilson, O. C., Bappu, M. K. V., 1957, *ApJ*, **125**, 661.
- W3 West, A. A. *et al.*, 2004, *AJ*, **128**, 426.
- Y1 Young, A., Skumanich, A., MacGregor, K. B., Temple, S., 1990, *ApJ*, **349**, 608.
- Z1 Zarro, D. M., Canfield, R. C., Strong, K. T., Metcalf, T. R., 1988, *ApJ*, **324**, 582.
- Z2 Zirin, H., 1988, *Astrophysics of the Sun*, Cambridge University Press.

5.10 HOMEWORK PROBLEMS

Problem 1

- Consider a stable sunspot on the Sun. Assuming the number density is approximately $1,017 \text{ cm}^{-3}$ in the photosphere (i.e., where optical depth in the continuum is ~ 1), compute the maximum magnetic field strength, B , that the sunspot can attain.
- How does the maximum field strength change for lower mass main-sequence stars? Why?
- Sunspots are observed on the Sun with $B > 3,000$ Gauss. Explain how this is possible. (You can assume the number you should have computed in part (a) was less than 3,000 Gauss.)

Problem 2

- Assume a dwarf of spectral type M3 has a starspot of radius 70,000 km, with a temperature 1,000 K lower than the stellar effective temperature. If the star is

located at a distance of 10 parsecs from the Sun, what is the bolometric magnitude variation as the starspot rotates in and out of view (assume that this is not a polar spot, but is at times completely visible and at other times completely invisible). Table 4.1 contains useful radius and temperature estimates. (Hint: model the star as consisting of two nesting shells, one 1,000 K hotter than the other, with the starspot a ‘window’ into the cooler shell.)

- (b) At what wavelength is the variation a maximum? Can this variation be detected with modern instrumentation?

Problem 3

- (a) Draw a plot of temperature vs. height in a mid-M dwarf atmosphere. Label the five major regions of the atmosphere (photosphere, temperature minimum region, chromosphere, transition region and corona). Indicate approximate column mass and continuum optical depth for each region.
- (b) Describe the dominant emission processes in each region of the atmosphere, including discussion of whether the emission is optically thick or optically thin.
- (c) Discuss where in the atmosphere NLTE effects are important, and for which radiative processes.
- (d) Why is the transition region so limited in spatial extent?

Problem 4

Using the 8-pc data available from the NLDS2 data website (<http://www.stsci.edu/~inr/nldsdat.html>), plot the fraction of stars that show H emission vs. spectral type. Note how many stars are in each spectral type bin. Compare your plot to Figure 5.18 (from [W3]) and comment on similarities and differences.

6

Brown dwarfs

6.1 INTRODUCTION

The existence of substellar mass, star-like objects was first considered seriously by Kumar ([K8] Section 3.3), who outlined their essential properties: no central energy source due to hydrogen fusion; pressure support against gravitational collapse provided by electron degeneracy; and short luminous lifetimes. Like low-mass M dwarfs, these objects, re-christened ‘brown dwarfs’ by Tarter [T1], are ideal baryonic dark matter candidates, and have been the target of a wide variety of surveys over the last three decades. As with any search for a newly-hypothesised source, the first task was determining whether any exist. All available techniques can be brought into play to tackle that question, without regard to complications due to biases and selection effects. However, once existence is confirmed, the emphasis of observational programs should switch to addressing specific issues through statistically well-controlled surveys. With the near-contemporaneous discoveries of Gl 229B and the planet circling 51 Peg, both brown dwarf and planet surveys have moved from the first to the second stage. At least in the case of brown dwarfs, the extensive catalogue of examples supplied by the DENIS, 2MASS, and SDSS projects are providing the data necessary for a transition from phenomenology to statistics.

This chapter summarises results from brown dwarf surveys obtained prior to mid-2004; Chapter 11 considers extrasolar planetary systems. After outlining the observational signatures of substellar-mass brown dwarfs, we review some of the earlier results (and false steps) before describing the first successes (not all immediately recognised as such). The field as a whole has undergone a revolution in the decade following the discovery of Gl 229B, with the identification of numerous isolated brown dwarfs in the general field and young clusters, including objects cooler than Gl 229B itself. This catalogue of substellar objects supplies an empirical description of how properties change with age and temperature, as discussed in the concluding sections of this chapter.

6.2 HOW TO RECOGNISE A BROWN DWARF

Brown dwarfs are defined by their mass. Unfortunately, that parameter is rarely measurable in a direct fashion (Section 9.3). A handful of brown dwarf binaries have been identified that may eventually provide direct calibration of the substellar regime, but so far no accurate mass measurements have been obtained. Lacking such data, we have to infer masses by comparing secondary indicators – luminosity and temperature – against theoretical predictions. The theoretical models described in Chapter 3 indicate that objects with $L < 10^{-4} L_\odot$ (see Figure 3.8) or $T_{\text{eff}} < 1,900 \text{ K}$ (see Figure 3.9 and Figure 6.1) are at or below the stellar mass limit. Classification is

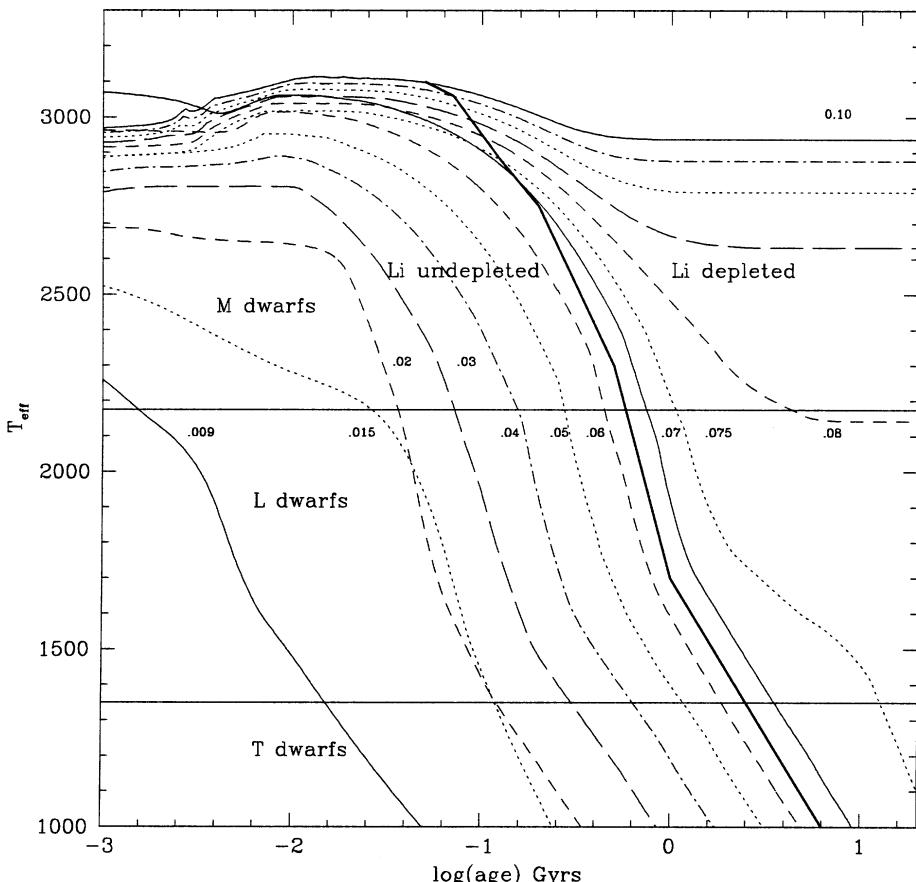


Figure 6.1. The lithium depletion line superimposed on evolutionary tracks for low-mass stars and brown dwarfs (labelled by mass in M_\odot) predicted using the Tucson models. Models to the right of the solid line have depleted lithium by over 90%. Brown dwarfs with masses below $\sim 0.065 M_\odot$ are predicted to retain lithium at the primordial abundance level throughout their lifetime.

more ambiguous for younger, hotter brown dwarfs, since current empirical and theoretical temperature calibrations are not sufficiently accurate that they can distinguish stars from brown dwarfs between 10^{-2} and $10^{-4} L_\odot$ (see Figure 3.12).

One technique that can be used to identify lower mass brown dwarfs is the lithium test – detection of the Li I 6,708-Å absorption line in mid- to late-M dwarfs [M2], [R1]. The primordial lithium abundance is estimated as $[Li] = 3.3$ on a logarithmic scale where $[H] = 12.0$: that is, there are 5×10^8 hydrogen atoms for every lithium atom. Lithium absorption is detected easily in M-type T Tauri stars with ages of a few million years, but, as discussed in Section 3.3.3, primordial lithium is destroyed by reaction 5^{II} of the PP chain. The reaction rate is temperature dependent, requiring a minimum temperature of 3×10^6 K. As a consequence, the rate of depletion varies with mass, and objects with masses below $0.055\text{--}0.065 M_\odot$ (some uncertainty remains in the model calculations – [B1], [B6], [B7]) retain lithium at the primordial abundance level throughout their lifetime. Figure 6.1 shows the lithium depletion line as a function of age and temperature superimposed on evolutionary tracks from the Tucson models. A $0.075 M_\odot$ dwarf, on the hydrogen burning threshold, is predicted to deplete lithium in $\sim 1.5 \times 10^8$ years, by which time its effective temperature is $\sim 2,800$ K, or spectral type M5. The implication is that any dwarf with a spectral type M6 or later and detectable lithium absorption is a substellar-mass brown dwarf.

Neither detailed models nor the lithium test were available to the first surveys for substellar objects. From the outset, however, it has been clear that brown dwarfs cool rapidly to effective temperatures below 2,500 K, so most investigations used broadband photometry to search for objects with extremely red colours. These programs were of two types: extensions of photometric parallax surveys for low-mass stars in the field; or targeted observations, searching for low-mass companions to known nearby stars. Initial studies were concentrated at red or far-red optical wavelengths (R, I passbands), but brown dwarfs emit most of their energy at wavelengths beyond 1 μm. Thus, the emphasis shifted to the near-infrared as soon as appropriate technology became available.

With the actual discovery of brown dwarfs, it became apparent that searching for objects that have red colours in *all* passbands is not an infallible discovery technique. In particular, methane absorption at temperatures below $\sim 1,300$ K leads to blue near-infrared colours. This behaviour was predicted by Tsuji in 1964 [T11], but the prediction was accorded scant recognition until Gl 229B brought matters forcibly to the attention of the astronomical community. Surveys that searched for sources with extremely red near-infrared colours could not detect T dwarfs. On the other hand, most of the L dwarfs discussed in Section 2.2.3 have $(J - K_S) > 1.3$, so those objects, which lie below the hydrogen-burning limit, meet the criteria adopted in the initial surveys.

Before moving on to discuss the results from these surveys, we need to consider an issue in classification: how does one discriminate between a low-mass brown dwarf and a high-mass planet? As a first criterion, we require that a ‘planet’ is a substellar-mass object bound dynamically to a much more massive companion: there are no ‘free-floating planets’. How do we distinguish between a brown dwarf binary

and a planetary system? In principle, we could classify using the mode of formation: planets form in a disk; brown dwarfs form as separate, accreting entities, like stars. In practice, unfortunately, we cannot reverse time and reconstruct the history of each low-mass companion, but must segregate systems based on their present properties. There have been suggestions that the brown dwarf/planet boundary should be set at the mass limit for deuterium burning ($\sim 0.013 M_{\odot}$ or ~ 14 Jupiter masses, M_J , at solar abundance, [B7]). This classification matches the fusion criterion that defines the stellar/brown dwarf boundary, and would apply to both single and multiple objects. Deuterium, however, is almost impossible to detect in objects at these low temperatures, so this creates another observationally imprecise division. In fact, as discussed further in Chapter 11, there may be a minimum in the mass distribution of companions at $10\text{--}30 M_J$. This might reflect a break between the mass distributions of brown dwarf and planetary companions. For the present, a grey area exists in classifying the lower extreme of the brown dwarf mass distribution. Here, we consider all companions with masses exceeding $15 M_J$ and all isolated T-type (Gl229B-like) dwarfs as brown dwarfs.

6.3 FIRST STEPS: SOME INTERESTING MISTAKES

Textbooks occasionally give the impression that science progresses in a smooth, orderly fashion, with one discovery leading naturally to the next. In fact, that is generally not the case: the path to greater knowledge follows a zigzag route, sometimes with unforeseen tangential diversions and the occasional U-turn out of a blind alley. The route toward the discovery of brown dwarfs is strewn with the carcasses of hopeful monsters. It is instructive to review some of those discarded hypotheses.

Many late-type M dwarfs have been suggested as brown dwarf candidates at one time or another. These include VB 10, LHS 2924 and the unusual dMe dwarf, PC0025 + 0447 (Section 5.5.4). Indeed, Kirkpatrick *et al.* [K2] suggested that the latest type M dwarfs might all be brown dwarfs: their survey seemed to indicate that those objects are closely confined to the Plane, with most lying in the southern hemisphere (i.e., between the Sun and the Galactic mid-Plane). That type of distribution would be consistent with a very young (i.e., substellar-mass yet visible) population. Subsequent surveys (e.g., [T4], [T7]), however, revealed a much more extended distribution of ‘ultracool’ dwarfs (spectral types later than M6), and the discovery of L and T dwarfs, coupled with the development of improved evolutionary models, sealed the fate of that hypothesis. Yet it is striking that such issues were still open to debate as recently as 1994.

6.3.1 VB 8B – the disappearing brown dwarf

Although eventually proven spurious, the hypothetical low-mass companion to Gl 644C (VB 8) had more impact on the field of brown dwarf surveys than many subsequent actual discoveries. The source was identified from one-dimensional

speckle scans of VB 8 [M1] (see Section 1.8.1). The visibility function suggested the presence of a companion, 3 magnitudes fainter than the primary ($K = 8.8$), at a separation of 1.0 arcseconds. Both this star and VB 10 had been previously suspected of having astrometric companions [H4], and the speckle data appeared to provide confirmation for VB 8. Both the low luminosity at $2.2\text{ }\mu\text{m}$ and the red infrared colours ($(H-K) \sim 1.3$) suggested a brown dwarf companion, and VB 8B was accepted as the first clear case of a substellar-mass dwarf. Assuming an age of $\sim 3.5\text{ Gyr}$ for the Wolf 629/630 system (Gl 643/644), VB 8B was inferred to have a mass between 0.065 and $0.085 M_{\odot}$, placing it just below the hydrogen-burning limit.

This apparent discovery prompted the first conference devoted exclusively to brown dwarfs (the 1986 Astrophysics of Brown Dwarfs Workshop, Fairfax, Virginia). VB 8B, however, did not long survive that meeting. Skrutskie *et al.* [S3] obtained direct K -band images of VB 8 using one of the first near-infrared cameras, equipped with a 32×32 element Rochester InSb array. Those data, taken at K - and L -band under excellent seeing conditions (a FWHM of 0.64 arcseconds in a 5-second exposure), showed no evidence of the expected companion, although simulations showed that such an object should have been detected easily even at separations as small as 0.6 arcseconds. The astrometric data for VB 8 were inconsistent with a short period, and hence with a rapid change in the projected separation between primary and hypothetical secondary.

The clinching non-observation of VB 8B, however, was made by Perrier and Mariotti [P2], who obtained new K - and L -band speckle interferometric scans and found no evidence for any companion brighter than $K \sim 14.5$ at the expected position. The inferred mass ratios and standard orbital dynamics rule out the possibility of the ‘companion’ lying along the line of sight to VB 8. The only conclusion possible is that the original detection was an observational artefact, probably due to the chromatic effects of atmospheric refraction. McCarthy *et al.* [M1] compared their original scans of VB 8 against data for a known point source, but at significantly different altitudes, and therefore with different image profiles. VB 8B thus became one in a long line of brown dwarf candidates that failed to survive rigorous scrutiny.

6.3.2 G 29-38 – dust or a brown dwarf?

Cool companions of white dwarfs can dominate the flux distribution at long wavelengths, leading to an apparent infrared excess. The first programs that used this method to search for low-mass stars and brown dwarfs were limited to aperture photometry (see Section 8.7.2). The most extensive such survey was undertaken by Zuckerman and Becklin [Z2], who targeted over 200 nearby white dwarfs at J , H and K . One of the earliest targets was G 29-38, a DA white dwarf that lies near the lower temperature boundary of the ZZ Ceti pulsational-instability strip, $T_{\text{eff}} \sim 11,500\text{ K}$. The star, which is also known as ZZ Psc, was identified in the Lowell proper motion survey and confirmed as a DA (hydrogen atmosphere) white dwarf (EG 159) by Eggen and Greenstein [E1]. Aperture photometry revealed significant excess flux at

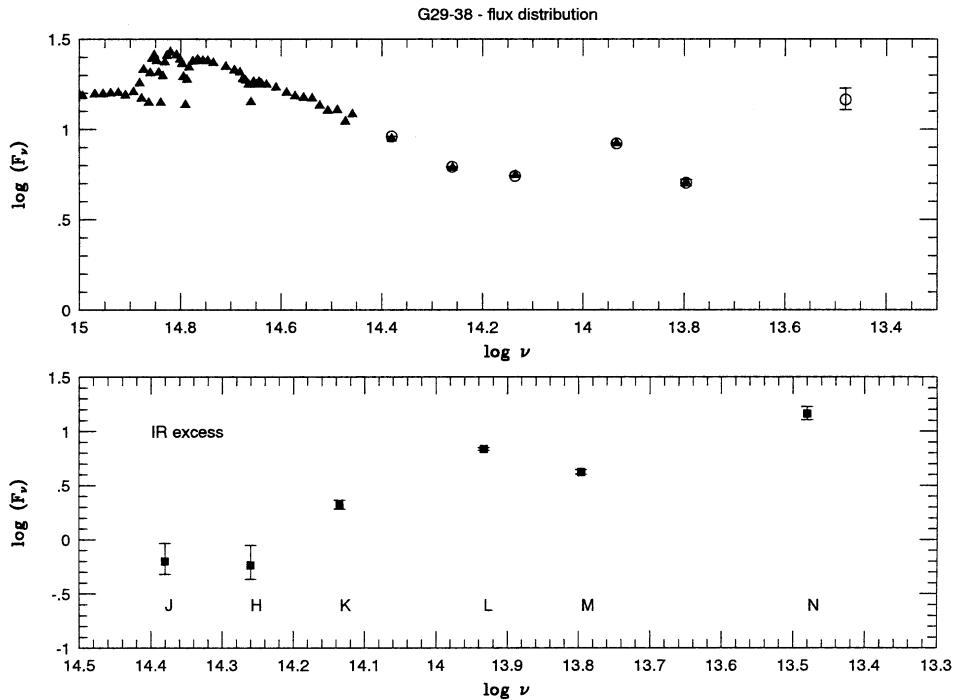


Figure 6.2. The flux distribution of G 29-38: (top panel) plots the combined flux distribution; (bottom panel) shows the residuals after the white dwarf contribution is subtracted.

infrared wavelengths [Z1]; the $(V-K)$ colour is 0.6 magnitudes redder than expected, with $(V-K)=+0.3$ rather than -0.3 magnitudes. Subtracting the white dwarf's contribution from the broadband photometry reveals a residual flux distribution that peaks at $\sim 3.5\text{ }\mu\text{m}$ (Figure 6.2).

Initially, the infrared excess was interpreted as being due to a low-mass companion. Greenstein [G6] estimated a black-body temperature in the range 1,100 to 1,500 K and a luminosity of $4 \times 10^{-5} L_\odot$, corresponding to a mass of 0.04–0.08 M_\odot , with the white dwarf and the brown dwarf making equal contributions at $2.2\text{ }\mu\text{m}$. Direct imaging showed that G 29-38 is unresolved at the 0.8 arcsecond level [T8]. One-dimensional speckle scans suggested that there was a slight elongation in the north–south direction, consistent with two point sources of equal luminosity (at $2.2\text{ }\mu\text{m}$) separated by $0.^{\hspace{-0.1em}\prime\prime}23$ [H1], but subsequent speckle observations by Kuchner *et al.* [K7] contradict that hypothesis. There is therefore no reliable evidence for a resolved companion.

As more photometric data were acquired, additional inconsistencies emerged. First, the implied radius of the companion is $\sim 0.15 R_\odot$ [Z1] – almost twice that predicted by models. Second, infrared spectroscopy [T9] fails to reveal any features in the K -band window (1.95 – $2.45\text{ }\mu\text{m}$): either CO or CH₄ absorption

should be visible for a cool dwarf companion. Finally, the system is detected at 10 μm [T2], [T9] at a level well above that expected for even an 1,100 K companion. No single temperature black-body can fit all of the photometric data, although one might accommodate the mid-infrared excess by adding dust to a low-temperature atmosphere.

The defining test of the companion hypothesis, however, rests with the variability of G 29-38. ZZ Ceti stars exhibit a complex range of flux variations due to non-radial pulsations with periods in the range of 200 to 1,000 seconds and amplitudes of up to 0.3 magnitudes. G 29-38 is one of the largest amplitude variables of this type. Typically, the variations are dominated by a single frequency mode (with a period of a few minutes), but numerous other lower amplitude variations at different frequencies are also present. In order to disentangle the complex interplay among these different modes, Nather *et al.* [N3] devised the concept of the Whole Earth Telescope (WET). This is a series of telescopes distributed around the world, with sufficient redundancy in sky coverage that a given star can be monitored continuously for 24 hours. Continuous coverage allows unambiguous determination of the intrinsic mode of variation, free of the aliasing problems that are introduced if observations are interrupted by, for example, an object rising and setting as viewed from a single site. The crucial point with regard to G 29-38 is that, if there were a binary companion, then the distance to Earth, and hence the travel time of light, would change with the same period as the binary orbit. This would produce a long-term modulation in the frequency of any stable intrinsic pulsation associated with the white dwarf.

In the analysis of the initial data set [W2], one particular mode of variation ($P = 615$ seconds) appeared to show exactly the behaviour expected if the star were in a binary orbit. The residuals (the observed minus calculated time of maximum for that mode) implied that the orbital period was 90 days and that the companion had a mass exceeding 0.3 solar masses – well above the brown dwarf limit. However, with such a massive companion, one would expect to see substantial radial velocity variations ($\pm 28 \text{ km s}^{-1}$) synchronised with the modulation in the pulsation. Those variations are not observed [G3]. Barnbaum and Zuckerman [B2] argue that lower amplitude ($\pm 5 \text{ km s}^{-1}$) variations are present with a period of 11.2 months, but Clemens *et al.* [C1] have demonstrated that non-radial pulsations can introduce apparent velocity shifts of this order.

Patterson *et al.* [P1] monitored G 29-38 photometrically at both blue and infrared wavelengths and found no evidence for long-period modulations of any stable modes visible at optical or infrared wavelengths, including the $P = 615$ second mode. Crucially, a number of the higher frequency variations are detected at *both* optical and infrared wavelengths, showing that the latter variations are driven by the white dwarf – probably by surface temperature variations due to non-radial pulsations. Finally, Kleinman *et al.* [K6] analysed WET data from a series of observing runs spaced over a period of five years. Several modes of variation, notably a 284-second periodicity, are evident throughout the full set of observations, and there is no evidence for orbital modulation. All of these results argue strongly against the brown dwarf hypothesis.

The infrared excess, however, is clearly real and has been verified by numerous observers – most recently by the ISO and Spitzer satellites. The only reasonable alternative to a discrete companion is circumstellar dust. This can account both for the presence of pulsations at infrared wavelengths, through reprocessing of light from the white dwarf, and for the fact that the infrared excess fails to match a single temperature black-body curve. However, there are also a number of complications involved in this interpretation of the available data.

First, maintaining temperatures of $\sim 1,000$ K requires that the dust particles are at radii of only $1 R_{\odot}$ from the white dwarf. At such small separations from the primary, the Poynting–Robertson effect, a net drag exerted on small particles by the absorption and re-emission of radiation, leads to short orbital lifetimes. Indeed, it was on this basis that Zuckerman and Becklin originally rejected the dust–cloud hypothesis. The Poynting–Robertson depletion timescale is given by

$$t_{PR} \sim 2 \times 10^4 r \text{ years} \quad (6.1)$$

where r is the radius of the particle in millimetres. Since periodic infrared pulsations are observed, the dust particles must be small enough to have thermal timescales that allow them to respond to the rapid changes in the surface temperature of the white dwarf. This sets an upper limit of ~ 100 mm on the particle radius [G2], implying a depletion time of less than 2 million years – over 100 times shorter than the cooling time of the white dwarf. Thus, we are either lucky enough to witness a rare, short-lived phenomenon (unlikely, but consider comet Shoemaker–Levy), or the radiative grains are being replenished from an external source.

This raises the matter of the origin of the dust. G 29-38 has a cooling time (as a white dwarf) of 1–2 Gyr, so one would expect any residual dust from the planetary nebula to have dispersed. Particulate matter at radii of a few R_{\odot} in the main sequence system should have been engulfed during the red giant stages of evolution. Graham *et al.* [G2] suggest that the dust may originate from a disk of relatively large (1–100-m diameter?) fragments, perhaps remnants of a disrupted asteroid, with boulder/boulder collisions supplying the reservoir of small fragments that produce the infrared excess. Whatever the source, the detection of neutral and ionised absorption lines of calcium and magnesium in the visible spectrum [Z3] suggests that dust is accreting onto the photosphere of the white dwarf. In summary, G 29-38 is clearly an object of considerable astrophysical interest, but it is unlikely that the system includes a substellar-mass brown dwarf.

6.4 FIRST GENERATION SURVEYS FOR BROWN DWARFS

Brown dwarf surveys in the 1960s, 1970s and even 1980s were part and parcel of the general effort to probe the boundaries of the lower main sequence. The surveys fall under two headings: searches for companions to known nearby stars; or wide-field proper-motion/photometric surveys for isolated dwarfs. These two techniques address different astrophysical issues. The large-scale surveys provide samples that can be used to study global parameters, such as the initial mass function or

Table 6.1. Benchmark ultracool dwarfs.

Star	Year	M_V	M_K	M_{bol}	Spectral type	Technique	Reference
Wolf 359	1918	16.54	9.17	12.3	M5.5	Proper motion – photographic	W4
VB 10	1944	18.65	9.95	12.95	M8	Companion – photographic	V3
VB 8	1960	18.00	9.76	12.9	M7	Companion – photographic	V4
RG 0050.5	1981	19+	10.89	13.8	M8	Photometry – photographic	R9
LHS 2924	1983	19.37	10.53	13.7	M9	Proper motion – photographic	P5
LHS 2397a	1985	18.79	9.96	13.15	M8	Proper motion – photographic	B23
LHS 2065	1987	19.15	10.29	13.5	M9	Proper motion – photographic	B23
Gl 569Bab	1987		9.44	12.6	M8.5/M9	Companion – infrared	F3, M8, L5
GD 165B	1988		11.64	14.8	L4	Companion – infrared	B25
Gl 229B	1995		15.60	17.65	T6.5	Companion – red/infrared	N2

Joint photometry listed for Gl 569Bab; the two components differ by 0.4 mag. at K (see Table 6.2), and therefore have bolometric magnitudes of ~ 13.0 and 13.4.

population kinematics; companion searches probe the properties of binary and multiple systems, and can also provide individual examples of unusual objects that can be used to tune subsequent large-scale surveys.

Table 6.1 maps out the milestones in the search for the bottom of the main sequence. VB 10, discovered in a photographic search for low-luminosity companions to known nearby stars, marked the bottom of the main sequence for almost 40 years. VB 10’s temperature was known to be $\sim 2,500$ K [G7]; (most) brown dwarfs would therefore be expected to be even cooler. Wien’s law (equation (1.29)) clearly indicates that the spectral energy distribution of such objects should peak at near-infrared wavelengths, beyond $1\text{ }\mu\text{m}$. However, the only wide-field infrared survey available during this period was the Two Micron Sky Survey, which was limited to sources brighter than 3rd magnitude at K (Section 1.6.2).

Surveys at mid-infrared wavelengths became possible with the launch of IRAS in 1983. The $12\text{ }\mu\text{m}$ observations offered the best prospect of detecting brown dwarfs, since they lie closest to the expected peak in the energy distribution. However, the survey sensitivity was limited by the short integration time and the 60-cm diameter of the telescope. As a result, even young, high-mass brown dwarfs are detectable to distances of only ~ 1 pc, while mature brown dwarfs would have to lie within 0.01 pc (2,000 AU) to meet the survey limits. With such a tiny sampling volume, the substellar mass function would have to rise more steeply than the Salpeter index before we could expect more than a handful of brown dwarf detections in the IRAS Faint Source Catalogue ($F_{12\mu} > 0.17$ Jy). Nonetheless, astronomers pored over the IRAS FSC – and failed to identify any brown dwarfs.

Given these technical limitations, photographic surveys remained the most effective means of searching for ultracool dwarfs until the advent of deep near-infrared sky surveys in the 1990s. Indeed, follow-up observations of stars from

Luyten's Half Second (LHS) survey, based on Palomar Schmidt plates, eventually turned up three ultracool dwarfs that were fainter than VB 10. The major breakthrough, however, came from a return to van Biesbroeck's techniques, and the identification of a number of extremely interesting low-luminosity companions of known nearby stars.

6.5 BROWN DWARFS REVEALED

A few key observations obtained in the late 1980s and early 1990s led to the verification of the existence of brown dwarfs. This section describes those observations. We have concentrated on the results that we believe had the most impact on this field. During the same period, a number of brown dwarf candidates were identified in very young open clusters. However, ambiguities in interpreting evolutionary models muted the impact of those observations. As described in Chapters 8 and 9, the focus is very much upon such objects in contemporary brown dwarf studies.

6.5.1 GD 165B

Searches for low-mass companions require observations that optimise the contrast between the primary and secondary in the potential binary system. This goal can be achieved either through high spatial resolution observations, which minimise the point-spread function of the primary star and therefore allow detection of a faint companion at small separation, or by making observations at wavelengths that minimise the flux difference between primary and secondary, and therefore maximise the chances of detecting the contribution from the lower luminosity component. As described in Section 8.7.2, white dwarfs are excellent targets for the latter type of survey since they have low luminosities but high effective temperatures. The presence of a cool companion therefore leads to unusual optical/IR colours, with the white dwarf dominating the combined flux distribution in the blue while the late-type companion contributes most light at long wavelengths.

GD 165 was identified as a white dwarf based on the blue ($m_{pg}m_v$) colour measured in the course of the Lowell proper motion survey. Subsequent observations by Greenstein [G5] confirmed it as type DA with a surface temperature of $\sim 13,400$ K, placing it near the hotter boundary of the ZZ Ceti instability strip. The star is also known as CX Boo and the current best estimate of the trigonometric parallax is $0''.0317 \pm 0.0025$, corresponding to a distance of 31.5 parsecs. Saffer *et al.* [S1] suggested that the white dwarf shows radial velocity variations due to a cooler white dwarf companion in a short-period orbit. However, those variations have not been confirmed [M10], and may well stem from shifts in the line core due to non-radial pulsations, as with G 29-38 [C1].

Like G 29-38 (Section 6.3.2), GD 165 was a target in Becklin and Zuckerman's IR-excess survey, and was among the first to be observed using an infrared array [B25]. Those observations revealed the presence of a faint companion at a separation

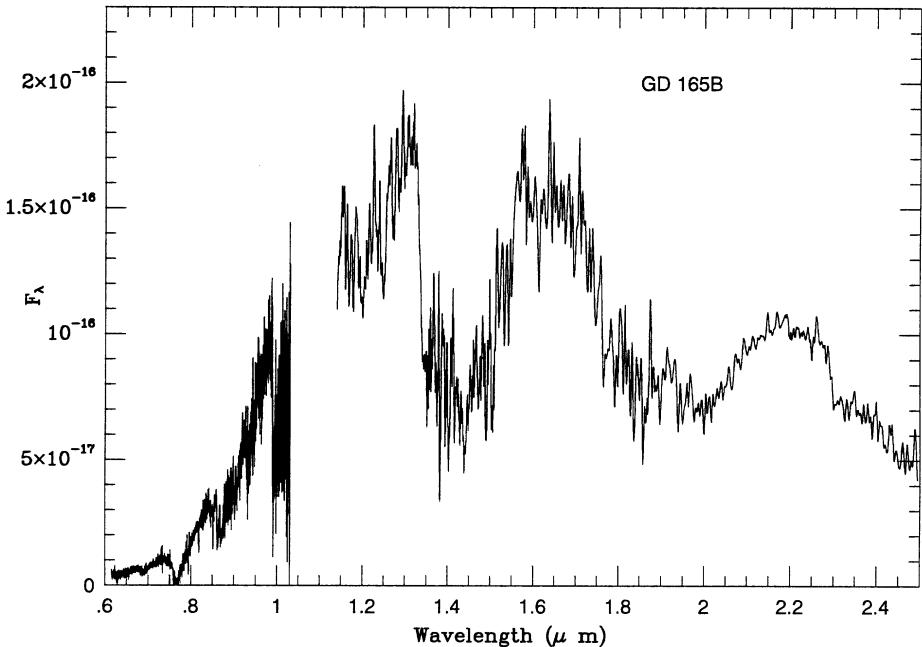


Figure 6.3. The optical and infrared spectrum of GD 165B, the first known L dwarf. Optical data are from [K4]; infrared data from [J1] (courtesy of the *Astrophysical Journal*).

of 3.5 arcseconds (110 AU). The companion has an unusually red ($J-K \sim 1.6$), suggesting a temperature well below 2,500 K. Subsequent astrometry confirms that GD 165B shares the proper motion of the white dwarf. There is therefore no question that it is a physical companion, as opposed to a foreground or background star or galaxy [Z2].

Based on the known distance to GD 165, the companion has $M_K = 11.66$ and $M_{bol} \sim 14.6$, or $L \sim 10^{-4} L_\odot$. This places GD 165B very close to the substellar regime. Black-body fitting to the flux distribution suggested a temperature close to 1,800 K. Infrared spectroscopy shows that the object has extremely strong H_2O absorption bands [J1], but is not qualitatively different from late-type M dwarfs (Figure 6.3). The first optical spectrum [K1] revealed markedly different spectral properties than are observed in late-type M dwarfs, such as LHS 2924 (spectral type M9). Instead of strong TiO and VO molecular absorption, the spectrum appears relatively smooth. At the time GD 165B was regarded as something of an individual oddity. The proximity of GD 165A made observations difficult, and potentially susceptible to errors in background subtraction, while it also seemed possible that atmospheric ‘pollution’ of GD 165B by mass loss from GD 165A might have occurred during the asymptotic giant branch phase.

It is now clear that GD 165B was the first L dwarf to be discovered (see Section 2.2.3). High signal-to-noise spectra show broad potassium absorption and metal hydride bands in the far red, characteristics of spectral type L4 [K4]. The

observational properties of L dwarfs are discussed further in Section 6.7. For the present, we note that the best-fit temperature derived for GD 165B is $1,900 \pm 100$ K. The mass depends on the inferred age of the system, which is the sum of the cooling time of the white dwarf and the main-sequence/giant branch lifetimes of the GD 165A progenitor. The latter depends on the mass of the progenitor, and that can be estimated from the present-day mass of the white dwarf, measured by fitting the Balmer line profiles to stellar models and determining the gravity [B5]. The initial/final mass relations derived by Weidemann [W1] and others indicate that low-mass white dwarfs are descended from low-mass progenitors. Current estimates for GD 165A indicate a mass of $0.56\text{--}0.65 M_{\odot}$, suggesting a progenitor mass of $1.2\text{--}3 M_{\odot}$, and main-sequence lifetimes from 4.6–0.4 Gyr. Adding the white dwarf cooling time of ~ 0.6 Gyr gives a likely age between 1 and 5 Gyr for GD 165B. This is consistent with a mass between 0.065 and $0.075 M_{\odot}$, indicating that GD 165B is probably a brown dwarf.

As an aside, both G 29-38 and GD 165B were discovered in the earliest stages of the Becklin–Zuckerman white dwarf survey. The final survey includes over 200 white dwarfs [Z2]. Those observations reveal ~ 20 previously unknown late-type stellar companions, but no further brown dwarfs (or dusty disks). The moral: only compute statistics when the complete sample has been observed.

6.5.2 HD 114762B

GD 165B was detected through direct imaging. An alternative technique, better suited to identifying close companions, is searching for reflex orbital motion in the primary star (Section 2.7.1). In most cases, one can only set a lower limit to the mass of the companion, since the orbital inclination, i , is usually unknown. However, under certain circumstances the mass can be constrained within reasonable limits.

HD 114762 is a 7th magnitude F9 dwarf ($V = 7.36$, $(B-V) = 0.56$) lying at relatively high Galactic latitude. The Hipparcos parallax ($0.^{\circ}0247 \pm 0.0014$) places it at a distance of 40.5 ± 2.5 parsecs. It is chromospherically inactive, slightly metal-poor ($[\text{Fe}/\text{H}] = -0.67$) and has a relatively low gravity, $\log g = 4.17$ [G4], suggesting that it has evolved slightly off the main sequence. Age estimates range from 5–10 Gyr [H2], [H5]. Recent high spatial resolution observations have revealed a faint stellar companion, HD 114762C [P3]. The companion is 7.3 magnitudes fainter than HD 114762A, corresponding to $M_K = 10.1$ and (probably) spectral type M8/M9, and lies at a separation of 3.3 arcseconds (135 AU).

HD 114762 was originally selected for extensive radial velocity measurements as a possible addition to an improved set of radial velocity standards. Those observations, obtained over a 10-year span, revealed unambiguous velocity variations of $\pm 0.6 \text{ km s}^{-1}$ with a period of 84 days [L1], [C2]. The non-symmetric nature of the velocity curve shows that the orbit has a significant eccentricity, $e = 0.38$ (Figure 6.4). The spectral type of HD 114762A indicates a mass of $1.17 M_{\odot}$, giving a projected semi-major axis of $a \sin(i) = 0.3$ AU and a companion mass of $M_2 \sin(i) = 0.011 M_{\odot}$. Clearly, the crucial parameter for determining whether this system harbours a brown

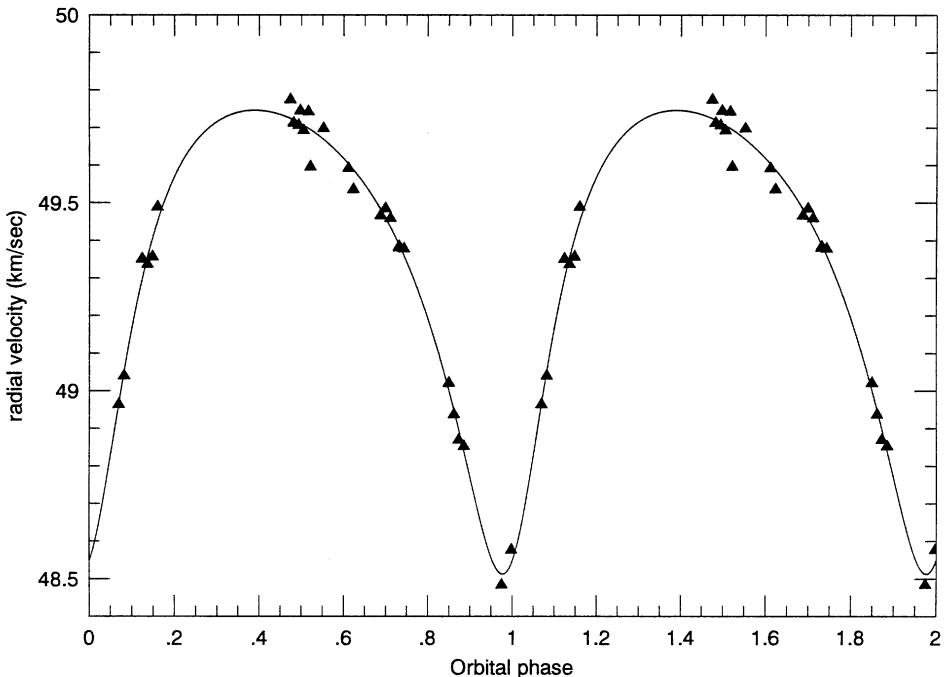


Figure 6.4. The radial velocity curve measured for HD 114762. The observations are taken from [C2] and have an accuracy of 34 m s^{-1} .

dwarf, or even a high-mass planet, is the orbital inclination: the companion has a mass exceeding the hydrogen-burning limit only if $i < 8.5^\circ$. Photometric monitoring shows no evidence of eclipses [R5], [H5], setting an upper limit of $i < 89^\circ$ for a Jupiter-sized companion. Stronger constraints can be set by using the measured line profiles to estimate the projected rotational velocity of the primary, $v \sin(i_{eq})$, and comparing that value with the average for similar F stars. This assumes, of course, that the companion has an orbit that is coplanar with the stellar equatorial plane. This is probably reasonable for either a planetary companion, which would be expected to form in a disk, or a brown dwarf, where the net angular momentum of the parent molecular cloud core might align the rotational and orbital axes in a system with such a small semi-major axis [H2]. The rotational velocity, derived from analyses of high-resolution spectra, is only 0.8 km s^{-1} [C2], [H2]. The intrinsic velocity depends to a limited extent on the age, ranging from 3.0 km s^{-1} at 5 Gyr to 2.3 km s^{-1} at 8 Gyr. The corresponding inclinations are $15\text{--}20$ degrees, with an uncertainty of at least ± 10 degrees.

Is HD 114762B a brown dwarf? The velocity data alone do not permit an unambiguous answer. If the orbit of HD 114762B is coplanar with the rotational plane, then the most likely value for its mass is $0.04\text{--}0.05 M_\odot$, although the uncertainties permit a mass as high as $0.08 M_\odot$ and as low as $\sim 0.025 M_\odot$. At the most likely inclination, the semi-major axis corresponds to an angular separation of

9 milliarcseconds. Thus, interferometric observations, either resolving the companion or, more likely, the reflex orbital motion of the primary, may eventually settle the issue of whether HD 114762B is a star, a brown dwarf or a planet.

6.5.3 Gl 229B

The identification of Gl 229B marks a pivotal point in brown dwarf research. Like GD 165B, this low-luminosity companion was discovered through direct imaging, in this case using a coronagraph to occult most of the radiation from the primary star. The characteristics of Gl 229B place it unequivocally below the hydrogen-burning limit and render it the prototype for spectral class T.

Gl 229 is an early-type M-dwarf, $M_V = 9.4$, lying at a distance of only 5.7 parsecs (see Appendix). As one of the nearest M dwarfs, the star was included in Henry's near-infrared speckle survey of stars within 8 parsecs of the Sun [H6]. Those observations showed no evidence for any companions within 20 AU (± 2 arcseconds). However, Gl 229 was also observed by T. Nakajima and collaborators, using the Johns Hopkins coronagraph on the Palomar 60-inch [N1]. This system uses an apodising mask and occulting disk to obscure the central 3 arcseconds and cut down scattered light, while tip–tilt correction (using the reflected image of the primary) provides image sharpening over the full 60×60 arcsecond field of view.

I-band images of Gl 229, obtained in October, 1994 revealed a potential faint companion (Figure 6.5), 13 magnitudes fainter ($M_I = 20.3$) and at 7.8 arcseconds separation [N2]. Subsequent measurements in October, 1995 confirmed that the companion shares the proper motion of the primary, while *K*-band observations on the Palomar Hale 200-inch showed that the fainter star has extremely red optical/infrared colours, $(I-K) = 6.2$ mag. [M9]. Integrating the broadband photometry gives a luminosity $L \sim 6 \times 10^{-6} L_\odot$, placing the object firmly in the brown dwarf regime.

Initially, the most surprising aspect of these observations was the infrared colours, $(J-H) \sim 0.0$, $(J-K) \sim -0.1$, strikingly different from late-type M dwarfs and GD 165B. Low-resolution spectroscopy [O2] supplied the answer to this puzzle, showing that the 1–2.5- μm spectrum exhibits strong molecular bands that cut into the longer wavelength regions of both the *H* and *K* atmospheric windows (Figure 6.6). As described in Section 2.2.4, those bands are due to methane, and Gl 229B is the prototype T dwarf. The prediction that cool brown dwarfs would have methane absorption dates back to the early 1960s ([T11]), but most observational astronomers had either forgotten, or never encountered, that prediction (although the spectra of Jupiter and Saturn should have given a clue).

Detailed comparison of the spectral energy distribution of Gl 229B with theoretical models leads to a temperature estimate of 960 ± 70 K [M3]; this is far too cool for a hydrogen-burning star and, like the luminosity, clearly identifies Gl 229B as a brown dwarf. As with GD 165B, the mass depends on the age, which is best estimated using the observed characteristics of the primary. Gl 229A is inactive chromospherically (no $\text{H}\alpha$ emission) and coronally (weak X-ray flux), and is



Figure 6.5. Ground-based and HST images of Gl 229B.

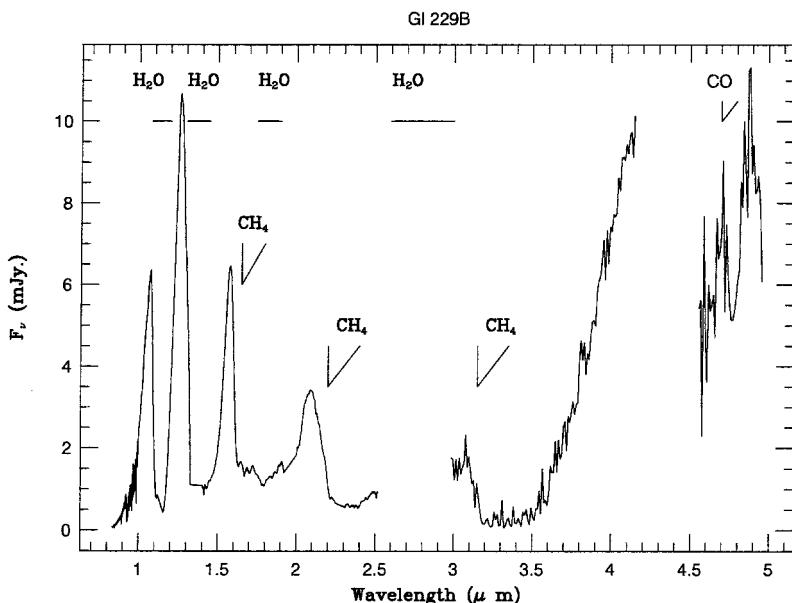


Figure 6.6. The spectral energy distribution of Gl 229B; the terrestrial atmosphere is opaque (H_2O absorption) between 2.52 and $3 \mu\text{m}$ and between 4.15 and $4.5 \mu\text{m}$ (from [O3], courtesy of B. Oppenheimer).

therefore at least 1 Gyr old, and may be as old as 5 Gyr. Thus, Gl 229B probably has a mass between 30 and 60 Jovian masses.

Spectroscopy of Gl 229B revealed some unexpected features. CO is detected at $4.7\text{ }\mu\text{m}$ (the (1–0) band [N4]), indicating abundances $\sim 1,600$ times higher than theoretical predictions [F1]. Similarly, atomic caesium is present in the optical spectrum even though it should be locked up in molecular CsCl. This might reflect mixing within the atmosphere, with material from deeper, hotter layers dredged up to the photosphere [O3]. Alternatively, the atmospheric structure could be extremely non-grey, allowing one to see to different depths, and different temperatures, at different wavelengths. Under those circumstances, the concept of a single effective temperature may not be valid. Brown dwarf atmospheres are considered in more detail in Section 6.7.3.

6.5.4 PPI 15

Surveys for brown dwarf members of open clusters have the advantage that age, abundance and distance are well determined (from observations of higher mass cluster members), so observational programs can focus on specific luminosities and colours. Moreover, young clusters offer the best prospect for detecting low-mass brown dwarfs, since substellar objects are hotter and brighter at younger ages. The closest cluster, the Hyades, has been surveyed thoroughly [L2], but with no success (Section 8.9.1). Given its proximity and age (~ 125 Myr), the Pleiades has also been a prime target of such surveys, which led to the first identification of an isolated brown dwarf.

The first Pleiades survey that reached sufficiently faint magnitudes to detect potential brown dwarf members was by Stauffer *et al.* [S4], [S5], using a CCD camera on the Hale 200-inch telescope. With limiting magnitudes of $V \sim 23$ and $I \sim 21$ over 0.4 square degrees in the central regions of the cluster, this survey identified 18 sources (the PPI sources) with colours and magnitudes that lie on an extrapolation of the Pleiades main sequence. Six of those sources lie close to the theoretical hydrogen-burning limit. Most are confirmed as late-type dwarfs ($> M_5$) by follow-up spectroscopy, with radial velocities consistent with cluster membership [H3], [S6].

The faintest of the PPI objects, PPI 15 ($I = 17.8$, $(V-I) = 4.6$), proves to be not only substellar in mass, but also a close binary system. An initial set of spectra obtained by Basri *et al.* [B3], using the HIRES spectrograph on the Keck I telescope, showed that Li I $6,708\text{ \AA}$ was present at moderate strength (equivalent width 0.5 \AA). This is consistent with lithium being depleted to 1% of the primordial value, a substantial depletion, but still consistent with a mass just below the hydrogen burning limit. Further observations revealed that PPI 15 is a double-lined spectroscopic binary, with a period of 6 days and an orbital eccentricity of 0.5. The two components are estimated to have masses of 0.065 and $0.059\text{ }M_\odot$ [B4].

As discussed in Section 8.9.2, follow-up observations of fainter candidates, notably by the Tenerife group (Rebolo, Martin, Zapatero-Osorio and collaborators),

have led to the identification of more Pleiades brown dwarfs at lower luminosities. Those objects have been used to estimate the underlying mass function, as described in Section 9.6.2. Moreover, the location of the lithium-depletion boundary in the cluster colour–magnitude diagram, between dwarfs that have suffered no lithium depletion, and objects with partial depletion, supplies an independent new method of estimating the age of the cluster.

6.5.5 Field brown dwarfs

A number of isolated brown dwarfs in the general field were also discovered in the early 1990s, prior to the large-scale infrared surveys described in the following section. The first field M dwarf with lithium absorption was discovered by Thackrah *et al.* [T3] through spectroscopic follow-up observations of a colour-selected photographic sample. Known as 296A (it lies within UK Schmidt survey field 296), the colours are relatively unremarkable, $(I-K) \sim 2.8$ at spectral type M6, but Li I 6,708 Å absorption is present with an equivalent width of 0.5 Å. Assuming an effective temperature of $\sim 2,700$ K, the presence of lithium implies an age between 20 and 200 Myr. Comparing the observed properties against theoretical models (e.g., [B1]), the M dwarf has a probable distance of ~ 50 parsecs and a mass in the range 0.05–0.095 M_{\odot} . Despite its youth, 296A is not associated with any obvious young cluster or association. It remains unclear whether it formed in isolation, or whether it was ejected from a cluster through dynamical interactions. Recent years have seen the detection of ~ 100 similarly isolated young stars and brown dwarfs; most appear to be members of low-density stellar associations lying within 100 parsecs of the Sun (see Section 6.8.2, Section 7.6 and [Z4]).

Brown dwarfs have also been identified serendipitously from proper motion surveys. The first isolated L dwarf, Kelu 1, was discovered in the course of a photographic survey aimed at finding low-luminosity white dwarfs [R6]. Rather than the featureless spectrum of a cool DC white dwarf, follow-up observations showed that Kelu 1 had a spectrum similar to GD 165B, with broad Na D lines, weak H α emission and Li I 6,708 Å, together with metal hydride bands in the far red. The detection of lithium confirms Kelu 1 as a brown dwarf with a mass less than 0.06 M_{\odot} . Trigonometric parallax measurements indicate a distance of 16 parsecs, implying $M_I \sim 15.8$ and $M_K \sim 10.8$, and Kelu 1 now serves as the primary standard for spectral class L2 (see Figure 2.4).

Even Luyten's proper motion catalogues are providing examples of brown dwarfs. LP944-20 is one of many M dwarfs in the New Luyten Two Tents (NLTT) catalogue. Re-discovered independently by Irwin, McMahon and Hazard in their photometric survey for high-redshift QSO candidates, and re-named BRI 0337-3535, low-resolution spectroscopy indicates a spectral type of M9.5 [K3]. The trigonometric parallax is 201 milliarcseconds, giving a distance of only 5 parsecs [T5], $M_K = 11.10$, and $M_{bol} = 14.3$ – significantly fainter than LHS 2065, LHS 2924 and LHS 2397a. There is another crucial difference: LP 944-20 has lithium absorption, equivalent width 0.53 Å [T6]. Like PPI 15, the abundance is depleted to $< 1\%$ of

the primordial value, leading to an age estimate of 0.5–1 Gyr, but still implying a mass between 0.06 and $0.07 M_{\odot}$, well below the hydrogen-burning limit.

6.6 SECOND GENERATION SURVEYS FOR BROWN DWARFS

The development of array detectors has permitted deep, wide-field surveys at near-infrared wavelengths. These arrays offer two advantages: first, larger areal coverage in a single exposure, with a higher spatial resolution than is possible with aperture photometry; and second, simultaneous measurement of the sky background, and therefore more accurate sky subtraction. At optical wavelengths, arrays of large-format CCDs provide both wide-angle coverage and 10 times more sensitivity than the photographic plates used for the Palomar and UK Schmidt sky surveys.

The DENIS and 2MASS projects (Section 1.6.2) have taken advantage of the new infrared technology to survey the sky at flux levels more than 10,000 times fainter than those achieved by the TMSS, extending near-infrared source catalogues from 5,600 to over 500 million objects. At the same time, the Sloan Digital Sky Survey (SDSS – Section 1.6.3) is imaging one-quarter of the sky to 24th magnitude at optical wavelengths. Together, these three surveys have revolutionised our understanding of L and T dwarfs.

6.6.1 L dwarfs in the field

DENIS provided the first intentional discoveries of L dwarfs in the general field. IJK photometry covering ~ 230 square degrees was analysed in a ‘Brown Dwarf Mini-survey’ as a preliminary search for unusually cool objects [D2]. None of the sources were as red as Gl 229B, but three objects have colours similar to those of GD 165B. All three have been confirmed as L dwarfs [T15], and one, DENIS 1228-1547, has strong (2.3 Å equivalent width) lithium absorption. None show evidence of strong methane absorption at near-infrared wavelengths [D2], [T10], although recent high signal-to-noise spectroscopy suggests that the coolest, DENIS 0205-1159, may have very weak absorption in the K band [N6]. Two of the three have been resolved as binaries, not unexpected given their bright apparent magnitudes (see Section 6.8.2 for further discussion). Under the classification system illustrated in Figure 2.4, their spectral types are L3, L5 and L7.

The overwhelming majority of currently known field L dwarfs have been identified from 2MASS data. This partly reflects the higher sensitivity of that survey, and partly a focused science program, carried on in tandem with the actual survey. In addition, the 2MASS team had access to the Keck 10-m telescopes and associated instrumentation, notably the Low Resolution Imaging Spectrograph (LRIS, [O1]) and the HIRES echelle spectrograph [V1]. That instrumentation provided the high signal-to-noise observations essential to characterising the optically faint sources identified by this program.

The initial set of 2-mass candidates was selected by combining optical and near-infrared data, where the optical data are scans of the POSS I and UK Schmidt

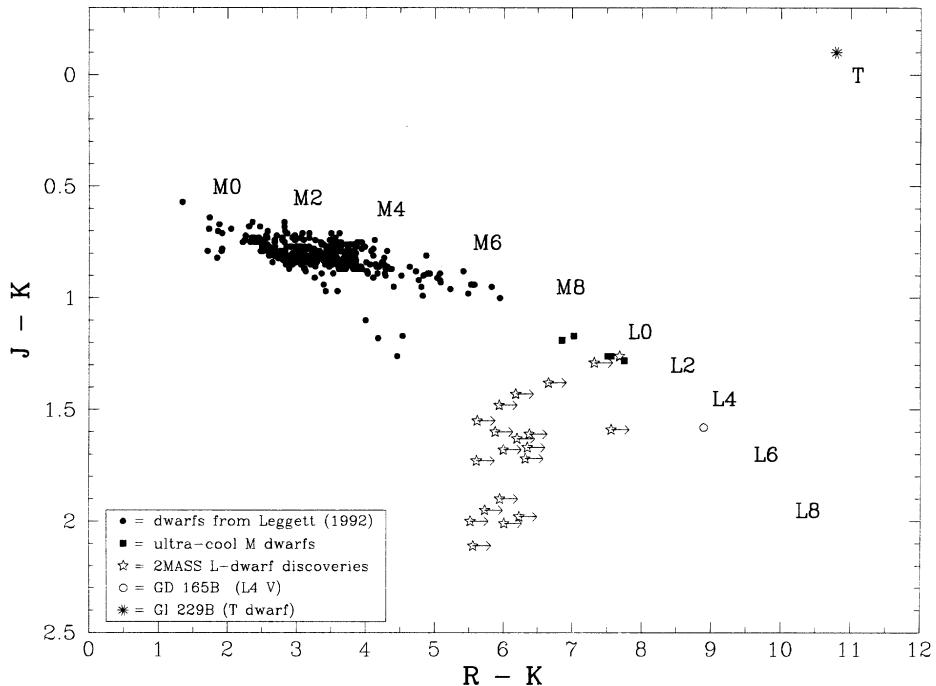


Figure 6.7. Selection criteria for the initial 2MASS L dwarf sample; sources with red ($J-K$) colours and non-detection on red POSS I plates were targeted for follow-up spectroscopy (from [K6], courtesy of the *Astrophysical Journal*).

surveys, compiled by the US Naval observatory, Flagstaff as the USNOA catalogue [M4]. Three criteria were adopted: $(J-K_S) > 1.3$; $K_S < 14.5$; and no optical counterpart visible on the red-sensitive POSS I E plates (Figure 6.7). The last ensures an $(R-K_S)$ colour redder than 5.5 magnitudes, corresponding to spectral types later than M6 (i.e., ultracool dwarfs). The first analysis covered 370 square degrees, larger than the DENIS Mini-survey, and extending to fainter apparent magnitudes. As a result, this investigation identified over 30 late-type M dwarfs and 20 L dwarfs [K6], spanning the full spectral range. Indeed, the latter objects laid the basis for the spectral sequence illustrated in Figure 2.4.

Combining the 2MASS sample with previous discoveries provided sufficient objects to map out the gross photometric and spectroscopic properties of L dwarfs. As a result, subsequent surveys were able to tune their search criteria to identify field L dwarfs in a more efficient manner. Thus, the next 2MASS-based survey, focused on late-type L dwarfs, turned up 67 new objects [K9], while the NStars census ([C3]; see Section 8.6.2) includes nearly 300 L dwarfs. As of October, 2004, over 400 L dwarfs have been confirmed spectroscopically. Approximately one-third of the sample has detectable lithium absorption, indicating masses

below $0.06 M_{\odot}$: these are unequivocally substellar-mass brown dwarfs, rather than very low-mass stars.

6.6.2 T dwarfs in the field

Isolated T dwarfs proved more difficult to find than L dwarfs – at least with 2MASS. While late-type L dwarfs have extremely red colours at all wavelengths, the onset of methane absorption, at temperatures below $\sim 1,300$ K, eats into the flux emitted at both H and K passbands (see Figure 2.10). As a result, T dwarfs evolve to bluer colours in the near infrared, and become indistinguishable from F, G and K dwarfs at those wavelengths (see Figure 2.17). T dwarfs also have extremely red optical-infrared colours (e.g., $(I-J) > 4.5$), and all save the nearest T dwarfs are too faint to be detected on all-sky photographic surveys. Thus, the only recourse for researchers using a purely infrared survey, like 2MASS, is to search for sources that have neutral/blue JHK colours and are also invisible on the POSS/UKST photographic sky surveys (see Figure 6.7). Unfortunately, genuine T dwarfs are outnumbered heavily by imposters. In particular, the ‘no optical counterpart’ method turns out to be excellent at finding asteroids, which have G-dwarf-like infrared colours and were not at the 2MASS position at the epochs of the photographic surveys.

Somewhat perversely, the optical Sloan Digital Sky Survey provides better opportunities for identifying T dwarfs than 2MASS; analysis of SDSS data led to the discovery of the first (SDSS 1624 + 0029, [S7]) and second (SDSS 1346 – 0031, [T13]) examples of isolated T dwarfs. Even though the SDSS passbands lie short of the peak in the flux distribution, they sample an ever-steepening part of the spectrum. The $(i-z)$ colour, in particular, increases monotonically with later spectral types, and T dwarfs have particularly distinctive $(i-z)$ colours. Moreover, SDSS has a sufficiently faint limiting magnitude that, in those passbands, it can detect T dwarfs at distances of ~ 30 parsecs. As an example, the T6 dwarf SDSS 1624 + 0029 has $i \sim 23.8$, $(i-z) \sim 3.5$; in comparison, late-type M dwarfs, such as LHS 2924 (M9), have $(i-z) \sim 1.5$, while late-type L dwarfs have $(i-z) \sim 2.5$. Indeed, T dwarfs are probably the reddest objects, either Galactic or extra-Galactic, at these wavelengths.

2MASS (or, rather, Adam Burgasser) started turning up T dwarfs shortly after the SDSS discoveries. Burgasser painstakingly identified and eliminated hundreds of candidate T dwarfs (neutral JHK colours, no optical data) before tracking down several bona fide examples [B9]. These initial discoveries include Gl 570D, a wide companion of a known nearby system (spectral types K3, M1.5, M3) at $d \sim 5.2$ parsecs, ~ 1 magnitude fainter and 200 K cooler than Gl 229B [B10]. As with Gl 229B, the low-chromospheric activity of the earlier type stars in the system points to a relatively old age of more than 2 Gyr, implying a likely mass of 50 Jupiter masses for Gl 570D.

Both the Burgasser-led 2MASS work and the SDSS T dwarf search are still underway at the time of writing. The 2MASS study is aimed mainly at mid- to late-type T dwarfs, T3–T7, with near-infrared selection criteria of $(J-H) < 0.3$ and $(H-K) < 0.0$ [B11]. There are simply too many spurious candidates at the redder

JHK colours that correspond to earlier type T dwarfs. Fortunately, this is not a problem faced by SDSS. As noted above, $(i-z)$ increases monotonically from M through L to T. Thus, SDSS was able to identify the first early-type T dwarfs [L7], three objects with unsaturated near-infrared methane absorption bands (two examples, T0 and T2, are shown in Figure 2.10). These objects are particularly important in understanding the temperature range spanned by the L/T transition, as discussed further in the following sections.

Finally, one of the most noteworthy discoveries of recent years belongs to neither 2MASS nor SDSS. The nearest known brown dwarfs, ε Indi Bab, were identified originally as an extremely red point source on UK Schmidt R- and I-band plate material [S8] (Figure 6.8). Further inspection showed that, not only does the object have significant proper motion, but that motion is shared by the known nearby star, ε Indi (or Gl 845), which lies only 3.6 parsecs from the Sun. ε Indi B is \sim 7 magnitudes fainter than the primary at K , with $M_K \sim 13.4$ and $(J-K) = 1.0$; those colours and magnitudes are consistent with an early-type T dwarf. In fact, high-resolution imaging shows that the companion is a T1 + T6 binary, with 0.73- arcsecond separation (2.65 AU) [M5], offering the potential for dynamical mass determination in the not too distant future.

The current T dwarf census includes 68 objects – approximately 20% the numbers in the current L dwarf census. Only four are companions of hydrogen-burning stars (Gl 229B, Gl 670D and ε Indi Ba and Bb), and only 14 (including ε Indi Ba) are early-type dwarfs, with spectral types T0–T2. The relatively small size of this sample underlines the fact that T dwarfs have very low luminosities, and are therefore detectable only within the immediate Solar Neighbourhood. We will consider the implications of these results for the substellar mass function in Chapter 9.

6.7 CHARACTERISTICS OF L AND T DWARFS

The most prominent characteristics of the optical spectra of L and T dwarfs are described in Sections 2.2.3 and 2.2.4, respectively, while Section 2.2.6 gives an overview of the main properties at near-infrared wavelengths. The variations in the relative strengths of the stronger features, namely molecular absorption by TiO, VO, CaH, FeH and CrH at optical wavelengths and FeH, H₂O, CO and CH₄ at infrared wavelengths, provide the basis for the spectral classification schemes outlined by Kirkpatrick *et al.* [K5] for L dwarfs and by Burgasser *et al.* [B12] and Geballe *et al.* [G7] for T dwarfs. This section gives more detailed consideration to the consequent effects on the colour–magnitude diagram, and outlines the physical processes that are likely to underlie the changes in the spectral energy distribution.

6.7.1 Photometric properties and the H–R diagram

Section 2.3 gives a broad overview of the photometric properties of low-mass stars and brown dwarfs. Figures 2.16 and 2.17, in particular, show how broadband

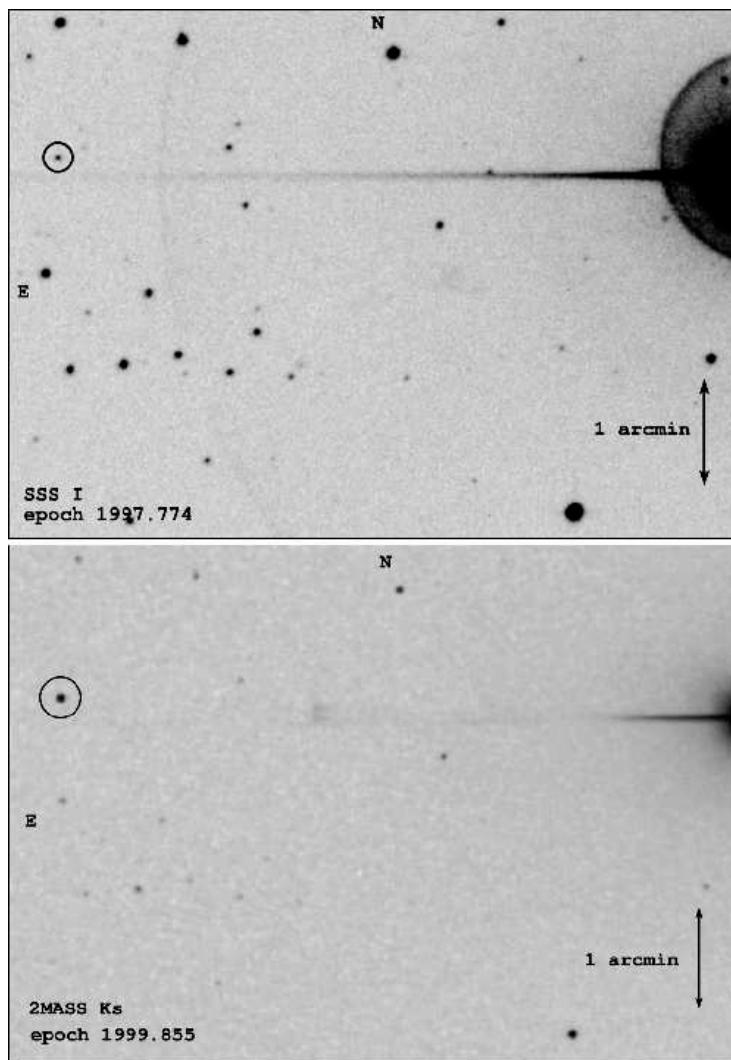


Figure 6.8. ε Indi Bab, the binary T dwarf companion to the nearby K5 dwarf, ε Indi A. The separation between the two components is 7 arcminutes (courtesy of R. Scholz and *Astronomy & Astrophysics*).

colours change with decreasing absolute magnitude through the L and T regime. Those diagrams only include systems with accurate distance measurements. Although trigonometric parallax data are available for over 50 systems ([D1], [V2]), those objects still represent a small subset of known L and T dwarfs.

An alternative way of tracking changes is to consider the variation with spectral type, since almost all known L and T dwarfs have types that are tied to the standard

systems [K5], [G7], [B12], and spectral type serves as a reasonable surrogate for effective temperature. Hawley *et al.* [H7] have averaged results for M, L and T dwarfs with SDSS and 2MASS photometry, and Figure 6.9 plots the mean colour as a function of spectral type for several combinations of these passbands. Among the more notable features are the reversal in $(r-i)$ at $\sim M9$ (due to weakening TiO absorption) and the turnaround in $(z-K)$ and $(J-K)$ at $\sim T0$ (due to the increasing strength of methane bands in the K passband). The $(i-z)$, $(i-J)$ and $(z-J)$ sequences are all close to monotonic with spectral type. Of these three, $(z-J)$ generally offers the best prospects for extending ground-based surveys, since T dwarfs are brightest at J and, allowing for the extremely red colours, the z passband is the most sensitive optical passband for ultracool dwarfs.

The onset of methane absorption has a significant effect on the luminosities of cool dwarfs at near-infrared wavelengths. The impact at H and K is obvious, but the J passband is also affected, in the opposite sense. Figure 6.10 shows the run of M_J and M_K with spectral type. With the acquisition of more parallax measurements, it became clear that the mid-type T dwarfs (T3–T5) are actually brighter in M_J than late-type L and early-type T dwarfs (L6–T2). This is at least partly due to the presence of higher opacities at longer and shorter wavelengths in T dwarfs, which force a higher proportion of the total flux into the relatively transparent 1.2- μm window. Note, however, that the effect is also present, to a lesser extent, in the K passband, which includes strong CH_4 absorption. We consider this behaviour further in the discussion of the L/T transition in Section 6.7.3.

Many of the brighter L and T dwarfs have ground-based observations in the L and, in a few cases, M passbands (3 and 5 μm). Figure 6.11 shows the colours as a function of spectral type. The $(K-L)$ colour increases monotonically with increasing spectral type, with significant dispersion at later types, as is the case with $(J-K)$. The $(L-M)$ colour becomes increasingly negative from $M0$ to $L5$, but that trend reverses at later types. The negative colours imply that there is more flux at 5 μm than predicted by a black-body extrapolation from 3 μm . Spitzer is providing more extensive mid-infrared observations of a much larger sample of ultracool dwarfs, generally confirming these results.

6.7.2 Bolometric corrections and the temperature scale

Section 2.3 emphasised the importance of observations at infrared wavelengths in mapping out the spectral energy distribution of cool dwarfs. Mid-infrared data are particularly important for ultracool dwarfs, since the cooler temperatures and the more diverse range of opacity sources leads to significant deviations from the simple Rayleigh–Jeans flux distribution at these wavelengths (Figure 6.11). Recent high-quality L and M photometry [G8], coupled with flux-calibrated near-infrared spectra, enable reliable measurement of bolometric magnitudes for a large sample of M, L and T dwarfs (Figure 6.12(a)), and those results, in turn, allow a closer examination of the likely temperature range spanned by spectral classes L and T.

Given a measurement of bolometric luminosity, the effective temperature can be derived using Equation (1.28), provided that one has an estimate of the radius.

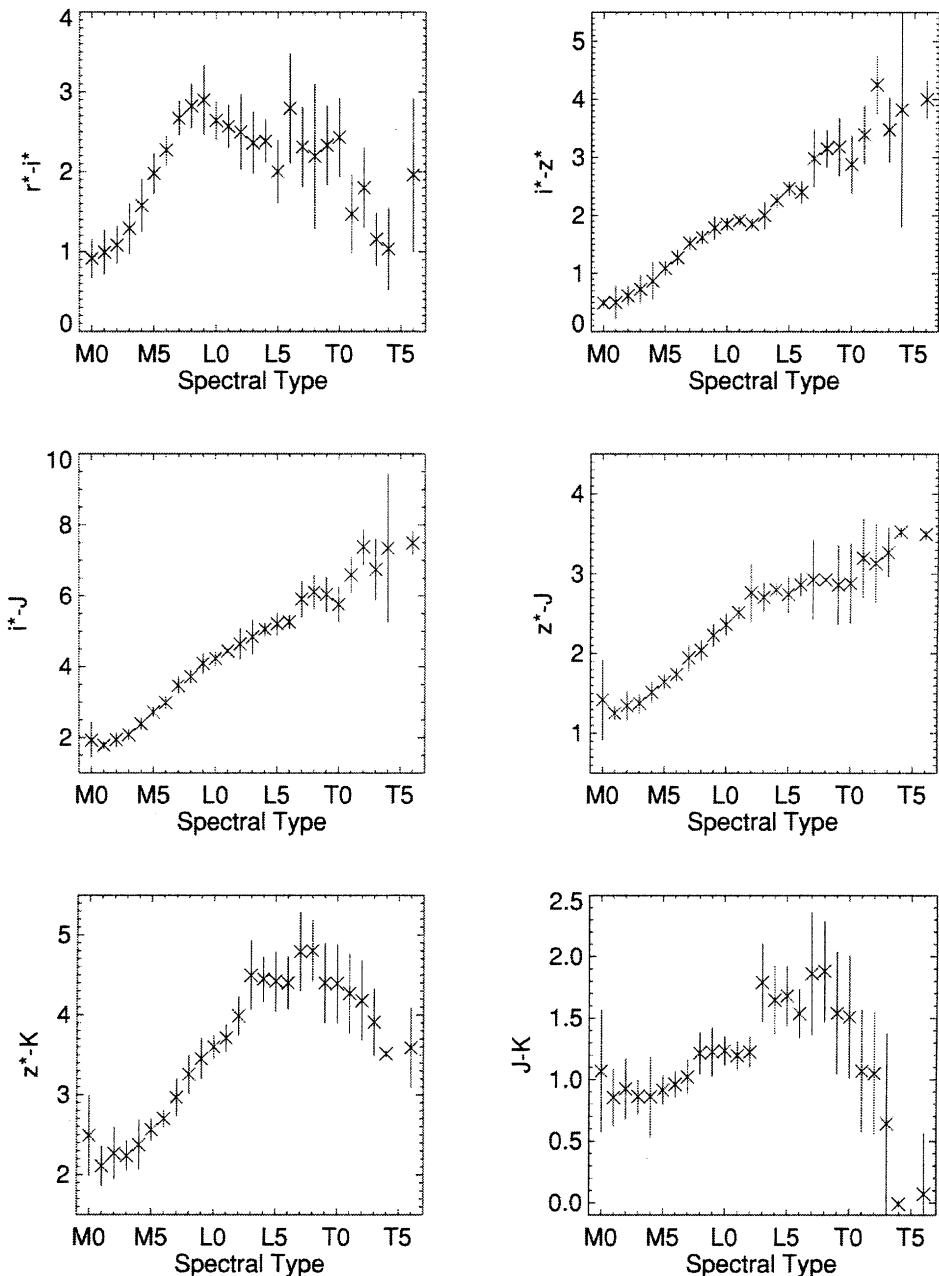


Figure 6.9. Far-red and near-infrared colours for late-type dwarfs as a function of spectral type (from [H7], courtesy of the *Astronomical Journal*). Note that the optical colours are on the SDSS/Gunn system (Table 1.1) and represents preliminary SDSS processing, as indicated by the * (see [H7]).

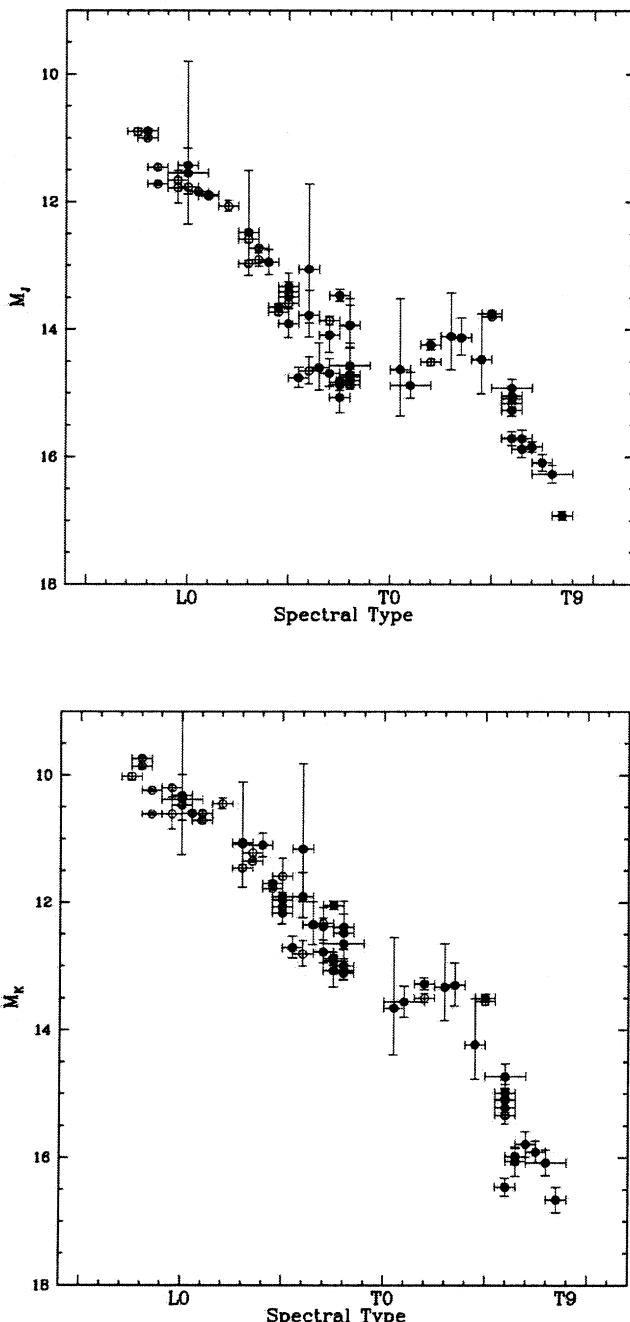


Figure 6.10. M_J and M_K as a function of spectral type for L and T dwarfs with accurate trigonometric parallax measurements (from [V2], courtesy of the *Astronomical Journal*).

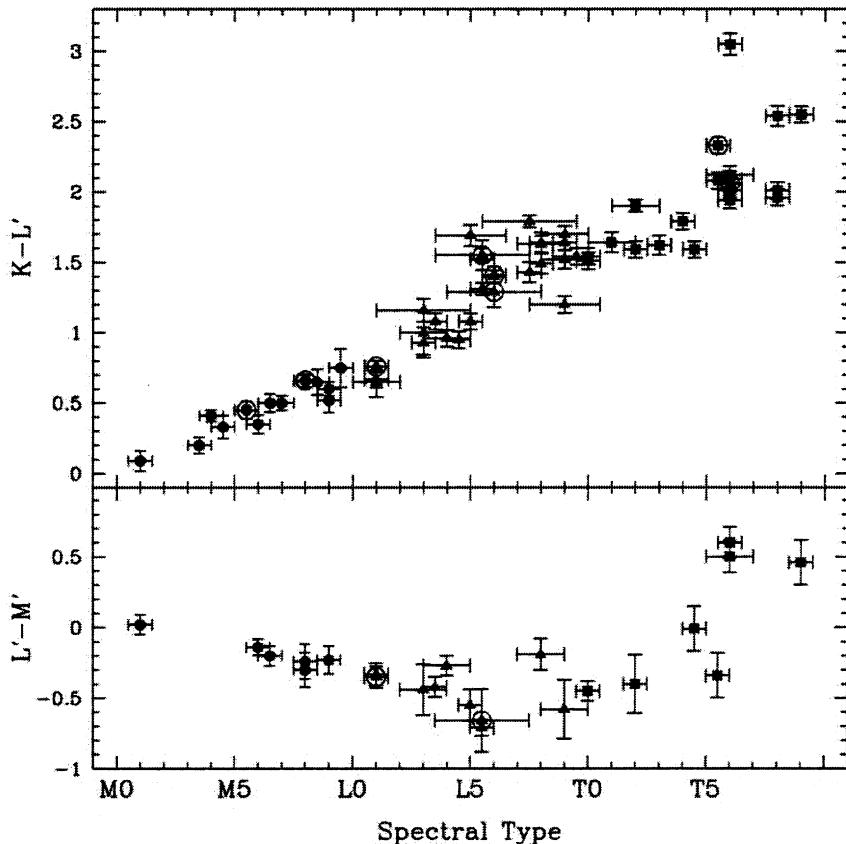


Figure 6.11. 3–5 μm colours of M, L and T dwarfs (from [G8], courtesy of the *Astronomical Journal*).

Lacking direct measurements, the radius is usually taken from evolutionary models (e.g., [B1], [B7]). For brown dwarfs, this demands an age estimate – at a given luminosity, young, low-mass brown dwarfs have larger radii than older, more massive objects. Since all of the calibrating sources are isolated brown dwarfs in the immediate Solar Neighbourhood, most studies adopt ages of 3–5 Gyr, corresponding to the average age of the local stellar population. Fortunately, the brown dwarf radius is not a very strong function of mass, varying by at most 30% (see Section 3.5.3). This corresponds to an upper limit of ~ 300 K in the uncertainty of the temperature estimates.

Figure 6.12 shows the results from the most recent study that uses M_{bol} to derive temperature estimates. The formal best-fit polynomial relation (from [G8]) is:

$$\begin{aligned} T_{eff} = & 14,322 - 5,128.7 \times SpT + 909.51 \times SpT^2 - 83.099 \times SpT^3 \\ & + 4.0323 \times SpT^4 - 9.8598 \times 10^{-2} \times SpT^5 + 9.5373 \times 10^{-4} \times SpT^6 \end{aligned} \quad (6.2)$$

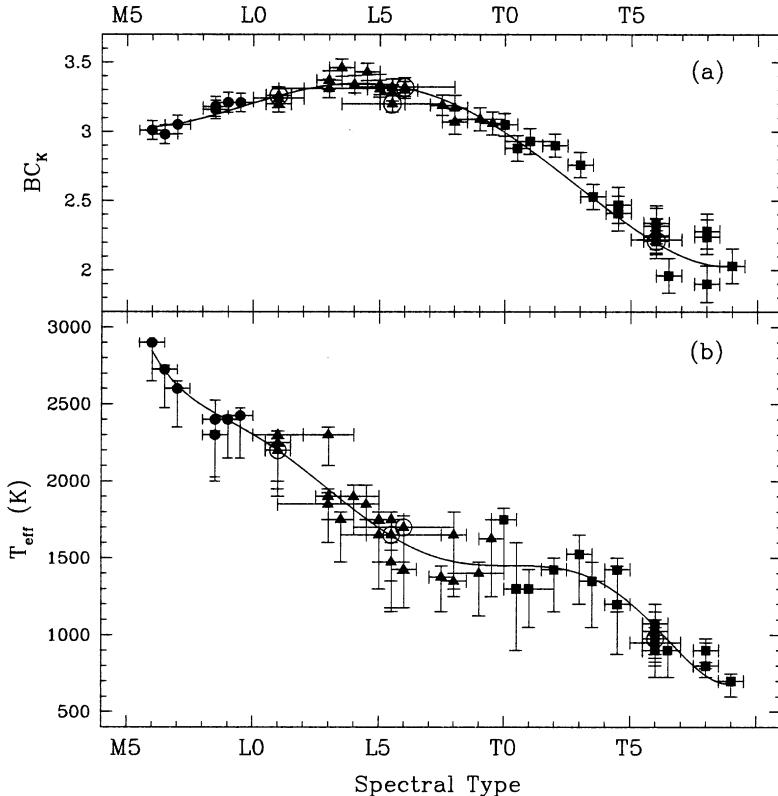


Figure 6.12. Bolometric corrections and effective temperatures of ultracool dwarfs as a function of spectral type (from [G8], courtesy of the *Astronomical Journal*). The solid lines plot 4th order and 6th order polynomial fits, respectively (see equation (2.4) for the BC_K calibration).

The temperature range spanned by spectral class L is close to prior estimates (e.g., [K5], [R3]; see Section 4.7), with $T_{eff} \sim 2,250$ K at L0 and $T_{eff} \sim 1,400$ K at L8. The variation with temperature is almost linear for \sim M7–L7 and T4–T8. However, these data suggest that the average temperature is effectively constant at $\sim 1,400$ K from spectral type L7–T3. Taken at face value, this implies that brown dwarfs spend a very short time evolving through this transition. Appearances may be deceiving, however, as will be discussed further in the following section.

We can match the temperature/spectral type calibration illustrated in Figure 6.12(b) against the theoretical evolutionary tracks shown in Figure 6.1 to estimate the relative mix of stars and brown dwarfs as a function of spectral type. In doing so, we need to bear in mind that brown dwarfs can appear at any temperature below $\sim 3,200$ K (spectral type \sim M4), although their residence times are short at higher temperatures. The Tucson models plotted in Figure 6.1 predict that $0.08 M_\odot$ stars settle onto the main sequence at a temperature of $\sim 2,100$ K, corresponding to

spectral type L1; $0.075M_{\odot}$ transition objects spend several Gyr at temperatures above 1,700 K, spectral type \sim L4. Thus, hydrogen-burning stars dominate at spectral types earlier than \sim L1; there is a mix of stars and brown dwarfs between L1 and L4, with the relative number of stars decreasing with increasing spectral type; and effectively all objects with spectral types L5 and later are brown dwarfs. In the Lyon models, the hydrogen-burning limit falls at a slightly lower mass ($0.072 M_{\odot}$), and $0.075M_{\odot}$ dwarfs are predicted to stabilise at \sim 2,000 K, but the overall picture is very similar.

6.7.3 Atmospheres, dust and the L/T transition

Temperature is the main parameter that drives spectral variations along the main sequence from O to M, and decreasing temperature also underlies most of the changes from L0 to T8. However, Figure 6.12(b) strongly suggests that some other factor is responsible for spectral evolution from L7 to T3. What are the physical changes that occur in brown dwarf atmospheres during the transition from spectral type M through L to T?

As summarised in Section 4.7, dust formation is an important process that affects spectral evolution in late M and L dwarfs. One of the main consequences is high atmospheric transparency at optical wavelengths, as the principal molecular opacity sources (TiO, VO) are removed to the solid phase [B8], [F1]. The atmospheres of L and T dwarfs are extremely non-grey, with the $\tau = 1$ level varying significantly in physical depth (and hence temperature) as a function of wavelength. The ‘photosphere’ lies at greater physical depth at $\lambda < 1 \mu\text{m}$, and the alkali atomic lines acquire increasing prominence due to high column density, high gas pressure (at $\tau = 1$) and increased contrast relative to the continuum. As the temperature decreases, those lines increase in strength until the parent elements form molecular species. In the case of the two most abundant elements, sodium ($[\text{Na}] = 6.31$) and potassium ($[\text{K}] = 5.13$), the lowest excitation (resonance) lines achieve white dwarf proportions in the coolest dwarfs, with exceedingly broad Lorentz damping wings caused by the high-pressure van der Waal’s broadening. The sodium D lines become a 1,500 Å wide bowl-like depression by spectral type L5 (Figure 6.13), while the potassium 7,665/7,699 Å doublet broadens to a width of 500 Å by L8 and exceeds 1,000 Å in the cooler T dwarfs (see further below). These resonance *atomic* lines become sufficiently strong that they affect the flux emitted in the *broadband V* and *R* passbands (Na D lines) and *I/i* and *z* passbands (K I doublet), to the extent that those colours can be used to test chemical equilibrium models [M6].

Dust plays a more active role in influencing the flux distribution at near-infrared wavelengths. Backwarming in late-type M dwarfs, where dust first makes its appearance, heats the atmosphere, dissociates H₂O, and, as a result, reduces the depth of the water bands [T12]. Dust continues to affect the near-infrared colours through the L dwarf sequence [T14]. Figure 6.14 compares the observed (M_J , $(J-K)$) diagram against brown dwarf models with dusty atmospheres, where dust is maintained at its depth of formation, and dust-free atmospheres, where the dust formation is taken into account in the equation of state and in the depletion of chemical species

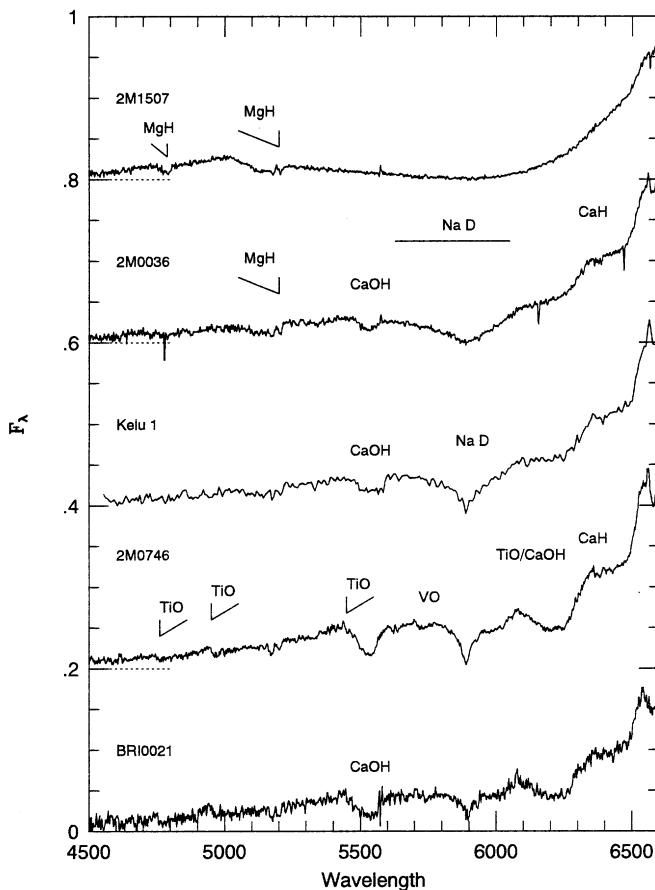


Figure 6.13. Optical blue/green spectra of late-type M and L dwarfs; the objects shown have spectral types of M9, L0.5, L2, L3.5 and L5 (from bottom to top). Note the behaviour of the sodium D lines in the latter types (from [R4], courtesy of the *Astronomical Journal*).

from the atmosphere, but gravitational settling is assumed to rapidly remove the condensates from the photosphere (the COND models). Both sets of models are by the Lyon group [C4]. The differences in the colour evolution predicted by the two models are substantial: at their reddest point, the dust-free tracks fail to match the ($J-K$) colours of even late-type M dwarfs, and show very limited colour evolution with decreasing luminosity; the dusty models are also initially too blue, but, in contrast to the dust-free models, evolve sharply redward, and predict that L dwarfs should achieve colours as red as ($J-K$) = 3.0 magnitudes.

The obvious conclusion to be drawn from Figure 6.14 is that, while dust is present in brown dwarf atmospheres, static models that leave dust where it forms are too simplified in approach. Brown dwarf atmospheres are more likely to be dynamical entities, with dust forming and ‘raining out’ to depths below the $\tau = 1$

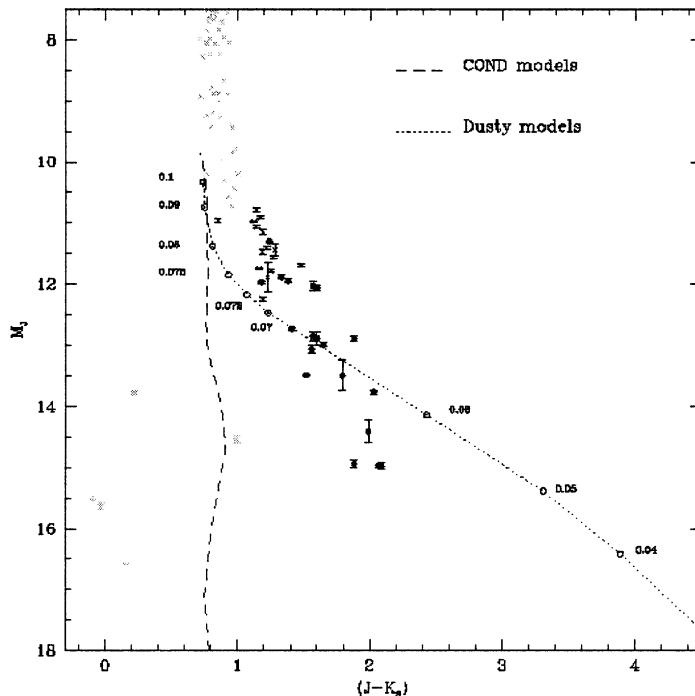


Figure 6.14. The empirical (M_J , $(J-K)$) diagram for M, L and T dwarfs compared to the predicted evolution of brown dwarfs with dust-free (dashed line) and dusty atmospheres (dotted line) (from [C4]). The individual points on the dusty isochrone are identified by the masses that correspond to an age of 1 Gyr.

‘photosphere’. Cloud formation and dissipation leads to the integrated spectrum sampling a range of depths within the atmosphere; thus, the infrared colours and bandstrengths are consistent with neither pure dust-free nor dusty models, but represent a mixture of the two extremes, with the relative proportion governed by the cloud filling factor.

In broad terms, the near-infrared colours of L dwarfs are correlated with spectral type; the later the type, the redder the colour. However, there is significant dispersion about the mean relation. This probably reflects differences in the dust distribution and cloud properties. Optical spectral types are based on the far-red spectral energy distribution, which is insensitive to dust. The dispersion in infrared colours is an argument against using infrared spectra to estimate spectral types; indeed, there is at least one dwarf (SDSS/2MASS 0423-0414) that has an optical spectral type of L7.5 [C3], but methane absorption and an infrared type of T0 [G7]. Several research groups are working actively on understanding the properties of clouds in ultracool dwarfs, with the most detailed analyses to date by Marley and collaborators [M6].

Clouds and dust almost certainly play a significant role in shaping the transition from spectral class L to class T. In particular, they may also be responsible for the

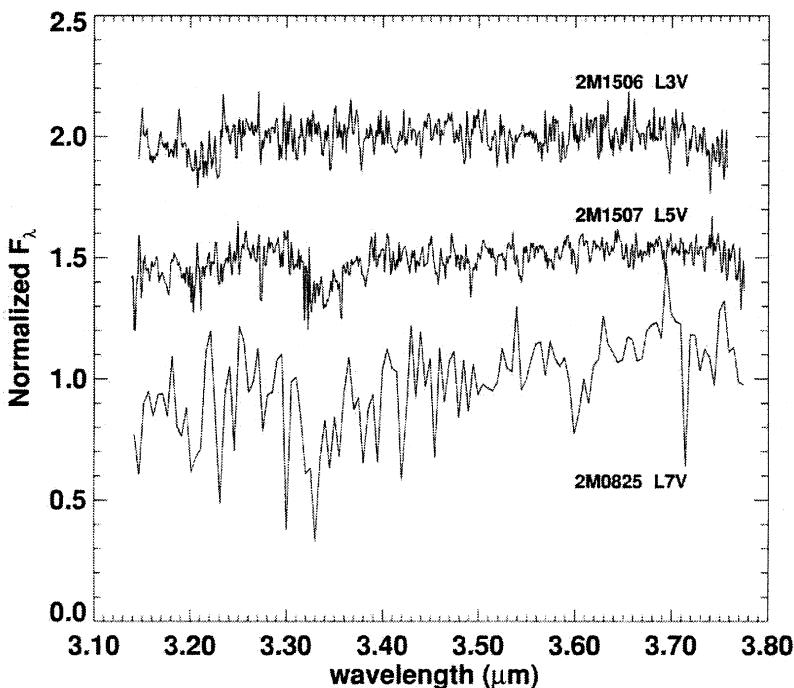


Figure 6.15. The Q branch of the fundamental ν_3 methane band at $3.3\text{ }\mu\text{m}$ in L dwarfs (from [N5], courtesy of K. Noll and the *Astrophysical Journal*).

apparent standstill in effective temperature between spectral types L7 and T3. Observationally, the L/T transition occurs with the appearance of strong overtone methane absorption between 1 and $2.5\text{ }\mu\text{m}$. Methane replaces CO as the dominant molecular repository for carbon when the temperature falls below a critical (pressure dependent) value. Methane absorption bands appear in the spectrum if there is sufficient column density of CH_4 above the $\tau = 1$ level in the atmosphere. That condition is not met for the 1.3 , 1.6 and $2.2\text{ }\mu\text{m}$ overtone CH_4 bands in most L dwarfs, although the Q branch of the stronger fundamental ν_3 bands are detected at $3.3\text{ }\mu\text{m}$ in mid- and late-type L dwarfs (Figure 6.15, [N5]). Recent high signal-to-noise spectroscopy suggests that the overtone bands are marginally detectable at spectral type L6.5, and are clearly present in L8 dwarfs, although the absorption is not sufficiently strong to have a significant impact on the broadband fluxes [N6]. (Note that this does *not* mean that L8 dwarfs need to be reclassified as type T, since the standard classification system is based on optical and near-infrared spectra.)

How can we account for the apparent standstill in the effective temperature/spectral type relation across the L/T boundary? There are drastic changes in the spectral energy distribution that must be tied to changes in the atmospheric temperature structure. The resolution of this conundrum probably lies in the fact that effective temperature is, to a large extent, a theoretical construct. The underlying

assumption is that a single temperature is appropriate to the $\tau = 1$ level at all wavelengths (i.e., (almost) all of the flux originates from the same physical depth in the atmosphere). But those circumstances do not apply in late L and early T dwarfs, where dust and λ -dependent opacity variations mean that the integrated flux is the sum of separate contributions from a range of physical depths (and hence temperatures) within the brown dwarf atmosphere. Theoretical models indicate that dust starts to settle below the $\tau = 1$ level in late-type L dwarfs. Thus, one can imagine that, as the brown dwarf cools and the dust settles, and the spectral energy distribution evolves from late L to early T, deeper, hotter layers within the atmosphere make a larger contribution to the emergent flux. The net result is that the luminosity is maintained (indeed, it increases at some wavelengths, e.g., M_J) with increasing spectral type. Since the model-based estimates of radius are essentially invariant, the effective temperature derived from Equation (1.28) formally holds constant between L7 and T3.

Theoretical models of T dwarfs have become relatively mature within the last few years (e.g., [B13]). Dust is expected to rain out completely to levels below the photosphere and have little influence on the emergent flux, although none of the current models match the depth of the near-infrared water bands, which are shallower than expected. This is reminiscent of the situation in late-type M dwarfs, and suggests that backwarming due to residual dust may still be a factor. At optical wavelengths, the spectrum is dominated by the Na D lines and K I doublet (Figure 6.16). Sodium is 20 times more abundant than potassium, and the D lines account for most of the absorption shortward of $1\,\mu\text{m}$. The exact line profile of those broad features remains unclear, however; adopting pure Lorentzians leads to significantly lower flux levels than observed for far-red wavelengths ($0.6\text{--}1.1\,\mu\text{m}$). Most models adopt truncated Lorentzians, limiting the extent of the wings in the line profile; the truncation mechanism is not known.

Theoretical models can be used to explore the consequences of changes in metallicity and gravity (i.e., mass, since brown dwarf radii vary so little). Lowering the metal abundance leads to more transparent atmospheres, and stronger atomic features, since the $\tau = 1$ level lies at higher pressure. These effects are similar to those observed in M (and L) subdwarfs (see Chapter 10). Lower gravity T dwarfs have deeper near-infrared water bands and redder ($J\text{--}K$) colours, due to weaker collisionally induced H_2 absorption (at $2.2\,\mu\text{m}$); both effects mimic cooler effective temperatures. We consider gravity indicators in M, L and T brown dwarfs in more detail in the following section.

Several interesting discrepancies between the models and the observations remain to be explained. Lithium, for example, is not detected in any T dwarfs (Figure 6.16, inset). This is surprising, since a substantial number of T dwarfs are expected to have masses below $0.06\,M_\odot$, the critical mass for lithium preservation. There are some observational indications that the Li I $6,708\,\text{\AA}$ line weakens in later-type L dwarfs, and it is possible that lithium may be depleted (to LiCl , LiH and LiOH), even though chemical balance models indicate that this should not be the case [L3]. In contrast, as noted in Section 6.5.3, Cs persists, even though it is predicted to be depleted. A further anomaly is the detection of CO at $4.7\,\mu\text{m}$ in

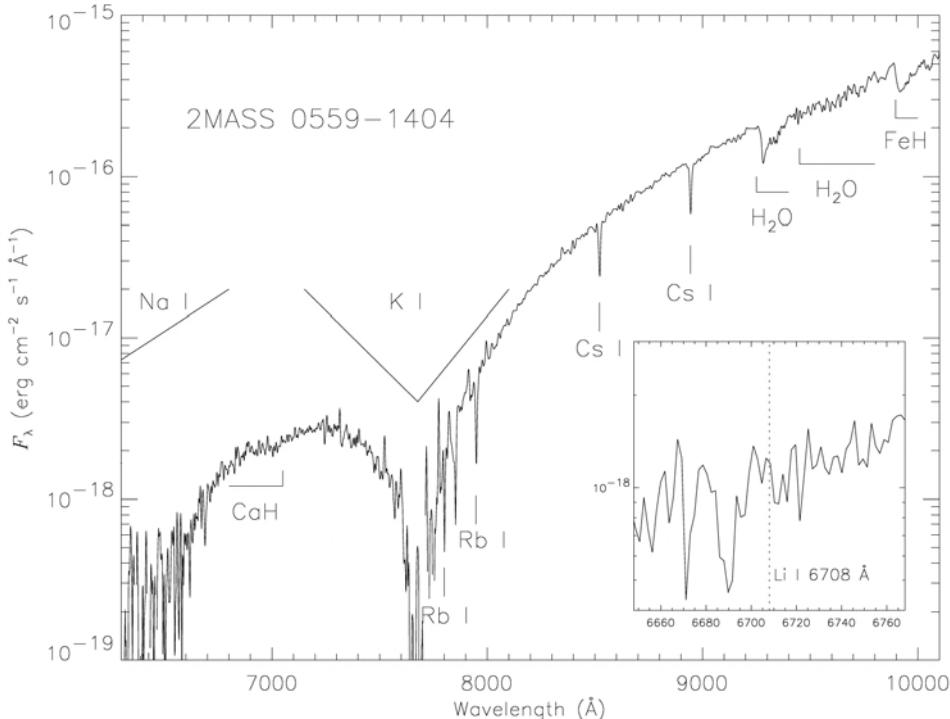


Figure 6.16. Far-red optical spectrum of the T5 dwarf, 2MASS 0559-1404. The most prominent molecular and atomic features are identified (from [B14], courtesy of the *Astrophysical Journal*).

Gl 229B [N4], since models predict abundances over 1,000 times lower than the value inferred from those observations. It is possible that these are also manifestations of the non-grey nature of T dwarf atmospheres.

6.7.4 Gravities and masses

The initial mass function, the number of stars forming per unit mass, is a key parameter for understanding star formation, and Chapters 8 and 9 focus on this issue. Deriving this function for hydrogen-burning stars is difficult, but tractable, since those objects conform to a mass–luminosity relation (Section 9.4). Brown dwarfs, on the other hand, fade rapidly with time along very similar cooling tracks; to a first approximation, all brown dwarfs, regardless of mass, look the same at some point or other during their lifetime.

Open clusters and young associations are therefore inviting targets for brown dwarf searches. First, age is a known quantity, so one knows roughly where to look (in colour–magnitude space). Second, young brown dwarfs have higher temperatures and luminosities, so low-mass objects are correspondingly more accessible. Set

against these advantages are the inherent problems of crowding and contamination from foreground and background stars and brown dwarfs; identifying complete samples, and determining whether individual objects really are members of the cluster or association. These problems are particularly acute in surveys of star-forming regions like Orion, ρ Ophiuchus and IC 348, which are sufficiently young that brown dwarfs with masses of $\sim 0.01 M_{\odot}$ ($\sim 10 M_J$) are detectable as late-type L or T dwarfs, but which are sufficiently distant (300–500 parsecs) that there is potential for contamination by foreground objects, even L and T dwarfs.

Almost all cluster surveys start by using photometric criteria to select candidate members. We will consider statistical issues associated with the derivation of luminosity functions and mass functions in Sections 8.9, 9.6 and 9.7; here, we consider low-mass brown dwarfs, and describe how spectroscopic features can be used to refine a photometrically selected sample and segregate genuine low-mass cluster members from interlopers.

As described in Section 3.5.3, electron degeneracy dictates that low-mass stars ($M < 0.15 M_{\odot}$) and brown dwarfs have radii within 10% of Jupiter's (see Figure 3.17). Young brown dwarfs ($\tau < 20$ Myr), which are still contracting toward their final configuration, have radii that are several times larger (see Figure 3.10). As a consequence, while the surface gravity of a main sequence $0.1 M_{\odot}$ M8 dwarf is $\log g \sim 5.3$, a 10-Myr old $0.01 M_{\odot}$ brown dwarf has $\log g < 3.5$. This substantial difference leads to significant changes in the strength of spectroscopic features due to a number of chemical species. Section 2.2.8 described some of the effects in early and mid-type M dwarfs; here we consider the effect on later type spectra.

Figure 6.17 compares the optical spectra of three M9.5 objects and two early-type L dwarfs [M9]. The former include a giant ($\log g \sim 0.5$), a young brown dwarf (KPNO-Tau 4, $\tau \sim 1$ –2 Myr, $\log g \sim 3.5$) and the field dwarf, 2M0149; the latter are G 196-3B ($\tau < 300$ Myr, see Table 6.2) and Kelu 1. As in earlier type M stars, decreasing gravity at M9.5 leads to weaker alkali lines (Na I, K I) and metal hydride bands (in this case, CaH), but stronger TiO and, at these cooler temperatures, VO absorption. Similarly, G 196-3B has significantly weaker alkali lines (Rb, Cs, Na) and FeH, CrH absorption than the older (and more massive) Kelu 1. The exception to this correlation is lithium, which is much stronger in G 196-3B; Kelu 1 lies close to the mass limit for lithium burning, and, in this case, the relative strengths reflect higher depletion of primordial material in the older, more massive field dwarf than in the younger, less massive object.

Similar effects are present at near-infrared wavelengths. Figure 6.18 shows J band spectra of the M9.5 giant and the two young, low-mass brown dwarfs. The alkali lines (the K I doublets and Na I) and the FeH bands are significantly weaker in the latter two dwarfs than in typical field dwarfs (see Figure 2.9). Moreover, weak TiO absorption is present, which is not the case in higher mass dwarfs. Besides these changes, there is some strengthening of the water bands with decreasing gravity; however, this effect is only significant at temperatures below $\sim 2,000$ K (spectral types later than \sim L1) [G14]. Observations are only starting to probe the full range of spectral variations; with the current focus on young clusters

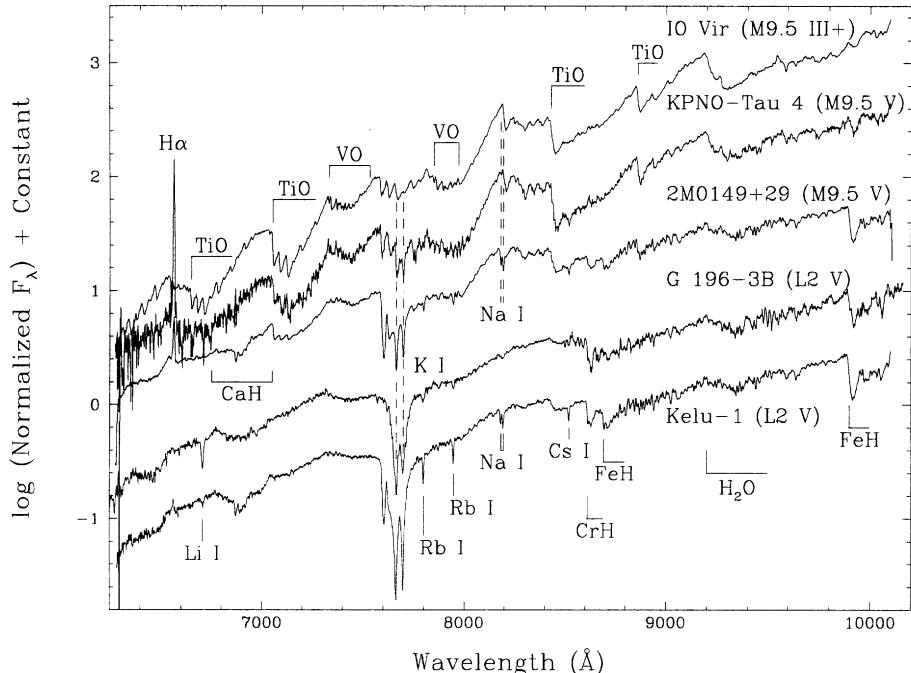


Figure 6.17. Gravity sensitive features in late-type dwarfs – the far-red spectral region of three M9.5 objects, a giant (IO Vir), a young brown dwarf (KPNO-Tau 4) and a field dwarf (2M0149); and two L2 dwarfs, the 300-Myr old G 196-3B and Kelu 1 (from [M9], courtesy of the *Astrophysical Journal*).

and associations, a more detailed picture of the spectroscopic behaviour at low gravities will emerge in the near future.

6.7.5 Activity and rotation

Section 5.6 described how the overall level of chromospheric and coronal activity declines substantially in ultracool dwarfs. X-ray emission (in the form of short-lived flares) has been detected from a number of brown dwarfs (e.g., [B18]), although all of the detected objects are ultracool M dwarfs. This is not surprising: as discussed in Section 5.6, brown dwarfs and stars have very similar atmospheric structures at the same temperature, and therefore behave in a similar fashion. The persistent and flaring radio emission detected from several ultracool dwarfs, including 2MASS036 + 1821 (L3.5) [B19], is more surprising and as yet unexplained.

H α emission is present in a number of early and mid-type L dwarfs, with the detection rate declining from $\sim 60\%$ at L0 to $< 5\%$ at L4 [K9]. The detection of an H α flare in 2MASS0144-0716, an L5 dwarf, shows that activity persists at later types, albeit at an even lower level [L4]. Interestingly, the active L0–L4 dwarfs appear to have velocity dispersions that are higher than inactive L dwarfs of the

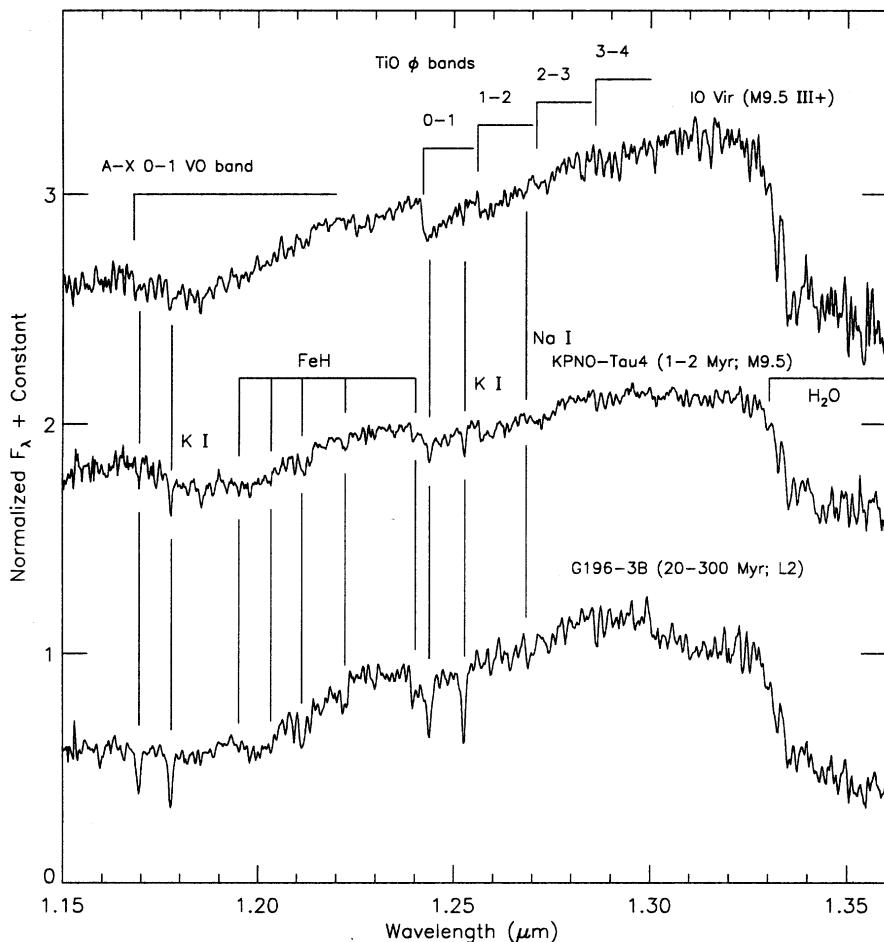


Figure 6.18. Gravity sensitive features in late-type dwarfs – near-infrared J band spectra of an M9.5 giant and two young brown dwarfs (from [M9], courtesy of the *Astrophysical Journal*).

same spectral type (i.e., the active L dwarfs are likely to be older and of higher mass). There are only two L dwarfs that have both $H\alpha$ emission and lithium absorption: Kelu 1 (L2) and SDSS/2MASS 0423-0414 (L7.5/T0 – see previous section).

Given the scarcity of active late-type L dwarfs, it is somewhat surprising that $H\alpha$ has been detected in emission in at least three T dwarfs: SDSS 1254-0122, a T2, 2MASS 1047+2124 and 2MASS 1237+6526, both T6.5 [B14]. The last is the most remarkable, with strong, well-defined emission (Figure 6.19), and an activity level over 10 times higher than any other T dwarf (Figure 6.20). Observations at several epochs show no evidence for variability in the emission line flux, as might be expected for chromospheric, particularly flaring, activity. Accretion of material from a close binary companion offers an alternative mechanism, but photometric

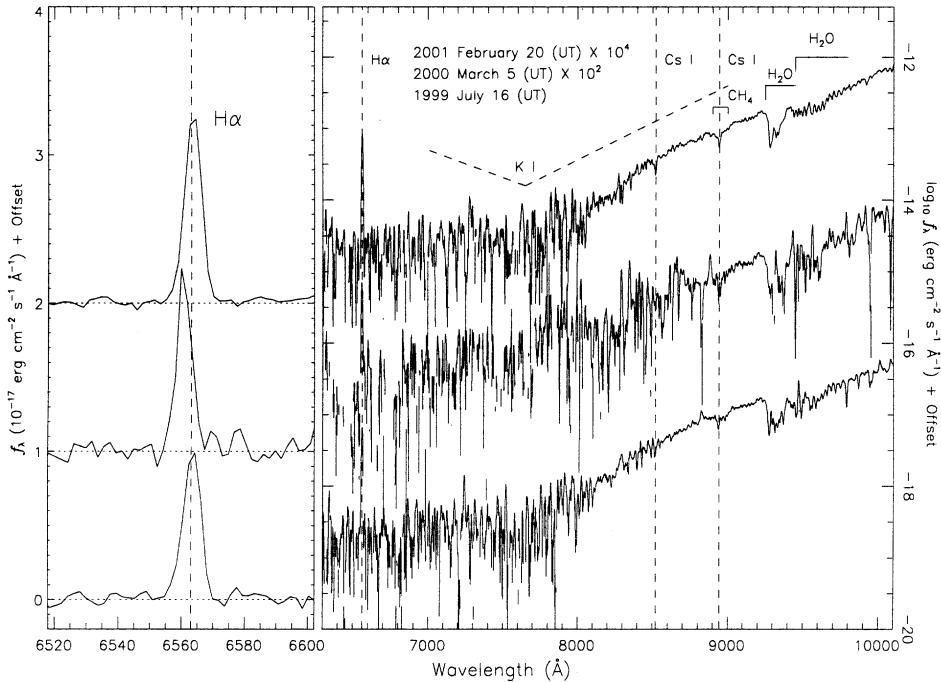


Figure 6.19. H α emission in the T6.5 dwarf, 2MASS 1237+6526 – the observations, from [B15], span 3 years, but show no evidence for significant variability (courtesy of the *Astrophysical Journal*).

monitoring shows no evidence for eclipses, while there is only marginal evidence for radial velocity variations [B15]. All of these characteristics are reminiscent of the peculiar M9 dwarf, PC0025+0447 (Section 5.6.1).

Activity and rotation are correlated for early M dwarfs, but, as described in Section 5.5.3, these parameters decouple in later type dwarfs. The breakdown in the correlation is underlined in ultracool dwarfs: over 25 have high-resolution spectroscopic observations [R4], [B16], and all have (projected) rotational velocities exceeding 10 km s^{-1} ; several, including Kelu 1, have $50 < v \sin i < 80 \text{ km s}^{-1}$. However, the overall level of activity ($f_{\alpha}/F_{\text{bol}}$, see Figure 6.20) is more than 2 orders of magnitude below the average for dMe stars, and some of the fastest rotators lack any detectable H α emission.

Chromospheric activity might also manifest itself in the form of star spots, which are potentially detectable through periodic photometric variations as they rotate into and out of the line of sight. Given a radius of $0.1 R_{\odot}$, a rotational velocity of 60 km s^{-1} corresponds to a rotational period of ~ 2 hours. The observed rotational velocities of $10\text{--}80 \text{ km s}^{-1}$ therefore correspond to rotational periods of ~ 12 hours to 100 minutes. Several photometric campaigns have been undertaken with the aim of detecting such variations (e.g., [G9], [B17]), with some

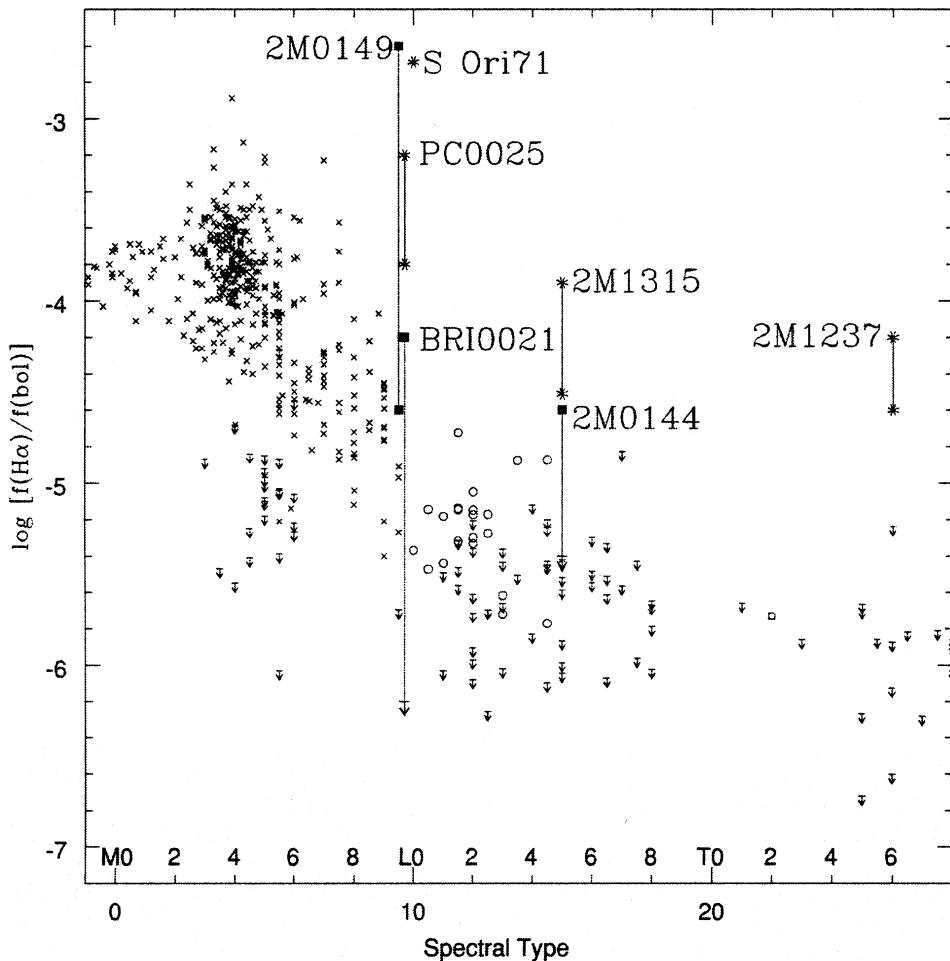


Figure 6.20. Activity in ultracool dwarfs – individual objects of interest, including 2M1237 and PC0025, are identified (from [L4], courtesy of the *Astrophysical Journal*).

success in a number of cases. Repeated observations, however, show that while the observed periods may be sustained on timescales of days, they are not recovered after weeks or months. The most likely explanation is that the photometric variability is due to clouds forming and dissipating in the low-temperature atmospheres, with covering factors of 10–15%. Arguing in favour of this hypothesis, variations in molecular bandstrengths have been detected in several L dwarfs. As an example, Figure 6.21 compares spectra taken at two different epochs of 2MASS 1523 + 3014 (L8), also known as Gl 584C; there are clear differences in the spectral regions longward of the 8,600 Å CrH band, and in the 9,300 Å H₂O band. The simplest explanation for these differences seems to be variation in the cloud cover on Gl 584C at the two epochs [K10].

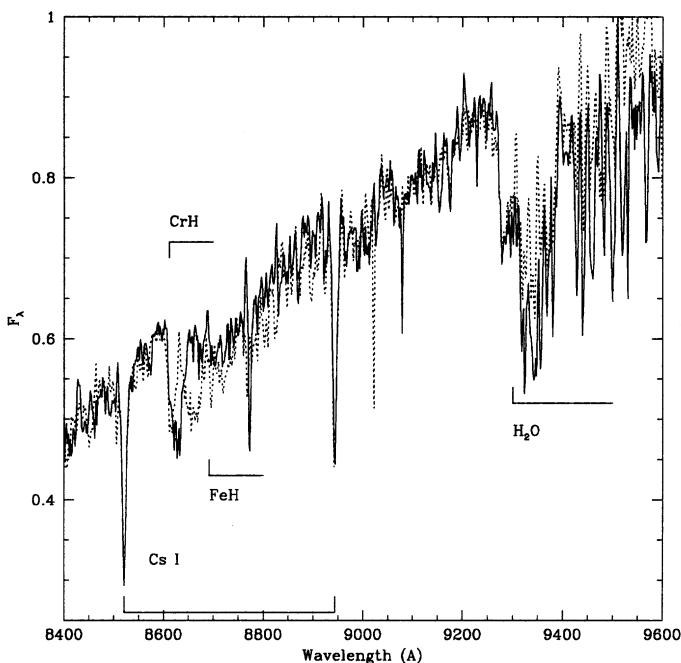


Figure 6.21. Changes in H_2O bandstrength in 2MASS 1523 + 3014, Gl 584C (from [K10]) from 24 Dec, 1998 (the ‘normal’ state) are plotted as a solid line, while observations from 25 Dec, 1998 are shown as a dotted line.

6.8 BROWN DWARFS AS COMPANIONS

Nerissa: When the Moon shines, we do not see the candle
 Portia: So doth the greater glory dim the less . . .

Shakespeare, *Merchant of Venice*

Searches for companions to known nearby stars have long proven an effective means of finding low-luminosity stars and brown dwarfs, as demonstrated by the list of discoveries in Section 6.5. As with surveys probing stellar binarity, these programs can use three main observational techniques: direct imaging, to find wide companions; long-term astrometry, to find companions at moderate separations; and radial velocity measurements, to find companions at the smallest separations. In each case, the sensitivity depends on several parameters: for imaging, the distance, separation, companion mass and age all affect detectability; for astrometry, the distance, separation and companion mass are important; and for radial velocity surveys, the companion mass and separation are directly relevant, while distance plays an indirect role through its effect on the apparent magnitude of the primary. Of these techniques, we consider primarily results from direct imaging in this section; both astrometric and radial velocity surveys have been tuned more toward finding extrasolar planets, and we consider results from those programs in Section 11.3.

6.8.1 Brown dwarf companions of main sequence stars

Photographic, proper-motion surveys afforded the first opportunity to search for low-luminosity companions with separations exceeding 2–3 arcseconds, and both the Lowell survey and Luyten’s Palomar surveys resulted in the detection of many wide, common proper-motion pairs. The most successful study in terms of detecting brown dwarf candidates was van Biesbroeck’s [V3], [V4] search for distant companions to nearby stars, which yielded some 20 systems, including VB7, an extremely cool white dwarf, and both VB8 (Gl 644C – spectral type M7) and VB10 (Gl 752B – M8), the archetypical ultracool M dwarfs.

Recent studies have concentrated on observations of stars within 10 parsecs, particularly late-type members of the 8-parsec sample (see Appendix). Systematic observational studies of the latter sample include the infrared speckle of M dwarfs [H6], wide-field *I*-band imaging [S2], ground-based *I*-band coronagraphy [O4] and both optical (WFPC2) and near-infrared (NICMOS coronograph) imaging with HST [G12]. In addition, high-resolution AO images have been obtained of a number of other nearby K and M dwarfs [D3], [C5], [B20]. While these observations have resulted in the identification of a number of stellar companions, only one brown dwarf has been discovered: LHS 2397aB, spectral type L6–L8 [F2].

Some of these techniques have also been applied to investigations of multiplicity in stars beyond the immediate Solar Neighbourhood. In particular, high-resolution red and infrared imaging with HST has been used to search for low-mass companions to M dwarfs in both the field [R8] and open clusters, such as the Hyades ([G1], [R4]), Pleiades [M7] and α Persei [P6]. As described in the following section, these observations probe angular separations of 0.1 arcsecond or more, corresponding to 5–15 AU for those targets. Again, a number of new stellar companions have been discovered, but no brown dwarfs, even though the open cluster observations had sufficient sensitivity to detect companions with masses as low as $0.05 M_{\odot}$.

High-resolution imaging observations are being made with increased frequency using ground-based adaptive optics systems. As a particular example, Gl 569B, the companion to Gl 569A originally discovered by Forrest *et al.* [F3], was resolved into two components, M8.5 and M9, by Martin *et al.* using an AO system on the Keck II telescope [M7]. The period of the ultracool system is only 2.5 years, and the orbit has been derived from repeated observations, leading to mass estimates of 0.055–0.08 and 0.05 – $0.07 M_{\odot}$ for the two components [L5]. Ground-based systems offer a substantial potential advantage in resolution compared to HST, since the larger apertures lead to smaller diffraction limits and higher angular resolution. At present, most studies require natural guide stars (see Section 1.8.2), and are therefore limited to relatively bright stars ($R < 14$ th magnitude), but laser guide star systems will become increasingly available in the future.

Several studies have targeted chromospherically active stars within 20–30 parsecs of the Sun. The underlying premise is that these stars are young (see Section 5.5.2), so brown dwarf companions are correspondingly more luminous, and easier to detect. Thus, the Gl 569 system has an age of \sim 300 Myr, so the brown dwarfs in that system are ultracool M dwarfs. Following this strategy,

Rebolo *et al.* [R2] discovered an L2 dwarf some 300 AU distant from the nearby (~ 21 pc) M dwarf, G 196-3. Assuming an age of 20–100 Myr, the companion has a mass between 15 and $50 M_J$. Similarly, observations of members of nearby young associations, such as TW Hydrae and the β Pic moving group, have produced at least two M-type brown dwarf candidates, TW Hya 5B and HR 7329B [Z4], and an L dwarf 2M1207 B [C7].

Finally, another strategy, which has become possible with the completion of 2MASS, is to search the vicinity of known nearby stars for sources that have near-infrared colours consistent with ultracool M, L or T dwarfs, and magnitudes consistent with the known distance of the target star. This method is limited to finding relatively wide systems, given the 2–3 arcsecond resolution of 2MASS (with larger limits for bright stars). Several L dwarf companions have been discovered using this technique ([K10], [W3], [G11]), and a systematic search of the immediate environs of all stars known to lie within 25 parsecs of the Sun is currently underway.

Table 6.2 lists the main characteristic of multiple systems that include a main sequence (or, for GD 165A, post main sequence) primary and at least one L or T dwarf secondary. It is important to emphasise that these discoveries stem from a range of different techniques, each with its own set of biases; none of these techniques have been applied in a consistent manner to stars beyond the 8-parsec sample, and in some cases there are non-uniformities even within that distance limit. Nonetheless, most of the companions lie at distances of more than 150 AU from the primary. We will consider these results in more detail in our discussion on extrasolar planets and the ‘brown dwarf desert’ (Section 11.4.2).

6.8.2 Ultracool binary systems

Ultracool dwarfs are themselves excellent targets for searches for brown dwarf companions (*vide* Portia’s remark). Most observational studies to date have relied on high-resolution imaging with HST, since very few ultracool dwarfs have sufficiently bright apparent magnitudes that they can serve as natural guide stars for ground-based AO systems (although a few systems have been resolved by direct imaging in good seeing conditions). The existence of such systems was established relatively quickly, since two of the three L dwarfs from the DENIS Mini-survey (DENIS 0205-1159 and DENIS 1228-547) prove to be binaries, with similar magnitude components separated by ~ 0.2 arcseconds ([M8], [K11]). To date, HST observations are available for over 160 late-M and L dwarfs ([R7], [G13], [B21]) and 10 T dwarfs [B22]. Table 6.3 lists the salient properties of the binaries identified from those programs, together with a number of ground-based discoveries. Combined, these data provide considerable insight into the properties of very low-mass binaries, and allow comparison with their more massive, hydrogen-burning counterparts.

Considering the data listed in Table 6.3, there are two clear areas of systematic behaviour: first, none of the ultracool systems have component separations that exceed 15 AU; second, almost all systems have mass ratios between 0.8 and 1.0 – ε Indi Bab and 2M1207 are outliers at $q = 0.5$ and 0.15 respectively. The restriction

Table 6.2. L/T binaries with main sequence or post main sequence primaries.

Primary	Spectral type		Secondary	Spectral type		Δ (AU)	M_2	q	Reference
	(Gyr)								
LHS 102A	M3.5	~5	LHS 102B	L5	190	0.075	0.2	G10	
			LHS 102C	L5	0.8–1.3	0.075	~0.9	G12	
GJ 1048A	K2	<1	GJ 1048B	L1	250	0.07	0.1	G11	
Gl 337A	G8	0.6–3.4	Gl 337C	L8	880	0.04–0.074	~0.08	W3	
G 196-3A	M2.5	0.3	G 196-3B	L2	340	0.02–0.05	0.06	R2	
Gl 417A	G0	0.1–0.3	Gl 417B	L4.5	1,950	0.02–0.05	0.04	K10	
			Gl 417C	L6	1.5	0.02–0.05	~1	B21	
LHS 2397a	M9	>3	LHS 2397aB	L7		0.065–0.08	0.9	F2	
GD 165A	DAZ4	1.2–5.5	GD 165B	L4	120	0.07–0.08	<0.06	K4	
Gl 564A	G2	<0.8	Gl 564B	L4	18	0.075	0.07	P4	
			Gl 564C	L4	0.9	0.065	0.85	P4	
Gl 584A	G1	1.0–2.5	Gl 584C	L8	3,610	0.05–0.075	0.06	K10	
Gl 618.1A	M0	0.6–11	Gl 618.1B	L2.5	1,090	0.06–0.08	0.1	W3	
HD 89744A	F7	1.5–3	HD 89744B	L1	2,460	0.075–0.08	0.06	W3	
Gl 779A	G1	1–3	Gl 779B	L4.5	14	0.055–0.08	0.06	L8	
Gl 229A	M0.5	>5	Gl 229B	T6.5	45	0.05–0.06	0.1	N2	
Gl 570A	K5	2–10	Gl 570D	T6.5	1,525	0.03–0.07	0.1	B10	
ε Indi A	K5	1–2	ε Indi Ba	T1	1,460	0.04–0.06	0.06	S8	
			ε Indi Bb	T6	2.65	0.02–0.04	0.5	M5	

Note: Column 1 lists the name of the primary and column 2 the spectral type; column 3 lists the age estimate for the system; column 4 identifies the low-mass secondary; column 5 lists its spectral type; column 6 gives the observed separation, in AU; column 7 lists the mass estimate for the secondary; column 8 shows the mass ratio, $q = M_2/M_1$ (M_3/M_2 for tertiary components); and column 9 gives the primary reference for each system.

to small separations was discussed implicitly in Section 2.7.1, in the broader context of the separation distribution of all binary systems as a function of total system mass. By concentrating on ultracool binaries, we are selecting systems with total mass $M_{tot} = M_1 + M_2 < 0.15 M_\odot$, lying at the base of the distribution shown in Figure 2.27. The physical mechanism responsible for this behaviour is not known, but it might reflect either constraints on the star formation process (low-mass wide binaries never form) or post-formation dynamical evolution (low-mass wide binaries are disrupted easily). Luhman has identified a possible counter example to this trend: a pair of $\sim 0.04 M_\odot$ M8 dwarfs separated by 240 AU in the Chamaeleon I star-forming region [L6]. Similarly, 2M1207 B [C7] is a $\approx 7\text{--}10 M_J$ L dwarf lying 60 AU from the $\sim 35\text{--}40 M_J$, ~ 10 -myr-old primary, a member of the TW Hydrae association. It is possible, however, that these objects formed separately within their clusters, rather than as dynamically associated binary pairs.

Unlike the absence of wide binaries, the strong preference for ultracool binaries with near-equal mass ratios could be a selection effect. However, closer examination of the underlying observations suggests that this is not the case. Figure 6.22 compares the magnitude differences for systems with HST images against the

Table 6.3. Ultracool binary systems.

System	Spectral type	Δ (AU)	$M_I(A)$	$M_I(B)$	q	Reference	Comments
D0205-1159	L7	7.3	16.8	17.5	0.95	K11	
D0357-4417		2.2	16.2	17.7	0.9	B21	
D1004-1146		6.8	15.0	15.65	1.0	B21	
D1228-1547	L5	5.1	16.85	17.25	1.0	M8	
D1441-0945		11.2	15.5	15.85	1.0	B21	
LP 415-20	M7/M9.5	3.6	10.7*	12.5*	0.85	S9	Hyad?
LP 475-855	M7.5/M9.5	8.6	10.8*	12.3*	0.9	S9	Hyad?
2M0746 + 2000	L0.5/L1.5	2.7	15.0	16.0	1.0	R7	
2M0850 + 1057	L6/L8	6.4	17.2	18.7	0.85	R7	
2M0856 + 2295		3.5	16.6	19.3	0.7	B21	
2M0920 + 3517	L6.5	1.5	18.3	19.2	0.95	R7	
2M1017 + 1308		3.0	16.2	17.0	0.95	G13	
2M1047 + 4026	M8/L0	2.9	11.0*	11.85*	0.9	C5	
2M1112 + 3518	L4.5	1.5	15.9	18.0	0.85	B21	G1 417BC
2M1127 + 7411	M8	9.0	13.5	14.3	0.95	G13, C5	
2M1146 + 2330	L3/L3	8.0	16.4	16.7	1.0	K11, R7	
2M1207	M8/L4	60	11.2	18.3	0.15	C7	TW Hya
2M1239 + 5515	L5	4.1	17.5	17.8	1.0	G13	
2M1311 + 8032	M8	9.8	13.4	13.8	1.0	G13, C5	
2M1426 + 1557	M8.5/L1	4.2	14.1	15.5	0.9	G13, C6	
2M1430 + 2915	L1	2.8	15.4	16.2	0.95	G13	
2M1449 + 2355	L0	13.1	14.0	15.5	0.9	G13	
2M1600 + 1708	L1.5	5.8	14.65	15.35	0.95	G13	
2M1728 + 3948	L7	3.1	18.4	19.1	0.95	G13	
2M1750 + 4424	M7.5/M9.5	4.9	10.8*	11.5*	0.9	S9	
2M2101 + 1756	L7.5	5.9	18.8	19.4	0.95	G13	
2M2140 + 1625	M8.5/L0	5.2	13.5	15.0	0.85	G13, C5	
2M2147 + 1431	M8	12.7	14.4	15.8	0.9	G13	
2M2206-2047	M8/M8.5	5.0	13.2	13.6	1.0	G13, C6	
2M2331-0406	M8/L3	15.0	13.5	17.4	0.7	G13, C6	
SD2335-0013		1.1	17.7	18.7	0.9	B21	
Gl 569Bab	M8.5/M9	0.9	11.2*	11.7*	0.9	L5	
Gl 564BC	L4/L4	0.9	12.6*	12.9*	0.9	P4	
2M1225-2739	T6/T8	3.2	20.1	21.6	0.7	B22	
2M1534-2952	T5.5/T5.5	1.0	18.6	19.1	1.0	B22	
LHS2397a	M8/L7.5	3.0			0.8	F2	
ε Indi Bab	T1/T6	2.65	14.3*	15.05*	0.5	M5	

Note: column 1 lists the name of the system; column 2 gives either the joint spectral type or individual estimates; column 3 gives the observed separation in AU; column 4 lists the I -band absolute magnitude for component A; column 5 gives the same data for component B (in several cases, marked with an asterisk, the J-band absolute magnitude, M_J , is listed rather than M_I); column 6 shows the estimated mass ratio, q ; column 7 gives the primary reference for each system; and column 8 lists additional comments.

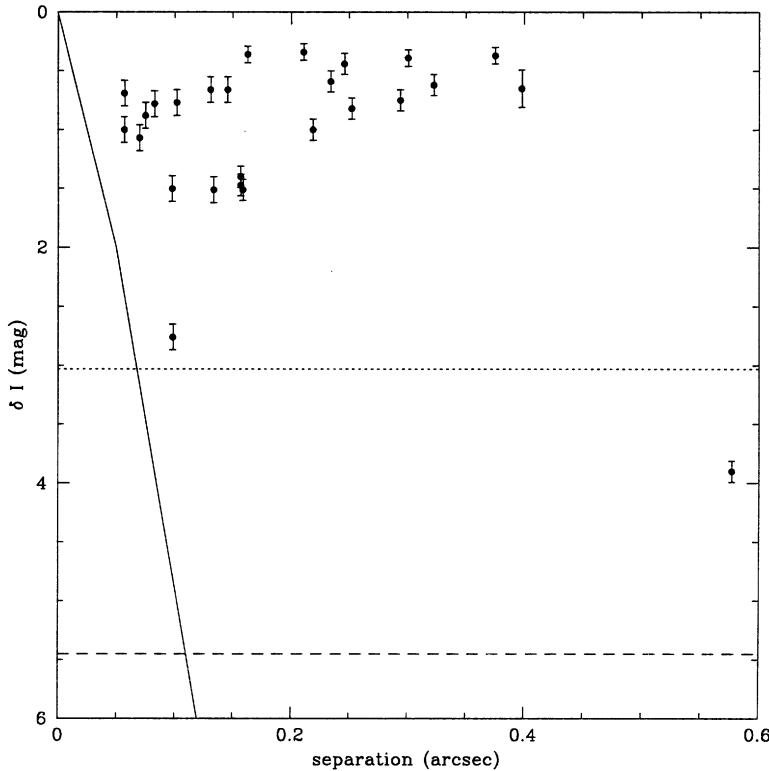


Figure 6.22. The distribution of component magnitude difference as a function of angular separation for ultracool dwarfs with HST observations. The solid line marks the formal detection limit set by instrumental resolution; the dotted and dashed lines mark the upper and lower sensitivity limits set by the individual observations. Almost all of the detected binaries lie well above even the brighter magnitude limit (from [B21]).

formal detection limits set by the angular resolution and the sensitivity of the observations. The overwhelming majority of the binaries are well above the formal detection limits (i.e., there is no obvious reason that systems with mass ratios as low as 0.4 should not have been found in these observations). The obvious conclusion is that such systems are rare. This preference for equal-mass systems echoes, and perhaps amplifies, a trend evident among the M dwarf binaries of the 8-parsec sample.

The overall binary fraction found in surveys of ultracool dwarfs is $\sim 20\%$. This is likely to be an overestimate of the true binary fraction, since all of these surveys are effectively drawn from magnitude-limited samples. A binary consisting of two components of equal brightness is twice as bright as a single object of the same spectral type; as a result, the effective distance limit in a magnitude-limited survey is higher by $2^{1/2}$, and the sampling volume for binaries is $2 \times 2^{1/2}$ larger than for single dwarfs. (By the same token, since binaries are brighter than single stars, it is not surprising

that two of the three brightest L dwarfs found by the DENIS survey prove to be binary.) Allowing for this bias gives a reduced binary fraction of $\sim 15\%$ for ultracool dwarfs [B22], [G13], [B21], significantly lower than the $\sim 30\%$ measured for M dwarfs (Section 2.7.1), but still higher than the 5% favoured by the disruptive brown dwarf formation mechanism described in Section 3.6.2. This calculation does not take small separation binaries into account ($\Delta < 5$ AU), but it is clear that the decrease in Δ_{\max} with decreasing mass (Figure 2.27) must be a significant factor. We consider the implications for star (and brown dwarf) formation mechanisms in Section 9.9.

6.9 BEYOND T: Y DWARFS AND THE FUTURE

The coolest T dwarfs currently known have effective temperatures estimated at ~ 650 – 700 K. However, substellar mass objects cool to much lower temperatures; indeed, most brown dwarfs in the Solar Neighbourhood are expected to have temperatures between 200 and 300 K (see Section 9.5). Different molecular species become prominent as the temperature decreases. In particular, at temperatures below 500 K, absorption due to ammonia (NH_3) appears at 1.5, 1.95, 2.95 and 10.5 μm ; this will probably serve as the signature molecule for spectral class Y. The methane bands at 2.2, 3.3 and 7.8 μm also strengthen considerably ([B24]; Figure 6.23). These strong absorption bands force flux into a few relatively transparent windows, notably the region between 3.5 and 5.5 μm – the *M* band. Surveys in this passband are likely to offer the best prospects for finding these very low-luminosity objects. A ground-based program would be extremely difficult (see Section 2.2.7); fortunately, several satellite mid-infrared observatories are either in place or planned for the near future.

Spitzer was launched in late 2003 and, as described in Section 1.6.2, can obtain imaging and spectroscopy at mid-infrared wavelengths. Spitzer will not undertake an all-sky survey, although specific science programs will compile deep imaging observations of a few hundred square degrees, and those data will be combed for candidate Y dwarfs. Depending on the form of the initial mass function, those programs should reveal anywhere between a handful to ~ 100 candidates.

WISE, the Wide-field Infrared Satellite Explorer, will address this issue directly through an all-sky survey in several passbands between 3 and 24 μm . Currently in the final stages of planning, WISE should be ready for launch in late 2009. Although the instrument has a relatively small aperture, it will be capable of detecting room temperature Y dwarfs within a few parsecs of the Sun; the main complication will be sifting through the extensive source catalogue to identify the relatively small number of good candidates.

The James Webb Space Telescope (JWST), the successor to HST, will likely provide the first opportunity for detailed inspection of any Y dwarfs identified by Spitzer or WISE. JWST will be an ~ 6 -m telescope placed in an Earth-trailing orbit (i.e., the telescope will orbit the Sun in close proximity to the Earth, rather than

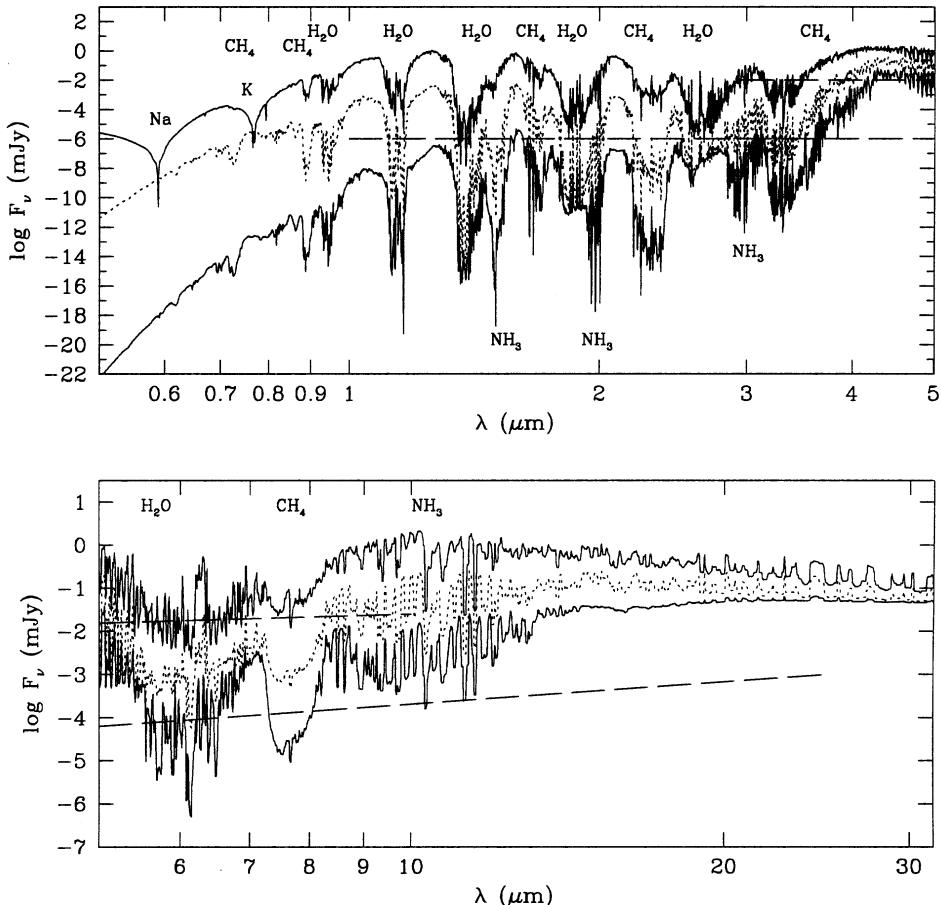


Figure 6.23. Model spectra of a $5 M_J$ brown dwarf at ages of 0.1, 0.3 and 3 Gyr; the corresponding effective temperatures are 590, 425 and 225 K. In each case, the flux levels are scaled to a distance of 10 parsecs, and the principal spectral features are identified. The dotted lines show the detection limit of IRAC and Spitzer (*upper set*) and the expected sensitivity of JWST (from [B24]).

circling the Earth itself). This configuration avoids problems such as terrestrial eclipses during each orbit and interference from the South Atlantic Anomaly (but renders servicing missions extremely difficult, except through the use of robotic technology). The telescope will be optimised for infrared observation, with minimal coverage shortward of $1\mu\text{m}$. The instrumentation will include imaging cameras and spectrographs that will cover both the near-infrared ($1\text{--}5\mu\text{m}$: NIRCam and NIRSpec) and the mid-infrared ($5\text{--}28\mu\text{m}$: MIRI). Figure 6.23 shows the expected sensitivities of the JWST suite of instrumentation. The overall field of view of any instrument will be very small (only $\sim 2 \times 2$ arcminutes with NIRCam and $\sim 110 \times 75$ arcseconds with MIRI), so it is unlikely that JWST will

be used directly to search for very cool brown dwarfs, except, perhaps, in distant star clusters. However, JWST will certainly be capable of detailed spectroscopic observations of Y dwarfs lying within 10 parsecs of the Sun.

6.10 SUMMARY

The main conclusions of this chapter can be summarised as follows:

- The existence of brown dwarfs has now been amply confirmed through direct observations.
- Wide-angle near-infrared surveys (DENIS, 2MASS) and deep optical surveys (SDSS) are uncovering numerous examples of L- and T-type brown dwarfs in the general field.
- Theoretical models are becoming increasingly mature. Dust and cloud formation play important roles in the evolution of the atmospheric properties of these cool dwarfs.
- Brown dwarfs are relatively rare as companions to main sequence (or post main sequence) stars; approximately 15% of ultracool dwarfs are near equal-mass binaries with separations of 5–15 AU.
- In the next few years, mid-infrared satellites, such as Spitzer, are likely to find brown dwarfs with temperatures as low as 300 K, almost certainly requiring the addition of a new spectral class, ammonia-rich Y dwarfs.

6.11 REFERENCES

- B1 Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P. H., 1998, *A&A*, **337**, 403.
 B2 Barnbaum, C., Zuckerman, B., 1992, *ApJ*, **396**, L31.
 B3 Basri, G., Marcy, G. W., Graham, J. R., 1996, *ApJ*, **458**, 600.
 B4 Basri, G., Martin, E. L., 1998, in *Brown dwarfs and extrasolar planets*, ASP Conf. Ser. 134, eds R. Rebolo, E. L. Martin, M. R. Zapatero-Osorio, p. 284.
 B5 Bergeron, P., Wesemael, F., Lamontagne, R., Fontaine, G., Saffer, R., Allard, N. F., 1995, *ApJ*, **449**, 258.
 B6 Bildsten, L., Brown, E. F., Matzner, C. D., Ushomirsky, G., 1997, *ApJ*, **482**, 442.
 B7 Burrows, A., Marley, M., Hubbard, W. B., Lunine, J. I., Guillot, T., Saumon, D., 1997, *ApJ*, **491**, 856.
 B8 Burrows, A., Sharp, C. M., 1999, *ApJ*, **512**, 843.
 B9 Burgasser, A. J., 1999, *ApJ*, **522**, L65.
 B10 Burgasser, A. J., 2000, *ApJ*, **531**, L57.
 B11 Burgasser, A. J., McElwain, M. W., Kirkpatrick, J. D., Cruz, K. L., Tinney, C. G., Reid, I. N., 2004, *AJ*, **127**, 2856.
 B12 Burgasser, A. J., *et al.*, 2002, *ApJ*, **564**, 421.
 B13 Burrows, A., Bugasser, A. J., Kirkpatrick, J. D., Liebert, J., Milsom, J. A., Sudarsky, D., Hubeny, I., 2003, *ApJ*, **573**, 394.
 B14 Burgasser, A. J., Kirkpatrick, J. D., Liebert, J., Burrows, A., 2003, *ApJ*, **594**, 510.

- B15 Burgasser, A. J., Liebert, J., Kirkpatrick, J. D., Gizis, J. E., 2002, *AJ*, **123**, 2744.
- B16 Bailer-Jones, C. A. L., 2004, *A&A*, **419**, 703.
- B17 Bailer-Jones, C. A. L., 2002, *A&A*, **389**, 963.
- B18 Berger, E., *et al.*, 2001, *Nature*, **410**, 338.
- B19 Berger, E., 2002, *ApJ*, **572**, 503.
- B20 Beuzit, J.-L., *et al.*, 2004, *A&A*, **425**, 997.
- B21 Bouy, H., *et al.*, 2003, *AJ*, **126**, 1526.
- B22 Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., Brown, M. E., Misky, C. L., Gizis, J. E., 2003, *ApJ*, **586**, 512.
- B23 Bessell, M. S., 1991, *AJ*, **101**, 662.
- B24 Burrows, A., Sudarsky D., Lunini, J. I., 2003, *ApJ*, **596**, 587.
- B25 Becklin, E. E., Zuckerman, B., 1988, *Nature*, **336**, 656.
- C1 Clemens, J. C., van Kerkwijk, M. H., Wu, Y., 2000, *MNRAS*, **314**, 220.
- C2 Cochran, W. D., Hatzes, A. P., Hancock, T. J., 1991, *ApJ*, **380**, L35.
- C3 Cruz, K. L., Reid, I. N., Liebert, J., Kirkpatrick, J. D., Lowrance, P. J., 2003, *AJ*, **126**, 2421.
- C4 Chabrier, G., Baraffe, I., Allard, F., Hauschildt, P., 2000, *ApJ*, **542**, 642.
- C5 Close, L. M., Siegler, N., Freed, M., Biller, B., 2003, *ApJ*, **587**, 407.
- C6 Close, L. M., Potter, D., Brandner, W., Lloyd-Hart, M., Liebert, J., Burrows, A., Siegler, N., 2002, *ApJ*, **566**, 1095.
- C7 Chauvin, G., *et al.*, 2004, *A&A*, **425**, L29.
- D1 Dahn, C. C., Harris, H. L., *et al.*, 2002, *AJ*, **124**, 1170.
- D2 Delfosse, X., *et al.*, 1997, *A&A*, **327**, L25.
- D3 Delfosse, X., Forveille, T., Beuzit, J.-L., Udry, S., Mayor, M., Perrier, C., 1999, *A&A*, **344**, 897.
- E1 Eggen, O. J., Greenstein, J. L., 1965, *ApJ*, **141**, 83.
- F1 Fegley, B., Lodders, K., 1996, *ApJ*, **472**, L37.
- F2 Freed, M., Close, L., Siegler, N., 2003, *ApJ*, **584**, 453.
- F3 Forrest, W. J., Skrutskie, M. F., Shure, M., 1988, *ApJ*, **330**, L119.
- G1 Gizis, J. E., Reid, I. N., 1996, *AJ*, **110**, 1248.
- G2 Graham, J. R., Matthews, K., Neugebauer, G., Soifer, B. T., 1990, *ApJ*, **357**, 216.
- G3 Graham, J. R., McCarthy, J. K., Reid, I. N., Rich, R. M., 1990, *ApJ*, **357**, L21.
- G4 Gratton, R. G., Carretta, E., Castelli, F., 1997, *A&A*, **314**, 191.
- G5 Greenstein, J. L., 1969, *ApJ*, **158**, 281.
- G6 Greenstein, J. L., 1988, *AJ*, **95**, 1494.
- G7 Geballe, T., *et al.*, 2002, *ApJ*, **564**, 466.
- G8 Golimowski, D. A., *et al.*, 2004, *AJ*, **127**, 3516.
- G9 Gelino, C. R., Marley, M. S., Holtzman, J. A., Ackerman, A. S., Lodders, K., 2002, *ApJ*, **577**, 433.
- G10 Goldman, B., *et al.*, 1999, *A&A*, **351**, L5.
- G11 Gizis, J. E., Kirkpatrick, J. D., Wilson, J. C., 2001, *AJ*, **121**, 2185.
- G12 Golimowski, D. A., Henry, T. J., Krist, J. E., *et al.*, 2004, *AJ*, **128**, 1733.
- G13 Gizis, J. E., Reid, I. N., Knapp, G. R., Liebert, J., Kirkpatrick, J. D., Koerner, D. W., Burgasser, A. J., 2003, *AJ*, **125**, 3302.
- G14 Gorlova, N. I., Meyer, M. R., Rieke, G. H., Liebert, J., 2003, *ApJ*, **593**, 1074.
- H1 Haas, M., Leinert, Ch., 1990, *A&A*, **230**, 87.
- H2 Hale, A., 1995, *PASP*, **107**, 22.
- H3 Hamilton, D., Stauffer, J. R., 1993, *AJ*, **107**, 1855.
- H4 Harrington, R. S., Kallarakal, V. V., Dahn, C. C., 1983, *AJ*, **88**, 1038.

- H5 Henry, G. W., Baliunas, S. L., Donahue, R. A., Soon, W. H., Saar, S. H., 1997, *ApJ*, **474**, 503.
- H6 Henry, T. J., 1994, Ph.D. thesis, University of Arizona.
- H7 Hawley, S. L., Covey, K. R., Knapp, G. R., *et al.*, 2002, *AJ*, **123**, 3409.
- J1 Jones, H. R. A., Longmore, A. J., Jameson, R. F., Mountain, C. M., 1994, *MNRAS*, **267**, 413.
- K1 Kirkpatrick, J. D., Henry, T. J., Liebert, J., 1993, *ApJ*, **406**, 701.
- K2 Kirkpatrick, J. D., McGraw, J. T., Hess, T. R., Liebert, J., McCarthy, D. W., 1994, *ApJS*, **94**, 749.
- K3 Kirkpatrick, J. D., Henry, T. J., Irwin, M. J., 1997, *AJ*, **113**, 1421.
- K4 Kirkpatrick, J. D., Allard, F., Bida, T., Zuckerman, B., Becklin, E. E., Chabrier, G., Baraffe, I., 1999, *ApJ*, **519**, 834.
- K5 Kirkpatrick, J. D., *et al.*, 1999, *ApJ*, **519**, 802.
- K6 Kleimann, S. J., *et al.*, 1994, *ApJ*, **436**, 875.
- K7 Kuchner, M. J., Koresko, C. D., Brown, M. E., 1998, *ApJ*, **508**, L81.
- K8 Kumar, S. S., 1963, *ApJ*, **137**, 1121.
- K9 Kirkpatrick, J. D., *et al.*, 2000, *AJ*, **120**, 447.
- K10 Kirkpatrick, J. D., Dahn, C. C., Monet, D. G., Reid, I. N., Gizis, J. E., Liebert, J., Gizis, J. E., Burgasser, A. J., 2001, *AJ*, **121**, 3235.
- K11 Koerner, D. W., Kirkpatrick, J. D., McElwain, M. W., Bonaventura, N. R., 1999, *ApJ*, **526**, L25.
- L1 Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M., Burki, G., 1989, *Nature*, **339**, 38.
- L2 Leggett, S. K., Hawkins, M. R. S., 1989, *MNRAS*, **238**, 145.
- L3 Lodders, K., 1999, *ApJ*, **519**, 793.
- L4 Liebert, J., Kirkpatrick, J. D., Cruz, K. L., Reid, I. N., Burgasser, A., Tinney, C. G., Gizis, J. E., 2003, *ApJ*, **125**, 343.
- L5 Lane, B. F., Zapatero-Osorio, M. R., Britton, M. C., Martin, E. L., Kulkarni, S. R., 2001, *ApJ*, **560**, 390.
- L6 Luhman, K. L., 2004, *ApJ*, **614**, 398.
- L7 Leggett, S. K., *et al.*, 2000, *ApJ*, **536**, L35.
- M1 McCarthy, D. W., Probst, R. G., Low, F. J., 1985, *ApJ*, **290**, L9.
- M2 Magazzu, A., Martin, E. L., Rebolo, R., 1993, *ApJ*, **404**, L17.
- M3 Marley, M. S., Saumon, D., Guillot, T., Freedman, R. S., Hubbard, W. B., Burrows, A., Lunine, J. I., 1996, *Science*, **272**, 1919.
- M4 Monet, D. G., *et al.*, 1998, *USNO A2.0 catalogue*.
- M5 McCaughrean, M. J., Close, L. M., Scholz, R.-D., Lenzen, R., Biller, B., Brandner, W., Hartung, M., Lodieu, N., 2004, *A&A*, **413**, 1029.
- M6 Marley, M. S., Seagar, S., Saumon, D., Lodders, K., Ackerman, A. S., Freedman, R. S., Fan, X., 2002, *ApJ*, **568**, 335.
- M7 Martin, E. L., Brandner, W., Bouvier, J., Luhman, K. L., Stauffer, J., Basri, G., Zapatero Osorio, M. R., Barrado y Navuesce, D., 2000, *ApJ*, **543**, 299.
- M8 Martin, E. L., Brandner, W., Basri, G., 1999, *Science*, **283**, 1718.
- M9 McGovern, M. R., Kirkpatrick, J. D., McLean, I. S., Burgasser, A. J., Prato, L., Lowrance, P. J., 2004, *ApJ*, **600**, 1020.
- M10 Maxted, P. F. L., Marsh, T. R., Moran, C. J., 2000, *MNRAS*, **319**, 305.
- N1 Nakajima, T., Durrance, S. T., Golimowski, D. A., Kulkarni, S. R., 1994, *ApJ*, **428**, 797.

- N2 Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., Golimowski, D. A., Matthews, K., Durrance, S. T., 1995, *Nature*, **378**, 463.
- N3 Nather, R. E., Winget, D. E., Clemens, J. C., Hansen, C. J., Hines, B. P., 1990, *ApJ*, **361**, 909.
- N4 Noll, K. S., Geballe, T. R., Marley, M., 1997, *ApJ*, **489**, L87.
- N5 Noll, K. S., Geballe, T. R., Leggett, S. K., Marley, M., 2000, *ApJ*, **541**, L75.
- N6 Nakajima, T., Tsuji, T., Yanagisawa, K., 2004, *ApJ*, **607**, 499.
- O1 Oke, J. B., *et al.*, 1995, *PASP*, **107**, 375.
- O2 Oppenheimer, B. R., Kulkarni, S. R., Matthews, K., Nakajima, T., 1995, *Science*, **270**, 1478.
- O3 Oppenheimer, B. R., Basri, G., Nakajima, T., Kulkarni, S. R., 1998, *AJ*, **113**, 296.
- O4 Oppenheimer, B. R., Golimowski, D. A., Kulkarni, S. R., Matthews, K., Nakajima, T., Creech-Eakman, M., Durrance, S., 2001, *AJ*, **121**, 2189.
- P1 Patterson, J., Zuckerman, B., Becklin, E. E., Tholen, D. J., Hawarden, T., 1989, *ApJ*, **374**, 330.
- P2 Perrier, M., 1987, *ApJ*, **317**, L27.
- P3 Patience, J., *et al.*, 2002, *ApJ*, **581**, 654.
- P4 Potter, D., Martin, E. L., Cushin, M. C., Baudoz, P., Brandner, W., Guyon, O., Neuhauser, R., 2002, *ApJ*, **567**, L133.
- P5 Probst, R. G., Liebert, J., 1983, *ApJ*, **274**, 245.
- P6 Patience, J., Ghez, A. M., Reid, I. N., Matthews, K., 2002, *AJ*, **123**, 2570.
- R1 Rebolo, R., Martin, E. L., Maguzzu, A., 1992, *ApJ*, **389**, L83.
- R2 Rebolo, R., Zapatero Osorio, M. R., Madruga, S., Bejar, V. J. S., Arribas, S., Licandro, J., 1998, *Science*, **282**, 1309.
- R3 Reid, I. N., *et al.*, 1999, *ApJ*, **521**, 613.
- R4 Reid, I. N., Kirkpatrick, J. D., Gizis, J. E., Dahn, C. C., Monet, D. G., Williams, R. J., Liebert, J., Burgasser, A. J., 2000, *AJ*, **119**, 369.
- R5 Robinson, E. L., Cochran, A. L., Cochran, W. D., Shafter, A. W., Zhang, E., 1990, *AJ*, **99**, 672.
- R6 Ruiz, M. T., Leggett, S. K., Allard, F., 1997, *ApJ*, **491**, L107.
- R7 Reid, I. N., Gizis, J. E., Kirkpatrick, J. D., Koerner, D. W., 2001, *AJ*, **121**, 489.
- R8 Reid, I. N., Gizis, J. E., 1997, *AJ*, **113**, 2246.
- R9 Reid, I. N., Gilmore, G. F., 1981, *MNRAS*, **196**, P15.
- S1 Saffer, R. A., Livio, M., Yungelson, L. R., 1998, *ApJ*, **502**, 394.
- S2 Simons, D. A., Henry, T. J., Kirkpatrick, J. D., 1996, *AJ*, **112**, 2238.
- S3 Skrutskie, M. F., Forrest, W. J., Shure, M. A., 1987, *ApJ*, **317**, L55.
- S4 Stauffer, J., Hamilton, D., Probst, R., Rieke, G., Mateo, M., 1989, *ApJ*, **344**, L21.
- S5 Stauffer, J., Hamilton, D., Probst, R. G., 1994, *AJ*, **108**, 155.
- S6 Stauffer, J., Liebert, J., Giampapa, M., Macintosh, B., Reid, I. N., Hamilton, D., 1994, *AJ*, **108**, 160.
- S7 Strauss, M., *et al.*, 1999, *ApJ*, **522**, L61.
- S8 Scholz, R.-D., McCaughrean, M. J., Lodieu, N., Kuhkbodt, B., 2003, *A&A*, **398**, L29.
- S9 Siegler, N., Close, L. M., Mamajek, E. E., Freed, M., 2003, *ApJ*, **598**, 1265.
- T1 Tarter, J. C., 1974, Ph.D. thesis, University of California at Berkeley.
- T2 Telesco, C. M., Joy, M., Sisk, C., 1990, *ApJ*, **358**, L17.
- T3 Thackrah, A., Jones, H., Hawkins, M. R. S., 1997, *MNRAS*, **284**, 507.
- T4 Tinney, C. G., 1993, *ApJ*, **414**, 279.
- T5 Tinney, C. G., 1996, *MNRAS*, **281**, 644.
- T6 Tinney, C. G., 1998, *MNRAS*, **296**, L42.

- T7 Tinney, C. G., Reid, I. N., 1998, *MNRAS*, **301**, 1031.
 T8 Tokunaga, A. T., *et al.*, 1988, *ApJ*, **332**, L71.
 T9 Tokunaga, A. T., Becklin, E. E., Zuckerman, B., 1990, *ApJ*, **358**, L21.
 T10 Tokunaga, A. T., Kobayashi, N., 1999, *AJ*, **117**, 1010.
 T11 Tsuji, T., 1964, *Ann. Tokyo Obs. ser. II*, **9**, 1.
 T12 Tsuji, T., Ohnaka, K., Aoki, W., Nakajima, T., 1996, *A&A*, **308**, L29.
 T13 Tsvetanov, Z. I., *et al.*, 2000, *ApJ*, **531**, L61.
 T14 Tsuji, T., Nakajima, T., Yanagisawa, K., 2004, *ApJ*, **607**, 511.
 T15 Tinney, C. G., Delfosse, X., Forveille, T., 1997, *ApJ*, **490**, L95.
 V1 Vogt, S., *et al.*, 1994, *Proc. S.P.I.E.*, **2198**, 362.
 V2 Vrba, F. J., *et al.*, 2004, *AJ*, **127**, 2048.
 V3 van Biesbroeck, G., 1944, *AJ*, **51**, 61.
 V4 van Biesbroeck, G., 1961, *AJ*, **66**, 528.
 W1 Weidemann, V., 1990, *ARAA*, **28**, 103.
 W2 Winget, D. E., *et al.*, 1990, *ApJ*, **357**, 630.
 W3 Wilson, J. C., Kirkpatrick, J. D., Gizis, J. E., Skrutskie, M. F., Monet, D. G., Houck, J. R., 2001, *AJ*, **122**, 1989.
 W4 Wolf, M., 1918, *Astr. Nacht.*, **206**, 237.
 Z1 Zuckerman, B., Becklin, E. E., 1987, *Nature*, **330**, 138.
 Z2 Zuckerman, B., Becklin, E. E., 1992, *ApJ*, **386**, 260.
 Z3 Zuckerman, B., Reid, I. N., 1999, *ApJ*, **505**, L143.
 Z4 Zuckerman, B., Song, I., 2004, *ARAA*, **42**, 685.

6.12 HOMEWORK PROBLEMS

Problem 1

LP 944-20 (M9), 2MASS 1527-16 (L5) and Gl 229B (T6) are all members of the 8-parsec sample listed in the Appendix. Using those dwarfs as templates, estimate the volume searched for each spectral type by surveys based on POSS II (northern hemisphere), 2MASS (all sky) and SDSS (π steradians). Discuss the optimum search technique(s) in each case.

Problem 2

You observe an object in the field, using the Double Imaging Spectrograph with the low-resolution grating at Apache Point Observatory (the DIS/LOW configuration), giving about 9 Å spectral resolution. The resulting spectrum is shown in Figure 6.24. Is this a brown dwarf? Describe your reasoning and what further observations, if any, are needed to answer the question.

Problem 3

A mid-type L dwarf is monitored photometrically over 3 nights and shows sporadic 0.03–0.05 magnitude variations in the *I*-band. These might be due to activity (spots), variations in the atmospheric dust content or an eclipsing companion. Explain what additional observations you would make to distinguish among those possibilities.

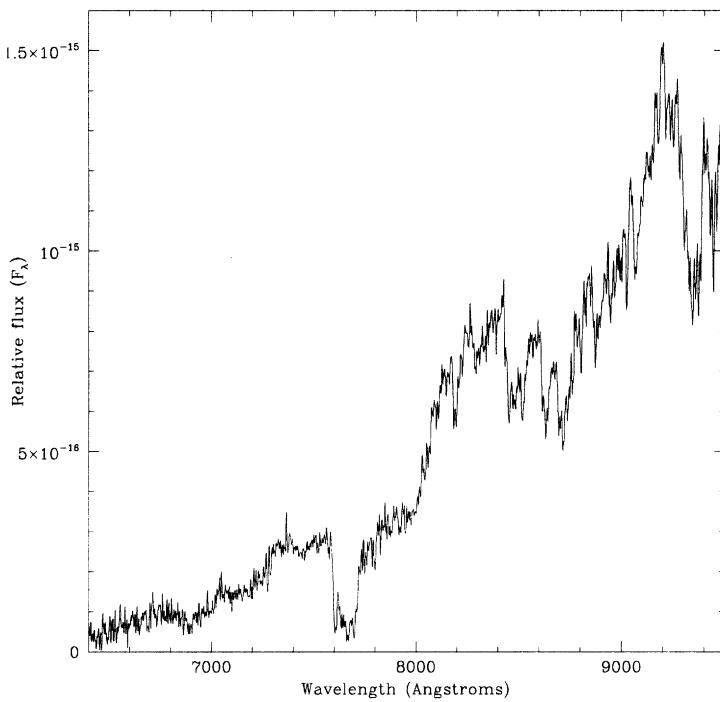


Figure 6.24. A far-red optical spectrum of an ultracool dwarf: see Problem 2.

7

A Galactic structure primer

7.1 INTRODUCTION

The first chapters of this book consider M dwarfs and brown dwarfs as individual objects, and use observations of their properties to study the relevant astrophysics. Low-mass stars are also potentially highly effective as probes of Galactic structure, since they are the most populous constituent of each stellar population. Until recently, their utility in that role has been restricted, both by difficulties in interpreting their emergent spectra in terms of luminosities, temperatures and abundances; and by their low intrinsic luminosities, which limit observations to nearby stars. However, with the development of the atmosphere models described in Chapter 4, together with substantial improvements in instrumental efficiency and the consequent availability of extensive, statistically well-defined samples, it is now becoming possible to better exploit their potential.

Detailed observations of late-type M dwarfs are limited to stars within 100 parsecs of the Sun, but early-type M dwarfs are accessible at distances of 2–3 kiloparsecs. The latter observations reach the transition region between the two main stellar populations of the outer Galaxy – the disk and the halo. Moreover, even though surveys of late-type stars may be confined to M dwarfs that are *currently* within a few hundred parsecs of the Sun, the majority of those stars have completed many Galactic orbits, and may have been formed at Galactic radii well separated from the current location of the Sun. Thus, local samples are still representative of global Galactic properties.

Our goal in this chapter is to provide context for the statistical studies of M dwarfs described in the following chapters. We do so by first reviewing the origins of the stellar populations concept and its contribution to our understanding of Galactic structure. Subsequent sections outline the basic properties of the disk and halo populations. The succeeding chapters describe how M dwarfs are used to probe aspects of Galactic structure, notably the underlying mass distribution.

7.2 STELLAR POPULATIONS

The realisation that stars in our Galaxy could be divided into stellar populations – groups of stars with well-characterised kinematic, spatial and chemical abundance distributions – marks one of the crucial steps in our understanding of galaxy formation. We now recognise that those distinct properties reflect significant differences in conditions prevalent during the formation and subsequent evolutionary histories of those groups of stars. Presenting the Galaxy as a combination of a small number of basic building blocks, rather than as a collection of 100 billion suns, is a striking simplification that opens the way toward theoretical investigation of the mechanisms underlying galaxy formation. However, like all simplifications, the population concept can obscure rather than illuminate, either through the broad, large-scale picture hiding important small-scale details, or through its over-use in a proliferation of physically almost-indistinguishable subcomponents.

The population concept owes its origin to Baade [B1], [B2], and excellent reviews of its development are given by Blaauw [B13] and Sandage [S5]. Sandage provides an illuminating account of how Baade's hypothesis grew through synthesis of results obtained over a period of thirty years and drawn from two areas of observational research: studies of high-velocity Galactic stars, originating with Charlier [C9]; and spectroscopic and photometric observations of extragalactic systems by Curtis, Hubble, Humason and others. The latter observations revealed significant differences in the structure of spiral and elliptical galaxies, with the ellipticals staunchly refusing to resolve into the individual stars clearly evident in the disks of the nearer spirals. It seemed that the brightest stars in the early-type systems (ellipticals) were both fainter and redder than the blue stars visible in the arms of spiral systems.

The identification of large numbers of high-velocity stars led to the realisation that there was a preferred direction of motion relative to the Sun – a consequence, we now understand, of the relative mean rotation of the disk and halo. Analysing data for several types of stars, Strömgren [S19], [S20] codified this velocity difference as the ‘asymmetric drift’ (preferred motion towards $l = 270^\circ$, $b = 0^\circ$, the direction opposite to Galactic rotation), and also showed that there was a significant correlation between the square of the U-velocity dispersion (radial from the Galactic centre) and the extent of the Strömgren drift. Lindblad [L5] and Oort [O2] demonstrated that the latter feature is a natural consequence of the distribution of kinetic energy between ordered rotation and velocity dispersion (pressure support) in kinematically-separate sub-systems: the lower the energy in rotation, the higher the velocity dispersion, and *vice versa*. The correlation is now usually plotted as rotational lag versus the total velocity dispersion (Figure 7.1).

Sandage has emphasised the key role played in this saga by Shapley's [S14], [S15] discovery of the nearby Sculptor and Fornax dwarf ellipticals. These were not only resolved into individual stars, but Baade and Hubble [B4] were able to identify RR Lyrae stars and, in Fornax, globular clusters associated with the system. This marked the first step towards drawing comparisons between the stellar content in Galactic globulars and ellipticals. The clinching observations came with Baade's

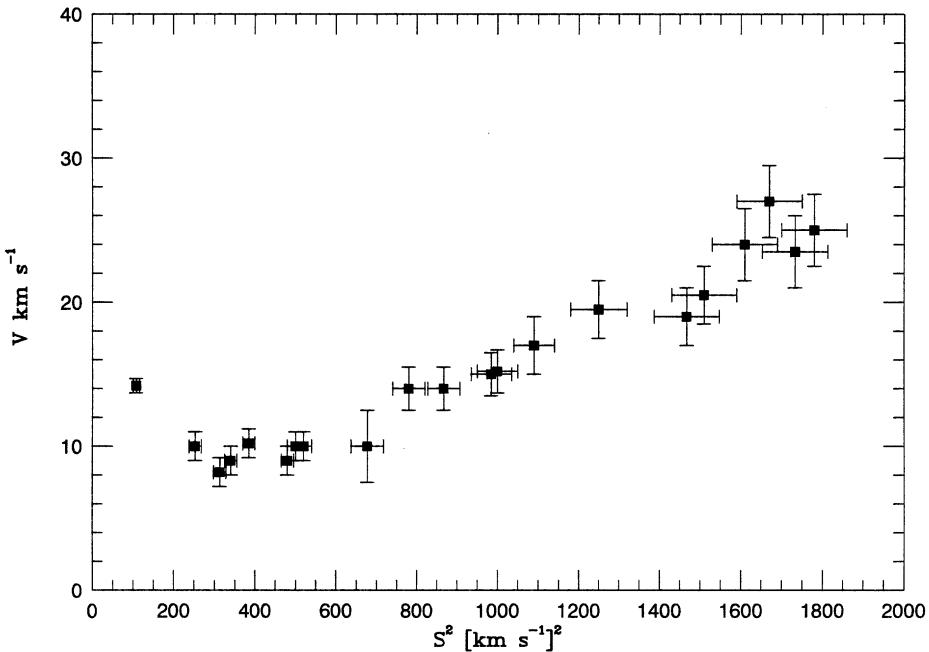


Figure 7.1. Strömgberg asymmetric drift: V is the mean rotational lag relative to the Sun and S^2 the total velocity dispersion ($\sigma_u^2 + \sigma_v^2 + \sigma_w^2$) for groups of stars from the *Hipparcos* catalogue. (Data from [D2].)

wartime photographic studies of the Andromeda spiral galaxy (M31) and its satellite dwarfs, M32, NGC 205, NGC 147 and NGC 185, made using the 100-inch Hooker telescope on Mount Wilson. For the first time, the brightest stars were resolved in both the central bulge of M31 and in the dwarf elliptical companions. Those stars proved to be K giants, with luminosities comparable to the brightest stars in Galactic globular clusters. Based on the similarity in the Strömgberg drift derived for globulars, RR Lyrae stars, short-period ($P < 210$ days) Mira stars and the high-velocity stars, Baade concluded that those Galactic objects are representative of the dominant stellar population in both spiral bulges and elliptical systems. In his own words [B2]: ‘This leads to the further conclusion that the stellar populations of the galaxies fall into two distinct groups, one represented by the well-known H–R diagram of the stars in the Solar Neighbourhood (the slow-moving stars), the other by that of the globular clusters. Characteristic of the first group (type I) are highly luminous O- and B-type stars and open clusters; the second (type II), short-period Cepheids [RR Lyrae stars] and globular clusters. Early-type nebulae (E-Sa) seem to have populations of the pure type II. Both types seem to co-exist in the intermediate and late-type nebulae. The two types of stellar populations had been recognised among the stars of our own Galaxy by Oort as early as 1926.’

From the outset, Baade realised that studies of nearby high-velocity stars represented the only available method of probing the characteristics of the lower

luminosity dwarf members of his Population II. Spectroscopic observations of the so-called subdwarfs and evolved high-velocity stars, such as RR Lyrae stars, had already revealed peculiarities, notably line-strengths weaker than expected for typical stars of the appropriate spectral type. These abnormalities were generally assumed to reflect different atmospheric conditions rather than different chemical composition. With neither a well-developed theory of stellar evolution, nor understanding of nucleosynthesis, it is not surprising that the standard paradigm in the early twentieth century was that all stars had abundances similar to those observed in the Sun.

Sandage identifies two papers which solidified the dichotomy in the characteristics of the disk and halo populations. Chamberlain and Aller's [C8] spectral analysis of the two archetypal subdwarfs, HD 19445 and HD 140283, produced the first metal abundances that were substantially below those found in the Sun;¹ and Roman [R8] demonstrated that subdwarfs share the common photometric characteristic of having bluer ($U-B$) colours (an 'ultraviolet excess') than disk dwarfs with the same ($B-V$) colour. Moreover, the extent of the ultraviolet excess was apparently dependent on the size of the Strömgren drift [R5]. Combined with observations of stars in globular clusters, these results provided a stronger link between the field subdwarfs and the Galactic globular cluster system, characterising them as members of a low-rotation, high-velocity-dispersion, metal-poor stellar population – the Galactic halo.

Baade originally proposed only two distinct stellar populations. However, the expanding empirical database accumulated through the 1950s revealed complexities, particularly among the local Population I stars, that were difficult to accommodate in this simple picture. The result was the development of a more detailed classification scheme at the 1957 Vatican conference on stellar populations [O1]. Three populations and five sub-populations were defined, each with its own distinct kinematics and abundance distribution, as summarised in Table 7.1. The original Population II and Population I were both subdivided, into respectively, the 'Halo' and 'Intermediate Pop II'; and the 'Extreme' (young) and 'Older Pop I'. The 'Disk' population was added as a component between the 'Intermediate Pop II' and the 'Older Pop I'. Various types of stars which were regarded as representative tracers of each component are included in Table 7.1.

These revisions represented a philosophical transition from the 'either/or' dichotomy of Baade's original scheme to a series of sub-populations with continuously varying properties, reminiscent of Lindblad's [L5], [L6] division of the Galaxy into kinematic sub-systems. Moreover, for the first time this classification was cast in

¹ Chamberlain and Aller's published results indicate abundances of 0.1–0.03 solar for the various elements studied, while both stars actually have less than one-hundredth the abundance of the Sun. Sandage reports that the initially-derived metallicities were indeed much lower, but given the prevailing climate of opinion of the time (as expressed by the journal referee), the temperature of each star was adjusted slightly, moved by either internal or external prompting, to give somewhat more palatable results (see also Chapter 8 in [C13]). HD 140283 is now known to be a halo subgiant, rather than a subdwarf.

Table 7.1. The Vatican stellar population scheme.¹

	Halo Pop. II	Intermediate Pop. II	Disk
Tracers	Subdwarfs Globulars RR Lyrae stars, $P > 0.4$ days	Stars with $ z > 30 \text{ km s}^{-1}$ Long-period variables Periods < 250 days	Galactic nucleus stars Planetary nebulae RR Lyrae stars, $P < 0.4$ days Weak-line stars
$\langle z \rangle$ (parsecs)	2,000	700	400
$\langle Z \rangle$ (km s^{-1})	75	25	17
Axial ratio, c/a^2	0.5	0.2	0.04?
Central concentration	Strong	Strong	Strong?
Distribution ³	Smooth	Smooth	Smooth?
Age (10^9 years)	6	5–6	1.5–5
	Older Pop. I	Extreme Pop. I	
Tracers	A-type stars Strong-line stars dMe stars	Gas Spiral structure Supergiants Cepheids, T Tauri	
$\langle z \rangle$ (parsecs)	160	120	
$\langle Z \rangle$ (km s^{-1})	10	8	
Axial ratio, c/a^2	?	0.01	
Central concentration	Weak	Weak	
Distribution ³	Patchy Spiral arms	Very patchy Spiral arms	
Age (10^9 years)	0.1–1.5	< 0.1	

¹ From [B13].² a is the semi-major axis measured in the Plane of the Galaxy; c is the semi-major axis of the density distribution perpendicular to the Plane; hence, c/a is the axial ratio viewed from an external location on the extension of the Plane.³ The total mass associated with each component in this model is $1.6 \times 10^{10} M_\odot$ for the halo; $4.7 \times 10^{10} M_\odot$ for the intermediate Population II and disk combined; $5 \times 10^9 M_\odot$ for the older Population I; and $2 \times 10^9 M_\odot$ for extreme Population I.

terms of an evolutionary scenario, with halo stars marking the first stage of star formation in the Galaxy, and extreme Population I the most recent. The remaining components represent intermediate stages of galaxy formation, with a broad trend between age and properties such as velocity dispersion and spatial distribution.

The revised, more gradualistic classification scheme devised at the Vatican conference provides a better approximation to the overall distribution of stellar properties. Paradoxically, however, repercussions from the codification into five sub-populations have led to almost as much confusion as illumination, particularly in recent years.

Dividing and classifying objects, establishing order in the face of apparent disorder, plays a prominent role during the early stages of many scientific disciplines. Classification schemes lay the foundation for subsequent understanding of why differences and similarities occur. Stellar spectral classification exemplifies the success possible with this approach.

Problems arise, however, when boundaries between adjacent sub-groups are less than clearly defined, as is the case with the Vatican system. While mean properties such as abundance and rotational velocity are different for each component, the distributions overlap, particularly among the disk and population I components. The result is that many individual stars cannot be assigned uniquely to one component. Applying slightly different criteria, such as emphasising kinematics over abundance, can lead to different effective definitions of a population, even though each is identified by name as the same component. All four of the spatially-flattened sub-populations in the Vatican scheme are subject to these vagaries of interpretation.

Given those concerns, the current best option in stellar population studies is to return to the simplicity of Baade's scheme, dividing the Galaxy at moderate (~ 1 kpc) to large radii into two major components, the disk and the halo, with the Galactic Bulge as a centralised, third component. The disk is less homogeneous than the halo, including stars spanning a significant range of age, kinematics and abundance, but whether this dispersion is the product of Galactic gradualism (that is, continuous evolutionary processes), or reflects the presence of discrete sub-populations, remains a subject of some debate. Section 7.5.3 presents some recent evidence that points to the latter alternative.

The simplest method of appreciating the relative extent of the different Galactic components is through observations of nearby external galaxies. Figure 7.2 presents an image of the edge-on spiral galaxy NGC 4565, an Sb system generally regarded as similar to our own Milky Way. The central bulge and the disk, with a prominent absorbing dust lane, are obvious; the halo population is not. While deep images have identified moderate numbers of globular clusters in this and other spirals, the field halo population lies below the current detection level in all external spiral galaxies. Based on observations of tracers such as metal-poor RR Lyrae stars within our own Galaxy, it probably describes a spherical distribution extending to radii beyond the edge of the NGC 4565 disk.

7.3 THE GALACTIC HALO

The stellar halo is an old, metal-poor population, characterised by little or no current star formation. The most prominent members of the halo are globular clusters, ~ 150 systems each consisting of 10^5 to 10^6 stars concentrated within radii of 10–30 parsecs. Figure 7.3 shows a typical example, the intermediate abundance cluster M13 in the constellation of Hercules. The sheer number of stars in these systems renders them excellent templates of the sparse field halo population, which comprises only $\sim 0.25\%$ of the stars in the Solar Neighbourhood.

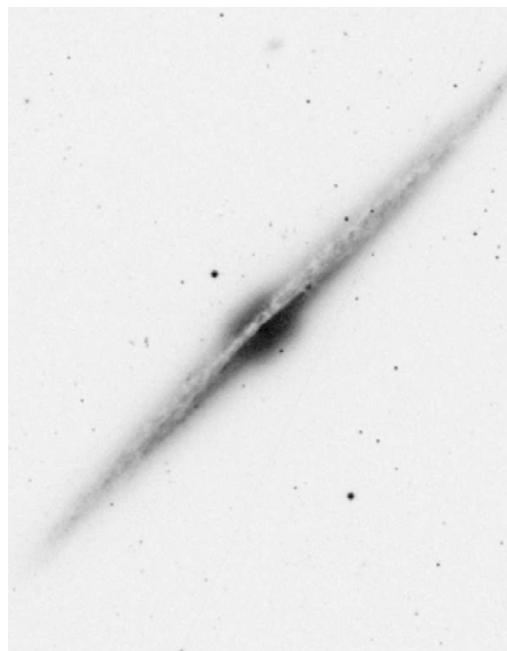


Figure 7.2. The edge-on spiral galaxy NGC 4565. (From the STScI Digital Sky Survey scans of a plate taken as part of the second Palomar Sky Survey, courtesy of Palomar Observatory and STScI.)

Figure 7.4 plots colour–magnitude diagrams for two globulars: M68 (from [W1]) at $[Fe/H] = -2.0$ [C6], one of the lower-abundance systems; and M5 [R2], an intermediate-abundance system with $[Fe/H] = -1.1$. In both clusters, the main sequence turn-off lies between $M_V = +3.5$ and $+4$, corresponding to a mass of $\sim 0.8 M_\odot$ and ages of $\sim 11\text{--}13$ Gyr. Evolved stars traverse well-populated subgiant and red giant branches (hydrogen shell-burning) before dropping onto the horizontal branch after the ignition of helium core-burning at the tip of the first giant branch (the helium flash). Depending mainly on mass and abundance, a horizontal branch star may evolve through the instability strip and pulsate as an RR Lyrae variable before initiating double shell-burning (hydrogen and helium burning). Subsequent evolution takes the star up the second, or asymptotic, giant branch, a phase terminated by envelope ejection and planetary nebula formation. (Further details on the evolution of these metal-poor, solar-type stars are given in the textbooks referenced in Chapter 3.)

Halo stars with masses approaching the hydrogen-burning limit can be identified within the Solar Neighbourhood. These are M subdwarfs, discussed in detail in Chapter 10. Studies of the halo population at large heliocentric distances are, of necessity, based on the more luminous population tracers, notably horizontal branch (HB) stars, particularly RR Lyrae stars, and globular clusters.



Figure 7.3. The globular cluster M92, from a plate taken with the 48-inch Oschin Schmidt telescope. (Courtesy of Palomar Observatory.)

Finally, a word about dark matter: the term ‘halo’ has been abducted by cosmologists to refer to the dark matter component that is held responsible for approximately 90% of the mass of galaxies like the Milky Way. The main observational evidence in favour of dark matter, flat rotation curves in spiral galaxies, requires that the distribution follow an R^{-2} density law, where R is the distance from the Galactic Centre. As described in Section 7.3.2, the stellar halo follows a much steeper density law. Moreover, the total mass of the stellar halo is $\sim 3 \times 10^9 M_\odot$, or approximately 0.67% the mass of the dark matter halo. The dark matter halo and the stellar halo are completely unrelated entities.

7.3.1 The metal-poor main sequence

The majority of halo stars in the Solar Neighbourhood are hydrogen-burning main sequence stars that have intrinsically lower luminosities than solar-abundance dwarfs of the same ($B-V$) colour. The term ‘subdwarf’ was coined by Kuiper [K5] to describe stars that lay between disk dwarfs and white dwarfs in the H–R diagram. Their existence was first suggested by Adams and Joy [A1], who identified three weak-lined ‘A-type’ stars – HD 19445, HD 219617 and HD 140283 – as having

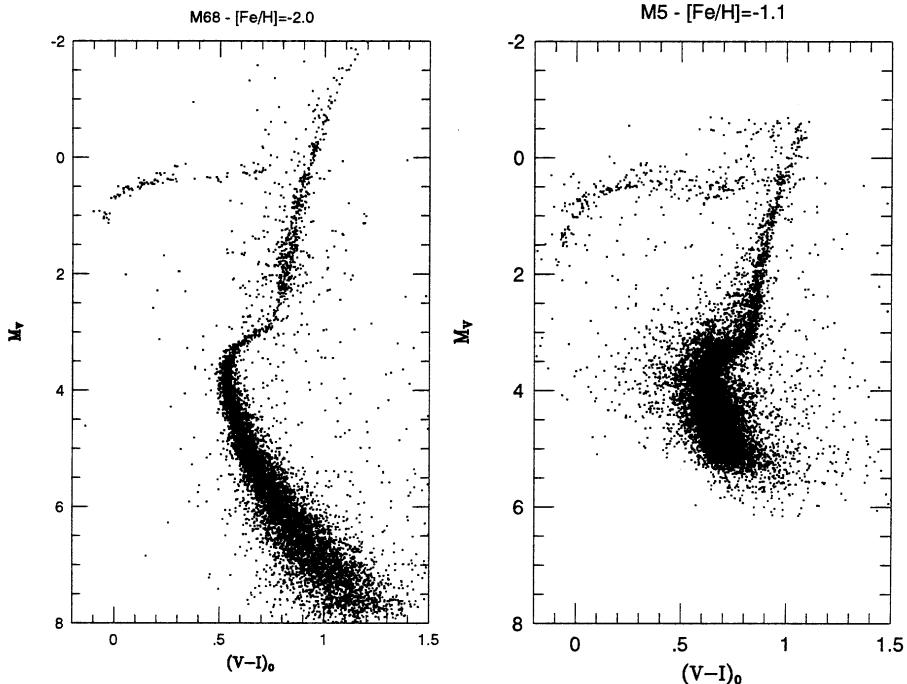


Figure 7.4. Colour–magnitude diagrams for the globular clusters M5 and M68. (M68 courtesy of A. Walker.)

unusual absolute magnitudes for their spectral properties. The latter two stars are now known to be extremely metal-poor subgiants.

By the 1950s, several other subdwarfs had been identified, but it remained unclear whether these stars were distinct in both the observational (M_V , $(B-V)$) and theoretical ((M_{bol}, T_{eff})) planes. Sandage and Eggen [S3] demonstrated that this was the case. As discussed in Chapter 4, decreased line and continuum absorption with decreased metal abundance leads to a star of given mass having a higher effective temperature and emitting a larger proportion of its flux in the optical régime (that is, a smaller bolometric correction, and a brighter M_V). Thus, while the subdwarf sequence is often characterised as lying ‘below’ the solar-metallicity main sequence in the colour–magnitude diagram, the true offset is primarily in colour (temperature), with lower abundance stars lying progressively further blueward of the disk dwarf sequence.

Theoretical models predict that the main sequence absolute magnitude of normal, single stars of a given colour (temperature) is a monotonic function of metallicity: the Russell–Vogt theorem outlined in Section 2.4. Until recently, few subdwarfs had both abundance determinations and parallaxes measured with sufficient accuracy to provide a strong test of that basic hypothesis. Observations from the ESA *Hipparcos* astrometric satellite project [E7] have improved the situation for F, G and early K-type subdwarfs. Figure 7.5 plots the (M_V , $(B-V)$) distribution

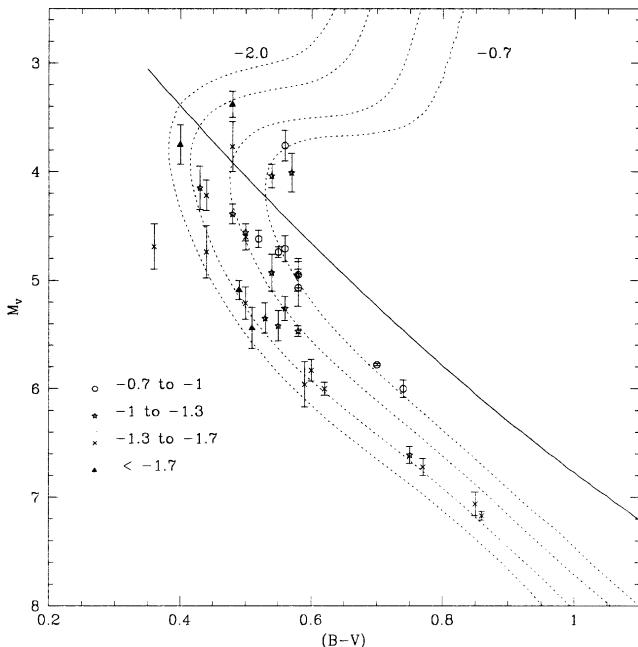


Figure 7.5. (M_V , $(B-V)$) data for subdwarfs with abundance determinations by [G3] or [A5], compared with 12-Gyr isochrones (dotted lines) predicted by [D1] for $[Fe/H] = -0.7, -1.0, -1.5$ and -2.0 . The disk main sequence is plotted as a solid line (data from [R4]).

described by subdwarfs and subgiants with $[Fe/H] < -0.7$ and parallaxes measured to a precision of at least 15% ($\sigma_\pi/\pi < 0.15$). The chemical abundances are derived from analysis of high-resolution echelle spectra [A5], [G3]. The mean (M_V , $(B-V)$) relationship defined by disk dwarfs in the immediate Solar Neighbourhood is plotted as a reference. By and large, data for FGK subdwarfs meet the theoretical expectations of a monotonic decrease in M_V with decreasing $[m/H]$.

7.3.2 The density distribution

The extended nature of the halo has been evident since Shapley's analysis of the globular cluster distribution. Globular clusters are large, luminous objects, and while a few distant ($R > 30$ kpc) or highly-obscured systems were identified in the last twenty years, most have been known since Dreyer's compilation of the New General Catalogue in 1885 (and many since Messier's 1785 catalogue of non-comets). The projected distribution in Galactic co-ordinates shows a clear concentration toward the Galactic centre, which led Shapley to formulate his Great Galaxy hypothesis. Distances can be estimated using a number of techniques, notably main sequence fitting, or calibrating the mean magnitude of non-variable horizontal branch stars and RR Lyrae stars. The latter variables had been identified in several clusters by Bailey in the later 1880s [B6], and Shapley's initial

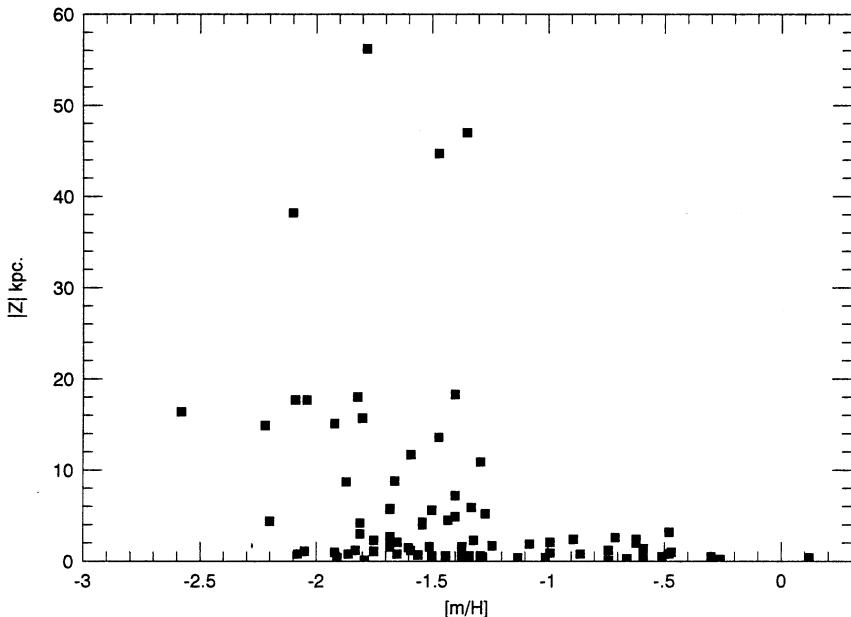


Figure 7.6. The present-day distribution of distance above the Galactic Plane of globular clusters, plotted as a function of abundance (from [Z2]). The low abundance systems extend to greater heights above the Plane.

distance-scale analysis was based on what we now know to be an erroneous RR Lyrae absolute-magnitude calibration. The inferred distance to the centre of the cluster distribution was ~ 20 kpc.²

Modern analyses of the inferred three-dimensional spatial distribution by Harris [H1] and Zinn [Z1] lead to the conclusion that the metal-poor ($[\text{Fe}/\text{H}] < -1$) clusters describe a near-spheroidal distribution, with the space density following a radial variation, $\rho(R) \propto R^{-3.5}$, where R is Galactocentric radius. Higher-abundance systems describe a much flatter spatial distribution (Figure 7.6), and have higher net rotation than the classical metal-weak clusters [Z2]. This apparent dichotomy has been confirmed by more recent observations [A3], and the two subsystems are often referred to as ‘halo’ (metal-poor, $[\text{m}/\text{H}] < -1$) and ‘disk’ clusters respectively. However, there are good grounds, notably the strong spatial concentration near the centre of the Galaxy, for associating the ‘disk’ system with the Galactic Bulge, rather than with the field-star disk population. This division may also be related to the two-component halo, discussed further in Section 7.3.6.

² Shapley’s Galaxy had a diameter about 30 times larger than the diameter of Kapteyn’s Universe, the prevailing model of the time. However, Shapley’s Milky Way was envisaged as a loose aggregation of sub-systems, with Kapteyn’s Universe being the local system, rather than as a coherent structure. In the so-called Great Debate, Shapley’s Great Galaxy was more accurate in scale, but Curtis’ Island Universe model was more accurate in morphology.

Globular clusters have completed as many as 100 Galactic orbits during their lifetimes. It is therefore inevitable that some systems, probably a substantial fraction of the initial population, have been disrupted through tidal interactions. The surviving systems are likely to be objects which have spent only a small fraction of the time in high-density regions near the Galactic Centre or close to the Galactic Plane. Thus, there is no guarantee that the present distribution of $\sim 10^5 M_\odot$ stellar aggregations provides an accurate reflection of the underlying field-star distribution, and it is important to supplement the cluster data with observations of field stars.

Blue horizontal branch (BHB) stars, especially RR Lyrae variables, are well suited as density tracers, since they are luminous, readily identifiable through their colours and/or variability; and their absolute magnitudes can be calibrated with some reliability. A complication is that the HB morphology depends on the core mass following red giant branch evolution, which varies as a function of metallicity and age [L3]. Metal-rich clusters, such as 47 Tucanae ($[m/H] = -0.7$), possess only a short, red horizontal branch, with few RR Lyrae stars. Moving to lower abundances, theoretical HB tracks for ages of 10 Gyr or more intersect the instability strip, and clusters such as M3 and M5 have significant numbers of RR Lyrae variables. Decreasing the abundance further leads to a continued migration towards higher temperatures until, at $[m/H] = -1.7$ to -1.9 , the horizontal branch lies fully blueward of the instability strip (an example being NGC 6397, $[m/H] = -1.8$). That trend reverses at lower abundances, mainly since lower mass-loss rates on the giant branch lead to higher core masses, and clusters such as M15 and M68 ($[m/H] \sim -2.1$) have nearly symmetric HB morphologies and well-populated instability strips.

In addition to the overall trend in morphology with metallicity, global analyses must address the ‘second parameter problem’: clusters of the same abundance have significantly different HB morphologies. This is exemplified by the intermediate-abundance pair M3 and M13, where M13 has an extended blue horizontal branch and almost no RR Lyrae variables, and by M92 and M68, where M92 has a more extended BHB than does M68. The origin of these differences remains a subject of debate (see, for example, [S8]), but age variations are usually indicted as responsible.

Regardless of the physical mechanism, these variations in HB morphology suggest that the density distribution of BHB stars might not match the underlying stellar population if the halo is chemically inhomogeneous. Fortunately, the latter does not appear to be the case. Most analyses derive density distributions consistent with power laws, $\rho(R) \propto R^{-n}$, with $2.5 < n < 3.5$ [K2], [C14], [C15], [S17]. Deep RR Lyrae surveys with SDSS suggest that the density law steepens sharply at distances of 50 kpc from the Galactic Centre [I3], comparable to the distance of the Large and Small Magellanic Clouds. It is likely that this marks the effective edge of the halo.

The shape of the halo is generally modelled as an axisymmetric spheroid. Most studies that are based on stars at distances of several kpc or more favour a relatively round system, with axial ratio $c/a \sim 0.9$ (where a is the semi-major axis in the Galactic Plane, and c , the axis perpendicular to the Plane). However, the structure appears more complex near the disk, with nearly one-half of the local subdwarf

population drawn from a more flattened component with $c/a \sim 0.55$ [H2], [S17]. There is little evidence for substantial metallicity differences between these components, but there are significant differences in the kinematics (see the following section). The existence of these two components, the far/outer halo and the inner halo, provides significant clues to the formation history of the Milky Way, as discussed further in Section 7.3.6.

7.3.3 Halo kinematics

Globular clusters, RR Lyrae variables, field horizontal branch stars, metal-poor giants and local subdwarfs have all been used to probe halo kinematics. In the case of the more luminous and more distant tracers, such as RR Lyrae variables and globular clusters, transverse motions are poorly determined. However, statistical techniques can be used to reconstruct the full three-dimensional motions from radial velocities, given a sample well-distributed on the sky.

Table 7.2 lists measurements of the halo kinematics, expressed as Schwarzschild velocity ellipsoids in the (U , V , W) co-ordinate system described in Chapter 1.

Table 7.2. The kinematics of the stellar halo.

	\bar{U}^a (km s $^{-1}$)	\bar{V}	\bar{W}	σ_U (km s $^{-1}$)	σ_V	σ_W	N_*	$\langle V_{\tan} \rangle^b$ (km s $^{-1}$)
Clusters ¹		-160		118	118	118	66	172
$\Delta S \geq 5$ RR Lyrae stars ²		-220		210	119	91	33	228
Halo RR Lyrae stars ³	9	-210	-12	168	102	97	162	209
Metal-poor stars ⁴		-200		153	93	107	887	201
Metal-poor stars ⁵	16	-217	-10	161	115	108	180	215
Subdwarfs ⁶		-180		130	105	85	452	177
Subdwarfs ⁷		-195		133	98	94	420	189
Subdwarfs ⁸		-20	-149	-4	140	100	67	33
Subdwarfs ⁹		-7	-181	-5	156	112	108	118
Subdwarfs ¹⁰		-20	-193	-3	152	104	95	150
Disk dwarfs ¹¹	-9	-22	-7	41	27	21	311	39

^a Typical uncertainties for the halo kinematics are 10–15 km s $^{-1}$.

^b The median tangential velocity for each sets of kinematics.

¹ Frenk and White [F3].

² Metal-poor RR Lyrae stars, from Oort [O3].

³ The Halo-3 statistical parallax analysis by Layden *et al.* [L2].

⁴ $[\text{Fe}/\text{H}] < -1.5$ field stars, Beers and Sommer-Larsen [B8].

⁵ Metal-poor giants and RR Lyrae stars, *Hipparcos* data, Chiba and Yoshii [C10].

⁶ Kinematically unbiased sample of field subdwarfs with $[\text{Fe}/\text{H}] < -1.2$, Norris [N3].

⁷ Kinematically unbiased sample of field subdwarfs with $[\text{Fe}/\text{H}] < -1.6$, Morrison *et al.* [M16].

⁸ Sommer-Larsen and Zhen [S17], local members of the inner halo.

⁹ Sommer-Larsen and Zhen [S17], local members of the outer halo.

¹⁰ Carney *et al.* [C5], nearby subdwarfs with $[\text{m}/\text{H}] < -1.5$.

¹¹ Nearby stars, single population, Hawley *et al.* [H3]: uncertainties ± 5 km s $^{-1}$.

Apparently the mean rotational velocity of the halo lags that of the Sun by 180–230 km s^{−1} which, since the net rotational velocity at the Solar Circle is estimated as between 200 and 220 km s^{−1}, implies that the halo has negligible mean rotation. The typical halo star has a substantial velocity relative to that of the Sun and subdwarfs make a disproportionately large contribution (relative to their space density) to catalogues of high proper-motion stars. As a result, proper motion surveys offer one of the most effective methods of identifying local members of the halo population (see Chapter 10).

There are indications, mainly from deep proper-motion surveys, that the net rotation of the halo decreases within increasing height above (or below) the Plane, and that the outer halo has significant retrograde rotation [M4], [C5]. In addition, there is growing evidence for kinematic sub-structure in the outer halo (i.e., systematic clumping in velocity phase space [M5], [H5]). The most likely explanation for this behaviour is that those stars share a common origin; they are tidal debris from disrupted satellite galaxies. The strongest evidence in favour of this hypothesis is the Sagittarius dwarf galaxy, which is currently merging with the Milky Way [I1]. Observations of M giants and carbon stars in the far halo have been used to trace tidal streams left by that system over its last four orbits of the Galaxy. The presence of this complex structure in the outer halo has clear implications for formation models; indeed, the outer halo may consist largely of debris from previous satellite mergers [M17].

7.3.4 The abundance distribution of the halo

Heavy elements originate from stellar nucleosynthesis. The first halo stars formed from protogalactic gas which was almost entirely H and He. The halo abundance distribution reflects how those stars and succeeding generations polluted their environment. Chapter 4 outlined the physical principles underlying the measurement of stellar metallicities, [m/H]. This section summarises the applications of those principles to study the chemical composition of halo subdwarfs.

Abundances are often expressed as [Fe/H] – the iron-to-hydrogen abundance relative to the solar value expressed in logarithmic units. The equivalence [m/H] ≡ [Fe/H] is valid for the majority of disk stars, where the relative abundances of individual elements remain approximately constant: that is, if iron is deficient by a factor of two relative to the Sun ([Fe/H] = −0.3), then the same holds for Mg, Ca, O and so on. That is not the situation for most halo stars. A number of species, notably the so-called α -elements (O, Ca, Ti, Mg, Si, S), have abundances which are enhanced in most³ subdwarfs with [Fe/H] < −1 (Figure 7.7). This has implications for formation timescales, as discussed in Section 7.3.5.

Since most halo stars have non-solar abundance ratios, the terms ‘metallicity’ and ‘[Fe/H]’ cannot be used interchangeably, as with disk dwarfs. Abundances are

³ As with most statistical correlations, this behaviour does not hold for *every* star. A number of subdwarfs are known with both [Fe/H] < −1 and solar element ratios, and with [Fe/H] ∼ −0.5 and enhanced [O/Fe] [E1].

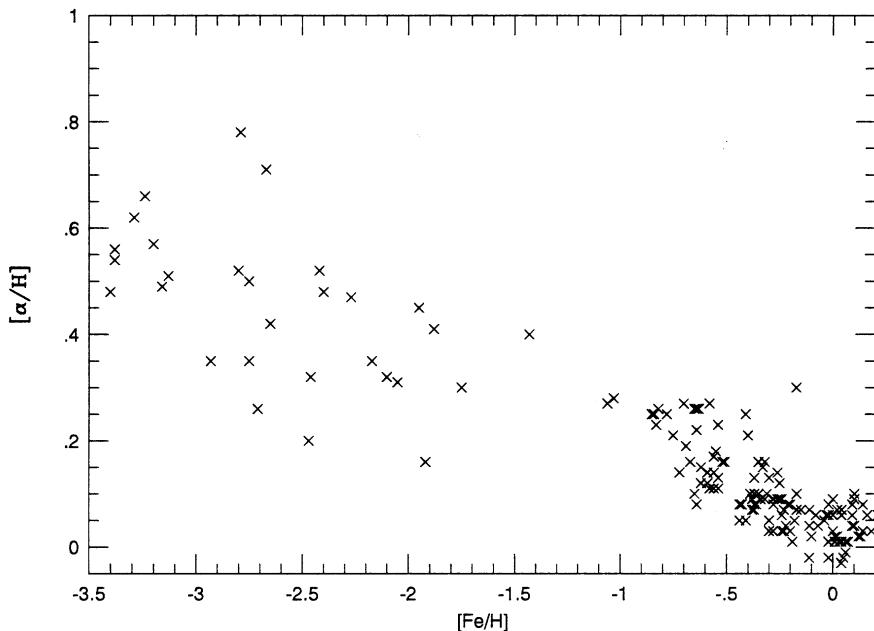


Figure 7.7. The variation in the α -element/iron abundance ratio as a function of iron abundance. (Data from [E1].)

therefore usually written as $[m/H]$, which can be approximately related to the iron abundance by

$$Z_M = Z_{Fe}(0.638 \times f_\alpha + 0.362) \quad (7.1)$$

where Z_{Fe} is the iron abundance (in linear units), f_α the α -element enhancement and Z_M the overall effective metal abundance [S1]. This is a useful relation, since many abundance analyses are based specifically on measuring the strength of Fe absorption lines.

Measuring abundances

Most studies of the halo abundance distribution are based on observations of either F, G subdwarfs or K-type red giants. In both cases, a variety of photometric and spectroscopic techniques are used to estimate $[m/H]$. The first studies [S2] centred on Roman's nearby high-velocity F, G subdwarfs, and used UBV photometry to quantify ultraviolet excess as

$$\delta(U-B) = (U-B)_{obs} - (U-B)_{std} \quad (7.2)$$

where $(U-B)_{std}$ is the colour of a solar-abundance dwarf with the same $(B-V)$ colour. Figure 7.8 illustrates the basis for this technique: the $(U-B)$ colour measures the relative flux emitted over the $3,300\text{--}3,900\text{\AA}$ and $4,100\text{--}4,900\text{\AA}$ regions. Metal-poor stars have weaker absorption lines, less line blanketing, and therefore more ultraviolet flux than solar-abundance dwarfs with the same $(B-V)$ colour (temperature).

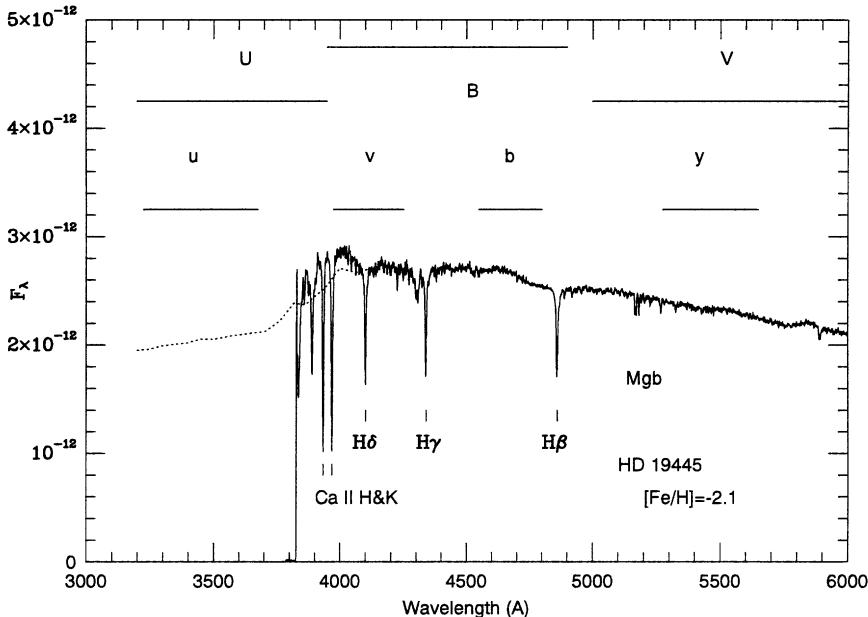


Figure 7.8. The UBV and $uvby$ photometric passbands matched against the spectral energy distribution of the F-type subdwarf, HD 19445. The solid line is the measured spectrum; the dotted line shows the extension to shorter wavelengths.

Figure 7.9 shows the $((U-B), (B-V))$ diagram outlined by photometric standards supplemented by data for subdwarfs with abundances $[m/H] < -1.5$.

The Strömgren $uvby$ photometric system [S21], [C12] provides a finer-tuned measurement of the same spectral characteristic, with the index

$$m_1 = (v - b) - (b - y) \quad (7.3)$$

measuring line blanketing in the violet relative to the blue–visual. As with $\delta(U-B)$, relative abundances are determined by comparing the metallicity index against the standard value at that temperature:

$$\delta m_1 = m_1(obs) - m_1(std) \quad (7.4)$$

In this case $(b - y)$ serves as the temperature indicator and the $m_1(\text{std})$ calibration is defined by the disk dwarf ($m_1, (b - y)$) relation [N1]. Strömgren indices have been measured for several thousand F, G and early K stars in both disk and halo.

In a similar manner, intermediate-band photometric systems have been devised to measure relative abundances of K-type stars. The two systems used most frequently are the DDO system [M1] and the Washington system [C3]. The former has six filters of bandwidth $\sim 100 \text{ \AA}$, centred at $\lambda\lambda 3,500, 3,800, 4,100, 4,200, 4,500$ and $4,800 \text{ \AA}$, usually supplemented by the ‘51’ filter, centred on the Mgb lines (Figure 7.10). The latter system employs four filters, $CMT_1 T_2$, with bandwidths close to $1,000 \text{ \AA}$ and centred at $\lambda\lambda 3,910, 5,085, 6,330$ and $8,050 \text{ \AA}$. As with UBV and

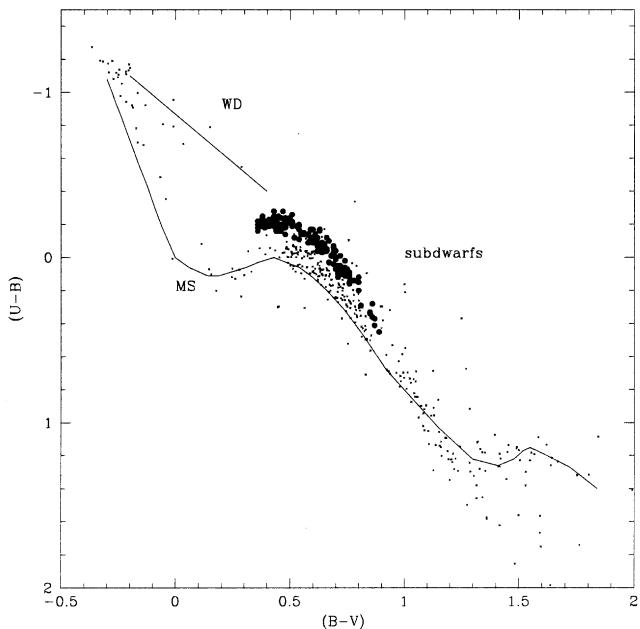


Figure 7.9. The $((U-B), (B-V))$ diagram. Crosses are photometric standards, predominantly disk stars; data for extreme subdwarfs are plotted as solid points. (Based on data from Carney *et al.* [C5].)

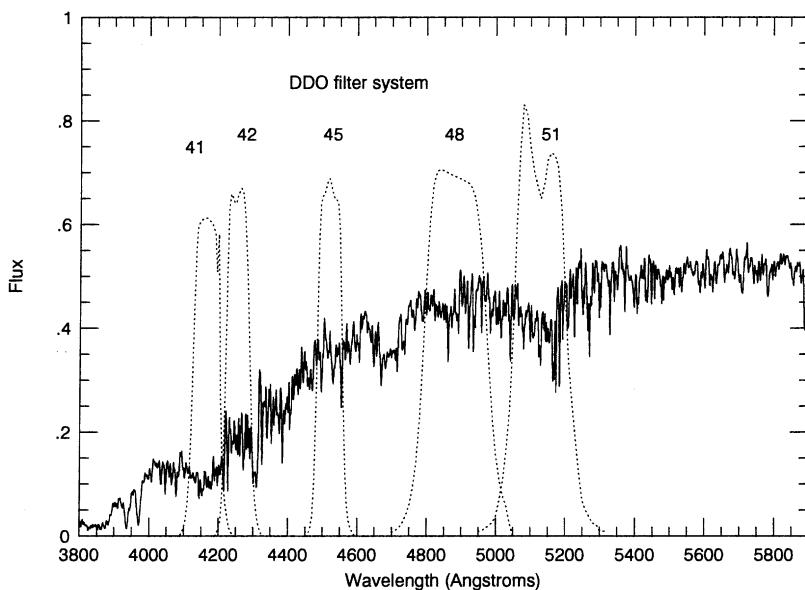


Figure 7.10. The passbands of the DDO photometric system matched against the spectrum of a K0 disk giant.

uvby data, relative abundances are determined by using one colour as a temperature estimator (usually DDO colour $C(45-48)$ and the Washington colour ($M-T_2$)) and matching an abundance-sensitive index against a standard relationship. The two systems are often used in conjunction, with temperature estimated from the broader-band Washington data and abundances from DDO indices. As Figure 7.10 shows, the DDO colour difference $C(41-42)$ provides a measure of the strength of CN absorption, while $C(48-51)$ measures Mg b/Mg H strength.

Considered in isolation, photometric indices provide only a relative ranking of stellar abundances. Setting those measurements on an absolute abundance scale requires detailed analysis of observations at higher spectral resolution. G dwarfs have a wealth of features at blue/visual wavelengths, notably Ca II H and K, the G-band (CH) and the Mg I Fraunhofer b triplet, as well as numerous atomic lines due to Fe. All vary significantly with decreasing abundance (Figure 7.11). The larger

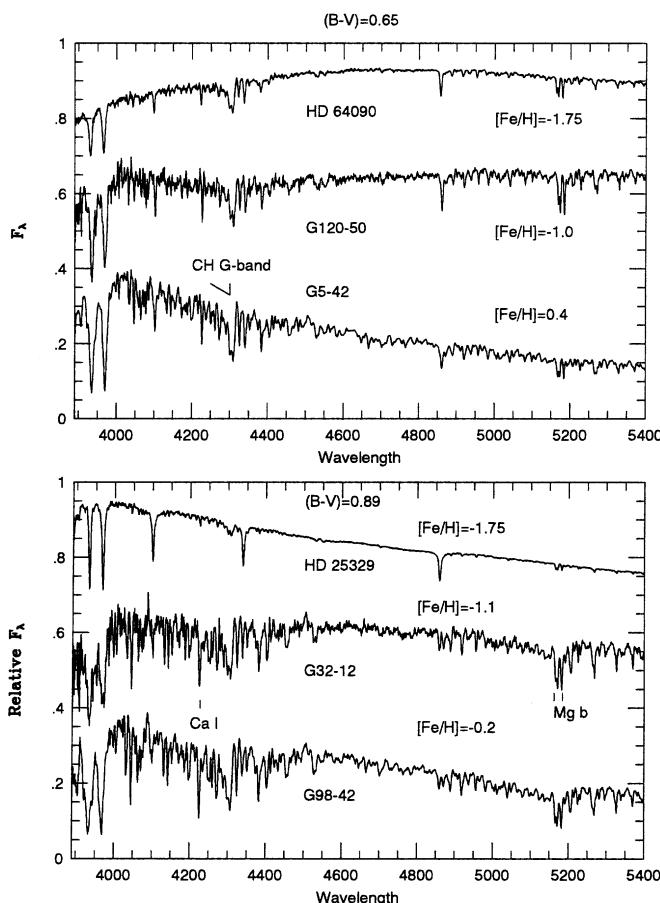


Figure 7.11. Spectra of early-type ($(B-V) = 0.65$) and late-type ($(B-V) = 0.89$) G dwarfs with $0 < [Fe/H] < -2$, showing the change in line strength with decreasing abundance.

variation in linestrengths between $[Fe/H] = -2$ and -1 (linear régime) than between -1 and solar abundance (flat or damped regime) illustrates the principle of curve of growth analysis (Figure 4.4). Narrowband spectrophotometric indices can be devised to measure specific features in low resolution spectra [R10], or individual line profiles can be analysed at spectroscopic resolutions $>20,000$ [A5], [G3], [C4]. Modern abundances are derived either through direct line analyses or by spectrum synthesis, matching the observed spectrum against theoretical predictions for a range of $[m/H]$. Stars calibrated in this manner serve as standards for lower-resolution spectroscopic and photometric systems.

The halo abundance distribution

High spectral-resolution abundance analyses provide the most accurate metallicity measurements, but require high signal-to-noise spectra and a substantial investment of time on 4-m class (or larger) telescopes. Photometric and spectrophotometric observations require fewer resources, and those techniques therefore provide the majority of metallicity estimates. Measuring uncertainties are typically given as $[m/H] = \pm 0.2$, but systematic offsets of similar magnitude in both zero-point and scale exist between different systems or different analyses based on the same technique [R4], [C11], [H6]. The latter often reflect the particular choice of standard stars as primary calibrators.

With an increasing number of stars having high-resolution spectroscopic observations, lower-resolution systems should be calibrated in a self-consistent manner in the near future. At present, analyses rest to a substantial extent on older observations, which often remain uncorrected for systematic errors. In particular, the fundamental reference point – the iron abundance of the Sun – was revised in the early 1990s. The pre-1991 reference value for the solar iron abundance was $Fe = 7.67$, where the abundance is expressed in logarithmic units with $H = 12.0$; the revised value [B12] is $Fe = 7.52$. In principle, this should not affect abundance determinations, since $[Fe/H]$ is a logarithmic ratio (scaled relative to the Sun), not an absolute value. However, systematic offsets of similar magnitude are present in some data sets based on pre-1991 calibrations (e.g., [N1], [C4]).

Figure 7.12 compares the abundance distribution of globular clusters [Z2] against data for nearby subdwarfs [L1]. Both abundance scales probably underestimate $[m/H]$ by 0.1 to 0.3 for $[m/H] > -2$. Deciphering the detailed distribution at $[m/H] > -1$ is complicated by the presence of ‘disk’ globular clusters and metal-weak disk dwarfs respectively. However, both distributions indicate a modal abundance of $[m/H] \sim -1.5$, or a metallicity $\sim \frac{1}{30}$ that of the Sun.

The field subdwarf distribution in Figure 7.12 extends to lower abundances than the cluster distribution, perhaps reflecting the destruction of the oldest clusters and their consequent absence from the present-day census. Few extremely metal-poor stars were known until recently, but large-scale objective prism surveys, aimed specifically at detecting stars with very weak lines, have now succeeded in discovering ~ 100 objects with $[Fe/H] < -3.0$ [B9], [C17]. The two stars with the lowest measured abundances are both giants: CD-38:245, with $[m/H] = -4.5$ [B10]; and

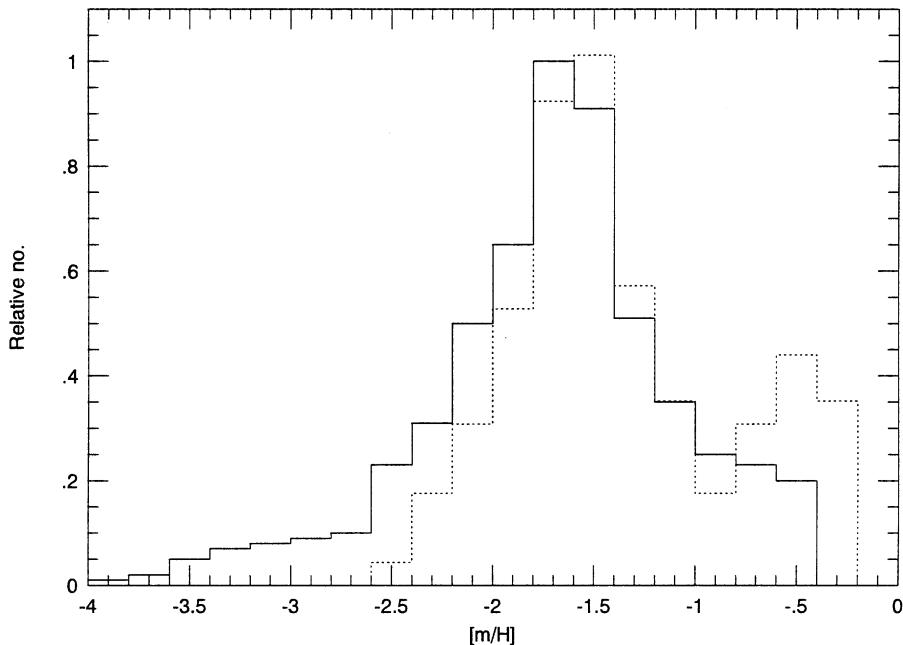


Figure 7.12. The halo abundance distribution. The solid line plots the field-star distribution from [L1]; the dotted line plots the cluster abundance distribution from [Z2].

HE 0107-5240, a recent discovery from the Hamburg/ESO objective prism survey, with $[m/H] = -5.3$ [C16], [C17]. Extremely metal-poor stars are expected to be rare: a mass of only several $\times 10^{-5} M_{\odot}$ in ‘metals’ per star is all that is required to raise abundances to $\frac{1}{1000}$ solar ($[m/H] = -3$) in a star-forming cloud. The mean abundance of the interstellar medium increased rapidly to $[m/H] = -2$ with the first generation of evolved stars; reaching higher abundances requires proportionately more mass in stellar ejecta, accumulated over longer timescales, which, in turn, allows more star formation. Thus, the number of stars increases with increasing abundance for $[m/H] < -1.4$. The decrease at higher abundances presumably marks the cessation of star formation in the halo. There is no evidence for significant variation in the mean abundance as a function of position in the halo.

7.3.5 The age of the Galactic Halo

This issue has two aspects: absolute age; and the range of ages spanned by members of the halo. The absence of OBA main sequence stars at large distances from the Galactic Plane indicates that star formation ceased at least 1 Gyr ago. Further quantifying that statement demands comparisons between observations and theoretical stellar models. The most effective method of deriving absolute ages is the determination of the mass of stars currently evolving onto the subgiant branch; that is, the mass at the main sequence turn-off. Although simple in principle, this

technique remains difficult to implement in practice. In contrast, more easily-calibrated methods are available for estimating the age distribution of halo constituents. We summarise those before discussing absolute ages.

Relative ages

Elemental abundances in field subdwarfs provide an important clue to the overall timescale of halo formation. At halo metallicities ($[m/H] < -1$), iron abundance (the abscissa in Figure 7.7) can be taken as a (non-linear) time-dependent variable. The α -elements form through nuclear reactions involving capture of α particles, which occur in type II supernovae – that is, supernovae with massive ($>8\text{--}9 M_{\odot}$) progenitors. In contrast, iron originates predominantly in type I supernovae, which form through thermonuclear runaway in accreting white dwarf stars in binary systems [I2]. The latter systems have evolutionary timescales of ~ 1 Gyr, since the intermediate-mass white dwarf progenitor must evolve through the red giant, Cepheid, asymptotic giant branch and planetary nebula phases, before forming a white dwarf; massive stars evolve to type II supernovae in a matter of a few million years. Thus, the high $[\alpha/\text{Fe}]$ ratios measured for halo stars suggest that most formed before substantial numbers of type I supernovae could contribute to the Galactic iron abundance; as type I supernovae became more common, the $[\alpha/\text{Fe}]$ ratio was reduced toward the solar value. This hypothesis implies that most of the Galactic halo formed in a span of only ~ 1 Gyr.

Globular cluster analyses suggest a similarly short formation period. Relative ages for clusters of similar abundance can be estimated using several techniques – notably by measuring the offset in colour between the turn-off and the base of the giant branch ($\Delta(B-V)_{TO}$ [V1], [S6]). The $(B-V)$ colour at the base of the giant branch is primarily a function of metallicity in old stellar systems (Figure 7.13). As the system ages, the main sequence turn-off moves to lower luminosities and redder colours, but the base of the giant branch maintains a constant colour, so the colour difference decreases. $\Delta(B-V)_{TO}$ can be measured directly from the observed colour–magnitude diagram, without requiring corrections for foreground reddening or knowledge of the exact cluster distance. Thus, this parameter provides a robust measurement of the cluster age *distribution* at a given metal abundance. Recent studies (see [S7]) indicate that the majority of clusters have similar ages, with a few ($\sim 10\%$) younger by 3–5 Gyr. Many of the younger clusters lie at Galactocentric radii exceeding 10 kpc (for example, Palomar 14 [S8]), a matter discussed further below.

Absolute age determinations

Absolute age measurements rest almost exclusively on globular cluster data, primarily comparison of the luminosity at the main sequence turn-off against evolutionary models. This procedure places stringent demands on both observation and theory. Observational requirements are accurate photometry; accurate measurement of foreground reddening, allowing determination of intrinsic colours and magnitudes; accurate abundance determination; and, above all, a reliable distance

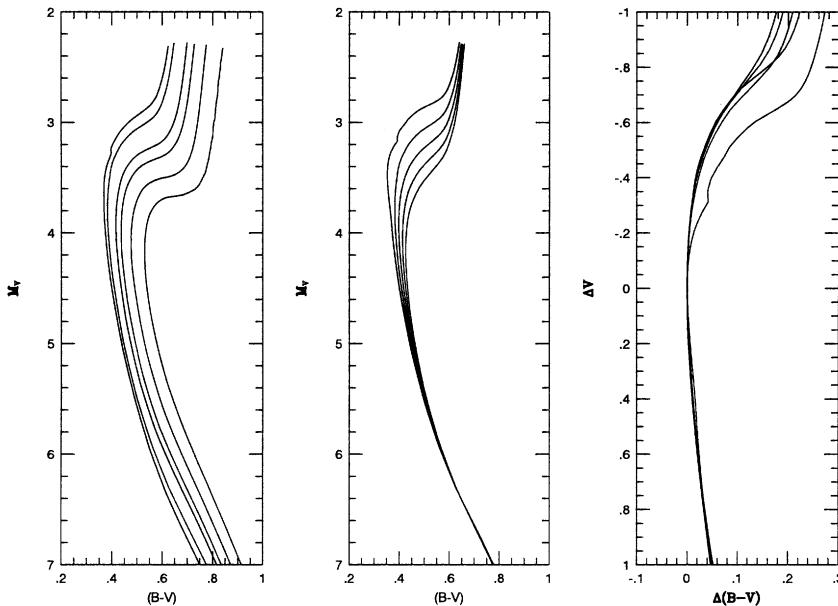


Figure 7.13. Theoretical isochrones for metal-poor globular clusters (from [D1]). (*Left*) 12-Gyr isochrones for abundances of $[Fe/H] = -0.7, -1.0, -1.3, -1.5, -2.0$ and -2.3 ; (*middle*) the evolution of an $[Fe/H] = -2.0$ system, plotting isochrones for age 10, 12, 14, 16 and 18 Gyr; (*right*) the change in the location of the giant branch with age at $[Fe/H] = -2$, scaling the isochrones to match at the main sequence turnoff.

estimate. Transforming the measured absolute magnitude, $M_V(TO)$ to an age requires accurate modelling of convection and helium diffusion; reliable opacities; an appropriate mix of chemical abundances; and the definition of the appropriate equation of state. As yet, no models fulfil all of these goals, but recent years have seen substantial improvements. Figure 7.13 presents the $(M_V, (B-V))$ colour-magnitude diagram predicted by the [D1] models for a range of age and metallicity.

The task of obtaining precise colour-magnitude data for globular clusters has been rendered straightforward by the availability of large-format CCDs on 4 m-class telescopes. Foreground reddening can be estimated to reasonable precision using far-infrared maps from either the IRAS or COBE satellites. Cluster distances are more problematic. The primary method of distance estimation is main sequence fitting [S4]: a fiducial sequence derived from the reddening-corrected cluster colour-magnitude diagram (that is, $(V_0, (B-V)_0)$) is matched against either model isochrones (the theoretician's route to distances) or an empirical $(M_V, (B-V)_0)$ sequence based on subdwarfs with accurate parallax measurements (the observer's route to distances). In either case, the offset $(V_0 - M_V)$ provides the distance estimate.

Both empirical and theoretical approaches to main sequence fitting must face significant obstacles to obtaining accurate distance estimates: in the former case, the calibration rests on only a handful of subdwarfs; the latter demands accurate models and accurate transformation from the theoretical to observational plane. Given these

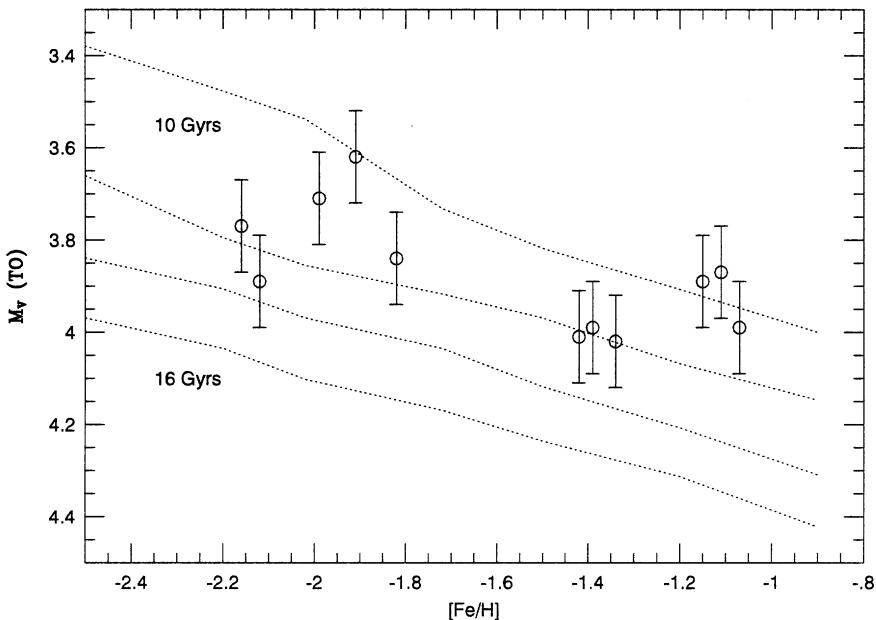


Figure 7.14. Age calibration for globular clusters: turn-off absolute magnitudes, deduced from main sequence fitting against subdwarfs with *Hipparcos* parallaxes, are matched against the predictions of the [D1] set of models. Ages in the range 11–14 Gyr are indicated (data from [R4]).

concerns, many studies (for example, [B15]) opt for semi-empirical methods, adjusting theoretical isochrones to match the few subdwarfs with accurate data. The availability of *Hipparcos* astrometry, and better parallaxes for a larger number of metal-poor stars, is improving matters.

Prior to 1995, most analyses favoured ages of 16 Gyr or more for low-abundance clusters such as M92 and M68, with intermediate-abundance systems, such as M5 or M13, 2 to 3 Gyr younger. Subsequently, improved stellar evolution models have become available which incorporate revised opacities and an improved equation of state [D1], [C7]. Compared to previous studies, the new models predict lower luminosities for main sequence stars of given mass, implying higher masses and younger ages for a given $M_V(\text{TO})$. Combined with *Hipparcos*-based main sequence fitting analyses [R4], which indicate an increase of $\sim 5\%$ in cluster distances (i.e., brighter $M_V(\text{TO})$), these new models reduce globular cluster age estimates to ~ 11 to 14 Gyr (Figure 7.14). There is no obvious trend towards younger ages with increasing abundance.

Summarising ages

Current analyses indicate that the bulk of the stellar halo was formed in a 1 to 2 Gyr burst of star formation $\sim 12.5 \pm 1.5$ Gyr ago. The presence of a small number of

younger globular clusters suggests that star formation continued at a reduced level for the succeeding 3 to 5 Gyr, primarily in the outer halo.

7.3.6 Forming the halo

The conclusions summarised in the previous subsections, concerning the kinematics, density distribution, abundance distribution and age distribution of the halo, provide baseline constraints for the two main models which have been proposed to describe the formation of the halo: the monolithic collapse model, originating with Eggen, Lynden-Bell and Sandage [E2]; and the fragmentary accretion model first proposed by Searle and Zinn [S13].

The ELS monolithic collapse model

The Eggen, Lynden-Bell and Sandage (ELS) model marks a landmark as the first serious analysis of galaxy formation. ELS combined kinematic and photometric data for 221 stars, drawn in almost equal numbers from two catalogues of, respectively, nearby stars and high-velocity stars. Reconstructing Galactic orbits from the space motions, the results showed that both orbital eccentricity and $|W|$, the velocity perpendicular to the Plane, were correlated strongly with the ultraviolet excess, $\delta(U-B)$: stars with high velocities and high eccentricities had stronger ultraviolet excess (that is, lower metallicity). ELS hypothesised that the observed relationships stemmed from an underlying correlation between age and abundance, with metal-poor stars having formed during the initial collapse of the largely gaseous protogalaxy. High-angular-momentum, metal-rich stars on nearly circular orbits (like the Sun) are denizens of the rotating disk which was the result of that collapse, chemically enriched by the metals produced by nucleosynthesis in massive stars of the first generation. ELS envisaged collapse by more than a factor of 10 in radius over a timescale of $\sim 10^8$ years or less. Since most subdwarfs form at large radius in this model, they are predicted to follow predominantly radial orbits, with large motions perpendicular to the Plane. This produces the observed ($\delta(U-B)$, eccentricity) and ($\delta(U-B)$, $|W|$) correlations.

In their original analysis, ELS proposed a rapid, dissipationless (non-collisional) collapse, which implies that there should be no abundance gradient in the resulting stellar halo. However, in a later analysis of a larger sample of subdwarfs, Sandage [S3] found evidence for a continuous increase in $|W|$ with decreasing $\delta(U-B)$. This, he argued, ‘shows that a chemical gradient exists in the Galactic halo such that stars with the largest $\langle |Z| \rangle$ (maximum attainable height above the Plane) have the lowest metal abundance. This gradient appears to be a natural consequence of the collapse of the halo towards the Plane, with the metal enrichment taking place as the collapse proceeds.’ That is, as the protogalaxy collapses, chemical enrichment from stellar mass-loss leads to increasing abundance in the interstellar medium, while star formation is confined within a volume of continuously decreasing radius. Each succeeding generation of star-forming clouds acquires higher angular momentum

as the collapse proceeds, leading to a smooth correlation between rotation and abundance.

This model, however, requires that the collapse be sustained over a time that is sufficient to allow the stellar ejecta from one generation of stars to become well-mixed in the gas clouds which form the succeeding generation. This demands a longer timescale than the $\sim 10^8$ years envisaged by ELS in their original model. Prompted by the observational results outlined by [S3], modified Galaxy-formation models were proposed where the collapse is slowed down by pressure support provided by winds from high mass OB stars and supernovae (see [Y1], for example). These models can prolong the collapse phase for several Gyr.

The SZ fragmentary accretion model

Searle and Zinn's [S13] model was conceived partly to address the complications encountered by the ELS model and partly as a result of more extensive observations of globular clusters in the outer halo. The latter revealed variations in the morphology of the colour–magnitude diagram (notably the horizontal branch) amongst clusters with similar metal abundance: the ‘second parameter’ problem, discussed in Section 7.3.2. This effect is particularly pronounced among the clusters at large Galactocentric radii. SZ proposed age as the second parameter, suggesting that the formation of the halo consisted of chaotic accretion of individual dwarf galaxy-sized (10^7 – $10^9 M_\odot$) gas clouds, rather than as a smooth monolithic collapse. Each accreted fragment has its own star-formation history, leading to a dispersion of properties amongst clusters and field stars once those fragments coalesce to form the halo. While most fragment accretion occurs during the initial 2 to 3 Gyr of the formation history, the Galaxy can continue to absorb low-mass systems at later epochs, providing a natural explanation for the presence of younger clusters at large radii.

A comparison of the ELS and SZ models

These models make different predictions about the expected properties of the halo population. The modified ELS model predicts a radial abundance gradient and a correlation between rotational velocity and abundance in response to the slow, unidirectional collapse. On the other hand, Searle and Zinn predict a dispersion in age- and abundance-dependent properties with radius, with no monotonic abundance gradient and no expectation of a correlation between rotation and abundance.

Observations tend to support the latter scenario. Halo star surveys undertaken subsequent to the [S3] analysis have resulted in the identification of halo stars with near-circular, high-angular-momentum orbits. Those stars escaped previous detection since they have low proper motions; their presence adds significant scatter to the smooth (velocity, $[m/H]$) correlations which underlie the ELS hypothesis. With the addition of these new data, the mean rotational velocity of halo subdwarfs appears to be independent of abundance for $[m/H] < -1.5$ (Figure 7.15). The weak correlation evident between $[m/H] = -1$ and -1.5 may stem from contamination by disk

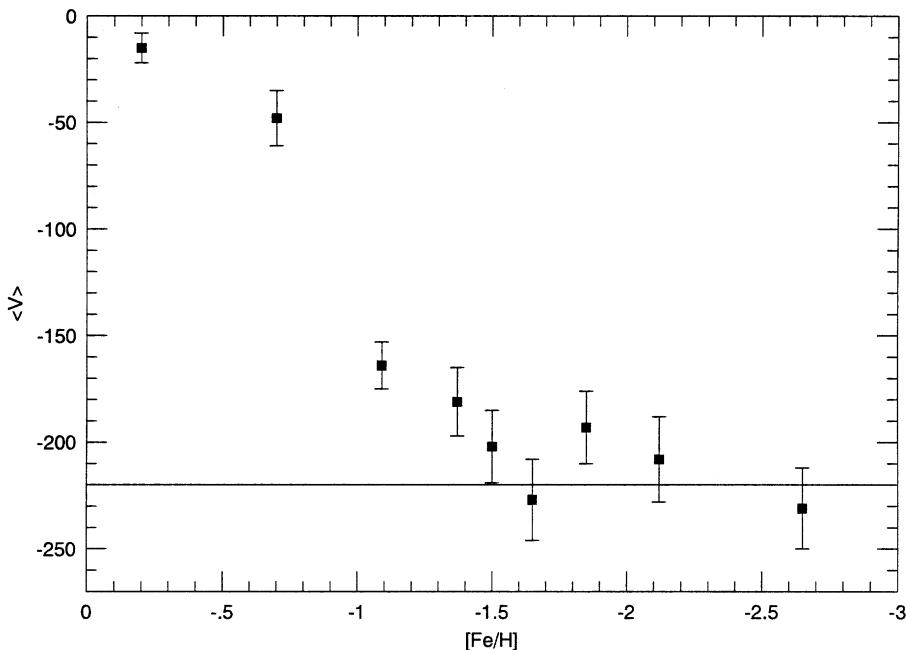


Figure 7.15. The correlation between rotation and abundance, from [C10] and [B8]. The near-constant rotation for $[m/H] < -1.5$ contradicts the predictions of halo models involving rapid, monolithic collapse.

stars. Most significantly, observations suggest that the outer halo has a net rotational velocity which is retrograde with respect to the motion of the Sun about the Galactic Centre [M4]. Clearly, this result cannot be accommodated within a monolithic collapse model. There is also direct evidence for the importance of accretion. Recent large-scale kinematic surveys have identified possible star streams in the halo, presumably remnants of disrupted fragments [M5], [J2].

In the final analysis, it is likely that both processes played a significant role in the formation of the Galaxy. Sandage has described the SZ scenario as ‘ELS with noise’. A plausible composite model might consist of a high-mass ($> 10^{10} M_{\odot}$) system undergoing monolithic ELS collapse to form the disk and inner halo, while simultaneously accreting smaller SZ fragments to form the less homogeneous outer halo. Indeed, the discovery of many interacting and merging extragalactic systems and the current engorging of the Sagittarius dwarf by our own Galaxy [I1] illustrates that the latter process continues to this day.

7.4 THE GALACTIC BULGE

M dwarfs in the Galactic Bulge have apparent magnitudes too faint to be accessible to detailed observations. However, for completeness, we include a brief summary of

the large-scale properties of that subsystem in our Galactic survey. A recent review of observational and theoretical studies is presented by Wyse *et al.* [W7].

Observational studies of the Bulge are complicated by the fact that the line-of-sight passes along the length of the disk and through substantial foreground obscuration. Far-infrared observations by the IRAS and COBE satellites provided the first clear picture of the large-scale structure of the Bulge, but questions remain about whether the central component is more closely related to the halo or the disk. In halo models, the Bulge forms during the initial collapse of the protogalaxy, either as the central component of the large-scale collapse in the ELS model, or as part of the inhomogeneous SZ merging-fragment model. The alternative is post-collapse formation. For example, Pfenniger and Norman [P4] have proposed that a strong central bar could first accumulate a concentration of stars, and then develop a bulge as gravitational resonance interactions drive stars out of the Galactic Plane.

Deciding amongst the various formation models is not straightforward. What are the constituents of the Galactic Bulge? The radically different colours of stars in, respectively, the disk and bulge of M31 provided the main stimulus for Baade's two-population hypothesis. His original identification of the Milky Way's Bulge as a metal-poor population was prompted by the discovery of RR Lyrae stars in the low-absorption field at ($l = 1^\circ$, $b = -4^\circ$) now known as 'Baade's window' [B3]. Arp's [A4] subsequent colour-magnitude work appeared to confirm that result, with the detection of a well-populated giant branch, red horizontal branch and a main sequence that was essentially truncated at G-type stars – morphological features clearly compatible with globular cluster characteristics. However, by the time of the Vatican conference, Morgan had shown that a significant component in the Bulge had strong-lined spectra, consistent with a near-solar abundance [M7], [M14], [M15]. Recent higher-resolution spectroscopic data confirm that the abundance distribution is comparable to that in the Solar Neighbourhood stars, although some stars have abundance ratio anomalies [M2]. In particular, the [Eu/Fe] ratio is enhanced in the Bulge stars, suggesting that Type II supernovae (sources of Eu) played a prominent role in the metal enrichment, and implying, as in the halo, a formation timescale of only ~ 1 Gyr.

One of the more extensive recent studies of the Bulge is Minniti's [M11], [M12] survey of the radial velocities and abundances of several hundred K giants in two fields at intermediate radii. His data show a strong correlation between these two parameters, with the $[\text{Fe}/\text{H}] < -1$ stars having no net rotation but a large line-of-sight velocity dispersion ($\sigma \sim 109 \text{ km s}^{-1}$), while the higher abundance stars have lower dispersion and significant mean rotation. The implication is that the former stars represent the innermost extension of the halo population, which also contributes Baade's RR Lyrae stars, while the metal-rich stars form a chemically- and kinematically-distinct component, perhaps related to the 'disk' globular clusters.

Synthesising these observations, it is likely that 'the Bulge' is a composite entity, including both the inner halo and a higher-abundance, rotationally-supported population. The latter probably formed rapidly, and may be a near-contemporary of the halo. The halo stars form during collapse, and therefore have radial orbits and little net rotation; enriched gas sinks towards the centre of the potential well during the

earliest phases of Galaxy formation, and, through dissipative collisions, forms a rotating system. Star formation within those rotating gas clouds produces Bulge stars. The issue, however, is far from settled.

7.5 THE GALACTIC DISK

The overwhelming majority of stars in the Solar Neighbourhood are members of the disk population. The total mass is estimated as $5 \times 10^{10} M_{\odot}$, over ten times the mass of the halo. The Sun lies in the outer disk between the Scutum and Perseus spiral arms, with the nearest star-forming regions (in Taurus) ~ 150 parsecs distant. The typical density of the inter-arm interstellar medium is ~ 0.5 atoms cm $^{-3}$, but the Sun resides within the ‘Local Bubble’, diameter ~ 100 parsecs, where the average density is lower by over a factor of 10 [F6], [P2]. Several other similar structures lie within 500 parsecs of the Sun, and are probably due to the effects of both winds from massive OB stars and blast-waves from supernovae [B15]. General star-counts (that is, the number of stars as a function of apparent magnitude) show that there are $\sim 10\%$ more stars towards the South Galactic Pole (SGP) than towards the North Galactic Pole (NGP). The simplest explanation for this asymmetry is that the Sun lies 35 parsecs above (north of) the mid-Plane of the disk [B16]. Finally, the Sun’s distance from the Galactic Centre (the Solar Radius) is usually taken as $R_{\odot} = 8.5 \pm 1.5$ kpc.

7.5.1 The kinematics of disk stars

The rapid rotation of the Galactic Disk is the result of conservation of angular momentum during Galactic collapse. The original protogalaxy had net rotation. Collisions between gas clouds during collapse dissipated energy in motion perpendicular to the present-day Galactic Plane, but preserved motion parallel to the Plane, producing ordered rotation. Disk stars show variation in mean kinematics, primarily velocity dispersion, as a function of spectral type. Table 7.3 presents results from a recent analysis of astrometric data obtained by the *Hipparcos* satellite [D2]. Stellar kinematics are usually expressed in the (U, V, W) velocity system (see Section 1.3.3), but there can be misalignments between the principle axes of the best-fit ellipsoid and the Galactic co-ordinate system. The parameters give in Table 7.3 for each group of stars are the velocity dispersions for the principal axes of the ellipsoid, with σ_1 closest to σ_U , σ_2 closest to σ_V , and σ_3 closest to σ_W . The vertex deviation, l_v , corresponding to the Galactic longitude of the direction of σ_1 , is also given. Finally, velocity dispersions in U, V and W for nearby M dwarfs (from [H2]) are listed for comparison [H3]. The solar motion derived from the *Hipparcos* analysis is

$$U_{\odot} = 10.00 \pm 0.32 \text{ km s}^{-1}$$

$$V_{\odot} = 5.23 \pm 0.62 \text{ km s}^{-1}$$

$$W_{\odot} = 7.17 \pm 0.38 \text{ km s}^{-1}$$

Table 7.3. Velocity dispersions of disk stars.

Spectral type	$(B-V)_{min}$	$(B-V)_{max}$	σ_1	σ_2	σ_3	l_v
B–A5	–0.24	0.14	14.4	9.6	5.5	30°3
B5–F0	0.14	0.31	20.2	9.6	8.1	22°8
F0–F2	0.31	0.41	22.4	11.9	9.4	19°8
F3–F5	0.41	0.47	26.3	15.9	12.2	10°2
F5–F7	0.47	0.53	30.5	18.5	13.4	6°8
F7–G0	0.53	0.58	33.0	21.9	15.1	1°9
G0–G3	0.58	0.64	37.7	23.6	21.3	10°2
G3–G5	0.64	0.72	38.2	24.0	20.9	7°6
>G5	0.72	1.55	37.3	26.1	18.3	13°1
M dwarfs	1.1	1.6	41	27	21	
Me dwarfs	1.1	1.6	27	20	15	
$\leq M3$	1.1	1.4	39	24	22	
$> M3$	1.4	1.6	42	29	22	
M dwarf g1	1.1	1.6	35	21	20	
M dwarf g2			52	36	32	

Data for spectral types B to G are from [D2]. M dwarf velocity dispersions are σ_U , σ_V , σ_W , from [H3], [G5]. g1 and g2 are velocity dispersions for the 2-Gaussian match to the latter dataset.

Table 7.3 shows that early-type stars have substantially lower velocity dispersions than GKM dwarfs. This behaviour was pointed out originally by Parenago [P1], and is referred to as ‘Parenago’s discontinuity’. Its origin lies in the dynamical evolution experienced by disk stars coupled with the decrease in main sequence lifetimes with increasing mass. Young stars emerge from the parent star-forming region with the space motion of the cloud, which generally lies within 10 km s^{-1} of the local circular velocity. As these stars orbit the Galaxy, they undergo gravitational interactions, particularly with massive (10^4 – $10^6 M_\odot$) molecular clouds, leading to random changes in motion, that is, scattering [J1]. Spitzer and Schwarzschild [S18] demonstrated that under such circumstances the overall velocity dispersion, σ_t , is expected to increase with $\sigma_t \propto \tau^{1/3}$, where τ is the age (see also [W3], [W4]). In particular, the scattering leads to dynamic heating of the disk, a higher velocity dispersion perpendicular to the Plane ($\sigma_3 = \sigma_W$ in Table 7.3) and a larger vertical scale height. The correlation between spectral type and σ_t follows: early-type stars have shorter main sequence lifetimes, and therefore, as a population, are subject to less scattering than GKM dwarfs with main sequence lifetimes comparable to, or exceeding the age of, the Galactic Disk.

The velocity dispersion/age dependence is also illustrated in comparing the kinematics of dM and dMe dwarfs. Since chromospheric activity declines relatively rapidly with time for early- and mid-type M dwarfs (Section 5.5.2), a sample of dMe dwarfs is biased towards stars younger than the average age of the disk. As Table 7.3 shows, those stars also have lower velocity dispersions.

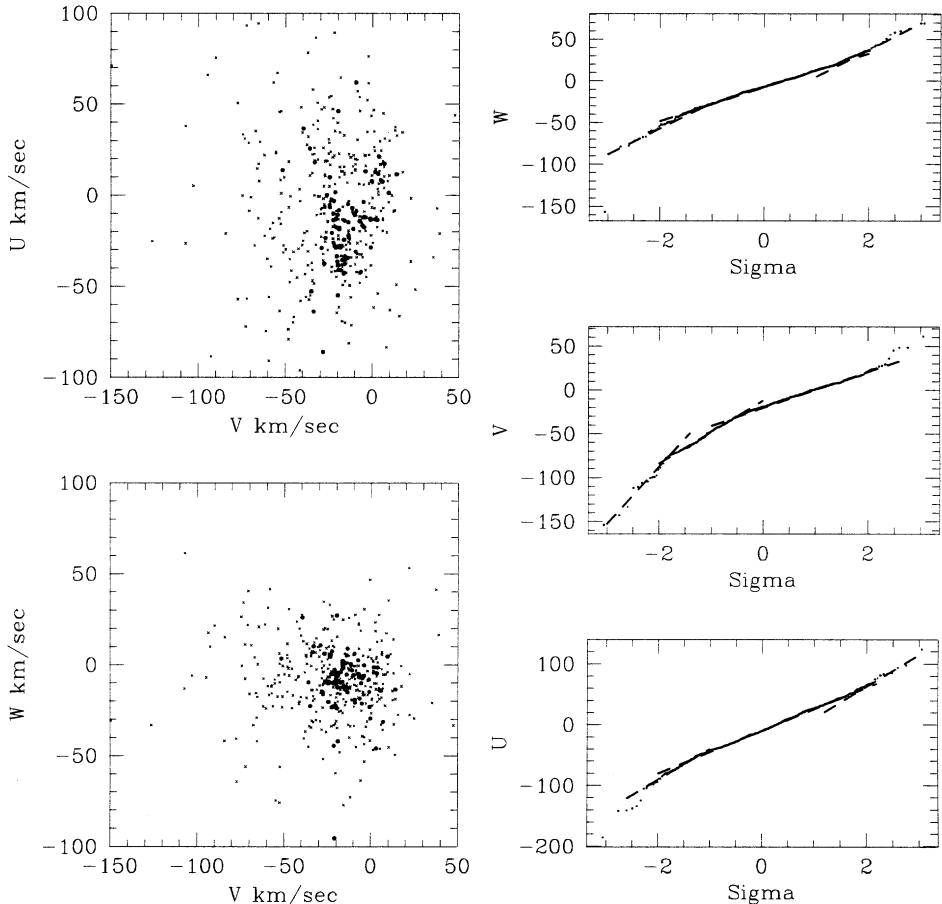


Figure 7.16. The left-hand panels plot the (V , U) and (V , W) velocity distributions of the nearby M dwarf sample, 514 systems with $8 < M_V < 15$. The right-hand panels show the probability plots for each component; in these co-ordinates, a Gaussian distribution follows a straight line, slope σ . Both the U and W distributions are well represented by two Gaussian distributions, with the lower dispersion component dominant in the core and the higher in the wings; the V distribution is more complex (data from [R11]).

Table 7.3 characterises the velocities of disk stars as a velocity ellipsoid, with Gaussian dispersion in each co-ordinate. That model is a simplification, as illustrated in Figure 7.16, where velocity data for a volume-complete sample of M dwarfs are plotted. These distributions prove to be poorly matched by single Gaussians [R3]; however, more detailed analysis suggests that the distributions can be deconvolved into two main components. The principal evidence favouring deconvolution is given by probability plots [L7], which use the inverse of the cumulative probability distribution. Consider a set of n data points, measurements of parameter x .

If we rank order those points, the cumulative probability at point i is given by:

$$F(i) = (i - \frac{1}{2})/n \quad (7.5)$$

$F(i)$ is non-linear for a Gaussian distribution; however, the inverse cumulative probability distribution, $F^{-1}(i)$, of a Gaussian is a straight line, with each point representing the deviation from the mean expected in units of the standard deviation. That is, plotting the rank ordered distribution, $F^{-1}(i)$, gives a straight line, slope σ , for a single Gaussian; the convolution of two Gaussians with the same mean will result in three linear segments, here the lower velocity component contributes the central segment, slope σ_1 , and the higher dispersion component contributes two linear segments, slope σ_2 , at higher and lower x .

As Figure 7.16 shows, probability plots for all three velocity components, U, V and W, can be represented as a series of linear segments. In the case of U and W, the data are well matched by a low-dispersion component in the core (g1 from Table 7.3) and a higher dispersion component (g2) in the wings; the latter component contributes approximately 10% of the stars and may well be the thick disk (see further below). The distribution in V is more complex, probably reflecting the contribution of Strömgren asymmetric drift, which offsets the average velocity of the different components. The data plotted are for local M dwarfs, but analyses of nearby FGK stars give almost identical results [R11], [R12].

Representing the stellar velocity distribution as a sum of two components is reminiscent of the division of disk stars into sub-populations in the Vatican Galaxy model (Table 7.1). However, it is likely that even these multi-component models are simplified representations of the local kinematics. Detailed studies of *Hipparcos* proper-motion data for nearby stars [D3] reveals clumping in velocity space. Those clumps are ‘moving groups’: some are stars that formed in clusters associated with a particular molecular cloud-complex. The Orion I association, for example, may give rise to such a moving group several hundred Myr in the future. Amongst the more prominent systems are groups associated with the Hyades, Pleiades and Coma clusters, and with Sirius [E3], [E4], [E5], [E6]. Other moving groups appear to have a dynamical origin, and represent the response to resonant gravitational interactions between individual stars and inhomogeneities in the Galactic potential (such as spiral arms and the Galactic bar) [D5]. In particular, the outer Lindblad resonance of the bar lies in the vicinity of the Sun, and that may account for the elongated structure in the U component locally.

Finally, an important consequence of the substantial velocities acquired by older disk dwarfs is that those stars migrate to regions in the disk far from their birthplace. With an overall velocity dispersion of $\sigma_U \sim 40 \text{ km s}^{-1}$, M dwarfs which are currently within the immediate Solar Neighbourhood may have originated from star formation regions at Galactic radii of 4–12 kpc. As a result, a *local* sample of disk dwarfs can provide a fair sampling of the properties of the Galactic Disk population as a whole.

7.5.2 The disk density distribution

The radial density law

The determination of the radial density distribution of the Galactic Disk is complicated by our vantage point within the disk. Observations of other spiral galaxies (such as NGC 4565, Figure 7.2) provide a clearer picture of the overall structure. The stellar density distribution in the disk of an external system is derived through surface photometry, making due allowance for contributions from the central bulge, star formation and dust absorption in spiral arms. The first extensive analysis was undertaken by Freeman [F2], who found that radial intensity profiles of several nearby spirals could be represented as exponential distributions, $I(r) \propto e^{-r/r_0}$, where r_0 is the scale length, estimated as 2–3 kpc. As an illustration of this technique, Figure 7.17 shows the radial surface-brightness profile, in magnitudes per square arcsecond (μ_I) of the face-on spiral galaxy NGC 214 (also known as UGC 438): the distribution is almost linear for $r > 15$ arcsec, corresponding to an exponential distribution. The rise in brightness near the centre is due to the bulge component in the galaxy.

Adopting the reasonable assumption that the Galactic Disk is similar to external spirals, models can be used to predict the expected surface brightness distribution and star-counts from our viewpoint within the disk. Those predictions are matched

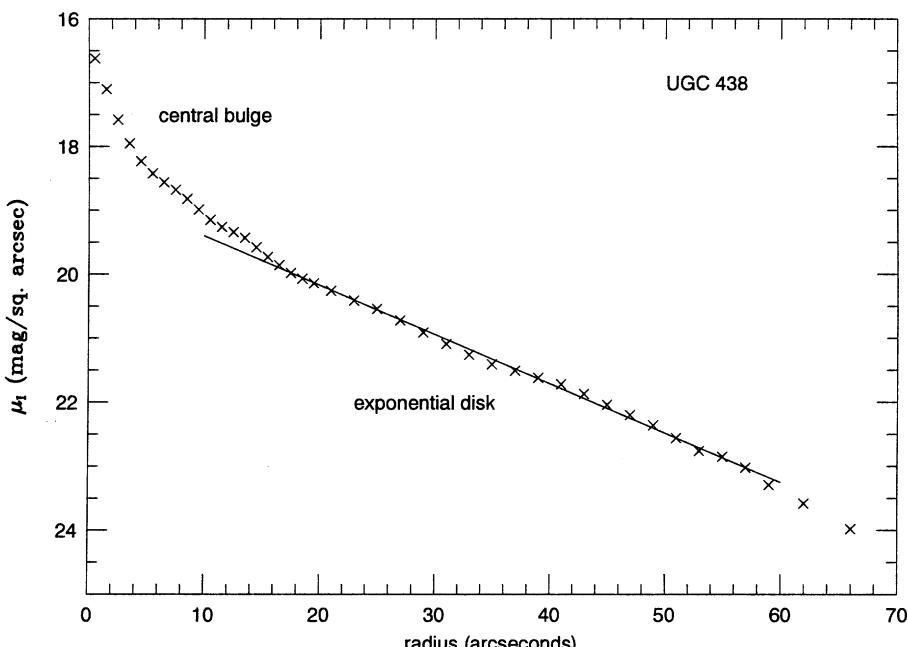


Figure 7.17. The radial surface brightness profile of the disk in the face-on spiral galaxy UGC 438 (NGC 214). (From [J3].)

against observations. Estimates of the disk scale length, R_0 , range from 2 to 6 kpc [R6], [K6]. However, the longer estimates are based on optical photometry (see [K3], for example) and are therefore vulnerable to bias due to inadequate compensation for reddening. Infrared data, less susceptible to interstellar absorption, indicate scale lengths between 2 and 3 kpc, consistent with $R_0 = 2.5$ kpc deduced from direct star-counts towards the anticentre [R6]. The latter data also show evidence for a sharp decrease in the number of disk stars at a distance of ~ 6 kpc beyond the Solar Radius: an edge to the Galactic Disk (although a few open clusters are known at larger radii). Similar truncations in the density distribution have been detected in other spiral galaxies [K4].

The vertical density law

The density distribution perpendicular to the Plane, $\rho(z)$, can be derived from direct star-counts, either by inverting the apparent magnitude–colour (or spectral type) distribution or by building Galaxy models. In the former case, the observations are used to infer absolute magnitudes, and hence distances, for each star, providing direct estimates of $\rho(z)$; in the latter, the Galactic stellar populations are parameterised (density laws, luminosity functions, colour–magnitude relationships) and predicted star-counts matched against observations spanning a range of directions (l, b). The model parameters are adjusted until suitable agreement is achieved. These two techniques can be viewed as complementary, using deductive and inductive methods of analysis.

The thickness of the disk depends on the vertical velocity dispersion, σ_w , which, as described in the previous section, increases with increasing age. Most molecular clouds and star-forming regions lie within 60 parsecs of the mid-Plane, and the short main sequence lifetimes of early-type (OBA) stars leads to their distributions being similarly restricted in z . Thus, Parenago’s discontinuity in (σ_t , spectral type) is reflected in $\rho(z)$, with few unevolved A-type stars reaching distances of $z > 500$ parsecs.

As with the radial density law, $\rho(z)$ is often modelled as an exponential, $\rho(z) \propto \exp(-z/z_0)$, with scale heights of $z_0 \sim 100$ parsecs for OBAF stars, and 250–350 parsecs for later-type dwarfs [M10], [B5]. This parameterisation has its origin in Oort’s summary of the Vatican conference (see Table 7.1), reinforced by Schmidt’s [S11], [S12] analysis of classical star-count data. Schmidt derived ‘equivalent widths’ for the thickness of the disk, where the equivalent width is ‘the quantity with which the spatial densities in the Plane must be multiplied to give the surface densities in the Plane.’ If the vertical density law is an exponential distribution, then the exponential scale height is half the equivalent width. Hence, Schmidt’s equivalent width estimate of 600 parsecs for stars with $M_V > +5$ is consistent with (but does not require) an exponential scale height of 300 parsecs.

Direct analyses of $\rho(z)$ reveal more complex behaviour. Gilmore and Reid [G1] reconstructed $\rho(z)$ from star-counts toward the South Galactic Pole, using an $(M_V, (V-I))$ relation to derive photometric parallaxes (see Section 7.5.2) for G and K stars. The resulting density law shows a clear change in slope at $z \sim 1.5$ kpc

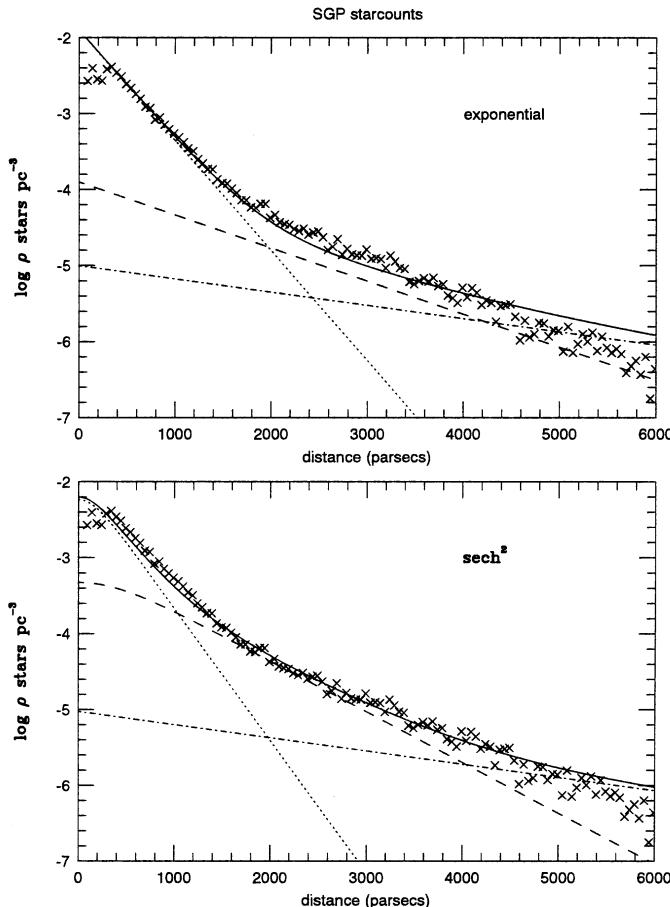


Figure 7.18. The stellar density distribution perpendicular to the Galactic Plane. The upper panel models the distribution as two exponentials, scale heights 300 and 1,000 pc, with 70 : 1 density normalisation; the lower, as two sech^2 distributions, $h_0 = 500$ and 1,400 pc, and a 30 : 1 normalisation. The dotted and dashed line shows the density distribution of the old disk and thick disk; the dot-dash line marks the halo; the solid line plots the combined $\rho(z)$.

(Figure 7.18). Such a feature can be expected in the transition between the highly flattened disk and the halo, but at heights of 5 kpc or more and at lower space densities. The additional stars between 1.5 and 5 kpc were interpreted by Gilmore and Reid as representatives of a separate stellar population, termed the ‘thick disk’, with a local density $\sim 1.5\%$ of the ‘classical’ old disk and a scale height $\sim 1,100$ parsecs.

Subsequent observations [F4], [R1], [R5], [S22], have confirmed the existence of these additional stars, although with somewhat different density parameters. The thick disk is observed directly at distances of more than 1–2 kpc above the Plane, so the local parameters rest on extrapolations over 1–3 scale heights. Estimates of the local density and the scale height are anti-correlated, with most recent studies

favouring smaller scale heights and higher density normalisations. We consider the thick disk in more detail in Section 7.5.4.

While the overall density law can be modelled using a double exponential, exponential distributions have a sharp cusp at $z = 0$. An alternative approach is outlined by Camm [C1], [C2], who demonstrated that an isothermal (single-Gaussian velocity dispersion) population follows a sech^2 distribution. Writing

$$\rho(z) = \rho_0 \text{sech}^2\left(\frac{z}{h_0}\right) \quad (7.6)$$

the density distribution tends to an exponential, scale height $h_0/2$ at large z ,

$$\rho(z) \propto \exp(-2z/h_0) \quad (7.7)$$

At small z , the gradient of the sech^2 function is less steep than an exponential, giving a physically more realistic density distribution.

7.5.3 The abundance distribution

Metallicities of disk stars are determined using techniques similar to those used in studying halo stars. In contrast to the halo, high-resolution spectroscopic analyses show that most (but not all) disk dwarfs have elemental abundance ratios close to scaled-solar values. The most extensive catalogues of metallicity determinations are based on Strömgren photometry of F and G dwarfs ([E1], [W6], [N5]) and analyses of the Mg b region of the spectrum [C4]. In broad terms, these studies find that $\sim 70\%$ of disk stars have abundances in the range $-0.3 < [\text{m}/\text{H}] < 0.15$. The remaining stars form a metal-poor tail that extends to $[\text{m}/\text{H}] \sim -1$, and are often associated with the thick disk.

Until recently, most analyses of the metallicity distribution in the Solar Neighbourhood derived an average value of $[\text{m}/\text{H}] \sim -0.1$ to -0.15 , suggesting that the Sun stood out as somewhat metal-rich. Those studies, however, were based on calibrations that underestimate $[\text{m}/\text{H}]$ and, in many cases, biased reference samples.

Haywood [H6, H7] has computed the most reliable determination of the disk chemical abundance distribution, based on a colour-selected sample of nearby stars. All stars have *Hipparcos* data, ensuring reliable distance and absolute magnitude estimates, and lie below the turn-off of the old disk main sequence ($M_V \sim 4$), ensuring that they are long-lived. These stars provide a fair sampling of the abundance distribution over the full history of disk star formation. The abundances are calibrated using Geneva photometry, a medium-band system that is well suited to G and K dwarfs. The metallicity distribution (Figure 7.19) has a median value coincident with the solar value (i.e., the Sun really is an average disk star). This result has been confirmed independently by Reid [R13], analysing a similar sample, but with metallicities calculated from Strömgren photometry.

Photometric abundance estimates measure the overall metal content; detailed element-by-element chemical abundances require high-resolution spectroscopy. Recent such analyses of disk dwarfs have revealed some extremely interesting trends [P5], [F7], [F8]. Figure 7.20 plots $[\text{Mg}/\text{Fe}]$ vs. $[\text{Fe}/\text{H}]$ abundance data for

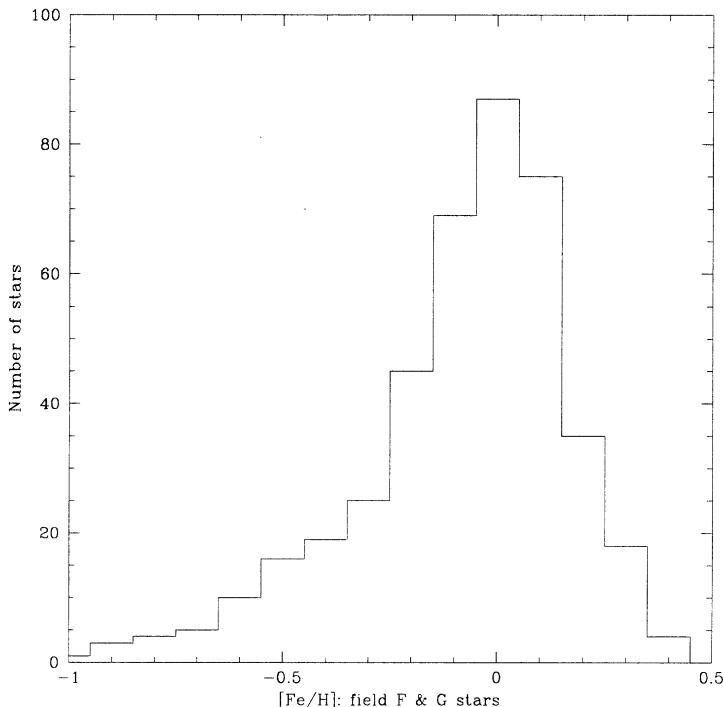


Figure 7.19. The metallicity distribution of an unbiased sample of long-lived stars in the Solar Neighbourhood; the median value matches the solar abundance (from [H6]).

~100 G dwarfs within 25 parsecs of the Sun; Mg is an α element, produced by Type II supernovae (see Section 7.3.4). Most stars have $[\text{Mg}/\text{Fe}] \sim 0$, matching the solar ratio, but 10–20% have enhanced Mg abundances, with a peak enhancement close to the median value for halo stars (Figure 7.7). Moreover, Fuhrmann [F8] has shown that the high $[\text{Mg}/\text{Fe}]$ stars also have significantly higher velocities – in fact, their velocity distribution is very similar to the M dwarf g2 component listed in Table 7.3. Combining these results strongly suggests that the enhanced-Mg, high-velocity dispersion stars are members of the thick disk. We consider the implications for disk formation history in Section 7.5.5.

7.5.4 The age of the Galactic Disk

The rapid rotation and high metallicity of the Galactic Disk point to its formation after the halo. In principle, the absolute age of the oldest disk stars can be determined by matching the position of the main sequence turn-off against theoretical isochrones, as in studies of halo globular clusters. In practice, the fact that star formation has continued throughout the existence of the disk complicates the task of identifying the oldest stars in the field. In general, results are consistent with an age of ~ 10 Gyr.

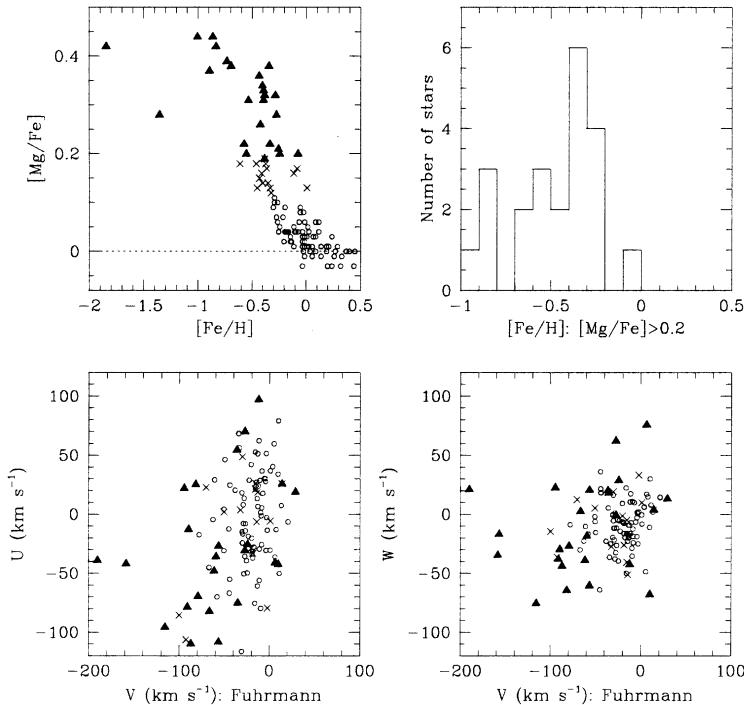


Figure 7.20. The upper left panel plots $[\text{Mg}/\text{Fe}]$ vs. $[\text{Fe}/\text{H}]$ for a sample of FG stars in the Solar Neighbourhood, while the top right panel plots the $[\text{Fe}/\text{H}]$ distribution for the $[\text{Mg}/\text{Fe}] > 0.2$ ‘thick disk’ stars. The lower two panels plot the (V, U) and (V, W) distributions, where stars are coded by $[\text{Mg}/\text{Fe}]$ as follows: solid triangles for $[\text{Mg}/\text{Fe}] > 0.2$; crosses for $0.2 > [\text{Mg}/\text{Fe}] > 0.15$; and open squares for $[\text{Mg}/\text{Fe}] < 0.15$.

A variation on this theme is to estimate the age of the oldest disk open cluster. This has the advantage of isolating a coeval, single-abundance group of stars, simplifying the age determination; on the other hand, open clusters are fragile, loosely-bound objects which are easily disrupted. Few old open clusters survive, and only a lower limit can be set on the age of the disk. At present, Berkeley 13 [P3] is the oldest known, with an age $\tau \sim 12$ Gyr, while NGC 7789 and NGC 188 have $\tau \sim 8$ Gyr [F5].

Finally, the white dwarf luminosity function, $\Phi_{WD}(M_V)$, can be used as an age estimator. Virtually all single stars (and most binary components) with masses less than $7\text{--}8 M_\odot$ become white dwarfs during their final, dying phase of evolution. Mestel’s [M8] early analytical models of these degenerate stars predicted that $L \propto (M/\tau_c)^{7/5}$, where τ_c is the cooling time, the time since the star became a white dwarf. If the disk has a finite age, $\Phi_{WD}(M_V)$ must drop to zero at a particular luminosity set by the finite cooling time available. Schmidt [S11] originally suggested this as a method of measuring the age of the disk. The main complication is the identification of a complete sample of low-luminosity degenerates, which are

photometrically similar to main sequence K dwarfs. Most recent studies of the white dwarf luminosity function [W5], [O4], [L4] favour an age between 8 and 12 Gyr for the oldest stars in the Galactic Disk. However, these analyses fail to take into account a complication introduced by the thick disk. This population is best characterised as a single-burst population, old and with a restricted range of ages (see Section 7.5.5). As a consequence, the thick disk has a higher proportion of white dwarfs (relative to main sequence stars) than the old disk: continuing star formation in the old disk replaces stars that evolve off the main sequence. Moreover, since the thick disk is old, most white dwarfs are concentrated at low luminosities. Combined, these two factors lead to the thick disk dominating the white dwarf luminosity function at low luminosities. As a result, age estimates assigned to the old disk are actually more appropriate to the thick disk. Re-analysing the luminosity function using revised white dwarf models, it seems unlikely that the thick disk is older than ~ 10 Gyr, while significant star formation within the old disk did not commence until ~ 8 Gyr ago [R12].

7.5.5 The star-formation history of the disk

One of the major questions in Galactic structure over the last 20 years is how the thick disk is related to the 300-pc scale height old disk (also known as the thin disk). Recent observational results, particularly the $[\text{Mg}/\text{Fe}]$ distribution shown in Figure 7.20, go a long way toward answering that question. It has been clear from the outset that the thick disk is an old population: star-counts show that there are very few A or F-type stars at heights of more than 1 kpc above the Plane. Moreover, the $[\text{Mg}/\text{Fe}]$ results indicate that, like the halo, the current population must have formed over a relatively rapid timescale, before Type Ia supernovae could start to increase the Fe abundance. However, thick disk kinematics show that the stars had settled into a rotating disk. What led to the cessation of star formation before $[\text{Mg}/\text{Fe}]$ could reach solar values?

The Sagittarius dwarf, and the star streams currently detected in the halo, may provide a clue to the answer to this question. Galaxy formation models indicate that mergers are likely to have been more common at earlier epochs, and could well have played a significant role in the evolution of the Milky Way. Hence, star formation in the thick disk may have been terminated by a major merger event; the satellite system not only disrupted the stellar disk, increasing the scale height to 1 kpc to match the present-day thick disk, but also triggered sufficient star formation (perhaps in the Bulge) that stellar winds blew out most of the remaining gas. That, in turn, led to a break in the star formation history of the disk, until the thin disk reformed. As discussed in Section 7.5.4, these events probably occurred ~ 10 Gyr ago, shortly after the formation of the halo [Q1].

Turning to more recent epochs and star formation in the present thin disk, we lack accurate measurements of the stellar birth rate, $B(t)$. The most obvious constituents of the disk population are open clusters. Their age distribution [F5] shows that star formation has been taking place continuously over the ~ 10 -Gyr lifetime of the Galactic Disk. However, these structures have masses of

$\sim 10^3\text{--}10^4 M_{\odot}$, with correspondingly low binding energy, and therefore can be disrupted by gravitational interactions on timescales of 1 Gyr or less (see Section 8.6.2). The majority of old ($\tau > 3$ Gyr) clusters which have survived to the present day have orbits which are significantly inclined to the Galactic Plane, leading to their spending the majority of their existence undisturbed by gravitational perturbations. Thus, the number of open clusters as a function of age is not a reliable guide to the overall stellar birth-rate.

Several techniques have been used to try to disentangle the star formation history from the properties of nearby stars. Assuming that there is a single-valued relationship between age and metallicity, the abundance distribution of disk stars could be used to infer the past star-formation activity [T1], [R7]. Unfortunately, detailed analyses show that there is significant variation in the abundances of stars formed at any given time in the disk. Figure 7.21 plots results for F and G dwarfs in the field, where ages have been estimated by matching the observed Strömgren colours against theoretical isochrones. These result in unreasonably high ages for some stars ($\tau > 12$ Gyr), but the relative ages are reliable. The main conclusion that should be drawn from this figure is that metallicity is not useful as an age indicator for abundances $[m/H] > -1$.

Attempts have also been made to use the white dwarf luminosity function, Φ_{WD} , to probe the star-formation history. Consider the case where there is a burst of star formation, duration τ_D between times τ_1 and τ_2 , superimposed on a nearly-constant background rate. White dwarf progenitors have masses from 1 to $7\text{--}8 M_{\odot}$, with corresponding main sequence lifetimes of 10 Gyr to 5×10^7 yr (0.05 Gyr). Thus, burst stars first make a contribution to Φ_{WD} at time $\tau_1 + 0.05$ Gyr, with the contribution increasing as lower-mass burst stars evolve through the red giant phase. Models [N2] show that the net effect is the appearance of a ‘bump’ in Φ_{WD} whose location can, in principle, be used to estimate both the duration of the starburst and the time since its occurrence. In practice, Φ_{WD} is defined to only moderate precision [W5], and while there is marginal evidence for a mild, recent (< 0.3 Gyr old) burst of star formation, the data are generally consistent with little variation in the average birth-rate over the last 4 Gyr.

Both lithium abundance and chromospheric activity have been employed as age indicators in late-F and G-type dwarfs. In those stars, the temperature at the base of the convective envelope is sufficiently high to allow partial lithium burning through the (p, α) process (see Section 3.3.3), with the result that the lithium abundance decreases with time. Early analyses [D4] favour a mildly decreasing SFR over the last 4–5 Gyr, but the mass-dependence of the age/(Li abundance) relationship complicates the analysis. Moreover, the surprising detection of lithium at low abundance levels in K dwarfs, where deep convection should have resulted in rapid lithium depletion, raises questions about the reliability of the technique [F1].

Chromospheric activity in solar-type stars is characterised through measurement of Ca II K emission strength, the time dependence of which is generally approximated by the $t^{-1/2}$ Skumanich relationship (Section 5.5.2). Given accurate data for an unbiased sample of field stars, that age calibration can be used to reconstruct the disk star-formation history. Barry [B7] suggested that

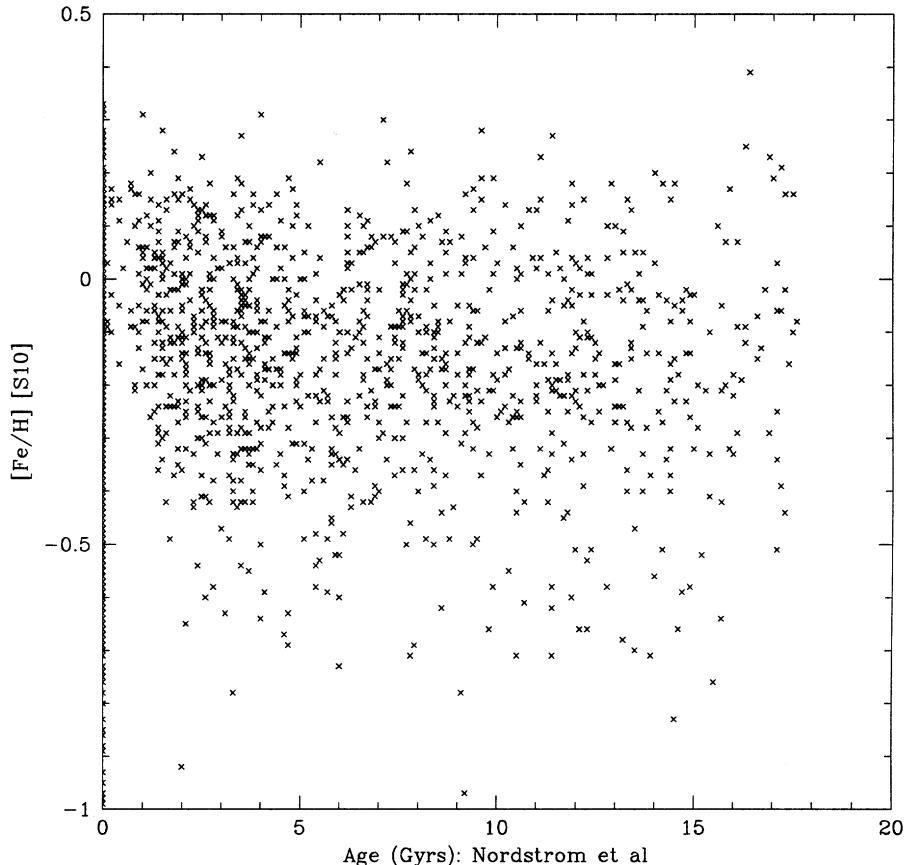


Figure 7.21. The age–metallicity relationship for stars in the Galactic Disk. The data plotted are for a volume-complete sample of FG stars with $d < 40$ pc from [N5].

there have been significant bursts in activity 4–5 Gyr before the present epoch, and within the last 10^9 years. Soderblom *et al.* [S16], however, have criticised that analysis, and argue that the available data [H4] can be represented by a constant star-formation rate. The distribution of chromospheric activity in M dwarfs is also consistent with only minor departures from constant star-formation over the history of the disk [G5]. In short, a reliable method for determining the detailed disk star-formation history remains to be identified, but there is little evidence for substantial variation.

7.6 A QUICK TOUR OF THE WIDER SOLAR NEIGHBOURHOOD

Most analyses of late-type stars and brown dwarfs concentrate, perforce, on the regions of the Galaxy nearest the Sun. Surveys for ultracool dwarfs (late-M, L

and T dwarfs) are restricted to distances of only 20–30 parsecs, but the nearest open clusters and young associations lie at greater distances; in particular, the nearest intermediate-age open cluster, the Pleiades, lies at a distance of ~ 130 pc, while most of the young star-forming associations targeted by brown dwarf surveys lie at distances between 300 parsecs (e.g., IC 348) and 450 parsecs (the Orion Nebula Cluster, σ Ori, λ Ori).

On a large scale, the Sun lies on the edge of the Orion Spur, between the Perseus spiral arm, ~ 2 kpc beyond the Sun in the outer regions of the disk, and the Sagittarius arm, ~ 1 kpc inward, toward the Galactic Centre. Most, but not all, major star formation sites lie within the spiral arms; indeed, in 1951, W. W. Morgan first demonstrated that the Milky Way had spiral arms by mapping the distribution of the nearest OB associations [G6]. Figure 7.22 shows an updated and extended version of Morgan’s map. Both the Orion Nebula and the star forming regions of the Orion OB1 Association are within the Orion spur, which is an extension of the inner Sagittarius arm, even though the Orion clouds lie at larger Galactic radii than the Sun.

Figure 7.23 provides a schematic map of the main Galactic features within ~ 500 parsecs of the Sun. Besides Orion, star formation is under way in the ρ Ophiuchus, Lupus and Chamaeleon dark clouds, and in the nearer, but much dispersed, Taurus clouds. Young open clusters, with ages up to 10 Myr, lie in or near most of these star-forming regions; these clusters have been observed intensively at infrared wavelengths over the last five years, with the primary aim of probing the structure of the mass function well below the hydrogen-burning limit (see Section 9.7).

There is evidence of substantial local star formation in the recent past (< 20 Myr) through the presence of several OB associations, notably ScoCen, Upper Scorpius, Perseus and the series of star-forming regions that form the Orion OB1 Association, extending beyond the Orion Nebula Cluster. Many of these associations and most of the nearby B and early A-type stars are associated with the large complex known as Gould’s Belt [G7], an ~ 600 -parsec diameter disk-like structure, ~ 60 parsecs thick and inclined by $17\text{--}22^\circ$ to the Galactic Plane. The presence of so many associations of similar age suggests correlated star formation, triggered by a particular event, 35–40 Myr before the present day. The origin remains unclear, but could possibly be by collision between a high-velocity HI cloud (~ 500 pc in diameter and a few $\times 10^5 M_\odot$) and the Galactic Disk, or by a chain reaction series of supernovae explosions (see [G8]).

The Sun is passing through Gould’s Belt. At present, it lies in a low gas-density (~ 0.001 atoms cm^{-3} , $\sim 10\%$ of the average density), 100-parsec diameter feature known as the Local Bubble. This feature appears to have been generated within the last 10 Myr by either a nearby supernovae or stellar winds from high-mass stars in the Scorpius–Centaurus star-forming region, which currently lies ~ 130 parsecs from the Sun. There are several other near-spherical, low-density regions within 500 pc of the Sun (see Figure 7.23), and it is likely that these too have been swept clean by winds generated by high-mass stars. The Sun has been travelling through the Local Bubble for the last ~ 5 Myr, headed towards the Great Rift in Aquila ($1 \sim 50^\circ$).

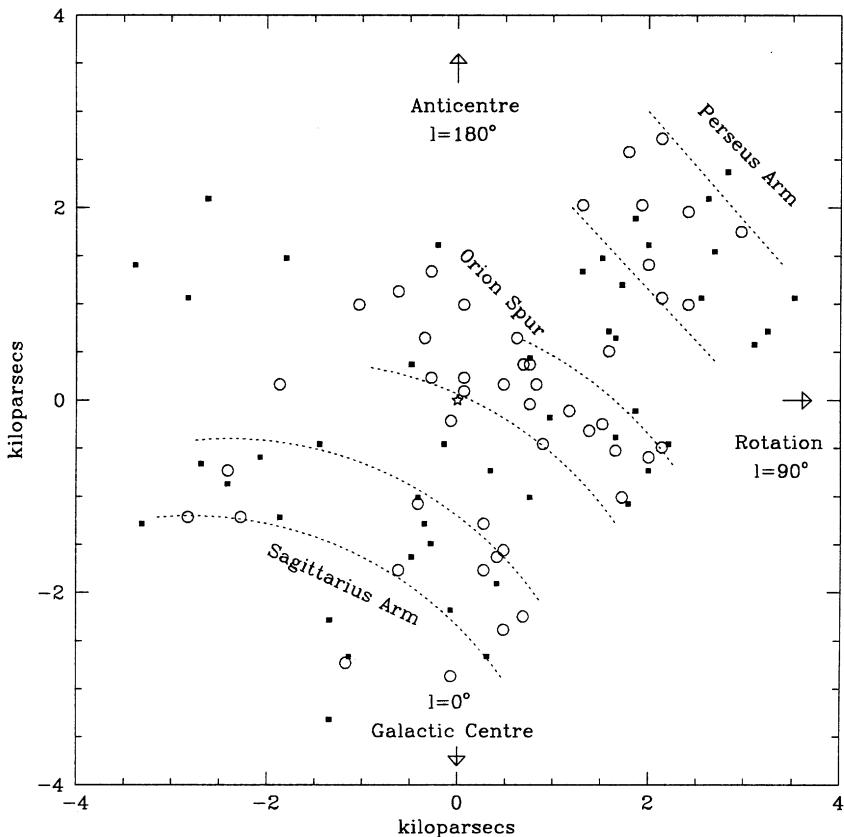


Figure 7.22. An outline of the main spiral arm complexes in the vicinity of the Sun; the Orion Spur merges the inner Sagittarius Arm. Circles mark OB associations and squares are open clusters (from [H8]); the location of the Sun is marked with a five-point star.

The nearest open cluster, the \sim 625 Myr old Hyades cluster, also lies within the Bubble, approximately 45 pc from the Sun, towards the Galactic anticentre. The younger Pleiades, age \sim 120 Myr, are on the edge of the Taurus clouds, the nearest extended star-forming region. The locations of those clusters, together with Praesepe, IC 2391 and several other intermediate-age open clusters targeted by luminosity function investigations (Section 8.9), are marked on Figure 7.23.

Finally, one of the more surprising discoveries of the last five years is the fact that, even though the Sun lies more than 130 parsecs from the nearest region of active star formation (the Centaurus–Crux Association within ScoCen), over 100 stars with ages between 10 and 50 Myr have been discovered within 60 parsecs of the Sun. These stars appear to be members of several loose associations: TW Hydrae ($\tau \sim$ 8 Myr); β Pictoris (12 Myr); Tucana/Horologium (30 Myr); η Chamaeleon (8 Myr); AB Doradus (50 Myr); and the Cha-Near group (10 Myr) [Z4]. While the origin of these young associations is not entirely clear, suspicions rest mainly on the

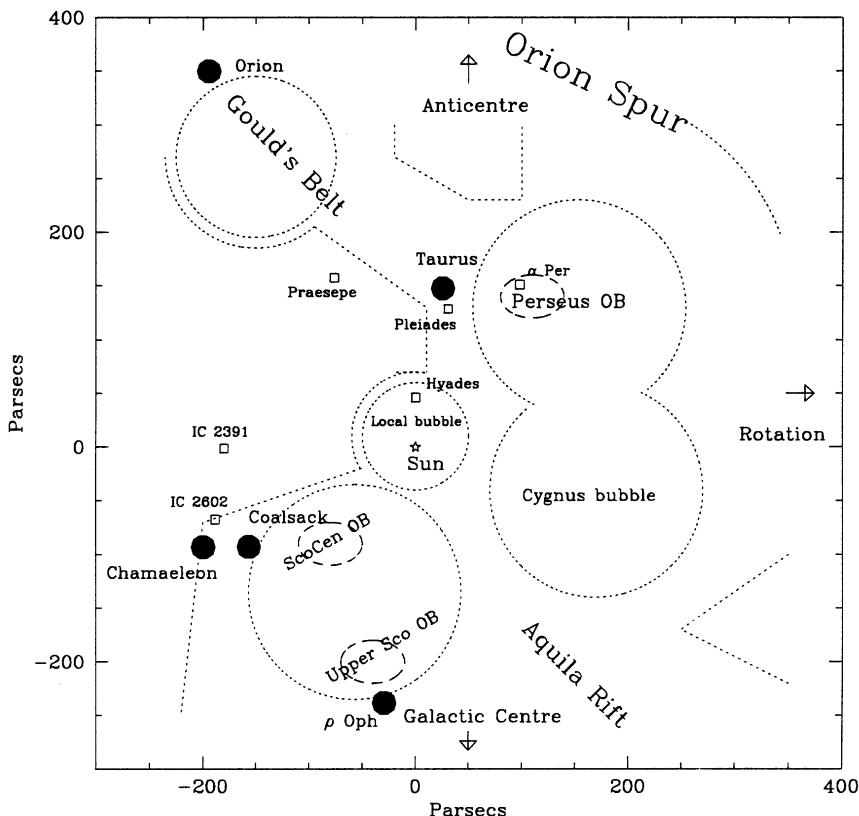


Figure 7.23. A schematic outline of the 800-parsec square region centred on the Sun. Open clusters are shown as open squares, star-forming regions as solid points and low gas-density bubbles as dotted circles.

ScoCen star-forming region. It is possible that the passage of a giant molecular cloud (or clouds) through a spiral arm \sim 60 Myr ago may have triggered a series of star-forming episodes that gave rise to the ScoCen complex. In any event, the members of these associations provide the best opportunity of investigating the circumstellar environments of young low-mass stars and brown dwarfs.

The overall structure of the Solar Neighbourhood, and, indeed, of much of the local Milky Way, should become much clearer towards the end of the next decade, when the ESA satellite, *Gaia*, completes its mission. *Gaia* (see <http://www.rssd.esa.int/gaia/>) is the direct successor to the extremely successful *Hipparcos* mission of the 1990s, and is scheduled for launch in 2011. Like *Hipparcos*, *Gaia* will measure accurate distances by obtaining extremely precise trigonometric parallaxes throughout its 5-year mission; however, while *Hipparcos* targeted only \sim 120,000 stars brighter than 13th magnitude and obtained milliarcsecond accuracy, *Gaia* aims to survey \sim 10⁹ stars to $V = 20$, with

an accuracy ranging from 4 μ arcseconds at $V = 10$ through $\sim 10 \mu$ arcseconds at $V = 15$, to 150 μ arcseconds at the survey limit. *Gaia* will also obtain multicolour photometry (broad and medium band filters covering 4,000 to 10,000 \AA) for all targets, and radial velocities to better than 10 km s^{-1} for stars brighter than $V \sim 17$. If successful, this extremely ambitious project will revolutionise our understanding of stellar evolution, Galactic structure, the cosmic distance scale and galaxy formation.

7.7 SUMMARY

Our aim in this chapter has been to provide an overview of current understanding of Galactic structure. We have

- reviewed the historical development of Galactic structure, with particular emphasis on the development of the concept of stellar populations;
- summarised current understanding of the kinematics, spatial distribution and abundances of the Galactic halo, and considered how those properties are interpreted in terms of the age and the initial stages of formation of the Galaxy;
- outlined the most significant features of the Galactic Bulge, and considered its relation to the disk and halo;
- summarised the main properties of the Galactic disk, particularly the complex spatial distribution perpendicular to the Plane and the results recently derived from analysis of *Hipparcos* data.

Pulling together the properties of the different stellar population, a plausible formation scenario might be as follows:

- 1 $\tau \sim 12 \text{ Gyr}$: formation of the native ELS halo during the rapid ($\sim 100 \text{ Myr?}$) dissipational collapse of the protogalactic cloud; the resultant system of metal-poor stars is predominantly pressure supported, with weak prograde net rotation and a moderate axial ratio ($c/a \sim 0.5$) (i.e., the inner halo).
- 2 $\tau < 12 \text{ Gyr}$: minor mergers of satellite systems add coherent star streams to the halo, primarily the outer halo; the merger rate is probably highest during the initial few Gyr, but continues to the present day (*vide Sagittarius*).
- 3 $\tau \sim 10\text{--}11 \text{ Gyr}$: dissipational collisions, coupled with angular momentum conservation, lead to gas clouds in the collapsing protogalaxy forming a rotating disk, with dynamical properties similar to the present-day thin disk. The thick disk stars form within that system, while star formation in the central regions of the Galaxy leads to the formation of the (metal-poor?) Bulge. Star formation and mass loss drives up the metallicity of the interstellar medium, and Type Ia supernovae start to increase the Fe abundance.
- 4 $\tau \sim 9\text{--}10 \text{ Gyr}$: a major merger disrupts the proto-thick disk, injecting sufficient energy to heat the disk and increase the scale height by a factor of ~ 3 . The merger may also prompt a sharp (and brief?) increase in star formation in the disk, and, perhaps, more prolonged star formation in the Bulge.

- 5 $\tau \sim 8$ Gyr to present day: the thin disk reforms and ‘normal’ star formation resumes in an interstellar medium now increasingly enriched by the products of Type Ia supernovae.

The following chapters are concerned primarily with a simple quantity: the stellar mass function, the number of stars and substellar-mass brown dwarfs as a function of mass. Although simple in concept, this parameter has wide-reaching consequences for both Galactic dynamics and the theory of star-formation. As we will describe, its derivation is also far from straightforward, and has absorbed substantial observational and theoretical effort throughout the latter half of the twentieth century.

7.8 REFERENCES

- A1 Adams, W. S., Joy, A. H., 1922, *ApJ*, **56**, 242.
- A2 Adams, W. S., Joy, A. H., Humason, M. L., Brayton, A. M., 1935, *ApJ*, **81**, 187.
- A3 Armandroff, T. E., 1989, *AJ*, **97**, 375.
- A4 Arp, H. C., 1955, *AJ*, **60**, 1.
- A5 Axer, M., Fuhrmann, K., Gehren, T., 1994, *A&A*, **291**, 895.
- B1 Baade, W., 1944, *ApJ*, **100**, 137.
- B2 Baade, W., 1944, *ApJ*, **100**, 147.
- B3 Baade, W., 1951, *Publ. Univ. Mich. Obs.*, **10**, 7.
- B4 Baade, W., Hubble, E., 1939, *PASP*, **51**, 40.
- B5 Bahcall, J. N., Soneira, R. M., 1980, *ApJS*, **44**, 73.
- B6 Bailey, S. I., 1917, *Harvard Annals*, **78**, pt. 2.
- B7 Barry, D. C., 1988, *ApJ*, **334**, 436.
- B8 Beers, T. C., Sommer-Larsen, J., 1996, *ApJS*, **96**, 175.
- B9 Beers, T., 2000, in *The First Stars*, eds. A. Weiss, T. G. Abel and V. Hill, Springer-Verlag.
- B10 Bessell, M. S., Norris, J., 1984, *ApJ*, **28**(5), 622.
- B11 Biémont, E., Badoux, M., Kurucz, R. L., Ansbacherm, W., Pinnington, E. H., 1991, *A&A*, **249**, 539.
- B12 Binney, J. E., Tremaine, S., 1998, *Galactic Dynamics*, W. H. Freeman & Company, San Francisco.
- B13 Blaauw, A., 1965, in *Galactic Structure*, Vol. 5, *Stars and Stellar Systems*, ed. A. Blaauw and M. Schmidt, p. 435.
- B14 Bolte, M., Hogan, C., 1995, *Nature*, **376**, 399.
- B15 Breitschwerdt, D., Egger, R., Freyberg, M. J., 1996, *Sp. Sci. Rev.*, **78**, 183.
- C1 Camm, G., 1950, *MNRAS*, **110**, 305.
- C2 Camm, G., 1952, *MNRAS*, **112**, 155.
- C3 Canterna, R., 1976, *AJ*, **81**, 228.
- C4 Carney, B. W., Laird, J. B., Latham, D. W., Kurucz, R. L., 1987, *AJ*, **94**, 1066.
- C5 Carney, B. W., Latham, D. W., Laird, J. B., Aguilar, L. A., 1994, *AJ*, **107**, 2240.
- C6 Carretta, E., Gratton, R. G., 1997, *A&AS*, **121**, 95.
- C7 Cassisi, S., Castellani, V., Degl'innocenti, S., Weiss, A., 1998, *A&AS*, **129**, 267.
- C8 Chamberlain, J. W., Aller, L. H., 1951, *ApJ*, **114**, 52.
- C9 Charlier, C. V. L., 1913, *Medd. Lund Obs. Ser. II*, No. 9, p. 92.
- C10 Chiba, M., Yoshii, Y., 1998, *AJ*, **115**, 168.

- C11 Clementini, G., Carretta, E., Gratton, R., Merighi, R., Mould, J. R., McCarthy, J. K., 1995, *AJ*, **110**, 2319.
- C12 Crawford, D. L., 1979, in ‘Problems of calibration of multicolour photometric systems’, ed. A. G. Davis Philip, *Dudley Obs. Report No. 14*, p. 23.
- C13 Croswell, K., 1995, *The Alchemy of the Heavens*, Doubleday, New York.
- C14 Chen, B., *et al.*, 2001, *ApJ*, **553**, 184.
- C15 Chiba, M., Beers, T., 2000, *AJ*, **119**, 2843.
- C16 Christlieb, N., *et al.*, 2002, *Nature*, **419**, 904.
- C17 Christlieb, N., *et al.*, 2004, *ApJ*, **603**, 708.
- D1 D’Antona, F., Caloi, V., Mazzitelli, I., 1997, *ApJ*, **477**, 519.
- D2 Dehnen, W., Binney, J. J., 1998, *MNRAS*, **298**, 387.
- D3 Dehnen, W., 1999, *AJ*, **115**, 2384.
- D4 Duncan, D. K., 1981, *ApJ*, **248**, 651.
- D5 Dehnen, W., 2000, *AJ*, **119**, 800.
- E1 Edvardsson, B., Andersen, J., Gustafsson, B., Lambert, D. L., Nissen, P. E., Tomkin, J., 1993, *A&A*, **275**, 101.
- E2 Eggen, O. J., Lynden Bell, D., Sandage, A. R., 1962, *ApJ*, **136**, 748.
- E3 Eggen, O. J., 1983, *AJ*, **88**, 642.
- E4 Eggen, O. J., 1986, *PASP*, **98**, 423.
- E5 Eggen, O. J., 1996, *AJ*, **111**, 1615.
- E6 Eggen, O. J., 1998, *AJ*, **116**, 1810.
- E7 ESA, 1997, *The Hipparcos Catalogue*, ESA SP-1200.
- F1 Favata, F., Micela, G., Sciortino, S., 1996, *A&A*, **311**, 951.
- F2 Freeman, K. C., 1970, *ApJ*, **160**, 811.
- F3 Frenk, C. S., White, S. D. M., 1980, *MNRAS*, **198**, 173.
- F4 Friel, E. D., 1987, *AJ*, **93**, 1388.
- F5 Friel, E. D., 1995, *ARA&A*, **33**, 381.
- F6 Frisch, P. C., York, D. G., 1983, *ApJ*, **271**, L59.
- F7 Fuhrmann, K., 1998, *A&A*, **338**, 161.
- F8 Fuhrmann, K., 2004, *Astr. Nachr.*, **325**, 3.
- G1 Gilmore, G. F., Reid, I. N., 1983, *MNRAS*, **202**, 1025.
- G2 Gilmore, G., Wyse, R. F. G., Kuijken, K., 1989, *ARA&A*, **27**, 555.
- G3 Gratton, R. G., Carretta, E., Castelli, F., 1997a, *A&A*, **314**, 191.
- G4 Gratton, R. G., Fusi Pecci, F., Carretta, E., Clementini, G., Cosri, C. E., Lattanzi, M., 1997b, *ApJ*, **491**, 749.
- G5 Gizis, J. E., Reid, I. N., Hawley, S. L., 2002, *AJ*, **123**, 3356.
- G6 Gingerich, O., 1985, in *The Milky Way Galaxy*, IAU Symposium 106, eds. H. van Woerden, R. J. Allen and W. B. Burton, Reidel, Dordrecht, p. 59.
- G7 Gould, B. A., 1874, *Proc. Am. Ass. Adv. Sci.*, **115**.
- G8 Grenier, I.A., 2004, Proc. of 39th Rencontres de Moriond, 2004, The Young Local Univers, eds. A. Chalabaev, Y. Fukui and T. Montmerle.
- H1 Harris, W. E., 1976, *AJ*, **81**, 1095.
- H2 Hartwick, F. D. A., 1987, in *The Galaxy*, eds. R. Carswell and G. Gilmore (Cambridge University Press, Cambridge).
- H3 Hawley, S. L., Gizis, J. E., Reid, I. N., 1996, *AJ*, **112**, 2799.
- H4 Henry, T. J., Soderblom, D. R., Donahue, R. A., 1996, *AJ*, **111**, 439.
- H5 Helmi, A., White, S. D. M., de Zeeuw, P. T., Zhao, H., 1999, *Nature*, **402**, 53.
- H6 Haywood, M., 2001, *MNRAS*, **325**, 1365.
- H7 Haywood, M., 2002, *MNRAS*, **337**, 151.

- H8 Humphreys, R. M., 1970, *AJ*, **75**, 602.
- I1 Ibata, R., Gilmore, G., Irwin, M. J., 1995, *MNRAS*, **277**, 781.
- I2 Iben, I., Tutukov, A. V., 1984, *ApJS*, **54**, 335.
- I3 Ivezić, Z., *et al.*, 2000, *AJ*, **120**, 963.
- J1 Jenkins, A., 1992, *MNRAS*, **257**, 620.
- J2 Johnston, K. V., 1998, *ApJ*, **495**, 297.
- J3 de Jong, R. S., van der Kruit, P. C., 1994, *A&AS*, **106**, 451.
- K1 Kent, S. M., Dame, T. M., Fazio, G., 1991, *ApJ*, **378**, 131.
- K2 Kinman, T. D., Suntzeff, N. B., Kraft, R. P., 1995, *AJ*, **108**, 1722.
- K3 van der Kruit, P. C., 1986, *A&A*, **157**, 230.
- K4 van der Kruit, P. C., Searle, L., 1982, *A&A*, **110**, 61.
- K5 Kuiper, G. P., 1939, *ApJ*, **89**, 549.
- L1 Laird, J. B., Rupen, M. P., Carney, B. W., Latham, D. W., 1988, *AJ*, **96**, 1908.
- L2 Layden, A. C., Hanson, R. D., Hawley, S. L., Klemola, A. R., Hanley, C. J., 1996, *AJ*, **112**, 2110.
- L3 Lee, Y. W., Demarque, P., Zinn, R., 1994, *ApJ*, **423**, 248.
- L4 Leggett, S. K., Ruiz, M. T., Bergeron, P., 1998, *ApJ*, 294.
- L5 Lindblad, B., 1925, *Upsala Medd.*, No. 3.
- L6 Lindblad, B., 1926, *Ark. Mat. Astron. Fys.*, **19B**, No. 7.
- L7 Lutz, T., Upgren, A. R., 1980, *AJ*, **85**, 1390.
- M1 McClure, R. D., 1979, in *Problems of Calibration of Multicolour Photometric Systems*, Dudley Observatory Report No. 13, ed. A. G. Davis Philip (Schenectady, New York), p. 83.
- M2 McWilliam, A., Rich, R. M., 1994, *ApJS*, **91**, 749.
- M3 McWilliam, A., 1997, *ARA&A*, **35**, 503.
- M4 Majewski, S. R., 1993, *ARA&A*, **31**, 575.
- M5 Majewski, S. R., Munn, J. A., Hawley, S. L., 1996, *ApJ*, **459**, L73.
- M6 Matteucci, F., Greggio, L., 1986, *A&A*, **154**, 279.
- M7 Mayall, N. U., 1946, *ApJ*, **104**, 290.
- M8 Mestel, L., 1952, *MNRAS*, **112**, 583.
- M9 Mihalas, D., Binney, J. J., 1981, *Galactic Astronomy*, W. H. Freeman and Company, San Francisco.
- M10 Miller, G. E., Scalo, J. M., 1979, *ApJS*, **41**, 513.
- M11 Minniti, D., 1995, *ApJ*, **459**, 175.
- M12 Minniti, D., 1995, *ApJ*, **459**, 579.
- M13 Minniti, D., Olszewski, W. E., Liebert, J., White, S. D. M., Hill, J. M., Irwin, M. J., 1995, *MNRAS*, **277**, 1293.
- M14 Morgan, W. W., 1956, *PASP*, **68**, 509.
- M15 Morgan, W. W., 1959, *AJ*, **64**, 432.
- M16 Morrison, H. L., Flynn, C., Freeman, K. C., 1990, *AJ*, **100**, 1191.
- M17 Majewski, S. R., 2004, *Proc. ASA*, **21**, 197.
- N1 Nissen, P. E., Schuster, W. J., 1989, *A&A*, **222**, 69.
- N2 Noh, H.-R., Scalo, J., 1990, *ApJ*, **352**, 605.
- N3 Norris, J., 1986, *ApJS*, **61**, 667.
- N4 Norris, J., 1987, *ApJ*, **314**, L39.
- N5 Nordstrom, B., *et al.*, 2004, *A&AS*, **418**, 989.
- O1 O'Connell, D. J. K., ed., 1958, *Specula Astronomica Vaticana*, Vol. 15, Specula Vaticana, Vatican City.
- O2 Oort, J. H., 1928, *Bull. Astr. Inst. Neth.*, **159**, 273.

- O3 Oort, J. H., 1965, in *Galactic Structure*, Vol. 5, *Stars and Stellar Systems*, eds. A. Blaauw and M. Schmidt, University of Chicago Press, Chicago, p. 435.
- O4 Oswalt, T. D., Smith, J. A., Wood, M. A., Hintzen, P., 1996, *Nature*, **382**, 692.
- P1 Parenago, P. P., 1950, *A. Zh.*, **27**, 150.
- P2 Paresce, F., 1984, *AJ*, **89**, 1022.
- P3 Phelps, R., 1998 *ApJ*, **483**, 823.
- P4 Pfenniger, D., Norman, C., 1990, *ApJ*, **363**, 391.
- P5 Prochaska, J. X., Naumov, S. O., Carney, B. W., McWilliam, A., Wolfe, A. M., 2000, *AJ*, **120**, 2513.
- Q1 Quillen, A. C., Garnett, D. R., 2001, in *Galaxy disks and Disk Galaxies*, eds. J. G. Funes and E. M. Corsini, *ASP Conf. Ser.*, **230**, 87.
- R1 Reid, I. N., Majewski, S. R., 1993, *ApJ*, **409**, 635.
- R2 Reid, I. N., 1996, *MNRAS*, **278**, 367.
- R3 Reid, I. N., Hawley, S. L., Gizis, J. E., 1995, *AJ*, **110**, 1838.
- R4 Reid, I. N., 1998, *AJ*, **115**, 204.
- R5 Robin, A. C., 1994, *IAU Symposium 161*, ed. H. T. MacGillivray *et al.*, p. 403.
- R6 Robin, A. C., Crézé, M., Mohan, V., 1992, *A&A*, **265**, 32.
- R7 Rochapinto, H. J., Maciel, W. J., 1997, *MNRAS*, **289**, 882.
- R8 Roman, N. G., 1954, *AJ*, **59**, 307.
- R9 Roman, N. G., 1955, *ApJS*, **2**, 195.
- R10 Rose, J. A., 1991, *AJ*, **101**, 937.
- R11 Reid, I. N., Gizis, J. E., Hawley, S. L., 2002, *AJ*, **124**, 2721.
- R12 Reid, I. N., 2005, *ARA&A*, **43**, in press.
- R13 Reid, I. N., 2001, *PASP*, **114**, 306.
- S1 Salaris, M., Chieffi, A., Straniero, O., 1993, *ApJ*, **414**, 580.
- S2 Sandage, A., Eggen, O. J., 1959, *MNRAS*, **119**, 278.
- S3 Sandage, A. R., 1969, *ApJ*, **158**, 115.
- S4 Sandage, A., 1970, *ApJ*, **162**, 841.
- S5 Sandage, A., 1986, *ARA&A*, **24**, 421.
- S6 Sarajedini, A., Demarque, P., 1990, *ApJ*, **365**, 219.
- S7 Sarajedini, A., 1997, *AJ*, **113**, 682.
- S8 Sarajedini, A., Chaboyer, B., Demarque, P., 1997, *PASP*, **109**, 1321.
- S9 Scalo, J. M., Miller, G. E., 1980, *ApJ*, **239**, 153.
- S10 Schuster, W. J., Nissen, P. E., 1989, *A&A*, **221**, 65.
- S11 Schmidt, M., 1959, *ApJ*, **129**, 243.
- S12 Schmidt, M., 1963, *ApJ*, **137**, 758.
- S13 Searle, L., Zinn, R., 1978, *ApJ*, **225**, 357.
- S14 Shapley, H., 1938, *Bull. Harvard Coll. Obs.*, No. 908.
- S15 Shapley, H., 1938, *Nature*, **142**, 715.
- S16 Soderblom, D. R., Duncan, D. K., Johnson, D. R. H., 1991, *ApJ*, **375**, 722.
- S17 Sommer-Larsen, J., Zhen, C., 1990, *MNRAS*, **242**, 10.
- S18 Spitzer, L., Schwarzschild, M., 1953, *ApJ*, **118**, 106.
- S19 Strömgren, G., 1922, *ApJ*, **56**, 265.
- S20 Strömgren, G., 1925, *ApJ*, **61**, 363.
- S21 Strömgren, B., 1966, *ARA&A*, **4**, 433.
- S22 Siegel, M. H., Majewski, S. R., Reid, I. N., Thompson, I. B., 2002, *ApJ*, **578**, 151.
- T1 Twarog, B. A., 1980, *ApJ*, **242**, 242.
- V1 Vandenberg, D. A., Bolte, M., Stetson, P. B., 1990, *AJ*, **100**, 445.
- W1 Walker, A. R., 1994, *AJ*, **108**, 555.

- W2 Wielen, R., 1974, *Highlights in Astronomy*, Vol. 3, p. 395, ed. G. Contopoulos, Reidel, Dordrecht.
- W3 Wielen, R., 1977, *A&A*, **60**, 263.
- W4 Wielen, R., Fuchs, K., 1983, in *Kinematics, Dynamics and Structure of the Milky Way*, ed. W. L. H. Shuter, Reidel, Dordrecht, p. 81.
- W5 Wood, M. A., Oswalt, T. D., 1998, *ApJ*, **497**, 870.
- W6 Wyse, R. F. G., Gilmore, G., 1995, *AJ*, **110**, 2771.
- W7 Wyse, R. F. G., Gilmore, G., Franx, M., 1997, *ARA&A*, **35**, 637.
- Y1 Yoshii, Y., Saiso, H., 1979, *PASJ*, **31**, 339.
- Z1 Zinn, R., West, M. J., 1985, *ApJS*, **55**, 45.
- Z2 Zinn, R., 1985, *ApJ*, **293**, 424.
- Z3 Unavane, M., Wyse, R. F. G., Gilmore, G., 1996, *MNRAS*, **278**, 727.
- Z4 Zuckerman, B., Song, I., 2004, *ARA&A*, **42**, 685.

7.9 HOMEWORK PROBLEMS

Problem 1

Schechter and Caldwell used the Las Campanas Swope 1-m telescope to obtain V , I CCD imaging of 40 square degrees towards the South Galactic Cap. The star catalogue derived from their observations is linked to the NLDS2 data website (<http://www.stsci.edu/~inr/nldsdat.html>). Download the catalogue and plot the $(V, (V-I))$ colour magnitude diagram; select an appropriate colour range, and use the appropriate $(M_V, (V-I))$ calibration from equation (2.1) or (2.2) to derive the density distribution perpendicular to the Plane. What corrections would be employed if the star catalogue were for a field at lower galactic latitude?

Problem 2

Our vantage point within the Milky Way allows us to determine the density distribution of the Galactic Halo from the number–magnitude distribution of individual halo stars. Those stars are too faint to be visible in most external galaxies: discuss how one might identify a halo population in those systems and how one might measure its density distribution.

Problem 3

The NLDS2 data website (<http://www.stsci.edu/~inr/nldsdat.html>) has links to datafiles with colour–magnitude data for three clusters: a metal-poor globular cluster, a metal-rich globular cluster and an old open cluster. Download the data, plot the colour–magnitude diagrams and determine which cluster is which; explain the reasoning behind your decisions.

Problem 4

Given the kinematics listed in Table 6.3, estimate the average relative age of B-A5, dM and dMe dwarfs. Explain how one might set those relative ages on an absolute scale.

Problem 5

Some dwarf galaxies are considered to have formed primarily through a single burst of star formation. What properties would you expect to characterise the abundance distribution of the stellar population in those systems? How would you search for an underlying, older stellar population?

8

The stellar luminosity function

8.1 INTRODUCTION

The stellar luminosity function, $\Phi(M)$ – the number of stars per unit absolute magnitude per unit volume – is one of the fundamental quantities required for understanding star formation and investigating the structure of our Galaxy. In recent years, most emphasis has been placed on determining the shape of $\Phi(M)$ at faint magnitudes – with the ultimate goal of assessing the likely dark-matter contribution of VLM stars and brown dwarfs. However, it was Galactic structure analyses, rather than an interest in the relative number of luminous and faint stars *per se*, that prompted the first derivation of the luminosity function.

The study of the stellar luminosity function has a long and chequered history of contentious debate, reviewed in the first part of this chapter. Later sections synthesise results from recent surveys to provide an estimate of the luminosity function defined by field stars in the Galactic Disk. Chapter 9 discusses the transformation of those results from luminosity to mass.

8.2 THE EARLY YEARS

William Herschel [H10] inaugurated the study of Galactic structure in the late 1770s through his programme of ‘star-gaging’ – counting the total number of stars that drifted through the 15-arcmin diameter field of view of his 20-foot reflecting telescope¹ within a set time, for a grid of reference fields distributed across the celestial sphere. His observations yielded the average surface density of stars brighter than the limiting magnitude of his telescope (about 12th magnitude), for Herschel had no

¹ The focal length of the telescope was 20 feet; the speculum mirror had a diameter of 18 inches.

means of obtaining accurate photometry nor, lacking distance estimates, had he any means of estimating intrinsic luminosities. Herschel adopted the assumption that the stars in his survey had the same average luminosity in each field, and that there was no significant intervening obscuration, and inferred the three-dimensional structure of the Galaxy from the observed surface densities. On that basis, he deduced that the Sun lay in a flattened system that bifurcated towards the Great Rift in Cygnus, a feature we now know to be due to absorption within the Plane of the Galaxy.

The opportunity to apply an approximate distance scale to these observations arrived with the first measurements of stellar trigonometric parallaxes. In 1847 John Herschel undertook observations of apparent brightnesses of 191 southern stars, including a small number that had measured trigonometric parallaxes. By the middle years of the nineteenth century, measurements had been obtained of stars with luminosities ranging from $20 L_\odot$ (Sirius) to $0.1 L_\odot$ (61 Cygni, or Gl 820). These results provided the first quantitative demonstration of the substantial range in intrinsic brightness spanned by stars in the Milky Way galaxy.

The main tool of Galactic structure studies in the late nineteenth century was stellar number-magnitude counts – indeed, this remains one of the most effective means of probing the stellar density distribution. Star-count analysis aims to invert the observed number counts as a function of apparent magnitude, and infer the underlying density distribution. To do so, due allowance must be made for the range of intrinsic luminosities. The formalism for this analysis – the fundamental equation of stellar statistics – was developed by von Seeliger [S7].

The observed number-magnitude distribution of stars, $A(m)$, can be expressed as an integral equation involving a density law, $D(r)$, a volume element, and the stellar luminosity function, $\Phi(M)$, corrected for line-of-sight absorption, $a(r)$,

$$A(m) = \omega \int \Phi(M - a(r)) D(r) r^2 dr \quad (8.1)$$

where ω is the solid angle covered by the observations.

von Seeliger's equation demonstrates that an estimate of $\Phi(M)$ is essential in determining the structure of the Galaxy, $D(r)$, the main focus of early star-count studies. The most influential proponent of this research was the Dutch astronomer, Jacobus Kapteyn, who oversaw the compilation and analysis of one of the most extensive photographic star catalogues yet undertaken, and, through his association with Hale and Mount Wilson Observatory, laid the foundations for the immensely fruitful influx of Dutch astronomers (including van Maanen, Luyten, Kuiper and Bok) into US observatories. Elected to the Chair of Astronomy and Theoretical Mechanics at Groningen in 1878, Kapteyn found himself in charge of an observatory with staff, but little in the way of instrumentation to allow him to assemble the extensive data essential for the pursuit of what he called the 'sidereal problem' – the distribution of stars in space.

Fortunately for Kapteyn, David Gill, Her Majesty's Astronomer at the Cape of Good Hope, soon found himself in entirely the opposite situation. Gill discovered the astronomical potential of photography in 1882, when he assisted a local amateur – Mr Allis – to obtain a picture of the Great Comet of that year,

mounting a camera on one of the observatory telescopes. Gill noticed that the photograph not only captured detailed structure in the cometary tail, but also revealed numerous faint stellar images. Realising the potential for stellar cartography, Gill circulated copies of this and other astronomical photographs among various leading astronomers, including Admiral Mouchez at Paris Observatory who, in turn, encouraged the brothers Henry to produce lenses suitable for astrophotography. The latter circumstance led to the initiation of the all-sky *Carte du Ciel* in 1887. Gill had started his own southern photographic survey – a project discussed in letters to Kapteyn in 1885 [M9]. Kapteyn responded by offering to help with the measurement and reduction of these plates – an offer which eventually led to the production of the *Cape Photographic Durchmusterung* in 1896. (In the meantime, the Board of Visitors of the Royal Observatory, Greenwich, had cancelled all funding for the project, and Gill was forced to devote almost a third of his salary for several years to the completion of the project. Further details are given by Murray [M9] and references therein.)

Kapteyn published one of the first observational determinations of the ‘luminosity curve’ of stars. While the techniques used in those early analyses have been superseded, the results not only track our evolving understanding of $\Phi(M)$, but also offer insight into some of the problems and errors encountered in more recent investigations. Before considering these results, a brief mention of the treatment of interstellar absorption is in order. von Seeliger explicitly used the term $a(r)$ in the star-count equation, but few early studies corrected observations for that effect. Kapteyn, in particular, noted [K3]: ‘Undoubtedly one of the greatest difficulties, if not the greatest of all, in the way of obtaining an understanding of the real distribution of the stars in space lies in our uncertainty about the amount of loss suffered by the light of the stars on its way to the observer.’ Kapteyn generally ignored absorption in his star-count analysis, and the effects are of little importance in luminosity function determinations where the majority of stars lie within 100 pc of the Sun. The question of general interstellar absorption (as opposed to the small, dark nebulae observed by Barnard and others) was not laid to rest until Trumpler’s [T5] studies of the distribution of open clusters.

8.2.1 The method of mean parallaxes

The most straightforward method of deriving $\Phi(M)$ is to count the number of stars as a function of absolute magnitude in a known volume. However, this method requires a complete sample, and it is only in recent years that statistically complete stellar samples can be compiled within even modest volumes. With accurate trigonometric parallaxes, and even photometry, available for only a handful of stars in the early years of this century, luminosity function derivations relied on less direct, statistical techniques.

The method of mean parallaxes – used by Kapteyn in his initial work on the luminosity function [K1] and one of the techniques included in van Rhijn’s [R16] analysis – assumes that one can use the sparse sample of stars with known trigonometric parallaxes as unbiased indicators of the distance distribution

underlying the general number–magnitude–proper motion distribution, $N(m, \mu)$. Taking known stars with measured parallaxes, Kapteyn calculated the mean parallax for a grid of points in apparent magnitude and μ , $\bar{\pi}(m, \mu)$. Within each bin, the distribution of true parallaxes is fitted to a Gaussian distribution² in $\log \pi$. Given the distribution of π for each $N(m, \mu)$ cell, one can integrate over μ to estimate the distance distribution of stars of a given apparent magnitude, and hence derive the absolute magnitude distribution. Summing the number of stars in an absolute magnitude bin within given parallax limits determines the luminosity function.

Figure 8.1(a) shows the luminosity curve derived by Kapteyn from his 1902 analysis (curve V from [K1]). This paper is also notable for establishing the ordinate used in early plots of the luminosity function (the logarithm of the local space density in stars $\text{mag}^{-1} \text{pc}^{-3} + 10$), and for defining absolute magnitude as the apparent magnitude a star would have at a distance of 10 parsecs (parallax of $0.^{\circ}1$),

$$M = m + 5 + 5 \log \pi \quad (8.2)$$

Kapteyn commented, [K1]: ‘The curve appears to reach a maximum about the absolute magnitude 10.5. Whether for fainter stars it will descend as rapidly or more rapidly, and whether it will soon reach a limit, below which no luminous stars exist, are questions for which a knowledge of the number of large proper motion stars of fainter than the ninth magnitude is required.’

Kapteyn’s method of presenting $\Phi(M)$ as the number of stars per unit absolute magnitude prompted one of the first controversies. G. Comstock, director of Washburn Observatory (Madison, Wisconsin), interpreted the densities as given per unit luminosity (that is, a linear, rather than logarithmic, scale). With that interpretation, Kapteyn’s results appeared to disagree sharply with Comstock’s own analysis of trigonometric parallax data [C2]. This prompted an extended and somewhat acrimonious exchange of articles in the journals, setting a prophetic precedent for the field.

8.2.2 The method of trigonometric parallaxes

This is the second method devised by Kapteyn, and was adopted subsequently by van Rhijn [R15], [R16], Bok and Macrae [B5], Trumpler and Weaver [T6], Starikova [S11] and is even discussed at length by Mihalas and Binney [M8]. Given the observed distribution $N(m, \pi)$ for a set of stars $\pi > \pi_{\lim}$, the aim is to correct both for errors due to random observational uncertainties and for systematic effects due to incompleteness in the sample.

Consider the number–magnitude–proper motion distribution, $N(m, \mu)$, for a set of stars with $\mu > \mu_{\lim}$, a subset of which have measured parallaxes. We can construct $N_\pi(M, \mu)$, the number–absolute magnitude–proper motion distribution for the

² Kapteyn gave no explicit justification for the choice of fitting function, but with a total of only 58 parallax stars, a Gaussian distribution is as reasonable a choice as any.

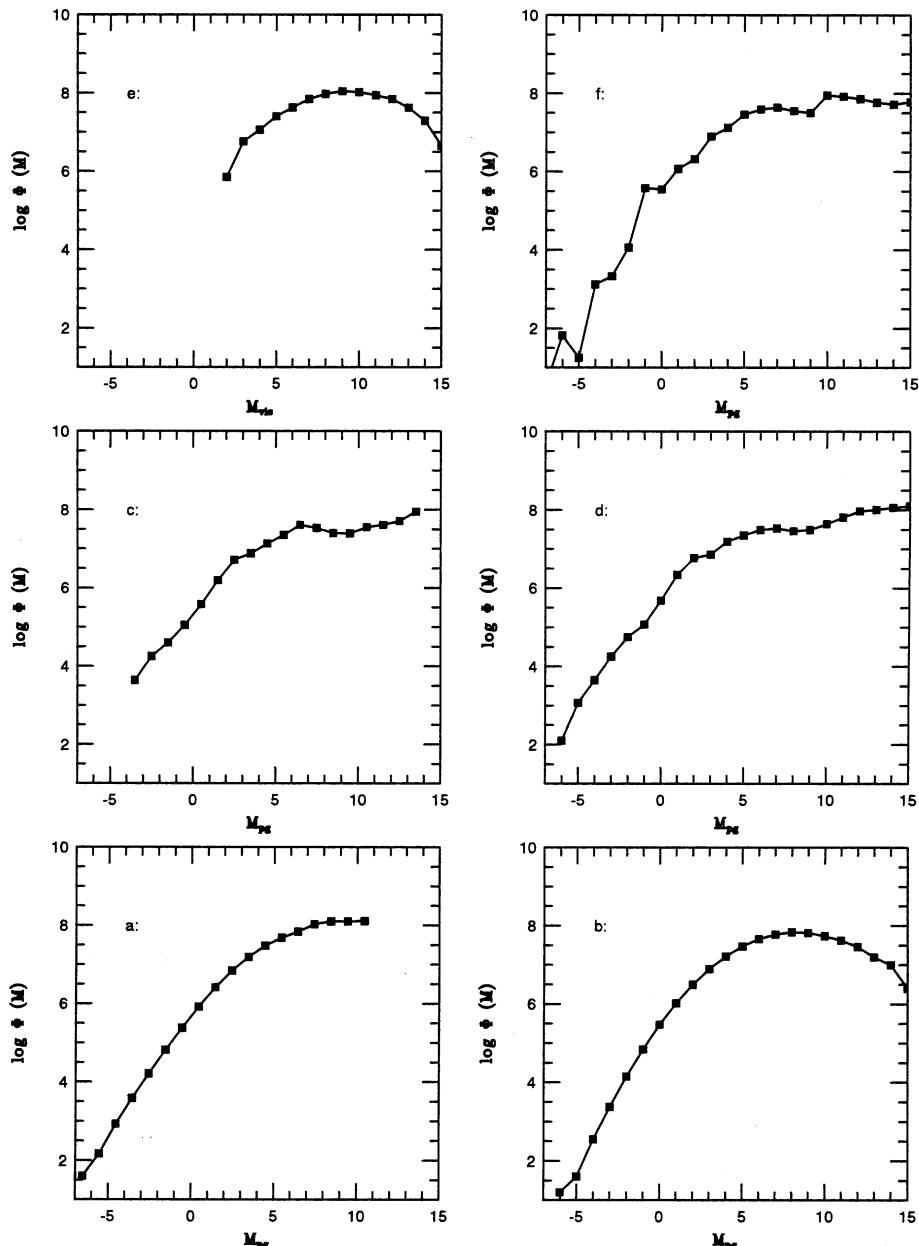


Figure 8.1. The changing form of the luminosity function. The results shown here are derived either using the method of mean parallaxes or the method of mean trigonometric parallaxes and are from: (a) Kapteyn [K1]; (b) Kapteyn and van Rhijn [K4]; (c) van Rhijn [R15]; (d) van Rhijn [R16]; (e) van Maanen [M3]; and (f) Starikova's [S11] analysis. Following the convention of the time, the ordinate is the logarithm of the space density (in stars $\text{pc}^{-3} \text{mag}^{-1} + 10$).

parallax stars. If those stars are an unbiased selection from the whole, then each star with known π represents f stars in total, where

$$f = \frac{N(m, \mu)}{N_\pi(m, \mu)} \quad (8.3)$$

Generally, analysis is limited to the range of $N(m, \mu)$, where f is small (<2).

Two further systematic corrections are required. First, the uncertainty in π is symmetric, but the volume of a shell of given width increases with π^{-2} . The result is that more stars with smaller true parallaxes (larger distances) are scattered into a parallax-limited sample than are scattered out (Section 8.4). The net result is that the average true parallax in a given interval $\pi \pm \delta\pi$ is smaller than would be the case if all measurements were error-free; that is,

$$\langle \pi_0 \rangle < \langle \pi \rangle$$

where π is the observed parallax, and π_0 is the true parallax.

The second correction takes account of stars within the distance limit, $\pi > \pi_{\lim}$, but with proper motions below the sample limit, $\mu < \mu_{\lim}$. These stars have a tangential velocity below the critical value for inclusion in the sample, either through projection effects (that is, most of the relative motion is in the radial velocity) or through their having a space velocity close to the solar motion. In either case, the number of such stars can be estimated by adopting a particular model for the local stellar kinematics (see [T6] and [M8]).

Figures 8.1(b)–8.1(e) show luminosity functions derived using this technique. van Maanen's [M3] study was based on analysis of 2,380 stars with $\mu > 0.^{\circ}5$ per year, 651 with parallax measurements – a significant improvement over the sparse data originally available to Kapteyn in 1902. In general, the later surveys, analysing data taken to faint limiting magnitudes, show the luminosity function peaking at fainter absolute magnitudes than the initial studies, with the space densities constant or slowly declining.

8.2.3 The method of mean absolute magnitudes

Luyten's analyses

This third statistical method was used extensively by Luyten [L7], [L8], [L11], primarily in the analysis of catalogues of high proper-motion stars. Following a suggestion by Hertzsprung, Luyten combined apparent magnitude and proper motion to produce a quantity analogous to absolute magnitude. At first Luyten referred to this quantity as the ‘Hertzsprung index’, M_μ , but later adopted the term ‘reduced proper motion’. Echoing equation (1.23), H , defined as

$$H = m + 5 + 5 \log \mu \quad (8.4)$$

Luyten calibrated this index against absolute magnitude using observations of stars with known trigonometric parallax. Initially, a magnitude term was included

in the equation, although Luyten later settled on a linear equation,

$$\bar{M}_H = a + bH \quad (8.5)$$

Luyten characterised the distribution of residuals about the mean relation as a Gaussian with an r.m.s. dispersion of ~ 1.2 magnitudes in absolute magnitude. Thus, H is useful only as a statistical predictor of absolute magnitude for a group of stars, and does not provide accurate estimates for individual cases. Nonetheless, Luyten applied this calibration to transform the observed distribution $N(H)$ from his proper-motion surveys to a distribution in absolute magnitude, M , using estimates of the velocity distribution to allow for stars with proper motions below the survey limit.

Luyten's initial calibration [L7] of equation (8.5) gave coefficients of $a = -0.69$ and $b = 0.54$ and, analysing data for 749 stars with $\mu > 0.^{\circ}5 \text{ yr}^{-1}$, he derived a distribution $\Psi(H)$ (not to be confused with the mass function, Chapter 9) which reached a maximum at $H \sim 12.5$, implying a maximum in the luminosity function at $M_{pg} \sim 8.5$ (Figure 8.2(a)). This was in broad agreement with the luminosity function derived some three years earlier by Kapteyn and van Rhijn. In contrast, Seares [S6] derived higher densities in his analysis combining proper-motion stars and a larger sample of stars with spectroscopic parallaxes. This result stems primarily from Seares' use of two linear relations to calibrate H against M , with coefficients $b \sim 0.67$ at $H < 8$, and $b \sim 1$ at larger H . The latter calibration places a larger number of stars at fainter absolute magnitudes than in Luyten's analysis, with a corresponding effect on $\Phi(M)$ (Figure 8.2).

Luyten [L8] disputed the validity of Seares' revised calibration, but his own luminosity function analyses evolved in the same direction. The final calibration adopted by Luyten is

$$\bar{M}_H = -3.5 + 0.86H \quad (8.6)$$

and this equation was used in his analysis of both the Bruce proper motion survey, covering the southern skies to $m_{pg} \sim 17.7$, the Palomar proper motion surveys [L5]. The latter extend from the north celestial pole to $\delta = -42^\circ$ and have a limiting magnitude of $m_{pg} \sim 20$. The luminosity functions from those surveys peak at $M_{pg} = 14.5$ and 15.7 magnitudes respectively, and are shown in Figure 8.2(c) and 8.2(d).

Reconsidering mean absolute magnitudes

Since accurate photometry is now available for many nearby stars with well-determined parallaxes, we can re-examine the calibration of \bar{M}_H . First, note that since

$$\mu = \frac{V_T}{\kappa r} \quad (8.7)$$

where V_T is the transverse velocity in km s^{-1} , r is the distance in parsecs and

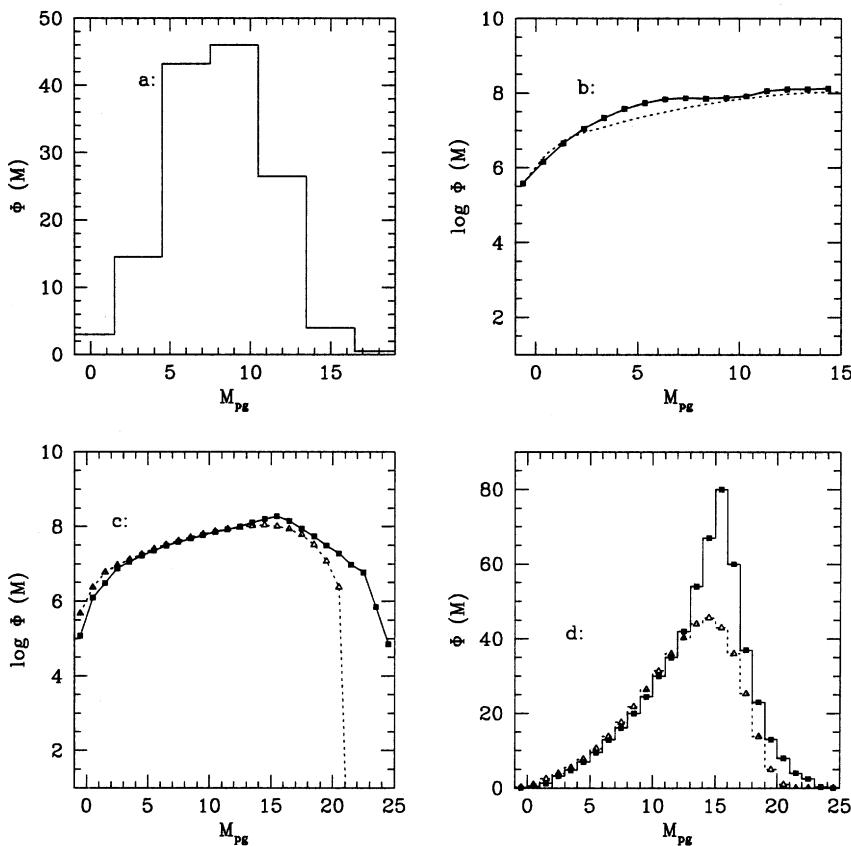


Figure 8.2. Luminosity functions derived using the method of mean absolute magnitude. The data are from: (a) Luyten [L8]; (b) Seares [S6] and (dotted line) Luyten [L9]; (c) Luyten [L9] (dotted line) and [L11] (solid squares); and (d) Luyten [L9], [L11] (triangles Δ and solid line). (a) and (d) present the results as the number of stars within 10 parsecs.

$\kappa = 4.74$, equation (7.4) can be rewritten as

$$\begin{aligned} \langle H \rangle &= m + 5 - 5 \log r + 5/\log V_T - 3.38 \\ &= M + 5\langle \log V_T \rangle - 3.38 \end{aligned} \quad (8.8)$$

This relationship implies that the coefficient ‘ b ’ in equation 8.5 should be 1.0, as in Seares’ faint calibration, unless the mean transverse velocity varies with M_V .

Figure 8.3 plots H_V against M_V for stars within 8 parsecs of the Sun. The best-fit relation is

$$M_V = -2.56 + 0.89H_V, \quad \sigma_M = 1.24 \text{ mag} \quad (8.9)$$

which implies that the average tangential velocity rises from $\sim 18 \text{ km s}^{-1}$ at $M_V \sim 0$ to 35 km s^{-1} at $M_V \sim 12$. This variation is not unreasonable, given the short main

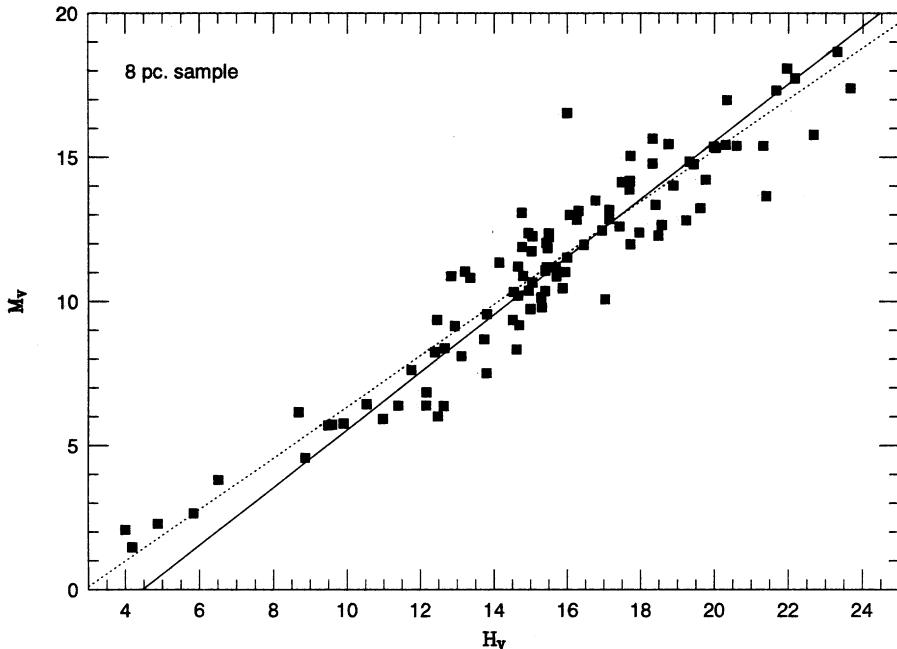


Figure 8.3. The relationship between reduced proper motion, H , and absolute magnitude, M_V , defined by modern observations of stars in the nearby-star catalogue with trigonometric parallaxes accurate to better than 10%. The dotted line is the best-fit relation, equation (8.9); the solid line is equation (8.10), where the slope is set to unity.

sequence lifetimes of early-type stars. However, a more reliable calibration of M_V versus H_V for later-type stars – where the main sequence lifetime exceeds the age of the Galactic Disk – follows if we substitute the appropriate value of $\langle \log V_T \rangle$ in equation (8.8). Averaging $\log V_T$ for stars in the 8-parsec sample gives

$$M_V = -4.46 + H_V \quad (8.10)$$

There is a further effect to consider in interpreting mean absolute magnitude analyses of proper-motion samples. Equation (8.8) shows that the relationship between H and M is single-valued only for a population with a simple uni-modal velocity dispersion. This circumstance holds for a volume-limited sample of nearby stars, since almost all are members of the Galactic Disk. Halo stars, however, have a higher velocity dispersion and negligible net rotation (a substantial solar motion), leading to a higher average tangential motion. As discussed further in Chapter 10, the number of stars contributed by a stellar population to a proper-motion sample is given by

$$N(\mu > \mu_{lim}) \propto \rho_0 V_T^3 \quad (8.11)$$

where ρ_0 is the local space density. Thus, high-velocity halo stars make a disproportionately large contribution to proper-motion selected samples. Moreover,

equation (8.10) does not provide the appropriate absolute magnitude calibration for halo stars [H5], [R4]. Since

$$\langle V_T(\text{halo}) \rangle \approx 6 \times \langle V_T(\text{disk}) \rangle$$

then

$$M_V(\text{halo}) = -8.3 + H_V \quad (8.12)$$

This is, applying an \bar{M}_H relationship calibrated against disk dwarfs to a sample of halo subdwarfs leads to absolute magnitudes which are too faint.

This miscalibration is unimportant for proper-motion surveys with bright magnitude limits (such as the Bruce survey), but becomes increasingly important for surveys extending to fainter magnitudes, such as the Luyten Palomar Schmidt surveys. As will be discussed further in Chapter 10, this bias can be turned to advantage by using broadband colours to isolate halo subdwarfs and degenerate white dwarfs in proper-motion surveys. With photometry of only modest accuracy, however (as in Luyten’s catalogues), the segregation between disk and halo is not evident, and an analysis that adopts a single calibration will necessarily produce an inflated estimate of the number of low-luminosity dwarfs. This effect is evident in Luyten’s Palomar Schmidt analysis, where the low-luminosity tail extends to beyond $M_{pg} = 20$ (Figure 8.2(d)): follow-up spectroscopy has shown that these stars are almost exclusively halo subdwarfs.³

8.2.4 Kuiper’s nearby-star luminosity function

In the late 1920s and early 1930s, the main direction of Galactic structure studies shifted towards determining the distribution of stars within the disk. In part this reflected limitations imposed by the available technology – low-sensitivity photographic plates and hand-driven measuring engines. More emphasis was placed on understanding the behaviour of the stellar luminosity function at brighter magnitudes, than on identifying stars at the lowest luminosities. The final pre-1950 study that is relevant to determining the form of the luminosity function at faint magnitudes is Kuiper’s analysis of a sample of 254 stars with $\pi \geq 0.^{\circ}095$ [K8] – a precursor of the more recent work by Wielen [W5], [W6] and Reid *et al.* [R11], [R18].

Many stars in Kuiper’s sample had trigonometric parallax measurements, and a number of the lower-luminosity stars had distances determined from spectroscopic parallaxes. Based on those data, Kuiper derived a luminosity function similar to Luyten’s results, with a maximum at $M_V \sim 10$ (Figure 8.4). Spectroscopic

³ Luyten was aware of potential problems with his analysis, particularly concerning the Palomar data. There were very few determinations of accurate parallaxes for stars fainter than ~ 16 th magnitude even in 1968, so Luyten had no means of checking the validity of his calibration at those magnitudes; nor were the parameters of the halo population well characterised at that time. Nonetheless, it is surprising that no attempt was made to understand why the coefficient, b , was not unity in any of the calibrations except Seares’ two-segment fit.

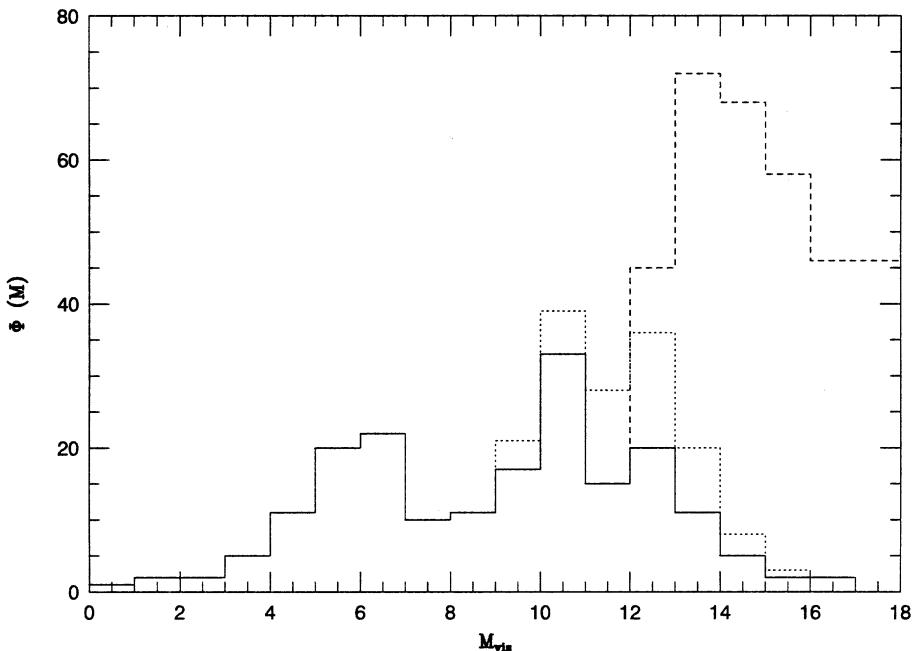


Figure 8.4. The three luminosity functions derived by Kuiper [K11] from the analysis of nearby stars. Histogram a (solid line) is based only on trigonometric parallaxes; histogram b (dotted line) includes stars with spectroscopic parallax distance estimated; and histogram c (dashed line) is from a ‘larger sample’ of proper motion stars.

observations only were available for a larger number of M dwarfs. Besides computing luminosity functions for stars within 10.5 pc (histograms a and b in Figure 8.4), Kuiper also presented preliminary analysis of proper motion catalogues, deriving much higher densities of late-type M dwarfs. This luminosity function is plotted as histogram c in Figure 8.4, and Kuiper commented that: ‘The difference c–b shows how much remains to be discovered even for the nearest stars.’ This passing remark sets the scene for the *cause célèbre* of M dwarfs and the missing mass.

8.3 THE 1970s: A plethora of M dwarfs?

‘If there is a generalisation to be drawn from this whole discussion . . . it is that the principles are easy but the practice is difficult.’

I. King, 1977 [K6]

In his 1966 review of the stellar luminosity function, McCuskey [M1] compared the results from available studies, and found, with two exceptions, excellent agreement. The first exception lay between the van Rhijn [R16] and Luyten [L9] functions for $7 < M_{pg} < 10$, with the van Rhijn $\Phi(M)$ dipping below Luyten’s data. This

discrepancy arises because Luyten smoothed his data to give a continuously increasing function [L8]. The second discrepancy lay at fainter magnitude, and had more far-reaching effects. Sanduleak [S1] had compiled a sample of $\sim 1,200$ late-type dwarfs from an objective prism survey of the North Galactic Cap. Determining distances to those stars using spectroscopic parallaxes, Sanduleak – like Kuiper before him – derived volume densities up to a factor of five higher than Luyten’s results in the interval $11 < M_B < 15$ (Figure 8.5(a)).

This result was not confirmed by a similar high-latitude survey by Klare and Schaifers [K8], and inevitably drew criticism from Luyten. However, one of the main astrophysical problems of the time concerned the search for the ‘missing mass’ in the Galactic Disk. Oort’s [O1], [O2] analyses, and others, of the motions of stars perpendicular to the Galactic Plane implied a gravitational force that required a local mass density, $\rho_{disk}(0)$, of $0.15 M_\odot \text{pc}^{-3}$. Summing the contributions to $\rho_{disk}(0)$ from the visible constituents of the Solar Neighbourhood gave a mass density of $\sim 0.06 M_\odot \text{pc}^{-3}$ from stars, and $0.03 M_\odot \text{pc}^{-3}$ from gas and dust, leaving $\sim 0.06 M_\odot \text{pc}^{-3}$ unaccounted for. Low-luminosity M dwarfs have a high mass-to-light ratio ($M/L > 10$), so numerous late-type M dwarfs could hide substantial mass without adding significantly to the total light in a system.

Support for Sanduleak’s $\Phi(M)$ appeared to come from results based on an entirely different technique. Weistropp [W2] had painstakingly undertaken iris diaphragm photometry of $\sim 13,500$ stars brighter than $V \sim 18$ within a 13.5 square-degree field towards the North Galactic Pole (NGP) – the same region surveyed by Sanduleak, although Weistropp’s data consisted of direct U , B and V plates taken with the 48-inch Oschin Schmidt at Palomar. Calibrating the iris photometry against a sequence of stars with known magnitudes (from photoelectric observations), she used the $(M_V, (B-V))$ relationship (Figure 2.16) to estimate photometric parallaxes (that is, given $(B-V)$, estimate M_V), and hence distances and space densities. Her results suggested that the Luyten function underestimated the number of faint M dwarfs by a factor of between 5 and 10 (Figure 8.5(b)). Further apparent corroboration was supplied by the identification of numerous very red (colour class +4) stars from Giclas *et al.* [G2] visual scans of blue and red plates from the Lowell Observatory survey, and in new objective prism surveys by Smethells [S9] and Sanduleak [S2].

The complication facing this hypothesis lay in explaining why these late M dwarfs were not included in Luyten’s proper-motion surveys. Assuming that the absolute magnitude estimates (from spectroscopy and $(B-V)$) were reliable, the only reasonable explanation seemed to be that those stars formed a low velocity-dispersion component in the Galactic Disk. With low space motions, the stars would also have low proper motions, and would therefore not contribute to a sample selected by proper motion. This hypothesis seemed to be consistent with proper-motion measurements of a subset of the Sanduleak M dwarfs by Pesch and Sanduleak [P4], Pesch [P3], Gliese [G5] and Murray and Sanduleak [M10]. Although each study was based on only ~ 20 stars, all had small proper motions, consistent with very low space motions *if the M dwarfs were actually at the distances indicated by the spectroscopic parallaxes*. Murray and Sanduleak, for example, estimated a transverse velocity dispersion of only 10 km s^{-1} , from which they inferred a vertical dispersion

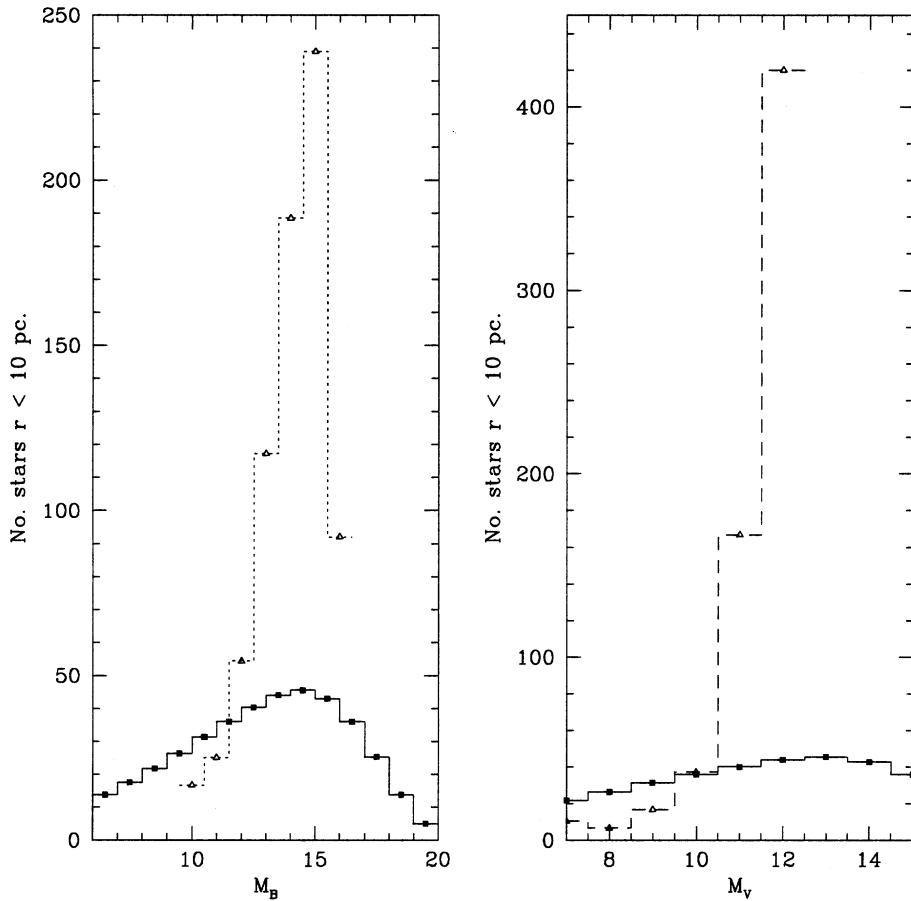


Figure 8.5. Proper motion-based luminosity functions compared with spectrophotometric analyses. The left panel compares Luyten's (1939) $\Phi(M_B)$ against Sanduleak's NGP objective prism survey (dotted line); the right panel compares $\Phi(M_v)$ from Luyten (1968) (solid line) against Weistrop's results (dotted line).

$\sigma_W < 8 \text{ km s}^{-1}$. The latter result implies a vertical density distribution similar to B stars, and a local mass density of $\sim 0.05 M_\odot \text{ pc}^{-3}$ – sufficient to account for almost all of Oort's ‘missing mass’. Gliese [G5] reached similar conclusions from analysis of the motions of 75 Lowell colour-class +4 stars towards the SGP.

While these results appeared to answer the ‘missing mass’ question, they posed conundra of their own. First, the existence of a substantial, low velocity-dispersion population of M dwarfs implies either a correspondingly large recent burst of star formation, confined, moreover, to low-mass stars, since an excess of solar-type stars would be very hard to hide; or that velocity dispersion and age were not correlated parameters [O3]. Second, while the presence of so much mass in a very thin disk might help explain the Oort ‘missing mass’, it raised severe problems for maintaining

the stability of the disk as a whole; essentially, there appeared to be insufficient kinetic energy to stop gravitational collapse [S3]. Finally, Luyten – who vehemently opposed the revised luminosity functions in a series of inimitable papers ([L12] *et seq.*) – pointed out that the number densities derived at faint absolute magnitudes implied at least 800 stars within 5 pc of the Sun (as compared with 56 known) and 29 stars within 2 pc, where only α Cen AB, Proxima Cen and Barnard's Star are known.

‘There is no trusting appearances.’
R. B. Sheridan

The M dwarf missing mass hypothesis started unravelling in the mid-1970s. Jones [J3] used narrowband photometry to estimate photometric parallaxes for 19 of the 21 stars in the [M10] SGP sample. He derived a mean parallax of 0''.009 while [M10] had estimated 0''.021, implying over a factor of two increase in the derived velocity dispersions. These results – later confirmed by broadband *BVRI* photometry by Weistropp [W3] – were not consistent with the Sanduleak stars being members of a low velocity-dispersion population, while the larger distances implied not only brighter absolute magnitudes than indicated by the spectroscopic parallaxes, but a lower space density.

The most significant blow, however, came from a reassessment of Weistropp’s photometry by Faber *et al.* [F1] and by Weistropp [W3] herself. This revealed the presence of systematic errors in both the $(B-V)$ colours of the standard stars and in the photographic $(B-V)$ colours used to infer absolute magnitudes and photometric parallaxes. The sense of the error was a colour term such that colours were systematically too red for the reddest stars. The error was only ~ 0.2 magnitudes at $(B-V) \sim 1.7$ magnitudes but the $(M_V, (B-V))$ colour-magnitude diagram has a slope of $\sim 8\text{--}10$ for M dwarfs, so this error led to an inferred M_V too faint by 1.5–2 magnitudes. Thus, a 15th magnitude M dwarf with $M_V = 11$ at ~ 60 pc was being misidentified as an $M_V = 13$ dwarf at 25 pc. Since the result of this error was both to overestimate the absolute magnitude (too faint) and underestimate the distance, the result was a substantial overestimate of the local density of late-type M dwarfs. Once the calibration was corrected, the number densities fell in line with those expected from a luminosity function like Luyten’s $\Phi(M)$.

Similar problems beset the objective prism surveys. Luyten [L12] was amongst the first to suggest, that, given the steep relation between M_V and spectral type (Figure 2.14), a small systematic error in classification could account for high inferred number densities. Indeed, Pesch (P3) had noted already that: ‘there seems to be a tendency to assign too late a type to objective-prism spectra of M stars near the plate limit.’

Accurate broadband photometry confirmed the presence of these systematic errors. Weistropp [W4] and Reid [R3] both obtained photoelectric *BVRI* data of the 186 M dwarfs catalogued by Smethells. A comparison between $(V-I)$ photometric parallaxes and Smethells’ $(M_V, \text{spectral type})$ relation shows that the spectroscopic absolute magnitude estimates are too faint by ~ 0.5 magnitudes at type K7, and by ~ 3 magnitudes for M5 and later. A number of these stars have since been observed by the astrometric satellite *Hipparcos*, and the resulting trigonometric

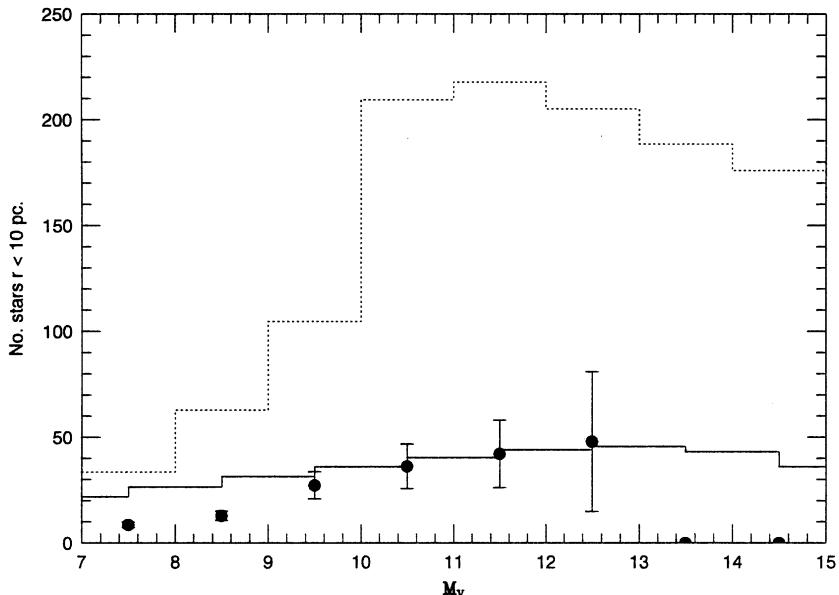


Figure 8.6. A comparison of the luminosity function derived by Smethells (dotted histogram); Reid's [R4] re-analysis of the same data, using $(V-I)$ photometry rather than low-resolution prism spectroscopy (solid points); and the Luyten [L11] $\Phi(M)$ (solid histogram).

parallaxes confirm the brighter absolute magnitudes of the $(V-I)$ calibration. Recalculating the luminosity function using the photometric parallaxes gives results indistinguishable from the Luyten or van Rhijn $\Phi(M)$ (Figure 8.6).

In a similar manner, BVR I photometry of stars from the NGP [S2] and SGP [P4] objective prism surveys demonstrated that only $\sim 10\%$ of the stars are actually fainter than $M_V \sim 13$, rather than the originally posited $\sim 50\%$ [R10], [P5]. Other surveys are similarly affected. Finally, catalogues of red dwarfs derived from visual estimates of colour classes (as in the Lowell survey) proved to be over-optimistic in the classification. Luyten originally pointed out that not only had he classified many of the Lowell +4 stars as type K and K-M in his surveys, but that a substantial number of the stars which lay in the overlap region between two Lowell fields were classified as colour class +3 (red), as opposed to class +4 (extremely red), on the second plate. Accurate photometry showed that nearly all of these stars were early-type M dwarfs at distances of 50–100 pc, rather than M5 to M8 dwarfs within a few parsecs of the Sun.

‘Everthing’s got a moral, if you can only find it.’
Lewis Carroll

Following the resolution of what Luyten unkindly termed the ‘Weistroop Watergrate’, the IAU devoted a joint discussion session of the 1976 General Assembly to the matter of the local density of M dwarfs. Opening that session, Gliese enquired

rhetorically whether all of the efforts of the last few years had produced any worthwhile result – or had they simply turned full circle, returning the subject to the position held in 1963? Answering his own question, Gliese argued that there had been some progress, notably the many observational programs that had been prompted by the controversy, several of which used innovative techniques. Weisstrop's survey, for example, was the first to apply faint, accurate number–magnitude star-counts to Galactic structure work on the wide-field scale made possible by 48-inch Schmidt plates.

One can also point to this incident as an excellent example of the scientific method actually at work – messier and more subjective than the idealised objective comparison of observation and theory, but achieving the same result. New observations appeared to solve a theoretical problem (in this case, disk ‘missing mass’), and cast doubt on an established viewpoint (the van Rhijn/Luyten $\Phi(M)$). These initial results were apparently supported by independent data, prompting a certain amount of uncritical bandwagoning, and a growing acceptance of the new hypothesis, despite problems (some valid, some histrionic) raised by the old guard. As time passes, the new result attracts more detailed scrutiny, and either withstands those more rigorous tests or, as in the present case, reveals an increasing number of inconsistencies and is rejected. Yet the attention generated by the now-invalidated hypothesis leads to better understanding of the subject as a whole.

A prime lesson learned from this affair is that systematic and random errors generally conspire to enhance the apparent number of unusual objects, particularly in large-scale surveys. One must be careful in sifting through such data, and highly sceptical of potentially exciting scientific discoveries. In short: be wary if your survey finds too much of what it was looking for.

8.4 SYSTEMATIC BIASES

Before summarising results from more recent luminosity function studies, two sources of systematic error should be considered.

8.4.1 Lutz–Kelker corrections

Any trigonometric parallax measurement has an associated uncertainty, usually assumed to be drawn from a Gaussian distribution, rms dispersion σ_π . The measured parallax is the best estimate of the distance to an individual star, but uncertainties can lead to systematic biases when data are combined in a sample based on the value of the parallax. Writing the true parallax as π_0 , the distribution of the observed parallax, π , for a given true value is

$$g(\pi|\pi_0) = \frac{1}{\sqrt{V\sigma_\pi}} \exp\left(-\frac{(\pi - \pi_0)^2}{2\sigma_\pi^2}\right) \quad (8.13)$$

where we have defined $V = 2\pi$ to avoid confusion. We require the inverse distribu-

tion; that is, the distribution of *true* parallaxes given the observed distribution of π , $g(\pi_0|\pi)$. To calculate this, the overall parallax distribution, $N(\pi)$ must be taken into account. If the number of stars in the range $\pi_0 \pm d\pi$ were constant for all π , then $g(\pi|\pi_0) = g(\pi_0|\pi)$. However, if the sample has uniform density, then the number of stars within a given shell, r to $r + dr$, is given by

$$N(r) = 2Vr^2dr \quad (8.14)$$

Hence, the number of parallaxes in the range π_0 to $\pi_0 + d\pi$ is

$$N(\pi_0) d\pi = 2V \frac{d\pi}{\pi_0^4} \quad (8.15)$$

The number of stars increases with decreasing parallax, so if we consider an observational sample defined by $\pi \pm d\pi$, a larger number of more distant (smaller true parallax than $\pi_0 - d\pi$) stars will be scattered into the sample than are either scattered out of the sample or are scattered into the sample from smaller distances (larger true parallax than $\pi_0 + d\pi$). If we select stars from this sample based on $\pi > \pi_{lim}$, then $\langle \pi \rangle > \langle \pi_0 \rangle$: the average distance is underestimated, and the true mean absolute magnitude is brighter than the formal value calculated directly from the individual parallaxes.

This effect was recognised by Russell and by Trumpler and Weaver, but was first quantified by Lutz and Kelker [L6], who showed that the resulting bias is a strong function of the accuracy of the parallax measurement. Writing $Z = \pi_0/\pi$, then for a uniform density distribution

$$g(\pi|\pi_0) \propto G(Z) = Z^{-4} \exp\left(\frac{-(Z-1)^2}{2(\sigma_\pi/\pi)^2}\right) \quad (8.16)$$

Integrating this distribution gives the mean offset as a function of σ_π/π , the Lutz–Kelker bias (Figure 8.7)

$$\Delta M_{LK} = \langle \bar{M}_t - M_{obs} \rangle = \left\langle 5 \log \frac{\pi}{\pi_0} \right\rangle \quad (8.17a)$$

where \bar{M}_t is the best estimate of the true absolute magnitude and the appropriate value to use in combining data statistically – for example, when deriving a mean colour–magnitude relationship.

Lutz and Kelker's original calculations were specifically for a parallax-selected sample drawn from a uniform distribution, $N_C(\pi_0) \propto \pi_0^{-4}$ or $N_C(r) \propto r^3$, where N_C is the cumulative distribution. Hanson [H4] showed that selection criteria, such as proper motion limits or magnitude limits, modify the underlying distribution $N(\pi_0)$ by removing stars at large distances. These are parallax-independent effects; as a result, they reduce the likelihood of including stars with $\pi < \pi_0$ in a parallax-selected sample, and therefore reduce the Lutz–Kelker bias. Hanson has derived the following general expression for the Lutz–Kelker bias corresponding to a parent

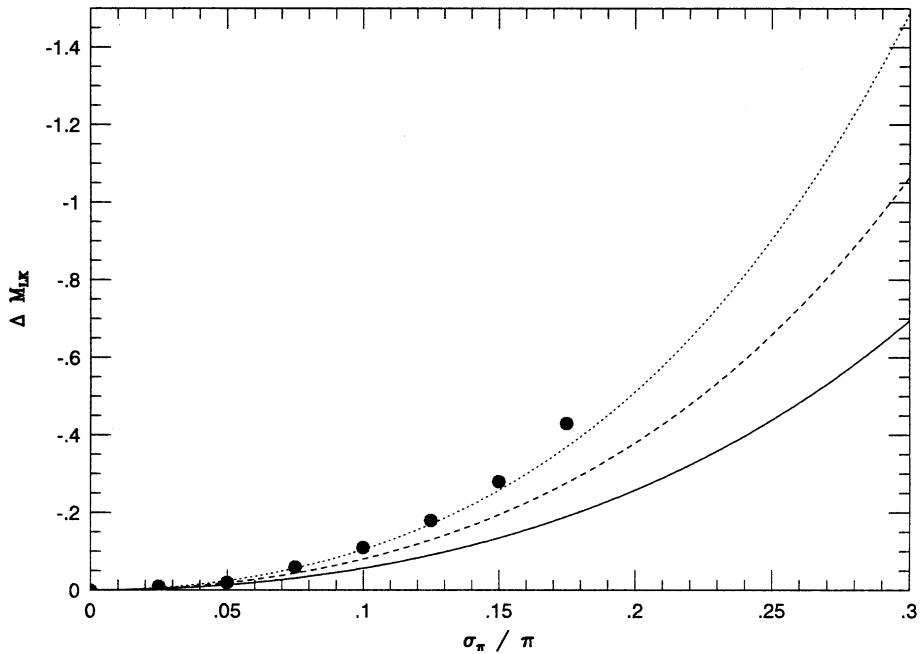


Figure 8.7. Lutz–Kelker corrections. The solid points mark the original [L6] calculations for a uniform density distribution; the solid line is Hanson’s approximation for a power-law parallax distribution, $N(\pi) \propto \pi^{-n}$, with $n = 2$; the dashed line for $n = 3$; and the dotted line for $n = 4$.

population with a power law distribution, $N(\pi_0) \propto \pi_0^n$:

$$\Delta M_{LK} = -2.17 \left[(n + 0.5) \left(\frac{\sigma_{\pi}}{\pi} \right)^2 - \left(\frac{6n^2 + 10n + 3}{4} \right) \left(\frac{\sigma_{\pi}}{\pi} \right)^4 \right] \quad (8.17b)$$

The corrections are smaller for smaller n (Figure 8.7), but still rise sharply with increasing σ_{π}/π for parallax precision worse than $\sim 15\%$.

Lutz–Kelker corrections have been misused extensively in the literature. Five important points should be made:

- LK corrections are appropriate only for a sample selected based on the value of the measured parallax. For example, if one selects all Cepheid variables in the *Hipparcos* catalogue with parallaxes measured to an accuracy of 20%, Lutz–Kelker bias is present; if the sample includes all Cepheids brighter than 10th magnitude, LK bias can be avoided.
- ΔM_{LK} depends on the parallax distribution of the parent sample, not the parallax distribution of the Galactic population from which that sample is drawn. The appropriate distribution can be estimated from the cumulative

proper motion distribution $N_C(\mu)$; since $\mu \propto \pi$, if $N_C(\mu)$ can be represented as a power law μ^{-n} , then the parallax distribution is $N_C(\pi) \propto \pi^{-n-1}$.

- If the parent sample is modified based on criteria that are independent of π , such as selection by apparent magnitude or spectroscopic line strengths, then ΔM_{LK} is also modified. There is no unique value for individual stars; ΔM_{LK} depends on context.
- The larger the LK correction, the larger the uncertainty in the correction; uncertainties are minimised by restricting analysis to stars with parallaxes measured to high precision ($\sigma_\pi/\pi < 0.1$).
- LK corrections are applied to statistical estimates derived from an ensemble of data; they do not provide improved distance estimates for individual objects.

8.4.2 Malmquist bias

Classical Malmquist bias

The most frequently applied statistical correction in astronomy is the classical version of Malmquist bias, which is important not only for number counts and luminosity function work – both Galactic and extragalactic – but also for calibrating distance determinations. Malmquist's [M4] original formulation was designed to estimate the frequency distribution of absolute magnitudes for stars of a given apparent magnitude, and given spectral type (or colour). Consider the case of stars of spectral type S, which have a mean absolute magnitude M_0 . Since there is an intrinsic dispersion to the (M, S) relationship, stars with apparent magnitudes in the range $m \pm \delta m$ magnitudes are drawn from a range of absolute magnitudes, and hence the magnitude limit $m + \delta m$ corresponds to a larger effective distance limit for the intrinsically more luminous stars. If the convolution of the density law and the volume element is such that the number of stars increases with distance, as is usually the case, then this leads to a magnitude-limited sample including a higher proportion of stars with $M < M_0$ than with $M > M_0$. As a result, the mean absolute magnitude of the observed sample is biased to a brighter absolute magnitude than M_0 .

Malmquist quantified the extent of the bias under the assumption that the intrinsic luminosity function for stars of a given spectral type could be characterised as a Gaussian, dispersion σ about a mean magnitude, M_0 ,

$$\Phi_S(M) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(M - M_0)^2}{2\sigma^2}\right) \quad (8.18)$$

If $A(m, r)$ is the joint distribution of stars as a function of apparent magnitude and distance, and $M(m, r)$ is the absolute magnitude distribution, then the mean absolute magnitude of the observed sample is given by

$$\bar{M}(m) = \frac{\int A(m, r) M(m, r) dr}{\int A(m, r) dr} \quad (8.19)$$

and $M(m, r) = m + 5 - 5 \log r$ (Malmquist assumed no interstellar absorption, but the equations can be re-written to take that into account; see [M8]). The number magnitude distribution is

$$A(m, r) dr = \omega \Phi_S(M) D_S(r) r^2 dr = dA_S(m) \quad (8.20)$$

where $D_S(r)$ is the density law, ω is the solid angle and $A_S(m)$ the number-magnitude distribution of stars of spectral type S. Substituting in equation (8.19) gives

$$\bar{M}(m) = \frac{\omega \int (m + 5 - 5 \log r) \Phi_S(M) D_S(r) r^2 dr}{A_S(m)} \quad (8.21)$$

with

$$A_S(m) = \omega \int \Phi_S(M) D_S(r) r^2 dr \quad (8.22)$$

Since $\Phi_S(M)$ is given by equation (8.18), we can write

$$\begin{aligned} \frac{dA_S(m)}{dm} &= \omega \int \frac{d\Phi}{dm} D_S(r) r^2 dr \\ &= -\frac{\omega(M - M_0)}{\sigma^2} \int \Phi_S(M) D_S(r) r^2 dr \end{aligned} \quad (8.23)$$

Combining equations (8.23), (8.22) and (8.21), we have

$$\sigma^2 \frac{dA_S(m)}{dm} = -A_S(m) \bar{M}(m) + M_0 A_S(m) \quad (8.24)$$

and so

$$\begin{aligned} \bar{M}(m) &= M_0 - \frac{\sigma^2}{A_S(m)} \frac{dA_S}{dm} \\ &= M_0 - \frac{\sigma^2}{\log_{10} e} \frac{d \log A_S}{dm} \end{aligned} \quad (8.25)$$

For a uniform density distribution, star-counts increase as $10^{0.6m}$, so this becomes

$$\bar{M}(m) = M_0 - 1.38\sigma^2 \quad (8.26)$$

The mean absolute magnitude of the sample, $\bar{M}(m)$, is therefore brighter than the mean absolute magnitude of all stars of spectral type S, M_0 , by $-1.38\sigma^2$. At high Galactic latitude ($b > 40$) the density law of the Galactic Disk leads to a shallower slope in the number counts. For a single exponential distribution,

$$A(m) \propto \left(\frac{z}{h} \right)^3 \exp \left(-\frac{z}{h} \right) \quad (8.27)$$

where h is the scale height, and z is the height above the plane, giving a Malmquist bias [S15] of

$$\bar{M}(m) = M_0 - \frac{0.2}{\log_{10} e} \sigma^2 \left(3 - \frac{z}{h} \right) \quad (8.28)$$

Thus, in the case of a simple exponential density law, $\rho(z)$, the Malmquist bias falls to zero at 3 scale heights. However, $\rho(z)$ in the disk is best represented as a convolution of several components (Chapter 7). Star-counts at high Galactic latitude have a typical logarithmic slope closer to 0.4 for magnitudes $13 < V < 20$, so

$$\bar{M}(m) = M_0 - 0.92\sigma^2 \quad (8.29)$$

In photometric parallax work, the dispersion in the main sequence colour–magnitude relation is typically ~ 0.4 magnitudes, giving Malmquist bias of ~ 0.15 magnitudes. That is, if we consider stars within a given range in colour, classical Malmquist analysis implies that the mean absolute magnitude within that range is 0.15 magnitudes brighter than would be inferred from the direct colour–magnitude relation.

Malmquist bias in a continuous distribution

Classical Malmquist analysis treats the luminosity function as if it were a series of discrete intervals, each characterised by a given spectral type or colour. The analysis allows only for uncertainties due to dispersion in the absolute magnitude, whereas the mean absolute magnitude is also affected by observational uncertainties in the measured colour. Stobie *et al.* [S15] have demonstrated that applying the classical Malmquist method to a photometric parallax survey does not provide an unbiased estimate of the luminosity function.

A direct consequence of Malmquist bias is that, by biasing the mean absolute magnitude to brighter values, one effectively samples a larger volume, leading to an increase in the number of stars, ϕ , assigned to a given bin in $\Phi(M)$. If we adopt the assumption of a uniform space density, then the fractional increase in the luminosity function is given by

$$\frac{\Delta\phi}{\phi} = \left(\frac{0.6\sigma}{\log_{10} e} \right)^2 - \left(\frac{0.6\sigma^2}{\log_{10} e} \right) \frac{\phi'}{\phi} \quad (8.30)$$

where ϕ' is $d\phi/dM$, the slope of the luminosity function. The first term arises from the additional volume sampled due to the underestimate in the mean absolute magnitude, and the second term measures the effect of sampling the luminosity function at a different value, $M_m < M_0$. The latter term is zero for a flat (uniform density) luminosity function.

The above correction is appropriate for standard Malmquist analysis, but allowing for uncertainties introduced by errors in the measured colour reduces the effective Malmquist bias [S15]. The fractional change in $\Phi(M)$ becomes

$$\frac{\Delta\phi}{\phi} = \frac{1}{2}\sigma^2 \left[\left(\frac{0.6}{\log_{10} e} \right)^2 - \left(\frac{1.2}{\log_{10} e} \right) \frac{\phi'}{\phi} + \frac{\phi''}{\phi} \right] \quad (8.31)$$

The final term is negligible and can be ignored, while the second term, in ϕ/ϕ , is identical to that in equation (8.30). However, the first term, tied to the change in sampling volume, is half the value derived from the classical analysis. The net result is that directly applying standard Malmquist methods will lead to *over-correction* of the observed densities, producing an *underestimate* of $\Phi()$ by a factor of $\frac{1}{2} \left(\frac{0.6}{\log_{10} e} \right)^2 \sigma^2$, or $\sim 15\%$ for $\sigma = 0.4$ magnitudes.

8.5 MODERN TIMES I: FIELD-STAR SURVEYS

‘If the fool would persist in his folly he would become wise.’

William Blake

Recent investigations of $\Phi(M)$ for low-mass dwarfs rely on one of three techniques: star-counts in the immediate Solar Neighbourhood; application of photometric parallaxes to deep, wide-field field-star surveys; and surveys for low-luminosity main sequence stars in open clusters. Each method offers its own advantages and disadvantages. This section concentrates on the first two techniques and their application to M dwarf surveys. The following section describes recent work on extending those statistical analyses to include lower luminosity M, L and T ultracool dwarfs in the Solar Neighbourhood. Open cluster studies are reviewed later in this chapter.

8.5.1 A census of the Solar Neighbourhood

Concentrating on the nearest stars to the Sun seems parochial at first sight. However, as discussed in Section 7.5.1, the gradual diffusion of Galactic orbits and the position of the Sun between spiral arms means that a local sample should include stars of all ages since the formation of the Galactic Disk ($\tau \sim 10$ Gyr), formed in molecular clouds at Galactic radii of $\sim 4\text{--}12$ kpc. Hence these stars provide a fair sampling of the average luminosity function of main sequence stars (average age, 5 Gyr) in the Galactic Disk population. Nearby stars also offer the best opportunity of discovering companions in multiple systems. As John Faulkner expresses it: ‘search locally, think globally.’

The most difficult step is to identify a complete sample of the nearest stars, particularly at faint magnitudes and low Galactic latitudes where the star density exceeds 1,000 stars deg⁻². Consequently, these analyses are restricted to the immediate Solar Neighbourhood. Kuiper’s catalogue (Section 8.2.4) was limited to $\pi > 0.^{\circ}095$, or $r < 10.5$ pc, and consisted mainly of stars brighter than 15th magnitude, identified as candidate nearby stars based on their having substantial annual proper motions. Larger telescopes, more sensitive detectors and new techniques led to more discoveries, and in the 1950s Wilhelm Gliese took on the task of compiling and maintaining a nearby-star catalogue.

Gliese's first compilation – published in 1957 as the *Catalogue of Nearby Stars* (CNS1 [G3]) – is limited to stars identified as being within 20 pc, and includes astrometric and photometric data for 1,094 stars in 915 systems (Gl 1 to Gl 915). By 1969, and the publication of the second catalogue (CNS2 [G4]), the list had expanded to include 1,890 stars in 1,529 systems, with a parallax limit of $\pi > 0.^{\circ}040$ ($r < 25$ pc). Prompted by the publication of the CNS2, Woolley (at that time the Astronomer Royal) oversaw the compilation of a rival RGO catalogue [W7], which includes a further 259 stars (in 216 systems) from diverse (and, in some cases, dubious) sources. Many of these additional stars, together with newer discoveries were included in a supplement to the CNS2, published by Gliese and Jahreiss [G6]. The most recent catalogue – the CNS3 [G7], includes data for approximately 3,820 stars in 3,264 systems (some binary companions are not listed separately in the current version of the new catalogue). Thus, the Solar Neighbourhood stellar census more than tripled over a 40-year period.

The stars in the CNS3 are drawn from many sources – primarily proper motion surveys and, in particular, Luyten's catalogue of stars with $\mu > 0.^{\circ}5\text{yr}^{-1}$ (the LHS catalogue). Other stars originate from objective prism surveys, wide-field photometric surveys, and, most recently, X-ray surveys, notably the ROSAT all-sky survey [S5]. Barely 60% of the systems had trigonometric parallax measurements when the catalogue was compiled, and of those only half were measured to a precision $\sigma_\pi/\pi < 15\%$. Most of the remaining stars had photometric or spectroscopic parallaxes, although some could be said to have entered the catalogue more through hearsay and rumour than through direct observational evidence of proximity. Uniform distance estimates were clearly required, and have been supplied from two sources: a spectroscopic survey of late-type dwarfs and astrometry by the *Hipparcos* satellite.

The PMSU spectroscopic survey ([R11], [H8]) includes 2,075 of the 2,200 candidate K and M dwarfs in the CNS3. 70 of these stars (mainly from the RGO catalogue) prove to be either misclassified earlier-type (AFG) stars or M giants – in either case lying at distances of 50–1,000 pc. Similarly, spectroscopic parallax estimates place many of the genuine K and M dwarfs beyond the nominal 25-pc limit of the CNS3, while *Hipparcos* astrometry of brighter FGK dwarfs also eliminates almost one-third as outwith the distance limits [R18]. This is not unexpected, given that both Malmquist and Lutz–Kelker bias tend to produce an overestimate in π and an underestimate in distance. Jahreiss is currently compiling the CNS4 based on these new data.

Both *Hipparcos* and PMSU are more effective in removing misclassified stars than in adding previously unrecognised nearby stars. The *Hipparcos* satellite measured parallaxes of around 118,000 stars to $V \sim 12$, including almost every star brighter than $V = 7.5$, but only a subset of fainter stars. The latter were pre-selected in 1982, so many nearby stars not suspected of being within 25 pc at that time were not observed by *Hipparcos*. Similarly, the PMSU sample was drawn from the CNS3 and did not add new candidates. Given these limitations, how complete is the nearby-star sample?

Completeness

Statistics are only as reliable as the parent sample. As already discussed in Chapter 2, the southern sky is less well surveyed than the north, as is evident when considering the distribution with declination of systems in the CNS3. Table 8.1 lists the total numbers for four equal areas, considering stars brighter than and fainter than $M_V = 7.5$. The incompleteness amongst faint stars at southern declinations is obvious.

The simplest method of assessing completeness is to determine where the cumulative distribution function of stellar systems, $(\log(N), \log(r))$, drops below a straight line, slope 3 – the relationship expected for a uniform distribution. This technique was applied by Wielen and collaborators [W5], [W6] in analysing the CNS2 and its supplement. Those studies are notable for identifying what has become known as the ‘Wielen dip’ at $M_V = +7$, a feature now recognised as introduced by the physics underlying the (mass, M_V) relationship (Sections 3.5.2 and 9.4). As noted in Section 8.2.3, this feature is present in van Rhijn’s analysis [R16], but was removed intentionally by Luyten. The completeness limits adopted in determining $\Phi(M_V)$ are listed in Table 8.2.

An alternative method of testing completeness is to use the V/V_{max} statistic devised by Schmidt [S4]. For each star, the ratio r^3/r_{max}^3 is calculated, where r is the observed distance, and r_{max} is the maximum distance that allows the star to remain in

Table 8.1. The declination distribution of CNS3 stars.

	$> +30^\circ$	$+30^\circ$ to 0°	0° to 30°	$< -30^\circ$
$M_V < 7.5$	256	218	236	228
$M_V \geq 7.5$	681	722	507	351

Table 8.2. Completeness limits for nearby-star analyses.

Wielen/CNS2			PMSU/CNS3		
M_V	δ_{lim}	r_{lim}	M_V	δ_{lim}	r_{lim}
≤ 7.5	All sky	20 pc	≤ 8.0	$\geq -30^\circ$	25 pc
7.5–9.5	$\geq -30^\circ$	20 pc	8.0–9.0	$\geq -30^\circ$	22 pc
9.5–11.5	$\geq -30^\circ$	10 pc	9.0–10.0	$\geq -30^\circ$	20 pc
11.5–17.5	$\geq -30^\circ$	5 pc	10.0–13.0	$\geq -30^\circ$	14 pc
			13.0–14.0	$\geq -30^\circ$	12 pc
			14.0–15.0	$\geq -30^\circ$	10 pc
			15.0–18.0	$\geq -30^\circ$	5 pc

the sample. If the sample is uniformly distributed, then as many stars are expected within the inner half of the volume as within the outer, and the average value of $\langle V/V_{max} \rangle = 0.5$. If the sample is incomplete at large distances,⁴ then $\langle V/V_{max} \rangle < 0.5$. This method can be adapted to construct a luminosity function, since each star makes a contribution of $1/V_{max}$ to the total density.

The V/V_{max} method was used by Upgren and Armandroff (U1) to test whether the ‘Wielen dip’ at $M_V \sim 7$ could be explained by incompleteness in the nearby-star sample. They found no evidence for significant differential incompleteness. Armandroff [A2] later extended this analysis to consider general star-counts at high Galactic latitude, and provided the first quantitative indication that the ‘Wielen dip’ is a global feature of the disk luminosity function. Reid *et al.* [R11] use both V/V_{max} and $(\log(N), \log(r))$ measurements to determine completeness limits for the PMSU survey (see Table 8.2).

Both of the $(\log(N), \log(r))$ and V/V_{max} tests identify distance limits within which a sample of stellar *systems* can be taken as statistically complete; they do not address the related question of whether every component in each system has been identified. This question is not amenable to simple analysis, since not every system has been subjected to the same degree of attention from high-resolution imaging and/or radial-velocity surveys. Reid *et al.* [R18] have used the *Hipparcos* database to compile a sample of 1,024 stars in 764 systems with $M_V < 8.0$ and $r < 25$ pc, complementing the PMSU M dwarf sample. While the brighter sample is statistically complete for *systems*, it is clear that as many as ~ 150 companions may await discovery, and more detailed follow-up observations for many stars are required. In addition, work is currently under way to extend the M dwarf survey to 20 parsecs, using 2MASS data in conjunction with proper motion observations [R20]. We describe this project further in the following section.

The 8-parsec sample

It is useful to define a nearby-star sample where all stars, regardless of luminosity, have a common distance limit. This data set then provides a reference sample that minimises many distance-dependent selection effects, such as resolving secondary components in multiple systems. Most studies have chosen a distance limit of 5 parsecs, but that gives a total of only ~ 50 stellar systems and very sparse statistics. We favour 8 parsecs, with the additional constraint that stars must lie north of $\delta = -30^\circ$; this is the northern 8-parsec sample described in Section 2.7 and listed in the Appendix. There may be some minor residual incompleteness in the sample, but the indications are that any missing stars are limited to spectral types later than M6. A decade of searching since the original definition has turned up a bare handful of additions, while almost as many of the original stars have been removed through improved parallax measurements [R19].

⁴ Note that the statistic is differential: a uniformly incomplete sample has $\langle V/V_{max} \rangle = 0.5$.

The luminosity function

Figure 8.8 compares the luminosity functions derived by [W6], by [R11], and from the 8-parsec sample listed in the Appendix. We also show the 5 parsec luminosity function. In each case the star-by-star luminosity function is plotted, counting each component separately. There is reasonable agreement among these three determinations, as expected, since there is substantial overlap between the underlying samples. Each shows a maximum in number density near $M_V \sim 12$. The peak is strongest in

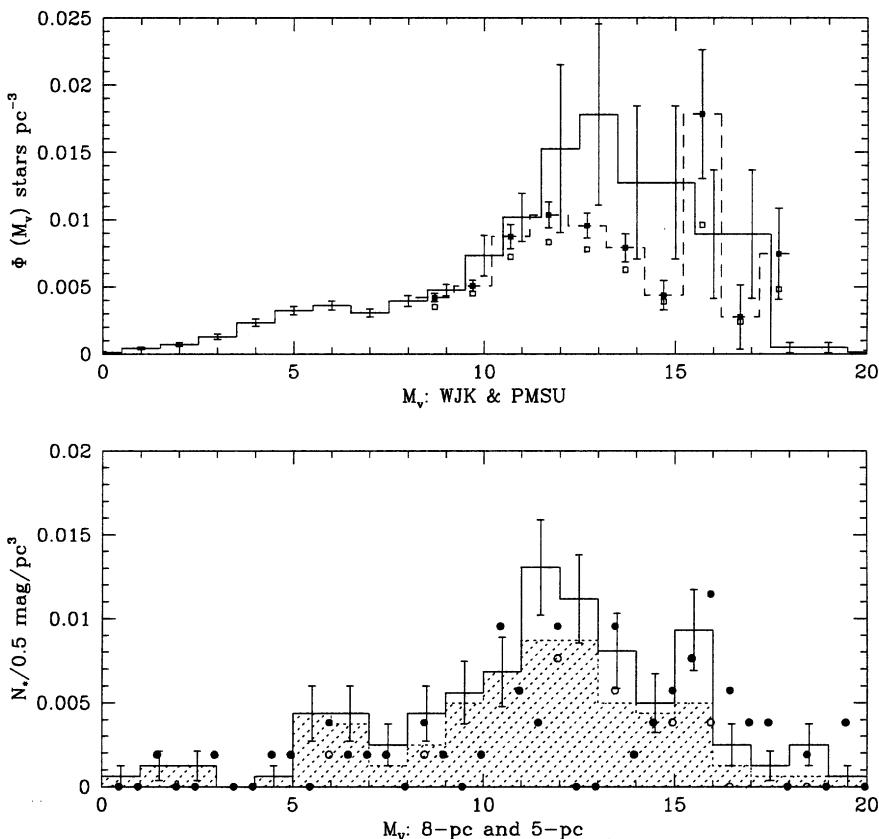


Figure 8.8. The V -band luminosity function, derived from a nearby star. The solid histogram in the upper diagram plots results from analysis of the CNS2 and its supplement [W6]; the solid points and dashed histogram shows the luminosity derived from the CNS3/PMSU data set [R11], with the open squares marking the contribution from single stars and primaries in multiple systems. The lower panel compares census data from the 8-parsec and 5-parsec samples: the 8-parsec data are shown as the solid histogram, with the contribution from single stars and primaries shaded; the 5-parsec data are plotted as solid points (all stars) and open circles (single stars and primaries). The error bars represent Poisson sampling statistics.

the Wielen *et al.* analysis; their sample includes a number of M dwarfs since removed to larger distance (or different classifications) by [R11]. The 5-parsec function has a strong spike in number density at $M_V = +16$, but a deficit at $M_V = +11$. The average star density is ~ 0.09 stars parsec $^{-3}$, corresponding to a median nearest neighbour distance of ~ 1.4 parsecs.

8.5.2 Photometric surveys

Surveys of the nearest stars offer the advantage that the objects are, by definition, the brightest (in apparent magnitude) of their respective classes. The main disadvantage is that the surface density is low (~ 1 system/150 square degrees for $r < 10$ pc), so analyses often have to rely on compiling a census by combining results from different individual surveys that use a variety of different techniques. This can lead to an inhomogeneous sample with the possibility of uneven selection effects.

Photometric surveys offer an alternative approach: distances to stars (and brown dwarfs) are estimated using appropriate magnitudes and colours, and homogeneous criteria can be applied to define a distance-limited sample. Selection by colour avoids the kinematic selection effects inherent in proper motion surveys, besides providing more accurate and objective absolute magnitude estimators than spectral type (from objective prism surveys). The main disadvantage of this method is that the changing slope of the main sequence leads to variation in the accuracy of the photometric distances at different spectral types. This can complicate analysis, as described further below.

Pencil-beam photometric surveys

Photometric parallax techniques were first applied to the study of the luminosity function using 48-inch Schmidt plate material. Although Weistrop's Palomar investigation suffered from systematic errors, her work revealed the potential of wide field imaging as a means of compiling a distance-limited sample of field M dwarfs. $(B-V)$ is not the colour of choice for M dwarf photometric parallaxes, given both the steep gradient $dM_V/d(B-V)$ and the faint B -band luminosities of late-type dwarfs. M8 dwarfs like VB10 are barely detectable beyond 10 parsecs on the POSS I O -band plates. Moreover, hand-measurement of positions and magnitudes for tens of thousands of objects presents a daunting task even for enthusiastic Ph.D. students.

These problems were ameliorated substantially by the development of both automated plate-measuring machines and the near-infrared sensitive IVN photographic emulsion in the late 1970s. Combined with suitable filters [B1], IVN plates provide wide-field I -band data, permitting surveys based on $(V-I)$ or $(R-I)$ colours. Schmidt plates, covering over 30 square degrees, can be scanned and digitised in 8–10 hours by machines such as COSMOS [S14] and APM [K5]. Those scans give positions accurate to ~ 1 arcsec, photometry to ± 0.1 magnitudes, and morphological information, allowing separation of stars and galaxies for 20,000–100,000 objects per plate.

Comparing the predicted absolute magnitude, $M_V(\text{phot})$, against the observed apparent magnitude, V , gives the photometric parallax, π_{pp} , where

$$\pi_{pp} = 10^{0.2(M_V(\text{phot}) - V - 5)} \quad (8.32)$$

The first low-mass star surveys were based on V (IIaD emulsion) and I -band plates [R10], [G1], [S15], while later studies combined R (IIIaF) and I -band material [H7], [L4], [T3]. The latter gain in depth ($R \sim 20.5$ against $V \sim 20$) at the expense of higher uncertainties in the absolute magnitude calibration ($\sigma_{V-I} \sim 0.34$ magnitudes in M_I against $\sigma_{R-I} \sim 0.44$ magnitudes). In each case, the analysis involves combining data from several plates in each passband (to eliminate photographic defects); determining mean colours for each object; selecting candidate M dwarfs and estimating absolute magnitudes and distances; setting an appropriate distance limit (typically 100 pc); and deriving space densities, with due allowance for the disk density distribution, $\rho(z)$. Most surveys cover only a single ~ 30 square degree field, although Tinney's [T3] dataset spans nine fields and ~ 270 square degrees.

Photographic analyses have been supplemented by three recent studies based on wide-field CCD imaging: a CCD-transit survey by Kirkpatrick *et al.* [K7], covering a strip of ~ 27.3 square degrees to a limiting magnitude of $R = 19.0$ magnitudes⁵; analysis of HST WFPC data by Gould *et al.* [G8], [G9], including 53 fields encompassing 0.28 square degrees to $I \sim 27$ magnitudes; and Martini and Osmer's [M5] $UBVRI_{75}I_{85}$ survey of 0.83 square degrees to $V = 22$, a by-product of a multicolour search for high-redshift QSOs. Those data sets provide more accurate photometry for individual objects, although the intrinsic dispersion in the calibrating photometric parallax relationship dominates the uncertainties in both photographic and CCD surveys. Both the deep WFPC observations and Martini and Osmer's analysis include M dwarfs at heights of 500–2,000 pc above the Plane, requiring simultaneous solution for $\Phi(M_V)$ and $\rho(z)$. By and large, the resulting luminosity functions are in good agreement (Figure 8.9).

Biases: binaries and absolute magnitude calibrations

A comparison between the luminosity function derived from the nearest stars and results from (mainly) photographic photometric parallax analyses shows good agreement at brighter magnitudes, but also some discrepancies fainter than $M_V \sim 12$. Specifically, the nearby star data sets (particularly the 5-parsec sample) predict almost a factor of three more stars in the $13 < M_V < 17$ range than the photometric surveys. Many of the nearby stars at these magnitudes are companions, suggesting that these stars might be systematically overlooked in the photometric

⁵ The [K7] scans cover 19.4 square degrees in the northern Galactic hemisphere, and 7.9 in the south, but the eight reddest M dwarfs ($M_I > 13$) lie in the latter portion of the survey. Since the Sun lies above (north of) the mid-point of the Galactic Plane, Kirkpatrick *et al.* suggested that this imbalance might indicate that the latest-type M dwarfs are tightly confined to the Plane, as one would expect for young, substellar objects. As discussed in more detail in the following chapter, this hypothesis is not supported by subsequent studies.

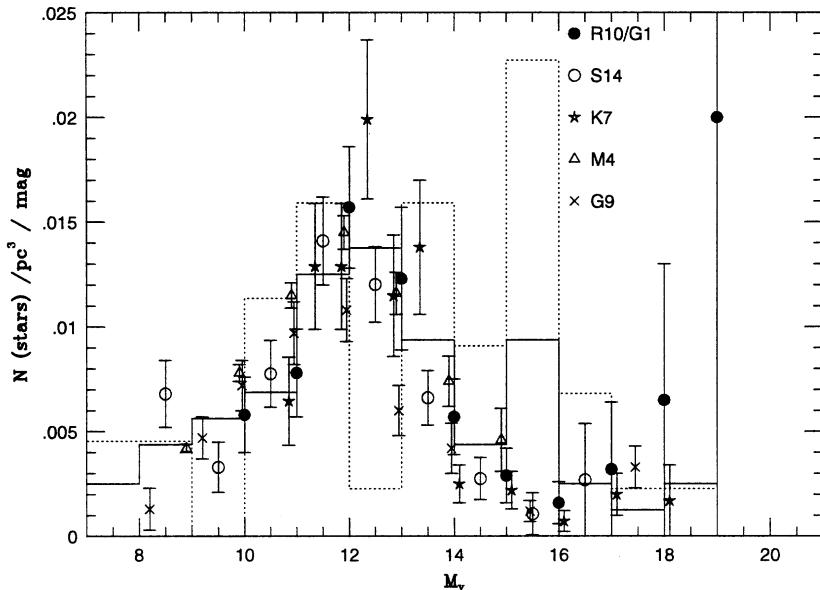


Figure 8.9. V -band luminosity functions derived from the photometric parallax surveys identified in the figure legend. The solid histogram plots $\Phi(M_V)$ from the 8-parsec sample, and the dotted line marks the 5-parsec luminosity function.

surveys; the stars in those surveys are typically at distances of 50–100 pc (much further for Gould’s HST survey), so close companions ($\Delta \sim 150$ AU) would be unresolved on the Schmidt plates. However, there are counter arguments: first, most of the nearby, low-luminosity companions are actually wide systems, like VB10, that would be easily resolved on the Schmidt plates; second, HST follow-up imaging of Schmidt-selected M dwarfs, at 10–20 times higher spatial resolution, did not reveal any new low-luminosity companions, although several near-equal magnitude systems were identified [R8]; and, third, binarism is a double-edged sword that both adds and subtracts from a photometric survey.

Expanding on the last point, consider a stellar sample selected to have photometric parallax $\pi_{pp} > \pi_{lim}$. An unresolved binary system has a higher (combined) luminosity than a single star with the same colour, but both binary and single star are assigned the same absolute magnitude in a photometric parallax calibration. As a result, the distance to the binary system is underestimated, with the consequence that the effective sampling volume for binaries is larger than for single stars. This effect is highest for an equal magnitude system, which is twice as bright as a single star, leading to an increase of $2 \times 2^{1/2}$ in the sampling volume. This more than compensates for unresolved ‘missing’ companions [R5].

In fact, the explanation for the discrepancy between the nearby star and photometric luminosity functions is more mundane, and serves as a warning of the potential effect of unrecognised systematic errors. Figure 8.10 compares the

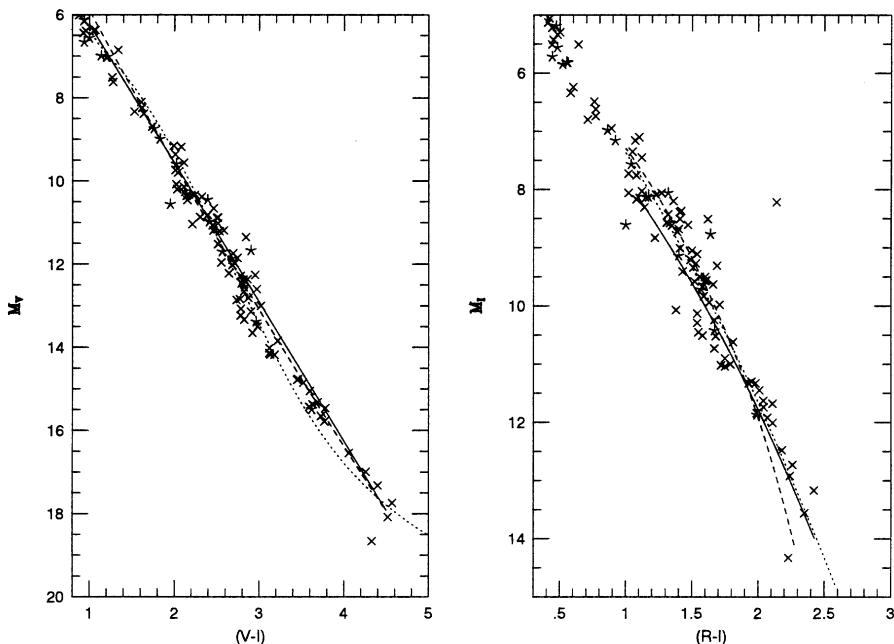


Figure 8.10. The $(M_V, (V-I))$ and $(M_I, (R-I))$ colour–magnitude diagrams for individual stars in the 8-parsec sample compared with the mean colour–magnitude relations used to infer photometric parallaxes. The three $(V-I)$ relations plotted are from [R10] (dotted line), [S15] (solid line) and [K7] (dashed line); the $(R-I)$ calibrations are from [H4] (solid line), [L4] (dotted line) and [T3] (dashed line).

photometric calibrations used in published photometric analyses against colour–magnitude data. It is clear that none are a perfect match to the observations; in particular, the relations take little account of the ‘step’ at $(V-I) \sim 2.9$ and $(R-I) \sim 1.6$ (see Section 3.5). Many simply adopt a linear relation that slices through the feature. This may seem a small discrepancy, but the net result is that stars with $3 < (V-I) < 4$ are assigned absolute magnitudes that are too bright, and distances that are too large; this leads to an underestimate of $\Phi(M_V)$ for $13 < M_V < 15$. In addition, the enhanced Malmquist bias produced by the increase in σ_m near $(V-I) = 2.9$ inflates the inferred number densities at $M_V = 12$. As Figure 8.11 shows, these two effects combine to give the discrepancy evident in Figure 8.9.

This result does not invalidate the photometric parallax technique; it simply means that one has to be careful in its application to avoid systematic mismatches to the data. Recent investigations (e.g., [R20]) use colour–magnitude relations (or spectral type calibrations) that take explicit account of the feature at $(V-I) \sim 2.9$. In general, we should emphasise that photometric parallax surveys only identify *candidate* nearby stars and brown dwarfs; the cosmic dispersion about any colour–magnitude relation guarantees that objects beyond the formal distance limit are included. Statistical corrections can be applied to the derived

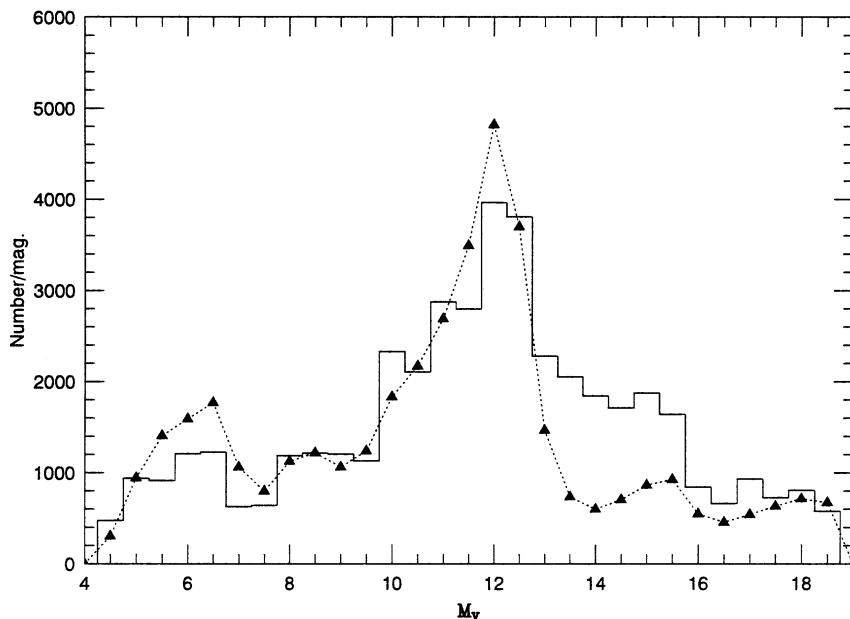


Figure 8.11. Systematic biases introduced in $\Phi(M_V)$ by using a linear ($M_V, (V-I)$) relationship. The histogram outlines the ‘true’ luminosity function in a model simulation; the triangles, the results of deriving distances and luminosities using a linear ($M_V, (V-I)$) relationship.

luminosity function, but more detailed follow-up observations (particularly trigonometric parallax measurements) are necessary to set any photometric sample on a reliable footing.

8.6 MODERN TIMES II: THE DIGITAL ERA

The first years of the 21st century have seen a sea change in photometric parallax surveys. Rather than rely on data sets compiled painstakingly from scans of individual Schmidt plates, or a patchwork of CCD images, current investigations can use large-scale photometric surveys such as DENIS, 2MASS and SDSS, which provide accurate, homogeneous data sets for substantial fractions of the celestial sphere. In the particular case of luminosity function studies, this means that we can effectively combine the nearby star and photometric approaches, using photometric criteria to identify candidate stars and brown dwarfs within the immediate Solar Neighbourhood ($d < 20$ to 30 pc), filling in gaps in the current census.

As described in earlier chapters, these new large-scale surveys have already had considerable impact on low-mass star investigations, particularly on brown dwarf research and the definition of spectral classes L and T (Chapter 6). To date, the main programs aimed directly at probing the luminosity function have been based on

2MASS data. These are the NStars programs summarised in [R20] and [C3], which have the ultimate aim of identifying all M and L dwarfs within 20 parsecs of the Sun. At the time of writing, the M dwarf program is still underway, but an ultracool M and L dwarf census has been completed.

8.6.1 M dwarfs

2MASS, by itself, is not well suited to identifying nearby M dwarfs. As Figure 2.17 shows, $(J-K)$ provides highly uncertain distance estimates for spectral types K5 to M7; supplementary optical photometry is required for those stars. In principle, such data are available for the whole sky from scans of the POSS II and UKST photographic surveys, while SDSS provides deeper, more accurate photometry for π steradians at high galactic latitude; combined with 2MASS, optical/infrared colours such as $(R-J)$ or $(I-J)$ would prove extremely effective in selecting nearby M (and L) dwarfs. At present, however, fully cross-referenced optical/IR data sets are not available, so the NStars M dwarf census has relied on other strategies to identify nearby stars, notably proper motion measurements, either from existing catalogues or from the 2MASS database itself.

The principal source of targets to date has been Luyten's NLTT catalogue, which includes $\sim 58,000$ stars with motions $\mu > 0.18 \text{ arcsec yr}^{-1}$. Those stars have received relatively little attention heretofore, partly because the low accuracy of NLTT photometry (m_{pg} and m_r , blue and red magnitudes, respectively) makes it difficult to segregate the most interesting objects and partly because, unlike the LHS stars, there are no finding charts. However, matched against 2MASS, the $(m_r - K_S)$ colours provide a sufficiently long baseline that the nearby star candidates can be segregated from more distant, background sources (see Figure 8.12). More accurate distances to those candidates are derived through a combination of photometric and spectroscopic parallax estimators [R20], with the eventual goal being measurement of trigonometric parallaxes. So far, over 350 systems have been added to the 20-parsec census from analysis of half of the sky [R21].

The NLTT survey becomes incomplete at relatively bright magnitudes: $m_r \sim 17$ for declinations accessible from Palomar; and $m_r \sim 15$ in the south ($\delta < -42^\circ$), where the main source is the Bruce survey. Several strategies are currently being employed to fill in the missing stars. First, as alluded to previously, the 2MASS database itself can be used to find proper motion stars: 2MASS data are matched against the USNO A catalogue, taken from scans of POSS I (1950–1958) and first epoch UKST plate material (1978–1984). The matching algorithm simply looks for optical counterparts within 5 arcseconds of the 2MASS position, so stars with moderate motions ($\mu \sim 0.1\text{--}0.2 \text{ arcsec yr}^{-1}$) lack optical counterparts. Thus, one can search for bright objects with M dwarf near-infrared colours, but no optical counterpart. Work is currently underway on this project, supplementing the 2MASS NLTT surveys [R21].

The second approach is to repeat Luyten's all-sky proper motion survey, but with digital scans of the photographic plates, rather than using analogue techniques. Lepine, Shara and Rich [L15], [L16] are surveying the northern sky, using digital

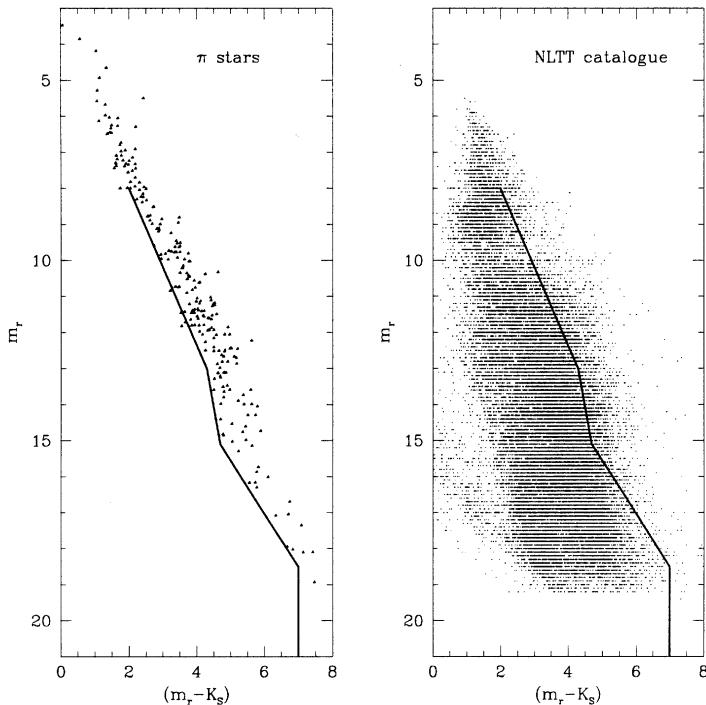


Figure 8.12. Finding nearby M dwarfs in the NLTT: the left-hand panel plots the $(m_r, (m_r - K_s))$ colour-magnitude relation for nearby stars with accurate parallax data, moving all of those stars to 20 parsecs. The solid line marks a lower boundary to that distribution. The right-hand panel plots the $(m_r, (m_r - K_s))$ diagram for $\sim 23,000$ NLTT stars, superimposing the nearby star selection criteria from the left-hand panel. Approximately 8% of the sample meets these criteria, but only ~ 350 stars survive more detailed scrutiny.

subtraction of POSS I and POSS II plate material to reveal objects with significant motions. Their survey will eventually extend south of the celestial equator, with a proper motion limit of $\sim 0.1 \text{ arcsec yr}^{-1}$ and a catalogue of over 100,000 stars. At the same time, scans of UKST plates by the SuperCOSMOS scanning machine are being used to survey the southern hemisphere for stars with $\mu > 0.18 \text{ arcsec yr}^{-1}$, matching the NLTT [P10]. Both the Lepine *et al.* and SuperCOSMOS surveys are cross-referenced against 2MASS, so nearby stars can be identified following the 2-step process employed in the NStars NLTT survey: select candidates based on optical/IR colours (see Figure 8.13); and refine distance estimates using more accurate photometric and spectroscopic parallax estimators. Again, the final goal is a trigonometric parallax measurement for each candidate.

Finally, the release of the full SDSS data set will simplify the search process, at least for 25% of the sky. SDSS provides accurate, reliable optical photometry *ab initio*. One can therefore skip step 1 of the two-phase analysis, and move directly to estimating photometric parallaxes, either directly from the SDSS

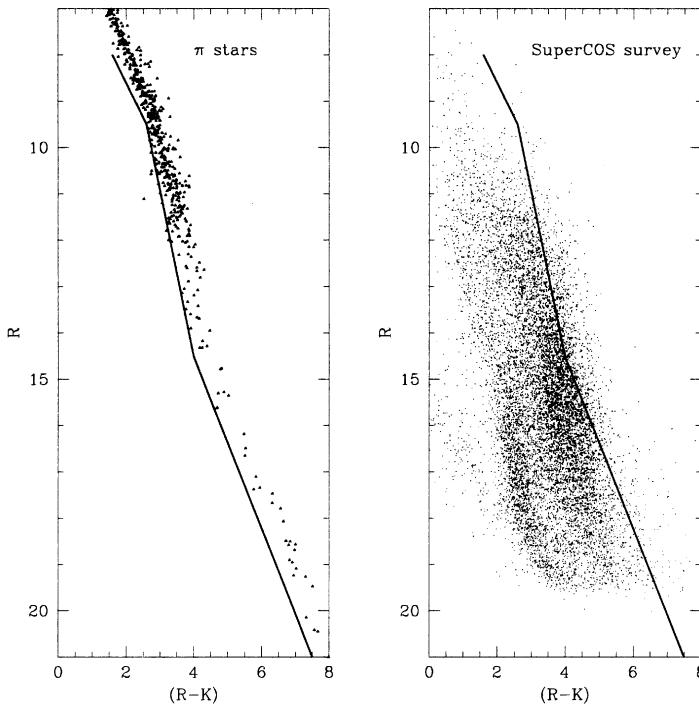


Figure 8.13. Finding nearby M dwarfs in the SuperCOSMOS proper motion survey. As in Figure 8.12, the left-hand panel uses nearby stars with accurate parallax data to define $(R, (R-K))$ criteria designed to select stars within 20 parsecs; the right-hand panel applies those criteria to the full survey. Again, $\sim 10\%$ of the sample are nearby star candidates; follow-up observations are still in progress.

optical photometry, or from optical/IR colours obtained by cross-referencing against 2MASS. Moreover, since SDSS has a much fainter limiting magnitude than the Schmidt surveys, the combined catalogue can be searched for L and T dwarfs (see Chapter 6). Investigations along these lines are being actively pursued at the University of Washington (e.g., [B15]).

8.6.2 Ultracool M and L dwarfs

Paradoxically, completing a census of ultracool dwarfs presents fewer problems than searching for early and mid-type M dwarfs. As Figures 2.17 and 6.14 show, dwarfs later than M7 have distinctive $(J-K)$ colours, and can therefore be identified directly from the 2MASS survey without recourse to other data sets. $(J-K)$, however, is a relatively poor photometric parallax calibrator; the dispersion about the best fit colour-magnitude relation is ~ 0.6 magnitudes, corresponding to distance uncertainties of $\sim 30\%$. Spectral types exhibit a much better correlation with absolute magnitude (see Figure 8.14) provided that those types are derived from the far-red

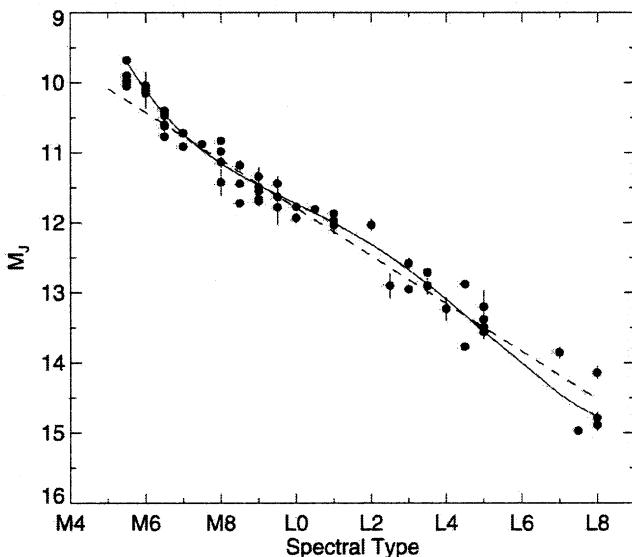


Figure 8.14. The (M_J , spectral type) relation for ultracool dwarfs; the solid line is the best fit relation listed in the text; the dashed line is from [D2] (from [C3], courtesy of the *Astronomical Journal*).

optical spectrum, rather than from infrared bandstrengths. This probably reflects the fact that the near-infrared spectrum is susceptible to the presence of dust in the atmosphere, while the far-red flux is more representative of fundamental properties (see Section 6.7.3).

A survey for ultracool dwarfs in the Solar Neighbourhood is therefore also a two-step process: identify nearby candidates through their location in the (J , ($J-K$)) colours (see Figure 6.14); then use optical spectroscopy to refine the distance estimates. As a rough guide, a search for ultracool dwarfs within 20 parsecs among the 200 million point sources in the 2MASS database yields approximately 3,000 candidates; most are interlopers, including carbon stars, M giants and highly reddened young stars; only ~ 600 are late-type dwarfs, including ~ 250 L dwarfs [C3].

So far, we have expressed the luminosity function primarily in terms of visual magnitudes, $\Phi(M_V)$. This is not really feasible for L dwarfs, which are extremely faint at optical wavelengths, and in many cases lack appropriate photometry. However, thanks to 2MASS, most nearby stars (and brown dwarfs) have accurate near-infrared data, and absolute magnitudes for ultracool M and L dwarfs can be derived from the M_J /spectral type relation plotted in Figure 8.14:

$$\begin{aligned} M_J = & -4.410 + 5.043 SpT - 0.6193 SpT^2 \\ & + 0.03453 SpT^3 - 0.0006892 SpT^4 \end{aligned} \quad (8.33)$$

where $SpT = 5$ at M5, 10 at L0 etc. [C3]. Note that this relationship is not valid for spectral types later than $\approx L7$ (see Figure 6.10). Applying this absolute magnitude

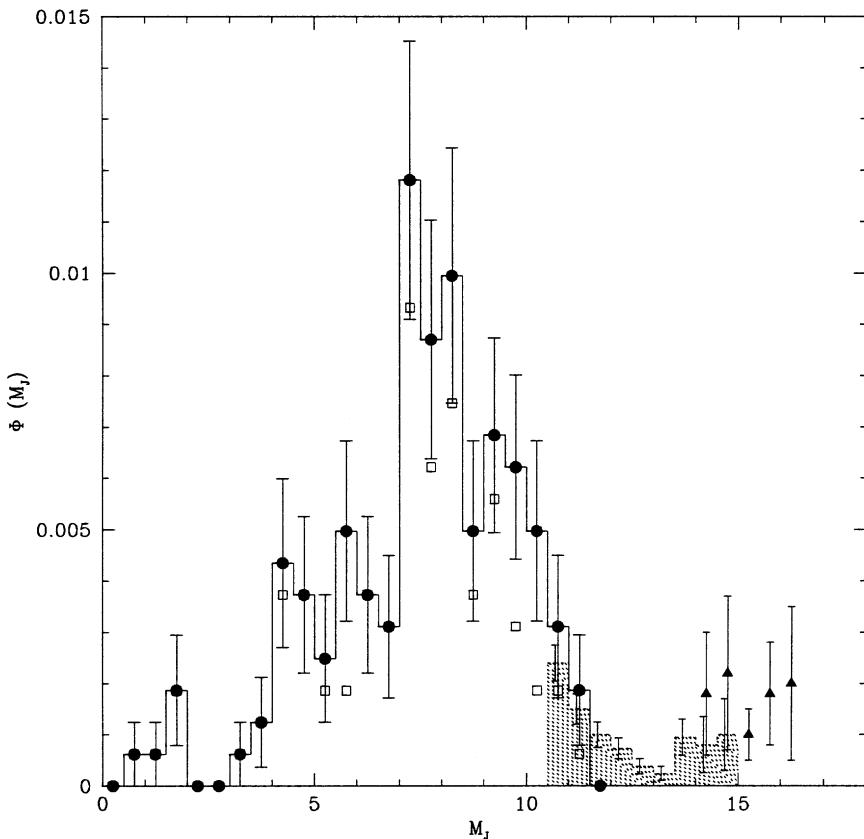


Figure 8.15. $\Phi(M_J)$, the J-band luminosity function: data for the 8-parsec sample are plotted as solid points with error bars; the NStars ultracool M and L statistics are shown as a shaded histogram; and the open triangles mark the likely contribution of T dwarfs.

calibration to the NStars sample, we find that a total of 80 ultracool M and 80 L dwarfs lie within 20 parsecs of the Sun. Figure 8.15 shows the 8-parsec luminosity function transformed to $\Phi(M_J)$, and adds the contribution from the NStars ultracool survey.

There is good continuity between the conventional 8-parsec data and the ultracool statistics in Figure 8.15, suggesting that there is no gross incompleteness. The most striking feature, however, is the small number density of L dwarfs relative to the M dwarfs. In broad terms, we expect to find between 1,600 and 1,700 M dwarfs within 20 parsecs of the Sun; in contrast, there are only 80 L dwarfs, a ratio of at least 20 to 1. There are likely to be no more than 6 L dwarfs in the 8-parsec sample; two are currently known, 2MASS1507-1657 at 7.3 pc and DENIS0255-4700 at 5 pc.

Why are there so few L dwarfs? There are two main reasons: first, brown dwarf lifetimes; and, second, the mass limits for *stellar* L dwarfs. All brown dwarfs become

L dwarfs at some stage in their evolution, but, with no stable energy source, they evolve relatively rapidly through the L dwarf temperature regime. As Figure 6.1 shows, even higher mass brown dwarfs spend only $\sim 2\text{--}3$ Gyr as L dwarfs, while the evolutionary lifetimes are less than 0.2 Gyr for masses below $0.04 M_{\odot}$. The cooling rates decrease at lower temperatures, leading to the upturn in number density at $M_J > 14$ (spectral types later than L5). Stellar L dwarfs are longer lived, but only stars with masses below $\sim 0.08 M_{\odot}$ cool to temperatures below $\sim 2,200$ K, and objects with $M < 0.075 M_{\odot}$ are brown dwarfs. In contrast, M dwarfs have masses between $0.08 M_{\odot}$ and $\sim 0.60 M_{\odot}$. Thus, L dwarfs are fated to be a perpetual minority in the Galactic disk.

8.6.3 T dwarfs

Brown dwarfs evolve through spectral type L to become T dwarfs (Figure 6.1). Cooling rates are even slower at these lower temperatures, and T dwarfs are therefore expected to be much more common than L dwarfs. However, T dwarfs are also extremely faint and difficult to find in the general field, pushing both 2MASS and SDSS to the limits of their technical capabilities (see Section 6.6.2). As a result, current statistics are sparse, and it is not clear that the selection effects inherent in either survey are fully understood.

Nonetheless, attempts are underway to determine the local number density of T dwarfs [B9]. Initial results from these wide-field investigations, although liable to significant uncertainties, indicate that T dwarfs continue the upward trend in number densities evident in the late L dwarfs. We have integrated those data into the J -band luminosity function plotted in Figure 8.15. Note that T0–T5 dwarfs overlap in absolute magnitude with the latest type L dwarfs in M_J (Figure 6.10), so they contribute to the increase in $\Phi(M_J)$ at $M_J > 14$. Current statistics are limited to T dwarfs brighter than $M_J \sim 17$, although we expect the luminosity function to extend to substantially fainter magnitudes, as discussed further in the following section. Within those limits, our best estimate is that T dwarfs are four or five times more common than L dwarfs in the Solar Neighbourhood.

8.7 SPECIALISED SURVEY METHODS

8.7.1 M dwarfs in front of dark clouds

The majority of techniques used to search for field M dwarfs are most effective at moderate and high latitudes, where neither crowding nor interstellar absorption play a strong role. Herbst and Dickman [H9], however, devised a variation on the photometric parallax method that not only works best at low latitudes, but, further contradicting convention, relies on the presence of substantial interstellar absorption along the line of sight. The principle is simple: a relatively nearby, high-density molecular cloud is identified using, for example, millimetre-wavelength CO observations. Accurate, multicolour photometry is obtained for stars lying

within the central high-density region of the cloud. Since the cloud has substantial absorption at visual wavelengths, most of those stars can be expected to lie in front of the cloud, while luminous stars visible through the cloud will have colours that are severely reddened. The cloud serves as a backdrop to provide a catalogue of main-sequence stars within the foreground volume. Absolute space density measurements demand a distance estimate to the cloud, but relative densities (determining the shape of $\Phi(M)$) do not, since the same effective volume is sampled at all absolute magnitudes.

The main limitations of this method are twofold. First, obscuration within a molecular cloud complex is patchy, and observations obtained along sight-lines with moderate reddening run the risk of including background, reddened stars in the catalogue. This limits analysis to dense cloud cores where $A_V > 10$ magnitudes. This restriction leads to the second, and crucial problem: the total solid angle subtended by high-density, high-obscuration clouds is very small. [H9] discuss results for a single cloud in the Scorpius–Centaurus region, covering an area of 220 square arcmin and enclosing only 25 cubic parsecs (against $1,608 \text{ pc}^{-3}$ enclosed by the northern 8-parsec sample). More recently, Jarrett *et al.* [J1] used CCD photometry to probe to fainter magnitudes towards clouds in the Taurus and Ophiuchus star-forming regions ($r \sim 150 \text{ pc}$), covering 2,000 square arcmin and ~ 200 cubic pc. A total of only 22 foreground stars with $9 < M_V < 18$ contribute to the luminosity function derived in the latter study. Thus, while an interesting technique, this is not an effective method of determining $\Phi(M)$.

8.7.2 M dwarf companions to white dwarfs

White dwarf stars serve as targets for two types of low-mass companion searches: wide common proper-motion pairs, or unresolved companions at smaller separations, identified through infrared observations. For all except the first few million years of their existence, white dwarfs have bolometric luminosities comparable with those of late-K and M dwarfs. The scarcity of metals in their high-gravity atmospheres leads to energy distributions close to the appropriate black-body curve. Since the cooling time to reach a temperature of even 5,000 K is several Gyr, most white dwarfs have flux distributions which peak at optical wavelengths. If such a star is in a binary with a low-mass main sequence companion, the white dwarf contributes the majority of the flux in the blue, but the M dwarf dominates at near-infrared and longer wavelengths. Indeed, even a low-luminosity VLM companion ($M_V \sim +18$) produces a substantial infrared excess as compared with that of an isolated white dwarf of the same luminosity (Figure 8.16). White dwarf/red dwarf binaries therefore stand out through their having unusual optical/infrared colours.

By and large, this technique has proven more useful in identifying individual low-luminosity dwarfs than in contributing to our knowledge of the luminosity function. The first major survey based on this method was by Probst [P7], who observed 100 white dwarfs at K, identifying seven with cool companions, including two discoveries. A more extensive study was undertaken by Zuckerman

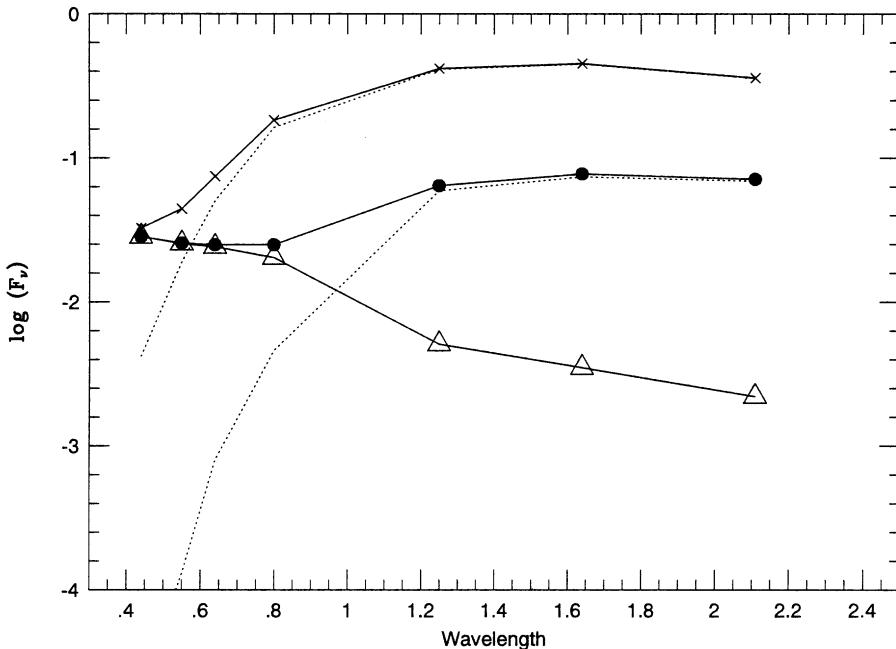


Figure 8.16. The spectral energy distribution of binary systems including a white dwarf and a later-type M dwarf. The white dwarf data (open triangles) are for GJ 1037 ($M_V = 12.9$, $T_{eff} \sim 9,800$ K); the composite distributions include Barnard's Star (crosses) and VB 10 (solid points). The dotted lines show the energy distributions of the M dwarfs alone.

and Becklin [Z1], who obtained aperture photometry or near-infrared imaging of ~ 200 white dwarfs, with the primary aim of identifying very low-mass companions. That aim was realised, since those observations led not only to the discovery of the dusty brown dwarf, G29-38 (Section 6.3.2), but also the identification of the first L dwarf, GD 165B (Section 6.5.1).

Both G29-38 and GD 165B were early discoveries of the [Z1] project; none of the degenerates observed subsequently have L dwarf companions. Twenty stars have M dwarf companions, suggesting a relatively low frequency of high mass-ratio binaries among intermediate mass stars ($8 M_\odot > M > 1.5 M_\odot$). The K-band luminosity function derived from this small sample rises more steeply towards faint magnitudes than field data, but this is probably a selection effect, since early-type M dwarfs are sufficiently bright at optical wavelengths that they affect the overall flux distribution in these binary systems. As a result, early-M/white dwarf binaries have colours that are similar to G or K-type dwarfs. Systems with those colours were not included in the Becklin and Zuckerman survey.

The [Z1] survey is based on bright, nearby white dwarfs, identified in either the Lowell or Luyten proper motion surveys. SDSS, with accurate photometry from u to z , will provide an effective means of searching for more distant ‘mixed’

WD/dM binaries, which are revealed by their inconsistent ultraviolet and far-red colours. Preliminary results are given in [R23] and [S19], while an extensive survey comprising almost 1,000 systems with SDSS spectra is being carried out by Silvestri [S20]. The complication of extragalactic contaminants, mainly quasars, is obviated in the latter survey. However, the difficulty of defining a complete sample makes it unlikely that these surveys will contribute significantly to the definition of $\Phi(M)$.

8.8 M DWARFS IN OPEN CLUSTERS

Open clusters are potent tools for investigating evolutionary phenomena in stars, since cluster members are products of a single star-forming event of known metallicity and of measurable age. In particular, open clusters offer the prospect of testing for variations in $\Phi(M)$ in different environments.

Cluster ages and distances are generally determined by comparing the location of the observed main sequence turn-off against theoretical models (see Figure 7.13). The accuracy of the derived results rests on the accuracy of the models. Recent observations have highlighted the importance of the level of convective overshoot adopted in models of intermediate-mass stars (those near the turn-off): that is, the extent to which a convective element travels beyond the point of neutral buoyancy (see equation (3.49)). This parameter determines the amount of fuel available for hydrogen-burning; the greater the overshoot, the longer-lived the star and the older the age associated with a given turn-off luminosity. Until recently, cluster ages were derived (at least in the USA) assuming little or no convective overshoot. The latter half of the 1990s, however, saw a revision in the age scale; for example, the age of the Pleiades has been revised from 70 Myr to \sim 125 Myr [S17]. The older ages are broadly consistent with those derived from model calculations that include significant convective overshoot (see [M2], and references therein). The cluster ages cited here are linked to the revised scale.

Only a limited number of Galactic clusters are suitable for studies of the luminosity function of lower main sequence stars for the following reasons:

- Detecting stars at or below the hydrogen-burning limit is practical only in the nearest (< 150 pc) clusters – not only because VLM dwarfs are intrinsically very faint, but also because sifting through the foreground and background stars to identify the cluster members is significantly more difficult in more distant clusters.
- Stochastic dynamical effects due to gravitational interactions with other massive objects gradually disrupt open clusters, stripping lower-mass stars preferentially and changing the luminosity function. Mass segregation contributes to this effect at low masses.
- Low-mass stars in open clusters younger than \sim 20 Myr are still evolving onto the main sequence, with luminosities which exceed the main sequence values. Thus, while well-suited to searching for substellar-mass brown dwarfs (Chapter 9), comparison with field-star data is problematic; reconstructing the mass

function is highly dependent on the accuracy of evolutionary models (as discussed in the following chapter).

Given these caveats, the prime targets are the four best-known clusters – the Hyades, the Pleiades, Praesepe and α Per – and these clusters have received the most attention. Before discussing the results, we review techniques used to segregate cluster members from the dominant background of Galactic field-stars.

8.8.1 Cluster membership criteria

Most open clusters lie close to the Galactic Plane, and an average cluster has a total of only a few hundred members. Consequently, the contrast between cluster and field can fall to below 1:1,000 at faint apparent magnitudes. Identifying those few low-luminosity members requires application of a series of astrometric, photometric and kinematic tests, each further refining the initial sample. Like Tamino, each would-be cluster member has to pass all of the tests; rejection accompanies failure at any step.

Proper motions and radial velocities

The velocity dispersion of stars in an open cluster is extremely low; Gunn *et al.* [G10] measure $\sigma \sim 0.23 \text{ km s}^{-1}$ for the Hyades. This is a requirement for the cluster to survive as a discernible entity. The potential energy of a star, mass m_i , in a cluster with n stars, is given by

$$\Omega_i = -Gm_i \sum \frac{m_j}{r_{ij}} \quad (8.34)$$

so the kinetic energy required for that star to escape the cluster is $-\Omega_i$ ([C1], Chapter 5)). The average kinetic energy required to escape the cluster is

$$\bar{E}_\infty = -\frac{2}{n}\Omega = \frac{Gm^2 n}{\bar{R}} \quad (8.35)$$

where \bar{R} is the average radius of the cluster, and m is the average stellar mass. Since $\bar{E}_\infty = \frac{1}{2}m\bar{v}_{\text{esc}}^2$, this reduces to

$$\bar{v}_{\text{esc}}^2 = 4\sigma^2 \quad (8.36)$$

where σ is the rms velocity dispersion in the cluster. Hence, stars with a velocity only twice the observed velocity dispersion are expected to escape from the cluster in a matter of a few Myr. Surviving cluster members occupy a correspondingly reduced velocity range.

The practical consequence of the limited range in space velocities is that kinematic criteria can be used to identify cluster members. Radial velocities present one means of selection, but large-scale observations of faint ($V > 16$) stars at the required precision ($< 1 \text{ km s}^{-1}$) remain impractical, while in some cases (such as the Pleiades) the cluster velocity lies within the field star distribution (see [S13]). Proper motions, readily measurable from wide-field photographic plates, offer a more tractable option as the first step towards distinguishing cluster from field.

Open clusters have typical diameters of 15–20 pc, corresponding to angular diameters of 10° – 30° in the three nearest systems. The proper motion of an individual star is the projection of the space velocity onto the plane perpendicular to the line of sight, so while the cluster stars have identical space velocities, the changing perspective across the cluster leads to different tangential motions, with the individual vectors directed towards a particular point on the sky – the convergent point (CP) [B6].⁶ Individual stellar motions are resolved into two components directed towards ($\mu_{u'}$), and perpendicular to (μ_i), the direction of the CP. The dispersion in the latter component, σ_{μ_i} , provides a measure of the intrinsic velocity dispersion of cluster members. If the CP has equatorial co-ordinates (α_P, δ_P) , then the expected direction of motion, θ (measured north through east) of a star at position (α_i, δ_i) is given by

$$\cos \theta = \frac{(\sin \delta_P - \sin \delta_i \cos \lambda_i)}{\cos \delta_i \sin \lambda_i} \quad (8.37)$$

where λ_i is the angular separation between the star and the CP.

The changing perspective across a cluster also leads to a change in the magnitude of the proper motion. The proper motion of an individual member, $\mu_{u'}$, at distance, r_i , is given by

$$\mu_{u'}^i = \frac{V_S \sin \lambda_i}{\kappa r_i} \quad (8.38)$$

where V_S is the cluster space motion and $\kappa = 4.74$. To allow for this projection effect, the observed motions, $\mu_{u'}$, are adjusted to the cluster centre:

$$\mu_u = \mu_{u'} \frac{\sin \lambda_C}{\sin \lambda_i} \quad (8.39)$$

where λ_C is the angular separation of the cluster centre and the CP, and μ_u the corrected motion.

All cluster members have effectively the same space motion V_S , so the proper motion of an individual member depends inversely on its distance (see Equation 8.38). As a result, the intrinsic depth of the cluster produces a range of proper motions, μ_u^i . This is the inverse to the ‘finger of God’ effect in redshift surveys of galaxy clusters: in galaxy clusters, the velocity dispersion produces an apparent spread in distance along the line of sight; in open clusters, the spread in distance along the line of sight leads to an apparent spread in tangential motions, μ_u^i . Indeed, if V_S is known, r_i can be determined from $\bar{\mu} - \mu_u^i$. The amplitude of this effect decreases with cluster distance: for the Hyades, $0.^{\circ}08 \leq \mu_u^i \leq 0.^{\circ}15$ (Figure 8.17); for the more distant Pleiades, $0.^{\circ}035 \leq \mu_u^i \leq 0.^{\circ}05$. In fact, defining $\Delta\mu_u = \mu_{\max}^i - \mu_{\min}^i$, then $\Delta\mu_u/\bar{\mu}_u$ measures the ratio between cluster diameter and mean distance.

⁶ The intrinsic cluster velocity dispersion means that the proper-motion vectors for all cluster members actually intersect in a small area, rather than a unique point [P2]. This dispersion is measurable only with high-accuracy astrometry such as the *Hipparcos* Hyades data, and can be taken into account.

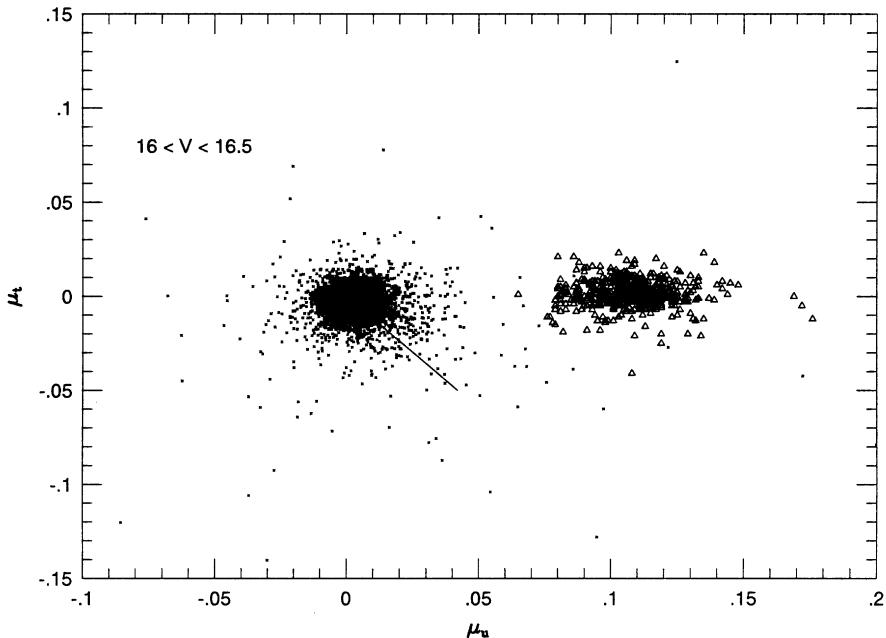


Figure 8.17. The proper motion distribution in (μ_α, μ_t) for stars with $16 < V < 16.5$ in a 50 square degree region near the Hyades. Candidate cluster members ($8 < V < 20$) are plotted as open triangles. The line indicates the proper-motion vector due to reflex solar motion. (Data from [R6].)

Proper-motion selection is effective if the mean cluster motion is distinct from that of the local field-stars. This holds for the Hyades (Figure 8.17) and, to a lesser extent, the Pleiades ($\bar{\mu}_\alpha = 0.^w02 \text{ yr}^{-1}$, $\bar{\mu}_\delta = -0.^w04 \text{ yr}^{-1}$) and Praesepe ($\bar{\mu}_\alpha = 0.^w029 \text{ yr}^{-1}$, $\bar{\mu}_\delta = -0.^w007 \text{ yr}^{-1}$). In the case of the Hyades, an initial sample of $\sim 400,000$ stars with $V < 20$ is reduced to ~ 400 candidate members [R6]. Some field-stars, however, have proper motions consistent with cluster membership, and the smaller the offset between cluster and field, the greater the degree of contamination. Other techniques must be employed to eliminate these interlopers.

Photometric selection

Open cluster members should outline a well-defined main sequence in the H–R diagram. Given accurate multicolour photometry, and distance estimates based on either individual measurements or mean cluster properties, a colour–magnitude diagram can be constructed for the candidate cluster members. Figure 8.18 shows the results of applying this procedure to the proper-motion selected Hyades candidates from Figure 8.17. Approximately two-thirds of the stars lie on a main sequence that is slightly redder than the field-star relationship. Most stars lying above the Hyades sequence (brighter in M_V at a given colour) are binaries in the cluster.

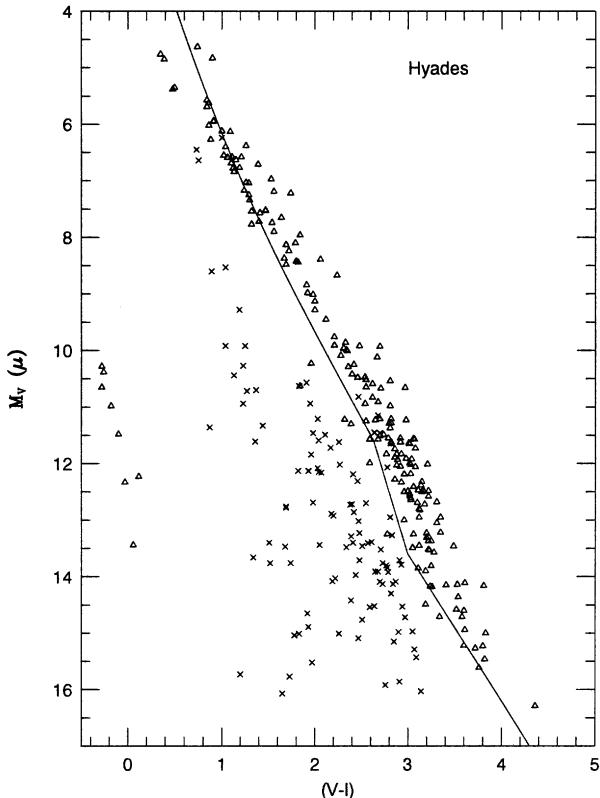


Figure 8.18. The $(M_V, (V-I))$ colour–magnitude diagram for the Hyades proper-motion candidate members. Open triangles identify stars with photometry consistent with cluster membership; crosses indicate background field-stars. The solid line marks the mean colour–magnitude relationship for nearby stars (see Figure 2.16) (data from [R6]).

Over a quarter of the proper-motion candidates, however, lie well below the main sequence. Eggen [E1] originally suggested that these stars might constitute ‘Hyades subdwarfs’ – an hypothesis which we now know to be incompatible with stellar evolution theory. Hanson and Vasilevskis [H6] demonstrated that none of the apparently sub-luminous stars have radial velocities consistent with cluster membership, and they are undoubtedly background stars with velocities which are sufficiently high to produce proper motions matching Hyades members.⁷

Spectroscopic membership criteria

Proper-motion astrometry and photometry are techniques that use integrated starlight, and are therefore efficient methods of examining large samples of faint

⁷ One star with discrepant photometry in Figure 8.18 is a Hyades member: HZ9, a white dwarf/red dwarf binary, $M_V = 10.6$, $(V-I) = 2.0$.

stars. Even after applying these criteria, however, some field-stars remain misclassified as cluster members. Eliminating these stars demands more detailed observations, but the sample size is now sufficiently small that such observations are feasible. Radial velocity provides an additional criterion; measurement of stellar activity, either chromospheric ($H\alpha$ emission) or coronal (X-ray emission), offers another option for late-type stars in relatively young clusters such as Praesepe, the Hyades and the Pleiades [P8], [P9], [R12], [S13]. Stars that survive all of these steps can be confidently accepted as cluster members.

8.8.2 Dynamical effects and the present-day luminosity function in clusters

Once a membership list has been compiled for a cluster of known distance, construction of the luminosity function is straightforward. This represents the *present-day luminosity function* for the cluster, and two time-dependent processes must be taken into account before comparing results with other studies: stellar evolution and dynamical evolution of the cluster. The former is a systematic change in the mass–luminosity relationship with age, which can be taken into account using theoretical models (see Chapter 3). Dynamical evolution, in contrast, is a stochastic effect which increases with time and is less susceptible to analysis.

Large open clusters, such as the Pleiades or the Orion star-forming region, have masses of $\sim 3,000 M_\odot$ at formation. With diameters of ~ 10 parsecs, individual stars are bound only weakly to the cluster. Most clusters have orbits that lie close to the Galactic Plane, bringing them in close proximity to massive giant molecular cloud complexes (GMCs). Stars on the near-side of the cluster receive a greater gravitational impulse from the GMC, and the additional energy may unbind stars from the cluster potential. The net result of this tidal effect is that an open cluster is gradually pulled apart over the course of time as stars ‘evaporate’ from the cluster, leading to the observed scarcity of old open clusters [F3]. A cluster which has an orbit with low inclination to the Galactic Plane is likely to survive for only a few orbits (perhaps several $\times 10^8$ years), while those few clusters in high-inclination orbits can survive for several Gyr.

If stellar evaporation were to occur at the same rate for stars of all masses, then the only effect on the luminosity function would be a change in the zero-point. However, this is not the case. As stars in a cluster relax to match the gravitational potential, we expect equipartition of energy; that is, the kinetic energy, $\frac{1}{2}mv^2$, should be the same for all groups of stars. This implies that lower-mass stars must have a higher velocity dispersion and, with higher σ_V , these stars have a more extended spatial distribution. Mass segregation occurs, with more massive stars and binaries confined to the central regions of the cluster.⁸ Since lower-mass stars spend more time at large distances from the cluster centre, they are more likely to be stripped from the cluster by tidal interactions with massive bodies. Thus, we expect the ratio

⁸ There are also indications that high-mass stars form preferentially near the cluster core; for example, the Trapezium stars in the Orion Nebula Cluster and the BA stars in the Pleiades.

between the number of low-luminosity and high-luminosity stars to decrease with time.

Both Spitzer [S10] and Chandrasekhar [C1] derive analytical expressions for the relaxation timescales for star clusters. Spitzer gives the following relationship:

$$\tau_R = 8.0 \times 10^5 \frac{N^{0.5} R^{1.5}}{\langle m \rangle^{0.5} (\ln N - 0.5)} \text{ yr} \quad (8.40)$$

where N is the total number of stars, R is the cluster radius in parsecs, and $\langle m \rangle$ is the average mass in solar units. This implies relaxation times of $\sim 10^8$ yr for rich open clusters such as the Pleiades.

Spitzer's calculations, however, consider only internal dynamical evolution (that is, gravitational interactions between cluster members) and are more appropriate to $10^5 M_\odot$ globular clusters in the Galactic halo than $10^3 M_\odot$ open clusters within the disk. The latter clusters are more loosely bound gravitationally, and external encounters can have a proportionately more significant effect. More recent calculations rely on N-body simulations, which model both the mean Galactic gravitational potential and short-term effects due to encounters with GMCs. Details of the computational techniques used to model the evolving cluster potential can significantly influence the final results [Z2]. Moreover, the binary fraction adopted is also important, since binaries can affect cluster dynamics by transferring energy to other stars during close encounters. In a binary-single star encounter, the binary becomes more tightly bound while the single star acquires additional orbital energy, and therefore spends more time in the cluster halo where it is more vulnerable to tidal stripping. Available models [T1], [F4], [Z2] indicate typical lifetimes of 2–4 Gyr for clusters with 1,000–2,000 stars, with significant mass segregation occurring in a few hundred Myr – a slightly longer timescale than derived in analytical calculations. Further detailed theoretical work on this complicated problem is required.

Empirically, there is clear evidence for mass segregation in clusters as old as the Hyades ($\tau \sim 625$ Myr). Figure 8.19 (*top*) plots the line-of-sight distance distribution of stars with $M_V < 8.5$ and $8.5 \leq M_V \leq 12$ for Hyades members from Reid [R7]. The higher-mass stars are more concentrated towards the cluster core. The effect is even more pronounced in older clusters, such as the $\sim 3.5 \times 10^9$ -yr-old NGC 2420, lying at $(m-M)_0 = 11.95$ magnitude. The other diagrams in Figure 8.19 plot colour-magnitude and luminosity functions for two fields in NGC 2420, the first near the cluster centre, the other offset by 10 arcmin (~ 7 parsecs). This cluster has a particularly rich binary population, evident in the doubling of the main sequence in the central regions. In the outer field, however, there are few binaries and no stars within two magnitudes of the turn-off. Even in the central field, $\Phi(M_V)$ reaches a maximum at $M_V \sim 6$ – six magnitudes brighter than in the field. Clusters as young as the Pleiades (~ 125 Myr) are less affected by dynamical evolution, but are unlikely to be entirely immune from such effects.

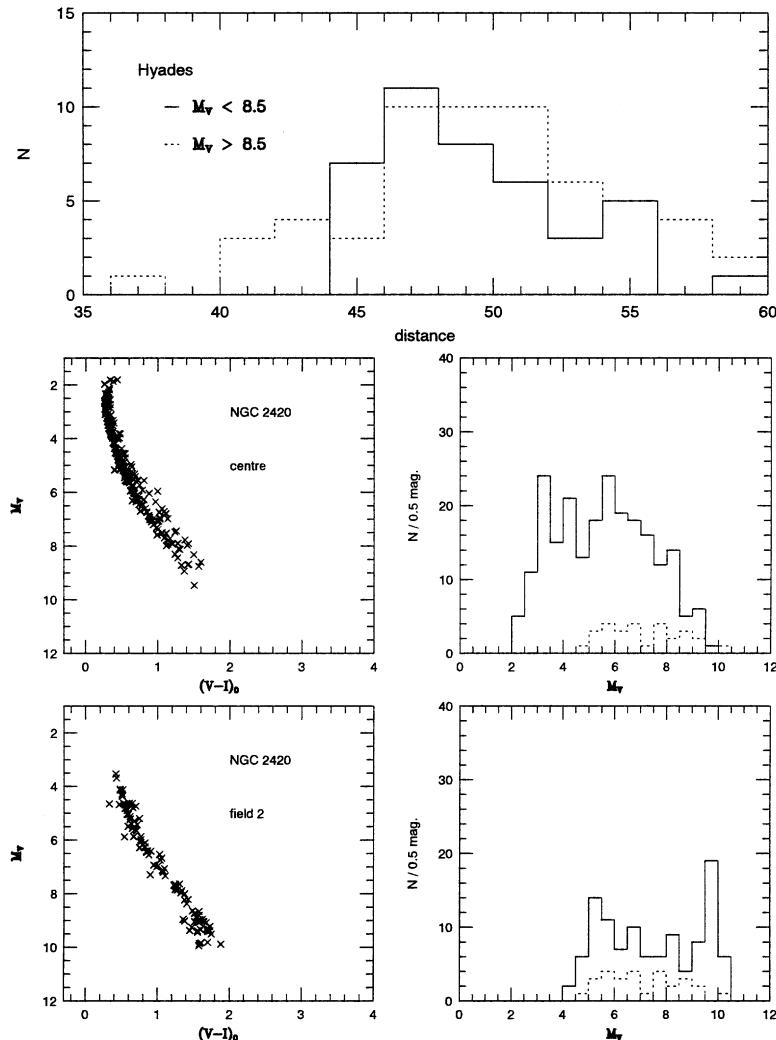


Figure 8.19. Mass segregation in older open clusters. The *top* panel compares the radial density distribution of Hyades stars with $M_V < 7$ and $7 \leq M_V \leq 12$; the *lower* panels compare colour-magnitude diagrams and $\Phi(M_V)$ for two regions in NGC 2420. Note the scarcity of bright stars and binaries in field 2 (the outer field).

8.9 $\Phi(M)$ FOR NEARBY CLUSTERS

8.9.1 The Hyades

The nearest substantial open cluster – the Hyades – was first identified as a coherent moving group by Boss [B6]. With a mean motion of $\bar{\mu} = 0.^{\circ}11 \text{ yr}^{-1}$, the cluster has

been the target of numerous proper-motion surveys [B8], [A1], [H3], [P1], [L14], [R6]. The cluster lies at moderately low latitude toward the Galactic anticentre ($l = 180^\circ$, $b = -22^\circ$).

One hundred and fifty of the brightest cluster members were observed by the *Hipparcos* satellite, and Perryman *et al.* [P2] combine the measured space motion of $V_S = 45.72 \text{ km s}^{-1}$ with astrometric data for 134 stars within the 10-parsec diameter cluster core to derive a distance of $46.34 \pm 0.27 \text{ pc}$ ($(m-M)_0 = 3.33$) to the centre of mass. Matching the colour–magnitude diagram against the appropriate theoretical models (the cluster is mildly metal-rich, $[\text{Fe}/\text{H}] = +0.11$ [B4]), they infer an age of $625 \pm 25 \text{ Myr}$.

The lower main sequence

Due to its proximity, the Hyades cluster subtends a substantial solid angle (>300 square degrees), and most surveys cover only a fraction of the cluster. The exception is Luyten’s Palomar data set [L14], but many of the 929 candidate members generated by that survey still lack accurate photometry and spectroscopy, and residual field star contamination rules out any detailed analysis. The most useful wide-field survey combines scans of POSS I and UK Schmidt $B + V$ plates, covering ~ 112 square degrees (one-third of the cluster) to $V \sim 19.5$ [R6]. Follow-up VRI photometry [R7, L3] and spectroscopy [R12], [S13] of the proper motion candidates have eliminated most, if not all, field star contaminants, giving a clean sample of ~ 400 Hyads. Many cluster members with spectral types later than M3 have high-resolution imaging by HST in a search for close binary systems [R9], while radial velocity measurements have been used to search for spectroscopic systems [R22]. Those observations indicate that the binary fraction is between 23 and 30%, similar to the binary frequency in field M dwarfs.

Figure 8.20(a) compares $\Phi(M_V)$ for the Hyades cluster against the 8-parsec data set, with the two distributions scaled to match at $M_V = +7$. At 625 Myr, the Hyades is significantly younger than the average disk dwarf in the Solar Neighbourhood, and, as a result, the Hyades has proportionately more stars brighter than $M_V = +4$. Both luminosity functions reach their maximum near $M_V = +12$, but the peak is less pronounced in the cluster than in the field: comparing the relative number of stars at $M_V = +12$ and $+6$ in the field, the fainter stars outnumber the brighter by 3 to 1; the ratio is only 3:2 for the Hyades. This is undoubtedly due to dynamical evolution of the cluster population, with the lower mass stars having suffered more attrition through tidal stripping.

Dynamical models predict that stars ejected from a cluster remain in its vicinity for a significant period of time. Those stars contribute to the Hyades moving group [E2]: stars with Hyades-like properties (colour–magnitude relation and chromospheric/coronal activity), but with space motions that differ by $2\text{--}5 \text{ km s}^{-1}$ from the cluster proper. Analysing the population of white dwarfs currently in the cluster, Weidemann *et al.* [W1] suggest that up to two-thirds of the higher mass ($> 1 M_\odot$) stars from the original protocluster have been lost. Since the total mass of the present cluster is $\sim 400 M_\odot$, this implies that the original system probably

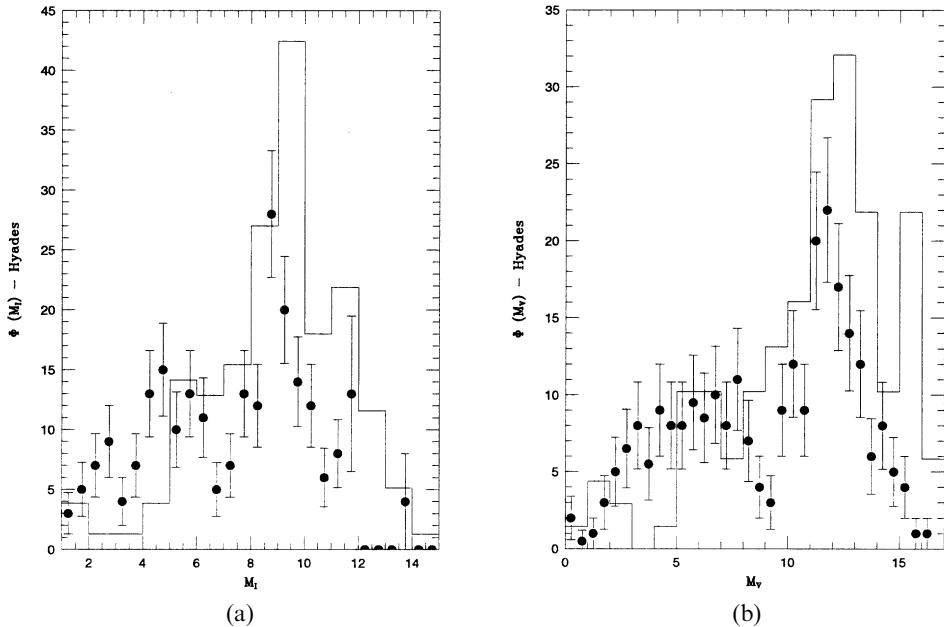


Figure 8.20. (a) $\Phi(M_I)$, including fainter stars from [B7] and [L5]; (b) $\Phi(M_V)$ for the Hyades, from [R6] and [R12] (solid points). Both data sets include low-mass binary companions identified by HST imaging. The histogram plots $\Phi(M)$ for the 8-parsec data, scaled to match at $M_V = +7$ and $M_I = +5$.

exceeded $1,200 M_{\odot}$ in mass, with 3,000–4,000 members, comparable in size to the Orion cluster.

Brown dwarfs in the Hyades

The Hyades luminosity function drops sharply beyond the peak at $M_V \sim 12$. The lowest luminosity cluster members identified in the [R6] survey have $M_V \sim 15.5$. Leggett and Hawkins [L5], Bryja *et al.* [B7] and Dobbie *et al.* [D3] have pushed the search for low-mass members to fainter luminosities in the central regions of the cluster. The Bryja *et al.* analysis combines scans of POSS I E plates and of Luyten's 1962 Palomar plates (also red sensitive) to search for proper motion members fainter than $m_r = 20$. While a dozen candidates were identified, with potential visual absolute magnitudes fainter than 16.5, only two survived spectroscopic follow-up observations: Bryja 804, spectral type M6, $M_V = 15.75$; and Bryja 262, or LP 415-20, spectral type M6.5. Recent AO observations of the latter star have resolved it as a binary, separation 0.09 arcsec and $\delta K = 0.7$ magnitudes. However, none of these objects is substellar.

Leggett and Hawkins [L5] also used red-sensitive plate material to probe to fainter magnitudes, combining scans of multiple R and I -band UK Schmidt

plates. They identified 12 candidates, and one of those stars, LH0418+13, is probably the faintest Hyades cluster member currently known: the star has colours and chromospheric activity consistent with membership, but still requires astrometric confirmation. LH0418+13 is not, however, a brown dwarf. Several of the remaining M dwarfs from the [L5] survey may actually be brown dwarfs, since Li I 6,708 Å is clearly visible in their spectra; however, they lie well beyond the Hyades, and are probably associated with the more distant Taurus star-forming region [R13].

Finally, Dobbie *et al.* [D3] have used CCDs to survey the central 10.5 square degrees to $I \sim 20.3$, corresponding to $R \sim 22.5$. They identified 20 objects with colours and magnitudes consistent with cluster members; however, 19 of the 20 are ruled out based on their proper motions, while the twentieth is RHy 297, one of the faintest members from the [R6] survey.

Thus, despite extensive surveys, the net result is that the only likely brown dwarf member of the Hyades is an unresolved spectroscopic companion. RHy 403, one of the fainter cluster members at spectral type M5, is a single-lined binary with a period of approximately 1.275 days [R22]. Assuming that the primary has a mass of $0.15 M_{\odot}$, the system mass function (see Equation 9.21), combined with the non-detection of the secondary in optical spectra, places the secondary mass between 0.07 and $0.095 M_{\odot}$. Infrared spectroscopy and measurement of the astrometric orbit (see Section 9.3.1) offer the best opportunities to further constrain the secondary mass.

The scarcity of low-mass stars and apparent absence of brown dwarfs are undoubtedly both manifestations of the cluster's dynamical evolution. It is likely that a fraction of the lowest mass members of the original cluster are still present in an extended halo. In principle, those objects could be recovered using the new generation of digital sky surveys. For example, an L0 Hyades brown dwarf would have $J \sim 15$ and $K \sim 13.7$, well above the limiting magnitude of 2MASS. The main complication, as always, is finding those cluster members in an area of 200–300 square degrees, particularly since the Hyades lies at a relatively low galactic latitude, where reddened objects will be a significant contaminant. At present, this remains a project awaiting an intrepid graduate student.

8.9.2 The Pleiades

The Pleiades cluster is younger than the Hyades and approximately three times more distant, with an on-sky diameter of $\sim 20^{\circ}$. The B-type stars on the upper main sequence form a compact and distinctive group, identified by both Greek and Chinese civilisations over 3,000 years ago. The cluster was recognised as a bound system in the late-19th century. The first extensive proper motion surveys were completed by Trumpler [T4] and Hertzsprung [H11]. Like the Hyades, it lies at moderate latitude ($l = 166^{\circ}$, $b = -23^{\circ}$). The cluster has an abundance close to that of the Sun [P6], and has served as a fiducial for stellar evolution theory and investigations of the Galactic and extragalactic distance scale.

Traditional main sequence fitting and convergent point analyses place the cluster at a distance of 132 ± 10 parsecs [L1]. This result was challenged by analysis of

Hipparcos parallax data, which favoured a larger parallax, 8.6 ± 0.24 arcseconds, corresponding to a distance of 118 pc [L2, M6]. This revision poses severe problems for stellar evolutionary theory, since the solar-abundance Pleiades would then be subluminous with respect to the Sun. While the theoretical models can be adjusted to match the shorter distance (by increasing the helium abundance, for example), none of those adjustments appears viable [P6]. Moreover, there is no evidence that any field stars share the anomalous luminosities implied for the Pleiades. Under those circumstances, attention centred on a possible systematic bias at the 1 milliarcsecond level in the *Hipparcos* Pleiades data.⁹ Recently, several independent investigations of the Pleiades distance have been completed. These include trigonometric parallax measurements by HST and orbital analysis of resolved and eclipsing binary stars [Z3, M11]; all arrive at distances that are within 1–2% of the traditional value of 132 parsecs.

Main sequence stars and cluster morphology

Trumpler and Hertzsprung's proper motion surveys identified ~ 500 cluster members. Since those initial studies, the cluster has been the subject of many similar surveys [L1], [S12], [H1], [M7], [M10], [A3]. Most are based on photographic plates and are either limited to relatively bright magnitudes ([L1], $B < 15$), or cover only the central regions of the cluster [S9], [M10]. The exception is the recent survey by Adams *et al.* [A3], which combines 2MASS data with proper motions from scans of POSS I and POSS II plate material to cover the full cluster. The [A3] survey is complete to $R \sim 19$. The [H1] survey extends to $R \sim 20.5$ (close to the hydrogen-burning limit in the cluster, as discussed further below), but covers only the central 25 square degrees. Since the cluster lacks a distinctive mean radial velocity with respect to the local Galactic field, candidate members are verified based on their photometric properties and chromospheric/coronal activity.

The Pleiades is significantly younger than the Hyades, so one expects that the cluster has suffered less dynamical evolution. Observations confirm this expectation. While the bright BA-type stars and higher mass binaries are more centrally concentrated than the cluster G dwarfs [R1], there is evidence for only limited mass segregation [A3]. Nonetheless, there are indications that the luminosity function is somewhat flatter in the core than in the halo, suggesting that lower mass ($< 0.5 M_\odot$) stars may lie preferentially in the outer regions of the cluster. In that case, a higher proportion of those stars may have evaporated from the cluster potential due to dynamical interactions with external objects.

The [A3] analysis of the 2MASS-selected sample gives a core radius of 1.0–1.3 degrees (2.3–3 pc) and a tidal radius between 5.8 and 6.8 degrees (13.4–15.7 pc). There is some evidence for a slightly extended halo of low-mass stars, and the

⁹ Note that there is no suggestion that this bias affects *all* *Hipparcos* data. There are several circumstances that are peculiar to the Pleiades data set, such as reliance on astrometry of very bright stars and fewer epochs of observations, that are probably responsible for the systematic bias.

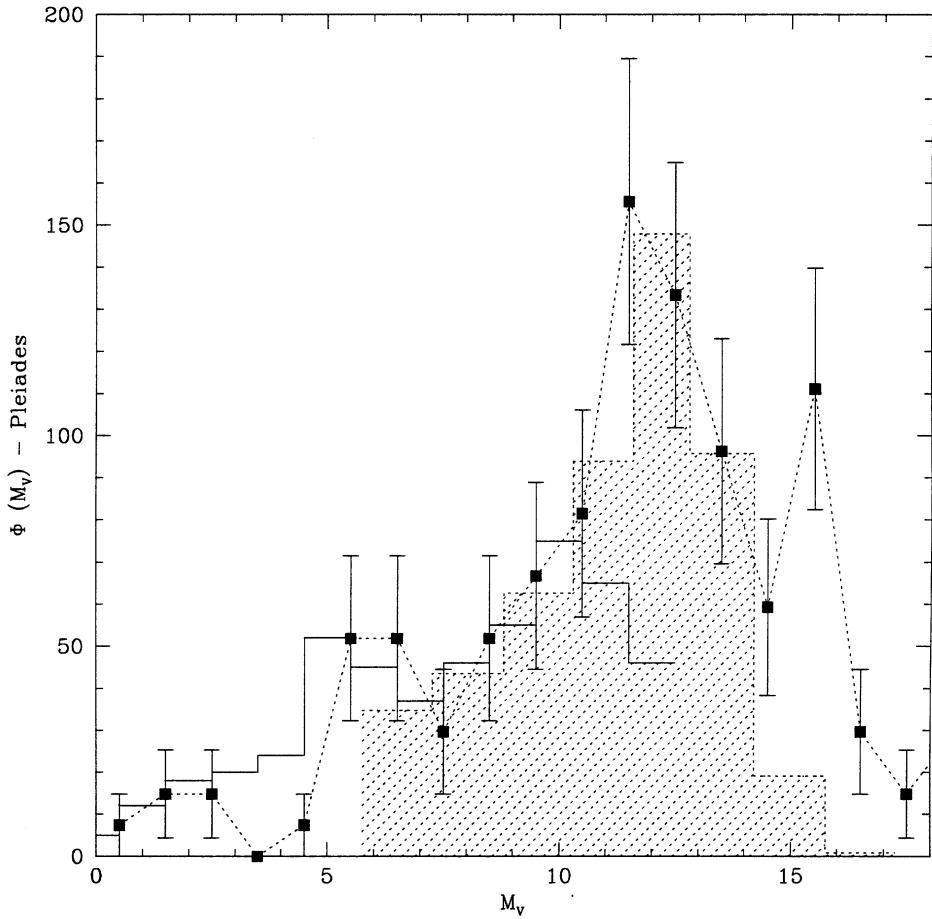


Figure 8.21. The solid histogram plots the Pleiades luminosity function from [S12], while the shaded histogram gives the results derived by the deeper [H1] survey; the latter survey becomes incomplete at $M_V > 14$. The solid squares outline the 8-parsec luminosity function scaled to match the Pleiades data at $8 < M_V < 10$; the error bars show the Poisson uncertainties.

cluster as a whole appears to become more elliptical with increasing radius, probably due to distortion by the Galactic tidal field. The present-day cluster is estimated to have a total mass of $\sim 800 M_\odot$ in stars with masses exceeding $0.1 M_\odot$ [A3].

Given the absence of strong dynamical evolution, one might expect that the Pleiades luminosity function would be closer to that observed in the field. This is borne out by observations. Figure 8.21 combines results from the [S12] and [H1] surveys, the most extensive surveys that have optical data, and compares the composite luminosity function against the 8-parsec data set, scaling the two to match at $M_V = +6$. The [H1] survey becomes incomplete at $M_V > 14$; taking that into account, the field and cluster luminosity functions are in reasonable agreement.

As discussed further in the following chapter, this suggests that the mass functions of the Pleiades and the general field are broadly consistent.

Brown dwarfs in the Pleiades

In contrast to the Hyades, the Pleiades has a substantial brown dwarf population; indeed, the binary PPI 15 was the first brown dwarf identified using the lithium test, as described in Section 6.5.4. Several deep imaging surveys have been completed over the last decade [B10], [Z4], [P11], [D4], covering from 1–6 square degrees near the cluster core to limiting magnitudes of $18.6 < I < 21.8$. Together, these surveys identify a total of ~ 80 candidate brown dwarf cluster members. Jameson *et al.* [J4] have synthesised the results from these surveys, and conclude that approximately half are non-members. The remaining cluster members include at least one L dwarf (Roque 25, [M12]), and the faintest objects are likely to have masses of only 0.035 to $0.04 M_{\odot}$.

The Pleiades brown dwarfs allow us to map out the lithium depletion boundary in the cluster, providing an independent estimate of the age. Current theoretical models (e.g., [B11], [B12]) predict the rate of lithium depletion with time, and hence the evolution of the location of the lithium-depletion boundary as a function of mass. Spectroscopy of Pleiades brown dwarfs places the boundary at $I \sim 17.8$ ($M_I \sim 12.2$) in this cluster (Figure 8.22), corresponding to an age of at least 125 Myr [S17], consistent with stellar evolution models that include a modest amount of convective core overshoot in upper main sequence stars.

8.9.3 Praesepe

Praesepe is, figuratively, the middle child among the three nearest clusters. Similar in age (650 Myr) and abundance ($[Fe/H] \sim +0.12$) to the Hyades cluster, Praesepe has received less attention than either that cluster or the younger Pleiades, probably because it lies at a distance of ~ 180 parsecs. It is at higher latitude than either the Hyades or Pleiades ($l = 206^{\circ}, b = +32^{\circ}$), and has a distinctive radial velocity ($V_r = 36 \text{ km s}^{-1}$). Initially described by Ptolemy, the first extensive proper motion surveys were completed by van Rhijn [R14] and Klein–Wassink [K9], with the latter cataloguing 160 likely members brighter than 14th magnitude. There are more recent surveys by Jones and Stauffer [J2], covering the central ~ 16 square degrees in V and I to $V \sim 17.5$; by Hambly *et al.* [H2], who use UK Schmidt R and I -band plates to reach $R \sim 20$ over the central 19 square degrees; by Pinfield *et al.* [P12], [P13], who use CCD imaging to reach $I \sim 21.5$ over the central square degree of the cluster; and by Adams *et al.* [A4], who, as in their Pleiades survey, combine 2MASS data with the USNO-A catalogue to cover over 100 square degrees to $R \sim 19$.

Like the Hyades, Praesepe shows evidence for significant dynamical evolution. Mass segregation is clearly present [R2], and the luminosity function drops sharply at faint magnitudes. The core radius is 1.1° and the tidal radius $\sim 5.3^{\circ}$, corresponding to 3.5 and 16 parsecs, respectively. The total mass of the cluster at the present time is estimated as between 400 and 600 M_{\odot} , comparable to the Hyades, and, like that

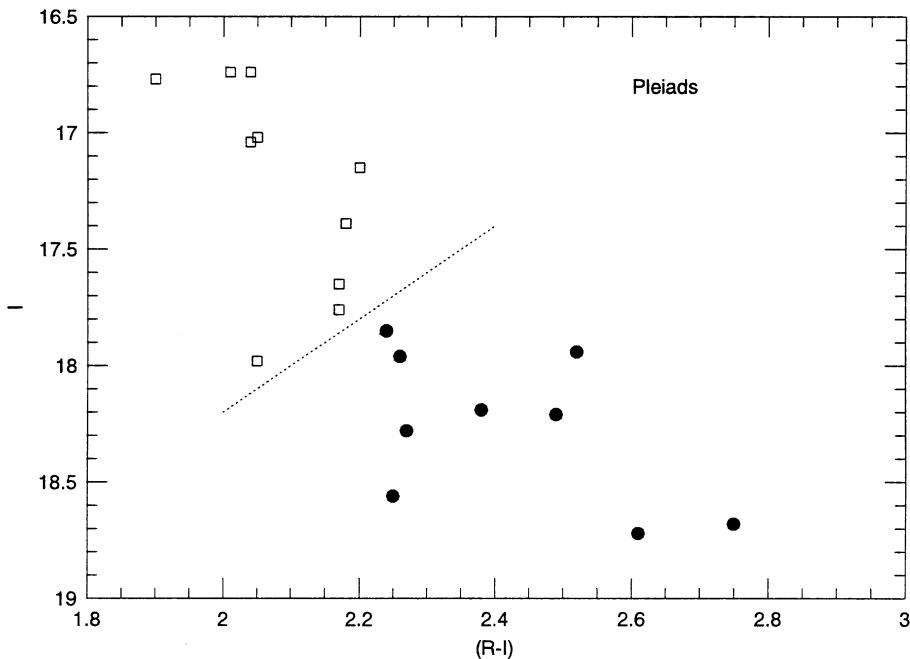


Figure 8.22. The lithium depletion boundary in the Pleiades. Solid points mark dwarfs with detected lithium absorption; open squares are cluster members with no measurable lithium. (Data from [S17].)

cluster, the progenitor cluster was two to three times more massive. Holland *et al.* [H12] suggested that there was evidence for a secondary density concentration, perhaps due to a collision with another, much smaller cluster; however, no such feature is evident in the more extensive 2MASS survey [A4]. Even the deepest surveys (e.g., [P13]) barely reach the hydrogen-burning limit and, at present, there are no known substellar members of Praesepe.

8.9.4 α Persei

α Persei is the youngest of the nearby clusters, with an age of ~ 90 Myr. It lies at a distance of ~ 183 parsecs, similar to Praesepe, but much closer to the Galactic Plane ($l = 147^\circ, b = -7^\circ$). As a result, there is moderate foreground reddening ($A_V \sim 0.3$ magnitudes), and field star contamination is a particular issue, although the cluster sequence lies well redward of the field at later spectral types. As with the Pleiades, the cluster radial velocity is similar to that of the field, so membership is generally confirmed based on the colours and the level of chromospheric or coronal activity.

The first proper motion survey for cluster members was completed by Heckman *et al.* [H13], reaching a limiting magnitude of $V \sim 12$. Prosser [P14] extended the cluster catalogue to 18th magnitude stars in the 1980s and 1990s using photographic

plates, identifying cluster members using both proper motions and colour/activity selection at fainter magnitudes. More recently, Stauffer *et al.* [S18] have used CCD R and I -band imaging to reach $I \sim 20$ over 6 square degrees; the deepest survey to date has been completed by Barrado y Navascués *et al.* [B13], who obtained R and I -band data to $I \sim 22$ within a 3 square-degree region near the cluster core. They find that the luminosity function peaks at $M_I \sim 10$, dropping sharply at fainter magnitudes, broadly consistent with the Pleiades function. There is relatively little evidence for significant mass segregation in this cluster.

Both CCD-based surveys of α Persei reach sufficiently low luminosities that they include cluster members with partially-depleted or undepleted lithium abundances. The transition occurs at $M_I \sim 11.5$, spectral type M6, corresponding to a mass of $\sim 0.085 M_\odot$ [S18]. Theoretical models indicate that the hydrogen-burning limit ($0.075 M_\odot$) lies at $M_I \sim 12.0$, or $I \sim 18.2$, thus the deepest photometric surveys extend well into the brown dwarf regime. We consider the implications for the mass function in Section 9.6.2.

8.9.5 Summary

The results outlined in this section highlight the potential pitfalls of using open clusters as probes of $\Phi(M_V)$. Both the Hyades and Praesepe show that, even at the relatively young age of 600 Myr, dynamical evolution modifies the stellar distribution to the extent that it is not possible to recover the initial properties. On the other hand, rich clusters with ages less than ~ 150 Myr, such as the Pleiades and α Persei, are close to intact in dynamical terms, at least at sub-solar masses, and those clusters are more suitable for statistical analysis.

To date, the Pleiades is the only moderately young cluster that has received sufficient observational attention for the derivation of reliable luminosity (and mass) functions that extend well below the substellar limit. Besides α Persei, several other younger clusters are currently under study, notably NGC 2516 ($\tau \sim 100$ Myr), NGC 2457, IC 2602 and IC 2391 (all $\tau \sim 50$ –60 Myr). The lithium depletion boundary has been detected at $M_I \sim 10.2$ ($\sim 0.12 M_\odot$) in both NGC 2457 [O4] and IC 2391 [B14], but observations are restricted to a fraction of the central regions, even in those clusters. Initial analyses of IC 2602 [F2] and IC 2391 [R17] suggested that both of those clusters have surprisingly few mid-type M dwarfs, perhaps indicating a flatter mass function than in the Pleiades; more detailed investigations of both clusters are currently under way. With the growing availability of large format CCD arrays, full scale surveys of these, and other, youthful clusters, extending well into the substellar mass regime, should be completed in the next five years.

8.10 THE LUMINOSITY FUNCTION OF THE GALACTIC DISK

This chapter has concentrated on the methods used to study the luminosity function, $\Phi(M)$, and has summarised recent results from studies of field stars and members of

Table 8.3. Visual luminosity function.

M_V	Φ (stars $M_V^{-1} \text{ pc}^{-3}$)	M_V	Φ (stars $M_V^{-1} \text{ pc}^{-3}$)	M_V	Φ (stars $M_V^{-1} \text{ pc}^{-3}$)
-0.5	4.6×10^{-5}	6.5	2.8×10^{-3}	13.5	7.3×10^{-3}
0.5	7.6×10^{-5}	7.5	2.2×10^{-3}	14.5	4.5×10^{-3}
1.5	2.3×10^{-4}	8.5	3.2×10^{-3}	15.5	5.0×10^{-3}
2.5	5.5×10^{-4}	9.5	5.0×10^{-3}	16.5	2.5×10^{-3}
3.5	1.2×10^{-3}	10.5	7.4×10^{-3}	17.5	2.0×10^{-3}
4.5	2.3×10^{-3}	11.5	7.7×10^{-5}	18.5	2.5×10^{-3}
5.5	2.2×10^{-3}	12.5	8.2×10^{-3}	19.5	6.0×10^{-4}

Table 8.4. J-band luminosity function.

M_J	Φ (stars $0.5 M_J^{-1} \text{ pc}^{-3}$)	M_J	Φ (stars $0.5 M_J^{-1} \text{ pc}^{-3}$)	M_J	Φ (stars $0.5 M_J^{-1} \text{ pc}^{-3}$)
0.25	6.2×10^{-5}	5.25	2.5×10^{-3}	10.25	5.0×10^{-3}
0.75	3.0×10^{-4}	5.75	4.5×10^{-3}	10.75	3.0×10^{-3}
1.25	4.0×10^{-4}	6.25	5.0×10^{-3}	11.25	2.0×10^{-3}
1.75	5.0×10^{-4}	6.75	5.2×10^{-3}	11.75	1.2×10^{-3}
2.25	7.0×10^{-4}	7.25	1.1×10^{-2}	12.25	7.0×10^{-4}
2.75	8.5×10^{-4}	7.75	9.0×10^{-3}	12.75	4.0×10^{-4}
3.25	1.5×10^{-3}	8.25	9.5×10^{-3}	13.25	3.0×10^{-4}
3.75	1.9×10^{-3}	8.75	6.1×10^{-3}	13.75	9.0×10^{-4}
4.25	2.6×10^{-3}	9.25	7.0×10^{-3}	14.25	1.7×10^{-3}
4.75	3.1×10^{-3}	9.75	6.2×10^{-3}	14.75	2.2×10^{-3}

open clusters. The two main techniques currently employed in surveying the field are, first, a census of stars in the immediate Solar Neighbourhood, and, second, deep pencil-beam photometric surveys, using colours to deduce luminosities, and hence distances and space densities. These two approaches give results that are in good agreement, once the appropriate colour–magnitude relations are adopted for the photometric parallax analyses. In particular, the ‘Wielen dip’ at $M_V = +7$ has been confirmed in star-count data extending more than 1 kpc above the Galactic Plane. The luminosity function derived from the Pleiades cluster, the only well-studied open cluster that is sufficiently young to be dynamically intact, is also closely consistent with the Solar Neighbourhood data. Taken together, these results provide a solid basis for assuming that the luminosity function derived for nearby stars is representative of the ‘mature’ ($\tau > 0.5$ Gyr) population of the Galactic Disk.

We have combined results from these various surveys to derive ‘best estimates’ of $\Phi(M_V)$ (Table 8.3) and $\Phi(M_J)$ (Table 8.4). At bright visual magnitude ($M_V < 8$), the densities are from the [R18] Hipparcos-based 25-parsec data set; at fainter

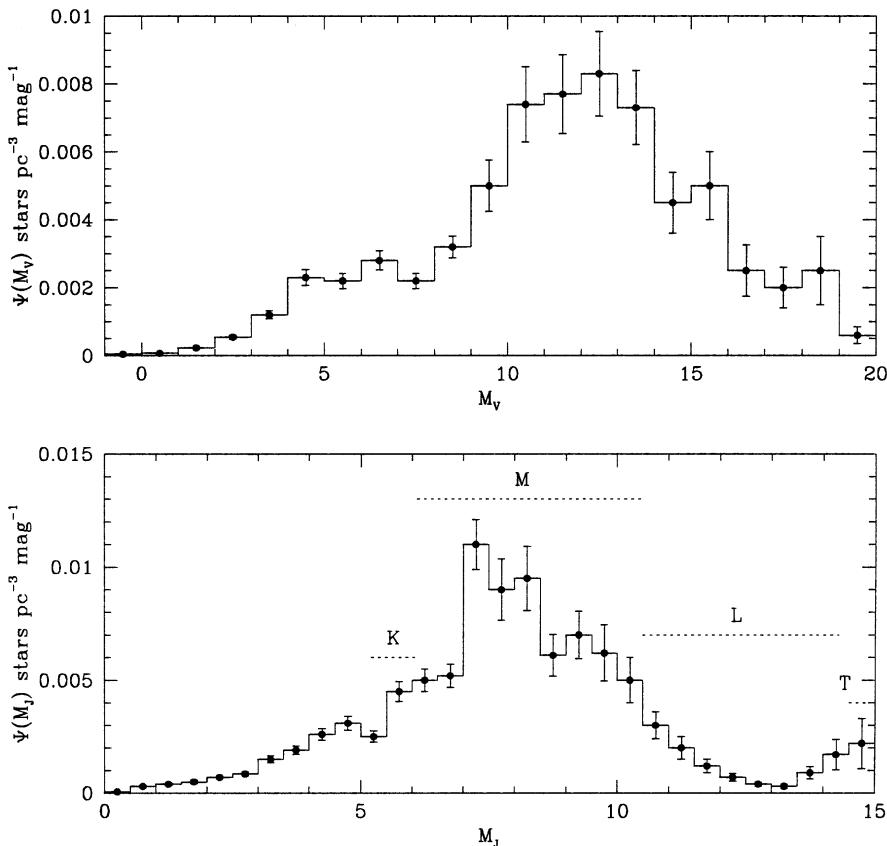


Figure 8.23. The current best estimates of the V-band and J-band luminosity functions for stars and brown dwarfs in the Solar Neighbourhood.

magnitudes, the statistical values are derived from the [R11] analysis and results for the 8-parsec sample. The J -band luminosity function is drawn primarily from the NStars survey [C3], [R21], combined with the T dwarf data from Burgasser's 2MASS survey [B9]. Both luminosity functions are plotted in Figure 8.23.

8.11 REFERENCES

- A1 van Altena, W. F., 1969, *AJ*, **74**, 2.
- A2 Armandroff, T. E., 1983, *IAU Colloquium 76, The Nearby Stars and the Stellar Luminosity Function*, ed. A. G. Davis Philip and A. R. Upgren, p. 229.
- A3 Adams, J. D., Stauffer, J. R., Monet, D. G., Skrutskie, M. F., Beichman, C. A., 2001, *AJ*, **121**, 2053.
- A4 Adams, J. D., Stauffer, J. R., Skrutskie, M. F., Monet, D. G., Portegies Zwart, S. F., Janes, K. A., Beichman, C. A., 2002, *AJ*, **124**, 1570.

- B1 Bessell, M. S., 1986, *PASP*, **98**, 1303.
B2 van Biesbroeck, G., 1944, *AJ*, **51**, 61.
B3 van Biesbroeck, G., 1961, *AJ*, **66**, 528.
B4 Boesgaard, A. M., 1990, *ApJ*, **351**, 467.
B5 Bok, B. J., Macrae, D. A., 1941, *Annals New York Acad. Sci.*, **42**, 219.
B6 Boss, L., 1908, *AJ*, **26**, 31.
B7 Bryja, C., Humphreys, R. M., Jones, T. J., 1994, *AJ*, **107**, 246.
B8 van Bueren, H. G., 1952, *Bull. Astr. Inst. Neth.*, **11**, 385.
B9 Burgasser, A. J., 2004, *ApJS*, **155**, 191.
B10 Bouvier, J., et al., 1998, *A&A*, **336**, 490.
B11 Burke, C. J., Pinsonneault, M. H., Sills, A., 2004, *ApJ*, **604**, 272.
B12 Bildsten, L., Brown, E. F., Matner, C. D., Ushomirsky, G., 1997, *ApJ*, **482**, 442.
B13 Barrado y Navascues, D., Bouvier, J., Stauffer, J. R., Lodieu, N., McCaughrean, M. J., 2002, *A&A*, **395**, 813.
B14 Barrado y Navascues, D., Stauffer, J. R., Jayawardhana, R., 2004, *ApJ*, **614**, 386.
B15 Bochanski, J. J., Hawley, S. L., Reid, I. N., Corey, K., Munn, J. A., Tinney, C. G., 2004, *BAAS*, **205**, 5506.
C1 Chandrasekhar, S., 1942, *Principles of Stellar Dynamics*, Dover Publications Inc., New York.
C2 Comstock, G. C., 1910, *Astr. Nachr.*, **185**, 297.
C3 Cruz, K. L., et al., 2003, *AJ*, **125**, 354.
D1 Dahn, C. C., Liebert, J., Harrington, R. S., 1986, *AJ*, **91**, 621.
D2 Dahn, C. C., et al., 2002, *AJ*, **124**, 1170.
D3 Dobbie, P. D., Kenyon, F., Jameson, R. F., Hodgkin, S. T., Hambly, N. C., Hawkins, M. R. S., 2002, *MNRAS*, **329**, 543.
E1 Eggen, O. J., 1950, *ApJ*, **112**, 144.
E2 Eggen, O. J., 1984, *AJ*, **89**, 830.
F1 Faber, S. M., Burstein, D., Tinsley, B. M., King, I. R., 1976, *AJ*, **81**, 45.
F2 Foster, D. C., Byrne, P. B., Hawley, S. L., Rolleston, W. R. J., 1997, *A&AS*, **126**, 81.
F3 Friel, E. D., 1995, *ARA&A*, **33**, 381.
F4 de la Fuente Marcos, R., 1997, *A&A*, **322**, 764.
G1 Gilmore, G. F., Reid, I. N., Hewett, P. C., 1986, *MNRAS*, **213**, 257.
G2 Giclas, H. L., Burnham, R., Thomas, N. C., 1972, The Lowell Observatory Proper Motion Survey, *Lowell Obs. Bull.* 158.
G3 Gliese, W., 1957, *Veröff. astr. Rechen Inst. Heidelberg*, **8**.
G4 Gliese, W., 1969, *Veröff. astr. Rechen Inst. Heidelberg*, **22**.
G5 Gliese, W., 1972, *QJRAS*, **13**, 138.
G6 Gliese, W., Jahreiss, H., 1979, *A&AS*, **38**, 423.
G7 Gliese, W., Jahreiss, H., 1991, *Preliminary Version of the Third Nearby Star Catalogue CNS3*, Strasburg.
G8 Gould, A., Bahcall, J. N., Flynn, C., 1996, *ApJ*, **465**, 759.
G9 Gould, A., Bahcall, J. N., Flynn, C., 1997, *ApJ*, **482**, 913.
G10 Gunn, J. E., Griffin, R. F., Griffin, R. E. M., Zimmerman, B. A., 1988, *AJ*, **96**, 198.
H1 Hambly, N., Hawkins, M. R. S., Jameson, R. F., 1993, *MNRAS*, **263**, 647.
H2 Hambly, N. C., Steele, I. A., Hawkins, M. R. S., Jameson, R. F., 1995, *MNRAS*, **273**, 505.
H3 Hanson, R. B., 1975, *AJ*, **80**, 379.
H4 Hanson, R. B., 1979, *MNRAS*, **186**, 875.

- H5 Hanson, R. B., 1983, *IAU Coll.* 76, p. 51, ed. A. R. Upgren, L. Davis Press, Schenectady.
- H6 Hanson, R. B., Vasilevskis, S., 1983, *AJ*, **88**, 844.
- H7 Hawkins, M. R. S., Bessell, M. S., 1988, *MNRAS*, **234**, 177.
- H8 Hawley, S. L., Gizis, J. E., Reid, I. N., 1996, *AJ*, **112**, 2799.
- H9 Herbst, W., Dickman, R. L., 1983, in *Nearby Stars and the Stellar Luminosity Function*, p. 187, eds. A. R. Upgren and A. G. Davis Philip, L. Davis Press, Schenectady.
- H10 Herschel, W., 1785, *Phil. Trans.*, **75**, 213.
- H11 Hertzsprung, E., 1934, *Bull. Astr. Inst. Neth.*, **7**, 258.
- H12 Holland, K., Jameson, R. F., Hodgkin, S., Davies, M. B., Pinfield, D., 2000, *MNRAS*, **319**, 956.
- H13 Heckman, V. O., Dieckvoss, W., Kox, H., 1956, *ANac*, **283**, 109.
- J1 Jarrett, T. H., Dickman, R. L., Herbst, W., 1994, *ApJ*, **424**, 852.
- J2 Jones, B. F., Stauffer, J. R., 1991, *AJ*, **102**, 1080.
- J3 Jones, D. H. P., 1973, *MNRAS*, **161**, 19P.
- J4 Jameson, R. F., Dobbie, P. D., Hodgkin, S. T., Pinfield, D. J., 2002, *MNRAS*, **335**, 853.
- K1 Kapteyn, J. C., 1902, *Publ. Astr. Lab. Gröningen*, No. 11.
- K2 Kapteyn, J. C., 1909, *ApJ*, **29**, 46.
- K3 Kapteyn, J. C., 1914, *ApJ*, **40**, 43.
- K4 Kapteyn, J. C., van Rhijn, P. J., 1924, *ApJ*, **52**, 23.
- K5 Kibblewhite, E. J., Bridgeland, M. T., Bunclark, P., Irwin, M. J., 1984, *Proceedings of the Astronomical Microdensitometry Conference*, ed. D. A. Klingsmith, NASA Conf. Publ. 2317.
- K6 King, I. R., 1977, *Highlights of Astronomy*, Vol. 4, ed. E. A. Müller, Reidel, Dordrecht.
- K7 Kirkpatrick, J. D., McGraw, J. T., Hess, T. R., Liebert, J., McCarthy, D. W., 1994, *ApJS*, **94**, 749.
- K8 Klare, G., Schaifers, K., 1966, *Astr. Nachr.*, **289**, 81.
- K9 Klein-Wassink, W. J., 1927, *Publ. Kapteyn Astr. Lab.*, No. 41.
- K10 Kroupa, P., 1995, *ApJ*, **453**, 358.
- K11 Kuiper, G. P., 1942, *ApJ*, **95**, 201.
- L1 van Leeuwen, F., 1982, PhD thesis, Leiden University.
- L2 van Leeuwen, F., Hansen Ruiz, C. S., 1997, in *Hipparcos Venice '97*, eds. B. Battrick and M. A. C. Perryman, ESA, p. 689.
- L3 Leggett, S. K., Dahn, C. C., Harris, H., 1994, *AJ*, **108**, 944.
- L4 Leggett, S. K., Hawkins, M. R. S., 1988, *MNRAS*, **234**, 1045.
- L5 Leggett, S. K., Hawkins, M. R. S., 1989, *MNRAS*, **238**, 145.
- L6 Lutz, T. E., Kelker, D. H., 1973, *PASP*, **85**, 573.
- L7 Luyten, W. J., 1923, *Lick Obs. Bull.*, **11**, 1.
- L8 Luyten, W. J., 1924, *ApJ*, **62**, 8.
- L9 Luyten, W. J., 1939, *Publ. Obs. Minn.*, **2**, 121.
- L10 Luyten, W. J., 1941, *Annals New York Acad. Sci.*, **42**, 201.
- L11 Luyten, W. J., 1968, *MNRAS*, **139**, 221.
- L12 Luyten, W. J., 1974, *Proper Motion Survey with the 48-inch Schmidt Telescope*, XLVI.
- L13 Luyten, W. J., 1977, *Highlights in Astronomy*, Vol. 4, p. 89, ed. E. A. Müller, Reidel, Dordrecht.
- L14 Luyten, W. J., Hill, G., Morris, S., 1981, *Proper Motion Survey with the 48-inch Schmidt Telescope*, LIX, University of Minnesota, Minneapolis.
- L15 Lepine, S., Shara, M. M., Rich, M. R., 2003, *AJ*, **124**, 1190.
- L16 Lepine, S., Shara, M. M., Rich, M. R., 2003, *AJ*, **126**, 921.

- M1 McCuskey, S. W., 1963, *Vistas in Astronomy*, **7**, 141.
- M2 Mazzei, P., Pigatto, L., 1988, *A&A*, **193**, 148.
- M3 van Maanen, A., 1937, *ApJ*, **85**, 26.
- M4 Malmquist, K. G., 1936, *Stockholm Obs. Medd.*, Nr. 26.
- M5 Martini, P., Osmer, P. S., 1998, *AJ*, **116**, 2513.
- M6 Mermilliod, J.-C., Robichon, N., Arenou, F., Lebreton, Y., 1997, in *Hipparcos Venice '97*, eds. B. Battinck and M. A. C. Perryman, ESA, 643.
- M7 Meusinger, H., Schilbach, E., Souchay, J., 1996, *A&A*, **312**, 833.
- M8 Mihalas, D., Binney, J., 1981, *Galactic Astronomy*, W. H. Freeman & Company, San Francisco.
- M9 Murray, C. A., 1988, *IAU Symposium 133: Mapping the Sky*, p. 143, eds. S. Débarbat, J. A. Eddy, H. K. Eichhorn and A. R. Upgren, Reidel, Dordrecht.
- M10 Murray, C. A., Sanduleak, N., 1972, *MNRAS*, **157**, 273.
- M11 Munari, U. *et al.*, 2004, *A&A*, **418**, L31.
- M12 Martin, E. L., Basri, G., Gallegos, J. E., Rebolo, R., Zapatero-Osorio, M. R., Bejar, V. S., 1998, *ApJ*, **499**, L61.
- O1 Oort, J. H., 1932, *Bull. Astr. Inst. Neth.*, **6**, 249.
- O2 Oort, J. H., 1965, *Galactic Structure*, ed. M. Schmidt, University of Chicago Press.
- O3 Oort, J. H., 1974, *Highlights of Astronomy*, Vol. 3, p. 317, ed. G. Contopoulos, Reidel, Dordrecht.
- O4 Oliveira, J. M., Jeffries, R. D., Devey, C. R., Barrado y Navascues, D., Naylor, T., Stauffer, J. R., Totten, E. J., 2003, *MNRAS*, **342**, 651.
- P1 Pels, G., Oort, J. H., Pels-Kluyver, H. A., 1975, *A&A*, **43**, 423.
- P2 Perryman, M. A. C., Brown, A. G. A., Lebreton, Y., Gómez, A., Turon, C., Cayrel de Strobel, G., Mermilliod, J.-C., Robichon, N., Kovalevsky, J., Crifo, F., 1998, *A&A*, **331**, 81.
- P3 Pesch, P., 1972, *ApJ*, **177**, 519.
- P4 Pesch, P., Sanduleak, N., 1978, *AJ*, **83**, 1090.
- P5 Pesch, P., Dahn, C. C., 1982, *AJ*, **87**, 122.
- P6 Pinsonneault, M. H., Stauffer, J., Soderblom, D. R., King, J. R., Hanson, R. B., 1998, *ApJ*, **504**, 170.
- P7 Probst, R. G., 1983, *ApJ*, **274**, 237.
- P8 Prosser, C. F., Stauffer, J. R., Kraft, R. P., 1991, *AJ*, **101**, 1361.
- P9 Prosser, C. F., Randich, S., Stauffer, J. R., Schmitt, J. H. M. M., Simon, T., 1996, *AJ*, **112**, 1570.
- P10 Pokorny, R. S., Jones, H. R. A., Hambly, N. C., Pinfield, D. J., 2004, *A&AS*, **421**, 763.
- P11 Pinfield, D. J., Hodgkin, S. T., Jameson, R. F., Cossburn, M. R., Hambly, N. C., Devereux, N., 2000, *MNRAS*, **313**, 347.
- P12 Pinfield, D. J., Hodgkin, S. T., Jameson, R. F., Cossburn, M. R., von Hippel, T., 1997, *MNRAS*, **287**, 180.
- P13 Pinfield, D. J., Dobbie, P. D., Jameson, R. F., Steele, I. A., Jones, H. R. A., Katsianannis, A. C., 2003, *MNRAS*, **342**, 1241.
- P14 Prosser, C. F., 1994, *AJ*, **107**, 1422.
- R1 Raboud, D., Mermilliod, J.-C., 1998, *A&A*, **329**, 101.
- R2 Raboud, D., Mermilliod, J.-C., 1998, *A&A*, **333**, 897.
- R3 Reid, I. N., 1982, *MNRAS*, **201**, 51.
- R4 Reid, I. N., 1984, *MNRAS*, **206**, 1.
- R5 Reid, I. N., 1991, *AJ*, **102**, 1428.
- R6 Reid, I. N., 1992, *MNRAS*, **257**, 257.

- R7 Reid, I. N., 1993, *MNRAS*, **265**, 785.
- R8 Reid, I. N., Gizis, J. E., 1997, *AJ*, **113**, 2246.
- R9 Reid, I. N., Gizis, J. E., 1997, *AJ*, **114**, 1992.
- R10 Reid, I. N., Gilmore, G. F., 1982, *MNRAS*, **201**, 73.
- R11 Reid, I. N., Hawley, S. L., Gizis, J. E., 1995, *AJ*, **110**, 1838.
- R12 Reid, I. N., Hawley, S. L., Mateo, M., 1995, *MNRAS*, **272**, 828.
- R13 Reid, I. N., Hawley, S. L., 1999, *AJ*, **117**, 343.
- R14 van Rhijn, P. J., 1916, *Publ. Kaptyen Astr. Lab. Gröningen*, No. 26.
- R15 van Rhijn, P. J., 1925, *Publ. Kaptyen Astr. Lab. Gröningen*, No. 38.
- R16 van Rhijn, P. J., 1936, *Publ. Kaptyen Astr. Lab. Gröningen*, No. 47.
- R17 Rolleston, W. R. J., Byrne, P. B., 1997, *A&AS*, **126**, 357.
- R18 Reid, I. N., Gizis, J. E., Hawley, S. L., 2002, *AJ*, **124**, 1190.
- R19 Reid, I. N., 2004, in *The IMF@50*, eds. E. Corbelli, F. Palla and H. Zinnecker, Kluwer.
- R20 Reid, I. N., Cruz, K.L., 2002, *AJ*, **123**, 2806.
- R21 Reid, I. N., et al., 2004, *AJ*, **128**, 463.
- R22 Reid, I. N., Mahoney, S., 2000, *MNRAS*, **316**, 817.
- R23 Raymond, S. N., et al., 2003, *AJ*, **125**, 2621.
- S1 Sanduleak, N., 1965, PhD thesis, Case Institute of Technology.
- S2 Sanduleak, N., 1976, *AJ*, **81**, 350.
- S3 Schmidt, M., 1974, *Highlights of Astronomy*, Vol. 3, p. 450, ed. G. Contopoulos, Reidel, Dordrecht.
- S4 Schmidt, M., 1975, *ApJ*, **202**, 22.
- S5 Schmitt, J. H. M. M., Fleming, T. A., Giampapa, M. S., 1995, *ApJ*, **450**, 392.
- S6 Seares, F. H., 1924, *ApJ*, **59**, 310.
- S7 von Seeliger, H., 1898, *Abhandlungen der K. Bayer. Akademie d. Wiss., II. Kl.*, **19**, 564.
- S8 Simons, D. A., Henry, T. J., Kirkpatrick, J. D., 1996, *AJ*, **112**, 2238.
- S9 Smethells, W. G., 1974, PhD thesis, Case Western Reserve University.
- S10 Spitzer, L., 1940, *MNRAS*, **100**, 396.
- S11 Starikova, G. A., 1960, *Soviet Astr.*, **4**, 451.
- S12 Stauffer, J. R., Klemola, A., Prosser, C., Probst, R., 1991, *AJ*, **101**, 980.
- S13 Stauffer, J. R., Liebert, J., Giampapa, M., Macintosh, B., Reid, I. N., Hamilton, D., 1994, *AJ*, **108**, 160.
- S14 Stobie, R. S., 1986, *Patt. Rec. L.*, **4**, 317.
- S15 Stobie, R. S., Ishida, K., Peacock, J. A., 1989, *MNRAS*, **238**, 709.
- S16 Siegler, N., Close, L. M., Mamajek, E. E., Freed, M., 2003, *ApJ*, **598**, 1265.
- S17 Stauffer, J. R., Schultz, G., Kirkpatrick, J. D., 1998, *ApJ*, **499**, L199.
- S18 Stauffer, J. R., et al., 1999, *ApJ*, **527**, 219.
- S19 Smolcic, V., et al., 2004, *ApJ*, **615**, L141.
- S20 Silvestri, N., et al., 2004, *BAAS*, **205**, 12005.
- T1 Terlevich, E., 1987, *MNRAS*, **224**, 193.
- T2 Thé, P. S., Staller, R. F. A., 1974, *A&A*, **36**, 155.
- T3 Tinney, C. G., 1993, *ApJ*, **414**, 279.
- T4 Trumpler, R. J., 1920, *PASP*, **32**, 43.
- T5 Trumpler, R. J., 1930, *Lick Obs. Bull.*, 420.
- T6 Trumpler, R. J., Weaver, H. F., 1953, *Statistical Astronomy*, University of California Press.
- U1 Upgren, A. R., Armandroff, T. E., 1981, *AJ*, **86**, 1898.
- W1 Weidemann, V., Jordan, S., Iben, I., Casertano, S., 1992, *AJ*, **104**, 1876.
- W2 Weistrop, D. W., 1972, *AJ*, **77**, 849.

- W3 Weistrop, D. W., 1976, *ApJ*, **204**, 113.
- W4 Weistrop, D. W., 1980, *AJ*, **85**, 738.
- W5 Wielen, R., 1974, *Highlights of Astronomy*, Vol. 3, p. 395, ed. G. Contopoulos, Reidel, Dordrecht.
- W6 Wielen, R., Jahreiss, H., Krüger, R., 1983, *IAU Colloquium 76, The Nearby Stars and the Stellar Luminosity Function*, eds. A. G. Davis Philip and A. R. Upgren, p. 163.
- W7 Woolley, R., v.d.r., Epps, E. A., Penston, M. J., Pocock, S. B., 1970, *Royal Obs. Bull.*, No. 5.
- Z1 Zuckerman, B., Becklin, E. E., 1992, *ApJ*, **386**, 260.
- Z2 Zwart, S. F. P., Hut, P., Makino, J., McMillan, S. L. W., 1998, *A&A*, **337**, 363.
- Z3 Zwahlen, N., North, P., Debernardi, Y., Eyer, L., Gallind, F., Groenewegen, M. A. T., Hummel, C. A., 2004, *A&A*, **425**, L45.
- Z4 Zapatero-Osorio, M. R., *et al.*, 1999, *A&AS*, **134**, 537.

8.12 HOMEWORK PROBLEMS

Problem 1

Malmquist bias plays an important role in all statistical analyses that are based on samples with implicit or explicit magnitude limits. We can explore its effects through numerical simulations. Use a random number generator to generate a sample of 100,000 stars with a uniform density distribution and distances, $d < 50$ parsecs from the Sun (hint: the distances have a uniform (flat) distribution in density, d^3). Assign all of the stars absolute magnitude $M = 12$; determine the number of stars with apparent magnitudes, $m < 14$, and check that the mean absolute magnitude is $\langle M \rangle = 12$. What is the corresponding distance limit? To simulate the effects of dispersion, σ , in absolute magnitude, we generate absolute magnitude deviations from a Gaussian distribution and adjust each absolute magnitude. (Examples of both uniform and Gaussian random number generators are available at the NLDS data website, <http://www.stsci.edu/~inr/nldsdat.html>.) Adjust the absolute magnitudes using corrections generated for dispersions, $\sigma = 0.1, 0.2, 0.3, 0.5, 1.0, 1.4$ magnitudes; in each case, determine the mean absolute magnitude for stars with $M < 14$. The dispersion, σ , is identical to the σ in equation 8.26, and the deviations from $\langle M \rangle = 12$ should reproduce the predictions of that equation at small σ ; explain why this simulation underestimates the corrections at large σ .

Problem 2

We can also use numerical simulations to examine the effects of Malmquist bias on density estimates. Taking the data from each simulation constructed in the previous question, assign each star an absolute magnitude $M = 12$ and use the apparent magnitudes to estimate photometric distances; determine how the apparent density of stars within 25 parsecs varies with σ .

Problem 3

Show that starcounts (the number magnitude distribution) increase as $10^{0.6m}$ for a uniform density distribution.

Problem 4

The NLDS data website, <http://www.stsci.edu/~inr/nldsdat.html>, lists simulated position, photometry and proper motions for 300 members of Corona Borealis, a hypothetical open cluster centred at (16 h 10 m, +25°). Data are also listed for 900 field stars in the same area. Use photometric and proper motion criteria to identify the likely cluster members; estimate the age of the cluster. If the cluster has a convergent point at (19 h 10 m, +65°) and a true space motion (relative to the Sun) of 56 km s⁻¹, estimate the distance to the centre of the cluster and its approximate dimensions.

9

The mass function

9.1 INTRODUCTION

The number of stars (and brown dwarfs) per unit mass – the stellar mass function, $\Psi(M)$ – describes how a molecular cloud redistributes its material to form stars. This parameter is also necessary for determining the mass distribution and the total stellar mass in star clusters and external galaxies. It is therefore fundamental to both star-formation theory and Galactic structure.

Deriving $\Psi(M)$ is more demanding than determining $\Phi(M)$. Given a well-defined sample of stellar systems, individual distances and apparent brightnesses can be measured, and $\Phi(M)$ calculated directly. In contrast, masses are currently known for relatively few stars, limited almost exclusively to components of resolved or eclipsing binary systems. As a result, indirect techniques must be employed to estimate masses for the overwhelming majority of stars, even among those within 8 parsecs of the Sun.

The necessity of relying on indirect mass estimates would not be a complication in determining $\Psi(M)$ if there were a simple, single-valued relation between a directly measurable quantity and mass. Unfortunately, this is not the case. In main sequence stars, luminosity depends strongly on mass, but is also dependent to a lesser extent on factors such as age and chemical composition. As already discussed in Chapter 8, luminosity evolution invalidates direct comparison between $\Phi(M)$ for nearby stars and results for young open clusters and star-forming regions, where stars are still in the process of contracting onto the main sequence. Conversely, the low metal abundance of elderly subdwarfs in the Galactic halo population leads to their having luminosities and colours significantly different from those of solar-abundance dwarfs of the same mass.

Despite the observational and theoretical difficulties, the stellar mass function can be determined for stars in a range of environments. The opening sections of this chapter outline the terminology used in describing the mass function and the

methods used to measure stellar masses. The latter sections summarise results derived for stars in the Galactic Disk: the nearest stars; intermediate-aged clusters such as the Pleiades; and young associations. Metal-poor subdwarfs and the halo mass function are discussed in Chapter 10.

9.2 DEFINING THE MASS FUNCTION

9.2.1 Basic terminology

The following symbols are used to represent the stellar luminosity function and mass function:

$$\Phi(M) = \frac{dN}{dM}: \text{ the luminosity function in stars pc}^{-3} \text{ mag}^{-1}$$

$$\Psi(M) = \frac{dN}{dM}: \text{ the mass function in linear units, stars pc}^{-3} M_{\odot}^{-1}$$

$$\xi(M) = \frac{dN}{d \log M}: \text{ the mass function in logarithmic units, stars pc}^{-3} \log(M_{\odot})^{-1}$$

A luminosity function is transformed to a mass function by applying the appropriate mass–luminosity relationship (MLR), dL/dM ,

$$\frac{dN}{dM} = \frac{dN}{dL} \frac{dL}{dM} \quad (9.1)$$

where the luminosity function is given in general by dN/dL . Strictly speaking, the term *mass–luminosity relationship* applies to the correlation between mass and bolometric luminosity. The expression, however, is often used to refer to the mass calibration for individual passbands, such as the mass– M_V relationship, just as *luminosity function* is used to describe the number of stars per unit M_V , M_I , M_K , in addition to per unit luminosity.

The linear and logarithmic mass functions are related as follows:

$$\begin{aligned} \Psi(M) &= \frac{dN}{dM} = \frac{dN}{d(\log M)} \frac{d(\log M)}{dM} \\ &= \log_{10} e \frac{dN}{d(\log M)} \frac{d(\ln M)}{dM} \\ &= \frac{0.4343}{M} \xi(M) \end{aligned} \quad (9.2)$$

Finally, the *star-forming mass distribution* is defined as

$$\Xi(M) = \frac{M \times \Psi(M)}{\int_{\min}^{\max} \Psi(M) dM} \quad (9.3)$$

This function describes the mass budget for star formation – the fraction of the total mass devoted to forming stars of a specific individual mass.

9.2.2 The Salpeter mass function

The first serious attempt to deduce the form of the initial mass function for disk stars was carried out by Salpeter [S1] in the mid-1950s. Many of the basic concepts, including the parameterisation of the mass function as a power-law, were introduced in this classic paper, albeit using different terminology. Salpeter defined the *original mass function*, $\xi^i(M)$, as

$$\xi^i(M) = \frac{T_0}{d(\log M)} \frac{dN}{dt} \quad (9.4)$$

where dN is the number of stars in the mass range $d(\log M)$ created in time dt , and T_0 is the age of the Galaxy (which Salpeter took to be 5 Gyr). $\xi^i(M)$ (or $\Psi^i(M)$ in linear units) is now known as the *initial mass function*, or IMF.

Salpeter combined the van Rhijn [R7] and Luyten [L8] analyses to derive a present-day mass function, $\Psi(M)$. With no direct measurement of the Galactic star-formation history, Salpeter assumed that the present star formation rate was characteristic of times past, and used estimates of the main sequence lifetimes of massive stars to allow for stars that had evolved off the main sequence. The observed mass function could then be transformed to the initial mass function (Figure 9.1), matched reasonably well by the approximation

$$\xi^i(M) \sim 0.03 \left(\frac{M}{M_\odot} \right)^{-1.35} \text{ stars pc}^{-3} [\log(M)]^{-1} \quad (9.5)$$

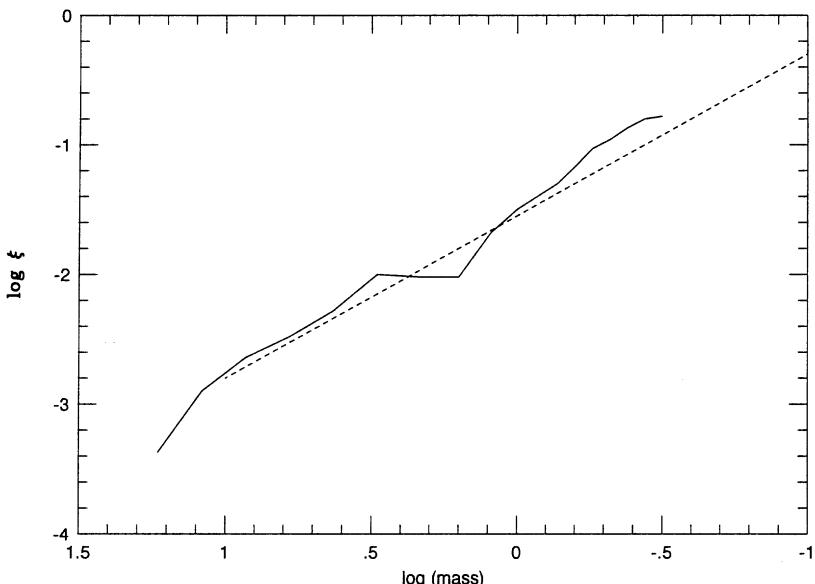


Figure 9.1. The Salpeter mass function. The solid line plots Salpeter’s empirical results, and the dotted line is a power law mass function, $x = -2.35$.

In linear units, this is

$$\Psi^i(M) \propto M^{-2.35} \text{ stars pc}^{-3} M_{\odot}^{-1} \quad (9.6)$$

Power-law mass functions are often written using the terminology $\xi(M) \propto M^{x+1}$ or $\Psi(M) \propto M^{-\alpha}$, where $\alpha = -x$. A power law mass function with slope $\alpha = 2.35$ is known as a Salpeter function; a mass function with slope $\alpha > 2.35$ is described as being ‘steeper’ than the Salpeter function, while $\alpha < 2.35$ is ‘flatter’ than the Salpeter value.

The main developments in studying the mass function since Salpeter’s initial work are surveyed in review articles by Scalo and others [M7], [S2], [R1], [S3]. The first of these papers [M7] introduced much of the terminology currently in use in the subject, particularly in the analysis of stars with masses exceeding $1 M_{\odot}$.

9.2.3 The mass function for intermediate- and high-mass stars

While M dwarfs represent the main focus of this book, their properties must be placed in a more general context. This is particularly relevant in studying the mass function, where one of the main goals is assessment of the relative proportion of mass bound *in perpetuo* as low-mass stars, and the proportion recycled in the interstellar medium as massive stars evolve. Unlike M dwarfs, stars more massive than $\sim 1 M_{\odot}$ have lifetimes shorter than the age of the Galactic Disk, and the observed luminosity function includes only a fraction of the total number formed during the history of the Milky Way. The ‘missing’ earlier generations of stars have evolved to become white dwarfs, neutron stars and black holes. In addition, higher-mass main sequence stars have a lower velocity dispersion than M dwarfs (hence Parenago’s discontinuity, Section 7.5.1), and are therefore confined more closely to the Galactic Plane. Both effects, together with possible variations in the star-formation history of the Milky Way, must be taken into account when deriving an estimate of the initial mass function. Indeed, it is these corrections, rather than uncertainties in the mass–luminosity relationship, which are the greatest source of uncertainty in $\Psi(M)$ at supra-solar masses.

In observational terms, Miller and Scalo [M7] define the present-day mass function for main sequence stars as

$$\xi_{ms}^{PD}(\log M) = \Phi(M_V) \left| \frac{dM_V}{d \log(M)} \right| 2H(M_V) f_{ms}(M_V) \quad (9.7)$$

where $dM_V/d \log(M)$ is the slope of the (M_V , mass) relationship; $H(M_V)$ is the exponential scale height of the vertical density distribution of stars, absolute magnitude M_V ; and $f_{ms}(M_V)$ is the fraction of stars with absolute magnitude M_V which are on the main sequence. ξ_{ms}^{PD} can also be written as

$$\xi_{ms}^{PD}(\log M) = \int_{T_0-\tau}^{T_0} C(\log M, t) dt \quad (9.8)$$

for stars with main sequence lifetimes, τ_{ms} , shorter than the age of the Galactic Disk, T_0 , or

$$\xi_{ms}^{PD}(\log M) = \int_0^{T_0} C(\log M, t) dt \quad (9.9)$$

for stars with $\tau_{ms} > T_0$. The latter expression holds for M dwarfs.

$C(\log M, t)$ is the creation function, the convolution of the stellar birthrate, $B(\log M, t)$, and the initial mass function, $\xi^i(\log M, t)$. If we assume that the latter two functions are separable – that is, that the initial mass function is time invariant – then the creation function can be written as

$$C(\log M, t) = \frac{\xi^i(\log M)}{T_0} \frac{B(t)}{\langle B \rangle} \quad (9.10)$$

where $\langle B \rangle$ is the time-averaged total birthrate,

$$\langle B \rangle = \frac{\int_0^{T_0} B(t) dt}{T_0} \quad (9.11)$$

Hence

$$C(\log M, t) = \frac{\xi^i(\log M)}{T_0} b(t) \quad (9.12)$$

where $b(t)$ is the relative birthrate, which expresses the star-formation rate at time t in terms of the average rate of the history of the Galactic Disk. Combining these relationships, the present-day mass function can be written as

$$\xi_{ms}^{PD}(\log M) = \frac{\xi^i(\log M)}{T_0} \int_{T_0 - \tau_{ms}}^{T_0} b(t) dt \quad (9.13)$$

for stars with $\tau_{ms} < T_0$, and

$$\xi_{ms}^{PD}(\log M) = \frac{\xi^i(\log M)}{T_0} \int_0^{T_0} b(t) dt = \frac{\xi^i(\log M)}{T_0} \quad (9.14)$$

for stars less massive than $\sim 1 M_\odot$ which have main sequence lifetimes longer than the age of the disk.

The rapid variation in main sequence lifetime with increasing mass means that results deduced for $\xi^i(\log M)$ at high masses are dependent on the functional form adopted for $B(t)$. While the problem can be constrained by requiring that there are no discontinuities in the inferred $\xi^i(\log M)$, most studies of higher-mass stars tend to concentrate on young open clusters and associations, representing individual episodes of star formation of determinate age, rather than the field. However, few Galactic associations lie within the range of direct trigonometric parallax measurement, so there are uncertainties in the distance calibration. In addition, small number statistics and interstellar reddening contribute to the uncertainties in $\xi^i(M)$. Finally, spectroscopic observations are required to determine masses for O and B stars ($\geq 6 M_\odot$). With temperatures exceeding 15,000 K, the *UBVRI* passbands lie on the long-wavelength Rayleigh-Jeans tail, and show little variation with spectral type.

Indeed, as the peak radiation shifts to shorter wavelengths, O stars and B stars overlap in luminosity at visual wavelengths, so even M_V becomes a poor indicator of intrinsic luminosity, and hence of mass.

Observations of stars in open clusters show that the initial mass function has a steep slope at masses exceeding $\sim 2 M_\odot$. There is significant dispersion in the slope measured in different clusters, probably due in part to the difficulties outlined above. Scalo [S3] has assembled more than 50 separate measurements of the slope of the stellar mass function (fitted as a power law) at masses of $2 M_\odot$ or above, and the average value is $\langle \alpha \rangle \sim 2.5 \pm 0.4$, close to Salpeter's original result. There is no obvious correlation with either metal abundance or environment; neither, it should be emphasised, is there strong evidence that the data require a power-law representation.

A full discussion of the complexities involved in determining $\Psi^i(M)$ at high masses is outside the scope of this book. Where it is necessary to extend analysis to stars more massive than $1 M_\odot$, we characterise the high-mass distribution as a power law, index $\alpha = 2.5$ (slightly steeper than Salpeter), normalising the densities to match the $\Psi(M)$ at lower masses.

9.3 MEASURING STELLAR MASSES

As noted in the introduction, direct mass measurement is possible for only a small number of individual stars. In the near future it may prove possible to use gravitational lensing techniques, but for the present, binary stars provide the only effective means of determining stellar masses. The stellar mass function cannot be derived directly from those few stars with known mass, but, if that sample is representative of the general population, those stars can be used to calibrate main sequence mass-luminosity relations.

Mass determination for stars in a binary system requires measurement of seven orbital elements: the semi-major axis, a ; the period, P ; the time of periastron passage, T_0 ; the eccentricity, e ; the longitude of periastron passage, ω ; the angle of the ascending node, Ω ; and i , the inclination of the orbital plane with respect to our line of sight. An orbital inclination of 90° corresponds to an edge-on (eclipsing) binary. These parameters can be determined using two observational techniques, direct imaging and radial velocity observations. Detailed descriptions of binary star orbit analysis may be found in [B4], [C1].

9.3.1 Astrometric binaries

If both stars in a binary system are resolved, then the relative orbit of the secondary about the primary can be measured directly. Binary star astrometry began with William Herschel, who in 1776 started recording visual measurements of the separation (ρ) and position angle (θ , measured from north through east) of double stars, with the assumption that the fainter star was more distant, and could therefore provide a reference for the determination of the parallax of the 'nearer' brighter star.

Instead of measuring parallax, Herschel detected orbital motion in some systems, including α Geminorum (Castor) and γ Virginis.

Visual measurements, primarily using bifilar micrometers, continued throughout the nineteenth and even twentieth century, with extensive catalogues compiled by (amongst others) Wilhelm and Otto Struve at Dorpat and Pulkovo Observatories, Robert G. Aitken and William J. Hussey at Lick Observatory and S. W. Burnham at Yerkes and other observatories. While photography extended observations to fainter stars, and allowed for more accurate astrometry of well separated systems, visual observations still held the edge for resolved binaries with small ($<2''$) angular separations. Exposure times of a few minutes produce images with an angular resolution equivalent to average seeing. In contrast, an experienced visual observer can take advantage of fleeting moments of superb seeing to measure a close double which would be unresolved on a photographic plate. Thus, double star observers such as George van Biesbroeck were making visual observations with large reflectors – such as the McDonald Observatory 82-inch – as late as the 1960s and early 1970s. In recent years, higher quantum-efficiency detectors, such as CCDs, and techniques such as speckle interferometry and adaptive optics (Section 1.8.1) have replaced visual observations.

The relative orbit of a binary is an ellipse with the primary star at one focus. Since the orbital plane is generally inclined with respect to the plane of the sky, observations map the projected orbit, which is also an ellipse, but with the primary no longer at an apparent focus. Measuring the latter's displacement from the focus of the apparent orbit allows determination of the inclination. Once i has been determined, the eccentricity and semi-major axis derived from the apparent orbit can be corrected to their true values. Finally, knowledge of the distance to the system is essential when converting the angular measurement of the semi-major axis, α , to a linear measurement, a .

Given a and P , the total mass of the binary system follows from Kepler's third law:

$$(M_1 + M_2) = \frac{a^3}{P^2} \quad (9.15)$$

where mass is measured in solar masses, a is measured in astronomical units and P in years. The determination of individual masses demands measurement of the relative distance of each star from the barycentre of the system. This requires absolute positional astrometry – measurement of (α, δ) for each component, rather than of the relative orientation (θ) and separation (Δ) of the two components. The contribution of parallax and proper motion can then be computed and removed to recover the individual orbits. The semi-major axes are inversely proportional to the masses:

$$\frac{M_1}{M_2} = \frac{a_2}{a_1} = \frac{\alpha_2}{\alpha_1} \quad (9.16)$$

In many binaries the lower-luminosity component is either fainter than the magnitude limit or is unresolved from the primary star, and the existence of a companion is deduced only by the presence of systematic positional residuals

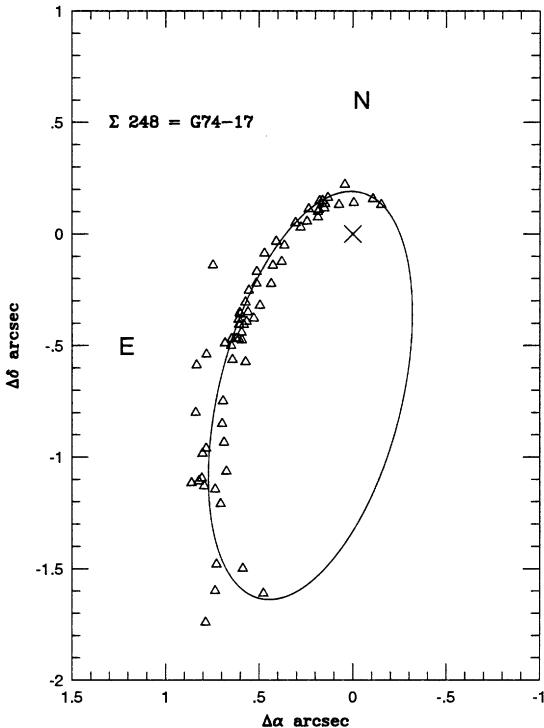


Figure 9.2. The relative orbit of the G/K dwarf binary $\Sigma 248$. The cross marks the position of the primary and observations span over 160 years, from W. Struve's original measurements in 1830 (at lower left) to speckle imaging data from 1991. (Data and orbit from [T2].)

when solving for parallax and proper motion. It was through this technique in 1840 that F. W. Bessel deduced that Sirius had an invisible companion of comparable mass (the white dwarf, Sirius B). Under such circumstances, absolute astrometric data can be used to determine the orbit described by the primary star about the barycentre. If the distance is known, and the orbital inclination estimated, then the semi-major axis, a_1 , of the primary-star orbit can be calculated. Kepler's third law can be written in the form

$$\frac{M_2^3}{(M_1 + M_2)^2} = \frac{a_1^3}{P^2} \quad (9.17)$$

and the individual masses follow from

$$M_1 = \frac{1 - B}{B^3} \frac{a_1^3}{P^2}$$

$$M_2 = \frac{1}{B^2} \frac{a_1^3}{P^2} \quad (9.18)$$

where $B = M_2/(M_1 + M_2)$ is the scale factor.

These equations can be solved for (M_1, M_2) only if B is known, which requires at least one observation of *both* components. If absolute positions for both the primary and the barycentre are known, the separation between the two components determines the relative distance of each from the barycentre, and hence the mass ratio. In the absence of direct detection, the only option is the use of indirect methods (such as setting an upper limit to the luminosity of the companion) to constrain the mass of both components.

A potential complication is introduced in unresolved systems because the photocentroid is measured: the intensity-weighted centroid of the point-spread function due to both components. If the secondary makes a significant contribution to the total flux, then the measured centroid is offset from the actual position of the primary, and α_1 is underestimated. In the extreme case of equal components, the photocentroid shows no evidence for orbital motion. This bias can be compensated if the relative luminosities of the two components are known; for example, modern high-resolution imaging techniques might be used to detect a low-luminosity companion in a star with a long history of photographic astrometry. However, it is often necessary to use observations acquired with a wide variety of photographic emulsions, electronic detectors and filters, each with their own particular spectral response; the contribution from the secondary then varies with wavelength in unequal mass systems. The derivation of an accurate orbit from such data is a process fraught with uncertainty.

Until recently, atmospheric seeing restricted astrometry to systems where the components had separations of at least 1–2 arcsec. Even at distances of 5–10 parsecs, this demands orbits with semi-major axes exceeding 5 AU and periods of decades, requiring long time-series of accurate astrometric observations. As recently as 1985, only a handful of M dwarfs could be characterised as having mass estimates of even moderate precision: just 10 stars had masses known with a formal uncertainty better than 20% [M2]. With the development of infrared speckle interferometry, complemented by Fine Guidance Sensor (FGS) observations on the Hubble Space Telescope and, recently, AO observations, it has become possible to obtain astrometric observations of systems with separations as low as 0''.1 and periods as short as 10 years [B3]. Higher precision mass measurements stem from more precise astrometry and more complete orbital coverage.

9.3.2 Spectroscopic binaries

The alternative to direct imaging is the determination of stellar orbits by monitoring radial velocities. If velocities for both stars are measurable – either through their being resolved or because they have similar luminosities and form a double-lined spectroscopic binary – then all orbital elements except inclination can be calculated. The mass ratio follows directly from the ratio between the two velocity amplitudes (K_1, K_2 – Figure 9.3), since

$$\frac{M_1}{M_2} = \frac{K_2}{K_1} \quad (9.19)$$

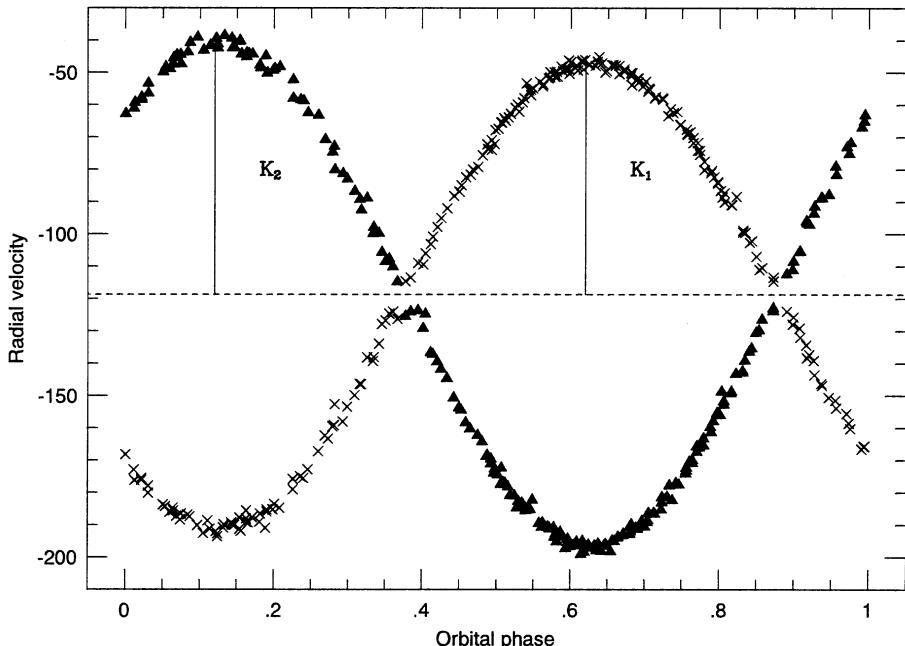


Figure 9.3. Radial velocity curves for the double-lined spectroscopic binary CM Draconis (from [M4]). The ratio of the semi-amplitude velocities of the two curves, K_1 and K_2 , is inversely proportional to the mass ratio.

Integrating each velocity curve over the full cycle allows computation of the projection of each orbit onto the line of sight (perpendicular to the plane of the sky) and hence $a_1 \sin(i)$ and $a_2 \sin(i)$. If the system is eclipsing, then the inclination must be close to 90° and individual masses can be determined. For non-eclipsing systems, at least one direct image of both components is required to determine i . Marcy and Moore [M2] provide an excellent discussion on synthesising results from radial velocity observations and astrometry of the close binary Gl 623 AB. Lacking such data, lower limits on the individual masses are given by

$$(M_1 + M_2) \geq \frac{(a'_1)^3}{P^2} \quad (9.20)$$

where $a'_1 = a \sin(i)$, the value for the semi-major axis deduced from the velocity data.

If the secondary component is faint, the binary is single-lined, allowing measurement only of the radial-velocity curve of the primary. In this case, the mass function can be determined:

$$\frac{M_2^3 \sin^3(i)}{(M_1 + M_2)^2} = \frac{(a'_1)^3}{P^2} \quad (9.21)$$

The amplitude of the reflex motion, K_1 , is given by

$$K_1 = \frac{M_2 \sin(i)}{M_1^{2/3} \sqrt{1 - e^2}} \left(\frac{2\pi G}{P} \right) \quad (9.22)$$

where i is the inclination of the pole of the orbit to the line of sight, e is the orbital eccentricity, and P is the period. Expressing the masses in solar units and the period in years, this becomes

$$K_1 = 29.79 P^{-1/3} \frac{M_2 \sin(i)}{M_1^{2/3} \sqrt{1 - e^2}} \text{ km s}^{-1} \quad (9.23)$$

Again, such observations allow limits to be set on the mass of the secondary star, but do not permit direct mass-determination.

Only four M dwarf eclipsing binaries are known: YY Gem [L1], CM Dra [M4], GJ 2069Aab [D3] and BW3 V38 [M1]. Of these, the last may be an interacting system, and it currently lacks radial-velocity data.

9.3.3 Mass determination from gravitational lensing

Light is deflected when it passes through a gravitational field. The classic example is the positional change of $1.^{\circ}75$ measured for stars in the vicinity of the Sun during the 1919 solar eclipse [D4]. A massive object also acts as a lens, amplifying the total flux delivered to a given observer from a background source to an extent which depends on the angular separation of the sources (the impact parameter) and the mass of the object [R2]. Paczyński [P1] originally pointed out the potential of this (microlensing) technique for detecting compact objects that *might* be constituents in the invisible dark-matter Galactic halo. Several projects (MACHO, EROS, DUO, OGLE) are underway or have been completed, with the goal of using statistical techniques to estimate the mass distribution of these hypothetical objects. Microlensing also offers the possibility of direct mass-measurement of isolated stellar objects.

The basic geometry of gravitational microlensing is illustrated in Figure 9.4(a) (following [P2]). Consider a source, S , at distance D_S from an observer, O . The light from that source passes close to an object, L , mass M , at distance D_L . The object acts as a gravitational lens, deflecting light from S through angle α , so that the source appears to be at position S' . In fact, if L is effectively a point source, then S will produce two images (as with QSO 0957+561, [W3]). If

$$D = \frac{(D_S - D_L)D_L}{D_S} \quad (9.24)$$

is the effective lens distance,

$$R_g = \frac{2GM}{c^2} \quad (9.25)$$

the gravitational radius of object L , and

$$R_E = \sqrt{2R_g D} \quad (9.26)$$

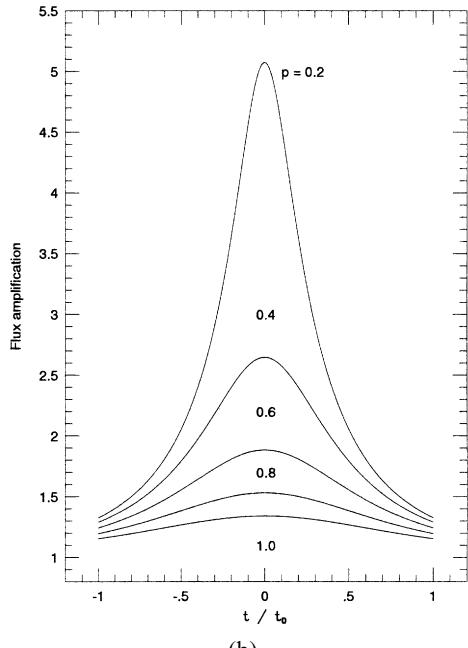
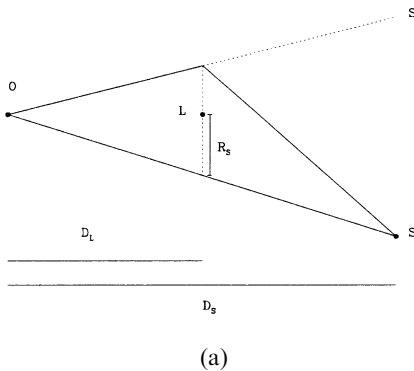


Figure 9.4. (a) Lensing geometry: O is the observer, S is the source, L is the lens and S' the observed position of the source from O . (b) Flux amplification for a range of impact parameters (see equation (9.28)).

the linear Einstein radius, then the positions of the images are offset by

$$R_{+-} = 0.5 \left[R_S \pm \sqrt{(R_S^2 + 4R_E^2)} \right] \quad (9.27)$$

where R_S is the distance between the lens, L , and the undeviated light path at distance D_L (Figure 9.4(a)). Since the lens is moving with respect to the background object, R_S changes with time; at closest conjunction, $R_S = R_{\min}$. Gravitational lensing by L amplifies the flux of the background object, S , by a factor A , given by:

$$A = \frac{u^2 + 2}{u\sqrt{(u^2 + 4)}}, \quad \text{where } u = \frac{R_S}{R_E}; \quad p = \frac{R_{\min}}{R_E} \quad (9.28)$$

The maximum amplification of a lensing event depends on the impact parameter, p , the minimum separation of L and S expressed in terms of the Einstein radius (i.e., the value of u when $R_S = R_{\min}$; see Figure 9.4(b)).

In the case of a Galactic source, the angular diameter of the Einstein ring (in milliarcseconds) is given by

$$r_E = \frac{R_E}{D_L} = 0.902 \sqrt{\left(\frac{M}{M_\odot}\right) \left(\frac{10 \text{ kpc}}{D_L}\right) \left(1 - \frac{D_L}{D_S}\right)} \text{ mas} \quad (9.29)$$

Equation (9.28) shows that the source amplification depends on u , the angular separation of L and S . In the case of a Galactic source, this parameter varies on relatively rapid timescales due to differences in the proper motions of S and L . The characteristic time for this variation is given by the time to traverse one Einstein radius, or

$$t_0 = 0.214 \text{ yr} \sqrt{\frac{M}{M_\odot} \frac{D_L}{10 \text{ kpc}} \left(1 - \frac{D_L}{D_S}\right)} \frac{200 \text{ km s}^{-1}}{V_T} \quad (9.30)$$

with V_T the transverse velocity of the lens.

Figure 9.4(b) shows the predicted light amplification for unresolved sources at various impact parameters. The amplification is wavelength-independent, providing a key property used to identify lensing events in the vast databases accumulated by the various lensing surveys. The average mass of the lensing objects is

$$\langle M \rangle = \frac{c^2 V_{\text{rms}}^2}{G D_S} t_{0,\text{av}}^2 \quad (9.31)$$

where $\log t_{0,\text{av}} = \langle \log t_0 \rangle$ and V_{rms} is the tangential velocity dispersion of the lenses. Current microlensing surveys are directed towards either the Magellanic Clouds or the Galactic Bulge (depending on the time of year) – targets that provide many background sources with known distance. The velocity dispersion of the lenses is usually taken as $\sim 200 \text{ km s}^{-1}$, as might be expected for a non-rotating, pressure-supported population (Section 7.3.3).

The lensing events toward the Magellanic Clouds have durations of 16–229 days, with most events in the range 20–60 days. Assuming halo-like velocities, this implies typical masses for the lenses of $0.5 \pm 0.3 M_\odot$ [A2], [A3]; indeed, the absence of short-period events (few hours or less) sets extremely stringent limits on the potential dark matter contribution from brown dwarfs or planetary-mass objects [G8]. The nature of the Massive Compact Halo Objects (MACHOs) that are responsible for the observed microlensing events remains unclear. The frequency of the events, a total of 17 over 5.7 years, indicates that, if these objects have the inferred masses, then they account for less than 20% of the dark matter complement of the Milky Way [A3], [G8]. It remains possible that most of the events are due to self-lensing within the Magellanic Clouds [S9], and dark matter may remain as inscrutable as ever.

Equations (9.28–9.30) show that we can estimate a mass for the lensing object, L , if both L and S are resolved. This requirement is met for one LMC event, MACHO-LMC-5, where both objects were resolved in HST WFPC-2 observations taken over 6 years after the microlensing event. The lens is an M dwarf at a distance of ~ 200 parsecs (hence the significant motion), and Gould *et al.* [G6] derive a mass of $0.097 \pm 0.016 M_\odot$, the first mass estimate for an isolated star (apart from the Sun).

This technique may eventually be applied to other lensed sources, particularly those toward the Galactic Bulge, where most lenses are expected to be foreground disk stars, offering the potential for separation between lens and source in a relatively short time frame. Alternatively, astrometric measurement of the curved paths described by lensed source(s), as the lens moves across the line of sight, provide a means of estimating the mass of L . The displacements are measured in milliarcseconds, and therefore require high-resolution interferometry and absolute astrometry; several experiments of this type are planned for the forthcoming SIM Planetquest mission, scheduled for launch in 2009 (see Section 11.6.2).

9.3.4 Summary

Mass estimates are currently available for stars in more than 100 systems, both eclipsing binaries (see the compilation by [A5]) and nearby astrometric binaries [H6], [H7]. The majority of known eclipsing systems are spectral type G or earlier, whereas most astrometric analyses centre on K and M binaries. This dichotomy does not necessarily imply that eclipsing systems are more common amongst early-type stars: rather, it reflects the spectral type distribution of stars brighter than 10th magnitude, which have been studied in more detail than stars at fainter apparent magnitudes.

In general, orbital parameters are determined to higher precision for the eclipsing binaries than for the astrometric systems. The majority of the former have relatively short periods (days or hours) and large velocity amplitudes ($>50 \text{ km s}^{-1}$, Figure 9.3). Radial velocities can be measured to a precision of better than 100 m s^{-1} , while the short periods allow for complete phase coverage. In contrast, local astrometric binaries have typical separations of one to several arcsec and periods of decades to centuries. While speckle interferometry and HST FGS observations can measure the relative positions with a typical precision of $\sim 0.^{\circ}05$, the long periods demand integration of older, lower-precision data into the solution. Even the inclusion of the latter observations does not guarantee coverage of the full orbit (Figure 9.2). Thus, astrometric mass determinations are generally less accurate.

9.4 MASS–LUMINOSITY RELATIONSHIP

Stellar luminosity functions usually exhibit significant structure. Equation (9.1) shows that this structure need not be an intrinsic property of the underlying mass distribution, but can be introduced by features in the mass–luminosity conversion. As described in Section 3.4, changes in stellar structure – notably due to H_2 dissociation, the disappearance of the radiative core and the onset of degeneracy at low masses – lead to variations in the mass–temperature and/or mass–radius relationship, with consequent effects on the (M, L) relationship. Variations in both the overall energy distribution and the growth and decline of specific spectral

features, especially molecular bands, have a strong influence on the absolute magnitude/mass relation for specific passbands. Accurate definition of the relevant mass–luminosity relationship (MLR) is therefore vital in obtaining an accurate estimate of $\Psi(M)$ for a given population.

9.4.1 The mass–luminosity relation for field stars

Figures 9.5 and 9.6 plot mass–luminosity data for main sequence stars with masses measured to better than 20% accuracy. The relevant observational data are listed in Table 9.1. Ideally, one would use these masses and luminosities to derive reliable

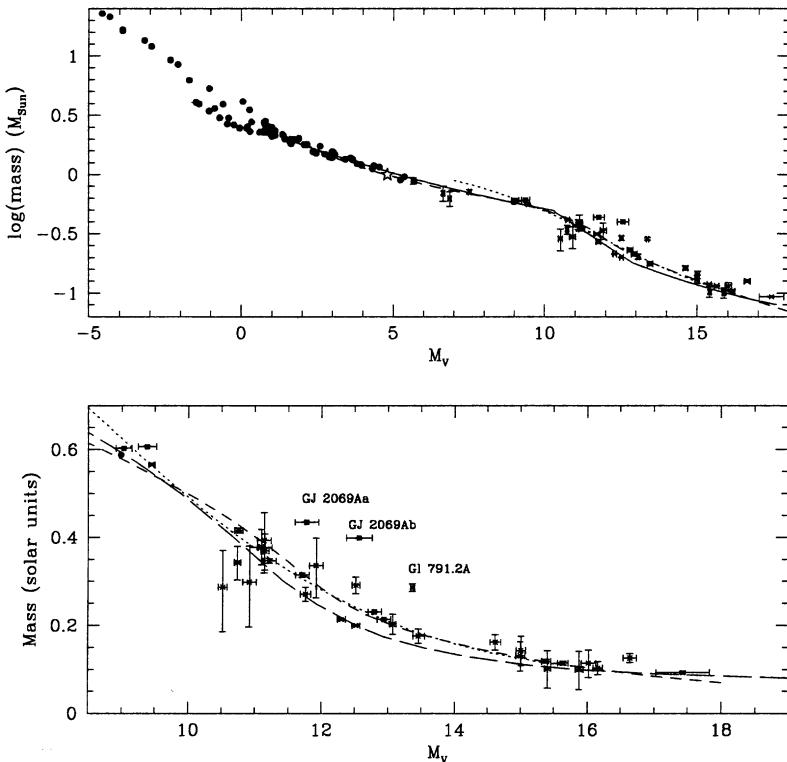


Figure 9.5. The mass–luminosity relation at visual wavelengths. Data for eclipsing binaries from Table 9.1 (YY Gem, CM Dra and CU Cnc) are plotted as solid squares, while astrometric results are shown as crosses. The upper panel includes results for high-mass eclipsing binaries from [A1] (solid points), and the Sun’s location is marked by a 5-point star; the lower panel gives an expanded view of the low-mass régime and identifies the most egregious outliers. The mass–luminosity relations shown are as follows: the solid line is the [H6] calibration (only in the upper panel); the short-dashed line is the [K3] semi-empirical relation (both panels); the dotted line is the recent [D5] empirical calibration (both panels); and the long-dashed line shows the 1-Gyr isochrone from the [B1] Lyon models (lower panel only).

Table 9.1. Low-mass binary systems.

System	Spectral type	Mass (M_{\odot})	M_V	M_K	a	e	Reference
Gl 278CD	M0	0.603 ± 0.001	9.03	5.1			YY Gem, 1, 2
	M0	0.607 ± 0.001	9.38	5.2			
Gl 630.1AB	M4.5	0.231 ± 0.001	12.82	7.7	0.0175	0.005	CM Dra, 3
	M4.5	0.214 ± 0.001	12.94	7.75			
GJ 2069Aab	M2	0.433 ± 0.001	11.78	7.1	0.0174	0.0	CU Cnc, 2, 4
	M3	0.399 ± 0.001	12.57	7.7	0.0189		
Gl 22AC	M2	0.361 ± 0.039	10.56	6.50	4.95	0.05	5, 6
		0.128 ± 0.014	13.64	8.34			
Gl 65AB	M5.5	0.102 ± 0.010	15.41	8.76	5.12	0.62	4, 5, 6
	M5.5	0.100 ± 0.010	15.87	9.16			
Gl 67AB	G1.5	0.736 ± 0.231	4.45	3.04	7.73	0.42	5
		0.221 ± 0.071	12.0	7.47			
Gl 166BC	DA	0.564 ± 0.019			333.52	0.41	7
	M4.5	0.177 ± 0.026	12.68	7.49			
Gl 234AB	M4.5	0.203 ± 0.011	13.07	7.64	4.17	0.40	4, 5, 6
		0.103 ± 0.004	16.16	9.26			
Gl 340AB	K3	0.696 ± 0.101	6.64	4.21	11.34	0.35	5
	K3	0.625 ± 0.090	6.86	4.29			
Gl 352AB	M3	0.287 ± 0.060	10.52	6.10	5.11	0.05	5
		0.298 ± 0.062	10.92	6.33			
Gl 469AB	M3.5	0.24	11.63				6
		0.17	13.22				
Gl 473AB	M5	0.143 ± 0.011	15.01	8.40			8
	M5	0.131 ± 0.010	15.00	8.84			
Gl 508AB	M0.5	0.865 ± 0.179	8.29	5.58	14.22	0.225	5
		0.606 ± 0.126	10.00	5.73			
Gl 570BC	M1	0.566 ± 0.003	9.45	5.39	0.821	0.76	4, 5
		0.377 ± 0.002	11.09	6.57			
Gl 632AB	M2.5	0.343 ± 0.030	10.74	6.46	2.06	0.566	9
		0.114 ± 0.009	16.02	9.33			
Gl 644ABab	M3	0.416 ± 0.006	10.76	6.35			2
		0.347 ± 0.005	11.22				
Gl 661AB		0.314 ± 0.004	11.77				
	M3.5	0.379 ± 0.035	11.10	6.36	4.45	0.80	
Gl 702AB		0.369 ± 0.035	11.15	6.78			
	K0	0.888 ± 0.056	5.67	3.87	22.83	0.50	
Gl 704AB	K5	0.713 ± 0.029	7.52	4.53			
	F7	0.741 ± 0.187	4.12	2.55	16.9	0.74	
Gl 725AB	K5	0.490 ± 0.124	7.48	4.61			
	M3	0.393 ± 0.061	11.14	6.68	48.5	0.53	
Gl 747AB		0.336 ± 0.052	11.92	7.21			2
	M3	0.214 ± 0.001	12.30	7.53			
		0.200 ± 0.001	12.52	7.63			

Table 9.1.—(continued)

System	Spectral type	Mass (M_{\odot})	M_V	M_K	a	e	Reference
GJ 748AB	M3.5	0.26 0.17	11.30 13.11	6.7			6
GJ 791.2AB	M4.5	0.286 ± 0.006 0.126 ± 0.003	13.37 16.64				4
GJ 831AB	M4.5	0.291 ± 0.013 0.162 ± 0.007	12.52 14.62	7.08 8.36			2, 6
GJ 860AB	M3 M4	0.271 ± 0.010 0.176 ± 0.007	11.76 13.46	6.95 8.32	9.46	0.41	4, 5
GJ 866ABC	M5.5	0.119 ± 0.011 0.115 ± 0.001 0.093 ± 0.001	15.39 15.64 17.43	8.96			2, 4
GJ 1081AB	M3.5	0.25 0.17	11.49 13.16				6
GJ 1245AC	M5.5	0.128 ± 0.021 0.074 ± 0.013	15.37 18.55	8.96 9.99	3.60	0.32	5
GJ 2005AD	M6	0.10 0.07	16.26 18.72				6
G 250-29AB	M3	0.28 0.19	11.04 12.65				6

References cited: 1 – Lacy [L1]; 2 – Segransan *et al.* [S10]; 3 – Metcalfe *et al.* [M4]; 4 – Delfosse *et al.* [D3]; 5 – Henry and McCarthy [H6]; 6 – Henry *et al.* [H7]; 7 – Reid [R4]; 8 – Torres *et al.* [T3]; 9 – Marcy and Moore [M2].

empirical calibrations that can also constrain theoretical models; in practice, the latter is possible only at masses above $1 M_{\odot}$, where most of the calibrators are eclipsing systems with well determined masses. Most systems with $M < 1 M_{\odot}$ are astrometric binaries, and the mass measurements have lower precision, leading to weaker constraints on the models; indeed, the adopted mass–luminosity calibrations must rely, to some extent, on the theoretical predictions.

Almost all of the binaries listed in Table 9.1 lie within 10 parsecs of the Sun. As noted in Section 9.3, when we use these stars to calibrate the mass–luminosity relation, we are making the assumption that they are representative of the Galactic Disk. Any bias toward, for example, young stars still in pre-main sequence contraction, or metal-poor stars, would lead to a corresponding systematic bias in the masses estimated for main sequence stars in the field. It is unlikely that any significant effect is present in the calibrators listed in Table 9.1. Even though these are local stars, the velocity dispersion of the local stars effectively guarantees that they are representative of the wider disk population (see Section 7.5.1).

On the other hand, there are several clear outliers in Figures 9.5 and 9.6, notably GJ 791.2A and both components of GJ 2069Aab. All three stars are subluminous for their mass. Direct spectroscopy of GJ 2069A [D3] suggests that at least this system is significantly metal-rich ($[m/H] \sim +0.25$ dex). A significant number of stars still have

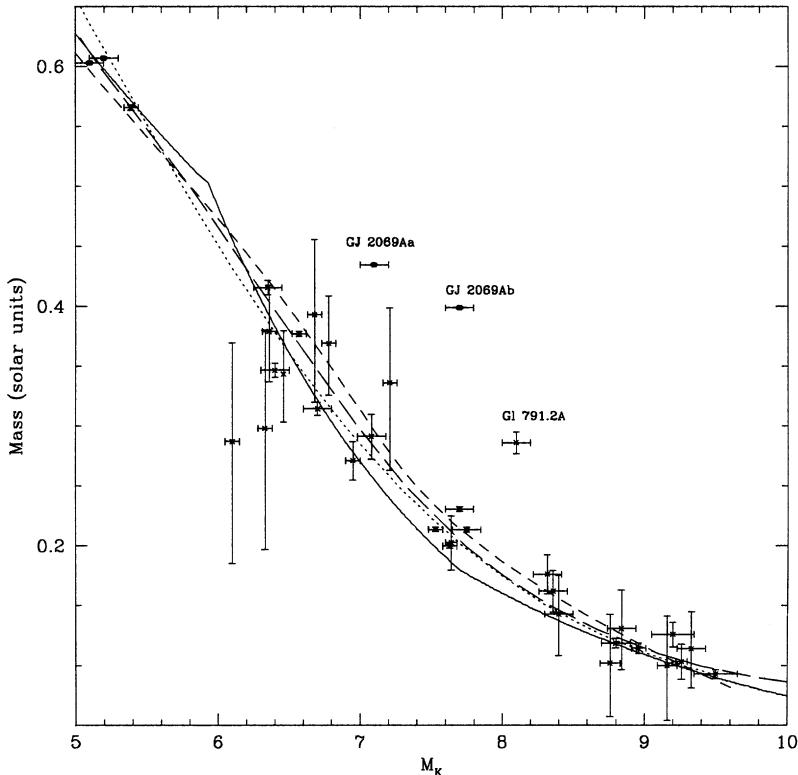


Figure 9.6. The K -band mass–luminosity relation; the data are from Table 9.1, using the same symbols as in Figure 9.5. Similarly, the relations shown are the empirical calibrations given by [H6] (solid line) and [D3] (dotted line); the [K3] semi-empirical relation (short-dashed line); and the 1-Gyr isochrone from the Lyon [B1] models (long-dashed line).

masses measured to relatively low precision; this is particularly the case for the mass range $\sim 0.5 < M/M_{\odot} < \sim 1.3$. Those orbital determinations will gradually improve, as high-precision astrometry spanning a longer baseline becomes available, but additional data points for other stellar systems are also required. Those should become available within the next decade, as one of the main science programs of SIM Planetquest (see Section 11.6.2) is focused on this issue.

Henry and McCarthy [H6] put together the first reliable calibration at subsolar masses, based primarily on their own speckle observations of many stars listed in Table 9.1. Those results were the standard reference in the 1990s. Within the last few years, new data and improved mass estimates have become available for several systems [S10], allowing refinement of the empirical calibrations [D5]:

$$\begin{aligned} \log(M/M_{\odot}) = 10^{-3} \times (0.3 + 1.87 \times M_V + 7.6140 \times M_V^2 \\ - 1.6980 \times M_V^3 + 0.060958 \times M_V^4) \quad 9 < M_V < 17 \end{aligned} \quad (9.32)$$

$$\log(M/M_{\odot}) = 10^{-3} \times (1.8 + 16.12 \times M_K + 13.205 \times M_K^2 - 6.2315 \times M_K^3 + 0.37259 \times M_K^4) \quad 4.5 < M_K < 9.5 \quad (9.33)$$

In addition, Kroupa *et al.* [K3] have used semi-empirical techniques to calculate mass–luminosity relationships. Figures 9.5 and 9.6 show these various calibrations. We also plot theoretical (mass, M_V) and (mass, M_K) 1-Gyr isochrones [B1]; even the lowest mass stars are on the main sequence at this age, so those models should be representative of the field population.

The differences between the various mass–luminosity calibrations appear relatively slight, but can have a significant effect on derivations of the stellar mass function. The original [H6] calibration is in good agreement with the [B1] models at visual wavelengths. However, both deviate from the [K3] calibration and the more recent [D5] empirical analysis, in the sense that the latter relations give higher masses (by 10–20%) for the same luminosity. There is less divergence at near-infrared wavelengths, but the overall uncertainties are at least $\pm 10\%$ for masses $M < 0.8 M_{\odot}$.

Even allowing for the uncertainties, it is clear that there are several significant changes in slope of the MLR. These are most obvious in the [H6] empirical calibration, which explicitly fits separate power laws to the mass ranges $M > 0.5 M_{\odot}$, $0.5 > M > 0.18 M_{\odot}$, and $M < 0.18 M_{\odot}$. Similarly, the [K3] semi-empirical relation has points of inflection at $M_V \sim 8$ ($M \sim 0.65 M_{\odot}$) and $M_V \sim 12$ ($M \sim 0.3 M_{\odot}$). These changes of slope match specific physical phenomena: the influence of H₂ dissociation and the emergence of H[−] as a substantive opacity source (0.5–0.6 M_{\odot}); the development of full convection ($\sim 0.25 M_{\odot}$) and increasing degeneracy ($< 0.2 M_{\odot}$).

The significance of non-linearities in the MLR can best be assessed by comparing its derivative against the luminosity function in the appropriate passband [K5]. If dM/dL is large, then a small range in luminosity corresponds to a large range in mass; in contrast, a numerically small value for the derivative implies that a given interval in mass is distributed over a large range in luminosity. Hence, a monotonic mass function can be redistributed to produce a maximum in $\Phi(M)$ in the former case, and an extended minimum in the latter. Figure 9.7 compares the first derivatives of the [B1] and [K3] (M_V , mass) relationships against $\Phi(M_V)$ for nearby stars (from Chapter 8). The maximum in the latter function at $M_V \sim 12$ lies close (but not exactly coincident with) peaks in the former functions, while both the Wielen dip and the extended tail at low luminosities match minima in the dM/dM_V relations.

None of the MLRs discussed above provide a definitive calibration for the lowest-luminosity stars ($M_V > 15, M < 0.1 M_{\odot}$) which lie close to the brown dwarf domain. As discussed in Chapter 6, this leads to consequent uncertainty in the exact location of the hydrogen-burning limit in the observational plane. Indeed, well-known low-luminosity dwarfs, such as LHS 2065 and LHS 2397a, have been proposed as possible brown dwarfs. It is now clear that these M dwarfs are

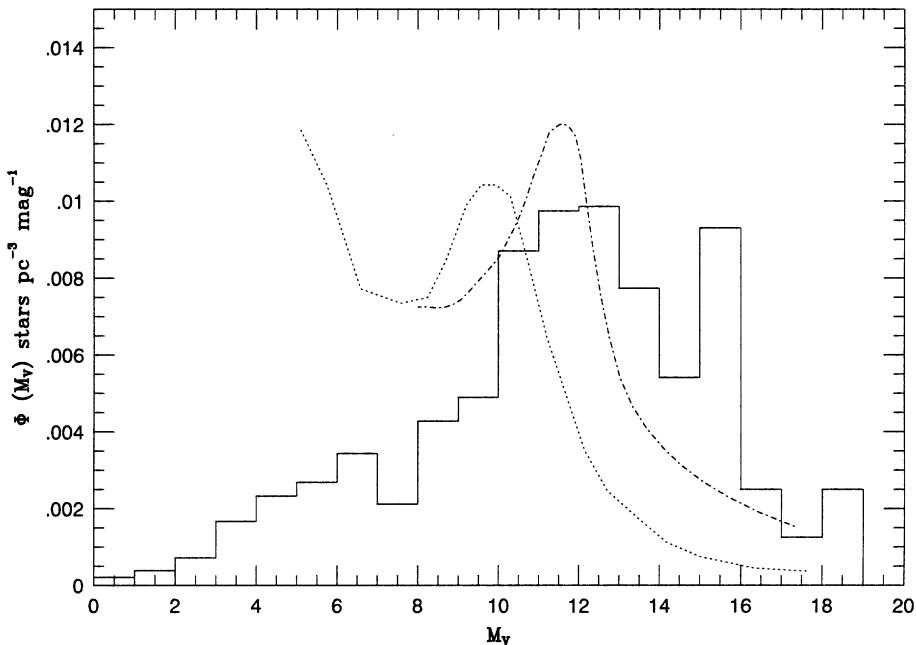


Figure 9.7. A comparison of the derivatives of the [B1] (dotted line) and [K3] (dash-dot) (mass, M_V) relations ($d(\text{mass})/dM_V$), and $\Phi(M_V)$ for the nearby stars. The peak in the latter corresponds approximately with maxima in the former.

fusion-powered, and that the stellar/substellar boundary lies at $M_V > 19$, $M_K > 10.5$ and spectral types \sim L3 to L4 (for main sequence stars).

9.4.2 The mass–luminosity relation for pre-main sequence stars

At present, few young stars have reliable empirical mass determinations. As a result, there is only limited calibration of the mass–luminosity relation during the pre-main sequence (PMS) phase. This situation does not stem from a lack of candidates: numerous pre-main sequence binary systems are known (the binary fraction in Taurus is close to 100% for solar-type stars [G3], [G4]), but few are accessible to orbital determinations. The nearest region with extensive star formation is the Taurus association, lying at a distance of \sim 150–200 parsecs; the nearest massive star-forming region is Orion, at \sim 500 parsecs. At those distances, even high-resolution imaging techniques are limited to resolving systems with separations exceeding \sim 20 AU, and orbital motions are readily measured only for high-mass systems (e.g., [M3], [W8]). Moreover, the relatively few PMS spectroscopic binaries with orbital analyses (e.g., [C3], [H12]) all have components with masses comparable to the Sun. Improved adaptive optics techniques offer hope for the future, and Stassun and collaborators have begun an observational program targeting low-mass PMS eclipsing binaries (see, e.g., [S14]).

Given the scarcity of empirical calibrators, investigations of the mass function in young ($\tau < 20$ Myr) clusters must rely exclusively on theoretical models for the MLR. There are, however, techniques that can use multiple pre-main sequence systems to test at least some aspects of these models, provided that the systems are spatially resolved.

Colour–magnitude data for intermediate-age and old open clusters, such as the Pleiades and Hyades, can be matched against theoretical predictions, since stars in these clusters are effectively coeval. Similar data for extremely young clusters show substantially more dispersion, even after allowing for effects due to differential reddening. This probably reflects the fact that star formation is not an instantaneous event, but persists for 2–3 Myr within a given region – a small fraction of the age of a cluster like the Pleiades, but a significant fraction of the whole in an active star-forming region, such as Orion. The increased dispersion in the H–R diagram precludes a simple test of pre-main sequence models. However, components of a binary (or multiple) star system can be expected to form essentially simultaneously. If temperatures and luminosities of each component can be determined, and if the distance is known with moderate accuracy, then all are expected to fall on the same isochrone.

The first studies of this type were undertaken by Hartigan *et al.* [H3], who analysed data for 39 multiple star systems, including the interesting quadruple GG Tau, which was recently re-analysed by [W4] and [P6]. The four T Tauri stars in that system form a double binary, with GG Tau Aa and Ab, separated by 0''.25, or 35 AU, lying 10''.1 from the 1''.48 (207 AU) separation, lower-luminosity pair, GG Tau Ba and Bb. All have radial velocities consistent with their belonging to a single system. The brighter pair have spectroscopic observations by the Faint Object Spectrograph on the HST, while the wider pair were observed using the Keck 10-m telescopes, allowing estimation of the effective temperatures; luminosities are known from infrared photometry. The former parameters are somewhat problematic for the lower-luminosity pair Ba and Bb, which have spectral types of M5 and M7. As described in Section 2.2.8 (and illustrated in Figure 2.12), the lower gravity of T Tauri stars leads to optical spectra resembling M giants. Mid-to-late-type M giants have higher temperatures, by 200–300 K, than M dwarfs of the same spectral type, and it remains unclear whether the dwarf or giant temperature scale is more appropriate for pre-main sequence stars.

Figure 9.8 compares empirically-derived parameters for the GG Tau quadruple against two sets of isochrones (from [B1]). The models differ in the value of the mixing-length factor (denoted as α , just to confuse matters; see Section 4.3) adopted for stars more massive than $0.6 M_{\odot}$: $\alpha = 1$, the standard model; or $\alpha = 1.9$, favoured by more recent studies. The figure also shows the temperature range bracketed by the dwarf and giant scales for the components of the low-mass binary. In either case, all four stars can be matched against a single isochrone, with ages of 5 Myr ($\alpha = 1$) or 2 Myr ($\alpha = 1.9, M < 0.6 M_{\odot}$) by choosing appropriate intermediate values of T_{eff} for the low-mass pair. Both analyses imply a mass close to $0.05 M_{\odot}$ for the lowest-mass object in the system, significantly below the H-burning limit. Once a reliable calibration of the temperature scale for low-mass

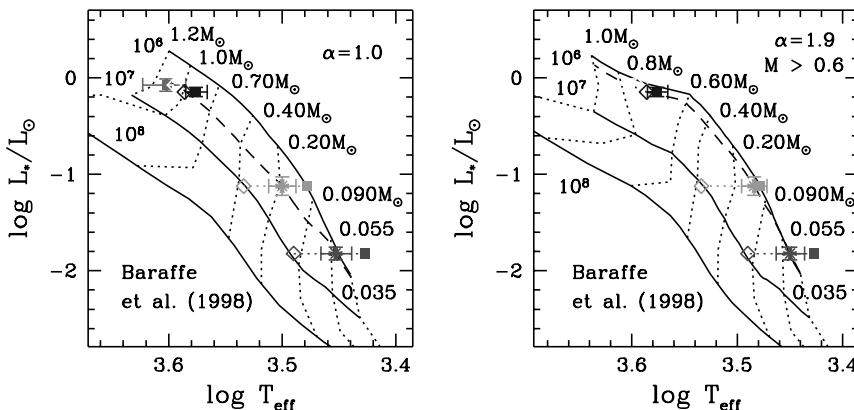


Figure 9.8. A comparison of the location of the four components of the pre-main sequence quadruple GG Tauri against theoretical isochrones computed by Baraffe *et al.* All four stars are expected to fall on the same isochrone. The calculations in the left panel are for a mixing length (α) of 1.0; the right panel shows models with $\alpha = 1.9$ for masses above $0.6 M_\odot$. The solid square and open diamond mark the upper (giant) and lower (dwarf) temperature estimates for the low-mass pair; the asterisks mark the inferred temperatures based on extrapolating the isochrone which best matches the higher-mass pair. (From [W4], courtesy of the *Astrophysical Journal*.)

T Tauri dwarfs becomes available, this technique, applied to other systems, will provide more stringent tests of the accuracy of pre-main sequence models.

9.5 THE FIELD-STAR MASS FUNCTION

Investigations of $\Psi(M)$ in the general Galactic Field can be placed under two headings: statistical determinations, which compute the luminosity function as an intermediate step; and more direct analyses, which use the mass–luminosity relationship to estimate individual stellar masses before combining the data to estimate $\Psi(M)$. Both approaches are liable to similar systematic uncertainties arising from uncertainties in the MLR. Rebinning luminosity function data can smooth over important features, while analyses based on the star-by-star approach introduce uncertainties through small number statistics. Regardless of the approach, however, the analysis stands or falls on the reliability of the parent sample: if that sample is unrepresentative, then the derived mass function will be biased and unreliable.

9.5.1 Statistical analyses of $\Psi(M)$

The majority of published studies derive $\Psi(M)$ from luminosity function data, usually $\Phi(M_V)$, either through direct application of the mass–luminosity relationship or through the inverse approach of adjusting $\Psi(M)$ (via the MLR) to match $\Phi(M_V)$,

Many of those also reach erroneous conclusions regarding the form of $\Psi(M)$, identifying a maximum at a mass of $\sim 0.25 M_{\odot}$ – a result which can be traced to various inaccuracies in the adopted luminosity functions.

In their influential paper, Miller and Scalo [M7] use the direct method to transform $\Phi(M_V)$ to $\xi(\log M)$, adopting a variety of star-formation histories and age estimates for the Galactic Disk. Several functional forms were used to represent their results, with the most frequently cited being the log-normal form,

$$\xi(\log M) = C_0 \exp[-C_1(\log M - C_2)^2] \quad (9.34)$$

where $C_0 = 106.0$, $C_1 = 1.09$ and $C_2 = -1.02$ for a constant stellar birth-rate in a 12-Gyr-old Galactic Disk. The Miller–Scalo function is a close match to a power-law with index $\alpha = 2.5$ for masses between 1 and $10 M_{\odot}$. However, unlike a power-law representation, the log-normal form reaches a maximum density (here at $\sim 0.25 M_{\odot}$) and declines thereafter. An alternative way of representing this functional form is:

$$\xi(\log M) = C_0 \exp(-(\log M - \log(M_0))^2/(2\sigma^2)) \quad (9.35)$$

where C_0 sets the density zeropoint, M_0 is the characteristic mass corresponding to the peak of the mass function and σ gives the width of the distribution. In averaging the available luminosity function data to derive the $\Phi(M_V)$ which underlies their mass function determination, Miller and Scalo follow Luyten in smoothing over the Wielen dip, and also underestimate the number density of stars at $M_V \sim 12$. The former error was corrected by Scalo [S2] in his reanalysis of the problem, but both he and Rana [R1] adopt (mass, M_V) relationships that fail to take into account the significant changes in slope below $0.6 M_{\odot}$ [K4]. Their smooth relationships redistribute stars in mass to give the apparent maximum of $\Psi(M)$.

The photometric parallax studies of the luminosity function described in Chapter 8 use direct star-by-star calibration to estimate $\Psi(M)$. The resulting mass functions are in general agreement, showing a peak in number density at $\sim 0.25 M_{\odot}$, a subsequent decline, and the suggestion of a rising function near the hydrogen-burning limit [R3], [H4], [S7], [T1]. Unfortunately, the good agreement stems from incorporating the same systematic error – an absolute-magnitude calibration which smoothes over the ‘step’ in the main sequence at $(V-I) \sim 2.9$. As was demonstrated in Section 8.5.2) (Figures 8.10 and 8.11), using a linear $(M_V, V-I)$ relationship to estimate absolute magnitudes in that region leads to stars being misplaced in $\Phi(M_V)$, enhancing the peak at $M_V \sim 12$, and decreasing the apparent numbers of fainter stars. Thus, even though the (mass, M_V) relationship used in these analyses is more accurate than the Miller/Scalo relationship, the final result is similarly incorrect.

Kroupa *et al.* [K2], [K3] pioneered the use of the inverse approach, varying $\Psi(M)$ to match a fiducial luminosity function. Their later study takes as its reference a smoothed version of the [W5] nearby-star luminosity function, supplemented by the results for VLM dwarfs within 5.2 parsecs [D1]. Using this as a constraint, they adjust the form of the mass function to minimise χ^2 in the

residuals between the required (mass, M_V) relationship and empirical data on binary star masses (from [P4]). Fitting a three-segment power-law, they derive

$$\alpha = 2.7, M > 1.0 M_\odot; \quad \alpha = 2.2, 1.0 > M > 0.5 M_\odot; \quad \alpha = 1.3, M < 0.5 M_\odot \quad (9.36)$$

The change in slope at $\sim 0.5 M_\odot$ may be partly due to incompleteness in the reference luminosity function. Adopting a similar approach, a single power-law fit to the mass range $0.1\text{--}1 M_\odot$ gives limits of $0.3 < \alpha < 0.9$, constrained mainly at the lowest masses [H5]. Reanalysis of the [K3] data resulted in a somewhat flatter relationship than the original study, with $0.66 < \alpha < 1.44$ for masses below $0.5 M_\odot$ [K4].

9.5.2 The nearby stars

The simplest, and most straightforward, method of deriving $\Psi(M)$ is to define a volume-limited sample of stars and brown dwarfs, and determine masses on an object-by-object basis. In the case of brown dwarfs, the absence of a unique mass–luminosity relation complicates matters, as discussed further in the following section, but we can apply this method to the stellar samples that form the basis of the nearby star luminosity function described in Section 8.6.

Figure 9.9, from [R8] shows the results of this exercise. The mass functions are based on two MLRs: the semi-empirical calibration from [K3]; and a purely empirical calibration, that combines the [D5] relation(s) for low-mass stars (Equations 9.32 and 9.33) with a polynomial fit to the empirical (mass, M_V) data plotted at $M > 0.8 M_\odot$. The main difference between the two calibrations is at near-solar masses: the [K3] relation passes through the Sun, but lies below stars (underestimated masses) at $3 < M_V < 4.5$; the polynomial relation passes above the Sun (mass overestimated) but is a better fit at higher luminosities. The net result is that the [K3]-calibrated mass function has a distinct change in slope at $\sim 0.7 M_\odot$, while the break lies nearer $\sim 1.0 M_\odot$ in $\Psi(M)$ derived from the polynomial calibration.

With this exception, the mass functions plotted in Figure 9.9 are in reasonable agreement. Both are relatively flat (in $dN/d\log(M)$) at low masses and steepen sharply at higher masses; note that these measure the present-day mass function (see Section 9.2.3). Following Salpeter (and many other subsequent analyses), we can represent the distributions as power laws, $\Psi^{\text{PD}}(M) \propto M^\alpha$ (all the while bearing in mind that these are simply mathematical fitting functions; the physical mechanisms that underlie the mass function are discussed further in Section 9.8). Taking this approach, the mass function derived from the [K3] MLR can be matched with power law indices: $\alpha = 1.2$ for $0.1 < M/M_\odot < 0.7$; $\alpha = 2.8$ for $0.7 < M/M_\odot < 1.8$; and $\alpha = 5.2$ for $1.8 < M/M_\odot$. The alternative MLR gives a mass function that can be represented with power law indices: $\alpha = 1.3$ for $0.1 < M/M_\odot < 1.0$; and $\alpha = 5.3$ for $1.0 < M/M_\odot$, or, in absolute terms:

$$\log(\xi^{\text{PD}}(\log M)) = -(0.3 \pm 0.15) \log M - (2.3 \pm 0.1) \quad M < 1 M_\odot \quad (9.37)$$

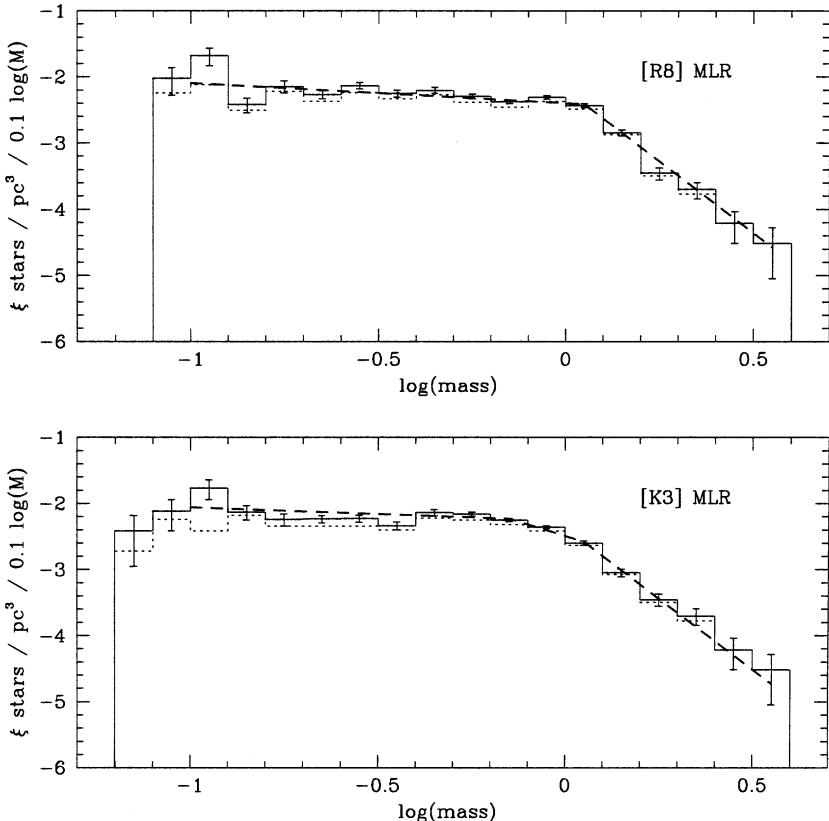


Figure 9.9. Two estimates of the present-day mass function for field stars. The solid histogram in the upper panel plots $\xi^{\text{PD}}(M)$ derived from the [R8] nearby-star sample using the empirical MLR cited in that paper; the lower panel shows the results of applying the [K3] semi-empirical mass–luminosity relation to the same data set. In both cases, the dotted histograms plot results for single stars and primaries in multiple systems, and the dashed lines show the best-fit power law relations.

and

$$\log(\xi^{\text{PD}}(\log M)) = -(4.3 \pm 0.3) \log M - (2.3 \pm 0.15) \quad M > 1M_\odot \quad (9.38)$$

These present-day mass functions are expressed in units of stars per $0.1 \log M_\odot$ per cubic parsec.

9.5.3 The substellar mass function in the field

Deriving the likely form of the mass function below the hydrogen-burning limit is more complicated. Brown dwarfs cool and fade on timescales that are rapid by astronomical standards, and which are mass dependent, with low-mass brown dwarfs evolving at faster rates along essentially the same cooling tracks as their higher mass counterparts. Moreover, while it may eventually prove possible to

calibrate the gravity dependence of certain spectroscopic features, and hence assign mass estimates to individual brown dwarfs, at present the lithium test is the only viable means of crudely segregating low and high-mass brown dwarfs.

Under such circumstances, one cannot construct a mass function either through object-by-object analysis of a volume-complete sample or by simple transformation of the luminosity function. Rather, the approach that must be adopted is a variant of the inverse analysis method outlined by [K3]. The observed luminosity function (e.g., $\Phi(M_J)$) serves as the observational constraint; assuming a particular mass function, $\Psi(M)$, and a formation history, or birth rate $B(\tau)$, we generate a model population of brown dwarfs with known age and mass; theoretical tracks are then applied to deduce the luminosity distribution, and that distribution is matched against the observed luminosity function.

This method was first used by [R9] in their analysis of the L dwarf surface densities derived from the initial 2MASS surveys. Assuming a constant birth rate function, they concluded that the observations available at that time were broadly consistent with a power law mass function with index, $1 < \alpha < 1.5$ – potentially a simple extension of the low-mass star relation into the substellar regime. Most significantly, this analysis provided definitive evidence that the substellar mass function is much flatter than the Salpeter function, and, as a result, there can be no question of brown dwarfs contributing significant dark matter.

Subsequent investigations [A6], [B6] have extended the analysis to include data from more recent L dwarf surveys [C4] and initial estimates of the T dwarf luminosity function [B7]. In general, these models find that the results are relatively insensitive to the assumed birth rate – at least for the currently observed range of spectral types. Assuming a constant star formation rate, Figure 9.10 shows the predicted age and mass distribution of ultracool dwarfs as a function of spectral type. As one might expect, given the relative lifetimes, low-mass stars and higher mass brown dwarfs are the predominant population, even among early-type T dwarfs; the space densities of L0-T4 dwarfs therefore only provide weak constraints on $\Psi(M)$.

Figure 9.11 compares the observed luminosity function, $\Phi(M_J)$, for ultracool dwarfs against theoretical predictions, based on power law representations of the substellar mass function. In interpreting this figure, it is important to bear in mind the incompleteness of current T dwarf surveys; nonetheless, it seems likely that $\Psi(M)$ flattens (i.e., $\alpha < 1$) in the substellar regime (Kroupa favours $\alpha = 0.3$, [K7]). Improved statistics for T dwarfs and, particularly, measurement of the space densities of field dwarfs at lower luminosities (reaching the Y dwarf regime) are required for a more definitive answer; however, analyses of young clusters and associations offer an alternative means of probing the substellar regime, as described further in Section 9.7.

9.5.4 The initial mass function of the Galactic Disk

A number of adjustments need to be applied to the present-day mass function

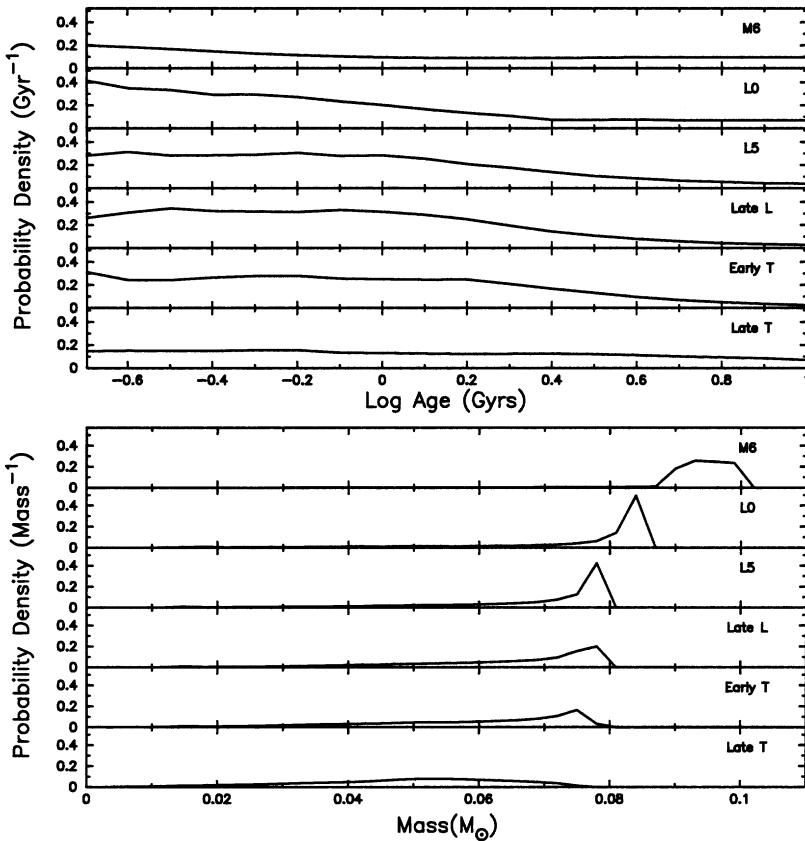


Figure 9.10. Predicted age and mass distributions for ultracool dwarfs as a function of spectral type; the models assume a constant birth rate and are based on the [B8] evolutionary tracks (from [A6], courtesy of the *Astrophysical Journal*).

(PDMF) to transform it into an estimate of the initial mass function (IMF). There are three major corrections to the stellar PDMF (see Section 9.2.3 and [R8]):

- First, intermediate and high-mass stars have lifetimes shorter than the age of the Galactic Disk, and the PDMF must be adjusted to allow for stars that have evolved off the main sequence. The lifetimes (in Gyr) can be estimated using the following relation (from [S11]):

$$\log \tau_{\text{MS}} = 1.015 - 3.491 \log M + 0.8157 (\log M)^2 \quad (9.39)$$

- Second, the density distribution of intermediate and high-mass stars has a smaller scale height perpendicular to the plane than sub-solar mass stars, since the higher mass stars are, on average, much younger (see Section 7.5.2). As a result, the Solar Neighbourhood includes a higher proportion of the present-day

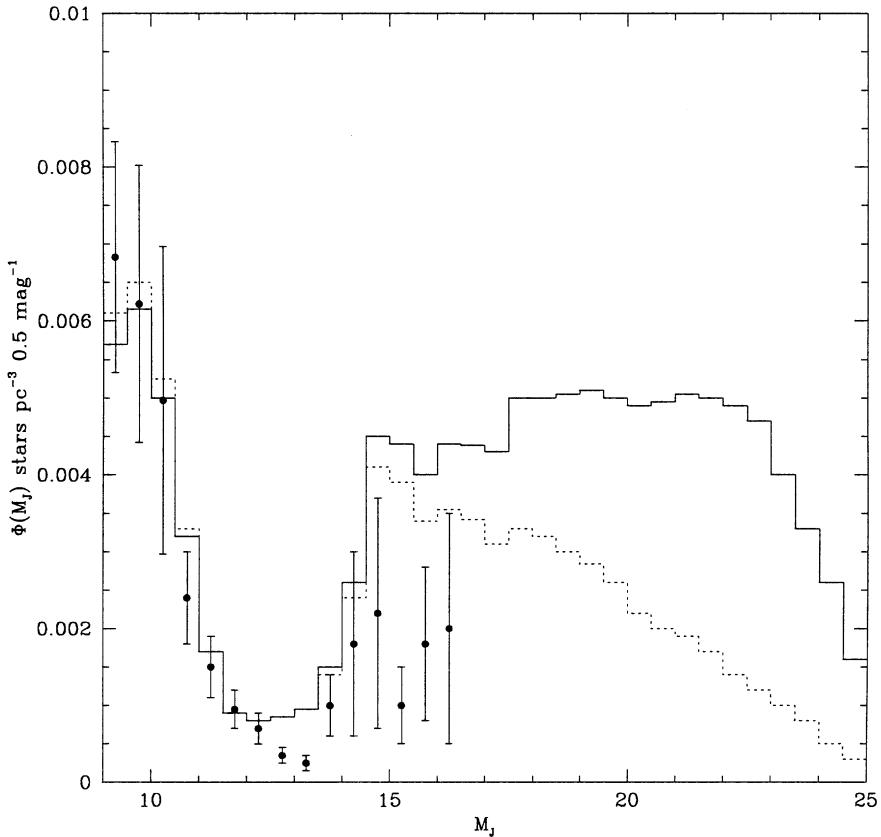


Figure 9.11. A comparison between the observed J -band luminosity function for ultracool dwarfs (solid points, from [C3] and [B7]) and theoretical predictions based on a constant birth rate and a power law mass function ($M < 0.1 M_{\odot}$) with $\alpha = 0.8$ (solid line) or $\alpha = 0.4$ (dotted line). The current T dwarf luminosity function is based on an incomplete sample, but the observed number densities may favour the flatter mass function (from [A6], courtesy of the *Astrophysical Journal*).

total population. This bias is corrected by applying the appropriate multiplicative scale factor to the local densities.

- Third, not all Solar Neighbourhood stars are members of the disk population; in particular, approximately 10% of stars later than spectral type \sim F5 are members of the thick disk. The space densities in the PDMF need to be adjusted accordingly.

As with the analysis of the substellar IMF discussed in the previous section, step 1 requires that one adopt an estimate of the stellar birth rate. Lacking strong evidence to the contrary, a constant star formation rate over the lifetime of the Galactic Disk is generally assumed.

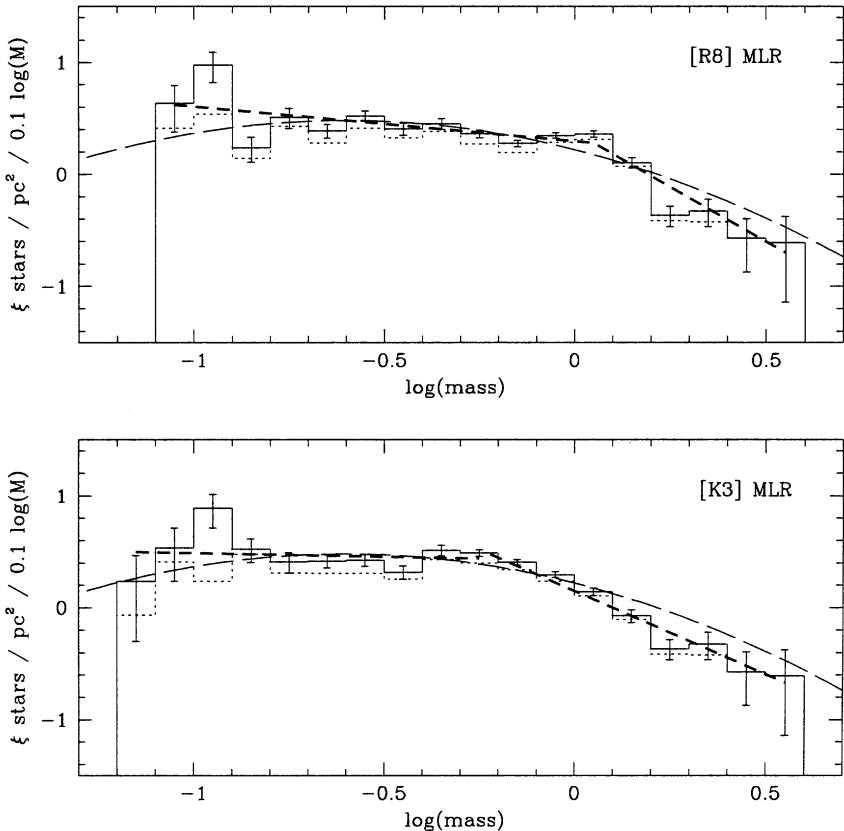


Figure 9.12. The initial mass function in the field: as in Figure 9.9, the upper panel plots results based on the [R8] empirical mass–luminosity calibration, while the lower uses the [K3] semi-empirical relation; the solid and dotted histograms have the same meaning as in Figure 9.9. In both cases, the stellar regime can be represented as a combination of two power laws, with $\alpha \sim 1.3$ at low masses and $\alpha \sim 2.9$ at higher masses (short-dashed lines). The long-dashed line is the log-normal mass function favoured by Chabrier [C6].

Figure 9.12 shows the results of applying these corrections to the nearby star PDMF; we show the derived IMF for both mass–luminosity relations represented in Figure 9.9. For the moment, we limit discussion to the stellar regime. Under this analysis (from [R8]), both MLR calibrations give an IMF that can be represented as a two-component power law, with $\alpha \sim 1.3 \pm 0.1$ at low masses ($M < 0.7 M_\odot$ [K3], [K7] or $M < 1.0 M_\odot$ [R8]) and a near-Salpeter index at higher masses. Expressed as a star formation rate, the [R8] calibration gives:

$$\xi^i(\log M) = -0.3 \log M - (0.35 - \log T_D) \text{ stars pc}^{-2} (0.1 \log M)^{-1} \text{ Gyr}^{-1} \quad M < 1 M_\odot \quad (9.40)$$

and

$$\xi^i(\log M) = -1.9 \log M$$

$$- (0.35 - \log T_D) \text{ stars pc}^{-2} (0.1 \log M)^{-1} \text{ Gyr}^{-1} \quad M > 1 M_\odot \quad (9.41)$$

where T_D is the age of the disk (in Gyr).

While we have chosen to characterise the IMF as a combination of two power laws, we recognise that it is likely that this is simply a mathematically convenient means of representing the data, rather than a reflection of any underlying physical process. As discussed in the previous section, current data suggest that the IMF flattens at substellar masses (see also [K7] and the following sections). This suggests that the overall form of $\Psi(M)$ may be closer to the functional form proposed originally by Miller and Scalo [M7], although a pure log-normal function cannot match the field star IMF at intermediate and high masses ($M > 1 M_\odot$). Chabrier [C5] has proposed a specific log-normal relation for the IMF at low masses:

$$\xi^i(\log M) = 0.076 \exp(-(\log M - \log(M_0))^2/(2\sigma^2)) \quad (9.42)$$

with $M_0 = 0.25 M_\odot$ and $\sigma = 0.55$. This gives a reasonable match to the observed function at low masses ($M < 1 M_\odot$), including the turnover at substellar masses; however, as Figure 9.12 shows (see also [C5]), this formalism still fails to match observations at intermediate and high masses.

It is possible that the IMF is a composite function, matching a log-normal distribution at low masses and a Salpeter-slope power law at higher masses. However, we should bear in mind that, even though this type of prescription may provide a good representation of the empirical results, without understanding the underlying physical processes, these mathematical relations are merely ‘Just so’ stories, telling us what the IMF looks like, not *why* it looks that way. We will return to this issue in Section 9.8; for the moment we note that the power law representations outlined above (or the [C5] log-normal distribution at low masses) provide a handy means of comparing observations of different star-forming environments, both within the Milky Way and in external galaxies.

9.6 MASS FUNCTIONS OF OPEN CLUSTERS

Reliable mass functions can be derived only for open clusters with reliable luminosity functions. For detailed surveys of low-mass stars, the cluster must be near enough to allow observations to extend to close to the hydrogen-burning limit, and sufficiently young that mass segregation and evaporation have not modified the cluster mass function significantly from its original form. The only cluster that meets these criteria and has received the necessary observational attention is the Pleiades, although both the Hyades and, to a lesser extent, Praesepe, offer some insight into dynamical evolution in clusters.

9.6.1 The Hyades, Praesepe and older open clusters

Observations of the Hyades and Praesepe show that both clusters exhibit strong mass segregation, as described in Sections 8.9.1 and 8.9.3. This dynamical evolution has almost certainly led to extensive evaporation of cluster members, and, inevitably, preferential depletion of low-mass members. The mass function of the present-day Hyades is significantly flatter than the field function; fitting the distribution with a power law gives $\alpha = 0.5$ for $1 > M/M_{\odot} > 0.1$ [R6], much flatter than the field function. There are no isolated brown dwarf members known in either cluster.

Initial observations of Praesepe suggested that the mass function was steeper, with $\alpha \sim 1.0$ to 1.5 [H2], [W6], [P1]. However, those analyses were based on photometrically selected candidates from an area of only ~ 1 square degree in the cluster centre, with no follow-up spectroscopic confirmation. Subsequent investigations, based on more extensive samples with the appropriate ancillary membership confirmation, notably the 2MASS survey [A7], find that the Praesepe mass function can be modelled as a power law with $\alpha \sim 1.5$ for $M > 0.4 M_{\odot}$, but flattens sharply at lower masses, with $\alpha \sim 0$ at $0.1 M_{\odot}$. Again, this differs radically from the PDMF in the field.

The bottom line is that clusters older than 600 Myr have been subject to such extensive dynamical evolution that it is essentially impossible to reconstruct the initial mass function for stars less massive than the current turn-off ($\sim 1.2 M_{\odot}$).

9.6.2 The Pleiades and intermediate-age open clusters

With an age of ~ 125 Myr and a distance of ~ 130 parsecs, the Pleiades is an ideal laboratory for investigating the mass function. Observations extend below the hydrogen-burning limit, and the cluster is sufficiently young that mass segregation is still relatively limited in scope. The cluster is sufficiently old, however, that stars with masses exceeding $\sim 0.4 M_{\odot}$ have contracted onto the main sequence, and, as a consequence, their masses can be estimated using field-star mass-luminosity relations (Figures 9.5 and 9.6). Lower mass stars are still over-luminous compared with the field, but the offset is relatively small. Figure 9.13 compares theoretical (mass, M_V) relations from the Lyon models [B1] for ages of 10^8 , 10^9 and 10^{10} years. There are only marginal differences between the two old isochrones, as we noted in Section 9.4.1, and the 100-Myr isochrone is offset by only a factor of 2 in luminosity at $0.1 M_{\odot}$. On the other hand, the 100-Myr models place the hydrogen-burning limit at $M_V \sim 15$, or over 5 magnitudes brighter than the 1-Gyr isochrone. This underlines the advantages that even intermediate-age clusters hold in probing the brown dwarf régime.

There is no single comprehensive Pleiades survey that spans the full mass range of cluster members. Figure 9.14, however, combines results from five sources: Meusinger *et al.*'s [M5] proper motion survey of the higher mass stars, complete to $M_V \sim 12$; Stauffer *et al.*'s [S6] survey, which includes fainter M dwarfs; Hambly *et al.*'s [H1], [H11] deep proper motion survey of the inner cluster regions, extending

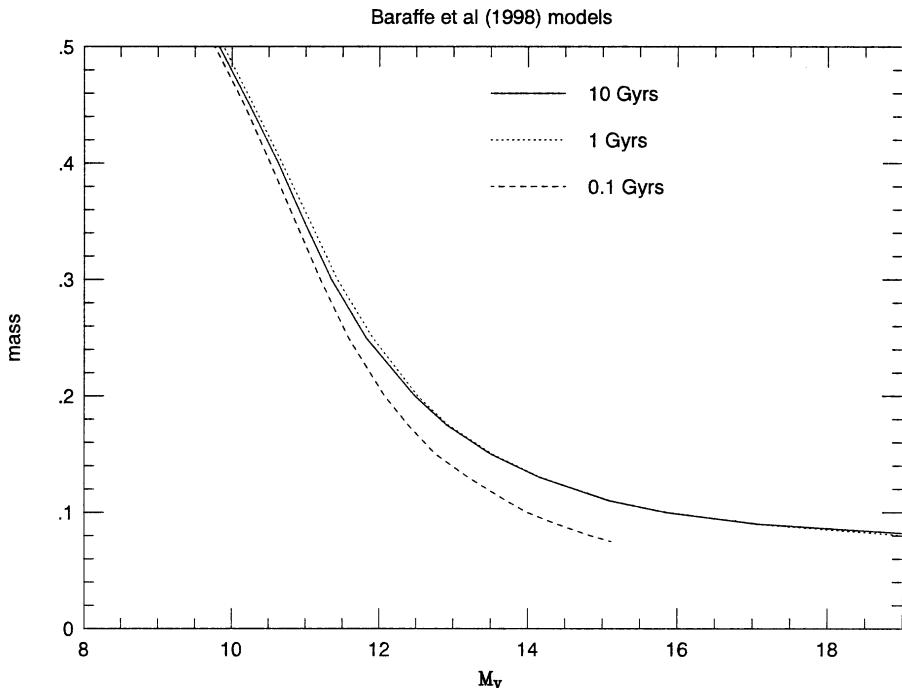


Figure 9.13. Theoretical (mass, M_V) relations for solar-abundance dwarfs with ages 0.1, 1 and 10 Gyr; as discussed in the text, the latter two relations are almost indistinguishable.

below the hydrogen-burning limit; and the 2MASS-based all-cluster survey by Adams *et al.* [A7]. Finally, we include an estimate of the space density of high mass ($M > 0.05 M_\odot$) brown dwarfs from Bouvier *et al.*'s observations [B9].

There is sufficient overlap between the surveys (at least at stellar masses) to scale them all to a common zeropoint. Once scaled, the overall form of $\xi(\log M)$ is broadly consistent with the field mass function. At high masses ($M > 1.2 M_\odot$), the distribution is steep, consistent with a power law, $\alpha \sim 3$ (note that the turn-off mass of the Pleiades cluster is $\sim 4 M_\odot$). At lower masses, $\xi(\log M)$ is flat, consistent with a power law index, $\alpha \sim 1$, with some indications of a turnover below $\sim 0.2 M_\odot$. Bouvier *et al.* [B9] derive a somewhat flatter mass function at low masses, with $\alpha \sim 0.6$, although their data set is not as extensive as the 2MASS Pleiades survey.

The turnover at low masses has been interpreted as evidence that the mass function has a log-normal distribution [H11], [A7], although, as with the field IMF, the formalism proposed by Chabrier [C6] is not a good match to the data at high masses and may over-predict brown dwarf number densities (see Figure 9.14). On the other hand, Moraux *et al.* [M9] have derived masses using slightly different mass-luminosity relations, and they find closer agreement with a log-normal relation with $M_0 = 0.25 M_\odot$ and $\sigma = 0.52$, close to the [C6] prescription. It remains possible that the relatively steep turnover at low masses may be due to preferential evaporation of low-mass cluster members.

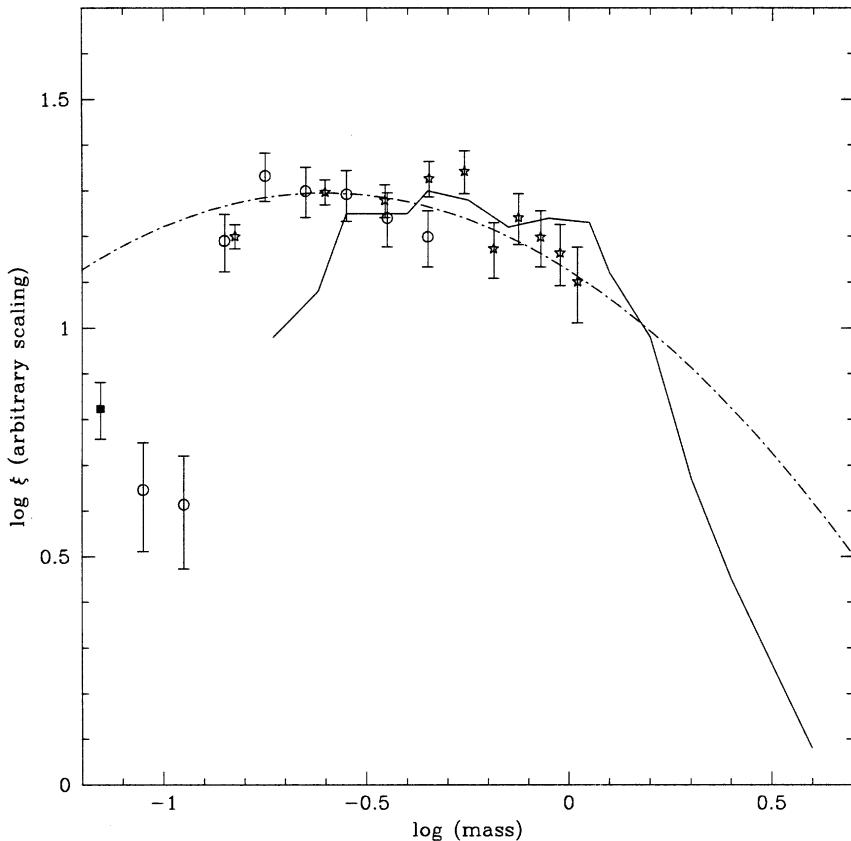


Figure 9.14. The Pleiades mass function, derived by combining results from the Meusinger *et al.* [M5] survey of the brighter cluster members (solid line); the [R5] compendium of K and M dwarfs from [S6], [H1], [H11] (open circles – note that this sample is incomplete in the highest mass bin); and the [A7] analysis of the 2MASS-selected data set (stars). The solid square is the high-mass brown dwarf density estimate from [B9]. The dash-dot line plots a scaled version of the Chabrier field IMF [C5].

Only a handful of other intermediate-age clusters have sufficient observations to allow reliable derivation of the mass function below $\sim 0.8 M_{\odot}$. Barrado y Navuesces *et al.* [B10] have used their survey of the central regions of α Persei to compute $\Psi(M)$ for masses between $\sim 0.7 M_{\odot}$ and $\sim 0.04 M_{\odot}$. They derive a mass function that can be represented as a power law, with $\alpha = 0.6$, with a possible dip from ~ 0.07 to $0.05 M_{\odot}$. This is suspiciously close to the hydrogen-burning limit, and may be tied more to the adopted mass-luminosity calibration (from the [B1] models) than to an actual feature in the PDMF. In broad terms, the results are not inconsistent with the Pleiades (and field) mass functions, but it would be extremely interesting to extend this analysis to lower masses and a larger fraction of the cluster.

9.7 PROTOCLUSTERS AND STAR-FORMING REGIONS

Surveys of clusters younger than ~ 20 Myr offer both advantages and disadvantages for studies of $\Psi(M)$ at subsolar masses (see [M6]). On the positive side, low-mass stars are at their most luminous during this phase of pre-main sequence contraction (Section 3.4.3). A $0.3 M_{\odot}$ star has a luminosity of $\sim 0.1 L_{\odot}$ at age 10 Myr, as compared with a luminosity of $\sim 0.013 L_{\odot}$ at 100 Myr, or $\sim 0.008 L_{\odot}$ on the main sequence. These objects can therefore be identified at substantially greater distances from the Sun than is possible for their older counterparts in Pleiades-like clusters.

Young clusters offer the further advantage of being largely free of the effects of dynamical evolution. Some degree of mass segregation exists in even the youngest systems, with the highest-mass stars often more centrally concentrated than the average cluster member. Given typical crossing times of between a few $\times 10^5$ – 10^6 yr, this concentration is more likely to be a result of the conditions prevailing during the earliest stages of star formation rather than a consequence of dynamical relaxation. Solar-type and lower-mass stars appear to follow nearly identical radial density distributions (as in the Pleiades), thereby minimising potential biases due to incomplete areal coverage.

Youth, however, also carries its disadvantages. Since the cluster members are still in the pre-main sequence contraction stage, and no empirical mass estimates are yet available (Section 9.4.2), theoretical calculations offer the only method of estimating a mass–luminosity relationship. Applying these calibrations demands reliable age estimates, which are complicated by the fact that clusters do not form instantaneously: that is, cluster members span a range of ages. The resultant uncertainty in individual ages leads to corresponding uncertainty in mass, which is highest at these young ages when stars evolve so rapidly.

Further complications can arise from the presence of circumstellar disks in some systems. The material in those disks usually dominates the energy distribution in the mid- and far-infrared, and, given sufficiently high temperatures, may even make a significant contribution to the observed flux at shorter wavelengths, leading to an overestimate of both the photospheric bolometric luminosity and the mass. Finally, and most importantly, protoclusters are still embedded within the remains of the parent molecular cloud. The highly-variable obscuration caused by dust within that cloud is a serious impediment to an accurate interpretation of observations, at least at optical wavelengths.

Dust obscuration amounts to a loss of a few magnitudes at visual wavelengths for a typical few-Myr-old star cluster (such as IC 348, [L2]), and can reach levels of $A_V \sim 20$ – 50 magnitudes, or more, in the denser regions of embedded protoclusters such as NGC 2024 (in Orion) or ρ Ophiuchi. The absorption is variable on scales of $\sim 1,000$ – $10,000$ AU, or ~ 1 – 10 arcsec for the nearer clusters. Under such circumstances, optical surveys are capable of identifying only the most luminous and least obscured cluster members, and are therefore poorly suited to providing catalogues for statistical analysis. Fortunately, the scattering properties of interstellar dust lead to significantly less absorption at longer wavelengths. In particular, the absorption at $2.2 \mu\text{m}$, A_K , is almost a factor of 10 less than that at $0.5 \mu\text{m}$. Typical

cloud temperatures are between 50 and 100 K, so dust *emission* peaks at $\lambda \sim 60 \mu\text{m}$, but is negligible at near-infrared wavelengths.

Given these circumstances, star-formation regions have long been recognised as interesting targets for infrared observations. However, while even the earliest scans led to notable discoveries – such as massive protostars like the Becklin–Neugebauer object and the surrounding Kleinmann–Low nebula in Orion [B2], [K1] – initial surveys were limited to either bright sources or small solid angles, and often both. It is only with the development in the mid-1990s of large-format infrared arrays and high-sensitivity spectrographs that it has become possible to undertake studies capable of detecting protostars with masses below $0.1 M_\odot$ over the entire area of major star-formation regions. As a result, the full potential of infrared studies remains to be realised. Nonetheless, preliminary results are intriguing.

In general, investigations of the stellar mass function in young, star-forming regions have followed two broadly complementary lines of attack: statistical analysis of deep star-counts, and more detailed source-by-source analyses. Those methods are described in more detail in the following two sections, and Section 9.7.3 uses observations of the stellar-mass members of IC 348 and the Orion Nebula Cluster (ONC) as practical examples of their application. The last few years have seen detailed surveys of a number of young clusters and associations, with particular attention focused on the substellar-mass members. Section 9.7.4 summarises the main results from those studies, and considers the overall implications for the initial mass function.

9.7.1 Infrared imaging surveys

Near-infrared number-magnitude counts offer a straightforward means of probing the stellar content in obscured, star-forming regions. The luminosity function of cluster members can be determined statistically by comparing source counts centred on the cluster against counts made within nearby, off-cluster fields. If the cluster distance is known, the apparent luminosity function $\phi(m_K)$ can be transformed to $\Phi(M_K)$. Since this technique is based on direct imaging, it offers the possibility of obtaining a complete census of even the lowest-luminosity cluster members through a series of simple and efficient observations. With the current generation of infrared detectors, data can be obtained covering an entire cluster in a matter of only a few nights on an intermediate-sized telescope.

There are, however, complications in deriving a mass function from the resultant K -band luminosity function:

- The (mass, M_K) relationship must be appropriate for the age of the cluster, τ_C , and therefore requires both an accurate estimate of τ_C and reliable pre-main sequence mass-luminosity relationships.
- The cluster stars are unlikely to be exactly coeval. Rather than the cluster forming in a single burst at time $T = T_0 - \tau_C$, where T_0 is the present time, individual stars span a range of ages, $\tau_i = \tau_C \pm \Delta\tau$.
- While working in the near-infrared minimises the effects of obscuration, differential reddening (either foreground or within the cluster itself) is likely to be

present at the 0.1–1.0 magnitude level in A_K . Moreover, there may be significant differences between the total obscuration in cluster and off-cluster fields: dust within the young cluster usually leads to higher reddening of background stars. This can produce systematic errors in $\phi(m_K)$.

- Emission from circumstellar disks can contribute significantly to the flux in the thermal infrared ($\lambda > 2 \mu\text{m}$) in young protostars.
- Source counts alone cannot distinguish single and multiple stars.
- All of these effects combine to make it difficult to distinguish cluster members from field stars.

Many of these problems can be addressed: optical and near-infrared colours can be used to map differential reddening, while excess radiation at longer wavelengths (above the predicted photospheric flux) measures possible contributions from hot circumstellar dust.

The usual technique is to compute the expected K -band luminosity function (KLF) based on an estimated initial mass function, age and star-formation history. The last two parameters can be determined to some extent from photometric and/or spectroscopic data, although a degree of guess-work is also often required. The predicted luminosity function is matched against the observed KLF (in the observational plane), and the input parameters adjusted until reasonable concordance is achieved. Initial analyses [Z2] were based on single-burst star-formation models, but more recent studies [L2], [L3] have adopted more complex (and more realistic) star-forming histories. In general, this approach is effective at ruling out inappropriate models, but only identifies consistent (rather than unique) solutions.

9.7.2 H–R diagram analyses

An alternative method of studying young clusters is to use spectroscopic and photometric observations to estimate bolometric luminosities and effective temperatures for each cluster member. Given these data, each star can be placed on the H–R diagram, making due allowance for foreground reddening and circumstellar dust emission, while eliminating foreground and background field stars. Comparison with pre-main sequence evolutionary tracks permits the estimation of masses and ages on a star-by-star basis (see Figure 9.8), whilst the mass function and star-formation history follow from summation of the individual results.

In principle, this approach offers higher precision than the statistical KLF analysis. There are, however, significant practical obstacles – notably in obtaining spectroscopic data of the requisite accuracy for the faintest, and most highly obscured, cluster members. Simultaneous observations of tens of candidate cluster members, using optical or near-infrared multi-object spectrographs can go some way towards addressing the latter problem, but photon scarcity limits the full-scale application of this method.

9.7.3 Two case studies: IC 348 and the Orion Nebula Cluster

To illustrate the relative merits of these two types of analysis we consider their application to IC 348 and the Orion Nebula Cluster (the Trapezium), two well-

studied, young star-forming regions. Observations of these clusters demonstrate the complementary nature of the techniques.

IC 348

This cluster lies at a distance of ~ 320 pc ($(m-M)_0 = 7.5$ magnitudes) in the Perseus Molecular Cloud. Originally catalogued by Dreyer, the cluster has 40–50 members detectable at optical wavelengths, and the brightest star α Persei (type B5) is surrounded by a prominent reflection nebula (Figure 9.15). Herbig's [H8] spectroscopic observations of strong H α emission in 16 stars within this region confirmed the presence of a young cluster. Infrared scans [S8] reveal a moderate to large population of optically-invisible sources, embedded in what is presumably the remnants of the parent molecular cloud [K6]. The general characteristics point to a young age for the system – initially estimated as between 5 and 20 Myr [S8]. More recent studies [H9], [L6], [L7] find that most of the stars have ages between 1 and 5 Myr.

Lada and Lada [L2] surveyed the central 0.1 square degree of this cluster, reaching limiting magnitudes of 16.5, 15.5 and 14.5 in the J , H and K passbands. These observations cover most of the cluster core, but not the outer regions; the cluster has a diameter of ~ 30 arcmin (~ 3 parsecs). Of the 600+ sources detected in the [L2] survey, most are concentrated in a relatively small region south of α Persei, where the source density rises to ~ 250 stars pc $^{-2}$. The contamination from background and foreground stars is estimated through observations of control fields, offset by 1° from the cluster centre. The resulting K -band luminosity function (Figure 9.16) has a broad peak at $M_K \sim 4.5$, or $K = 11.5$, well above the limiting magnitude of the survey. As in studies of other young clusters, the brighter source

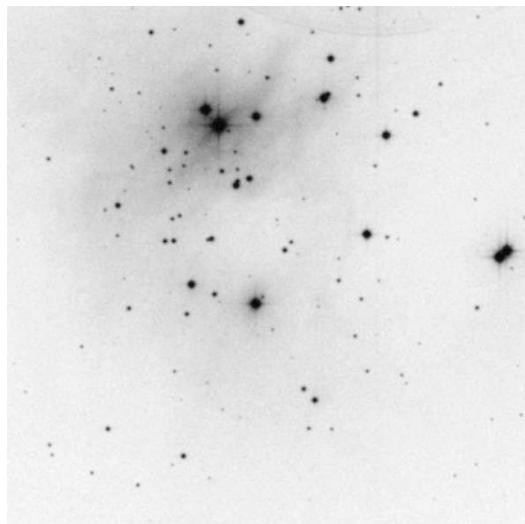


Figure 9.15. An R-band image of the cluster IC 348. The field is 10 arcmin square. (Courtesy of Palomar Observatory/STScI.)

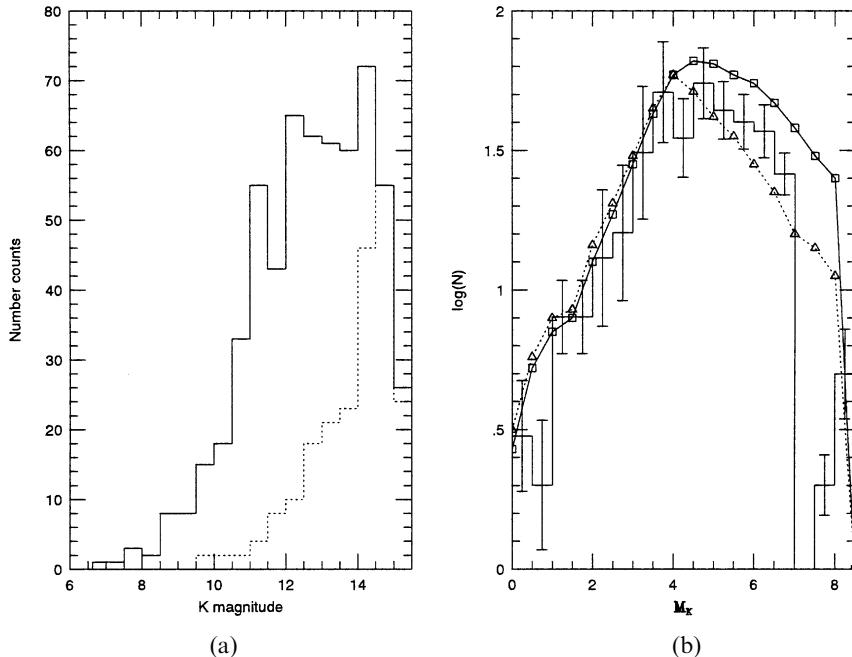


Figure 9.16. (a) K -band source counts towards IC 348. The dotted histogram shows the likely contribution from background/foreground sources [L2]; (b) the K -band luminosity function (histogram) matched against predictions based on a log-normal mass function (open triangles) and a truncated mass function (open squares – from [L3]).

counts follow a near-power law distribution with magnitude, index 0.38 at K and 0.32 at J . This is much steeper than those measured for either intermediate-age open clusters or field stars at similar magnitudes and reflects a compression of the mass–luminosity relationship: a smaller interval in L spans a larger range in M .

Rather than attempting a direct transformation of $\Phi(M_K)$ to $\Psi(M)$, [L2] and [L3] invert the analysis, using computer simulations to predict $\Phi(M_K)$ given $\Psi(M)$ and a star-formation history. Starting with a Miller–Scalo log-normal mass function, pre-main sequence tracks from [D2] are used to estimate (L, T_{eff}) as a function of (mass, age). Transforming to M_K via temperature-dependent bolometric corrections, the resulting KLFs are compared to the observed luminosity function. The star-count data match models where star formation has progressed at a relatively uniform rate over the last 5–10 Myr [L4]; single-burst (coeval) star-formation models produce luminosity functions which are either too narrow or fail to match the observed slope of the source counts at bright magnitudes. Recent spectroscopic analysis of cluster members [L6], [L7] has refined these timescales: as indicated above, the majority of the cluster stars prove to be younger than ~ 3 Myr. Results favour a mass function matching the Miller–Scalo formulation to $\sim 0.2 M_\odot$, but are better represented by a power law $0.3 < \alpha < 0.8$ at lower masses.

The Orion Nebula Cluster

This is the young star cluster surrounding θ^1 Ori – the four O-type stars of the Trapezium that ionise the Orion nebula. Lying at a distance of 470 ± 70 pc [G1], the Orion Nebula Cluster (ONC) is one of the most prominent of a series of active star-forming regions in the Orion OB association [G2]. Immediately behind the cluster lies the L1640 molecular cloud whose very substantial density leads to extinctions of $A_V > 80$ magnitudes, effectively eliminating any contribution by background stars to even near-infrared source counts. The ONC itself is ~ 3 parsecs in diameter, and lies primarily within the low gas-density region excavated by winds from the Trapezium stars (the overall morphology of the area is described by Zuckerman [Z3]). Proper-motion studies [J1] show that the overwhelming majority of stars within ~ 20 arcmin of the Trapezium are cluster members, simplifying source-count analysis. The star density rises to $\sim 20,000$ stars pc^{-3} in the central ~ 0.3 parsecs.

The Orion complex has been the subject of numerous spectroscopic and photometric investigations of both the stellar and gaseous content (see [G2] and [H10] for summaries), including recent high-resolution Hubble Space Telescope imaging [P5] which provides direct observations of the ‘silhouettes’ of circumstellar disks (‘proplyds’ – Figure 3.19, in colour section) around a number of cluster stars. The cluster was an early target of infrared observations [A4]; the relatively narrow KLF has been modelled as a single burst of star formation, aged 10^6 years, with a Miller–Scalo mass function [L2].

The most detailed study of the ONC, however, was undertaken at optical wavelengths. Hillenbrand [H10] compiled V, I photometry for 1,600 of the 3,500 sources identified within the central 5 pc of the cluster, as well as spectroscopy for more than 980 of the brighter stars. While the photometric sample includes only 40% of the complete population, the spatial distribution of the observed sources is similar to the complete sample, and there is no evidence for significant differences between the constituent stars in the two datasets. Thus, the optically-selected sample is probably a fair subset of the population in the ONC cluster.

Hillenbrand’s estimate of the ONC mass function is derived using the star-by-star approach. Effective temperatures are estimated from the spectral types, while M_{bol} is derived from the photometry allowing each star to be placed on the H–R diagram. Comparison with theoretical evolutionary tracks (Figure 9.17) give the age and mass of each star; and the mass function and star-formation history follow by combining those individual results. As in the star-count analyses, age and mass calibrations rest entirely upon the theoretical models, and choosing a different set of models, or different transformations from the observational to theoretical plane, can lead to different conclusions. Restricting analysis to the lower main sequence stars plotted in Figure 9.17, comparison with either the [D2] or the more recent [D6] tracks indicates that the average age of ONC members is somewhat less than 1 Myr, but with an overall spread in age of at least 2 Myr.

Initial analysis of these data led to the derivation of a log-normal mass function, peaking at $\sim 0.3 M_\odot$, similar to the results for IC 348 shown in Figure 9.16. There

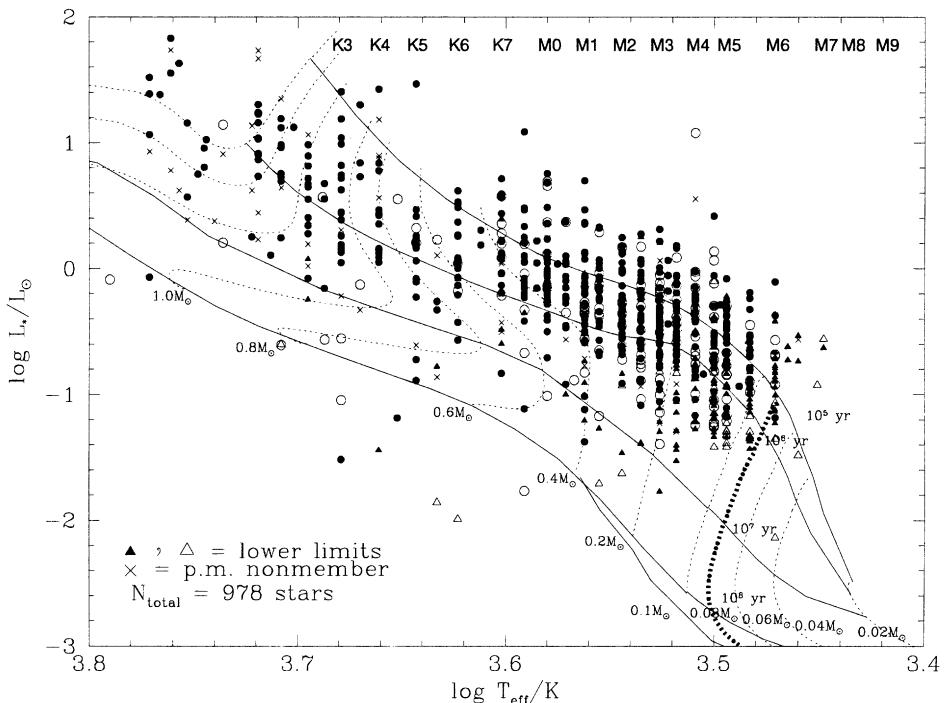


Figure 9.17. The H–R diagram for lower-luminosity stars in the Orion Nebula Cluster. Filled symbols indicate stars with spectroscopic observations; open symbols have only photometric data (From [H10], courtesy of L. Hillenbrand and the *Astronomical Journal*.)

was also some evidence for a dip near $\sim 1 M_\odot$. As discussed further below, more recent analyses, using improved bolometric corrections and a revised effective temperature scale, indicate that $\xi(\log M)$ continues to rise through $0.1 M_\odot$, with a slope close to a power law, $\alpha \sim 1.8$.

Summary and addenda

The two studies outlined above highlight the relative merits of the photometric and spectroscopic routes to the mass function: infrared imaging provides a straightforward means of covering large areas and a statistically well-defined sample of candidate cluster members, but ambiguous membership criteria and age/mass estimates; the H–R diagram analysis (including spectroscopy) supplies more insight into star formation history, and more reliable membership criteria and masses for individual objects, but requires significantly more resources (i.e., telescope time). Fortunately, instrumental advances over the last few years, particularly the development of multi-object near-infrared spectrographs, have made it much easier to obtain detailed follow-up observations of individual candidates, and most recent studies combine both approaches. Thus, Muench *et al.* [M10] have used wide-field infrared imaging to survey a 20.5×20.5 arcmin field centred

on IC 348 to $K \sim 18.5$, three magnitudes fainter than the [L2] survey, while Luhman *et al.* [L9] combined I and z-band imaging with optical and near-infrared spectroscopy to verify cluster membership for the [M10] photometric candidates. Similarly, there have been deeper, wider-field surveys of the ONC [H13, M11]. We consider results from these more recent studies further in the following section.

9.7.4 The IMF in young clusters and associations

Over the last five years, many of the star-forming regions within 500 parsecs of the Sun (see Figure 7.23) have been targets of surveys for low-mass star and brown dwarf members. In most cases, those programs combine deep far-red/near-infrared imaging with follow-up spectroscopy (optical or near-infrared) to identify reliable catalogues of cluster members; however, a few surveys rest either on only near-infrared photometry, or on photometry coupled with low signal-to-noise spectra, and the membership catalogues from those surveys are likely to include greater contamination from foreground and background field stars (and brown dwarfs). Here, we summarise the results for a representative subset of clusters, spanning a range of star-forming environments (see Figure 9.18).

IC 348. This cluster has been well observed in recent years, as described above. While there is significant absorption behind the cluster, it is not sufficient to eliminate all background sources, and both foreground and background objects contaminate photometric samples. However, with an age of 2–3 Myr, late-type cluster members are easily recognised through the presence of high levels of chromospheric activity and spectroscopic signatures of low gravity (see Section 2.2.8 and 6.7.4). There are almost 300 spectroscopically confirmed members, including at least 23 with spectral types between M6 and M9. Matched against theoretical models, those spectral types correspond to masses between 0.08 and $\sim 0.025 M_{\odot}$ [L9].

The Orion Nebula Cluster. The stellar (and brown dwarf) population in the core of the ONC has been surveyed extensively at near-infrared wavelengths over the last five years [H13], [M11], [L10], [L11], [S13]. The area covered by individual surveys ranges from ~ 2 arcmin square (0.3×0.3 pc) for the HST NICMOS survey [L10] to ~ 7 arcmin square (1×1 pc) for the [L11] ground-based survey. All data sets reach 18th magnitude at K , well below the hydrogen-burning limit; as a guide, a moderately reddened ($A_V \sim 5$ mag) $0.08 M_{\odot}$ dwarf has $K \sim 14.5$ and spectral type \sim M6, while a $0.02 M_{\odot}$ dwarf, subjected to similar reddening, has $K \sim 17$ and is an M9 or early L dwarf.

There is substantial overlap in the membership catalogues of the different surveys, which are all centred near the Trapezium; however, each investigation employs different analysis techniques. It is reassuring that there is broad agreement in the overall shape of the mass function at masses above $\sim 0.03 M_{\odot}$. [M10] have characterised the distribution as three power law segments, with $\alpha \sim 2.2$ for $M > 0.6 M_{\odot}$, $\alpha \sim 1.15$ for $0.6 M_{\odot} > M > 0.12 M_{\odot}$ and $\alpha \sim 0.25$ for $0.12 M_{\odot} > M > 0.025 M_{\odot}$. The power law indices, and the mass limits, for the

higher mass segments are close to the values listed for the field IMF in Equations (9.40) and (9.41) (see also [K7]). Other ONC surveys [H13], [L10] find that the near-Salpeter slope extends to lower masses and favour a functional form closer to a log-normal distribution, but concur in the flattening below $\sim 0.3 M_{\odot}$.

The results diverge more significantly at the lowest masses: both the [M10] and [L11] analyses find a steep rise in $\Psi(M)$ at $\sim 0.025 M_{\odot}$, followed by a sharp drop in densities at masses just below the deuterium-burning limit at $\sim 0.013 M_{\odot}$ (see Figure 9.18, ONC panel); the [L10] and [H13] studies, however, find no evidence for this behaviour. The first-mentioned investigations are based on photometric analyses (although see [L12]), comparable to the IC 348 case study outlined above; in contrast, the [L10] study uses spectroscopy and the H–R diagram analysis, while the [H13] analysis, although photometric, takes account of the likely age range in ONC stars. The recent spectroscopic observations by [S13] may have resolved this conflict: they find significant numbers of older (~ 10 Myr) M dwarfs in the core of the ONC; the masses of those M dwarfs are underestimated in the photometric surveys, which assume $\tau \sim 2$ Myr, leading to inflated numbers of very low-mass brown dwarfs.

Further studies of the ONC are currently underway, including a Legacy optical/infrared imaging survey with the Hubble Space Telescope that covers 450 square arcminutes of the cluster, or almost ten times the area surveyed by [M10]. In addition to probing the spatial distribution of the low-mass stars and brown dwarfs in the cluster, this project will also provide an unparalleled survey of the circumstellar environments across the cluster: proplyds, disks, jets and binary systems.

σ Orionis. This cluster is part of the Orion OB 1b association, which has constituents with ages between 1 and 7 Myr and an average distance of ~ 400 parsecs. The *σ Ori* cluster itself appears to have $\tau \sim 5$ Myr. The cluster was initially identified optically, as a concentration of B-type stars near the eponymous brightest member (spectral type O9.5), lying just below ζ Orionis, the western terminator of Orion’s belt. The extensive complement of lower mass stars was first recognised from ROSAT X-ray observations of Orion’s belt. [S12] surveyed the brighter cluster members, and estimate that there are approximately 160 members with masses between 0.2 and $1.0 M_{\odot}$ (based on the [B1] models), concentrated within a 3–5-pc radius. Extrapolating to lower masses using the [K3] IMF, they calculate a total mass of $\sim 225 M_{\odot}$ for the cluster. Crucially, there is little foreground reddening ($A_V < 0.5$ mag), rendering this cluster an inviting target for brown dwarf surveys; however, the presence of only moderate reddening behind the cluster means that field star contamination is a significant issue.

Several deep imaging surveys at far-red (I , z) and near-infrared (JHK) wavelengths have been undertaken by the Tenerife group, leading to the identification of ~ 150 candidate low-mass members [B13], [Z4], [B14]. Spectroscopic follow-up observations have been obtained for a subset of these sources, confirming many as cool dwarfs with spectral types later than M5, significant H α emission and spectroscopic signatures indicating low gravity. Among the likely members are several objects with spectra consistent with M9 or early L dwarfs; assuming a cluster age of ~ 4 Myr,

those objects are likely to have masses of less than $0.03 M_{\odot}$. The mass function is derived from these observations by comparing the photometric properties against the [B1] models. While these results nominally extend below the deuterium-burning limit, one should bear in mind that none of the (potentially) lowest mass objects are confirmed as members, and that these bins are the most susceptible to field-star (and field brown dwarf) contamination.

ρ Ophiuchus. This compact star-forming region lies at a distance of ~ 160 parsecs. The cluster includes 200–300 members, with a relatively low density of ~ 200 stars pc^{-3} in the core. Most of the members have optical and/or infrared spectroscopy, and a comparison with theoretical tracks (e.g., [B1], [D6]) indicates ages between 0.1 and 1 Myr (i.e., comparable with the ONC [L13]). The cluster has also been surveyed at mid-infrared wavelengths by ISOCAM on the Infrared Space Observatory (ISO), allowing the detection of circumstellar disk emission around many members [B12]. A number of substellar-mass candidates are known, although relatively few have spectroscopic observations.

λ Orionis. This cluster, originally catalogued in Ptolemy's Almagest, is part of the series of OB associations in a giant molecular cloud complex to the north of the Orion OB1 association. The complex lies at a distance of ~ 450 parsecs and appears to have been forming stars for the last ~ 7 Myr; winds from high-mass OB stars (and, perhaps, supernovae) started to disrupt the molecular cloud 1–2 Myr ago (see [D7] for a thorough discussion). *λ Orionis* itself is an O8 giant, and the surrounding stars form a loose cluster ~ 5 Myr old. The [D7] spectroscopic survey identifies over 250 pre-main sequence stars in the region, including stars with spectral types earlier than $\sim \text{M}4$ (masses ~ 0.3 – $0.4 M_{\odot}$). Barrado y Navascués and collaborators have extended the census to much lower masses with a deep *R*, *I* imaging survey covering 42×28 arcmin (4.9×3.25 pc) centred on *λ Orionis*, although, to date, only 20% of the candidates have spectroscopic data [B15]. As in most other studies, masses are derived by matching photometric properties against predictions from the [B1] models.

Taurus. Finally, the molecular clouds of the Taurus–Auriga association offer a radically different star-forming environment. Lying at a distance of ~ 140 parsecs, and stretched over ~ 100 square degrees on the sky, these clouds are the epitome of a low-density star-formation region, with only 1–10 stars pc^{-3} , $\sim 10^4$ times lower in density than Orion. Moreover, unlike the other clusters and associations described here, and despite an age of only 1–2 Myr, there are no OB stars; the earliest type stars are spectral type $\sim \text{G}0$, corresponding to masses of only 1.0–1.2 M_{\odot} . There are small pockets of pre-main sequence stars distributed throughout the $\sim 30 \times 10$ pc cloud complex [G7]. Star formation is proceeding under relatively quiescent conditions, without the disrupting effects of violent winds from massive OB stars and supernovae.

The sheer angular size of the Taurus–Auriga complex hampered surveys until recently. The 2MASS catalogue, however, provides a means of searching for very low mass stars and brown dwarfs; a $0.08 \cdot M_{\odot}$ dwarf member of Taurus has spectral

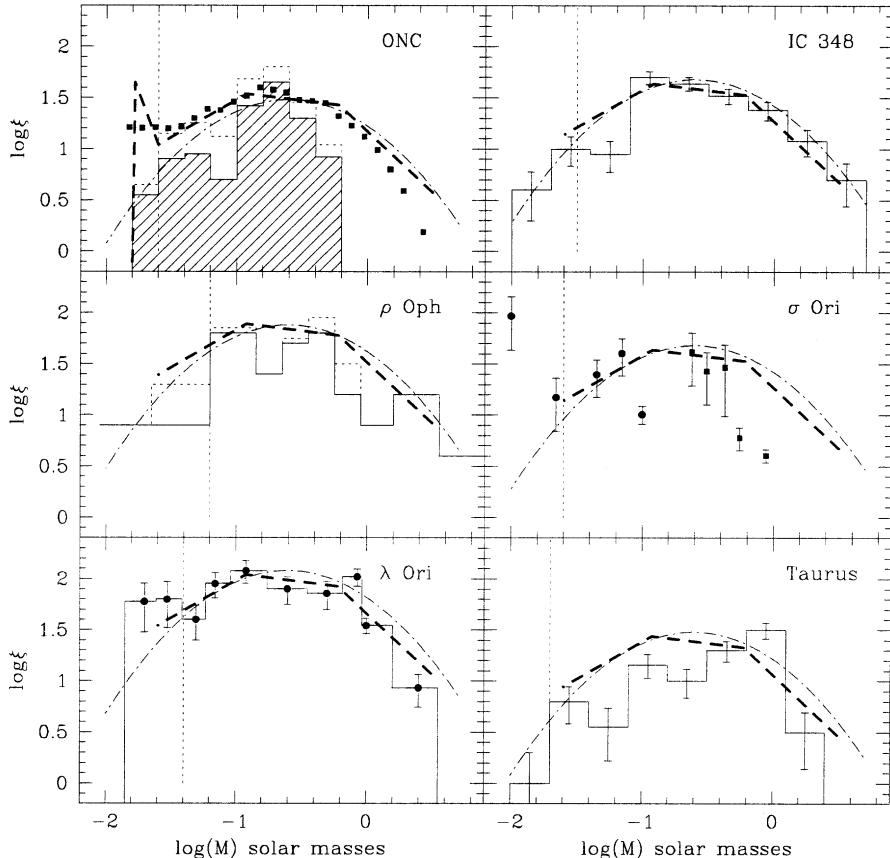


Figure 9.18. Mass functions for a representative sample of young clusters and associations. Data are taken from the following sources: ONC, [H13] ($A_V < 10.0$ sample – solid points), [S13] (histogram) and [M10] (dashed line); IC 348, [L9]; ρ Oph, [L13]; σ Ori, [S12] (solid squares) and [B13] (solid points); λ Ori, [B15]; and Taurus [B13]. The dotted histograms plotted for the ONC and ρ Ori have been corrected for completeness. In each case, the vertical dotted line marks the completeness limit, the error bars reflect Poisson counting uncertainties, the dashed line reproduces the higher mass power law segments of the [M10] ONC relation and the dash-dot line shows the log-normal relation favoured by Chabrier [C5] for the field IMF.

type $\sim M6$, and apparent magnitudes of $I \sim 14$, $K \sim 10$; a $0.02-M_\odot$ dwarf has spectral type $\sim M9$, and $I \sim 17.5$, $K \sim 13.5$. Briceño *et al.* [B10] have compiled the most extensive survey to date, combining 2MASS data with both optical imaging, centred on the concentrations of pre-main sequence stars identified by [G7], and spectroscopy of candidate members. The faintest member identified in this survey has spectral type M9.5, and a likely mass of $0.015 M_\odot$.

The mass functions derived for these six young clusters or associations are plotted in Figure 9.18. In each case, we have superimposed the three higher mass power

law segments from the [M10] analysis of the ONC, scaled appropriately in density (i.e., we omit, as probably spurious, the spike near the deuterium limit). As noted above, this distribution is similar to the power law representation of the field-star IMF (Section 9.5.4). We also plot scaled versions of Chabrier's [C5] log-normal parameterisation of the field-star IMF (Equation 9.42).

In broad terms, both the power laws and the log-normal distribution reproduce the gross properties of the individual mass functions. The [C5] function is not entirely consistent with either the ONC or IC 348 data; a better match might be obtained with a lower characteristic mass ($M_0 \sim 0.15 M_\odot$) and a smaller value of σ . On the other hand, the composite power law characterises well the observations of IC 348, ρ Oph and λ Orionis, although one might argue, somewhat paradoxically, that it is less consistent with the [S13] data for the ONC. The σ Orionis cluster is more problematic; this may stem partly from the fact that there is no overlap between the two observational data sets used to construct $\xi(M)$. The mass function for Taurus, however, is clearly inconsistent with both the [C5] log-normal relation and the [M10] composite power law: the mass function peaks at $\sim 0.8 M_\odot$ and matches a power law index $\alpha \sim 0.1$ at lower masses. There also appear to be fewer brown dwarfs (relative to stars) in Taurus than in higher density star-forming regions, such as IC 348 [L9].

The anomalous nature of the Taurus IMF has been recognised for a number of years, and is generally believed to be related to the extremely low density of star formation. However, the overwhelming majority of disk stars are believed to originate from clusters, rather than loose associations, and Figure 9.18 suggests that there is relatively little variation in the IMF in those systems. The theoretical implications of these results are considered further in the following section.

9.8 THEORETICAL INTERPRETATIONS OF THE IMF

Understanding $\Psi^i(M)$, and how that function varies as a function of other parameters, is a key step towards understanding the overall star formation process. Considerable progress has been made in identifying the different stages in the formation of an individual star from a cloud core (see Section 3.6). However, we have only very limited understanding of the physical processes that govern the formation of a stellar ensemble (i.e., what decides how a molecular cloud redistributes its material to form stars of different masses). A wide variety of mechanisms have been proposed, including fragment coalescence (a bottom-up approach) [S4], [M8]; fragmentation and accretion (a top-down approach) [A1], [Z1], [N1]; interactions of accretion and protostellar outflows [S5]; fractal networks within clouds [L4], [E1]; and competitive accretion and ejection in small groups [C7], [B16]. None of these models is completely successful in explaining the observed form of the IMF. Full consideration of these competing theories is outwith the scope of the present volume, and we refer the interested reader to

more specialised reviews (e.g., [P6]). However, we briefly review the implications of the results summarised in Sections 9.5.4, 9.6.2 and 9.7.4.

A key parameter in many star formation theories is the Jeans mass, defined as:

$$M_{\text{Jeans}} = \left(\frac{5kT}{G\bar{\mu}} \right)^{3/2} \left(\frac{3}{4\pi\rho} \right)^{1/2} \quad (9.43)$$

where T and ρ are the temperature and density of the cloud, m_H is the mass of the hydrogen atom and $\bar{\mu}$ is the mean molecular weight. This can be rewritten as:

$$M_{\text{Jeans}} = 400(T^3/n)^{1/2} \quad (9.44)$$

where n is the number density in molecules cm⁻³. At the Jeans mass, gravitational forces just exceed gas pressure, and the cloud collapses.

In principle (that is, under idealised conditions), the Jeans mass represents the lowest mass object that can form in this environment. However, a typical molecular cloud core has a temperature of 10 K and a density of $\sim 10^6$ molecules cm⁻³, and the corresponding Jeans mass is $\sim 3 M_\odot$ (i.e., under idealised conditions, the Sun should not exist, never mind M, L or T dwarfs). The Jeans mass can be lowered if the potential energy gained through collapse is radiated away, leading to higher densities at (approximately) the same temperature. Eventually, the internal opacity of the collapsing fragment increases to the point that hydrostatic equilibrium is achieved, halting further collapse. This process is opacity-limited fragmentation [H14].

Opacity-limited fragmentation is predicted to produce a mass spectrum that is close to a power law, increasing toward lower masses, with an effective lower mass limit between 0.01 and 0.001 M_\odot . Molecular clouds, however, are far from the idealised, smooth, tranquil environments that would allow this process to proceed as if each collapsing fragment were within its own isolated cocoon. Density non-uniformities, magnetic fields, turbulence and ordered motions due to stellar winds all complicate matters in real clouds, and the internal motions, in particular, are likely to be more important than opacity in setting the mass distribution of fragments, as originally proposed by Arny [A8].

Adams and Fatuzzo [A1] echo and amplify this argument, emphasising the complexity of molecular cloud structure. Indeed, they argue that, ‘the Jeans mass has nothing to do with the masses of forming stars.’ Instead, they propose that the final stellar masses from individual collapsing fragments are determined by a number of parameters – T , ρ , M_{gas} , composition and turbulence – each with its own distribution. The resultant mass function is basically the result of the convolution of each distribution under the central limit theorem, so the final IMF is almost accidental in nature. However, this model predicts a functional form close to a log-normal distribution, and therefore fails to match the near-Salpeter slope at high masses.

Recently, considerable attention has focused on the competitive accretion model for star formation favoured by [C7]; as discussed in Section 3.6.2, this interest is stimulated mainly by the high value associated with the Jeans mass in molecular

clouds, and the consequent apparent impediment to brown dwarf formation. Bate and Bonnell [B16] have examined the likely form of the IMF generated by this process. Their simulations show that, as with most mechanisms, it is possible to reproduce some of the observed features; however, there are also a number of significant problems. As with the [A1] accidental IMF, the mass spectrum generated by competitive accretion is close to log-normal in form, and therefore does not provide a good match at high masses. Moreover, the characteristic mass is close to $0.1 M_{\odot}$, and, while the models agree with observations in predicting larger numbers of brown dwarfs in higher density environments (cf. IC 348 vs. Taurus), they also predict too many brown dwarfs relative to stars.

Perhaps the most promising mechanism, at least at the present time, is turbulent fragmentation, long championed by Larson [L4], [L14]. In this model, protostellar cores form as gas is compressed by supersonic turbulent flow, leading to a broad distribution of gas densities [P7]. The mass distribution of the protostellar cores is a power law (for scale-free turbulence), but collapse, and star formation, requires that gravitational forces exceed the thermal and magnetic energies (i.e., the cores have to exceed the local Jeans mass, M_{Jeans} , which depends on the local gas density). The net result is that $\Psi^i(M)$ is predicted to be a power law at high masses (all cores have masses exceeding M_{Jeans}); the probability that $M_{\text{core}} > M_{\text{Jeans}}$ declines at lower masses, and $\Psi^i(M)$ turns over.

Recent investigations of the IMF predicted by turbulent fragmentation models find a slope close to the Salpeter index at high masses, but a functional form closer to a log-normal distribution below $\sim 1 M_{\odot}$ [N1], although the decline in number densities in the low-mass star/brown dwarf regime is somewhat steeper than observed. The low-mass tail on the distribution is controlled by parameters such as the average gas density and the Mach number of the supersonic flow. The models predict that this should vary in the sense that lower density regions produce fewer low-mass objects, broadly consistent with the comparison between observations of the ONC and Taurus.

9.9 THE STAR-FORMING MASS DISTRIBUTION AND THE LOCAL MASS DENSITY

Multiplying the initial mass function – the number of stars per unit mass – by mass, and normalising that function by the total mass of stars formed, gives $\Xi(M)$. This function describes the fraction of the total star-forming mass of a given molecular cloud which becomes bound up in stars of a given mass, providing a measurement of the relative efficiency of the star-formation process. In the case of the field-star IMF, $\Xi(M) \propto M^0$ for $0.1 < M/M_{\odot} < 1$; that is, mass is equally divided among stars in this mass range. Integrating $\Psi(M)$, stars with masses between the hydrogen-burning limit and $1 M_{\odot}$ account for $\sim 48\%$ of the total star-forming mass. Most of the residual mass is in $1-10 M_{\odot}$ stars, with (intrinsically rare) higher mass stars adding only $\sim 1\%$ of the total. Brown dwarfs are extremely numerous, but regardless

of which mass function is adopted, they contribute only 2–3% of the star-forming mass budget.

Integrating $\Psi(M)$ also allows an upper limit to be set on the contribution made by stars to the local mass density, ρ_0 . Subsolar-mass stars contribute $\sim 0.036 M_\odot \text{pc}^{-3}$, with half of the mass density in M dwarfs. A straightforward integration of $\Psi(M)$ for higher-mass (1–100 M_\odot) stars would indicate that those stars make a similar total contribution. However, the calculation fails to take into account the fact that most of those stars have evolved off the main sequence, and have recycled much of their material to the interstellar medium. As a result, high-mass stars make a relatively small contribution to ρ_0 at any given time. In the local inter-arm region, 1–10 M_\odot stars contribute only $\sim 0.004 M_\odot \text{pc}^{-3}$ to the total mass density. Stellar remnants make a similarly small contribution, amounting to $\sim 0.005 M_\odot \text{pc}^{-3}$, giving a total stellar mass density of $0.045 M_\odot \text{pc}^{-3}$.

As discussed in Section 8.3, one of the main stimulants in the surge of interest in M dwarfs in the 1970s was the discrepancy between the Oort dynamical mass density and the summed contribution from known constituents of the Solar Neighbourhood: the hypothesis of ‘missing mass’ in the Galactic Disk. How do matters stand at the turn of the century? The observed mass density remains largely unchanged: interstellar gas and dust contributes $\rho_{\text{ism}} \sim 0.03$ to $0.05 M_\odot/\text{pc}^3$ which, combined with the stellar mass density, gives $\rho_{\text{obs}} \sim 0.075$ to $0.095 M_\odot/\text{pc}^3$. Dynamical estimates, however, have changed. The most recent, based on Hipparcos data, derives a mass density of $\rho_{\text{dyn}} \sim 0.076 \pm 0.015 M_\odot/\text{pc}^3$ [C2], significantly lower than Oort’s $0.15 M_\odot/\text{pc}^3$, and consistent with the observed value. Thus, there is no longer a need to invoke significant quantities of dark matter within the Galactic Disk.

9.10 SUMMARY

We have reviewed the definition of the stellar mass function, $\Psi(M)$, and its importance in understanding star formation mechanisms and Galactic structure parameters, such as the local mass density. At present, masses can be derived directly only for stars in binary systems with well-determined orbits, although microlensing projects may offer additional data in the future. Those stars are used to calibrate empirical mass–luminosity relationships, which can then be used to derive the field star mass function from the observed luminosity function. The most recent studies indicate that the mass function for disk stars can be well represented by a power law, $\Psi(M) \propto M^{-1}$ for masses between 0.1 and $1 M_\odot$.

Theoretical mass–luminosity calibrations offer the only means of determining $\Psi(M)$ for young stars still in the process of contracting onto the main sequence. There are reasonable grounds for suspecting that variations in the form of $\Psi^i(M)$ occur from one star-forming region to another. The strongest support for this hypothesis comes from comparing observations of dense, populous clusters, such as Orion and IC 348, against the low-density star-forming clouds in the Taurus–Auriga association. Star formation theories still abound for the origin of the IMF,

with no clear outright winner to date. Models based on supersonic turbulent fragmentation appear to show promise.

9.11 REFERENCES

- A1 Adams, F. C., Fatuzzo, M., 1996, *ApJ*, **464**, 256.
- A2 Alcock, C., *et al.*, 1997, *ApJ*, **479**, 119.
- A3 Alcock, C., *et al.*, 2000, *ApJ*, **542**, 281.
- A4 Ali, B., Depoy, D., 1995, *AJ*, **109**, 1379.
- A5 Andersen, J., 1991, *Astr. Ast. Rev.*, **3**, 91.
- A6 Allen, P. R., Korner, D. W., Reid, I. N., Trilling, D., 2005, *ApJ*, in press.
- A7 Adams, J. D., Stauffer, J. R., Skrutskie, M. F., Monet, D. G., Portegies Zwart, S. F., Janes, K. A., Beichman, C. A., 2002, *AJ*, **124**, 1570.
- A8 Arny, T., 1971, *ApJ*, **169**, 289.
- B1 Baraffe, I., Chabrier, H., Allard, F., Hauschildt, P. H., 1998, *A&A*, **337**, 403.
- B2 Becklin, E. E., Neugebauer, G., 1967, *ApJ*, **147**, 799.
- B3 Benedict, G. F., McArthur, N., Nelan, E. *et al.*, 1994, *PASP*, **106**, 327.
- B4 Binnendijk, L., 1960, *Properties of Double Stars*, University of Pennsylvania Press, Philadelphia.
- B5 Boss, A. P., 1991, *Nature*, **351**, 298.
- B6 Burgasser, A. J., 2004, *ApJS*, **155**, 191.
- B7 Burgasser, A. J., 2005, *ApJ*, in press.
- B8 Burrows, A., Marley, M., Hubbard, W. B., Lunine, J. I., Guillot, T., Saumon, D., 1997, *ApJ*, **491**, 856.
- B9 Bouvier, J., *et al.*, 1998, *A&A*, **336**, 490.
- B10 Barrado y Navascues, D., Bouvier, J., Stauffer, J. R., Lodieu, N., McCaughrean, M. J., 2002, *A&A*, **395**, 813.
- B11 Briceño, C., Luhman, K. L., Hartmann, L., Stauffer, J. R., Kirkpatrick, J. D., 2002, *ApJ*, **580**, 317.
- B12 Bontemps, S., *et al.*, 2001, *A&A*, **372**, 173.
- B13 Bejar, V. J. S., *et al.*, 2001, *ApJ*, **556**, 830.
- B14 Barrado y Navascues, D., *et al.*, 2003, *A&A*, **404**, 171.
- B15 Barrado y Navascues, D., Stauffer, J. R., Bouvier, J., Jayawardhana, R., Cuillander, J.-C., 2004, *ApJ*, **610**, 1064.
- B16 Bate, M. R., Bonnell, I. A., 2004, *MNRAS*, **355**, 728.
- C1 Couteau, P., 1981, *Observing Visual Double Stars*, trans. A. H. Batten, The MIT Press, Cambridge, Massachusetts.
- C2 Crézé, M., Chereul, E., Bienaymé, O., Pichon, C., 1998, *A&A*, **329**, 920.
- C3 Covino, E., Frasca, A., Alcalá, J. M., Paladino, R., Sterzik, M. F., 2004, *A&A*, **427**, 637.
- C4 Cruz, K. L., *et al.*, 2003, *AJ*, **125**, 354.
- C5 Chabrier, G., 2003, *PASP*, **115**, 763.
- C6 Chabrier, G., 2004, in the *IMF@50*, ed. E. Corbelli, F. Palla, and H. Zinnecker, Kluwer, Dordrecht.
- C7 Clarke, C., 1998, *The Stellar Initial Mass Function*, ASP Conf. Ser., **142**, 189.
- D1 Dahn, C. C., Liebert, J., Harrington, R. S., 1986, *AJ*, **91**, 62.
- D2 D'Antona, F., Mazzitelli, I., 1994, *ApJS*, **90**, 467.
- D3 Delfosse, X., Forveille, T., Mayor, M., Burnet, M., Perrier, C., 1999, *A&A*, **341**, L63.

- D4 Dyson, F. W., Eddington, A. S., Davidson, C. R., 1920, *Mem. RAS*, **62**, 291.
- D5 Delfosse, X., Forveille, T., Segransan, D., Beuzit, J.-L., Udry, S., Perrier, C., Mayor, M., 2000, *A&A*, **364**, 217.
- D6 D'Antona, F., Mazzitelli, I., 1997, *Mem. Soc. Astr. It.*, **68**, 807.
- D7 Dolan, C. J., Mathieu, R. D., 2001, *AJ*, **121**, 2124.
- E1 Elmegreen, B. G., 1997, *ApJ*, **486**, 944.
- G1 Genzel, R., Reid, M. J., Moran, J. M., Downes, D., 1981, *ApJ*, **224**, 884.
- G2 Genzel, R., Stutzki, J., 1989, *ARA&A*, **27**, 41.
- G3 Ghez, A. M., Weinberger, A. J., Neugebauer, G., Matthews, K., McCarthy, D. W., 1995, *AJ*, **110**, 753.
- G4 Ghez, A. M., McCarthy, D. W., Patience, J. L., Beck, T. L., 1997, *ApJ*, **481**, 378.
- G5 Gizis, J. E., Reid, I. N., Monet, D. G., 1999, *AJ*, **118**, 997.
- G6 Gould, A., Bennet, D. P., Alves, D. R., 2004, *ApJ*, **614**, 404.
- G7 Gomez, M., Hartmann, L., Kenyon, S. J., Hewett, R., 1993, *AJ*, **105**, 1927.
- G8 Greist, K., 2002, *Proc of International conference DARK2002*, eds. H. V. Klapdor-Kleingrothaus, R. D. Viollier, Springer, Berlin.
- H1 Hamblin, N. C., Hawkins, M. R. S., Jameson, R. F., 1993, *MNRAS*, **263**, 647.
- H2 Hamblin, N. C., Stelle, I. A., Hawkins, M. R. S., Jameson, R. F., 1995, *MNRAS*, **273**, 505.
- H3 Hartigan, P., Strom, K. M., Strom, S. E., 1994, *ApJ*, **427**, 961.
- H4 Hawkins, M. R. S., Bessell, M. S., 1988, *MNRAS*, **234**, 77.
- H5 Haywood, M., 1994, *A&A*, **282**, 444.
- H6 Henry, T. J., McCarthy, D. W., 1993, *AJ*, **106**, 773.
- H7 Henry, T. J., Franz, O. G., Wasserman, L. H., Benedict, G. F., Shelus, P. J., Ianna, P. A., Kirkpatrick, J. D., McCarthy, D. W., 1999, *ApJ*, **512**, 864.
- H8 Herbig, G. H., 1954, *PASP*, **66**, 19.
- H9 Herbig, G. H., 1998, *ApJ*, **497**, 736.
- H10 Hillenbrand, L. A., 1997, *AJ*, **113**, 1733.
- H11 Hamblin, N. C., Hodgkin, S. T., Cossburn, M. R., Jameson, R. F., 1999, *MNRAS*, **303**, 835.
- H12 Hillenbrand, L. A., White, R. J., 2004, *ApJ*, **604**, 741.
- H13 Hillenbrand, L. A., Carpenter, J. M., 2000, *ApJ*, **540**, 246.
- H14 Hoyle, F., 1953, *ApJ*, **118**, 513.
- J1 Jones, B. T., Walker, M. F., 1988, *AJ*, **95**, 1755.
- K1 Kleinmann, D. E., Low, F. J., 1967, *ApJ*, **149**, L1.
- K2 Kroupa, P., Tout, C. A., Gilmore, G. F., 1990, *MNRAS*, **244**, 76.
- K3 Kroupa, P., Tout, C. A., Gilmore, G. F., 1993, *MNRAS*, **262**, 545.
- K4 Kroupa, P., 1995, *ApJ*, **453**, 358.
- K5 Kroupa, P., Tout, C. A., 1997, *MNRAS*, **287**, 402.
- K6 Kutner, M. L., Machnik, D. E., Tucker, K. D., Dickman, R. L., 1980, *ApJ*, **237**, 734.
- K7 Kroupa, P., 2002, *Science*, **295**, 82.
- L1 Lacy, C. H., 1979, *ApJ*, **228**, 817.
- L2 Lada, E. A., Lada, C. J., 1995, *AJ*, **109**, 1682.
- L3 Lada, E. A., Lada, C. J., Muench, A., 1998, in *The 38th Herstmonceux Conference*, ed. G. Gilmore, Cambridge University Press.
- L4 Larson, R. B., 1992, *MNRAS*, **256**, 641.
- L5 Leggett, S. K., Hawkins, M. R. S., 1989, *MNRAS*, **238**, 145.
- L6 Luhman, K., Rieke, G. H., Lada, C. J., Lada, E. A., 1998, *ApJ*, **508**, 347.
- L7 Luhman, K., 1999, *ApJ*, **525**, 466.

- L8 Luyten, W. J., 1941, *Annals New York Acad. Sci.*, **42**, 201.
- L9 Luhman, K. L., *et al.*, 2003, *ApJ*, **593**, 1093.
- L10 Luhman, K. L., *et al.*, 2000, *ApJ*, **540**, 1016.
- L11 Lucas, P. W., Roche, P. F., 2000, *MNRAS*, **314**, 858.
- L12 Lucas, P. W., Roche, P. F., Allard, F., Hauschildt, P. H., 2004, *MNRAS*, in press.
- L13 Luhman, K. L., Rieke, G. H., 1999, *ApJ*, **525**, 440.
- L14 Larson, R. B., 1981, *MNRAS*, **194**, 809.
- M1 Maceroni, V., Rucinski, S. M., 1997, *PASP*, **109**, 782.
- M2 Marcy, G. W., Moore, D., 1989, *ApJ*, **341**, 961.
- M3 Mathieu, R. D., Ghez, A. M., Jensen, E. L. N., Simon, M., 1999, in *Protostars and Planets IV*, ed. V. Mannings, A. Boss and S. Russell, University of Arizona Press.
- M4 Metcalfe, T. S., Mathieu, R. D., Latham, D. W., Torres, G., 1996, *ApJ*, **456**, 356.
- M5 Meusinger, H., Schilbach, E., Souchay, J., 1996, *A&A*, **312**, 833.
- M6 Meyer, M. R., Adams, F. C., Hillenbrand, L. A., Carpenter, J. M., Larson, R. B., 1998, in *Protostars and Planets IV*, ed. V. Mannings, A. Boss and S. Russell, University of Arizona Press.
- M7 Miller, G. E., Scalo, J. M., 1979, *ApJS*, **41**, 513.
- M8 Murray, S. D., Lin, D. C., 1996, *ApJ*, **467**, 728.
- M9 Moraux, E., Kroupa, P., Bouvier, J., 2004, *A&A*, **426**, 75.
- M10 Muench, A. A., *et al.*, 2003, *AJ*, **125**, 2029.
- M11 Muench, A. A., Lada, E. A., Lada, C. J., Alves, J., 2002, *ApJ*, **573**, 366.
- N1 Nordlund, A., Padoan, P., 2002, *Lecture Notes in Physics*, **614**, 271.
- P1 Paczyński, B., 1986, *ApJ*, **304**, 1.
- P2 Paczyński, B., 1996, *ARA&A*, **34**, 419.
- P3 Pinfield, D. J., Hodgkin, S. T., Jameson, R. F., Cossburn, M. R., von Hippel, R., 1997, *MNRAS*, **287**, 180.
- P4 Popper, D. M., 1980, *ARA&A*, **18**, 115.
- P5 Prosser, C. F., Stauffer, J. R., Hartmann, L., Soderblom, D. R., Jones, B. F., Werner, M. W., McCaughrean, M. J., 1994, *ApJ*, **421**, 517.
- P6 Palla, F., Stahler, S. W., 2001, *ApJ*, **553**, 299.
- P7 Padoan, P., Juvela, M., Bally, J., Nordlund, A., 2001, *ApJ*, **553**, 227.
- R1 Rana, N. C., 1987, *A&A*, **184**, 104.
- R2 Refsdal, S., 1964, *MNRAS*, **128**, 295.
- R3 Reid, I. N., 1987, *MNRAS*, **225**, 873.
- R4 Reid, I. N., 1996, *AJ*, **111**, 2000.
- R5 Reid, I. N., 1998, in *The Stellar Initial Mass Function*, ASP Conf. Ser. vol. 142, ed. G. Gilmore and D. Howell, p. 121.
- R6 Reid, I. M., Hawley, S. L., 1999, *AJ*, **117**, 343.
- R7 van Rhijn, P. J., 1936, *Publ. Kapteyn Astr. Lab. Gröningen*, No. 47.
- R8 Reid, I. N., Gizis, J. E., Hawley, S. L., 2002, *AJ*, **124**, 1190.
- R9 Reid, I. N. *et al.*, 1999, *ApJ*, **521**, 613.
- S1 Salpeter, E. E., 1956, *ApJ*, **121**, 161.
- S2 Scalo, J. M., 1986, *Fund. Cos. Phys.*, **11**, 1.
- S3 Scalo, J. M., 1998, in *The 38th Herstmonceux Conference*, ed. G. Gilmore, Cambridge University Press.
- S4 Scalo, J. M., Pumphrey, W. A., 1982, *ApJ*, **258**, L29.
- S5 Silk, J., 1995, *ApJ*, **438**, L41.
- S6 Stauffer, J. R., Klemola, A., Prosser, C., Probst, R., 1991, *AJ*, **101**, 980.
- S7 Stobie, R. S., Ishida, K., Peacock, J. R., 1989, *MNRAS*, **238**, 709.

- S8 Strom, S. E., Strom, K. M., Carrasco, L., 1974, *PASP*, **86**, 798.
- S9 Sahu, K. C., Sahu, M. S., 1998, *ApJ*, **508**, L147.
- S10 Segransan, D., Delfosse, X., Forveille, T., Beuzit, J.-L., Udry, S., Perrier, C., Mayor, M., 2000, *A&A*, **364**, 665.
- S11 Schaller, G., Schaerer, D., Meynet, G., Maeder, A., 1992, *A&AS*, **96**, 269.
- S12 Sherry, W. H., Walter, F. M., Wolk, S. J., 2004, *AJ*, **128**, 2316.
- S13 Slesnick, C. L., Hillenbrand, L. A., Carpenter, J. M., 2004, *ApJ*, **610**, 1045.
- S14 Stassun, K. G., *et al.*, 2004, *ApJS*, **151**, 357.
- T1 Tinney, C. G., 1993, *ApJ*, **414**, 279.
- T2 Torres, G., 1995, *PASP*, **107**, 524.
- T3 Torres, G., Henry, T. J., Franz, O. G., Wasserman, L. H., 1999, *AJ*, **117**, 562.
- W1 Walker, M. F., 1969, *ApJ*, **155**, 447.
- W2 Warren, W. H., Hesser, J. E., 1977, *ApJS*, **34**, 207.
- W3 Walsh, D., Carswell, R. F., Weymann, R. J., 1979, *Nature*, **279**, 381.
- W4 White, R. J., Ghez, A. M., Reid, I. N., Schultz, G., 1999, *ApJ*, **520**, 811.
- W5 Wielen, R., Jahreiss, H., Krüger, R., 1983, *IAU Colloquium 76, The Nearby Stars and the Stellar Luminosity Function*, ed. A. G. Davis Philip and A. R. Upgren, p. 163.
- W6 Williams, D. M., Rieke, G. H., Stauffer, J. R., 1995, *ApJ*, **454**, 144.
- W7 Williams, D. M., Comeron, F., Rieke, G. H., Rieke, M. J., 1995b, *ApJ*, **454**, 144.
- W8 Woitas, J., Kohler, R., Leinert, Ch., 2001, *A&A*, **369**, 249.
- Z1 Zinnecker, H., 1984, *MNRAS*, **210**, 43.
- Z2 Zinnecker, H., McCaughrean, M. J., Wilking, B., 1993, in *Protostars and Planets III*, ed. by E. H. Levy and J. I. Lunine, University of Arizona Press, p. 429.
- Z3 Zuckerman, B., 1973, *ApJ*, **181**, 863.
- Z4 Zapatero-Osorion, M. R., *et al.*, 2002, *A&A*, **384**, 937.

9.12 HOMEWORK PROBLEMS

Problem 1

Outline four methods of searching for binary stars. Describe which are most likely to identify systems suitable for measuring the masses of individual components.

Problem 2

Rhy 403 is a single-lined binary in the Hyades cluster with $M_V = 14.4$ and normal colours for its spectral type (M4.5). The system has a velocity semi-amplitude of 40 km sec^{-1} and a period of 1.275 days. Determine the system mass function and the projected semi-major axis. Estimate the mass of the primary using Figure 9.5 and discuss how one can set upper and lower limits on the mass of the companion. Is the companion likely to be a white dwarf if the faintest white dwarfs in the Hyades cluster have $M_V = 12$?

Problem 3

Derive equation 9.30, the characteristic timescale for a microlensing event.

Problem 4

An open cluster has 15 members with masses between 0.9 and $1.0 M_{\odot}$. Assuming a composite power-law mass function (i.e., equations (9.37) and (9.38), scaled appropriately), estimate the total mass contributed by stars with masses between the hydrogen-burning limit and $20 M_{\odot}$. Extrapolate the low-mass tail into the brown régime assuming power-law mass functions with $\alpha = 1.5, 1.0, 0.5$ and 0.0 ; in each case, estimate the number of cluster brown dwarfs with masses exceeding $0.01 M_{\odot}$.

Problem 5

Using the empirical mass–luminosity calibration given in equations (9.32) and (9.33), estimate masses for stars in the hypothetical Corona Borealis cluster (see <http://www.stsci.edu/~inr/nldsdat.html>) and derive the mass function for that cluster. Explain why that function does not match the field mass function shown in Figure 9.12.

10

M dwarfs in the Galactic Halo

10.1 INTRODUCTION

Approximately 99.7% of the stars in the immediate vicinity of the Sun are members of the Galactic Disk. The remaining stars belong to the Galactic Halo – the fossil remnants of the first extensive burst of star formation in the history of the Galaxy (see Chapter 7). A clear distinction should be drawn between the stellar halo population – made of baryonic material and having a total mass of only $\sim 10^9 M_\odot$ – and the dark-matter halo, which is believed to be the dominant contributor to the Galactic potential. The dark-matter halo is held responsible for the relatively flat Galactic rotation curve, but its constituents have not yet been identified; they are the targets of gravitational lensing surveys. The present chapter concentrates on the stellar halo, and illustrates how observations of the lower-mass halo subdwarfs provide insight into the structure of the oldest stellar population so far identified in the Galaxy. The nature of the dark-matter halo is a subject in itself, and recent investigations are summarised elsewhere [T1], [E2].

Low-mass halo subdwarfs are best identified either in the immediate vicinity of the Sun – where they have relatively bright apparent magnitudes and can therefore be studied in detail, but are outnumbered substantially by dwarfs in the Galactic Disk – or by probing the lowest-mass members of globular clusters, where subdwarfs dominate the star-counts, but have faint apparent magnitudes ($V > 22$). The high source density, and consequent image crowding, also complicates ground-based observations of globulars, and, as with open cluster studies (Section 8.8.1), dynamical evolution, mass segregation and tidal stripping remain important concerns in statistical analyses. The availability of the fully operational Hubble Space Telescope has gone a long way toward resolving these problems. In combination, observations of local stars and deep surveys of globular clusters provide complementary means of studying stars near the hydrogen-burning limit in the stellar halo.

10.2 HALO SUBDWARFS IN THE FIELD

As members of an old, and presumably kinematically well-mixed population, the nearby subdwarfs should provide a representative cross-section of the Galactic Halo – at least the inner halo (see Section 7.3.2). Studying the statistical properties of these stars, notably the luminosity and mass functions, demands the identification of a well-defined sample. Low-resolution objective-prism surveys and narrowband photometric surveys [B3] provide a means of searching for metal-poor F and G subdwarfs, looking for stars with weak Mg II H and K absorption lines. Similar surveys for later type subdwarfs can be undertaken (e.g., SDSS spectroscopic data are being used to search for stars with strong hydride features [W2]). To date, however, most studies have concentrated on sifting through proper motion catalogues to identify suitable samples of halo stars.

10.2.1 Photometric properties of M subdwarfs

Individual examples of nearby, late-type subluminous stars have been known since the discovery of Kapteyn’s Star in the late nineteenth century. Most well-known subdwarfs are high-velocity stars, identified through their having large proper motions, and many have accurate trigonometric parallax measurements. Given the variety of original sources, and consequent heterogeneous selection criteria, these subdwarfs scarcely constitute a sample suitable for statistical analysis. However, they do provide an indication of the distribution in the H–R diagram.

Figure 10.1 plots the $(M_V, (V-I))$ and $(M_V, (B-V))$ H–R diagrams for late-type subdwarfs with parallax measurements. Photometric data for nearby stars are also plotted, providing a reference sequence for the Galactic Disk. As with the FGK subdwarfs, the most subluminous stars, lying up to 3 magnitudes below the disk sequence, are also the most metal-poor. The subdwarf sequence has less dispersion in colour in the $(M_V, (B-V))$ diagram than in $(V-I)$, and crosses the disk sequence at $M_V \sim 11$. Later type subdwarfs have redder $(B-V)$ colours than disk dwarfs of the same absolute magnitude, leading to an offset of 0.1–0.2 magnitudes between disk and halo stars in the $(B-V)/(V-I)$ two-colour diagram (see Figure 2.21). [W2] find a similar offset in the SDSS $(g-r)$ colour.

A similar colour offset is evident between disk and halo in the $(J-H)/(H-K)$ diagram (Figure 2.22). As described in Section 4.6.4, metal-poor subdwarfs have increased opacity between 1 and $2\mu\text{m}$ due to the pressure-induced dipole of H_2 . This reduces the flux emitted at those wavelengths, particularly in the H -band ($1.6\mu\text{m}$), leading to subdwarfs having bluer $(J-H)$ colours than disk dwarfs with the same $(H-K)$ colours. However, as with BVI colours, the offset amounts to less than 0.1 magnitude for even the most extreme subdwarfs, while uncertainties in the near-infrared colours are typically ± 0.03 magnitude. Given these circumstances, broadband colours alone cannot provide reliable disk/halo discrimination, and offer only crude abundance estimates.

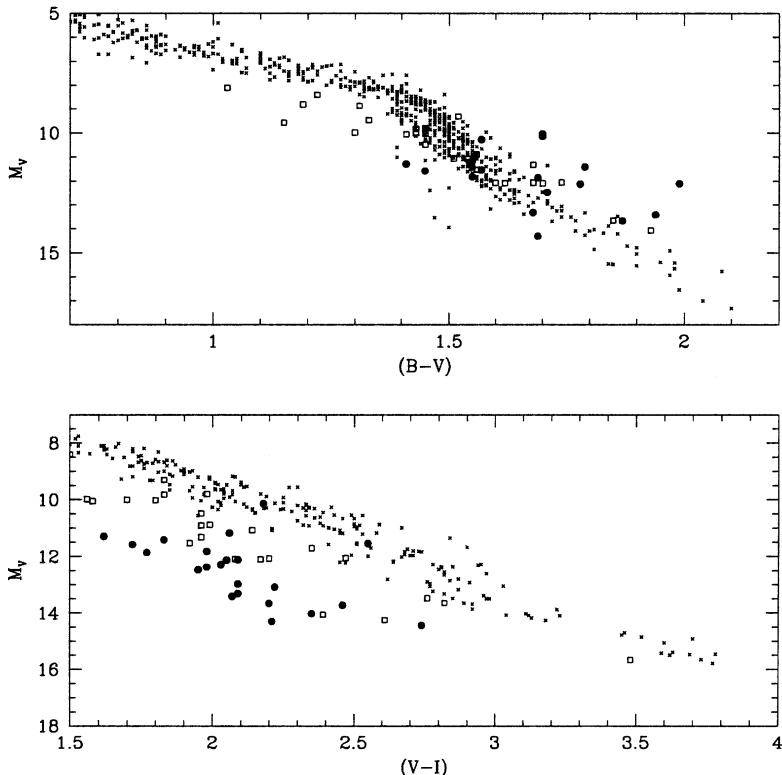


Figure 10.1. The $(M_V, (B-V))$ and $(M_V, (V-I))$ H-R diagrams for late-type subdwarfs. Disk dwarfs are plotted as crosses, sdM dwarfs (Section 10.3) are open squares and esdM dwarfs are filled circles. (From [G1], courtesy of J. Gizis and the *Astronomical Journal*.)

10.2.2 Searching for local subdwarfs

Proper-motion surveys are biased towards stars with high tangential velocities. These biases can be modelled if the underlying velocity distribution is known, as is the case for the local disk and halo populations. Indeed, the bias is a distinct advantage in halo subdwarf identification, since the higher average velocity of halo stars amplifies the number of subdwarfs relative to disk dwarfs in a proper-motion selected sample. This is particularly important, given the extremely low proportion (about 1 in 400; see Section 10.4.4) of halo stars in a *volume-limited* sample.

The relative enhancement of halo stars within a proper-motion survey can be estimated as follows: consider a survey limited to stars with $\mu > \mu_{\text{lim}}$. A star with transverse velocity, V_T , is just detected at a distance given by

$$r_{\text{lim}} = \frac{V_T}{\kappa \mu_{\text{lim}}} \quad (10.1)$$

where $\kappa = 4.74$. A stellar population with solar motion (U, V, W) and velocity

dispersions ($\sigma_U, \sigma_V, \sigma_W$) has a characteristic transverse velocity, $\langle V_T \rangle$. Thus, a characteristic distance can be defined for the detection of members of that population in the survey:

$$r_C = \frac{\langle V_T \rangle}{\kappa \mu_{\text{lim}}} \quad (10.2)$$

The distribution with distance of stars selected by proper motion depends on the form of the velocity distribution. Given a Schwarzschild velocity ellipsoid (that is, Gaussian velocity distributions in each co-ordinate), 25% of the population members with $r < r_C$ are predicted to have $\mu > \mu_{\text{lim}}$, and these stars represent 80% of the contribution made by that population to the proper-motion survey (Figure 10.2). Hence, if a sample includes N_{tot} stars from a given stellar population, with known kinematics which provide a reliable estimate of r_C , then,

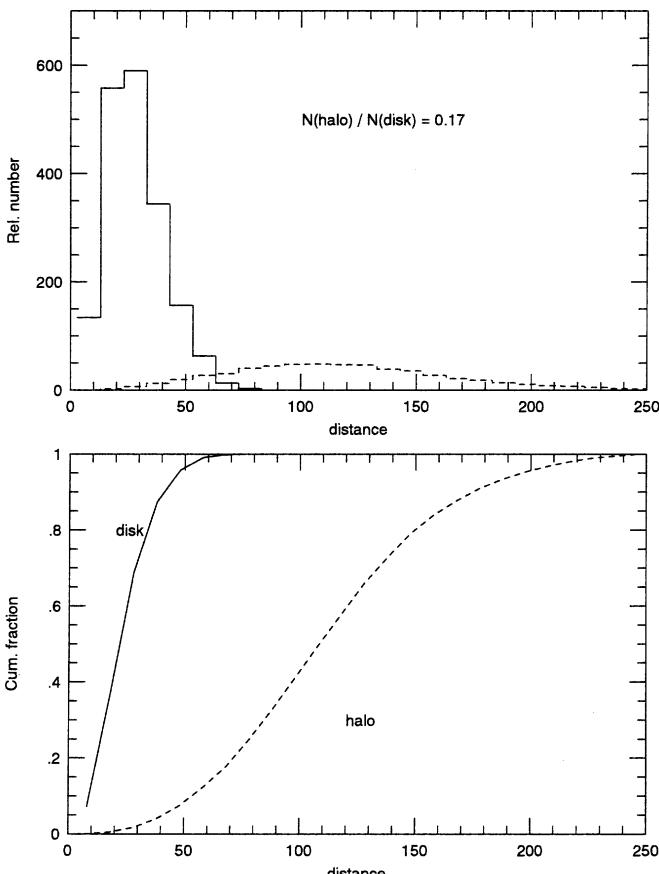


Figure 10.2. The expected distance distribution of a sample of stars selected by proper motion. Disk stars are indicated with solid lines, halo stars with dashed lines. Distances are given in parsecs.

in the absence of any other selection effects, the space density of that population can be estimated from

$$\rho = \frac{N_{\text{tot}} \times 0.8}{0.25 \times r_C^3} \quad (10.3)$$

Since all stars drawn from a given population have similar kinematics, intrinsically fainter stars have fainter apparent magnitudes, on average, than the more luminous members of that population.

Assume a mix of two stellar populations with different kinematics – the disk and the halo. Those two populations have average tangential velocities and characteristic distances of $(\langle V_T \rangle_d, r_d)$ and $(\langle V_T \rangle_h, r_h)$ respectively. The relative number of stars that each population contributes to a proper-motion survey is given by

$$\frac{N_h}{N_d} = \frac{\rho_h r_h^3}{\rho_d r_d^3} \quad (10.4)$$

where ρ_d and ρ_h are the local space densities of the two populations. Since $d_C \propto \langle V_T \rangle$, this ratio can be written as

$$\frac{N_h}{N_d} = \frac{\rho_h \langle V_T \rangle_h^3}{\rho_d \langle V_T \rangle_d^3} = \frac{\rho_h}{\rho_d} \times A_h \quad (10.5)$$

where A_h is the halo amplification factor.

The specific case of interest is the local mixture of halo and disk stars. The kinematic data listed in Table 7.3 show that disk dwarfs have an average tangential velocity of only $\sim 30 \text{ km s}^{-1}$. In contrast, the various representations of the kinematics of the halo predict average tangential velocities of $175\text{--}215 \text{ km s}^{-1}$. This implies an amplification factor, $A_h \sim 100$, for halo subdwarfs in a proper-motion limited sample. Thus, although the *number* density ratio of disk to halo stars is $\sim 400:1$, the higher tangential motions of the halo population can reduce this ratio to values $\sim 4:1$ in proper-motion catalogues.

The full amplification factor, however, does not apply under all circumstances. The calculation above assumes that the stellar sample is defined only based on the proper motion; that is, the sample is proper-motion limited. Other factors can impose a lower effective distance limit; for example, the limiting apparent magnitude of the survey may set the detection limit, leading to a flux-limited sample. In this case, the effective amplification factor for a high-velocity population is reduced.

Consider the case of a sample drawn from the Luyten Half-Second (LHS) catalogue ($\mu_{\text{lim}} = 0.5 \text{ arcsec yr}^{-1}$). Luyten compiled this catalogue by matching the POSS I survey plates against second-epoch plate material, taken with the Palomar Schmidt, which has an effective limiting magnitude of $m_r \sim 19$. Given the average tangential velocities cited above, the characteristic distances for detection of disk and halo stars are ~ 32 and $\sim 140 \text{ pc}$ respectively. These distances correspond to distance moduli of 2.5 and 5.8 magnitudes respectively. As a result, the effective sampling volume for disk dwarfs with $M_r > 16.5$ ($M_V > 18$) and halo subdwarfs with $M_r > 13.2$ ($M_V > 14.5$), is set by the apparent magnitude limit, rather than by the

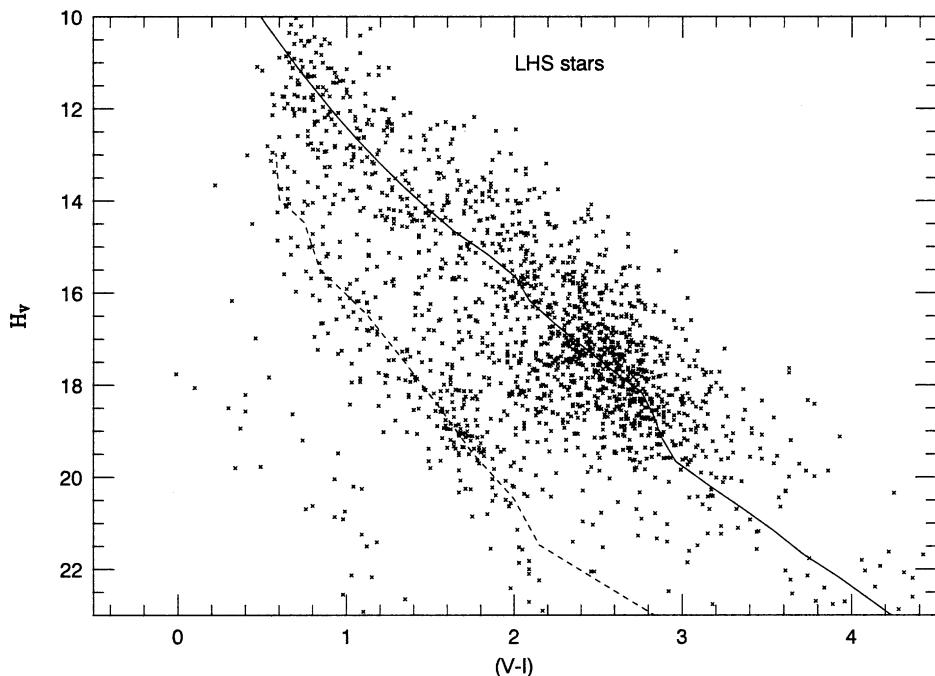


Figure 10.3. The $(H_V, (V-I))$ reduced proper motion diagram for stars in Luyten’s LHS Catalogue. The solid line and dashed line mark the expected location of the disk and halo main sequences at their respective mean tangential velocities. White dwarf stars contribute to the sequence at blue colours and high H_V .

proper-motion limit. The lowest luminosity stars are therefore drawn from a smaller volume, and each star carries more weight in determining both $\Phi_{\text{halo}}(M_V)$ and $\Psi_{\text{halo}}(M)$.

10.2.3 The reduced proper-motion diagram

Proper-motion data provide only the first cut in identifying nearby halo stars. Even with the optimum amplification factor, subdwarfs are still outnumbered by disk dwarfs by a factor of 4:1. Further observations are required to separate the two populations. Moreover, the subdwarfs themselves span a substantial range in abundance and consequently – as Figure 10.1 emphasises – a range of colour-magnitude (and mass–luminosity) relationships. A luminosity function based on *all* subdwarfs within a given volume combines stars drawn from different abundance-dependent mass intervals in a given interval, ΔM_V . It would be ideal to group subdwarfs by abundance, and compare mass functions for intermediate and extreme metal-poor stars; but this demands a means of both differentiating halo subdwarfs from disk dwarfs, and determining approximate metallicities.

Photometric data offer some assistance in segregating disk and halo stars within a proper-motion selected sample. Reduced proper-motion diagrams (Section 8.2.3) are effective if data are available for the right passbands. The disk main sequence, halo subdwarf and white dwarf sequences converge in the $(H_V, (B-V))$ and $(H_r, (m_{pg}-m_r))$ planes as the colour saturates at low luminosities. As a result, these diagrams are of only limited usefulness in identifying low-temperature subdwarfs. However, the sequences are well-separated in $(H_V, (V-I))$ or $(H_r, (R-I))$, allowing identification and elimination of the majority of late-type disk dwarfs from proper-motion selected samples. Figure 10.3 shows an example, plotting $(V-I)$ data for stars from the LHS Catalogue.

10.3 CLASSIFYING LATE-TYPE SUBDWARFS

10.3.1 Cool Subdwarfs

Section 4.6.4 describes how a reduction in metal abundance affects the emergent spectrum in low-mass dwarfs with effective temperatures below 4,000 K. In essence, TiO absorption decreases in strength with decreasing [m/H], but the hydride bands are largely unaffected. The relative strength of those bands (TiO versus hydride), measured using either narrowband photometry or directly from spectra, provides an effective method of verifying the identity of late-type subdwarfs and quantifying their probable abundance. Magnesium hydride is the strongest feature present at optical wavelengths in the G and early-K halo subdwarfs [A1], [C6], while CaH is the most prominent absorber in cooler subdwarfs.

Narrowband indices can be designed to measure the flux within, and adjacent to, individual spectral features. The flux ratio between the on-band and sideband ('pseudocontinuum') measurements provides an estimate of the strength of the feature. As described in Section 2.2.8, the CaH and MgH bands have long been known as dwarf/giant discriminators [O1], and photometric indices designed to measure CaH strength (primarily the 6,880 Å band) were originally devised to take advantage of the gravity sensitivity [J1]. Hydride bandstrength can also be used to identify metal-poor late-type dwarfs (see Section 2.4.1). Following Mould's [M3] prediction that the CaH:TiO bandstrength ratio should be abundance-sensitive, narrowband photometry was used to identify a number of late-type dwarfs with disk-like motions, but subnormal TiO strength [M4]. The objects were classified as 'metal-weak old disk stars'. Similar techniques were later used to classify some 50 high proper-motion stars as disk, intermediate (I) or halo (H) [H1].

Spectrophotometry permits a more finely-tuned measurement of variations in TiO and hydride bandstrength. The first extensive spectroscopic observations of M-type subdwarfs were undertaken by Bessell [B4], whose spectra confirmed the dominance of CaH in the coolest subdwarfs. These qualitative results have been quantified by Gizis [G1], who uses the narrowband indices listed in Table 10.1 (see also Figure 2.20) to measure the relative strengths of CaH and TiO features in the 6,200–7,500 Å region. Figure 10.4 plots these measurements for the parallax subdwarfs from Figure 10.1. The stars have been grouped into three

Table 10.1. Molecular bandstrength indicators.

Band	Sideband 1	On-hand	Sideband 2
TiO 1	6,703–6,708	6,718–6,723	
TiO 2	7,043–7,046	7,058–7,061	
TiO 3	7,079–7,084	7,092–7,097	
TiO 4	7,115–7,120	7,130–7,135	
TiO 5	7,042–7,046	7,126–7,135	
CaH 1	6,345–6,355	6,380–6,390	6,410–6,420
CaH 2	7,042–7,046	6,814–6,846	
CaH 3	7,042–7,046	6,960–6,990	

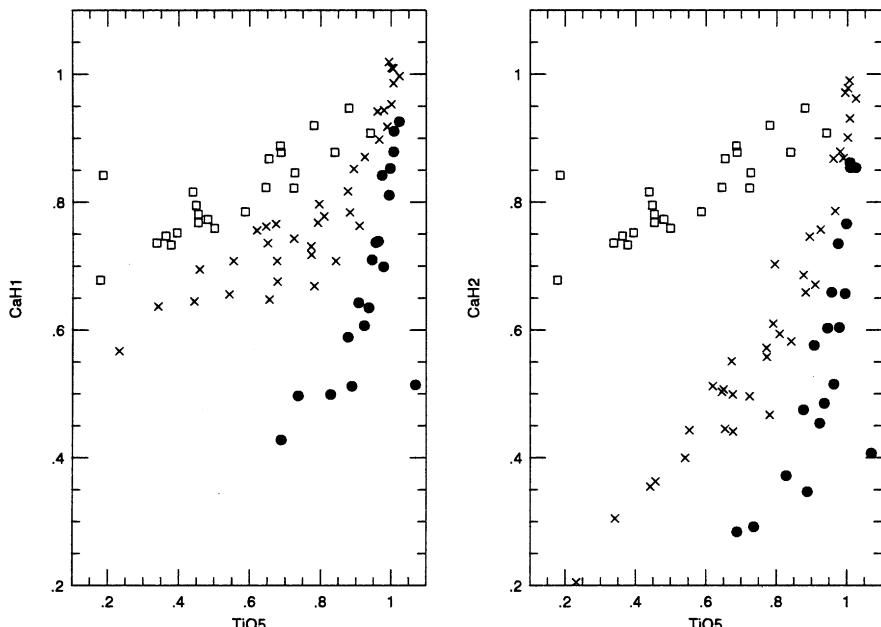


Figure 10.4. The narrowband CaH/TiO classification system. Open squares are disk dwarfs, sdM subdwarfs are marked as crosses, and esdM subdwarfs are solid points. (From [G1], courtesy of J. Gizis and the *Astronomical Journal*.)

categories: disk-like; intermediate-bandstrength subdwarfs, class sdM; and extreme subdwarfs, class esdM.¹ The latter two categories correspond approximately to class (I) and class (H) stars in the [H1] nomenclature. A comparison of Figures 10.1 and 10.4 shows that the edsM subdwarfs are also the most subluminous stars in the colour-magnitude distributions.

¹ The majority of the [M4] metal-weak disk stars – including the well-known nearby high-velocity star, Barnard’s Star (G1 699) – are classified as disk-like.

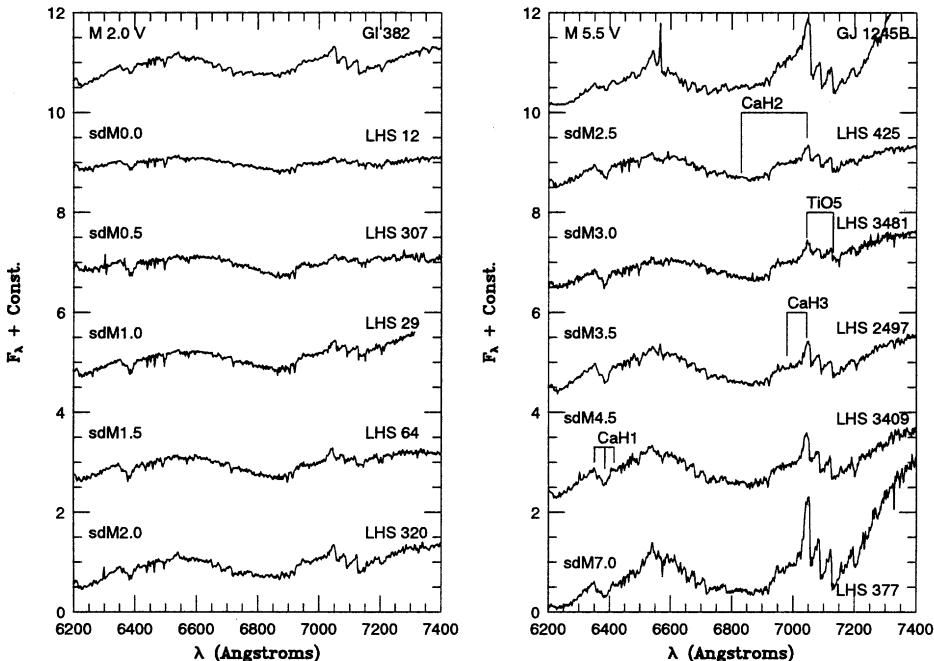


Figure 10.5. The sdM spectroscopic sequence [G1]. A disk dwarf is shown for comparison at the top of each panel. (From [G1], courtesy of J. Gizis and the *Astronomical Journal*.)

Gizis uses these narrowband spectrophotometric indices to define a spectral classification system for M subdwarfs which is comparable in utility to the system defined by Kirkpatrick *et al.* [K3] for near-solar abundance disk dwarfs. The strong dependence on metal abundance exhibited by the TiO bandheads renders those features less useful for unambiguous classification purposes, and Gizis ties his spectral-type scale to the abundance-insensitive CaH bandstrength. The latter offers a more direct correlation with effective temperature. The resultant sdM and esdM spectral sequences are plotted in Figures 10.5 and 10.6.

The relative distribution of the sdM and esdM stars in the colour–magnitude plane indicates that the esdM subdwarfs have lower metal abundance than the sdM subdwarfs. As described in Section 4.6, atmospheres computed for low-temperature metal-poor dwarfs [A3] allow the determination of preliminary quantitative estimates. These comparisons suggest an average metal abundance of $\langle [m/H] \rangle \sim -2$ for the esdM dwarfs, and $\langle [m/H] \rangle \sim -1.3$ for the sdMs. These estimates are supported both by spectroscopy of three sdM companions of FG subdwarfs of known abundance [G2], and by matching the observed colour–magnitude distribution against theoretical isochrones (Figure 3.15). Recent spectral synthesis work by Woolf and Wallerstein [W1] promises to refine these metallicity estimates, and provide more accurate calibration of the CaH/TiO bandstrength relations.

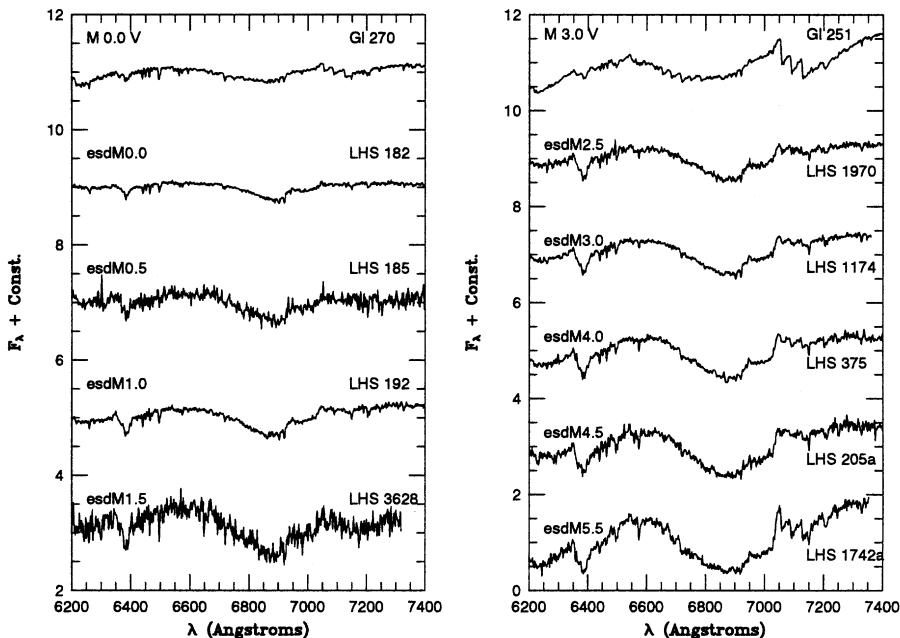


Figure 10.6. The esdM spectroscopic sequence. A disk dwarf is shown for comparison at the top of each panel. (From [G1], courtesy of J. Gizis and the *Astronomical Journal*.)

Gizis' classification therefore provides an effective method of dividing the local halo stars into moderately metal-poor and extremely metal-poor systems. The fact that mildly metal-poor stars, such as Barnard's Star, are excluded from even the sdM category demonstrates that contamination from disk dwarfs is limited to negligible proportions. This approach therefore opens the way for a measurement of statistical properties of the halo – notably the form of the luminosity function – as a function of abundance.

10.3.2 Ultracool subdwarfs

The coolest subdwarf contributing to the original Gizis classification scheme is LHS 377, spectral type sdM7. Over the last five years, new discoveries have extended the subdwarf sequence through sdM to sdL, and perhaps even to T-type subdwarfs. The late-type sdM and esdM dwarfs all stem from new proper-motion surveys, based on the digitisation of old photographic plate material from the Palomar and UK Schmidt telescopes: the latest type esdM dwarf, APMPM 0559-2903, was identified using APM scans of southern Schmidt plates [S3]; Lépine and collaborators have catalogued several intermediate subdwarfs with spectral types later than LHS 377 in their proper-motion survey based on digital subtraction of POSS I and POSS II material [L2], [L3]; and Scholz *et al.* [S4] found the sdM9.5 dwarf SSSPM 1013-1356 using data from the SuperCOSMOS Sky Survey [H2]. In general terms, these

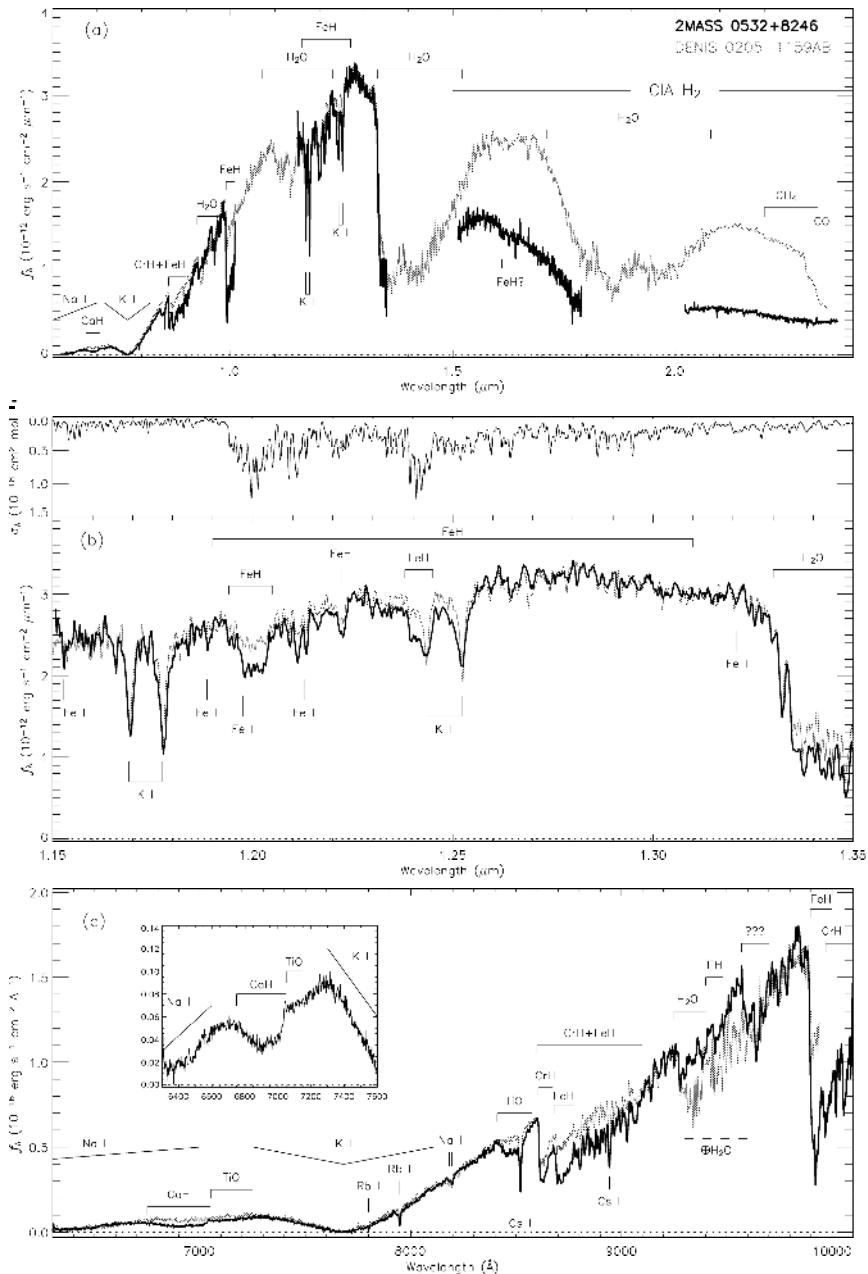


Figure 10.7. A comparison between spectra of 2M0532 + 82 (thick black line) and the L7 dwarfs, Denis 0205–1159. The upper panels (a) and (b) plot low and medium-resolution NIRSPEC data; panel (c) plots optical LRIS data, with the inset showing only 2M0532. In each case, the main atomic and molecular features are identified (from [B5], courtesy of the *Astrophysical Journal*).

later type M subdwarfs continue the trends evident in Figures 10.5 and 10.6, with decreasing TiO and VO band strength and strengthening metal hydride bands.

The first L subdwarf was discovered serendipitously, as a by-product of the 2MASS T dwarf survey described in Section 6.6.2. The object, 2MASS 0532+8246, has a high proper motion ($2.6 \text{ arcsec yr}^{-1}$) and relatively blue near-infrared colours, ($J-K_S=0.26$), accounting for its selection as a ‘blue’ source with no POSS I counterpart [B5]. The spectrum is shown in Figure 10.7: the far-red optical regions resemble late-type L dwarfs, with metal hydride bands, including TiH at $9,400 \text{ \AA}$, and broad sodium and potassium absorption; however, the infrared flux distribution is markedly different longward of the J band, and implies a significantly hotter temperature. The simplest explanation is that 2MASS 0532+8246 is metal-poor; the absence of extensive TiO and VO absorption, and consequent high atmospheric transparency, is not primarily because those elements have condensed to form dust, but because the atmosphere never included a high abundance of those elements.

With a tangential velocity estimated at 320 km s^{-1} (for a distance of $20 \pm 10 \text{ pc}$), 2M0532+82 clearly has halo kinematics. The temperature is not well known, but is likely to be less than $\sim 2,100 \text{ K}$. In that case, the mass is estimated to be less than $0.08 M_\odot$, which is just below the hydrogen-burning limit for a metallicity of $[\text{m/H}] < -1$. This would make the 2MASS 0532+8246 the first halo brown dwarf. Since its discovery, two other likely subdwarf Ls have been added to the roster [L4], [B6].

Moving to even cooler temperatures, the T dwarf 2MASS 0937+2931 has an unusual flux distribution, with very blue near-infrared colours and a very steep optical spectrum, and it has been suggested that these characteristics may stem from low metal abundance [B7]. It is also possible, however, that high surface gravity could produce the same type of effects. Since three other T dwarfs (out of a total of ~ 60) show similar anomalies, statistics seem to favour the latter option, but further observations are clearly warranted.

With the extension of the subdwarf sequence into spectral class L (at least), it becomes necessary to revisit the classification scheme. The standard Gizis calibration rests on measurements of TiO band strengths; those features have essentially disappeared (by definition) in L-type subdwarfs, although the metal hydride bands are still present. This subject is still under discussion, but one option being considered is including measurements at near-infrared wavelengths. Spectroscopy at those wavelengths is no longer a particularly arduous task, and Figure 10.8 shows that a wide range of atomic and molecular absorption features are available, particularly in the J passband. Expanding the late-sdM and sdL catalogue, and refining classification schemes, will be key areas of investigation over the next few years.

10.4 THE FIELD SUBDWARF LUMINOSITY FUNCTION

All except the most recent studies of the halo luminosity function predate quantitative abundance analysis of late-type subdwarfs, and therefore combine all

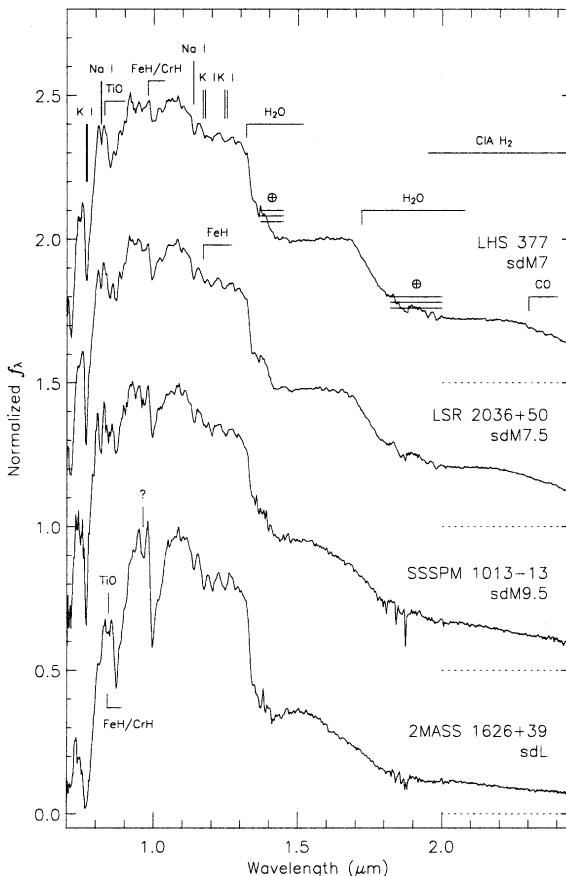


Figure 10.8. Infrared spectra of late-type sdM and sdL subdwarfs; the main atomic and molecular features are identified (from [B6], courtesy of A. Burgasser and the *Astrophysical Journal*).

stars within a single function, $\Phi_{\text{halo}}(M_V)$. Most analyses are based on proper-motion selected samples which, while biased towards including halo stars, still consist predominantly of disk dwarfs. The usual method of removing the disk contaminants is to determine transverse velocities – using either trigonometric or photometric parallaxes – and to include only stars with velocities exceeding some threshold, typically $V_T > 200\text{--}250 \text{ km s}^{-1}$. This criterion also restricts the sample to high-velocity halo stars, and directly computed space densities underestimate the true $\Phi_{\text{halo}}(M_V)$. However, Monte Carlo simulations can be used to compute the expected transverse velocity distribution for a given set of kinematics, $(U, V, W; \sigma_U, \sigma_V, \sigma_W)$, and hence determine the fraction of stars, f_c , with transverse velocities $V_T > V_{\text{lim}}$ [R5], [B1]. The results of recent analyses of halo kinematics are given in Table 7.2, and Figure 10.9 plots the expected tangential velocity distribution for two models. Appropriate correction factors, $\zeta_c = 1/f_c$, typically

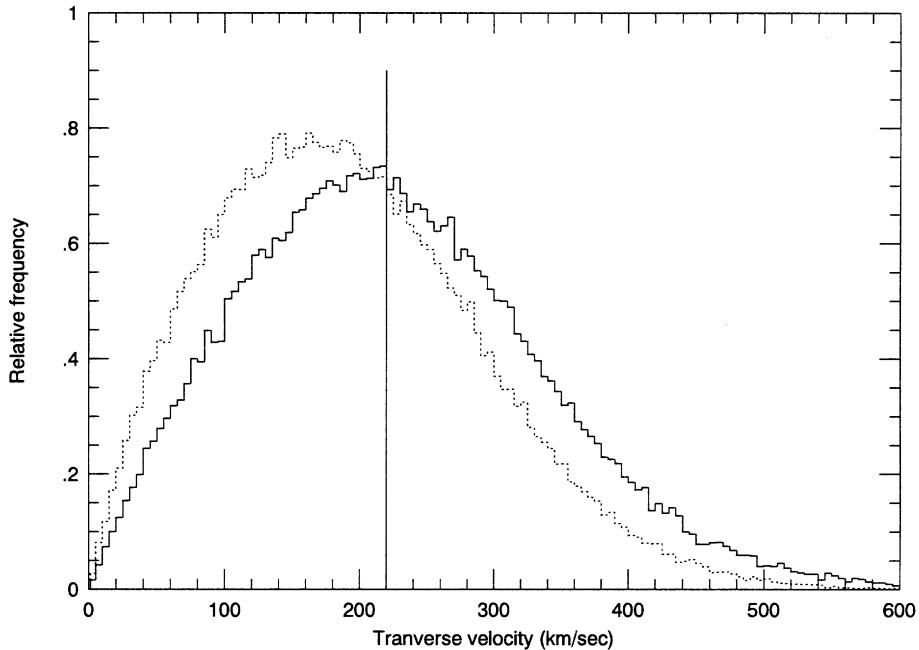


Figure 10.9. Tangential velocity distributions predicted by Monte Carlo simulations for kinematic models based on the Layden *et al.* RR Lyrae analysis [L1] (solid line), and Norris' [N1] observations of local subdwarfs (dotted line). In the former case, 51% of the stars have $V_T < 220 \text{ km s}^{-1}$, while the lower kinematics derived in the latter analysis place $\sim 63\%$ of the sample below that velocity cut-off.

ranging from 2 to 4, are deduced from these distributions and applied to the observed number-densities to estimate $\Phi_{\text{halo}}(M_V)$.

Deep star-count data can also be used to constrain the form of the luminosity function of field subdwarfs. Transforming these measurements to an estimate of the absolute space density of nearby subdwarfs is more complicated, since it requires accurate knowledge of the halo density distribution. This section summarises results deduced using these different techniques, and combines the most reliable data to obtain a best estimate of $\Phi_{\text{halo}}(M_V)$ for field subdwarfs.

10.4.1 Schmidt's derivation of the halo luminosity function

The first observational study of the halo luminosity function was carried out by Schmidt [S1]. His sample consisted of 126 stars, drawn primarily from the Lowell proper-motion survey, with $m_{pg} < 15.95$ and $\mu \geq 1.^{\circ}295 \text{ yr}^{-1}$. 121 of these stars had measured trigonometric parallaxes (in 1975), and distances to the remaining five stars were estimated using their photometry. Based on the V/V_{max} statistic (Section 8.5.1), this sample has a uniform distribution, indicating that it is likely

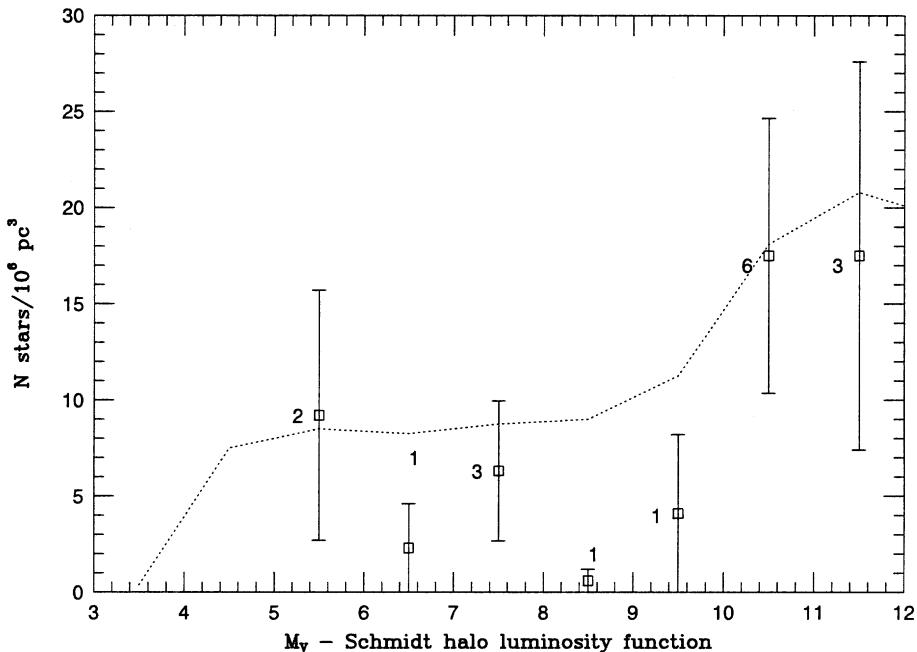


Figure 10.10. A comparison between Schmidt's halo luminosity function (transformed to M_V) and the nearby-star luminosity function scaled by a factor of 1/400 (dotted line).

to be complete.² Schmidt eliminated disk dwarfs by restricting analysis to stars with $V_T > 250 \text{ km s}^{-1}$, the median velocity of a stellar population with kinematics matching those derived by Oort [O2] for metal-poor RR Lyrae stars (Table 7.2). The corresponding correction factor is $\zeta_c \sim 2$.

Applying the cut-off in tangential velocity reduces the final sample to only 17 subdwarfs with $5 < M_V < 12$, and one white dwarf. Schmidt derived space densities from these stars using the $1/V_{\max}$ method: as in determining V/V_{\max} , one calculates the maximum distance to which a given object can be moved and still be retained in the survey. The distance limit is set by either the apparent magnitude limit or the proper-motion cut-off. Thus, the star G3-36 has $\mu = 2.^{\circ}40 \text{ yr}^{-1}$, $m_{pg} = 13.4$, and $r = 26.3 \text{ pc}$. Moving it to $r = 48.7 \text{ pc}$ reduces μ below the survey limit of $1.^{\circ}295 \text{ yr}^{-1}$, giving $V/V_{\max} = (26.3/48.7)^3 = 0.157$. On the other hand, the star G75-47, at $\mu = 1.^{\circ}42 \text{ yr}^{-1}$, $m_{pg} = 15.9$ and $r = 45.5 \text{ pc}$, has $r_{\max} = 46.5 \text{ pc}$ set by the apparent magnitude limit of $m_{pg} < 15.95$.

The density contribution of each star is given by

$$\Phi_{tot} = \zeta_c \times \sum \frac{1}{V_{\max}} \quad (10.6)$$

² No stars matching Schmidt's criteria have been discovered in the intervening 30 years.

and the total space density derived from Schmidt's halo star sample is 1.9×10^{-4} star pc $^{-3}$. Two-fifths of the number density rests with the single white dwarf in the sample (G259-21, also known as Gl 699.1), so the space-density of halo subdwarfs in the range $5 < M_V < 12$ is 1.2×10^{-4} stars pc $^{-3}$; or a number ratio relative to disk dwarfs in the same absolute magnitude range of $\sim 1 : 350$.³ Allowing for statistical limitations, there appear to be no radical differences between the shapes of the halo and disk luminosity functions (Figure 10.10).

10.4.2 $\Phi_{\text{halo}}(M_V)$ from proper-motion star samples

Schmidt's methods have been applied in several subsequent analyses which take advantage of larger, better calibrated proper-motion samples. Each adopts a different velocity criterion for segregating halo subdwarfs, different halo kinematics and, as a result, different values of ζ_c . With one exception the results are in reasonable agreement. The exception is Eggen's study [E1], where the halo selection criterion corresponds to $V_T > 85\text{--}130 \text{ km s}^{-1}$, leading to heavy contamination by moderate-velocity disk dwarfs and substantially higher inferred number densities. Setting aside those results, the other analyses can be combined to determine the local subdwarf luminosity function for $M_V > 4$.

Bahcall and Casertano [B1] provide the best statistics at bright magnitudes. They re-analyse Eggen's dataset of proper-motion stars, comprising stars with $V < 15$ and $\mu > 0.^{\circ}7$ south of $\delta = +15^\circ$, but with a tangential velocity cut-off of 220 km s^{-1} . This reduces the sample to only 94 stars and, since this is a southern sample derived partly from the Bruce proper motion survey ($m_{pg} < 15$), there may be incompleteness at fainter apparent (and absolute) magnitudes. Distances are based on photometric parallaxes (Eggen's M_I , ($R\text{--}I$) calibration) rather than trigonometric parallaxes. The halo kinematics adopted by [B1] have relatively high net rotation, ($V = -154 \text{ km s}^{-1}$; $\sigma_U = 140$, $\sigma_V = 100$, $\sigma_W = 75 \text{ km s}^{-1}$), and imply a correction factor of $\zeta_c = 3.0$ and a halo to disk density ratio of $1 : 300$ for stars in the absolute-magnitude range $4 \leq M_V \leq 11$. Casertano *et al.* [C3] reconsider this analysis with revised halo kinematics, corresponding to a higher median transverse velocity and a lower correction factor of 1.85. Figure 10.11 plots the [B1] results with the [C3] scaling.

Statistics at lower luminosities come from four major analyses, all based on proper-motion selected samples: Dahn *et al.*'s [D1] survey of stars from the LHS catalogue; Gould's [G5] analysis of halo subdwarfs in Luyten's NLTT catalogue; and two investigations based on new proper motion catalogues, by Gizis and Reid [G3] and Digby *et al.* [D12]. Both of the [D1] and [G3] studies include spectroscopic follow-up observations, and are therefore based on confirmed subdwarfs; the [G5] and [D12] surveys are on a larger scale, and rely on photometric selection. However, all four give luminosity functions in good agreement, as shown in Figure 10.11.

³ Note that this comparison does not take account of the changing mass– M_V relationships followed by stars of varying abundance.

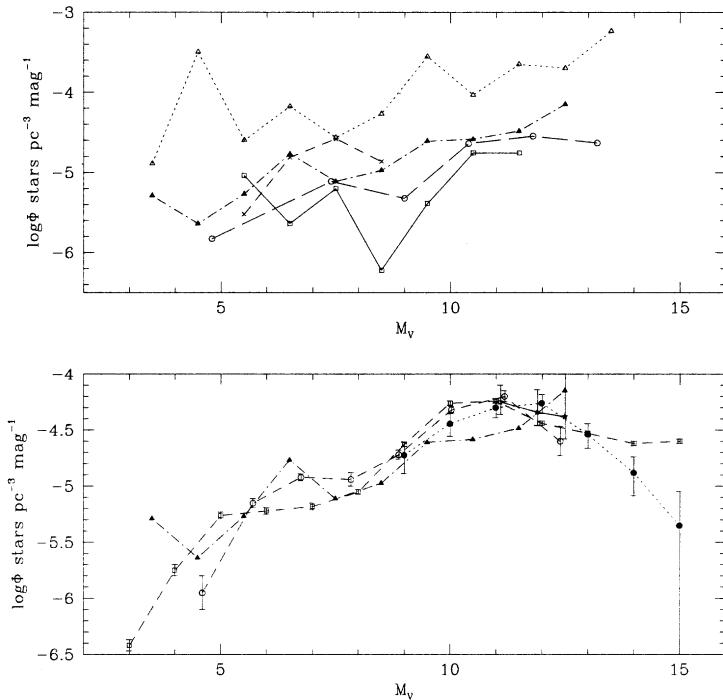


Figure 10.11. The halo luminosity function in the Solar Neighbourhood – the upper panel plots data from the analyses by Schmidt ([S1], open squares), Eggen ([E1], open triangles), Reid ([R1], crosses), Dawson ([D5], open circles) and Casertano *et al.* ([C3], solid triangles). The lower panel plots the [C3] data (solid triangles), Gizis and Reid ([G3], five-point stars), Dahn *et al.* ([D1], solid points), Gould ([G5], open squares) and Digby *et al.* ([D12], open circles).

Considering these investigations in turn, the [D1] survey is based on 114 LHS stars with $0.^{\circ}8 < \mu < 2.^{\circ}5$, $11 < m_r < 18.1$, $\delta > -20^\circ$ and $V_T > 220 \text{ km s}^{-1}$. Space densities were calculated using the $1/V_{\max}$ method, and scaled for incompleteness using a correction factor of $\zeta_c = 2.46$, based on the [M2] halo kinematics (see Table 7.2). Subsequent studies (e.g., [G5]) have favoured the [C3] kinematics as more appropriate ($\zeta_c = 1.85$), and the results plotted in Figure 10.11 are scaled accordingly. The [D1] survey was the first to provide statistically significant evidence that $\Phi(M_V)$ flattens at $M_V > 10$.

The [G3] analysis concentrates on halo subdwarfs in the absolute magnitude range $10 < M_V < 13$. The parent proper motion sample is drawn from scans of UKST and POSS plates that cover 300 square degrees, with selection criteria $\mu > 0.^{\circ}1$ and $R < 20$. Subdwarf candidates were identified using the reduced proper motion diagram (H_R , $(R-I)$), and confirmed spectroscopically. The relative numbers of esdM and sdM stars in the sample are in the ratio 2:5, broadly consistent with the relative numbers of F and G-type subdwarfs with

$[m/H] < -1.5$ and $[m/H] > -1.5$ [C1]. This suggests that there are no substantial changes in the shape of the luminosity function over this abundance range. The overall densities are in close agreement with [D1] (Figure 10.11).

The remaining two investigations are derived from much larger parent samples, but essentially apply the same analysis techniques as the [G3] survey. Thus, in Gould’s analysis of the NLTT, subdwarf candidates are selected using the $(H_V, (V-J))$ reduced proper motion diagram, where V is derived from the NLTT photometry and J from 2MASS. Some 4,600 of the $\sim 24,000$ NLTT stars in the revised NLTT are classed as subdwarfs, underlining the kinematic bias described in Section 10.2.2; space densities are derived using a maximum likelihood technique.

The [D12] analysis is based on a proper motion catalogue derived from the SDSS early release [S5], which covers 394 square degrees. The proper motions are calculated by cross-referencing the SDSS data against SuperCOSMOS scans of POSS I plates, and candidate subdwarfs are selected using the $(H_r, (r-i))$ reduced proper-motion diagram. While these stars lack spectroscopy, SDSS provides accurate $ugriz$ photometry, and the distribution of the 2,600 candidates in the $(u-g)-(g-r)$ plane shows that they have ultraviolet excess fluxes consistent with $[m/H] < -1$. Densities are derived using the $1/V_{\max}$ method. The SDSS data set is drawn from both the northern and southern Galactic Caps, but densities derived from each region are in excellent agreement, so a combined luminosity function is plotted in Figure 10.11.

10.4.3 Φ_{halo} from star-count analyses

An alternative to local surveys is to use deep star-counts to study halo stars *in situ* at large distances above the Plane. The vertical density law deduced from Galactic structure analyses (Sections 7.3.2 and 7.5.2) indicates that the halo component dominates the stellar distribution at heights of 5 kpc or more above the Plane. Selecting a sample at these distances avoids significant problems due to contamination by disk dwarfs, but also requires observations extending to faint apparent magnitudes: $V > 23$ and $I > 20$ for even the earlier-type M subdwarfs. These faint magnitudes make spectroscopic observations difficult and time-consuming. There is also the complication of transferring the observed number densities to the equivalent densities in the Plane.

The deepest star-count analysis combines V and I data from 53 HST Wide Field Camera images to derive $\Phi_{\text{halo}}(M_I)$ for the halo at $|z| > 6$ kpc [G4]. The majority of these stars have I magnitudes fainter than 21st magnitude. None have spectroscopic observations, although their inferred physical location in the Galaxy argues overwhelmingly that they are likely to be members of the halo. The general morphology of the derived luminosity function is consistent with local analyses. If one extrapolates the [G4] results to $z = 0$ using density laws derived from the star-count data, $\rho(r) \propto r^{-3.15 \pm 0.23}$, and an axial ratio $c/a = 0.82 \pm 0.13$, the implied space densities are a factor of two lower than those derived from local stars. The flattened inner halo ([S2]; see Section 7.3.2) lies unnoticed in deep star count studies, since it makes only a minor contribution at large heights above the Plane. Consequently, its contribution

to the local halo density cannot be extrapolated from those studies. Proper-motion surveys of local stars continue to provide a more effective method of probing Φ_{halo} .

10.4.4 The field halo luminosity function

Figure 10.12 plots the composite halo luminosity function, found by combining results from the recent investigations plotted in the lower panel of Figure 10.11. All analyses are scaled by the same correction factor, based on the [C3] halo kinematics. There is some disagreement between the [G5] and [D12] analyses at bright magnitudes, with the NLTT survey deriving lower number densities for $5 < M_V < 8$, but higher densities at brighter magnitudes. Given the higher precision of the SDSS photometry, we give more weight to the [D12] survey for $M_V > 5$, but the NLTT is likely to be more complete for the brighter stars, which are saturated in SDSS. At fainter magnitudes, the [B1] luminosity function is dis-

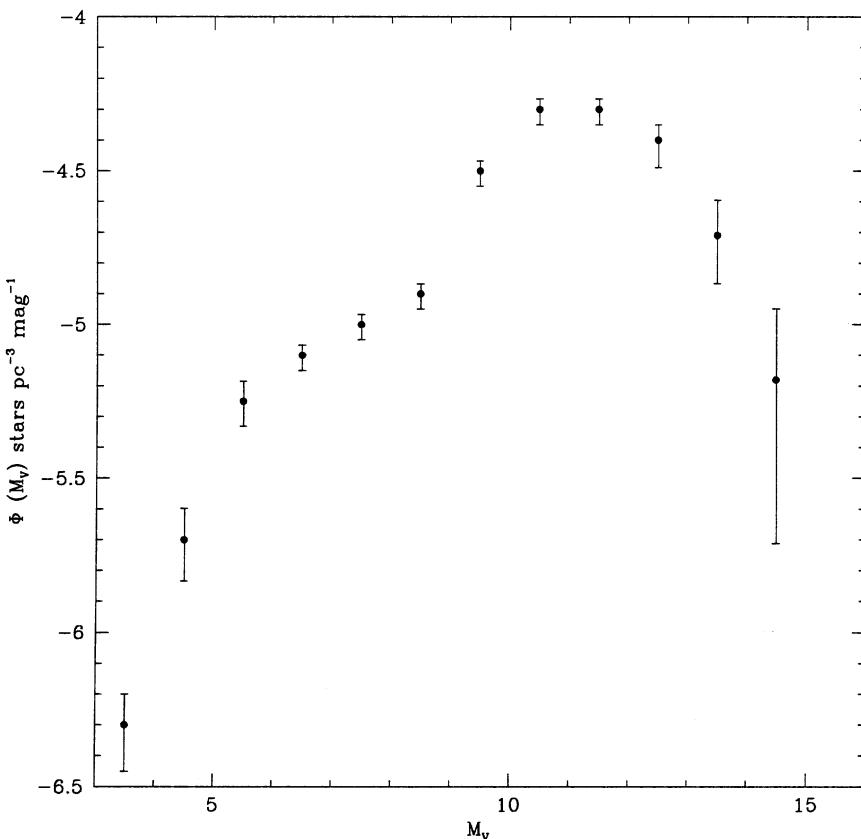


Figure 10.12. The halo subdwarf luminosity function for the Solar Neighbourhood. As described in the text, the number densities have been derived by combining results from individual surveys plotted in Figure 10.11.

crepant, probably reflecting incompleteness in the parent sample; the other surveys agree within 10% at these magnitudes. The inferred number density of halo subdwarfs in the Solar Neighbourhood with absolute magnitudes in the range $4 < M_V < 15$ is 2.36×10^{-4} stars pc $^{-3}$, implying a number ratio of $\sim 1:400$ relative to the local number density of disk dwarfs (Section 8.10).

10.5 LUMINOSITY FUNCTIONS IN GLOBULAR CLUSTERS

Globular clusters provide an alternative means of exploring the halo luminosity function. Cluster studies offer the advantage of dealing with a coeval population of known metal-abundance, and since the cluster population spans a broad abundance range, one can hope to identify significant metallicity-dependent trends. There are, however, compensating complications. First, internal dynamical evolution is expected to lead to mass segregation, while external gravitational interactions can strip the outermost stars from the cluster. Extensive simulations [R3], [K1] show that internal evolution least affects the mass distribution at intermediate radii near the half-light radius, r_h (the radius which encompasses half the total flux). As a result, luminosity functions measured at intermediate radii are reasonably close approximations to the global luminosity function in most globular clusters, and, as discussed further in the following section, dynamical models can be used to estimate the appropriate corrections to the derived mass functions.

A more important consideration is the necessity of obtaining accurate star-counts at faint magnitudes and high surface source-densities. With distances of 3–10 kpc for the nearer cluster systems, early-type M subdwarfs have apparent magnitudes of $I \sim 21$, while the hydrogen-burning limit lies at apparent magnitudes fainter than $I = 23$ and $V = 26.5$. The star density to $I = 22$ nd magnitude in even the outer regions of these clusters can exceed 2,000 stars per square arcminute, corresponding to an average star-to-star separation of ~ 1 arcsec. As a result, image crowding and confusion limits both the photometric and star-counting accuracy attainable with ground-based, seeing-limited observations.

Despite these substantial difficulties, considerable effort was devoted in the late 1980s and early 1990s to obtaining deep ground-based photometry – often in a single passband – of the nearer cluster systems. Many of these observations were undertaken by Richer, Fahlman and collaborators using the Canada–France–Hawaii Telescope [R3], [R4]. These studies employ sophisticated automated software to identify stars by finding local intensity maxima in the CCD frames, and determine magnitudes by fitting the full image profile. Comparably deep exposures of fields offset from the cluster are used to assess the contribution from background field-stars, while the completeness of the final stellar census is estimated by adding artificial stars and determining, as a function of magnitude, the fraction recovered by the analysis software. Summarising the results from these programmes – which extend to fainter than 23rd magnitude in the I -band – [R3] argued that the rising luminosity functions measured in most clusters provided clear evidence for a steeply-rising mass function, perhaps extending into the brown dwarf regime.

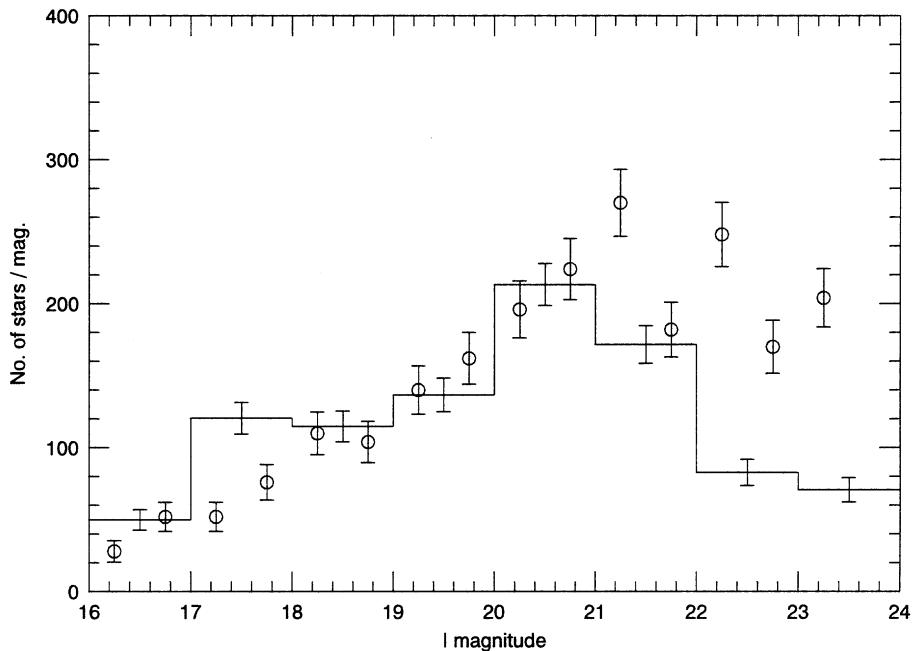


Figure 10.13. Luminosity functions derived for the nearby cluster NGC 6397, from ground-based observations [R3] (open circles), and using HST data [P1] (histogram).

This conclusion has not been borne out by higher spatial-resolution HST observations. Diffraction-limited images with a point-spread function of FWHM $\sim 0.^{\circ}09$ at $\lambda = 8,000 \text{ \AA}$ provide significant advantages in analysing observations of crowded fields. The first globular cluster studied in detail was NGC 6397 ($[\text{m}/\text{H}] \sim -1.7$), the second-nearest system to the Sun, observed by Paresce *et al.* [P1]. Their I -band luminosity function shows a maximum at $I \sim 20.75 \pm 0.25$ ($M_I \sim 8.0$) with an almost-symmetric decline in number densities towards fainter magnitudes (Figure 10.13). This is not due to incompleteness. The colour–magnitude diagram (Figure 10.14) shows a strikingly narrow main sequence which peters out well above the limiting magnitude reached by the observations.

The discrepancy between the ground-based and HST observations stems from two sources. First, severe crowding within the field led to an *overestimation* of the necessary corrections for incompleteness. Second, NGC 6397 lies at low Galactic latitude, and, as a result, field stars contribute significantly to the total number counts. This contribution can be assessed accurately from the HST colour–magnitude diagram. [R3], however, had observations in only the I passband, and estimated the contamination from observations of a control field, offset from the cluster. The latter observations underestimated the level of contamination, and the two errors combined to give an overestimate of the number of low-luminosity cluster members.

Baraffe et al, 1997

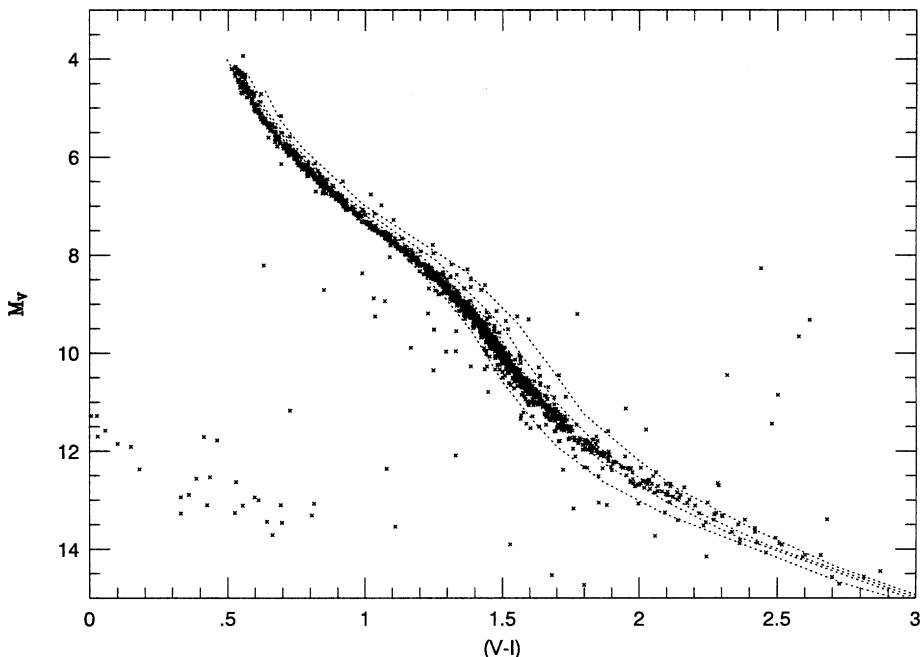


Figure 10.14. The main sequence of NGC 6397, from HST observations [C5], compared with theoretical isochrones for metal-poor stars. The lowest luminosity track has $[m/H] = -2.0$; isochrones for $[m/H] = -1.5, -1.2$ and -1.0 are also shown. (Data courtesy of A. Cool.)

As a bonus, it has proven possible to use HST images of NGC 6397 spanning a baseline of only 32 months to determine proper motions with sufficient accuracy to isolate cluster members [K2]. Analysis of those observations produces the colour-magnitude diagram plotted in Figure 10.14. The s-shaped morphology of the main sequence was predicted by D'Antona [D2], and stems from the same changes in internal structure that occur in solar-abundance disk dwarfs (see Figure 3.15). The change in slope at $M_V \sim 8$ marks the onset of H_2 association, and the corresponding decrease in the adiabatic gradient, while the second point of inflection, at $M_V \sim 13$ ($\sim 0.12 M_\odot$), occurs when pressure, rather than temperature, becomes responsible for the ionisation of the stellar envelope [D3]. Below this point, stars become increasingly degenerate, with the radius remaining nearly constant at $0.1 R_\odot$. The hydrogen-burning limit in NGC 6397 lies at $M_V \sim 14.5$ and $(V-I) \sim 2.6$, both brighter and bluer than in the disk, reflecting the lower opacities in these metal-poor stars. With an age of more than 11 Gyr, any brown dwarfs in these systems have cooled to temperatures substantially less than 1,000 K, but should be detectable with JWST.

By now, over half of the Galactic globular cluster population has been imaged by HST, although in most cases the exposures are relatively short, and the data scarcely extend more than 1–2 magnitudes below the main sequence turn-off. Only

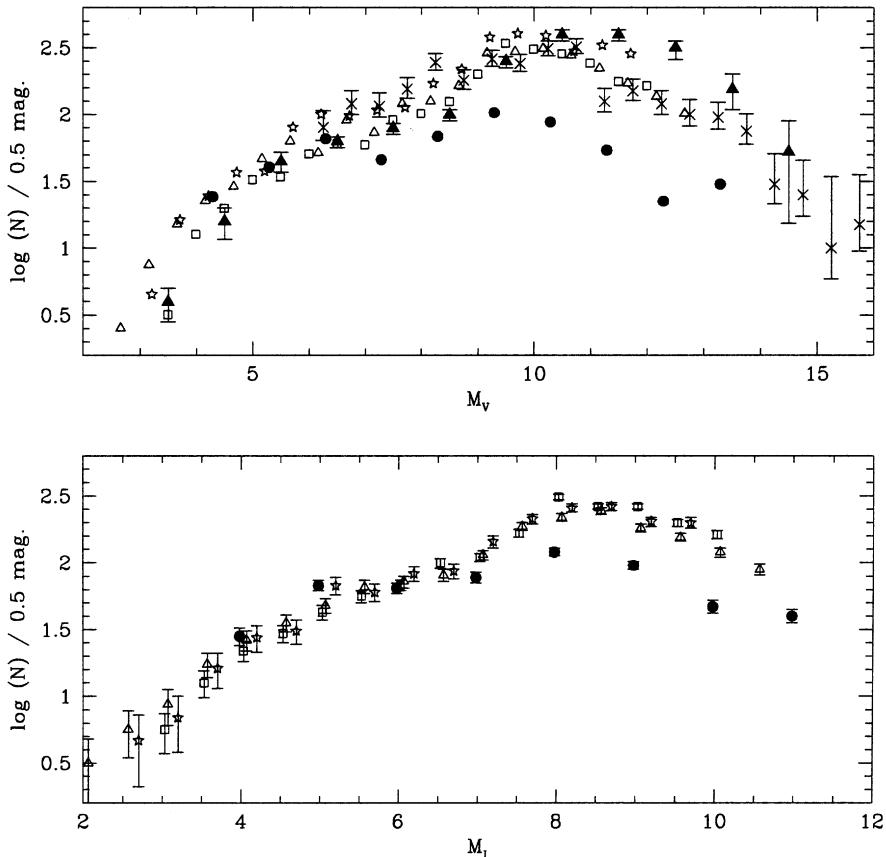


Figure 10.15. $\Phi(M_V)$ (upper panel) and $\Phi(M_I)$ (lower panel) for the clusters M15 (open squares), M30 (stars), M92 (open triangles), NGC 6397 (solid points) and M4 (crosses). The individual data sets are scaled to match at $M_V = +5$ and $M_I = +6$. The upper panel also plots the composite luminosity function for the field halo (solid triangles). (Cluster data from [P2] and [R6].)

a dozen or so of those clusters have deep images, providing accurate photometry of the lower main sequence (see, e.g., [K2], [P2], [P3], [P6]). Those observations, however, are sufficient to confirm mass segregation at the level predicted by dynamical models [D8], [M5], [P7]. The deepest observations currently available are of M4, the nearest globular cluster, which received the undivided attention of HST for 123 orbits in 2001. The effective integration times were ~ 36 hours in the V -band and ~ 53 hours in the I -band, resulting in photometry that reaches $V \sim 30$ and $I \sim 28$ [R6]. Although aimed primarily at mapping the cluster white dwarf sequence, the cluster main sequence is discernible to at least $V = 29$.

Luminosity functions derived from some of the deeper HST observations are shown in Figures 10.15 and 10.16. With the exception of M4 ($[m/H] \sim -1.2$), the

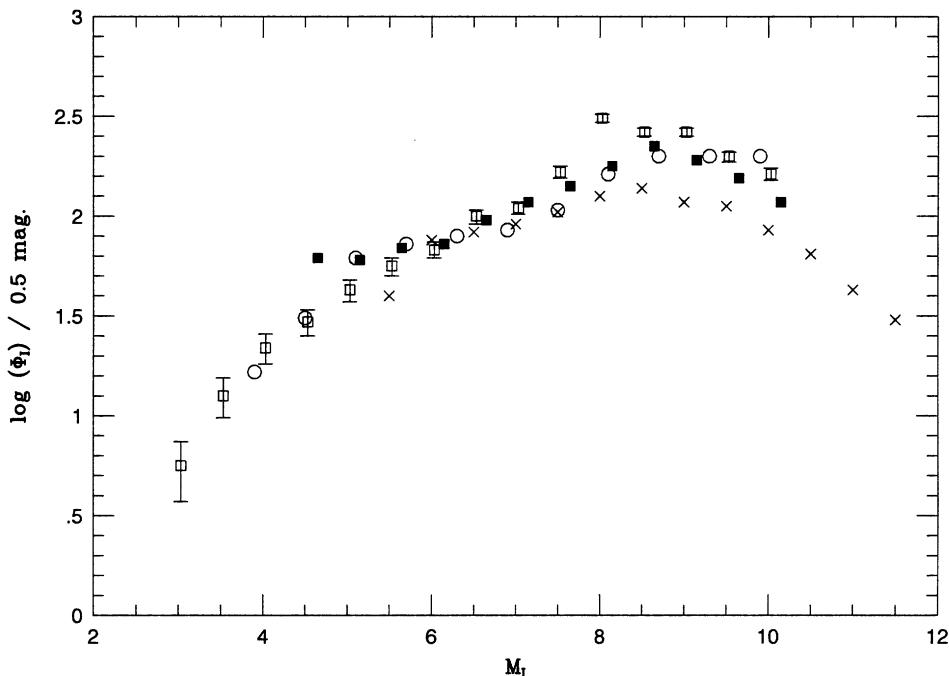


Figure 10.16. Luminosity functions for the metal-rich globular cluster 47 Tuc (solid squares [D7]) and the intermediate abundance clusters Ω Cen (open circles [P4]) and NGC 6752 (crosses [F1]). The luminosity function of the metal-poor cluster M15 (open squares [P2]) is plotted as a reference.

clusters plotted in the former figure are all metal-poor, with $[m/H] < -1.6$; the latter figure shows $\Phi(M_I)$ for M15 ($[m/H] \sim -2.15$) and several intermediate abundance clusters. Neither $\Phi(M_V)$ nor $\Phi(M_I)$ shows an obvious correlation between the morphology of the luminosity function and metallicity. Most luminosity functions peak at $M_I \sim 8.5$ or $M_V \sim 10$, with a subsequent relatively shallow decline in number density that appears to steepen noticeable at $M_I > 10$ and $M_V > 13$.

One cluster stands out from the rest, with a noticeably flatter luminosity function: NGC 6397. As noted above, this cluster lies at low Galactic latitude, at a height of ~ 1 kpc above the Plane. NGC 6397 ($[m/H] = -1.8$) follows an orbit that takes it through the Galactic Plane at a distance of only 2–3 kpc from the Galactic Centre, and reaches a maximum height of less than 1.5 kpc above the Plane [D9]. As a result, it is subject to frequent tidal shocks, which strip the outermost stars from the cluster. Since NGC 6397 has a relatively short dynamical relaxation time (~ 200 Myr), those stars are rapidly replenished from the cluster core, with mass segregation favouring low-mass stars as the replacements, and consequent systematic depletion of the lower main sequence. The other clusters plotted in Figure 10.15 and 10.16 follow Galactic orbits where the cluster spends relatively little time near the Plane, and, as a result, these clusters have suffered correspondingly less dynamical

attrition. Averaging the data for these systems is likely to provide a better estimate of the initial luminosity and mass functions.

10.6 THE MASS FUNCTION OF THE STELLAR HALO

The relationship between mass and luminosity depends on chemical composition. None of the binaries with known orbits listed in Table 9.2 has an abundance significantly below the solar value, so no empirical calibration of this dependence exists. However, several groups have recently computed sets of models of metal-poor dwarfs [D4], [A2], [B2]. Figure 10.17 compares their intermediate-abundance (mass, M_V) relationships, plotting the $[Fe/H] = -1.3$ results from [D4] with $[m/H] = -1.5$ data from the other groups. The different analyses produce similar results.

As discussed in Section 9.4, rapid changes in slope in the mass–luminosity relationship can introduce features into the mass function. The lower panel in Figure 10.17 compares the composite luminosity function (from Figure 10.12) against the first derivative of each of the three (mass, M_V) relationships plotted in the upper panel. Again, the three theoretical relationships are in reasonable agreement, but there is no obvious correlation between the maximum in dM/dM_V and the maximum at $M_V \sim 11.5$. Given the good agreement of the three theoretical analyses, further comparisons are restricted to the Lyon models [B2].

10.6.1 $\Psi_{\text{halo}}(M)$ in the field

Field subdwarfs span a substantial range in abundance. Figure 10.18 plots the (mass, M_V) and (mass, M_K) relationships predicted for abundances $[m/H] = -1.0$ to -2.0 ; the 1-Gyr solar abundance isochrone is plotted as a reference. The models preserve a consistent morphology, with decreased line-blanketing at lower abundances leading to brighter M_V for a given mass, while the mass limit for hydrogen-burning also increases. The effect is more pronounced at V than at K , as would be expected given the greater prominence of atomic and molecular features at optical wavelengths. If the metallicity of a subdwarf is well determined, this is not a significant issue; however, many stars contributing to recent luminosity function analyses lack spectroscopic abundance estimates. The consequent uncertainties in mass are negligible in the theoretical (mass, M_K) relation, but translate to uncertainties of 10–15% in mass using the M_V calibration. Unfortunately, while 2MASS provides near-infrared photometry for most subdwarfs discovered from the classical proper-motion surveys (Lowell, Luyten), many of the recent additions, particularly from SDSS, are too faint for reliable photometry. As a result, we still have to rely on the (mass, M_V) relation to calibrate $\Psi_{\text{halo}}(M)$.

In principle, field subdwarfs could be segregated by abundance to compute $\Psi_{\text{halo}}(M)$ for a range of metallicities; in practice, the total sample is too small, and the available data too inhomogeneous, to permit this approach. Faced with such

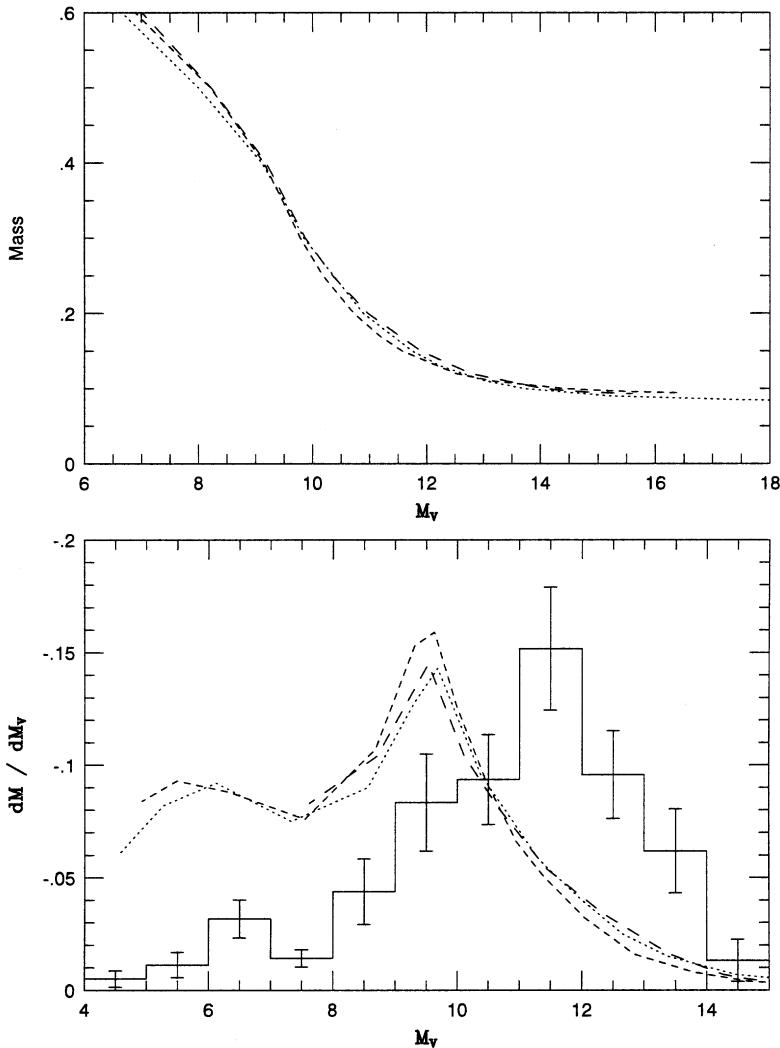


Figure 10.17. The upper panel plots ($\text{mass}-M_V$) relationships predicted by intermediate-abundance models from [D4] ($[\text{Fe}/\text{H}] = -1.3$, dotted line), [A2] ($[\text{m}/\text{H}] = -1.5$, dashed line) and [B2] ($[\text{m}/\text{H}] = -1.5$, long-dashed line). The results are in good agreement. The lower panel compares the first derivatives from those relationships to $\Phi_{\text{halo}}(M_V)$.

limitations, we can only estimate $\Psi_{\text{halo}}(M)$ for the field by applying a single mass-luminosity relationship to the composite luminosity function plotted in Figure 10.12. We use the Lyon (mass, M_V) relationship for $[\text{m}/\text{H}] = -1.5$, corresponding to the modal abundance of field subdwarfs (Section 7.3.4). The derived mass function is plotted in Figure 10.19, and is substantially steeper than the corresponding relationship for the Galactic Disk (Figure 9.9). Indeed, fitting the distribution

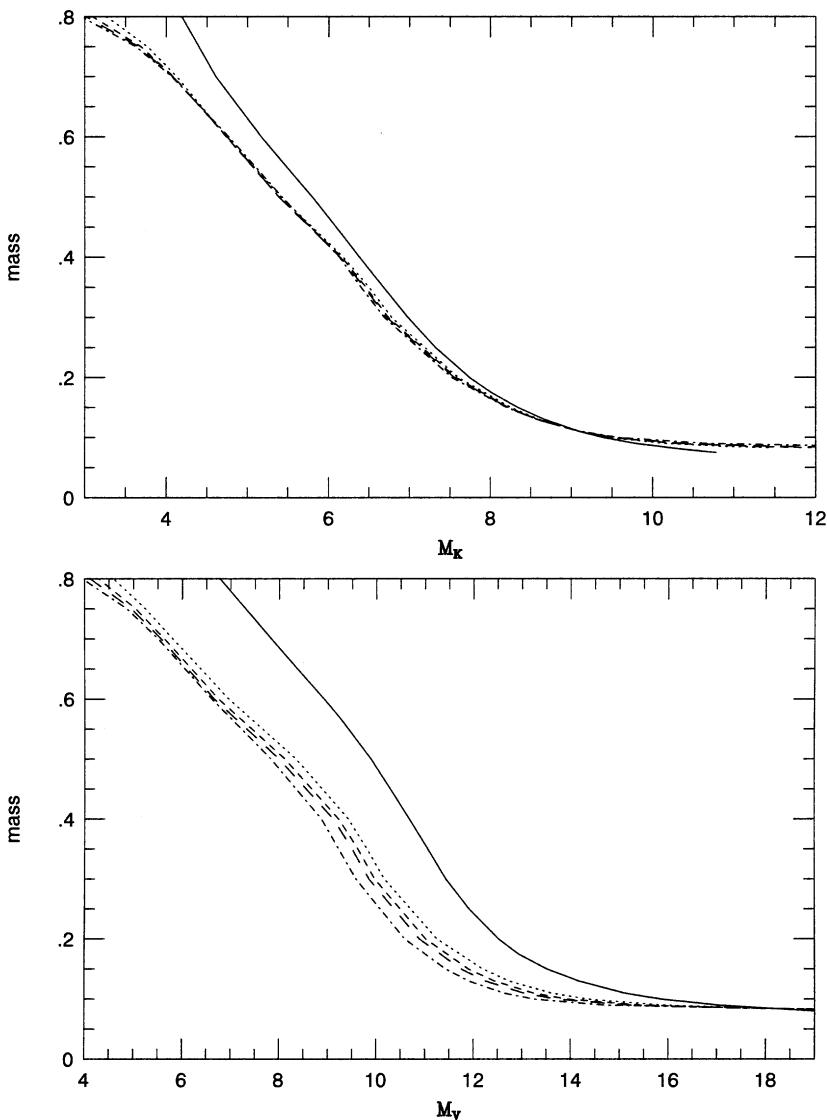


Figure 10.18. The (mass– M_V) and (mass, M_K) relationships predicted by the Lyon models for abundances $[m/H] = -1.0$ (dotted line), -1.5 (short dash), -1.8 (long dash) and -2.0 (dash-dot). The solar abundance relationship for $\tau = 1$ Gyr is shown as the solid line. Models are from [B2].

with a power law gives $\alpha \sim 1.5 \pm 0.25$ (where $\alpha = 2.35$ is the Salpeter value). The juncture between the [B1] and [D1] luminosity function data occurs at a mass of $\sim 0.45 M_\odot$, so the sharp increase below $0.2 M_\odot$ is not due to a mismatch in density zero-points.

10.6.2 $\Psi_{\text{halo}}(M)$ in globular clusters

The mass function for the Galactic Halo can also be estimated using globular cluster data. As with open clusters, globulars have the advantage that the system metallicity is well determined, so the appropriate mass–luminosity calibration can be applied; however, as with open clusters, dynamical evolution can distort the initial mass function (see the discussion of NGC 6397 in Section 10.5). Nonetheless, mass functions have been derived for over a dozen clusters (see [D7], [D8], [P2], [P3], [M1], [P4], [P5], [R6]. Figure 10.20 shows a representative subset of the derived mass functions, together with the mass function for halo field stars (from Figure 10.19). There is clearly some dispersion in the results, even allowing for uncertainties in the relative scaling. Matched against power laws, the mass functions for M4 and NGC 6397 are relatively flat, with $\alpha \sim 0.8$; M3 ($[m/H] \sim -1.5$) and M30 ($[m/H] \sim -2$) are significantly steeper, with $\alpha \sim 1.5$, close to the best-fit value for the field halo (Figure 10.19). We already noted that the luminosity function for NGC 6397 suggests that it has been subject to severe dynamical evolution. M4 also lies at low Galactic latitude and, while the luminosity function is comparable to the metal-poor clusters (Figure 10.15), it has higher metallicity; the mass function

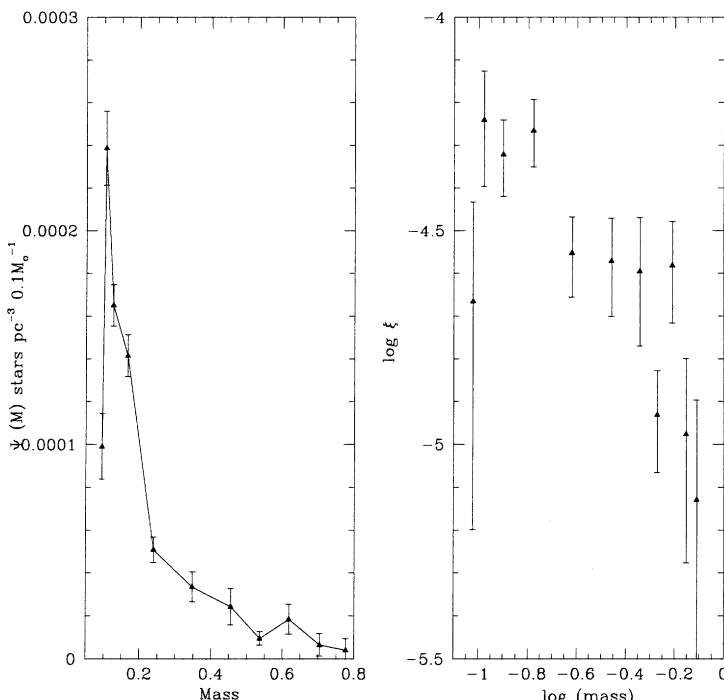


Figure 10.19. The halo mass function derived from field subdwarfs adopting the Lyon $[m/H] = -1.5$ mass–luminosity relationship. The left panel plots the linear function, dN/dM ; the right panel plots the logarithmic function, $dN/d \log M$.

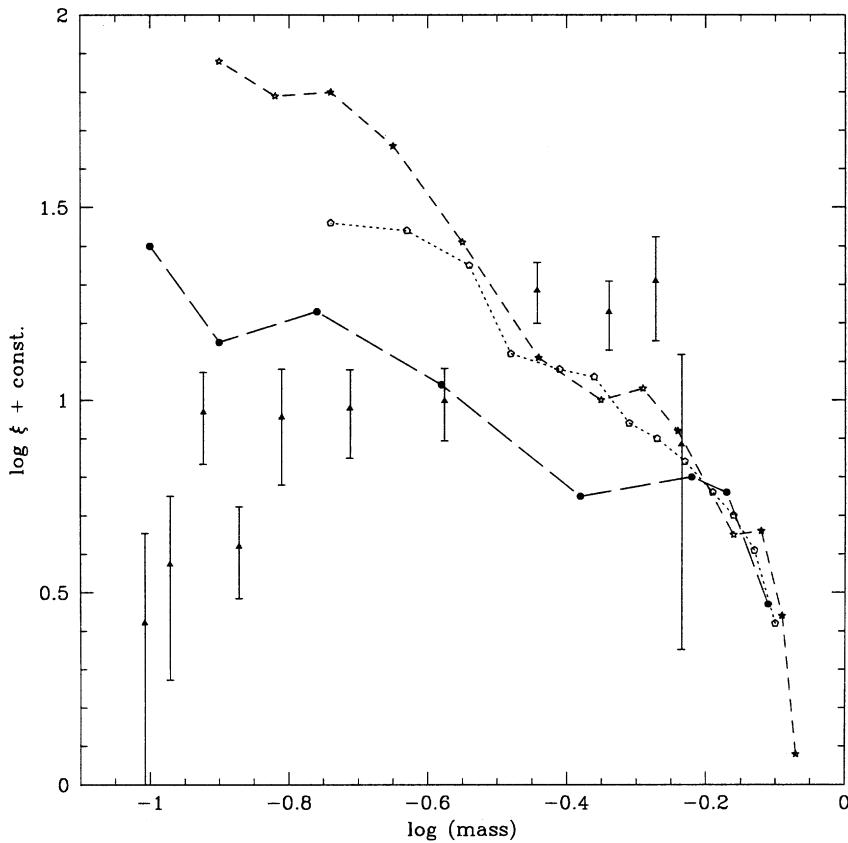


Figure 10.20. Present-day mass functions, $\Psi^{\text{PD}}(M)$, for the globular clusters NGC 6397 (solid points, [P2]), M3 (open pentagons, [M1]), M4 (crosses, [R6]) and M30 (star, [P2]). The field mass function is plotted as solid triangles. The cluster data have not been corrected for possible dynamical evolution.

shown in Figure 10.20 strongly suggests that it, too, has suffered extensive depletion of low-mass stars during its passages through the Galactic Disk.

The cluster data plotted in Figure 10.20 are present-day mass functions, representative of not just the current epoch, but also the particular cluster radii sampled by the observations. Dynamical evolution can lead to significant differences between that function and both the present-day *global* mass function and the initial mass function; these effects are known to be substantial for NGC 6397 (Figure 10.15), and Figure 10.20 suggests that M4 may also be affected. Reconstructing the initial mass function for those clusters is as fraught with uncertainty as reconstructing the Hyades mass function (see Section 9.6.1). However, the effects can be taken into account if only mild dynamical evolution is present, and the indications are that those circumstances hold at intermediate radii in most clusters.

Dynamical models can track the time evolution of mass segregation using a variety of methods, based on either direct N-body simulations [V1] or analytical techniques, such as the Fokker–Planck method [C4]. Stellar two-body interactions can disrupt a cluster through evaporation and/or core collapse, where the central density peaks sharply (such as in M15). Both evaporation and tidal stripping, due to shocks as the cluster passes through the Galactic Disk, lead to the total cluster mass decreasing with time; mass segregation promotes preferential depletion of low-mass stars in the outer regions.

The appropriate dynamical model for a given cluster is selected by comparing the predicted kinematics and density distributions (as a function of stellar mass) against observational determinations. Deriving these observational parameters is not straightforward. The density distribution for stars of different masses can be calculated from radial number counts of stars of different apparent magnitudes; but since we are observing a two-dimensional projection, some assumptions are required regarding isotropy before the three-dimensional distribution can be inferred. Similarly, while astrometry can, in principle, be used to determine the internal kinematics, the measurements require sub-milliarcsecond accuracy. Most current studies are based on radial velocity observations, which provide only a one-dimensional estimate of σ_V (and possible rotation) and are limited to the brighter cluster members, and therefore stars within a restricted mass range ($\sim 0.8 M_\odot$). Since the typical velocity dispersion near the centre of a globular cluster is only $\sim 10 \text{ km s}^{-1}$, high accuracy observations of large samples of stars are required.

Despite these stringent requirements, observational analyses have been undertaken of a number of clusters. These studies consider multi-mass models, in which the cluster stars are grouped in dynamical units by mass, with typically 20% of the total mass assigned to white dwarf remnants and a few percent to neutron stars. These analyses can be used to estimate both the total mass and the likely form of $\Psi^{\text{PD}}(M)$. As examples, [D10] derive a total mass of $\sim 6.6 \times 10^4 M_\odot$ for NGC 6397, and a mass function with the steep value of $\alpha \sim 1.9$ for $M > 0.4 M_\odot$, and the even steeper value $\alpha \sim 2.5$ for $0.1 < M/M_\odot < 0.4$. This is much steeper than the present-day mass function implied by direct observations of the low-mass cluster members (Figure 10.20). [D11] estimate a total mass of $\sim 5 \times 10^5 M_\odot$ for M15, with a mass function of slope $\alpha \sim 1.9$ to $\sim 0.3 M_\odot$, flattening sharply at lower masses. Finally, [P3] find that relatively small corrections for dynamical effects need to be applied to the slopes derived in their analyses of seven intermediate abundance and metal-poor clusters. It is significant that the strongest correlations derived in the last study are between α and dynamical parameters – the cluster destruction rate and the relaxation time. NGC 6397, with the flattest present-day mass function at low masses (Figure 10.20), has the shortest relaxation time and the highest disruption rate.

10.6.3 Summary

The halo mass function is not yet well determined. Analyses based on samples of nearby subdwarfs must allow for heterogeneous and poorly-determined abundances,

as well as relatively small sample-sizes and kinematic selection effects which underlie $\Phi(M_V)$; determinations from globular cluster data are vulnerable to the choice of distance scale and the effects of dynamical evolution. Both sets of analyses are forced to rely solely on theoretical mass-luminosity relationships, and while the available models are in good agreement, no empirical calibration is available to test for possible systematic errors. However, with all these caveats, current results suggest that the halo mass function is not identical in shape to the disk mass function. The halo mass function exhibits a steeper rise in number density between ~ 0.8 and $0.25 M_\odot$, but flattens and perhaps turns over at lower masses.

The mass density of the local halo, obtained by integrating the function plotted in Figure 10.19, is $7.8 \times 10^{-5} M_\odot \text{ pc}^{-3}$, or 0.26% of the mass density due to main sequence disk dwarfs. Stellar halo remnants, such as white dwarfs and neutron stars, probably contribute a further $2\text{--}3 \times 10^{-5} M_\odot \text{ pc}^{-3}$. If the mass function turns over at lower masses, the contribution from undetected halo brown dwarfs will produce only a marginal increase in this mass density.

10.7 M SUBDWARFS AND THE DISTANCE SCALE

In main sequence fitting, the distance to a star cluster is estimated by matching the cluster's colour-magnitude diagram against an absolute calibration defined by parallax stars. Traditionally, the latter stars lie on the upper main sequence, with spectral types no later than about K1. Only a small number of F and G subdwarfs have parallaxes determined to an accuracy of better than 10%, even after the completion of the *Hipparcos* astrometric survey. Since the halo mass function rises toward lower masses, the local volume encloses larger numbers of K and M dwarfs, offering the prospect of a more accurate definition of the main sequence relationship. Later type stars have been incorporated in a few distance studies of disk open clusters, but until recently their application to globular clusters was prevented by the necessity for both accurate abundance estimates for field stars and accurate photometry at faint magnitudes in the clusters. With the availability of atmosphere models [A3] to calibrate abundance and deep HST photometry for the clusters, main sequence fitting distances can now be measured using M subdwarfs.

NGC 6397 provides the best target for distance determination, since observations reach the hydrogen-burning limit. Lying at low Galactic latitude ($l = 338^\circ$, $b = -12^\circ$), the cluster is subject to substantial foreground reddening, but there is good agreement among the various indicators at $E_{B-V} = 0.18 \pm 0.02$. Thus, intrinsic colours can be calculated with a high degree of confidence, while the proper-motion study by King *et al.* [K2] provides effective elimination of field-star contamination.

NGC 6397 has a metal abundance of $[\text{Fe}/\text{H}] = -1.82$ [C2]. esdM subdwarfs are estimated as having abundance $\langle [\text{m}/\text{H}] \rangle \sim -2.0$, so these stars are well matched as distance calibrators. Figure 10.21 shows the results of fitting the cluster lower main sequence against 11 esdM subdwarfs with parallaxes measured

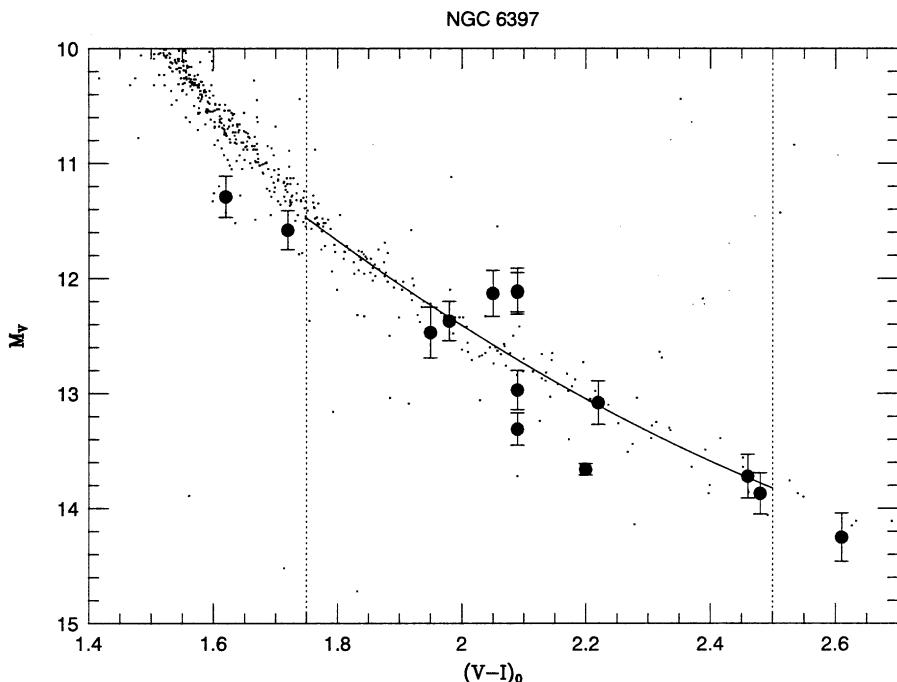


Figure 10.21. Main-sequence fitting, using M subdwarfs. The solid line is the mean relationship defined by the lowest-luminosity stars in NGC 6397; the distance modulus is derived by matching against field subdwarfs within the colour range isolated by the vertical lines. (From [R2], courtesy of the *Astronomical Journal*.)

to a precision of better than 10% [R2]. The significant dispersion in M_V among the field subdwarfs is not surprising, given that these stars must span a range of abundances. The derived distance modulus is $(m-M)_0 = 12.12 \pm 0.15$, which compares favourably with $(m-M)_0 = 12.24 \pm 0.1$ derived by matching the $(M_V, (B-V))$ cluster colour-magnitude diagram against F and G subdwarfs with Hipparcos parallax data. This is a technique in need of more refinement, but the results augur well for the future.

10.8 SUMMARY

As in the Galactic Disk, low-mass subdwarfs are the majority constituent of the halo. While the properties of these stars are less well understood than their higher-temperature counterparts, recent theoretical work, both on the atmospheres and on the internal structure, combined with the availability of better data, have opened the way to more quantitative analysis. Abundances can be estimated, both from the location of stars in the H-R diagram and from molecular bandstrengths, and the

abundance distribution appears to be consistent with that derived for more luminous subdwarfs. The overall luminosity function is well determined, although further work is required to establish both the overall mass function and the extent of any metallicity dependence. Present results suggest that $\Psi_{\text{halo}}(M)$ is somewhat steeper than the disk function over the range $0.8\text{--}0.25 M_{\odot}$, but may turn over at low masses and is poorly approximated as a power law. Current results in this area are at best provisional. Despite these uncertainties, however, and even given the current low-accuracy abundance estimates, we can use late-type subdwarfs as calibrators in main sequence fitting. The initial results from that analysis is in good agreement with the more conventional main sequence fitting studies based on F and G subdwarfs.

10.9 REFERENCES

- A1 Ake, T. B., Greenstein, J. L., 1980, *ApJ*, **240**, 859.
- A2 Alexander, D. R., Brocato, E., Cassisi, S., Castellani, V., Ciacio, F., Degl'Innocenti, S., 1997, *A&A*, **317**, 90.
- A3 Allard, F., Hauschildt, P. H., 1995, *ApJ*, **445**, 433.
- B1 Bahcall, J. N., Casertano, S., 1986, *ApJ*, **308**, 347.
- B2 Baraffe, I., Chabrier, G., Allard, F., Hauschildt, P., 1997, *A&A*, **327**, 1057.
- B3 Beers, T. C., Rossi, S., Norris, J. E., Ryan, S. G., Molaro, P., Rebolo, R., 1998, *Sp. Sci. Rev.*, **84**, 139.
- B4 Bessell, M. S., 1982, *Proc. ASA*, **4**, 417.
- B5 Burgasser, A. J. *et al.*, 2003, *ApJ*, **582**, 1186.
- B6 Burgasser, A. J., 2004, *ApJ*, **614**, L73.
- B7 Burgasser, A. J. *et al.*, 2002, *ApJ*, **594**, 510.
- C1 Carney, B. W., Latham, D. W., Laird, J. B., Aguilar, L. A., 1994, *AJ*, **107**, 2240.
- C2 Carretta, E., Gratton, R. G., 1997, *A&AS*, **121**, 95.
- C3 Casertano, S., Ratnatunga, K. U., Bahcall, J. N., 1990, *ApJ*, **357**, 435.
- C4 Chernoff, D. J., Weinberg, M., 1990, *ApJ*, **351**, 121.
- C5 Cool, A. M., Piotto, G., King, I. R., 1996, *ApJ*, **468**, 655.
- C6 Cottrell, P. L., 1978, *ApJ*, **223**, 544.
- D1 Dahn, C. C., Liebert, J. W., Harris, H., Guetter, H. C., 1995, in *The Bottom of the Main Sequence and Beyond*, ed. C. G. Tinney, Springer, Heidelberg, p. 239.
- D2 D'Antona, F., 1990, in *Physical Processes in Fragmentation and Star Formation*, ed. R. Capuzzo D'Locetta, C. Chiosi and A. Di Fazio, Kluwer, Reidel, Dordrecht, p. 367.
- D3 D'Antona, F., 1995, in *The Bottom of the Main Sequence and Beyond*, ed. C. G. Tinney, Springer, Heidelberg, p. 17.
- D4 D'Antona, F., Mazzitelli, I., 1995, *ApJ*, **456**, 329.
- D5 Dawson, P. C., 1986, *ApJ*, **311**, 984.
- D6 De Marchi, G., Paresce, F., 1995, *A&A*, **304**, 202.
- D7 De Marchi, G., Paresce, F., 1995, *A&A*, **304**, 211.
- D8 De Marchi, G., Paresce, F., 1996, *ApJ*, **467**, 568.
- D9 Dinescu, D. I., Girard, T. M., van Altena, W. E., 1999, *AJ*, **117**, 1792.
- D10 Drukier, G. A., 1995, *ApJS*, **100**, 347.

- D11 Dull, J. D., Cohn, H. N., Lugger, P. M., Murphy, B. W., Seitzer, P. O., Callanan, P. J., Rutter, R. G. M., Charles, P. A., 1997, *ApJ*, **481**, 267.
- D12 Digby, A. P., Hambly, N. C., Cooke, J. A., Reid, I. N., Cannon, R. D., 2003, *MNRAS*, **344**, 583.
- E1 Eggen, O. J., 1983, *ApJS*, **51**, 183.
- E2 Ellis, J., 2002, *Proc of International conference DARK2002*, eds. H. V. Klapdor-Klein-grothaus, R. D. Viollier, Springer, Berlin.
- F1 Ferraro, F. R., Carretta, E., Bragaglia, A., Renzini, A., Ortolani, S., 1997, *MNRAS*, **286**, 1012.
- G1 Gizis, J. E., 1997, *AJ*, **113**, 806.
- G2 Gizis, J. E., Reid, I. N., 1998, *PASP*, **109**, 1223.
- G3 Gizis, J. E., Reid, I. N., 1998, *AJ*, **117**, 508.
- G4 Gould, A., Flynn, C., Bahcall, J. N., 1998, *ApJ*, **503**, 798.
- G5 Gould, A., 2003, *ApJ*, **583**, 765.
- H1 Hartwick, F. D. A., Cowley, A. P., Mould, J. R., 1984, *ApJ*, **286**, 269.
- H2 Hambly, N. C., Davenhall, A. C., Irwin, M. J., MacGillivray, H. T., 2001, *MNRAS*, **326**, 1315.
- J1 Jones, D. H. P., 1973, *MNRAS*, **161**, 19P.
- K1 King, I. R., Sosin, C., Cool, A. M., 1995, *ApJ*, **452**, L33.
- K2 King, I. R., Anderson, J., Cool, A. M., Piotto, G., 1998, *ApJ*, **492**, L37.
- K3 Kirkpatrick, J. D., Henry, T. J., McCarthy, D. W., 1991, *ApJS*, **77**, 417.
- L1 Layden, A. C., Hanson, R. B., Hawley, S. L., Klemola, A. R., Hanley, C. J., 1996, *AJ*, **112**, 2110.
- L2 Lépine, S., Shara, M. M., Rich, M. R., 2002, *AJ*, **124**, 1190.
- L3 Lépine, S., Shara, M. M., Rich, M. R., 2003, *ApJ*, **585**, L69.
- L4 Lépine, S., Shara, M. M., Rich, M. R., 2003, *ApJ*, **591**, L49.
- M1 Marconi, G. *et al.*, 1998, *MNRAS*, **293**, 479.
- M2 Merritt, D., Meylan, G., Mayor, G., 1997, *AJ*, **114**, 1074.
- M3 Mould, J. R., 1976, *A&A*, **48**, 443.
- M4 Mould, J. R., McElroy, D. B., 1978, *ApJ*, **220**, 935.
- M5 Mould, J. R. *et al.*, 1996, **108**, 682.
- N1 Norris, J. E., 1986, *ApJS*, **61**, 667.
- O1 Ohman, Y., 1936, *Stockholm Obs. Ann.*, **12**, No. 3.
- O2 Oort, J. H., 1965, in *Galactic Structure*, ed. A. Blaauw and M. Schmidt, University of Chicago Press, p. 455.
- P1 Paresce, F., De Marchi, G., Romaniello, M., 1995, *ApJ*, **440**, 216.
- P2 Piotto, G., Cool, A. M., King, I. R., 1997, *AJ*, **113**, 1345.
- P3 Piotto, G., Zoccali, M., 1999, *A&A*, **345**, 485.
- P4 Pulone, L., De Marchi, G., Paresce, F., Allard, F., 1998, *ApJ*, **492**, L41.
- P5 Pulone, L., De Marchi, G., Paresce, F., 1999, *A&A*, **342**, 440.
- P6 Pulone, L., De Marchi, G., Covino, S., Paresce, F., 2003, *A&A*, **399**, 121.
- P7 Pasquali, A., De Marchi, G., Pulone, L., Brigas, M. S., 2004, *A&A*, **428**, 469.
- R1 Reid, I. N., 1984, *MNRAS*, **206**, 1.
- R2 Reid, I. N., Gizis, J. E., 1998, *AJ*, **116**, 2929.
- R3 Richer, H. B., Fahlman, G. G., Buonanno, R., Fusi Pecci, F., Searle, L., Thompson, I., 1991, *ApJ*, **381**, 147.
- R4 Richer, H. B., Fahlman, G. G., 1992, *Nature*, **358**, 383.
- R5 Richstone, D. O., Graham, F. G., 1984, *ApJ*, **277**, 227.
- R6 Richer, H. B. *et al.*, 2002, *ApJ*, **574**, L151.

- S1 Schmidt, M., 1975, *ApJ*, **202**, 22.
 S2 Sommer-Larsen, J., Zhen, C., 1990, *MNRAS*, **242**, 10.
 S3 Schweitzer, A. *et al.*, 1999, *A&A*, **350**, L62.
 S4 Scholz, R.-D., Lehmann, I., Matute, I., Zinnecker, H., 2004, *A&A*, **425**, 519.
 S5 Stoughton, C. *et al.*, 2002, *AJ*, **123**, 485.
 T1 *The 19th Texas Symposium on Relativistic Astrophysics and Cosmology*, 1998, eds. J. Paul, T. Montmerle and E. Aubourg, CEA, Saclay.
 V1 Vesperini, E., Heggie, D. C., 1997, *MNRAS*, **289**, 898.
 W1 Woolf, V. M., Wallerstein, G., 2005, *MNRAS*, **356**, 963.
 W2 West, A. A., *et al.*, 2004, *AJ*, **128**, 426.

10.10 HOMEWORK PROBLEMS

Problem 1

The Luyten New Two-Tenths (NLTT) proper motion survey has a proper motion limit of 0.18 arcseconds/year and an apparent magnitude limit of $V \sim 19$. Assuming a halo : disk ratio of 1 : 400, estimate the relative number of disk and halo stars in the NLTT with $M_V = 5$ and $M_V = 12$.

Problem 2

The HST Advanced Camera for Surveys has a field of view of ~ 10 square arcminutes. Modelling the halo density law as a 3,500-pc exponential with a cut-off at a distance of 50 kpc from the Galactic Centre, estimate the total number of halo stars in a single ACS field close to the Galactic Pole. Deep imaging observations with Hubble reach limiting magnitudes of $V \sim 29$ and $I \sim 28$; using those limits, estimate how many halo stars will actually be detected in the ACS observations (hint: estimate the distance limit for each magnitude bin of the halo luminosity function and sum the totals).

Problem 3

Assume that NGC 6397 has a true distance modulus of $(m - M)_0 = 12.2$ and an age of 10 Gyr; use the brown dwarf evolutionary tracks and colour-magnitude diagrams from Chapters 3 and 6 to estimate the likely I -band and K -band apparent magnitudes of the top of the brown dwarf sequence.

11

Extrasolar planets

Innumerable Suns exist; innumerable Earths revolve around those Suns in a manner similar to the way the seven planets revolve around our Sun.

De l'Infinito, Universo e Mundi ... Giordano Bruno

11.1 INTRODUCTION

Most astronomical investigations tackle narrow astrophysical questions, gradually pushing back the frontiers of astronomical knowledge. There are some issues, however, which have much wider ramifications. For example, observations of the binary pulsar PSR1913 + 16 provide stringent tests of general relativity [H7], [T1], while interest in the existence of brown dwarfs was stimulated in large part by their possible contribution to dark matter, and hence their relevance to galaxy formation. No study resonates as strongly as the search for planetary systems amongst the stars in our Galaxy. This fascination, both lay and scientific, stems from the obvious link with questions concerning the existence of extraterrestrial life and humankind's place in the Universe.

The latter decades of the 20th century saw the development of several programs designed to search for evidence for extraterrestrial life, notably SETI, the Search for Extraterrestrial Intelligence [L5]. Underpinning such projects, which seek civilisations similar to our own, is an equation devised by Frank Drake, one of SETI's pioneers:

$$N = R_* f_p n_e f_l f_i f_c L \quad (11.1)$$

where N is the number of civilisations capable of long-distance communication in the Galaxy; R_* is the formation rate of suitable stars; f_p the fraction of those stars with planets; n_e the number of terrestrial planets per system; f_l the fraction of terrestrial planets that support life; f_i the fraction of life-bearing terrestrial planets where

intelligence develops; f_c the fraction where technological civilisations are achieved; and L , the average lifetime of those civilisations.

Most factors in this equation are accessible only to indirect analysis, either purely theoretical (sometimes speculative) or laboratory-based simulations. It is instructive to recast the equation in a manner that separates these parameters from more observable factors:

$$N_l = N_* P_l P_p \quad (11.2)$$

where N_l is the number of life-bearing planets other than Earth at time T ; N_* the number of stars at time T , given by the convolution of $\Psi^l(M)$ and the star formation rate, with due allowance made for the stellar death rate; P_p the average probability of a star having a planetary system; and P_l the average probability of a planet sustaining life at time T . P_l collects in a single parameter such diverse factors as planetary mass, orbital semi-major axis and eccentricity, binarity (both stellar and planetary), age and stellar temperature, all of which may contribute to the likelihood of abiogenesis and the subsequent length of existence of any life forms.

The primary criterion for habitability is generally regarded as access to liquid water. This has led to the definition of the ‘habitable zone’, the circumstellar regions where planetary temperatures allow liquid water to be stable on the surface [K4]. (This definition also requires a planet that is sufficiently massive to retain a relatively dense atmosphere.) Under this definition, the habitable zone (HZ) in the Solar System, at the present time, extends from ~ 0.9 – 1.4 AU. However, there are additional factors that can influence habitability: atmospheric greenhouse effects can raise temperatures by more than 10°C , potentially extending the HZ to 2.4 AU, beyond the Martian orbit; salt water (brine) remains liquid at temperatures of -10 to -40°C ; and alternative heat sources may be present (e.g., Jupiter’s tidal heating is thought to maintain liquid oceans under the Europan ice cap). Thus, the range of habitable parameter space may be significantly larger than the simple habitable zone (see [K8], [K11], [D3], [W10] for more extensive discussion of these issues).

Philosophically, one could argue that there are only two interesting solutions to Equation (11.2): $P_l = 0$, $N_l = 0$, and Earth is a unique system; or $P_l > 0$, $N_l = \text{many}$. In other words, did life appear once, or many times? There is, of course, an intriguing subtext to the second option: how often does simple life (prokaryotic bacteria and archaea) turn into complex life (eukaryotic species)? This question has received considerable attention over the last few years, with widely divergent opinions on the matter (e.g., [W10], [D3]).

In any case, P_l remains largely a matter of theoretical conjecture (although the discovery of life on other bodies within the Solar System, such as Mars or Europa, would certainly have an impact on those conjectures). P_p on the other hand, is accessible to empirical determination, and the last decade of the twentieth century saw the first independently verified detections of planetary-mass companions to a variety of primary stars [M5], [M2]. None of these new discoveries is a rocky, Earth-mass planet within the habitable zone of its parent star, but the existence of numerous extrasolar gas giants lends more credence to the possible existence of

terrestrial analogues elsewhere – even in the local Galaxy. Here, we summarise the highlights of the various investigations; we refer the interested reader to Croswell’s [C8] excellent popular review of the subject, and the review articles by Koerner [K6], Marcy and Butler [M3] and Bodenheimer and Lin [B14], for more extensive discussion.

11.2 PLANET FORMATION

11.2.1 Classical scenarios

Until very recently, theoretical mechanisms for forming planetary systems were constrained by observations of one system – our own. The inherent pitfalls in dealing with what Wetherill [W7] terms ‘statistics of one’ were well appreciated by planetary scientists and cosmogonists. Nonetheless, models for the formation of our Solar System came to represent the standard paradigm for planetary formation.

The main characteristics of the Solar System are: first, a significant change in the mean composition with increasing distance from the Sun, with the terrestrial planets and asteroids at radii of less than 4 astronomical units (AU), gas giants at intermediate distances and icy planetismals at radii beyond 30 AU; second, nearly coplanar and low-eccentricity orbits for all nine planets and many lower mass objects; third, angular momentum vectors in both orbital motion and rotation that are well aligned with the direction of solar rotation. Taken together, these properties strongly suggest an origin within a disk formed by the collapse of the solar nebula. This concept of Solar System formation, first suggested by Kant [K3] and further elaborated by Laplace [L1], held sway during the nineteenth century, but was supplanted in the early years of the 20th century by the near-collision theory proposed originally by Buffon [B12], and revived by Chamberlin [C3] and Moultin [M7], and, later, by Jeans [J2] and Jeffreys [J3]. The latter hypothesis envisaged a close encounter with a passing star leading to a long spindle of material being drawn from the Sun by tidal forces, with the planets condensing from that spindle.

The near-collision model had (to some) the philosophical attraction of requiring a very rare, perhaps unique, event. In contrast, the nebular hypothesis renders planet formation part of the natural sequence in star formation. The latter hypothesis was revived by von Weizsäcker [W5], who postulated the formation of cellular vortices due to instabilities within the protoplanetary disk, partly as a means of accounting for the numerical progression of planetary semi-major axes known as Bode’s Law. The origin of modern theories, however, can be traced to Kuiper [K9], who not only suggested that the protoplanetary nebula was significantly more massive than the present-day sum of planetary masses, perhaps exceeding $0.1 M_{\odot}$, but also proposed that the gas giants are the result of gaseous accretion on solid protoplanetary cores. Safronov [S2] and Hayashi *et al.* [H2] further extended this concept of building planets through accretion of planetismals within a rotating disk of gas and dust. The identification in the 1980s of such disks associated with young T Tauri stars (see Section 3.6) provides strong support for this conceptual model.

Current Solar System formation models are well summarised by Lissauer [L12], Pollack *et al.* [P2] and Koerner [K6]. The original solar nebula is envisaged as having a diameter of ~ 100 AU and a mass exceeding $0.02 M_{\odot}$, based on adding cosmic proportions of hydrogen and helium to the current ‘metallic’ planetary masses. These estimates are consistent with masses and radii inferred from millimetre observations of T Tauri disks [B2]. HST near-infrared imaging shows that dust is present in the equatorial plane (Figure 3.20, see colour section); supplying the vital building blocks for planet formation.

Classical formation scenarios envisaged the planets forming by progressive accretion within a relatively quiescent solar nebula. Refractory silicate grains are expected to form as the temperature drops below 1,700 K, and their formation is expected to follow a well-ordered radial condensation sequence as the protoplanetary nebulae cools. The grains aggregate to form 1–10-km planetismals, which accrete material within the Hill sphere:

$$R_H = a_p \frac{M_p^{1/3}}{3M_*} \quad (11.3)$$

where a_p is the distance from the central star and M_p and M_* are the mass of the planetesimal and star, respectively. Simulations show that material within $4 R_H$ of the planetesimal can be perturbed onto eccentric orbits and eventually accreted [L11]. The planetismals grow to form $\sim 10^{-4} M_J$ (Jovian mass) planetary ‘embryos’ and, finally, terrestrial-mass planets [W6]. Recent observations of circumstellar disks show that they are turbulent, and grain formation is likely to progress in a less ordered, more stochastic fashion. Processes such as collisional adhesion [W3], [C9] are likely to play a vital role in the formation of planetismals that form the seeds for future planets.

The traditional method of forming gas giants is accretion onto a ‘super-embryo’, formed through the merging of 10–20 rocky embryos. The ambient temperature in the disk is expected to drop to < 100 K at radii exceeding 4 AU for solar-type stars. This allows ice to condense (mainly H_2O , but also CO_2 , CO , NH_3 , CH_4 and N_2), and the proto-gas giants accrete larger cores and massive envelopes [M6]. Envelope accretion, however, is predicted to require as much as 10^7 years [P2], a timescale that conflicts with the observed lifetime of optically thick disks in young stars (< 5 Myr, see Section 3.6.2). This issue is less of a problem for terrestrial planet formation, since the constituent planetismals are expected to form more rapidly, but clearly makes it difficult to reach Jovian masses. Moreover, since accretion is only effective at distances less than $4 R_H$, a planet on a circular orbit in an idealised quiescent disk will sweep out a low-density ring, and eventually run out material. Both of these difficulties led to suggestions that systems with Jovian-mass planets might be the exception, rather than the rule [W7]. The recent spate of planetary detections, however, shows that such systems are *not* rare and suggests that an alternative formation scenario may be required, as discussed in the following section.

Wetherill has pointed out that the existence of Jovian-mass planets at distances of several AU from the primary star may have a strong influence on the evolution of a solar system. Jupiter, and to a lesser extent Saturn, acts as a guardian of the inner

solar system, ejecting cometary-mass objects to large radii, and therefore reducing the potential for catastrophic Shoemaker–Levy-like collisions with the terrestrial planets. Uranus and Neptune perform similar tasks in the outer solar system, marshalling comets into the Kuiper and Oort belts. Thus, the development of complex life might be inhibited in a system lacking a Jovian-like planet at the requisite distance from the central star, since the higher rate of cometary and asteroid impacts could disrupt an ecosystem on relatively short ($\sim 10^4$ year) timescales.

11.2.2 Disk instability models

Boss [B8] has advocated an alternative process to the standard core accretion scenario, reviving the gravitational instability hypothesis proposed originally by Kuiper [K9] and later championed by Cameron [C10]. Under this mechanism, gravitationally unstable condensations form within the gaseous circumstellar disk, assuming the Toomre gravitational stability parameter:

$$Q = \frac{\kappa C_S}{\pi G \Sigma} \approx 1 \quad (11.4)$$

where C_S is the sound speed, κ the epicyclic frequency at some point on the disk, G the gravitational constant, and Σ the surface density. Once the condensations form, they collapse and grow in size as they accrete further material from the disk. Boss' calculations suggest that Jovian-mass (and larger) objects can form in a matter of only a few thousand years, well within the projected lifetime of circumstellar disks.

Planet formation *via* gravitational instability has also been explored by Mayer and collaborators, who have completed a sophisticated series of smooth particle hydrodynamical (SPH) simulations of protoplanetary disks [M8]. The simulations show that spiral instabilities tend to appear rapidly, and the disk fragments form dense clumps. These clumps merge to give a small number of protoplanets, which continue to accrete material from the disk and can acquire masses from 0.1 to ~ 7 Jupiter masses over a period of $\sim 1,000$ years. The disks themselves are disrupted on timescales of a few $\times 10^4$ years in some of the models, which is broadly consistent with observations of T Tauri stars, but may present challenges in supplying the ‘hot Jupiters’ in the current roster of extrasolar planets. However, the protoplanets often have orbits with significant eccentricities, leading to subsequent interactions and orbital evolution (see Section 11.4.2).

There are two main difficulties faced by this mechanism: first, forming the 0.02–0.03 M_J ice and rock cores that are required by models of the internal structure of Jupiter and Saturn; and, second, forming bodies that are smaller than 5–10 M_J . The first problem has been ameliorated to some extent by recent theoretical calculations, which suggest that gas giants may have core masses lower than standard values by a factor of three or more [G13]. However, the fractional metal content of the Solar System planets increases from ~ 4 times solar in Jupiter, to ~ 10 times in Saturn and ~ 40 times in Uranus and Neptune [W11]. While the gaseous fragmentation models can probably accommodate the lower mass cores, it is not clear that they can account for the total metal content of the outer planets. The second difficulty

arises because the gravitational instability mechanism is so efficient: again, forming gas giants as small as Uranus and Neptune is particularly problematic. Nonetheless, gravitational fragmentation offers a potential alternative to standard core accretion, and merits more extensive investigation.

11.2.3 Definitions: brown dwarfs vs. planets

As discussed in Chapter 6, it is likely that the brown dwarf mass spectrum extends to masses below $\sim 0.01 M_\odot$ ($11 M_J$), reaching what might be considered the planetary regime. How, then, does one distinguish between a low-mass brown dwarf and a high-mass planetary companion? The pivotal distinction between the two (despite the pronouncements of the IAU) rests with the mode of formation: brown dwarfs form by accretion within a giant molecular cloud in the same manner as hydrogen-burning stars; planets, by definition, form within the circumstellar disk of the protostellar nebula. That difference leads to a clear expectation of differences in chemical composition: as noted above, observations of Jupiter and Saturn suggest abundances of $Z \sim 0.02\text{--}0.06$ and $\sim 0.04\text{--}0.12$ respectively in the outer envelope, as compared with $Z \sim 0.02$ for the Sun. Unfortunately, even though there are plans for satellite missions that will provide images of the nearest systems (see Section 11.6), there are only very limited prospects of obtaining observations capable of measuring detailed chemical abundances.

In most cases, available observations of extrasolar planets allow only determination of a limited subset of orbital parameters, with the orbital inclination usually indeterminate. Kumar [in C8], Black [B6] and Boss [B9] have suggested that one can distinguish between the two types of companion on the basis of the orbital eccentricity, limiting ‘planets’ to objects in near-circular orbits. That definition, however, may well be biased by the ‘statistics of one’ argument – our Solar System may be atypical, and interactions between massive embryos have reasonable probability of scattering objects onto eccentric orbits [L11]. Both planets and brown dwarfs in close binary systems with stellar primaries are subject to similar dynamical interactions; the similarity in the distribution of orbital eccentricities therefore reflects a similar history of orbital evolution, rather than identical origins.

Given current observational limitations, it may be many years before we are able to do more than differentiate statistically between the two possible classifications.¹ Clearly, if one only has a measurement of $M \sin(i)$, any individual low-mass companion might be a brown dwarf (or even a low-mass star) in a low-inclination orbit. However, barring a cosmic conspiracy, the distribution of orbital inclination should be random, allowing a statistical estimate of the mass distribution. If the companions are predominantly brown dwarfs, one might expect a distribution that shows some continuity from the mass-function of companions above the hydrogen-burning limit; a distribution confined to near-Jovian masses, on the other hand, suggests a distinct, planetary origin. Current results (see Section 11.5) favour the latter alternative.

¹ The authors note that such predictions tend to be invalidated. We will happily accept the error if that proves to be the case for this prediction.

11.3 SEARCHES FOR EXTRASOLAR PLANETS

Most techniques used to search for planetary companions of main sequence and evolved stars are also used to search for stellar or brown dwarf binaries; planets, however, are 1–2 orders of magnitude less massive than a typical brown dwarf, with much lower luminosities, so the technical requirements are correspondingly stringent. In many cases, the observations only provide a mass ratio with respect to the primary star, either m/M_* or, more commonly, $m \sin(i)/M_*$, where m is the mass of the planet; M_* the mass of the star; and i the orbital inclination. If the distance of the parent star is known, M_* can be estimated from main sequence mass–luminosity relations or, if the primary is evolved, by comparing the position on the H–R diagram against theoretical isochrones (see [F2], [G9]). As will become clear in the following sections, most surveys have focused on solar-type stars.

11.3.1 Astrometric surveys

Planetary-mass companions are capable of introducing perturbations of no more than several milliarcseconds in the motions of even the nearest stars to the Sun. As an example, Figure 11.1 shows the residuals introduced in the Sun's motion by the planets in the Solar System, scaling those motions to a viewpoint 10 parsecs distant along the line of the ecliptic poles. Most of the ‘wobble’ is due to Jupiter, in its 11.86-year orbit; Saturn's contribution has approximately one-third the amplitude, with a period of 29.46 years, while Uranus and Neptune add further excursions, each

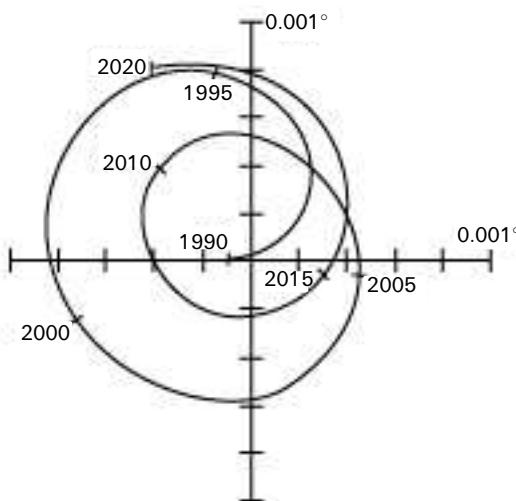


Figure 11.1. The astrometric displacement of the Sun produced by the Solar System planets, as seen from a distance of 10 parsecs. Jupiter is responsible for most of the residual motion, but the other planets (notably Saturn) are also influential, leading to the non-circular ‘wobble’.

one-quarter the amplitude of the Jovian wobble, and with periods of 84 and 169 years, respectively.

The semi-amplitude of the astrometric perturbation due to a single planet in a circular orbit is given by [G4]:

$$\begin{aligned} A &= \frac{m}{r} \left(\frac{P}{M_* + m} \right)^{2/3} \\ &= \frac{m}{M_J r} \left(\frac{P}{M_*} \right)^{\frac{2}{3}} \end{aligned} \quad (11.5)$$

where m is the planetary mass; M_J is the mass of Jupiter; M_* the mass of the primary star in solar units; r the distance in parsecs; and P the period in years. One can also express the detection limit m_d (planetary mass) for a given series of astrometric observations as:

$$m_d \propto \frac{\sigma r}{\sqrt{n}} \left(\frac{M_*}{P} \right)^{2/3} \quad (11.6)$$

where σ is the measurement uncertainty.

These equations show that the astrometric signature due to a planet, mass m , and, therefore, the likelihood of detection, increase with decreasing mass of the primary star and increasing semi-major axis (longer period). Orbital inclination, and hence mass, can be determined directly from the astrometric orbit. In practice, however, planets with $A > 1$ milliarcsec have periods of many decades, and accurate characterisation therefore requires observing programs of comparable duration. Additional obstacles to accurate interpretation of the observations are introduced if there is more than one massive planet, although the catalogue of observations acquired over the last decade suggest that a single planet has the largest dynamical influence in many systems (e.g., Jupiter is the dominant perturber within the Solar System – see Figure 11.1).

Despite the technical difficulties, astrometric searches for unseen companions have been undertaken for well over a century. There have been notable successes, such as Bessel's detections of the white dwarf companions to Procyon and Sirius, but there have also been many false alarms. The first purported detections of planetary-mass companions date to the 1940s, when Reuyl and Holmberg [R2] announced the identification of a 17-year period, $\sim 0.01 M_\odot$ companion to 70 Ophiuchi, while Strand [S6] claimed the discovery of a 4.8-year period, $0.008 M_\odot$ companion to one of the components of 61 Cygni. Similar claims were later made for, amongst others, AD Leo (Gl 388, [R3]), Lalande 21185 (Gl 411, [L9]), Stein 2051A (Gl 169.1A, [S7]) and G96-45 (GJ 1081, [B3]). All of these analyses were based on measurements of hundreds of photographic plates, spanning several decades. The hypothetical companions have orbital periods of 5–20 years and semi-major axes of 10–20 milliarcseconds. None has been confirmed by subsequent, more accurate observations ([L10], [H4]).

The classic case is Barnard's star (Gl 699), for which van de Kamp ([K1], [K2]) accumulated measurements of some 4,000 plates taken with the Sproul 24-inch

refractor between 1916 and 1974. Analysing those data, van de Kamp identified significant astrometric perturbations, which he ascribed to two Jovian-mass planets with orbital periods of 11.5 and 20–25 years respectively. Unfortunately, observations with the Allegheny 30-inch Thaw refractor, both photographic ([G2]) and with the photoelectric Multichannel Astrometric Photometer (MAP), fail to confirm the predicted motions. Similarly, observations made at McCormick Observatory by Fredericks and by Harrington with the US Naval Observatory 61-inch failed to detect significant perturbations, while Hershey [H5] identified similar wobbles in astrometric analyses of Sproul photographic plates of other (more distant) stars. It is likely that the original ‘detections’ stem from systematic positional errors, rather than true perturbations. The most recent observations limit possible planetary companions of Barnard’s Star to masses significantly below that of Jupiter [G6].

Current astrometric searches for planetary perturbations are based either on direct CCD observations [P3] or Ronchi-grating photometry [G5], with both techniques achieving accuracies of ~ 0.1 milliarcsec. At the time of writing, only one astrometric candidate still (barely) resists formal elimination: Gatewood [G7], reversing an earlier result of his own [G3], has claimed detection of perturbations in the secular acceleration S_a of Lalande 21185 (Gl 411). Secular acceleration is the change in proper motion, μ , arising from the changing heliocentric distance of a star:

$$S_a = \frac{2.05 \times 10^6}{r} \mu V_r \text{ arcsec yr}^2 \quad (11.7)$$

with r in parsecs and V_r , the radial velocity, in km s^{-1} : as a star gets nearer, the proper motion increases; as it recedes, μ decreases. This parameter can only be measured for the nearest stars with current technology, although future instrumentation will open wider vistas. In the case of Lalande 21185, the perturbations in S_a were attributed to two planets, 10 and $20 M_J$, but subsequent radial velocity measurements have failed to confirm any line-of-sight motions.

To date, astrometric surveys have had only a limited impact in searches for planetary companions. Astrometry, however, has played an important role in verifying the nature of other low-mass companions, as discussed further in Section 11.4.1, and several space missions are likely to make major contributions in the near future (see Section 11.6).

11.3.2 Photometric detections and direct imaging

Direct imaging

The simplest form of photometric detection of a planet would be an image of a nearby planetary system. Technically, however, this is an extremely challenging task, and current instrumentation is not capable of direct detection of Jupiter-like gas giants, even around the nearest stars. Planets shine by reflected light at optical

and near-infrared wavelengths, so their brightness scales as:

$$F_P = L_* A_P \frac{R_P^2}{r_P^2} \quad (11.8)$$

where F_P is the flux emitted by the planet, as detected at Earth, L_* is the stellar luminosity; A_P is the planetary albedo, the fraction of light reflected by the planet ($A_P = 0.4$ for Earth); R_P is the planet's radius; and r_P is the distance from the primary star. Viewed from beyond the Solar System, Earth (at its brightest) is fainter than the Sun by a factor of $\sim 2 \times 10^{-10}$, while Jupiter, five times more distant from the Sun, but with ten times Earth's radius and $A_P = 0.51$, is ten times brighter. Putting this in context, a hypothetical observer at a distance of 10 parsecs measures the apparent magnitude of the Sun as $V = 4.79$; Earth lies at a maximum separation of 0.1 arcsec with $V = 29.0$ (at 'full Earth'); Jupiter's maximum separation is 0.5 arcsec, and is no brighter than $V = 27$.

At mid-infrared wavelengths Jupiter is self-luminous, and the contrast (with respect to the Sun) drops to $\sim 10^{-4}$ (10 magnitudes) at 20–100 μm , offering better prospects for detection. Current detectors, however, lack the sensitivity and spatial resolution to permit surveys at those wavelengths, at least from the ground. The James Webb Space Telescope will offer the first serious prospect of searching the environs of nearby stars at these wavelengths. Other planned missions targeting the nearest planetary systems are discussed in Section 11.6.

One method of enhancing the chances of direct observations is to search for planets around young stellar systems. As with brown dwarfs, theoretical models predict that Jovian planets are substantially more luminous during the initial stages of formation. Several programs are currently underway, targeting young stars in the Solar Neighbourhood using ground-based AO imaging, HST observations and mid-infrared imaging with Spitzer. The most promising candidate uncovered to date is an L dwarf lying only ~ 0.8 arcsec from 2MASS 1207-3932, a young brown dwarf that is probably a member of the TW Hydreae association (see Section 7.6). Assuming an age of ~ 8 Myr and a distance of ~ 70 parsecs, 2M1207 is estimated to have a mass of $\sim 25 M_J$; the spectral type of the candidate companion, based on infrared spectroscopy, is $\sim \text{L}5$, which would imply a mass of $\sim 5 M_J$ [C11]. The L dwarf lies at a distance of ~ 60 AU from the primary – beyond the Kuiper belt in the Solar System. This places it at ~ 4 times the separation of the widest binary brown dwarfs (see Section 6.8.2), and well beyond the likely extent of the circumstellar disk of 2M1207. Further candidates are likely to emerge in the near future from these, and other, investigations.

Eclipses and transits

An alternative to direct imaging is a search for planetary transits of the primary star. Jupiter has 1% the surface area of the Sun, so one would expect a corresponding dip in the total flux of a solar-type star during the transit of a gas giant. If the primary is a mid-type M dwarf, the eclipse can be as deep as 10%. In order for these

circumstances to pertain, the orbital inclination must satisfy the condition:

$$\tan(i) > \frac{a}{R_*} \quad (11.9)$$

where a is the orbital semi-major axis; and R_* the radius of the star. The duration of the eclipse depends on the orbital period and R_* ; for transits across the centre of the star, the duration is:

$$\tau_{tr} = 13R_* \sqrt{\frac{a}{M_*}} \text{ hours} \quad (11.10)$$

where R_* is in solar radii with a in AU and M_* in solar masses. Thus, if we were to observe the Sun from a star lying in the ecliptic plane, we would see Jovian transits lasting ~ 29.6 hours every 11.86 years. Earth, with a diameter 1% that of the Sun, would produce annual eclipses of depth 0.01%, each lasting for 13 hours. Ideally, one should observe at least two (and preferably three) transits to verify the reality of the observations (unless other supporting evidence is available, see below). This sets a practical limit on searches for long-period (multi-year) systems; moreover, concerns about instrumental and atmospheric stability in ground-based systems are another barrier to searching for long-duration events (although see Section 11.6 for a discussion of the *Kepler* space mission). Consequently, this technique is best suited to detecting planets in short-period orbits. Note that, given the similarity in radii, this technique (in isolation) cannot distinguish between a gas-giant planet, brown dwarf or a very low-mass star as the potential companion.

How common are transiting systems? Defining $\tan(i_*) = a/R_*$, the probability that an orbit lies within the range $i_* < i < 90^\circ$ is given by $\cos(i_*)$. Following Marcy and Butler [M3], if we consider the case of a Jovian-sized planet orbiting a solar-type star at $a = 0.1$ AU, eclipses occur if the inclination exceeds 87.3° ; if orbital inclinations are random, then $\sim 4.7\%$ of systems should meet this criterion. Current results suggest that $\sim 3\%$ of G dwarfs have planets with $a < 0.1$ AU, suggesting that $\sim 0.1\%$ of all G dwarfs should experience this type of eclipse. As a comparison, if we consider a mid-type M dwarf (M2 to M4) as the primary, then $R \sim 0.2 R_\odot$ (Table 2.1), and the orbital inclination must exceed 89.25° for a transit at $a = 0.1$ AU. Again assuming random inclinations, only 1.7% of mid-type M dwarfs with suitable planetary companions are expected to meet this criterion; mid-type M dwarfs outnumber G dwarfs by a factor of $\sim 3:1$, but the overall frequency of M dwarf planetary systems has yet to be established.

Over the past few years, a number of dedicated transit surveys have been set in operation. These typically involve small telescopes with wide-field CCD cameras, monitoring hundreds to thousands of stars each night for photometric variations. Most surveys have yet to bear fruit, and the first transiting system, HD 209458, was discovered through photometric monitoring of a system already known, from radial velocity observations, to harbour a hot Jupiter [C4], [H8]. Since the primary is a 6th magnitude star, this system has received intense scrutiny, particularly using instrumentation on the Hubble Space Telescope; Figure 11.2 shows the HST light curve.

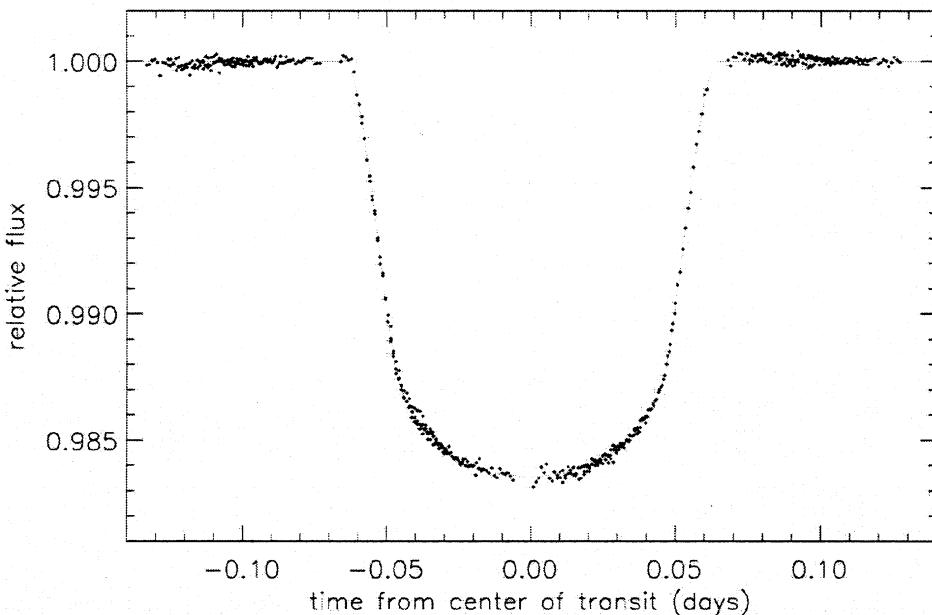


Figure 11.2. HST observations of the planetary transit on HD 209458 (from [B16], courtesy of the *Astrophysical Journal*). The concave light curve during eclipse is due to limb darkening on the host star.

These observations have resulted not only in accurate measurements of the radius of HD 209458b, but also direct detection of its atmosphere (see Section 11.5.3).

Subsequent to the identification of HD 209458 as a transiting system, a few candidates have emerged from the dedicated transit surveys (e.g., TrES 1, [A4]). In order to maximise the chances of detecting a transit, these programs need to cover tens of thousands of stars; consequently, this requires targeting fainter stars than those surveyed by the radial velocity programs. Hence, TrES-1 is a 12th magnitude K0 dwarf rather than a 6th magnitude G dwarf.

The most prolific source of transiting planets, however, is the OGLE microlensing project, which has discovered over a dozen candidates; most have proven spurious, primarily grazing eclipses by M dwarf companions, but a handful have been confirmed through radial velocity monitoring [B5]. OGLE's primary search areas are towards the Galactic Bulge, and the exposure times are optimised for relatively faint stars. As a result, the host stars in the transit systems are all 15th to 16th magnitude, much fainter than HD 209458, and lie at distances of several kiloparsecs. Consequently, while these systems provide valuable information on the frequency of planetary systems in a different region in the Galaxy, we know much less about the parent stars. In the near future, ESA's satellite COROT will target \sim 10,000 stars for detailed photometric monitoring, while *Kepler* will survey 100,000 A–K dwarfs in the outer Galaxy (see Section 11.6.1).

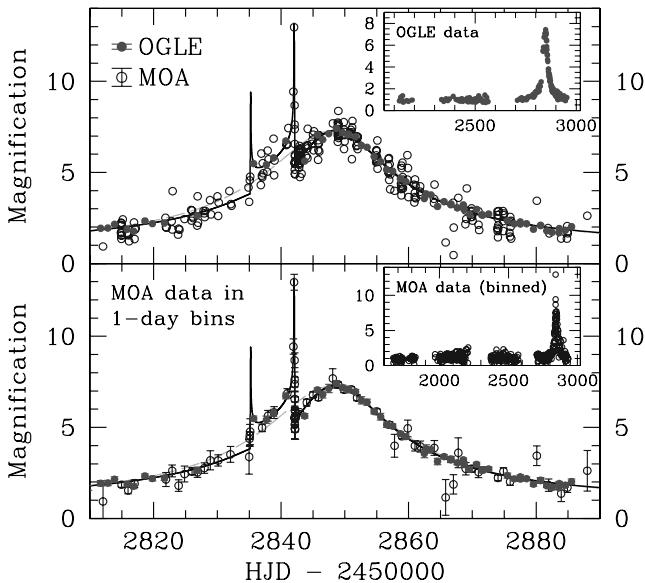


Figure 11.3. A microlensing planet: the figure plots the light curve of the OGLE 2003-BLG-235/MOA 2003-BLG-53 lensing event. The long-term curve is due to the lensing star; the sharp spikes at $HJD \sim 2835$ and 2842 are attributed to a planetary companion of that star (courtesy of [B15] and the *Astrophysical Journal*).

Microlensing

Section 9.3.3 summarised how microlensing can be used to estimate masses for isolated stars. The same technique can also be used to search for planetary companions: if the foreground lens is a star with one or more planets, those planets will also produce microlensing events, if there is appropriate alignment between the proper motion vector of the lens and the planetary orbital plane. This is basically equivalent to a lensing event produced by a high mass-ratio binary (or multiple) system. The net result is a light curve that shows a long-term symmetric variation due to the lensing star, with a short duration blip(s) superimposed as the Einstein ring of the companion planet(s) passes across the line of sight. Figure 11.3 shows the light curve for the sole candidate detected to date, OGLE 2003-BLG-235/MOA 2003-BLG-53 [B15]. This is believed to be due to a $1.5 M_J$ companion at a distance of ~ 3 AU from an M dwarf lying ~ 5 kpc from the Sun, towards the Galactic Bulge.

As with transit surveys, microlensing experiments allow one to probe the properties of planets around different types of stars in a different environment; in this case, the predominant lenses are K and M dwarfs lying in the inner Galaxy. However, there are also disadvantages: there is little prospect of obtaining any more information on the planets save the statistical inference of the mass and likely separation derived from analysis of the microlensing light curve. All of those quantities have substantial uncertainties that can only be beaten down through the acquisition of a

large catalogue of objects – and we have only one event after approximately a decade of microlensing surveys.

Protoplanetary disks and post-planetary debris disks

While direct detection of planets lies beyond the bounds of current technology, potential planetary environments have been identified. As described in Section 3.6, molecular gas has been detected in association with many young T Tauri stars, and there is now overwhelming evidence that the gas settles into rotationally supported circumstellar disks, obvious candidate sites for planetary formation. Millimetre observations are capable of tracing disk evolution until typical ages of 3–5 Myr, at which point the disk is sufficiently depleted of gas that it becomes optically thin and falls below current detection thresholds. It is only recently that it has become possible to trace the subsequent evolution through far-infrared observations of thermal radiation from the surviving dusty disk.

HR 4796 is a nearby ($r \sim 67$ parsecs) 10^7 year old A0 star that, like 2M1207-32, is probably a member of the TW Hydrae association; there is a wide M dwarf companion. Jura [J4] originally pointed out that IRAS observations reveal a substantial far-infrared excess, and suggested that this star has a residual dust disk. Mid-infrared observations [K7], [J1] have succeeded in resolving the excess as a dusty disk at an inclination of 72° and extending at least 100 AU from the central star. Significantly, the 20- μm flux distribution does not increase monotonically with decreasing radius, but is best fit by a disk with substantially reduced densities at radii of less than 55 ± 15 AU (that is, a central hole). That model has been confirmed by near-infrared observations with the NICMOS camera on HST, and a second system with similar characteristics, HD 141569 (spectral type A0), has been identified (Figure 11.4). This strongly suggests that relatively massive planetismals have already formed and are transferring material from the inner to the outer disk. Residual emission at smaller radii, detectable at 12 and 20 μm , suggests the presence of hotter (200–300 K) dust. The outer disk lies at radii comparable to the Kuiper belt in the Solar System, while the inner particles have properties reminiscent of zodiacal dust.

Debris disks represent a later stage in disk evolution, corresponding to ages between 10^7 and 10^8 years; the gaseous protoplanetary disk has largely dissipated, leaving a low-density dust disk. β Pictoris (see Figure 11.5, colour section), a nearby A star with an age of ~ 20 Myr, was one of the first systems to be identified [S4]. The dust has a spatial distribution comparable to the zodiacal disk in the Solar System [A3], with a radius of ~ 400 AU and a nearly exponential density profile perpendicular to the rotational plane. The infrared flux distribution implies the presence of grains spanning a wide range of sizes, from $< 1 \mu\text{m}$ to at least several millimetres [T2], [Z1]. The Poynting–Robertson effect and radiation pressure lead to orbital lifetimes of 10^3 to 10^6 years for smaller grains [A2], much shorter than the age of the primary star. This clearly suggests that the particles are being replenished continually, perhaps by collisions between larger planetismals [W4]. Similar considerations apply to almost all other stars with known debris disks.

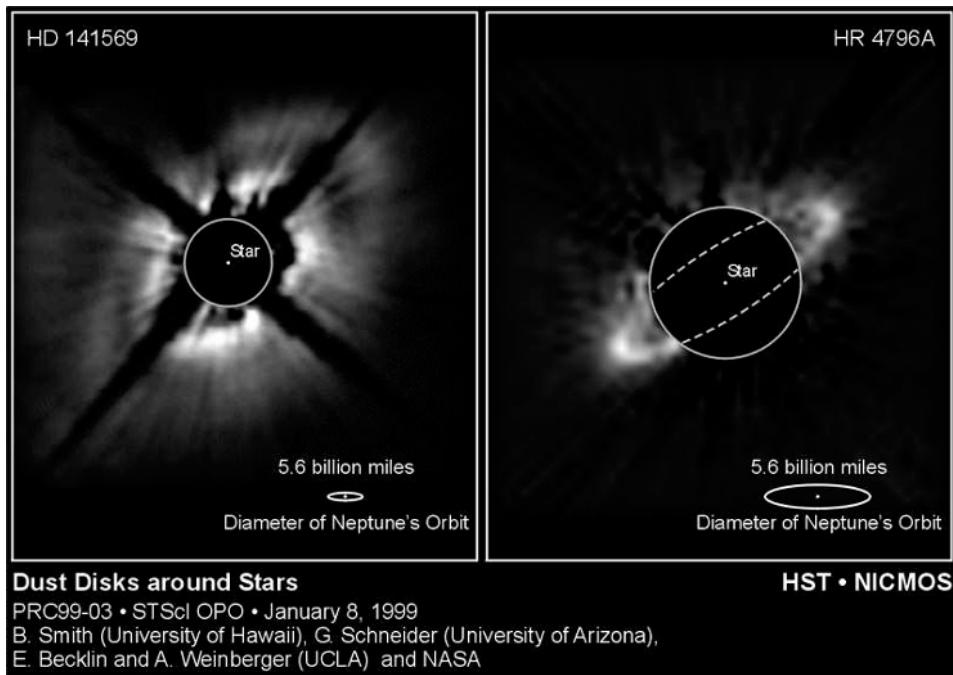


Figure 11.4. Near-infrared observations by the NICMOS camera on HST of dust disks around two nearby stars (courtesy of the Space Telescope Science Institute).

Mid-infrared observations of β Pic detect the $10\text{-}\mu\text{m}$ emission feature, showing that silicate grains are present in the disk. However, the high albedo in the outer regions of the disk (0.5 ± 0.2) indicates the presence of more reflective material, probably ‘dirty ices’, as found in Solar System comets. Most significantly, there is strong evidence that the complex structure evident in Figure 11.5 (see colour section) is due to the presence of at least four non-coplanar dust rings, at radii between 14 and 80 AU [W13], and for lower grain density in the inner portion of the disk ($r_p < 30$ AU). Both characteristics are best explained through gravitational interactions with several planetary companions. Indeed, rapid changes in the Ca II H and K absorption line profiles of β Pic have been attributed to planetismals passing in front of the stellar disk. Any major planets in this system are probably undergoing sustained bombardment from planetismals and comets, as happened to Earth during the early stages of the evolution of the Solar System.

Debris disks are easiest to detect around early-type stars; however, as far-infrared and sub-millimetre detector technologies have improved, dusty disks have also been detected around cooler, solar-type stars. Thus, ε Eridani is an $\sim 0.85 M_{\odot}$ K dwarf, age ~ 800 Myr, lying only 3.2 parsecs from the Sun (see Appendix). IRAS observations revealed a small far-infrared excess [A5], and sub-millimetre observations have confirmed the presence of extensive warm dust [G14]. The emission peaks

at distances between 35 and 75 AU from the star (comparable to the Kuiper Belt in the Solar System) and shows distinct asymmetries in brightness. Similarly, HD 107146 is a G2 dwarf with an age between 80 and 200 Myr. Sub-millimetre observations show that cool dust ($T \sim 50$ K) is present [W12], and optical imaging with the ACS coronagraph on HST (viewing reflected light, rather than thermal emission) reveals an almost face-on dusty disk with a central hole (Figure 11.6, see colour section).

Thus, most of the debris disks with detailed imaging share common morphological characteristics, notably a central hole and azimuthal asymmetries in the disk at larger radii. The most likely explanation of these features is embedded planetary systems. Indeed, a $0.86 M_J$ companion of ε Eri has been detected in an orbit with a semi-major axis of 3.4 AU [H9], and another planet at $a \sim 40$ AU is suspected to be responsible for the observed asymmetries in the dust ring. Reversing the search process, approximately 25% of stars known to have extrasolar planetary companions show 60–100 μm excesses, either in IRAS observations or in more recent Spitzer imaging. The most likely origin of this radiation is cool dust at radii exceeding ~ 40 AU.

Finally, M dwarfs have disks too. AU Mic (Gl 803) is a very young (10–20 Myr old) M0 dwarf at a distance of less than 10 parsecs from the Sun. AU Mic is one of the few M dwarfs detected by IRAS, and shows a pronounced excess at long wavelengths. Sub-millimetre observations at 450 and 850 μm confirmed the presence of cool dust [L15], and deep optical coronagraphic imaging showed that the dusty disk is almost edge-on to our line of sight and extends more than 200 AU from the central star ([K10] and Figure 11.6, see colour section). There is no evidence for excess radiation at wavelengths shorter than 25 μm , and this has been interpreted as evidence for an inner edge to the disk at a radius of 10–15 AU; as with higher mass stars, planets are suspected to be responsible for the lower density material in the inner disk.

11.3.3 Detection using Doppler techniques

Radial velocity surveys have proven to be the most successful method of identifying candidate planetary systems. Between 1994 and 2004, more than 130 probable extrasolar planets were identified around more than 110 nearby stars. However, the first planetary-mass companions, and the only terrestrial-mass objects found to date, were detected using a different type of Doppler observation: pulsar timing.

Pulsar planets

Pulsars are rapidly rotating neutron stars, the remnants of type II (massive star) supernovae. Initially rotating with periods of less than one millisecond, pulsars gradually spin down with time. During the early phases, abrupt changes can occur in the pulse period, prompted by changes in the internal structure, but such irregularities have not been detected in older systems, with periods of several milliseconds or

Table 11.1. The pulsar planets.

Planet	Period	a	$m \sin(i)$	e
PSR 1257+12a	25.262 days	0.19 AU	$0.015 M_E$	0.0
PSR 1257+12b	66.5419	0.36	3.4	0.018
PSR 1257+12c	98.2114	0.47	2.8	0.026
PSR 1257+12d	~ 170 yr	40	100	?
PSR 1257+12e?	~ 3 yr?	?	cometary mass?	?
PSR 1620+26c	~ 100 yr	23	$2.5 M_J$?

more. The latter sources serve as remarkably accurate celestial clocks, where binary motion reveals itself as systematic periodic residuals in the pulse arrival.

The first ‘detections’ of pulsar planets all proved to be spurious, with the most plausible (PSR B1529-10: [B1]) arising from data reduction errors: adopting the wrong position for the pulsar and failing to allow for the eccentricity of the Earth’s orbit led to incomplete correction for the Earth’s orbital motion ([L14], see also [C8]). Almost simultaneously with the retraction of that ‘detection’, Wolszczan and Frail [W9] announced the discovery of two very low-mass companions to PSR B1257+12. Subsequent observations have not only confirmed their existence, but also detected the signature of their mutual gravitational interaction and identified a third planetismal. All three are on short-period, near-circular orbits and have masses closer to those of the terrestrial planets than the gas giants (Table 11.1). Two additional longer period companions are suspected to be present, including one likely to have a mass comparable to Solar System comets.

The second pulsar planetary system is particularly interesting because it is a member of the globular cluster M4 ($[m/H] \sim -1.4$). PSR 1620+26 has a white dwarf companion, detected through HST observations, and a planetary-mass companion, detected through secular changes in the pulsar orbital motions [S10]. Sigurdsson *et al.* suggest that the planet was originally in orbit about the white dwarf, when it was an isolated main sequence star, but dynamical interations within the cluster led to an exchange reaction with a neutron star binary. The planet and its primary (the white dwarf in the present system) were captured by the neutron star, while the latter’s original companion was ejected from the system. This scenario is somewhat speculative, but, regardless, confirmation of the presence of a planetary-mass object in M4 is important, given the apparent preference for metal-rich systems among the Doppler planetary systems in the Solar Neighbourhood (see Section 11.5.4).

The origin of the pulsar planets remains unclear. Given the relatively short lifetime of the supernova progenitor, the standard model of formation within the protostellar disk may not apply. Moreover, one would expect the substantial mass-loss during the supernova event to have severely disrupted any planetary system. Thus, it is possible that these terrestrial-mass objects originate in a ‘born-again’

circumstellar disk, formed either by material accreted from a now-defunct companion, or during the merger of two white dwarfs which gave rise to the parent neutron star. If either of these scenarios holds, the pulsar planets may have little to do with conventional planetary systems.

Spectroscopic radial velocity surveys

Solar-type stars constitute the majority of the targets of high-precision radial velocity programs. In part this reflects the underlying aim of finding planetary systems like our own, but technical considerations also favour observations of these stars. The reflex motions induced by planetary companions are not easy to detect, and require high signal-to-noise spectra with reliable velocity signatures (i.e., bright stars with lots of strong, narrow absorption lines).

The reflex radial velocity induced by a companion of mass m is given by:

$$\begin{aligned} K &= \left(\frac{2\pi G}{P} \right)^{1/3} \frac{m \sin(i)}{(M_* + m)^{2/3}} (1 - e^2)^{1/2} \\ &= \left(\frac{Ga}{(M_* + m)} \right)^{1/2} (1 - e^2)^{1/2} m \sin(i) \end{aligned} \quad (11.11)$$

where K is the semi-amplitude of the velocity variation, and the other symbols have their usual meaning. In most cases, $M_* \gg m$, so $(M_* + m) \approx M_*$. Considering the Solar System, Jupiter alone produces a near-sinusoidal variation of $\pm 12.5 \text{ m s}^{-1}$ in the radial velocity of the Sun as measured from a vantage point in the plane of the ecliptic. Hence, observational precision better than 4 m s^{-1} , maintained over a period of ~ 10 years, is required for a 3σ detection of Jovian analogues in other solar-type stars.

Acquiring velocity measurements of such high precision demands high signal-to-noise spectra; even with 10-m telescopes, this limits observations to bright ($V < 12$ magnitude) stars, favouring main sequence stars earlier than spectral type K or giant stars. Evolution beyond the main sequence can disrupt or destroy existing planets, while many giant stars pulsate, complicating detection of orbital motion; this makes giants less favourable targets for planetary searches. High-velocity accuracy also requires the presence of many narrow lines in the stellar spectrum; this excludes stars hotter than $\sim 8,000 \text{ K}$ (A-type or earlier), which have few metallic absorption features, and tend to have rapid rotation, leading to broad lines and poorly determined line centroids. Hence, F, G and K dwarfs are selected naturally as the prime targets for planet searches, with a few bright M dwarfs also within reach of high-precision measurement.

Planetary detection also demands spectrographs that can achieve long-term instrumental stability better than 3 m s^{-1} . Section 1.7 outlines the main technical characteristics of spectrographs. The usual mode of observation is to intersperse programme star exposures with spectra of emission-line calibration lamps, usually combinations of He, Ne, Ar, Hg and Th. While suitable for most purposes, systematic errors in wavelength calibration at the $\sim 100 \text{ m s}^{-1}$ level can arise from a number

of sources, including time-dependent physical flexure of the spectrograph; thermal expansion of optical and mechanical elements; and differences in the optical beam due to the way that the star (a point source) and the arc lamp (a diffuse source) illuminate the slit.

Most potential mechanical distortions are eliminated if the spectrograph is located in a separate, temperature-controlled room, rather than on the telescope itself. In practice, high-resolution spectrographs designed for 8/10-m class telescopes have room-sized dimensions, and it is therefore not possible to mount them directly on the telescope. Most optical distortions can be taken into account if the wavelength calibration is determined simultaneously with the programme star observations. Several techniques have been used: adapting a method used in photographic spectroscopy, Mayor's team [M5] use optical fibres to feed light from both the target star and a calibration lamp to adjacent positions on the slit, allowing simultaneous observation. The use of fibres for both star and calibration source ensures similar illumination of the slit, and radial velocities can be measured to an accuracy, $\varepsilon_v \sim 4 \text{ m s}^{-1}$.

Comparable precision can be achieved if the wavelength calibration is imprinted directly on the stellar spectrum. The simplest approach is to take advantage of the natural grid of absorption lines due to H₂O, O₂ and OH provided by the Earth's atmosphere, but natural variations in temperature and pressure (in our atmosphere) limit velocity uncertainties to $\varepsilon_v \sim 20 \text{ m s}^{-1}$. More stable calibrations result from introducing a gas-filled cell, usually either hydrogen fluoride [C1], [C2] or iodine [M1] into the spectrograph beam immediately below the slit. This superimposes a fine grid of narrow absorption lines onto the stellar spectrum, allowing the radial velocity to be determined to an accuracy of $\varepsilon_v \sim 1\text{--}3 \text{ m s}^{-1}$ [B13]. An alternative technique, which may see more use in the near future, is dispersed fixed-delay interferometry [G19]. Originally proposed by Erskine [G20], this method couples a Michelson interferometer with a low-dispersion spectrograph, producing a long-slit spectrum whose intensity is modulated by a large-scale fringing pattern, produced by the stellar absorption lines. Doppler shifts in the absorption lines produce corresponding phase shifts in the fringing. The main advantage of this method over traditional echelle spectroscopy is that optical fibres can be used to observe multiple objects simultaneously. An initial test run with a prototype instrument has succeeded in detecting the planetary companion to 51 Peg [V1].

Finally, the ultimate limit on velocity accuracy is set by the intrinsic stability of the photosphere of the target star. Saar *et al.* [S1] have completed a detailed series of observations of F, G and K main sequence dwarfs, and find that the r.m.s. dispersion of residuals about the mean velocity is well correlated with the measured rotation period, $\sigma \propto P^{-1.1}$. Stars with rotation periods of 10 days typically have $\sigma \sim 10 \text{ m s}^{-1}$, with the intrinsic dispersion rising to $\sigma \sim 40 \text{ m s}^{-1}$ at rotation periods of ~ 3 days. These apparent velocity variations are generally ascribed to starspots, plages and other symptoms of stellar activity, which introduce changes in the stellar line-profile by changing the flux distribution over the stellar surface. Similarly, convective motions in giant stars can lead to velocity 'jitter' exceeding 0.5 km s^{-1} . The added noise not only makes it more difficult to detect orbital motion, but, as discussed

Table 11.2. Radial velocity searches for planets.

Survey	N (stars)	Duration (yr)	ε_v (m s^{-1})	Detections	Status
Walker <i>et al.</i> [W1]	21	12	13	0	Complete
Elodie [P5]	320	11	15	17	Continuing
Coralie [Q1]	1,650	6	3–7	38	Continuing
ESO-HARPS [P6]	TBD	1	< 2	2	Continuing
Lick [C12]	120	15	5–10	27	Continuing
Keck/Marcy [V2]	530	10	1–4	30	Continuing
AAO [T4]	200	6	3	17	Continuing
Keck/subdwarf [S11]	50	2	< 5	0	Continuing
McDonald/HET [C13]	170	3	1–3	5	Continuing
N2K [F5]	14,000	1	7–10	1	Continuing

further in Section 11.4.1, can sometimes mimic the periodic velocity variations produced by planetary-mass companions. Radial velocity surveys generally aim to minimise such problems by selecting stars with low-chromospheric activity (weak Ca II H and K or H α emission); effectively, this biases samples to stars older than ~ 1 Gyr. In addition, known stellar binaries with separation of ~ 10 AU or less have received little attention so far from current surveys; not only are those systems likely to have enhanced activity, but the presence of a massive object at small separation leaves little scope for stable planetary orbits.

Table 11.2 gives a synopsis of the major radial-velocity based planet searches undertaken over the last decade. Almost all of these programs are still underway; in some cases, the target list is still evolving, while, in others, the projects have only recently started acquiring observations. The duration listed in the table is with reference to the time of writing (December, 2004). Several planets have been observed and confirmed by more than one observing program.

Walker *et al.* [W1] completed the first survey that reached accuracies capable of detecting planetary companions. Their analysis of 21 stars failed to find solid evidence for planetary companions, although it now seems that one of their targets, γ Cephei, may indeed harbour a planet [H11]. Just as their negative results were published, Mayor and Queloz [M5] announced the identification of a companion to 51 Pegasi – 51 Peg b (planetary companions are denoted by lower case letters to distinguish them from stellar/brown dwarf companions).

Confounding expectations, 51 Peg b is a Jovian-mass planet in an intra-Mercurian orbit: the radial velocity variations have a period of only 4.23 days and a semi-amplitude of $\sim 60 \text{ m s}^{-1}$, well above the detection limit of previous surveys; the inferred mass $m_2 \sin(i)$ is $0.47 M_J$; and the projected semi-major axis is only 0.05 AU, or one-sixth the radius of Mercury's orbit. Both Hipparcos astrometry and subsequent interferometric observations [B7] rule out a stellar companion in a low-inclination (near face-on) orbit. A gas giant in such a short-period orbit was most unexpected: the surface temperature at that distance from the G5 primary is

~1,300 K, but there is insufficient mass in refractory elements in typical protoplanetary disks to construct a terrestrial planet of that mass.

In the decade following this seminal discovery, planetary companions of more than 120 other stars have been identified, including at least 14 systems with two or more planets and 30 ‘hot Jupiters’, with semi-major axes smaller than Mercury’s orbit. With the passage of time, and improvements in observational accuracy from 5 to 3 to 1 m s^{-1} , radial velocity programs are identifying planetary companions with longer periods and larger semi-major axes. At the present time, the surveys have just become capable of detecting true Jupiter analogues, and are starting to find companions with masses comparable to Neptune or Uranus, albeit in much shorter period orbits.

The initial impression might be that the flood of detections reflects the culmination of decade-long observing programs, but only ~25% of the currently known systems have periods longer than 2 years, while a comparable number have periods less than 2 months. Moreover, the detection of 51 Peg was based almost entirely on observations with the Elodie echelle spectrograph, mounted on the Haute-Provence 1.93-m, and several subsequent discoveries by Marcy and Butler used data obtained with the Hamilton echelle, mounted on the Lick Observatory 3-m telescope. Thus, contrary to expectations, the technological hardware – and the actual observations – for detecting extrasolar planets has been available for over 20 years. The main breakthrough came with the availability of greater computing power and more sophisticated analysis programs, combined with the knowledge that gas giants exist in short-period orbits.

11.4 INTERPRETING THE RESULTS

11.4.1 Alternative interpretations: pulsations, brown dwarfs or spots?

Stellar radial velocity variations can be produced by several mechanisms other than binary motion, including surface features (spots) produced by magnetic activity, convective motions and stellar pulsation. All of these possibilities were considered as alternative explanations for the radial velocity variations observed in 51 Peg. Moreover, the fact that, in most cases, observations only allow determination of the projected mass $m \sin(i)$ has led to persistent suggestions that the companions are brown dwarfs (albeit, of very low mass), rather than bona fide planets. None of these alternatives is a realistic model for 51 Peg, or most other subsequent discoveries.

Stellar pulsations can, in principle, give rise to long-term, stable variations. Gray [G10], [G12] noted apparent profile variations in the Fe I 6,253 Å line in 51 Peg that were not predicted by the binary model, but would be consistent with stellar oscillations. There are, however, severe problems with this mechanism as an explanation for the overall stellar behaviour.

Stellar pulsations occur in two modes: p-modes, or pressure (acoustic) waves; and g-modes, or gravity (buoyancy) waves [G8]. Purely radial pulsations can be ruled out, since, for 51 Peg, these would lead to variations of 10–15% in the stellar radius

(and concomitant effective temperature changes) that would be detected as periodic photometric variations. All G stars with potential planetary-mass companions with periods of less than 10 days have been monitored extensively, and (except for HD 209458, see Section 11.5.3) photometric variations of more than 0.2 millimagnitudes can be ruled out [H3]. Thus, any pulsations must be non-radial, with no significant global radius or temperature changes.

There are also significant problems involved in matching the observed periodicities. The frequency of harmonic oscillations for solar-type stars corresponds to periods of only 3 to 8 minutes for p-mode waves (as in the 5-minute Solar oscillations discovered by Leighton *et al.* [L4]), and only \sim 165 minutes for g-modes. Matching those variations against the observed periods requires the dominant mode of oscillation to be a high overtone. Willems *et al.* [W8] calculate that the 115th overtone in g-mode oscillations is required to match the 51 Peg 4.23-day period. That oscillation should be accompanied by oscillations at other frequencies; the additional oscillations are not present. Matching longer period variations, which now exceed 5 years, requires even higher overtones, and is correspondingly even less viable. Finally, subsequent high-resolution spectroscopy has failed to confirm the original hypothetical profile variations [B11], [G11], [M3], and pulsations have been effectively excluded as a reasonable alternative explanation for any planetary candidates. Convective motions are also unable to match both the range of periods of the extrasolar planetary systems and the long-term stability of those periods.

Surface features, such as star spots, can lead to spurious ‘detections’ of extrasolar planets. Stellar rotation contributes to the shape of the line profile; star spots are darker than the average photosphere; thus, as a spot moves from the approaching limb to the receding limb, it effectively changes the line profile. If there are relatively few spot systems, this can lead to a systematic change in the measured velocity; and if the spot systems are sufficiently long lived, this can give the appearance of orbital motion. G and K dwarfs have rotation periods less than \sim 30 days, so this mechanism cannot account for longer period variations. Shorter period systems can generally be identified, since spots also produce systematic variations in the level of chromospheric activity (e.g., Ca II H & K emission) with the same periodicity; moreover, observations at well-spaced epochs will usually lead to velocity variations of the same period, but with a significant phase shift, as new groups of star spots, at different longitudes, dominate the line profiles. For example, velocity variations detected in HD 195019 [F1] and HD 219542 [D1] are now known to have been the result of activity, not planetary companions.

Finally, brown dwarfs have been invoked in two ways as a possible alternative to planetary companions: first, since radial velocity surveys only measure $m_2 \sin(i)$, it is possible for a brown dwarf in a nearly face-on orbit to mimic a planetary-mass object; second, it has been proposed that the newly discovered planets are simply the low-mass extension of the companion star/brown dwarf mass function. However, the observed mass distribution of companions argues strongly against the second hypothesis, as discussed further in Section 11.5.1, and both statistics and direct observations exclude the first.

Han *et al.* [H10] argued that *Hipparcos* astrometry of several planet-host stars

shows evidence for orbital motion, as would be expected for a high-mass object in a face-on orbit. However, more rigorous astrometric analysis showed that they underestimated the inherent uncertainties in the *Hipparcos* data [P4], [Z2]. Moreover, it is remarkably unlikely that there are ~ 130 brown dwarfs in near face-on ($i < 5^\circ$) orbits, but less than a handful at higher inclinations. In fact, astrometric orbits for a few systems have been determined using HST, and the results confirm that the companions have very low mass. For example, astrometry of Gl 876 shows that the primary has a reflex motion with semi-major axis ~ 0.25 milliarcsec and an orbital inclination of $\sim 84^\circ$ (almost edge-on), confirming that the companions have near-jovian masses [B4].

To summarise, stellar pulsations are unable to match even the shortest period extrasolar planet candidates; convective motions are not viable, since they are not capable of matching the periodic nature of the velocity variations; chromospheric activity can produce cycle velocity variations with periods less than ~ 30 days, but cannot account for the long-term stability of the velocity variations in those systems; and brown dwarfs require a cosmic conspiracy to account for the required distribution of orbital inclinations. Planetary-mass companions are the simplest explanation for the bulk of the observations.

11.4.2 Formation conundra: Jovian planets in Mercurian orbits

The existence of giant planets at distances of 0.1 AU or less from the parent star is clearly at odds with the standard formation theory. Temperatures at those radii in the protoplanetary disk exceed 1,000 K, precluding the existence of any icy condensates. Refractory grains can exist, but there are two major problems with forming massive, rocky planets *in situ*: first, tidal forces, while insufficient in strength to disrupt a fully formed jovian-mass planet, are capable of preventing aggregation of smaller planetismals; second, the heavy element content in a ‘rocky Jupiter’ represents a significant fraction of the total mass of those elements in the parent star.

If giant planets cannot form at such small orbital radii, a possible solution is to form them at conventional distances and identify a mechanism that allows the orbit to evolve inward over the stellar lifetime. Perhaps surprisingly, such mechanisms have been known for well over a decade. Indeed, their efficiency is so high that discussion has centred more on how Jupiter has managed to maintain its position at the present orbital radius, than on whether orbital migration is possible [W2], [L6], [L8].

The two most efficient mechanisms of planetary migration are based on angular momentum coupling between the planet and the protoplanetary disk, with the planet losing momentum and spiralling inward. First, Lindblad resonances can form within the disk [A3] and interact with the planetary cores; second, massive planetismals ($M > 0.1 M_J$) sweep out gaps in the disk, and the net torque leads to inward migration [T3]. In both cases, theoretical simulations predict timescales of only $3\text{--}5 \times 10^5$ years for a jovian-mass planet formed at 5 AU to accrete onto the central star. This is $\sim 10\%$ of the expected lifetime of the parent disk. Lin [L6] suggests that a series of such planets form over the disk lifetime and march inward in succession,

with the surviving planets being those that form just before the disk dissipates. As a variant of this process, inward migration may cease at small radii either due to tidal interactions, transferring momentum from the star's rotation to orbital motion, or because the star has cleared a low-density hole in the central regions of the disk.

There are, however, problems with mechanisms acting over such rapid time-scales. The accretion time of 3×10^5 years is short compared with optimistic estimates of the formation time of giant planets (Section 11.2.1). The disk as a whole participates in the inward migration, so all material interior to the planet ($r_P < 5$ AU) accretes onto the central star. For a minimum-mass solar nebula, this corresponds to an accretion rate of $\sim 3 \times 10^{-8} M_\odot \text{ yr}^{-1}$. Observations of T Tauris [H1] indicate accretion rates at or above that level at ages less than ~ 1 Myr, with the accretion rate decreasing as t^{-2} . Maintaining the inner disk over its few $\times 10^6$ -year lifetime therefore demands a substantial reservoir of material ($0.05\text{--}0.10 M_\odot$) at larger radii. On the other hand, orbital migration allows the planet access to larger reservoirs of material, and can therefore increase the accretion rate; simulations indicate that the time required to build gas giants by core accretion can be shortened by a factor of 10 to $\sim 10^6$ years [A7].

Migration scenarios generally predict near-circular planetary orbits. This expectation is generally met for the short-period hot Jupiters, as discussed in Section 11.5.2, but another mechanism plays a strong role in shaping those orbits. At radii below ~ 0.15 AU, tidal interactions enforce orbital circularisation, and synchronisation of the planetary rotational and sidereal periods. Indeed, the hot Jupiter companion to τ Boo is sufficiently massive that it has spun up the stellar rotation period to match the planetary orbit.

In contrast to hot Jupiters, many of the planetary companions with larger semi-major axes have notably eccentric orbits. It is possible that this results from gravitational scattering among a number of giant planets, as mutual perturbations cause the orbits to overlap [R1], possibly culminating in planet/planet mergers [L7]. If so, additional giant planets in longer period orbits should be present in systems with a single detected planet in a highly eccentric orbit; in a few cases, a longer period planet has been detected. Other possible mechanisms for inducing eccentric orbits involve inhomogeneous structure and gravitational perturbations within the protoplanetary disk or interactions with (as yet undiscovered) stellar companions.

In summary, none of the proposed theoretical models for the dynamical evolution of protoplanetary disks provides a full explanation for the observed characteristics of the known planetary systems. However, initial efforts have identified a number of promising avenues for further exploration through more detailed theoretical and observational studies.

11.5 WHAT DO WE KNOW ABOUT EXTRASOLAR PLANETS?

The ~ 130 stars in the current catalogue of extrasolar planetary systems embrace a wide variety of properties (see <http://www.obspm.fr/planets> – the Extrasolar Planets Encyclopedia – for an up-to-date summary of detections). Almost all of the parent

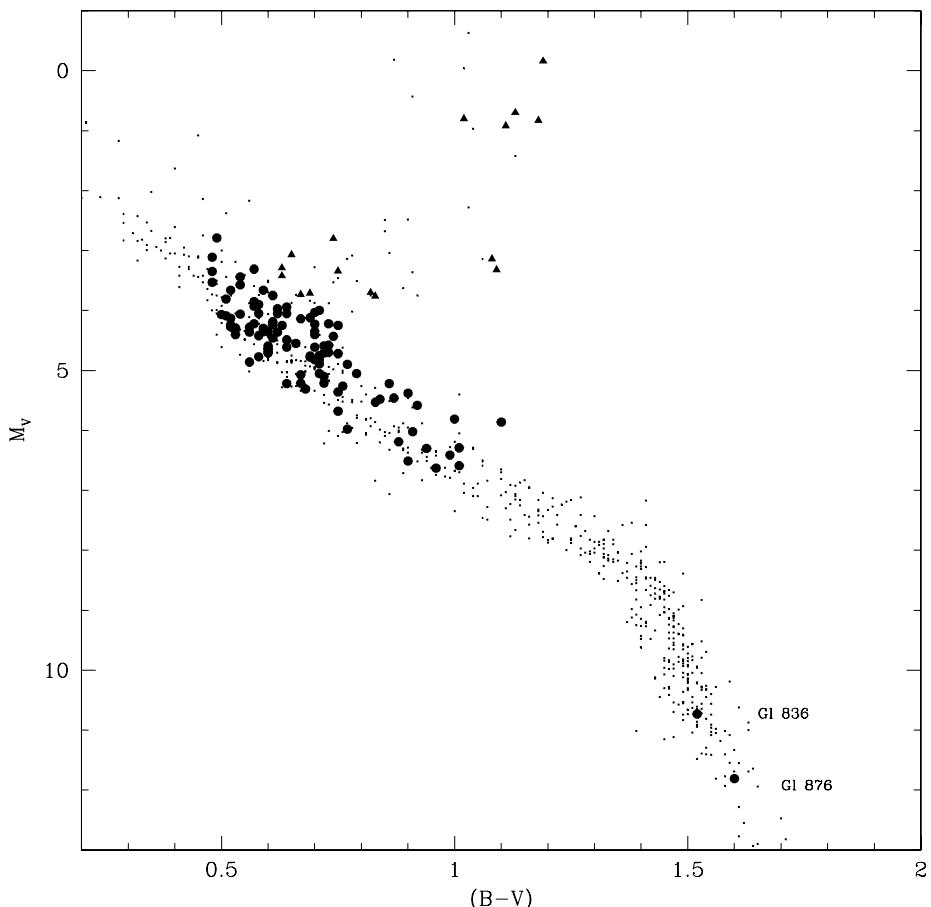


Figure 11.7. H–R diagram for the nearby stars; main sequence stars with planetary mass companions are plotted as solid points, while evolved systems are plotted as solid triangles.

stars are solar-like, spectral types F, G and K, and predominantly on the main sequence (see Figure 11.7). This is expected, since the vast majority (~ 120 systems) were discovered in radial velocity surveys, which preferentially target those stars. However, planetary companions have been detected around two M dwarfs: Gl 876, which has two companions on resonant orbits ($P = 30.1$ and 61.0 days); and Gl 836, which has one of the lowest mass companions identified so far, with $m_2 \sin(i) = 0.067 M_J$, or ~ 21 Earth masses. Planetary companions have been detected in stellar binary systems (e.g., 16 Cygni B); in stellar/brown dwarf systems (e.g., HD 41004AB, which includes a K0V dwarf, an M3 dwarf, a brown dwarf and at least one planet); around giant stars (e.g., HD 47536); and multiple planets have been found in more than 20 systems, including 4 companions to 55 Cnc. Almost all of the host stars lie within 50 parsecs of the Sun, but with the addition of transit and

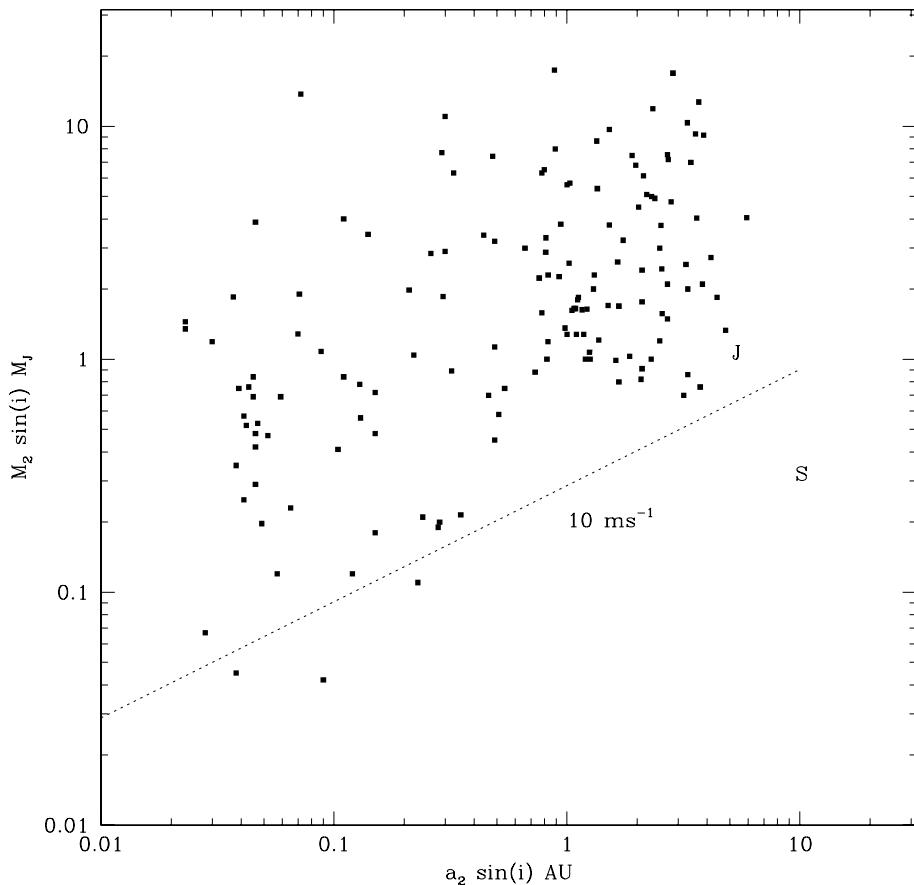


Figure 11.8. The mass/semi-major axis distribution for the known planetary systems. J and S mark the location of Jupiter and Saturn, respectively, and the dashed line marks the locus of systems producing velocity variations of semi-amplitude 10 m s^{-1} for a solar-mass primary star.

microlensing surveys to the observational arsenal, a handful of systems are being detected in the inner Galaxy, 4–6 kpc distant. The following sections provide a broad overview of the ensemble properties of this growing menagerie.

11.5.1 The mass distribution

The mass function for low-mass companions to solar-like stars $\Psi_C(M)$ is one of the most important parameters that can be derived from the current sample. Figure 11.8 plots the observed distribution of mass as a function of semi-major axis for all systems with planetary-mass companions; in both cases, we measure only the projected parameters $m_2 \sin(i)$ and $a \sin(i)$. This illustrates two selection effects in the present sample: first, lower mass companions become progressively less easy to

detect with increasing semi-major axis; second, only a handful of systems are known with semi-major axes exceeding 4 AU. Jupiter lies on the boundary of the current sample, and Saturn is below the detection limit.

Both selection effects stem from the fact that the overwhelming majority of systems have been detected through radial velocity surveys. Few surveys have been running for longer than 6 or 7 years; since detection requires observation through at least one cycle, this sets an upper limit on the semi-major axis for announced systems (there are unpublished candidates showing longer term trends in velocity that could well include Saturn-like systems). The lower mass limit, and the increasing trend with $a \sin(i)$, is tied to the overall accuracy of the radial velocity surveys. The dotted line in Figure 11.8 shows the approximate locus of systems that give rise to reflex motions of semi-amplitude 10 m s^{-1} in the parent star. While current radial velocity measurements are achieving accuracies close to 1 m s^{-1} , most data from the late 1990s has an uncertainty closer to 3 m s^{-1} . Thus, the dotted line effectively marks a 3σ detection limit for radial velocity surveys; statistics for objects with masses $M < 1 M_J$ clearly suffer from significant incompleteness.

The distribution of points in Figure 11.8 suggests another systematic effect: a scarcity of high-mass planets in orbits with small semi-major axis. This is extremely unlikely to be due to an observational selection effect, since high-mass companions produce larger reflex motions in the parent star, and are therefore easier to detect.

A key issue in understanding the formation of low-mass companions is determining whether $\Psi_C(M)$ is a single continuous function, spanning the stellar, brown dwarf and planetary regimes, or whether the overall distribution is a composite function. Distinguishing between these alternatives offers strong clues as to the relative origins of, particularly, brown dwarfs and planets (Section 11.2.3). From the outset, results from radial velocity surveys have suggested that the mass distribution is discontinuous. Stellar companions are not uncommon, even at small separations, among FGK stars (see, e.g., [D2]), and an increasing number of planetary-mass objects are being detected. In contrast, while brown dwarfs are not uncommon as wide (>100 AU) companions to solar-type stars [G15], very few have been identified at small separations (<10 AU) from solar-type main sequence stars. The deficit at small separations has been termed the ‘brown dwarf desert’ by Marcy.

Figure 11.9 illustrates this phenomenon, plotting mass against semi-major axis (or separation for the wider systems) for low-mass companions to solar-type stars ($4 < M_V < 7$) lying within 25 parsecs of the Sun. The masses for the known companions are either estimated from orbital parameters or, for wide stellar systems, from the spectral type of the companion; for brown dwarfs, the age estimate for the primary star is used to constrain the mass of the companion. Most of the planetary systems lack measurement of the orbital inclination; if we assume a random distribution of inclinations, the most likely value of the mass is:

$$\langle m \rangle = \frac{4}{\pi} m \sin(i) \quad (11.12)$$

Similar considerations apply to the semi-major axes, and both parameters plotted in Figure 11.9 take this adjustment into account.

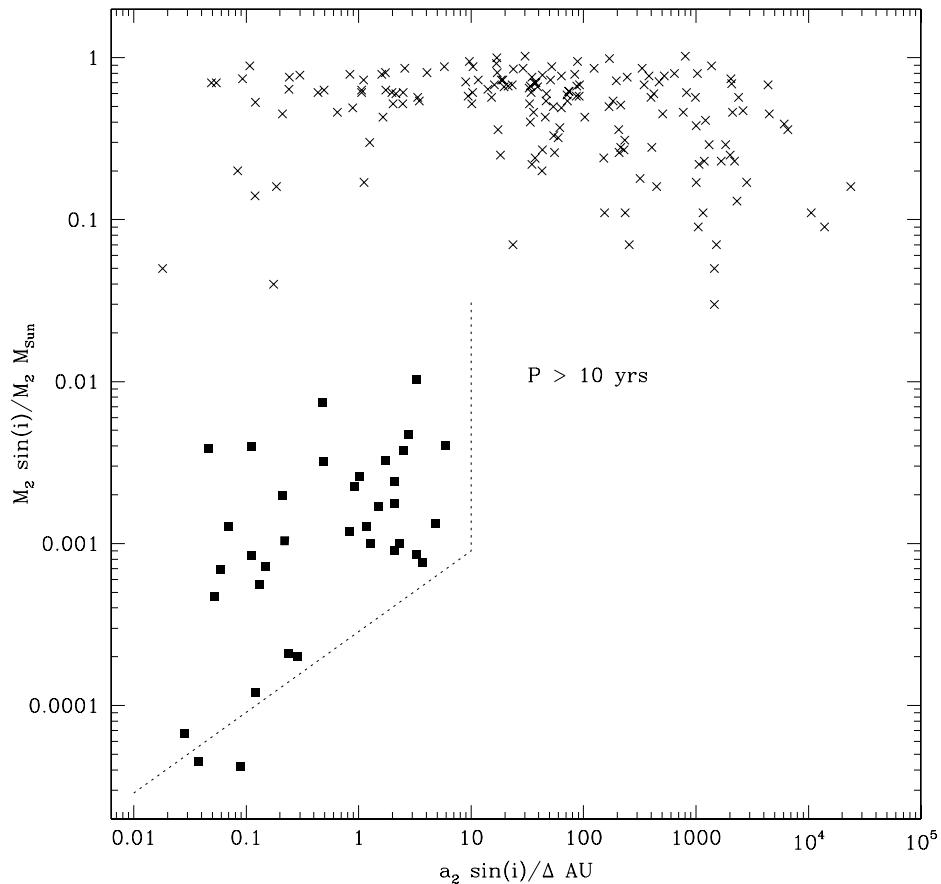


Figure 11.9. The brown dwarf desert: the mass distribution of companions to FGK stars ($4 < M_V < 7$) within 25 parsecs of the Sun plotted as a function of semi-major axis/observed separation. Stars and brown dwarfs are plotted as crosses; candidate extrasolar planets are plotted as filled squares. The effective selection constraints on the planetary systems are shown by the dashed lines.

There are a total of 415 systems within 25 parsecs that meet the absolute magnitude criteria: 140 stars have confirmed stellar companions, mainly at separations more than 50 AU; 30 systems have planetary-mass companions; but only a handful have known brown dwarf companions, and most lie at wide separations. It is important to emphasise that 90% of the FGK stars in this 25-parsec sample are included in at least one of the radial velocity surveys listed in Table 11.2, providing stringent detection limits for close ($a \sin(i) < 10 \text{ AU}$) companions. Thus, the scarcity of companions with masses between 0.01 and $0.1 M_\odot$ cannot be ascribed to observational incompleteness or systematic selection effects. It is clear that there is a significant minimum in $\Psi_C(M)$ at these masses.

Figure 11.10 plots $\Psi_C(M)$ for solar-type stars. The upper panel plots data for

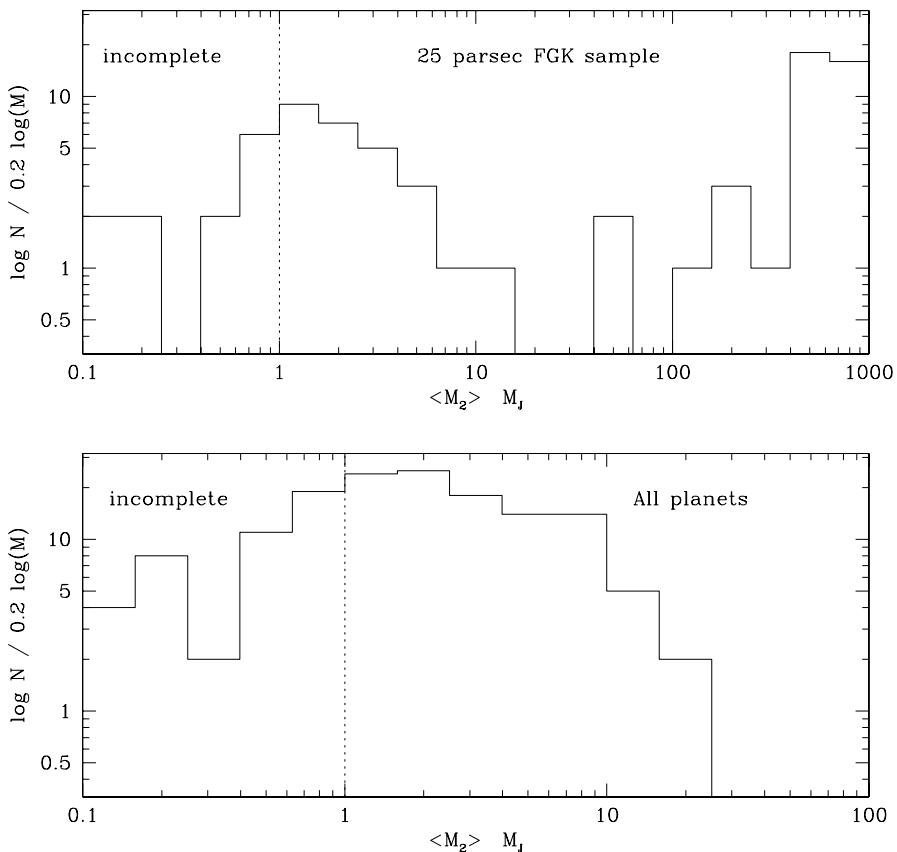


Figure 11.10. The mass function of low-mass companions to solar-like stars. The upper panel plots $\Psi_C(M)$ for the 25-pc companions from Figure 11.9, limiting the sample to separations $\Delta < 10$ AU; the lower panel plots $\Psi_{PC}(M)$ for all of the planetary companions shown in Figure 11.8. The dotted lines mark the low-mass regions where the sample is known to be incomplete.

the 25-parsec sample shown in Figure 11.9, including only companions at separations of less than 10 AU from the parent star, while the lower panel plots the mass distribution of all planetary mass companions known at the time of writing (December, 2004). As in Figure 11.9, the masses are scaled by a factor of $4/\pi$ to give the expectation value of the mass for each system. Two conclusions can be drawn from the 25-parsec distribution. First, there is a striking minimum in the companion mass function between ~ 10 and $100 M_J$ (~ 0.01 and $0.1 M_\odot$), the brown dwarf desert. This division strongly suggests that stellar/brown dwarf companions and planetary-mass companions of solar-type stars are the product of different formation mechanisms. Second, the companion mass function at stellar masses (at least for close binary systems) is clearly inconsistent with the field star mass function (Figure 9.9); there is a general scarcity of low-mass stellar companions

within 10 AU of solar-type stars. While there may be a brown dwarf desert, there are only a few more oases in the M dwarf regime.

While there is no definitive explanation of the origin of the brown dwarf desert, it seems likely that it reflects an intrinsic property of the star formation process: specifically, the difficulty of forming a low-mass object in close proximity to a high-mass object. Several studies (e.g., [A6]) have suggested that competitive accretion between two protostellar cores at small separations will result either in two stars of similar mass (i.e., a high-q binary system), or the disruption and assimilation of the lower mass core by the higher mass core. Brown dwarfs are particularly vulnerable, since their masses are comparable with the mass of the protoplanetary disk around a solar-type star; as a result, systems at small separation (< 50 AU) migrate inward rapidly and merge with the primary star.

The current sample of planetary-mass companions, shown in the lower panel of Figure 11.10 is clearly incomplete at low masses. Nonetheless, the mass function, offers some interesting pointers to the likely form of the underlying distribution, particularly at masses above $1 M_J$. The data suggest that $\Psi_{\text{PC}}(M)$ increases steeply as the mass falls from 30 to $10 M_J$. The distribution flattens at lower masses, although $\Psi_{\text{PC}}(M)$ is still increasing; indeed, the data are consistent with a power law, $\Psi_{\text{PC}}(M) \propto M^{-1}$, at masses between 10 and $1 M_J$. It is possible that incompleteness accounts for most, perhaps all, of the apparent turnover below $1 M_J$.

11.5.2 The orbits: semi-major axes, eccentricities and multiple systems

Among the planets in the Solar System, Pluto, Mercury and Mars have the orbits with highest eccentricity, $e = 0.253$, 0.206 and 0.093 , respectively. Saturn has the most eccentric orbit of the gas giants, with $e = 0.053$. In contrast, the current catalogue of extrasolar planets includes systems with eccentricities as high as 0.8 . Figure 11.11 shows the distribution of eccentricity as a function of orbital period and mass $m_2 \sin(i)$ for the full sample. As noted in Section 11.2.3, the overall distribution is more reminiscent of stellar binaries than the Solar System planets, but this almost certainly represents similar dynamical processes rather than similar origins.

Tidal forces tend to circularise planets in short-period orbits, accounting for the absence of systems with $e > 0.2$ and $P < 80$ days. Orbital circularisation is also responsible for the (relatively) low-eccentricity orbits of almost all sub-Jovian-mass planets; in order to be detected by the radial velocity programs, these planets must be in short-period ($a < 0.3$) orbits. The shortest period systems, all from the OGLE transit survey, have orbital periods of only 1.2 – 1.7 days, and semi-major axes between 0.022 and 0.03 AU (almost 10 times smaller than Mercury's orbit). There are no high-mass planets ($M > 8 M_J$) in orbits with $e < 0.2$ in the current sample. This at least partly reflects the absence of those companions in short-period orbits (Figure 11.8).

Approximately 15% of extrasolar planetary hosts harbour at least two detected planets. Statistically, there is no evidence that the properties of these multiple-planet systems differ in any significant way from the (current) single-planet hosts. This is not surprising, since many of the latter may well include as yet undetected lower

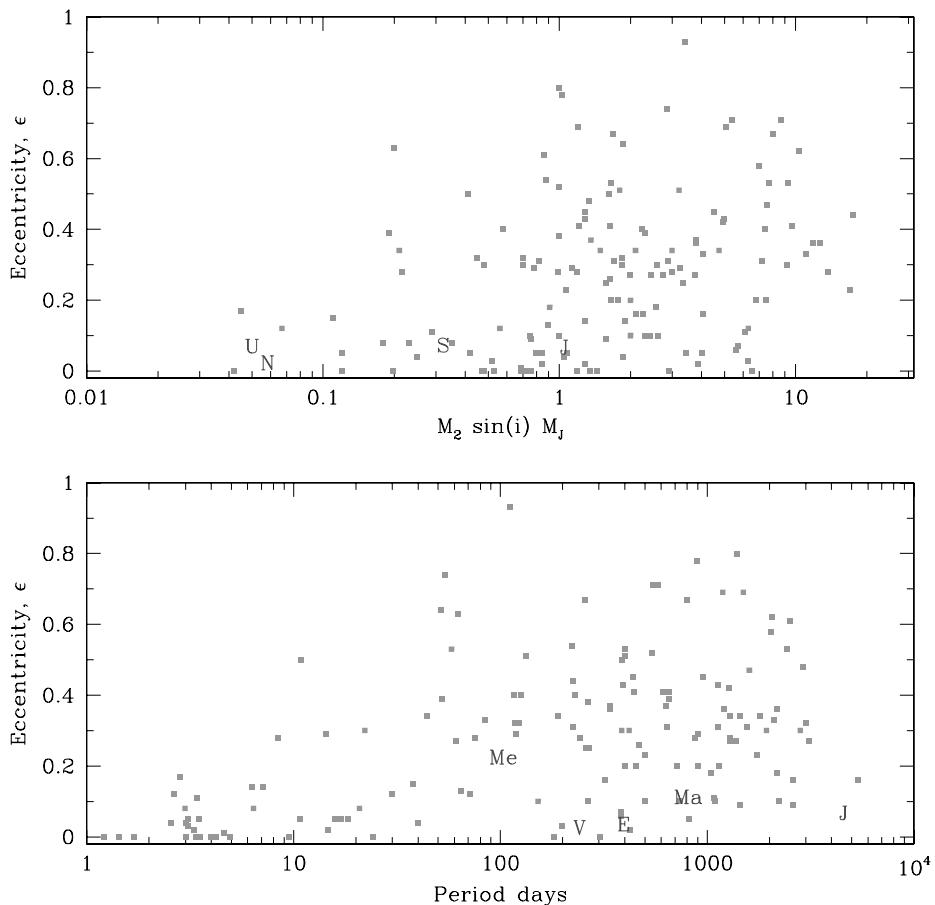


Figure 11.11. The distribution of eccentricity as a function of mass (upper panel) and period (lower panel) for extrasolar planetary systems; the letters show the locations of the planets in the Solar System.

mass and/or longer period planets. In a few cases, the multiple planets are on resonant orbits, but many systems show no sign of this effect. Some systems include planets on orbits with substantial eccentricities (e.g., 55 Cnc), but simulations show that they are dynamically stable. Similarly, all known planetary companions of stars in binary systems are on stable orbits, despite the limited parameter space available in such systems.

11.5.3 Radii and atmospheric composition

The identification of planets in transiting systems not only allows a definitive measurement of the mass of the companion, but also provides an opportunity to determine the physical size of the planet and probe its atmospheric structure and

composition. HD 209458, as the brightest such system, has attracted the most attention; the remaining five systems currently known are more than 5 magnitudes fainter (~ 10 magnitudes for the OGLE systems), and therefore allow less scope for detailed observations.

As a planet transits across the face of the primary star, light from that star passes through the outer regions of the planetary atmosphere on its way to Earth; this is analogous to the illumination of the Moon by sunlight refracted through the Earth's atmosphere during a total lunar eclipse. As the light passes through the planetary atmosphere, it can be absorbed by chemical species in the atmosphere, leading to slightly stronger absorption lines during transit. Detecting this effect demands observations of extremely high accuracy. To date, the only successful observation has been the measurement of a 0.01% increase in the equivalent width of the Na D lines in HD 209458 through HST optical spectroscopy; space-based observations have a clear advantage in this type of study, since they are not susceptible to terrestrial atmospheric variations and therefore offer much greater intrinsic stability. Ground-based attempts to detect CO (at infrared wavelengths) in HD 209458 have so far been unsuccessful; all the remaining systems are too faint for spectroscopic observations at this level.

The HST photometry of HD 209458 (Figure 11.2) shows that the planet obscures 1.6% of the flux from the parent star, indicating a radius of $1.347 R_J$ [B16]. Since the planet has a mass of $0.69 M_J$, this corresponds to a density one-quarter that of Jupiter (and one-half the density of Saturn), and implies that HD 209458b is mainly composed of H and He. Both the implied composition and the relatively large radius are unexpected; proposed explanations include inflation due to irradiation in the sub-Mercurian orbit, an extremely small core mass (~ 1 Earth mass) and an additional internal heat source, perhaps due to tidal heating [B14]. However, all of the transiting planets discovered subsequently by OGLE and by other transit surveys have smaller radii (~ 1.08 to $1.15 R_J$) and higher densities, much closer to the predictions of standard theoretical models; this is despite the fact that the OGLE systems are in orbits with smaller semi-major axes, and are therefore subject to greater irradiation and stronger tidal forces. As yet, there is no accepted explanation for the anomalous nature of HD 209458b.

11.5.4 Metallicities and kinematics of the host stars

Are stars with planets unusual in anyway? With ~ 130 examples, we can compare the statistical properties of the extrasolar planetary hosts against average field stars, particularly since most of the host stars are late-F, G or early-K main sequence stars. Reid [R4] and Santos *et al.* [S9] have examined the kinematics of a subset of these stars, and conclude that there is little to set them apart from the average field star. The overall velocity dispersion is $\sim 45 \text{ km s}^{-1}$, intermediate between dMe and dM stars (Table 7.3), and consistent with an average age of 3–4 Gyr.

The abundance distribution is another story. Gonzalez [G16] originally suggested that there was evidence that metal-rich stars were more likely to have planetary companions. Several subsequent studies arrived at similar conclusions

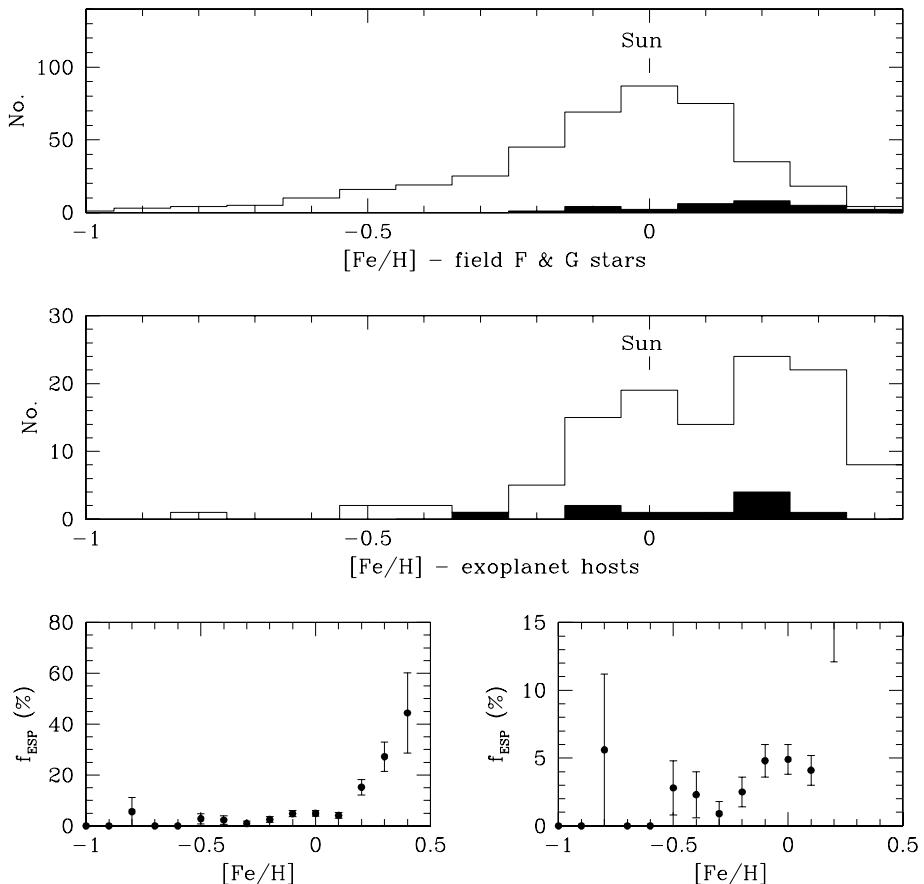


Figure 11.12. The metallicity distribution of extrasolar planet hosts and the frequency of planetary systems as a function of metallicity. The uppermost panel plots the abundance distribution of local FGK stars (the solid histogram shows the contribution from subgiants); the middle panel shows the metallicity distribution of stars known to have planetary companions (again, the contribution from evolved stars is shown as a solid histogram); the lower panels combine these distributions to show the frequency of planetary systems, f_{ESP} , as a function of metallicity (both lower panels plot the same data with different vertical scales).

[G17], [S8], [R4], [L17]; indeed, Laughlin [L16] suggested specific targets for radial velocity monitoring based on their metallicity, and follow-up observations have confirmed planetary companions in several cases.

This result can be quantified, since most of the planetary host stars have both spectroscopic and photometric (Strömgren) metallicity measurements (M dwarfs are notable exceptions), providing abundances accurate to better than 0.1 dex. Figure 11.12 compares the $[m/H]$ distribution of the planetary hosts against the field distribution (Section 7.5.3). There are clearly a higher proportion of metal-

rich stars among stars known to have planetary systems. The lower two panels plot the ratio between the two distributions: the data indicate that, while $\sim 5\%$ of solar metallicity FGK stars are likely to have a planetary companion detectable in current radial velocity surveys, the detection rate rises to almost 50% for FGK stars with metallicities more than twice solar.

There are relatively few local field stars at metallicities below -0.3 dex; nonetheless the available data suggests that the frequency of Doppler planets declines at these lower metallicities. Sozzetti *et al.* [S11] have recently initiated an observing program that specifically targets nearby metal-poor stars (see Table 11.2).

Globular clusters offer a happier hunting ground for metal-poor stars, and an 8.3-day photometric monitoring campaign by HST of $\sim 34,000$ stars in 47 Tuc ($[m/H] \sim -0.7$ dex) failed to find any transiting planets; approximately 30 detections would have been expected based on local statistics [G18]. On the other hand, PSR 1657+20c is in M4 (see Section 11.3.2). It remains possible that the dense stellar environment in globular clusters truncates protoplanetary disks, modifying both planet formation and the subsequent orbital distribution; nonetheless, the absence of detections tends to support the apparent decline in planetary frequency in sub-solar metallicity Solar Neighbourhood stars.

What is responsible for this correlation between metallicity and the occurrence of (detectable) planetary systems? Two broad scenarios have been proposed, along lines that mirror the nature vs. nurture debates in the biological sciences. Under one hypothesis [nature], (massive) planets form more readily around metal-rich stars because there is more material available for their construction; under the other [nurture], planetary hosts are metal rich because they have formed, and absorbed, planets.

Under the core accretion model, planet formation depends on the presence of grains to form the dirty ice planetismals that collect to form planetary embryos, which accrete envelopes to become gas giants. Under this scenario, a correlation between metal abundance and planet formation is not unexpected. On the other hand, metals play much less of a role in planet formation by disk fragmentation, so the observed correlation is more difficult to explain under that scenario.

The alternative hypothesis was originally suggested by Laughlin and Adams [L3], who noted that the migration mechanism proposed by Lin could lead to significant numbers of planets being accreted by the central star. They hypothesised that, since these gas giants would be expected to have higher than solar metallicity, this would lead to pollution of the outer envelope, and (apparently) super-solar metallicities. There are, however, several difficulties with this hypothesis:

- First, while it is clear that the Solar System gas giants have enhanced metallicities, it is not clear what fraction of that enhancement resides in refractory metals, and what fraction in ices (CNO). By and large, metallicity estimates of FGK stars rest more on the former than the latter.
- Second, Laws [L18] has analysed the relative abundances of 15 elements in stars with known planets, and finds that the correlation of those abundances with the

condensation temperatures of the individual elements is not consistent with the accretion hypothesis.

- Third, stellar pollution by accretion is viable only for stars with shallow surface convection zones, where the accreted material can alter significantly the average metal content (i.e., the metallic mass in the planet should be comparable to the metallic mass in the convection zone). This condition holds for spectral types earlier than \sim G0/1; the Sun (G2) has a convective zone of mass \sim 0.3 M_{\odot} , and the accretion of the entire protoplanetary disk would only increase the metal abundance by \sim 0.1 dex [S9]. Despite this, the prevalence of metal-rich planet hosts holds at later spectral types. Moreover, the degree of metal enrichment should increase with increasing temperature, since the depth (and mass) of the convective zone decreases at earlier spectral types. However, not only is there no evidence for this correlation among main sequence stars, even subgiants, which have substantially larger convection zones, show the same bias toward high [m/H].

Given these considerations, it seems more likely that nature, rather than nurture, accounts for the preference for metal-rich stars evident in Figure 11.12.

An interesting abundance anomaly has emerged in a comparative study of the two stars in the 16 Cygni system. Both are G2 dwarfs, with temperatures similar to the Sun. 16 Cygni A shows no evidence for radial velocity variations, while 16 Cygni B has a $1.7 M_J$ companion in an 800-day orbit. The two stars have almost identical heavy element abundances (16 Cygni A is slightly more metal-rich [L19]) with one exception: the lithium abundance in 16 Cygni A is $\log N(\text{Li}) = 1.27$ dex, 0.2 dex higher than the solar value; 16 Cygni B has $\log N(\text{Li}) = 1.48$ dex [K5]. As with brown dwarfs, this suggests that the convective zone in the latter star extends to higher temperatures (lower depth), leading to more rapid lithium destruction. However, the additional convective mixing must be limited, since beryllium, which is destroyed at $T \sim 3.5 \times 10^6$ K, rather than 2.5×10^6 K, has identical abundance in both stars [G1].

Both theoretical [P1] and observational [S5] studies suggest that differences in $\log N(\text{Li})$ can be driven by variations in stellar rotation, with deeper mixing occurring in slow rotators [M4]. Cochran *et al.* [C5] have proposed that observed differences between the 16 Cygni components may stem from differences in disk mass during the pre-main sequence phase, with more efficient braking of 16 Cygni B by magnetic coupling to a massive protoplanetary disk. While suggestive, this cannot provide a general rule for planetary formation, since the Sun (which clearly has a planetary system) has a lithium abundance, and presumably an angular momentum history, closer to 16 Cygni A than 16 Cygni B. Moreover, detailed analyses of other extrasolar planetary hosts have failed to find strong evidence for significant lithium depletion.

Finally, we should emphasise that the planetary frequencies illustrated in Figure 11.12 refer to solar-type stars. To date, detailed radial velocity results are only available for a relatively small number of M dwarfs. There are some initial indications that the overall frequency of Doppler detections may be significantly

lower for those stars ($\sim 2\%$ as opposed to $\sim 5\%$ for FGK stars), despite the fact that lower mass gas giants should be detectable around these lower mass stars. There is no conclusive evidence for a trend between planetary mass and stellar mass in the current sample, but, since the mass of the circumstellar disk must decline with M_* , it would not be surprising if lower mass stars have lower mass and/or fewer planetary companions. Further observations of a larger sample of M dwarfs are required to fully address this issue.

In terms of absolute numbers, while the frequency of planetary systems increases with increasing metallicity, stellar densities peak at solar metallicity and decline sharply at higher [m/H] (Figure 11.12). Thus, the local density of FGK stars with abundances within ± 0.15 dex of the Sun is ~ 0.0044 stars pc $^{-3}$; 5% of those stars have detectable planetary systems. Integrating around the Galaxy, we would expect to find $\sim 21,500$ such stars within an annulus 50 parsecs in diameter, centred at the Solar Radius (8 kpc); extending that calculation to Galactic radii of 7–9 kpc, we would expect 1.1 million Doppler detections [R4]. Extrapolating those results to M dwarfs, the planetary frequency may be lower by a factor of 2.5, but the overall number densities are higher by a factor of 8, so we would expect $\sim 70,000$ and ~ 3.5 million detections, respectively. Planetary systems are not particularly rare in the Milky Way.

11.6 THE FUTURE

The next decade should witness substantial progress in the characterisation of extra-solar planetary systems. Radial velocity surveys are now routinely achieving accuracies of $1\text{--}2\text{ m s}^{-1}$ for solar-type stars, and an increasing number of Neptunian-mass companions are being detected in longer period orbits. Moreover, the observational samples are being extended to include larger numbers of lower luminosity stars. Ground-based transit surveys are starting to produce results, albeit at a lower rate than most of the participants had anticipated, and interferometry is becoming routine at infrared wavelengths. This section outlines three key space projects that are likely to come to fruition over the next decade (or so), and which are likely to have a very substantial impact on our understanding of the nature, frequency and origin of extrasolar planetary systems. In addition to these projects, *Gaia* (Section 7.6) has sufficient astrometric accuracy to identify, from reflex motions, candidate jovian-mass systems within 100 parsecs of the Sun.

11.6.1 *Kepler* and COROT

The identification of a terrestrial planet transiting a solar-like star demands the ability to detect a flux decrease of $\sim 3 \times 10^{-4}$ to $\sim 5 \times 10^{-5}$ that lasts between 2 and 16 hours, and occurs at intervals separated by at least 10 months. That, in turn, requires that the stellar brightness is measured to better than 10^{-5} , or 0.01 millimagnitudes, and that the photometric system can remain stable at that level for years. This level of accuracy, precision and stability is impossible to achieve with

ground-based instrumentation, subject to the vagaries of Earth's atmosphere. However, the HST observations of HD 209458 demonstrate the exquisite accuracy that is possible in space-based photometry; both the COROT and *Kepler* missions have been designed to take advantage of these favourable circumstances, and search for terrestrial transits.

COROT is an ESA mission with two main aims: a search for planetary transits; and detailed asteroseismology of specific stars, using stellar pulsations to probe their internal structure. The satellite has a 27-cm telescope and a $2^\circ.8 \times 2^\circ.8$ field of view that is sampled by four CCD detectors, two dedicated to each program. Launch is scheduled for 2006, and the aim is to observe five fields, devoting 150 days to each field, and covering a total of 17.5 square degrees and $\sim 10,000$ stars with $11 < V < 16.5$ in the extrasolar planet survey. With the relatively small aperture, COROT is unlikely to be able to detect terrestrial planets lying in the habitable zones of solar-type stars, but it does offer the prospect of identifying a significant number of transiting gas giants and, potentially, terrestrial planets in sub-Mercurian orbits around K and M dwarfs [B17].

Kepler (see <http://www.kepler.arc.nasa.gov>) is a Discovery-class mission scheduled for launch in November, 2007. The satellite has a 0.95-m telescope, giving a field of view of 105 square degrees, coupled with a single instrument – an optical camera with a detector array of forty-two 2200×1024 -format CCDs (each pixel covers ~ 14 square arcsecs). *Kepler* will observe a single field in Cygnus, which includes $\sim 250,000$ stars brighter than 14th magnitude. About one-half of those stars are expected to be dwarfs with spectral types A to K, although 20% are likely to be binaries or active stars that are unsuitable for planetary transit detections. Early in the mission, approximately 100,000 stars will be identified as long-term targets, and photometric monitoring of these stars will be obtained continuously over the four-year mission lifetime.

Kepler's observational procedures have been designed to maximise photometric accuracy and efficiency, and to optimise the chances of detecting planetary transits. The observations will be made using white (unfiltered) light, maximising the number of photons detected by using the full wavelength response of the CCDs. In addition, the telescope is defocused slightly, so that a stellar image has a FWHM of 10 arcsecs, and covers more pixels on the detector. The CCDs must be read every 3 seconds to avoid saturation of bright stars, and the separate images are accumulated using on-board computer systems to give effective integration times of 15 minutes; this results in individual photometric measurements with accuracies better than 0.002% for 12th magnitude stars. It is not possible to either store all of the pixel data on board the spacecraft or transmit it to Earth, so small postage stamps centred on each target star will be accumulated, transmitted and saved in the final database.

Kepler's success rate will depend, obviously, on the prevalence of planetary systems in general, and terrestrial planets in particular. Making the optimistic assumption that most stars have associated planets, then *Kepler* could detect ~ 50 Earth-radius objects at distances of 1 AU from the parent star; if systems with planets 30% larger than Earth are common, *Kepler* could detect up to 200 planetary transits. Giant planets will be easily identified, and *Kepler* should even

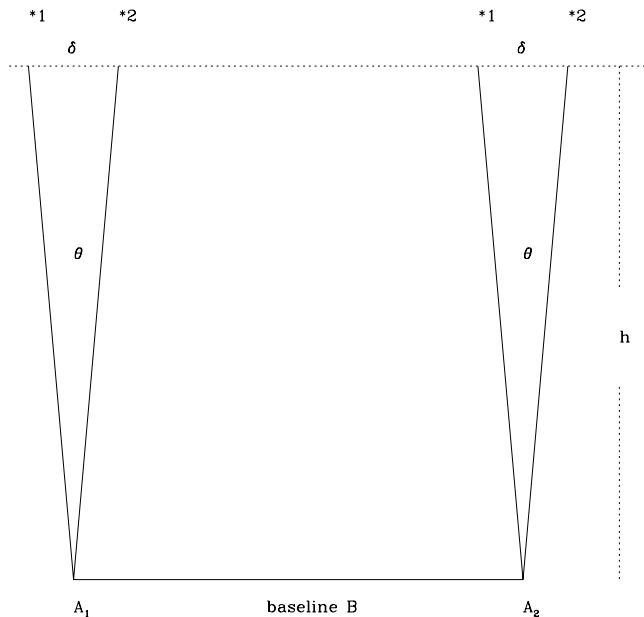


Figure 11.13. Schematic diagram of relative astrometry with a ground-based, long-baseline interferometer; $*1$ and $*2$ are the target and reference star, separated by an angular distance θ corresponding to a linear distance δ at the top of the atmosphere, height h .

detect the small amplitude photometric modulation produced by reflected light from non-transiting hot Jupiters.

11.6.2 Ground-based interferometry and the Space Interferometry Mission

Ground-based interferometry

Optical and near-infrared interferometry is set to play a key role in extrasolar planetary astrophysics. Major new facilities are being completed both on the ground, notably at the Keck Observatory and on Mt. Wilson by the Georgia State University Center for High Angular Resolution Astronomy (CHARA), and for future space missions, notably SIM Planetquest, the Space Interferometer Mission. These offer the prospect of attaining astrometric measurements at the microarcsecond level, possibly sufficient to detect perturbations due to terrestrial planets orbiting the nearest stars. Not surprisingly, significant technical difficulties must be overcome to achieve that goal.

Figure 11.13 shows a schematic diagram of an interferometer: two telescopes, A_1 and A_2 , separated by baseline B. Radio interferometry allows measurement of absolute positions over the entire sky, with the reference baseline set by the Earth's rotational motion. This technique cannot be transferred directly to optical and near-infrared wavelengths, since ground-based observations must be made

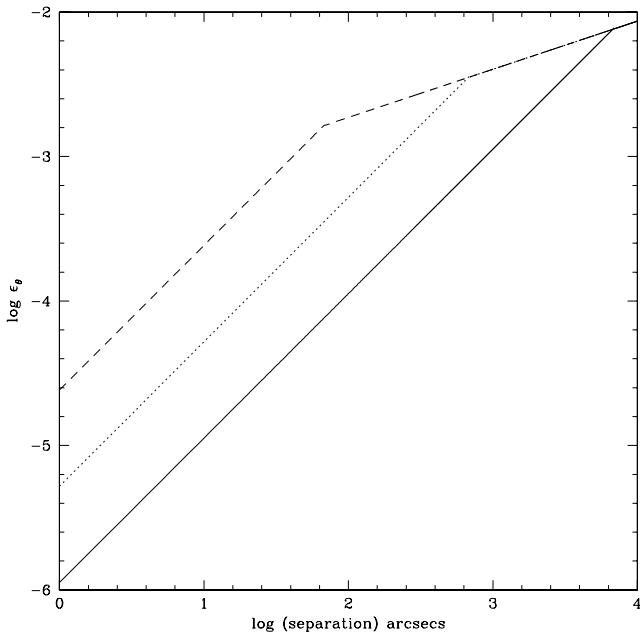


Figure 11.14. The predicted astrometric accuracy ε_θ in arcseconds, as a function of angular separation for interferometric observations made with baselines of 1 m (dashed line), 10 m (dotted line) and 100 m (solid line).

through the atmosphere (making it more difficult to correlate data over long time intervals and from separate observatories), while defining a reliable absolute reference frame is problematic in space-based interferometry. However, extremely precise *relative* astrometry is possible at optical wavelengths, measuring the position of the target (*1 in Figure 11.13) with respect to a local reference (*2), provided that observations are limited to small angular scales.

Angular separations measured using conventional ground-based astrometry (i.e., psf fitting to images) have uncertainties that increase with increasing angular separation of the target and the reference star, and those uncertainties are independent of the telescope aperture. This is because atmospheric turbulence is increasingly correlated on smaller angular scales. We can model the atmosphere as a plane-parallel layer, height h , with a Kolmogorov turbulence spectrum [S3]. If we consider two stars separated by an angular distance θ , then, in conventional astrometry, the uncertainty in the measurement of the separation is $\varepsilon_\theta \propto \theta^{1/3}$. In contrast, if those stars are observed using an interferometer with baseline B , and if the separation of the two stellar lightpaths at the top of the atmosphere $\delta = \theta h$ is less than B (Figure 11.13), then $\varepsilon_\theta \propto B^{-2/3}\theta$ (i.e., the measurement uncertainties decrease linearly with decreasing separation between the target and reference star, and also improve with longer baselines between apertures). The expected variation in ε_θ with θ for different baselines is plotted in Figure 11.14 (following [S3]). Accuracies are expected to exceed 10 μarcsec for 100-m baselines and target/reference separations

less than 10 arcsecs, and the initial results support those predictions [C6]. Putting these measurements in context, astrometry with this accuracy could provide $\sim 10\sigma$ detection of Jupiter at 100 parsecs.

Measurements of angular separation require simultaneous observations of both target and reference. These observations are generally made using a dual-beam interferometer, where the two stars are isolated at each aperture (A_1 and A_2 in Figure 11.13) and fed into separate beam-combiners. The angular separation is determined by using delay lines to measure the difference in pathlength required for constructive interference. The visibility of the fringes decreases if either source is resolved, and this sets a limit on the astrometric accuracy possible with interferometry (independent on any technical considerations associated with building giant baselines). Solar-type stars have an angular diameter of ~ 0.5 mas at 10 parsecs, effectively limiting optical systems to baselines $B < 250$ m or $B < 1,000$ m in the near-infrared (2.2 μ m).

Space interferometry

The critical limiting factor for ground-based interferometry is the small field of view available for finding reference stars, ~ 15 arcsecs for $B \sim 100$ m. Many potential targets, particularly at high galactic latitude, lack reference stars, even with the limiting magnitude of $K \sim 21.5$ afforded by Keck. Placing an interferometer in space eliminates atmospheric distortions, and allows access to reference stars at larger angular separations (degrees). Such instruments offer the potential of astrometric measurements made with 1 parsec precision [C7], [H6]. Moreover, several current projects aim to use interferometers for nulling (i.e., adjusting the pathlengths from the individual apertures to give destructive interference for on-axis sources). The net result is that light from the on-axis stellar primary is cancelled, allowing the detection of (off-axis) orbiting planets [B10], [A1]. Hot Jupiters, such as 51 Peg b, have temperatures between 2,000 and 1,300 K, comparable to L dwarfs; since their radii are also very similar (see Section 3.5.3), these planetary companions are expected to be 8–10 magnitudes fainter than the primary star in the J band, and only ~ 5 –6 magnitudes fainter at 5 μ m. With angular separations of ~ 2 milliarcsec at 25 parsecs, the nearest of those systems could be within the grasp of ground-based interferometers. Detecting terrestrial planets requires nulling with an efficiency of 1 part in 10^6 , which can only be achieved with space-borne instrumentation.

SIM Planetquest (see http://planetquest.jpl.nasa.gov/SIM/sim_index.html) will be the first space-based, long-baseline, Michelson interferometer, and will incorporate nulling (at a level of 1 part in 10^4). Scheduled for launch into an Earth-trailing orbit sometime after 2010, the current design for SIM envisages three collinear, optical interferometers providing observations with baselines from 0.5–10 m. The observing program is scheduled to last five years, during which time SIM will undertake a number of large-scale science programs, including a search for new planetary systems. SIM will establish a grid of $\sim 6,000$ reference stars over the full celestial sphere, and those stars are designed to provide wide-angle astrometric accuracies of ~ 4 parsec. The planetary search will use local ($\theta < 1^\circ$) reference stars and narrow-

angle astrometry, with the aim of achieving astrometric accuracies of ~ 1 parsec. The technical requirements are considerable: the pathlength between apertures must be measured to accuracies of $\sim 10^{-12}$ m (1 picometre). If these goals are met, the observations will not only provide accurate measurements of the inclinations (and hence masses) of all currently known planetary systems, but will also offer the prospect of detecting reflex motions due to terrestrial-mass companions in several nearby stellar systems [F3].

11.6.3 *Darwin* and the Terrestrial Planet Finder

COROT, *Kepler* and SIM are designed to detect planets indirectly, through their effect on either the brightness or the astrometric position of the host star; *Darwin* [F4] and TPF (see http://planetquest.jpl.nasa.gov/TPF/tpf_index.html and [B18]) aim to obtain direct images of the planets themselves. Both of these missions are at least ten years from launch, and therefore in the very early stages of development. Current plans for TPF envisage two separate missions: TPF-C, an optical/far-red coronagraph, which will be launched no earlier than 2014; and TPF-I, a long-baseline, multi-element interferometer, which will be launched after 2019. *Darwin*, ESA's planet-finder mission, is planned for launch in 2014, and is envisaged as consisting of several free-flying elements, providing baselines of 40–500 m, with each telescope feeding a central beam combiner. Clearly, this design will place extraordinary technical requirements on the mechanisms monitoring the individual beam pathlengths.

The primary goal of these missions is not just the detection of planets of terrestrial mass lying within the habitable zone of nearby solar-like stars, but observation at sufficient detail to determine whether those worlds are possibly habitable. This requires detection of potential biosignatures, spectroscopic features such as O₂, H₂O and the ‘red edge’ due to chlorophyll [B18]; that, in turn, demands at least narrowband ($R \sim 50$) photometry, or, preferably, low-resolution spectroscopy. TPF-C will use a series of deformable mirrors to feed ‘sharpened’ images to sophisticated coronagraphic masks, which will reduce the stellar brightness to close to the expected brightness of terrestrial planetary companions at separations of $\sim 0.05\text{--}1$ arcsec from the primary. The aim is to obtain observations at three epochs of 50–150 solar-type (late-F, G, early-K) stars within 15 parsecs of the Sun. In principle, a planetary companion could be detected (broadband imaging) with integration times of ~ 1 day; characterising the atmosphere through higher resolution observations will require significantly longer integrations.

TPF-I and *Darwin* will use nulling interferometry at infrared wavelengths to resolve planetary systems. Current designs for TPF-I include both a structurally connected system, with baselines of ~ 40 m, and a free-flying system, similar to *Darwin* [B18]. Both missions (which may eventually be combined into a single mission) are likely to concentrate on obtaining detailed observations of terrestrial planets identified by TPF-C.

11.7 SUMMARY

The intensive radial velocity surveys undertaken over the last decade have provided irrefutable evidence for the existence of extrasolar planets. Other techniques, such as photometric transit and microlensing surveys, have recently started to contribute to the planetary census. With ~ 130 host systems currently known, it is clear that the mass distribution of these very-low mass companions is not an extension of the mass function of stellar/brown dwarf companions. The obvious inference is that the planetary-mass objects formed as planets, within the circumstellar disk of the parent protostar. Most of the known systems have solar-type stars as primaries, but to a large extent this reflects the distribution of spectral types targeted for observation. There is a strong correlation between the metallicity of the host star and the likelihood of detecting Doppler companions; current indications are that gas giants form more readily at super-solar metallicities, although $\sim 5\%$ of sun-like stars have detectable companions. Lower mass stars also have planetary companions, as emphasised by the discovery of planets around Gl 876 and Gl 836, although the overall frequency of Jovian-mass gas giants may well be lower among M dwarfs. Many important questions remain to be settled concerning the actual formation process, but the overall implication is that planetary systems are not uncommon in the Galactic Disk.

11.8 REFERENCES

- A1 Angel, J. R. P., Woolf, N. J., 1997, *ApJ*, **475**, 373.
- A2 Artymowicz, P., 1988, *ApJ*, **335**, L79.
- A3 Artymowicz, P., 1993, *ApJ*, **419**, 166.
- A4 Alonso, R., *et al.*, 2004, *ApJ*, **613**, L153.
- A5 Aumann, H. H., 1988, *AJ*, **96**, 1415.
- A6 Armitage, P. J., Bonnell, I. A., 2002, *MNRAS*, **330**, L11.
- A7 Alibert, Y., Mordasini, C., Benz, W., 2004, *A&A*, **417**, L25.
- B1 Bailes, M., Lyne, A. G., Shemar, S. L., 1991, *Nature*, **352**, 311.
- B2 Beckwith, S. V. W., Sargent, A. I., Chini, R. S., Güsten, R., 1990, *AJ*, **99**, 924.
- B3 Behall, A. L., Harrington, R. S., 1976, *PASP*, **88**, 204.
- B4 Benedict, G. F., *et al.*, 2002, *ApJ*, **581**, L115.
- B5 Bouchy, F., *et al.*, 2004, *A&A*, **421**, L13.
- B6 Black, D. C., 1997, *ApJ*, **490**, L171.
- B7 Boden, A. F., *et al.*, 1998, *ApJ*, **504**, L39.
- B8 Boss, A. P., 1997, *Science*, **276**, 1836.
- B9 Boss, A. P., 1998, *Nature*, **393**, 141.
- B10 Bracewell, R. N., McPhie, R. H., 1979, *Icarus*, **38**, 136.
- B11 Brown, T. M., Kotak, R., Horner, S., Kennelly, E. J., *et al.*, 1998, *ApJS*, **117**, 563.
- B12 Buffon, G. L. L., 1745, *De la Formation des Planètes*, Paris.
- B13 Butler, R. P., Marcy, G. W., 1996, *ApJ*, **464**, L15.
- B14 Bodenheimer, P., Lin, D. C., 2002, *Ann. Rev. Earth & Planetary Sciences*, **30**, 113.
- B15 Bond, I. A., *et al.*, 2004, *ApJ*, **606**, L155.

- B16 Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., Burrows, A., 2001, *ApJ*, **552**, 699.
- B17 Bordé, P., Rouan, D., Léger, A., 2003, *A&A*, **405**, 1137.
- B18 Beichman, C., 2002, in *Bioastronomy 2002: Life Among the Stars*, IAU Symposium 213 (San Francisco; Astronomical Society of the Pacific), p. 451.
- C1 Campbell, B., Walker, G. A. H., 1979, *PASP*, **91**, 540.
- C2 Campbell, B., Walker, G. A. H., Yang, S., 1988, *ApJ*, **331**, 902.
- C3 Chamberlin, T. C., 1901, *ApJ*, **14**, 17.
- C4 Charbonneau, D., Brown, T. M., Latham, D. W., Mayor, M., 2000, *ApJ*, **529**, L45.
- C5 Cochran, W. D., Hatzes, A. P., Butler, R. P., Marcy, G. W., 1997, *ApJ*, **483**, 457.
- C6 Colavita, M. M., 1994, *A&A*, **283**, 1027.
- C7 Colavita, M. M., Shao, M., Rayman, M. D., 1993, *Appl. Opt.*, **32**, 1789.
- C8 Croswell, K., 1998, *Planet Quest*, Harcourt Brace.
- C9 Cuzzi, J. N., Dobrovolskis, A. R., Champney, J. M., 1993, *Icarus*, **106**, 102.
- C10 Cameron, A. G. W., 1978, *Moon Planets*, **18**, 5.
- C11 Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., Lowrance, P., 2004, *A&A*, **425**, L29.
- C12 Cummings, A., Marcy, G. W., Butler, R. P., 1999, *ApJ*, **526**, 890.
- C13 Cochran, W. D. *et al.*, 2004, *ApJ*, **611**, L131.
- D1 Desidera, S. *et al.*, 2004, *A&A*, **420**, L27.
- D2 Duquennoy, A., Mayor, M., 1991, *A&A*, **248**, 485.
- D3 Darling, D., 2000, *Life Everywhere*, Basic Books.
- F1 Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S., Apps, K., 1999, *PASP*, **111**, 50.
- F2 Fuhrmann, K., Pfeiffer, M. J., Bernkopf, J., 1997, *A&A*, **326**, 1081.
- F3 Ford, E. B., Tremaine, S., 2003, *PASP*, **115**, 1171.
- F4 Fridlund, C. V. M., 2002, *Adv. Sp. Res.*, **34**, 613.
- F5 Fischer, D. E. *et al.*, 2005, *ApJ*, in press.
- G1 García Lopez, R. J., Pérez de Taoro, M. R., 1997, *A&A*, **334**, 599.
- G2 Gatewood, G., Eichorn, H., 1973, *AJ*, **78**, 769.
- G3 Gatewood, G., 1974, *AJ*, **79**, 52.
- G4 Gatewood, G., 1976, *Icarus*, **27**, 1.
- G5 Gatewood, G., 1987, *AJ*, **94**, 213.
- G6 Gatewood, G., 1995, *Ap. Sp. Sc.*, **223**, 91.
- G7 Gatewood, G., 1997, *BAAS*, **28**, 885.
- G8 Gautschy, A., Saio, H., 1995, *ARAA*, **33**, 75.
- G9 Gonzalez, G., 1998, *A&A*, **334**, 221.
- G10 Gray, D., 1997, *Nature*, **385**, 795.
- G11 Gray, D., 1998, *Nature*, **391**, 153.
- G12 Gray, D., Hatzes, A., 1907, *ApJ*, **490**, 412.
- G13 Guillot, T., Chabrier, G., Morel, P., Gautier, D., 1994, *Icarus*, **112**, 354.
- G14 Greaves, J. S., *et al.*, 1998, *ApJ*, **506**, L133.
- G15 Gizis, J. E., Kirkpatrick, J. D., Burgasser, A., Reid, I. N., Monet, D. G., Liebert, J., Wilson, J. C., 2001, *ApJ*, **551**, L173.
- G16 Gonzalez, G., 1997, *MNRAS*, **285**, 403.
- G17 Gonzalez, G., 1999, *MNRAS*, **308**, 447.
- G18 Gilliland, R. L., *et al.*, 2000, *ApJ*, **545**, L47.
- G19 Ge, J., 2002, *ApJ*, **571**, L165.
- G20 Ge, J., 2003, *ApJ*, **593**, L147.
- H1 Hartmann, L., Calvert, N., Gullbring, E., D'Alessio, P., 1998, *ApJ*, **495**, 385.

- H2 Hayashi, C., Nakazawa, K., Adachi, I., 1977, *PASJ*, **29**, 163.
- H3 Henry, G. W., Baliunas, S. L., Donahue, R. A., Soon, W. H., Saar, S. H., 1997, *ApJ*, **474**, 503.
- H4 Henry, T. J., 1994, Ph.d. thesis, University of Arizona.
- H5 Hershey, J. L., 1973, *AJ*, **78**, 769.
- H6 Höglund, E., Fabricius, C., Makarov, V. V., 1997, *Exp. Ast.*, **7**, 101.
- H7 Hulse, R. A., Taylor, J. H., 1975, *ApJ*, **195**, L51.
- H8 Henry, G., Marcy, G. W., Butler, R. P., Vogt, S. S., 2000, *ApJ*, **529**, L41.
- H9 Hatzes, A., et al., 2000, *ApJ*, **544**, L145.
- H10 Han, I., Black, D. C., Gatewood, G., 2001, *ApJ*, **548**, L57.
- H11 Hatzes, A. et al., 2003, *ApJ*, **599**, 1383.
- J1 Jayawardhana, R., Fisher, S., Hartmann, L., Telesco, C., Pina, R., Fazio, G., 1998, *ApJ*, **503**, L79.
- J2 Jeans, J. H., 1928, *Astronomy and Cosmogony*, London, Cambridge University Press.
- J3 Jeffreys, H., 1929, *MNRAS*, **89**, 636.
- J4 Jura, M., 1991, *ApJ*, **383**, L79.
- K1 van de Kamp, P., 1969, *ApJ*, **74**, 241.
- K2 van de Kamp, P., 1977, *Vistas in Astr.*, **20**, 501.
- K3 Kant, I., 1755, *Allgemeine Naturgeschichte und Theorie des Himmels*, J. F. Petersen, Konisburg & Leipzig.
- K4 Kasting, J. F., Whitmire, D. P., Reynolds, R. T., 1993, *Icarus*, **101**, 108.
- K5 King, J. R., Deliyannis, C. P., Hiltgen, D. D., Stephens, A., Cunha, K., Boesgaard, A. M., 1997, *AJ*, **113**, 1871.
- K6 Koerner, D. W., 1997, *Origins Life Evol. Bio.*, **27**, 157.
- K7 Koerner, D. W., Ressler, M. E., Werner, M. W., Backman, D. E., 1998, *ApJ*, **503**, L83.
- K8 Koerner, D., LeVay, S., 1999, *Here be Dragons*, Cambridge University Press.
- K9 Kuiper, G., 1951, *Proc. Nat. Acad. Sci.*, **37**, 1.
- K10 Kalas, P., Liu, M. C., Matthews, B. C., 2004, *Science*, **303**, 1990.
- K11 Kasting, J. F., Catling, D., 2003, *ARA&A*, **41**, 420.
- L1 LaPlace, P. S., 1796, *Exposition du Système du Monde*, Circle-Sociale.
- L2 Latham, D. W., et al., 1989, *Nature*, **339**, 38.
- L3 Laughlin, G., Adams, F. C., 1997, *ApJ*, **491**, L51.
- L4 Leighton, R. B., Noyes, R. W., Simons, G. W., 1962, *ApJ*, **135**, 474.
- L5 Lepage, A. J., MacRobert, A. M., 1998, *Sky & Telescope*, **96**, 44.
- L6 Lin, D. N. C., 1986, in *The Solar System: Observations and Interpretations*, ed. M. G. Kivelson, Prentice Hall, p. 68.
- L7 Lin, D. N. C., Ida, S., 1997, *ApJ*, **477**, 781.
- L8 Lin, D. N. C., Papaloizou, J. C. B., 1986, *ApJ*, **309**, 846.
- L9 Lippincott, S. L., 1960, *AJ*, **65**, 445.
- L10 Lippincott, S. L., 1978, *Sp. Sci. Rev.*, **22**, 153.
- L11 Lissauer, J. J., 1993, *ARA&A*, **31**, 129.
- L12 Lissauer, J. J., 1995, *Icarus*, **114**, 217.
- L13 Lowrance, P. J., et al., 1999, *ApJ*, **512**, L69.
- L14 Lyne, A. G., Bailes, M., 1992, *Nature*, **355**, 213.
- L15 Liu, M. C., Matthews, B. C., Williams, J. P., Kalas, P. G., 2004, *ApJ*, **608**, 526.
- L16 Laughlin, G., 2000, *ApJ*, **545**, 1064.
- L17 Laws, C. et al., 2003, *AJ*, **125**, 2664.
- L18 Laws, C., 2004, Ph.d. thesis, University of Washington.
- L19 Laws, C., Gonzalez, G., 2001, *ApJ*, **553**, 405.

- M1 Marcy, G. W., Butler, R. P., 1992, *PASP*, **104**, 270.
- M2 Marcy, G. W., Butler, R. P., 1996, *ApJ*, **464**, L15.
- M3 Marcy, G., Butler, P., 1998, *ARAA*, **36**, 57.
- M4 Martin, E. L., Claret, A., 1996, *A&A*, **306**, 408.
- M5 Mayor, M., Queloz, D., 1995, *Nature*, **378**, 355.
- M6 Mizuno, H., 1980, *Prog. Theor. Phys.*, **64**, 544.
- M7 Moultin, F. R., 1905, *ApJ*, **22**, 165.
- M8 Mayer, L., Quinn, T., Wadsley, J., Stadel, J., 2004, *ApJ*, **609**, 1045.
- P1 Pinsonneault, M. H., Kawaler, S. D., Demarque, P., 1990, *ApJS*, **74**, 501.
- P2 Pollack, J. B., Hubickyj, P., Bodenheimer, P., Lissauer, J. J., Podolak, M., Greenzweig, Y., 1996, *Icarus*, **124**, 62.
- P3 Pravdo, S. H., Shaklan, S. B., 1996, *ApJ*, **465**, 264.
- P4 Pourbaix, D., 2001, *A&A*, **369**, L22.
- P5 Perrier, C. *et al.*, 2003, *A&A*, **410**, 1039.
- P6 Pepe, F. *et al.*, 2004, *A&A*, **423**, 385.
- Q1 Queloz, D. *et al.*, 2000, *A&A*, **354**, 99.
- R1 Rasio, F. A., Ford, E. B., 1996, *Science*, **274**, 954.
- R2 Reuyl, D., Holmberg, E., 1943, *ApJ*, **97**, 41.
- R3 Reuyl, D., 1943, *ApJ*, **97**, 186.
- R4 Reid, I. N., 2002, *PASP*, **114**, 306.
- S1 Saar, W. H., Butler, R. P., Marcy, G. W., 1998, *ApJ*, **498**, L153.
- S2 Safronov, V. S., 1969, *Evolution of the Protoplanetary Cloud and Formation of Earth and Planets*, Moscow, Nauka Press (also NASA TTF-677), 1972.
- S3 Shao, M., Colavita, M. M., 1992, *A&A*, **262**, 353.
- S4 Smith, B. A., Terrile, R. J., 1984, *Science*, **226**, 4681.
- S5 Soderblom, D. R., *et al.*, 1993, *AJ*, **106**, 1059.
- S6 Strand, K. Aa., 1943, *Proc. Amer. Phil. Soc.*, **86**, 364.
- S7 Strand, K. Aa., 1977, *AJ*, **82**, 745.
- S8 Santos, N. C., Israelian, G., Mayor, M., 2001, *A&A*, **373**, 1019.
- S9 Santos, N. C., Israelian, G., Mayor, M., Rebolo, R., Udry, S., 2003, *A&A*, **398**, 363.
- S10 Sigurdsson, S., Richer, H. B., Hansen, B. M., Stairs, I. H., Thorsett, S. E., 2003, *Science*, **301**, 193.
- S11 Sozzetti, A. *et al.*, 2004, *BAAS*, **204**, 6201.
- T1 Taylor, J. H., Weisberg, J. M., 1989, *ApJ*, **345**, 434.
- T2 Telesco, C. M., Becklin, E. E., Wolstencraft, R. D., Decher, R., 1988, *Nature*, **335**, 51.
- T3 Trilling, D., Benz, W., Guillot, T., Lunine, J. I., Hubbard, W. B., Burrows, A., 1998, *ApJ*, **500**, 428.
- T4 Tinney, C. G. *et al.*, 2001, *ApJ*, **551**, 507.
- V1 van Eykum, J. C., Ge, J., Mahadevan, S., DeWitt, C., 2004, *ApJ*, **600**, L79.
- V2 Vogt, S. S., Marcy, G. W., Butler, R. P., Apps, K., 2000, *ApJ*, **536**, 902.
- W1 Walker, G. A. H., Walker, A. T., Irwin, A. W., Larson, A. M., Yang, S. L., Richardson, D. C., 1995, *Icarus*, **116**, 359.
- W2 Ward, W. R., 1981, *Icarus*, **47**, 234.
- W3 Weidenschilling, S. J., Spaute, D., Davis, D. R., Marzari, F., Ohtsuki, K., 1997, *Icarus*, **128**, 429.
- W4 Weissman, P. R., 1984, *Science*, **224**, 987.
- W5 von Weizsäcker, C. F., 1944, *Z. Astrophys.*, **22**, 319.
- W6 Wetherill, G. W., 1990, *Ann. Rev. Earth Planet.*, **18**, 20.
- W7 Wetherill, G. W., 1994, *Astrophys. Sp. Sci.*, **212**, 25.

- W8 Willemse, B., Van Hoolst, T., Smeijers, P., Waelkens, C., 1997, *A&A*, **326**, L37.
- W9 Wolszczan, A., Frail, D. A., 1992, *Nature*, **355**, 145.
- W10 Ward, P. D., Brownlee, D., 2000, *Rare Earths*, Copernicus Press, Springer-Verlag New York.
- W11 Wuchterl, G., Guillot, T., Lissauer, J. J., 2000, in *Protostars and Planets IV*, eds. Mannings, V., Boss, A. P., Russell, S. S., Tucson, Arizona University Press, p. 1081.
- W12 Williams, J. P., *et al.*, 2004, *ApJ*, **604**, 414.
- W13 Wahhaj, Z., Koerner, D. W., Ressler, M. E., Werner, M. W., Backman, D. E., Sargent, A. I., 2003, *ApJ*, **584**, L27.
- Z1 Zuckerman, B., Becklin, E. E., 1993, *ApJ*, **414**, 793.
- Z2 Zucker, S., Maxeh, T., 2001, *ApJ*, **562**, 549.

11.9 HOMEWORK PROBLEMS

Problem 1

Derive equation 11.10, describing the duration of planetary transits.

Problem 2

HD 146233 (Gliese 616) is a close analogue of the Sun, lying at a distance of 14.0 parsecs ($\pi = 71.3$ milliarseconds). Suppose that this star is discovered to be a radial velocity variable, with a period of 119 days and semi-major amplitude 39 m sec^{-1} : calculate $a \sin(i)$ and $M_2 \sin(i)$, under the assumption that the companion has negligible mass compared with the primary. Suppose that the companion is, in fact, a $0.06 M_\odot$ brown dwarf: what is the orbital inclination, and what is the expected amplitude of the astrometric orbit of the primary star?

Problem 3

The subsolar temperature, T_{SS} , at distance r from the Sun is defined as the temperature of a blackbody in equilibrium with the ambient radiation field. This value can be calculated from equation 1.28, substituting r for R and T_{SS} for T_{eff} , and remembering that the luminosity is fixed for a given star (hint: scaling by taking ratios is always a good method). Estimate T_{SS} for Earth ($r = 1 \text{ AU}$) and Mars ($r = 1.52 \text{ AU}$); estimate the subsolar temperatures for Earth and Mars analogues around an M0 dwarf and an M9 dwarf.

Problem 4

Real planets only absorb a fraction of the incident radiation. The planetary albedo, A , gives the fraction of energy reflected; in the case of Earth, $A = 0.4$, so Earth absorbs 60% of the incident flux. This is equivalent to reducing the solar luminosity by 40%, with a corresponding effect on the ambient planetary temperature, T_A . Compute T_A for the Earth. The classical habitable zone is defined as the circum-

stellar region where water can exist in liquid form on a planetary surface given the appropriate atmospheric conditions; i.e. the ambient temperature lies between 0 and 100°C. Determine the limits of the habitable zone for the Sun, an M0 dwarf and an M9 dwarf.

Appendix

THE 8-PARSEC SAMPLE

The Sun's immediate neighbours have attracted detailed scrutiny for well over a century. Following van de Kamp [K1, K2], most catalogues set the distance limit at 5 parsecs. However, the last decade has seen both extensive surveys for additions to the local census (e.g., [C2], [R2]) and intense scrutiny of known nearby stars for previously unrecognised companions (e.g., [D1]). Consequently, it is not unreasonable to extend the local reference sample to include stars (and brown dwarfs) within 8 parsecs of the Sun.

The starting point for the present census was the preliminary version of the third Catalogue of Nearby Stars (the CNS3, [G2]), supplemented by spectroscopic observations from the Palomar/Michigan State University survey (PMSU: [R1], [H1], [G1]). More than a decade has elapsed since the compilation of the CNS3, and a handful of stars and brown dwarfs have been added to the local sample, either companions to known nearby stars (e.g., G 204-47B, Gl 229B, Gl 866C, Gl 845 Bab) or isolated low-luminosity dwarfs (e.g., LP 816-60, 2M1507-16, 2M1835+32, Denis1048-47). As one might surmise from their names, most of the latter dwarfs were identified from either the 2MASS or DENIS infrared surveys.

Tables A.1 and A.2 present photometry, spectral types and distance estimates for the 8-parsec sample (as of 31 December, 2004). These data, together with positions, proper motions and radial velocities, are also available at <http://www.stsci.edu/~inr/nlds2files.html>, while further information on individual objects can be obtained from the SIMBAD database (<http://simbad.u-strasbg.fr/Simbad>). Further information on the nearest stars is given at the RECONS (Research Consortium for Nearby Stars) website (<http://www.chara.hsu.edu/~thenry/RECONS/>).

The current 8-parsec sample includes 191 objects in 142 systems (including the Sun). As discussed in Section 2.7, we have divided this data set by declination:

Table A.1 lists the 109 systems that lie north of declination -30° , while Table A.2 lists the 32 southern systems; since the Sun never descends below $\delta = -24^\circ$, it is included in Table A.1. The northern sample comprises 151 objects, including 139 main sequence stars, 9 white dwarfs and 3 brown dwarfs; the southern sample, drawn from one-third the area of the northern sample, includes 33 main sequence stars, 3 white dwarfs and 3 brown dwarfs. The spectral types listed for these objects are from three main sources: the CNS3 for white dwarfs and early-type ($< K5$) stars; PMSU, for most M dwarfs; and Henry *et al.* [H2], for most late-type ($> M6$) dwarfs. In addition, spectral types are taken from the discovery papers for 2M1835+32 and all brown dwarfs. There are no giant stars in the 8-parsec sample; $\sim 70\%$ of the main sequence stars are M dwarfs, $\sim 15\%$ K dwarfs and $\sim 10\%$ G dwarfs; the most luminous star is Vega (Gl 721, $M_{bol} = 0.44$); the lowest luminosity star is 2M1835+32 (M8.5) at $M_{bol} = 13.1$, while Gl 570D (T8) is the coolest, faintest brown dwarf.

All of the photometry listed in Tables A.1 and A.2 is on the Cousins system at optical wavelengths and the CIT system at near-infrared wavelengths. Most data are taken from Bessel's [B1] and Koen *et al.*'s [K3] optical observations, and from Leggett's [L1] compilation of nearby star photometry (optical and IR), supplemented by 2MASS data. The latter have been transformed to the CIT system using the relations from [C1].

A few stars still lack near-infrared observations; in those cases, we have used one of three techniques to estimate K -band absolute magnitudes from optical photometry. First, we have used Leggett's data for nearby stars to define a $(V-I)/(V-K)$ relation:

$$(V-K) = 1.58(V-I) + 0.82 \quad \sigma_{V-K} = 0.15 \quad 1.0 < (V-I) < 3.5$$

Second, for stars lacking I -band data, we use the $(B-V)/(V-K)$ relation calibrated by stars from the ESO JHK list of standards [B2]:

$$(V-K) = 2.22(B-V) + 0.10 \quad \sigma_{V-K} = 0.12 \quad 0 < (B-V) < 1.1$$

Finally, in the case of stars with no colour information (mainly close companions), the K -band absolute magnitudes are estimated from the approximate relations:

$$M_K \sim 0.75M_V - 1.5 \quad 6 < M_V < 10$$

$$M_K \sim 0.50M_V + 1.0 \quad 10 < M_V < 18$$

Almost all of the stars in the present 8-parsec sample have accurate trigonometric parallaxes. The brighter stars were observed in the course of the *Hipparcos* survey, and therefore have milliarcsecond-accuracy astrometry, while most of the fainter stars have accurate ground-based data from either the US Naval Observatory or, in a few cases, the southern CTIOPI survey (see RECONS web page, cited above). A handful of stars still lack trigonometric parallax measurements, and in those cases (marked with a superscript S in column 11 of Tables A.1 and A.2) the distances are estimated by combining photometric and spectroscopic parallaxes.

REFERENCES

- B1 Bessell, M. S., 1991, *A&AS*, **83**, 357.
B2 Bouchet, P., Manfroid, J., Schmider, F. X., 1991, *A&AS*, **91**, 409.
C1 Carpenter, J. M., 2001, *AJ*, **121**, 2851.
C2 Cruz, K. L., *et al.*, 2003, *AJ*, **126**, 2421.
D1 Delfosse, X., *et al.*, 1999, *A&A*, **344**, 897.
G1 Gizis, J. E., Reid, I. N., Hawley, S. L., 2002, *AJ*, **123**, 3356.
G2 Gliese, W., Jahreiss, H., 1991, *Third Catalogue of Nearby Stars*, CNS3.
H1 Hawley, S. L., Gizis, J. E., Reid, I. N., 1996, *AJ*, **112**, 2799.
H2 Henry, T. J., Kirkpatrick, J. D., Simons, D. A., 1995, *AJ*, **108**, 1437.
K1 van de Kamp, P., 1953, *PASP*, **65**, 30.
K2 van de Kamp, P., 1977, *Vistas in Astr.*, **20**, 501.
K3 Koen, C., Kilkenny, D., van Wyk, F., Cooper, D., Marang, F., 2002, *MNRAS*, **334**, 20.
L1 Leggett, S. K., 1992, *ApJS*, **82**, 351.
R1 Reid, I. N., Hawley, S. L., Gizis, J. E., 1995, *AJ*, **110**, 1838.
R2 Reid, I. N., *et al.*, 2004, *AJ*, **128**, 463.

Table A.1. The northern 8-parsec sample.

Name	M_V	$(U-B)$	$(B-V)$	$(V-R_C)$	$(V-I_C)$	M_K	$(J-H)$	$(H-K)$	Sp.	r	ϵ_r	M_{bol}	Comments	
Sun	4.85	0.65	1.86	1.99	1.60	3.60	9.07	0.53	0.35	G2 V	4.70	0.08	4.67	
GJ 1002	15.47	1.26	1.71	1.39	2.77	7.99	0.64	0.18	M4	M5.5	0.09	11.97	G158-27	
GJ 1005 A	13.38	1.573	1.573	1.55	3.51	9.49	0.24	0.77	M4	M4	0.09	10.95		
GJ 1005 B	15.73	10.32	1.24	1.55	2.14	6.27	0.57	0.22	M1	M3.5	0.09	12.34		
Gl 15 A		13.31	1.40	1.79	1.24	2.82	8.21	0.60	0.24	M3.5	3.56	0.01	8.72	
Gl 15 B		6.36	0.58	0.88	0.51	0.96	4.26	0.48	0.09	K2 V	7.46	0.05	10.90	
Gl 33		4.58	0.02	0.57	0.40	0.67	3.09	0.33	0.06	G3 V	5.94	0.02	5.82	
Gl 34 A		8.64		1.39		4.91	0.58	0.11	K7 V	M	5.94	0.02	4.53	η Cass
Gl 34 B		14.22	0.02	0.55	0.26	0.49	13.28	0.12	0.07	DZ7	4.33	0.06	7.6	V Ma 2
Gl 35		5.81	0.09	0.69		4.18	0.45	0.06	G5 VI	M	7.55	0.03	5.62	μ Cass
Gl 53 A		11.6				6.70 ^b			M	M	7.55	0.02	9.7	
Gl 53 B		14.25	1.33	1.86	1.37	3.15	8.57	0.52	0.29	M4.5	3.72	0.04	11.42	
Gl 54.1		15.40	1.09	1.85	1.65	3.69	8.73	0.60	0.28	M5.5	2.68	0.02	11.70	
Gl 65 A		15.94								M6	2.62	0.03	12.09	UV Ceti
Gl 65 B		5.88	0.49	0.83	0.35	0.86	3.72	0.47	0.08	K1 V	7.47	0.05	5.42	
Gl 68		5.68	0.21	0.72	0.43	0.81	3.96	0.36	0.12	G8 Vp	3.65	0.01	5.25	τ Ceti
Gl 71		14.08	1.35	1.83	1.37	3.12	8.43	0.55	0.29	M4.5	4.45	0.06	11.28	
Gl 83.1		6.52	0.80	0.97	0.57	1.07	4.23	0.50	0.14	K3 V	7.21	0.05	5.95	
Gl 105 A		12.39	1.09	1.61	1.19	2.79	7.32	0.55	0.19	M4	7.21	0.05	10.01	
Gl 105 B		17.40	-0.18	2.40	2.09	4.51	9.23	0.80	0.54	M7	7.21	0.05	12.25	
Gl 105 C		11.18	1.11	1.56	1.10	2.45	6.38	0.56	0.20	M3	7.55	0.14	9.24	
Gl 109		SO 0253+1652	17.42	1.80	1.90	4.24	9.81	0.51	0.29	M6.5	3.55 ^s	0.5	12.85	
LP771-95		13.26		1.69	1.39	2.55	6.9 ^c			M3.5	6.86	0.18	9.85	
LP771-96		13.86					7.1 ^b			M3	6.86	0.18	10.1	
LP771-96A		15.86					7.5 ^b			M4	6.86	0.18	10.3	
Gl 144		6.19	0.58	0.88	0.50	0.94	4.08	0.37	0.07	K2 V	3.23	0.01	5.78	ϵ Eri
Gl 166 A		5.92	0.44	0.82	0.47	0.88	4.08	0.43	0.07	K1 Ve	5.03	0.02	5.56	40 Eri
Gl 166 B		11.01	-0.68	0.03						DA4	5.03	0.02		

GJ 166 C	12.68	1.03	1.63	1.24	2.87	7.45	0.48	0.28	M4.5	5.03	0.02	10.21
GJ 169.1A	12.35	1.21	1.64	1.21	2.81	7.03	0.65	0.26	M4	5.48	0.03	9.96
GJ 169.1B	13.76	-0.52	0.31						DC5	5.48	0.03	
LHS 1723	12.15		1.80	1.29	2.85	8.05	0.55	0.33	M4	7.51 ^s	2.25	9.32
GJ 205	9.19	1.21	1.47	0.97	2.07	5.09	0.68	0.19	M1.5	5.71	0.04	7.67
GJ 213	12.62	1.22	1.60	1.23	2.82	7.46	0.52	0.26	M4	5.87	0.11	10.31
GJ 223.2	15.40	0.77	1.77	0.97	1.87				DZ9	6.46	0.17	
GJ 1087	14.59	-0.14	0.60						DAP9	7.99	0.37	
LHS 1805	12.34		1.60	1.18	2.72	7.27	0.52	0.31	M3.5	7.52	0.16	10.00
G 99-49	12.68	1.10	1.68		7.39	0.60	0.23		M4	5.37	0.17	10.3
GJ 229 A	9.31	1.19	1.50	0.96	2.01	5.34	0.63	0.20	M0.5	5.77	0.04	7.86
GJ 229 B					20.6	15.6	-0.10	-0.10	T6	5.77	0.04	17.7
GJ 234 A	13.02	1.23	1.72	1.31	3.02	7.43	0.63	0.26	M4.5	4.09	0.03	10.33
GJ 234 B	16.5					9.19	0.68	0.41	M	4.13	0.03	12.2
GJ 244 A	1.48	-0.04	0.00	-0.01	-0.02	1.56	0.00	-0.02	A1 V	2.63	0.01	1.28
GJ 244 B	11.33		-0.03						DA2	2.64	0.01	
GJ 251	11.30	1.20	1.57	1.13	2.53	6.55	0.58	0.22	M3	5.57	0.06	9.29
GJ 268 A	13.10	1.19	1.70	1.33	3.04	7.56	0.60	0.28	M4.5	6.14	0.07	10.33
GJ 268 B	13.5					7.77	0.60	0.29	M	6.14	0.07	10.75
GJ 273	11.99	1.12	1.57	1.15	2.70	7.01	0.54	0.25	M3.5	3.79	0.02	9.71
GJ 280 A	2.65	0.03	0.42	0.25	0.49	1.63	0.17	0.04	F5 IV	3.50	0.01	2.67
GJ 280 B	13.0								DA	3.50	0.01	
GJ 285	12.31	0.97	1.61	1.26	2.95	6.85	0.60	0.26	M4.5	5.97	0.08	10.25
GJ 299	13.69	1.19	1.77	1.25	2.92	8.47	0.47	0.28	M4.5	6.84	0.12	11.13
GJ 300	13.26		1.58	1.22	2.90	7.83	0.64	0.25	M4	6.10	0.35	10.63
GJ 1111	17.10	2.11	2.05	2.00	4.26	9.46	0.58	0.36	M6.5	3.63	0.04	12.39
GJ 1116 A	15.47		1.84	1.67	3.78	8.84	0.57	0.28	M5.5	5.23	0.07	11.66
GJ 1116 B	16.33								M	5.23	0.07	
LHS 2090	17.21									6.0 ^s	0.9	12.62
GJ 338 A	8.69	1.20	1.41	0.85	1.74	5.13	0.61	0.21	K7	6.17	0.09	7.52
GJ 338 B	8.72	1.17	1.42			5.15	0.62	0.19	K7	6.17	0.09	7.55
LHS 6167	14.64			1.45	3.34	8.61	0.52	0.35	M4	6.71 ^s	0.96	11.50

(continued)

Table A.1—(continued)

Name	M_V	$(U-B)$	$(B-V)$	$(V-R_C)$	M_K	$(J-H)$	$(H-K)$	Sp.	r	ϵ_r	M_{bol}	Comments
G 161-71	14.76		1.41	3.34	8.62	0.59	0.30	M4	6.21 ^s	0.84	11.64	
Gl 380	8.16	1.28	1.38	0.85	1.62	4.78	0.62	K2 V	4.86	0.02	7.12	
Gl 382	9.80		1.50	0.99	2.17	5.63	0.64	M1.5	7.81	0.09	8.16	
Gl 388	10.87	1.06	1.53	1.09	2.51	6.16	0.64	M3	4.89	0.07	8.84	AD Leo
Gl 393	10.33	1.23	1.51	1.01	2.24	6.04	0.56	M2	7.23	0.11	8.62	
LHS 292	17.45		2.10	2.06	4.40	9.68	0.58	M6.5	4.54	0.07	12.59	
Gl 402	12.51	1.06	1.64	1.21	2.79	7.26	0.61	M4	6.80	0.23	10.14	
Gl 406	16.63	1.59	1.99	1.88	4.06	9.19	0.62	M6	2.39	0.01	12.39	Wolf 359
Gl 408	10.90	1.22	1.54	1.05	2.39	6.32	0.59	M2.5	6.66	0.07	9.02	
Gl 411	10.44		1.14	1.51	1.01	2.15	6.33	M2	2.54	0.01	8.84	Lalande 21185
Gl 412 A	10.33	1.18	1.54	1.00	2.02	6.33	0.54	M0.5	4.85	0.03	8.87	
Gl 412 B	16.01		2.08	1.66	3.77	9.42	0.57	M6	4.85	0.03	12.22	WX UMa
Gl 445	12.17	1.08	1.58	1.12	2.64	7.28	0.51	M3.5	5.38	0.04	98.97	
Gl 447	13.50		1.34	1.76	1.30	2.98	8.01	M4	3.34	0.02	10.88	Ross 128
Gl 1156	14.79	1.01	1.88	1.51	3.46	8.56	0.52	M5	6.54	0.13	11.48	
Gl 473 A	14.83	1.18	1.81	1.54	3.54	8.57	0.61	M5	4.39	0.08	11.43	
Gl 473 B	15.10		1.49	0.96	2.01	8.50	0.53	M7	4.39	0.08	11.45	
Gl 514	9.64		1.22	0.96	2.04	5.65	0.61	M0.5	7.63	0.08	8.18	
Gl 526	9.80	1.11	1.43	0.96	2.04	5.79	0.59	M1.5	5.43	0.04	8.31	
Gl 555	12.38	1.20	1.63	1.24	2.86	7.07	0.59	M4	6.14	0.10	9.92	
Gl 566 A	5.55	0.22	0.73		3.83 ^a			G8 V	6.78	0.03	5.42	ξ Boo
Gl 566 B	7.82	1.15	1.16		4.36 ^b			K4 V	6.78	0.03	7.28	
LHS 3003	18.02		1.50	2.17	4.52	9.90	0.60	M6.5	6.40	0.12	13.00	
Gl 570 A	6.90	1.06	1.10	0.70	1.34	4.32	0.58	K5 V	5.89	0.06	6.29	
Gl 570 B	9.24	1.22	1.48	1.00	2.12	5.08	0.61	M1	5.89	0.06	7.65	
Gl 570 C	11.6					6.58	0.72	M	5.89	0.06	9.5	
Gl 570 D					20.53	16.42	0.05	T8	5.89	0.06	17.5	
2M1507-16	23.56			3.86	6.25	12.06	0.90	L5	7.36	0.03	14.20	
Gl 581	11.58	1.21	1.60	1.10	2.51	6.86	0.61	M3	6.28	0.09	9.56	

Gl 625	11.00	1.28	1.60	0.99	2.21	6.74	0.57	0.23	M1.5	6.61	0.02	9.32
Gl 628	11.93	1.12	1.57	1.17	2.68	6.86	0.54	0.27	M3.5	4.24	0.03	9.70
Gl 643	12.72	1.35	1.69	1.19	2.73	7.70	0.51	0.30	M3.5	6.45	0.03	10.45
Gl 644 A	10.67	1.13	1.57	1.10	2.46	6.05	0.64	0.25	M3	6.45	0.03	8.70
Gl 644 B	10.8					6.1			M	6.45	0.03	8.9
Gl 644 D	12.0:					8.4			M	6.45	0.03	9.8
Gl 644 C	17.73		2.20	2.15	4.56	9.77	0.58	0.37	M7	6.45	0.03	12.70
G 203-47 A	12.37	0.89	1.46	1.20	2.81	7.10	0.6	0.24	M3.5	7.63	0.18	9.98
G 203-47 B	>15				>8				WD	7.63	0.18	
Gl 661 A	10.99	1.01	1.49	1.10	2.51	6.41	0.54	0.21	M3.5	6.40	0.05	8.96
Gl 661 B	11.37					6.72	0.49	0.25	M	6.40	0.05	9.4
Gl 663 A	6.18	0.49	0.86	0.49	0.93	4.18			K1 V	5.97	0.03	5.80
Gl 663 B	6.23		0.86			4.23			K1 V	5.97	0.03	5.82
Gl 664	7.44	1.08	1.16	0.70	1.28	4.58	0.50	0.20	K5 V	5.97	0.04	6.73
Gl 673	8.09	1.27	1.36	0.85	1.61	4.74	0.61	0.13	K5	7.72	0.06	7.05
Gl 687	10.94	1.04	1.50	1.09	2.50	6.24	0.57	0.22	M3	4.54	0.02	8.93
Gl 1221	15.21	-0.30	0.40						DXP9	6.14	0.11	
Gl 699	13.25	1.26	1.73	1.21	2.78	8.20	0.50	0.26	M4	1.83	0.01	10.90
LP 71-82	14.25					8.41	0.50	0.26	M4.5	7.09	1.07	Barnard's star
Gl 701	9.90		1.50	0.98	2.06	5.88	0.60	0.22	M1	7.80	0.09	8.40
Gl 702 A	5.48	0.51	0.86	0.51	0.96	3.68	0.47	0.06	K0 V	5.10	0.04	5.07
Gl 702 B	7.47		1.15			4.8			K5 V	5.10	0.04	6.92
Gl 1224	14.18		1.79	1.39	3.18	8.58	0.43	0.42	M4.5	7.70	0.20	11.28
LHS 3376	14.19		1.80	1.40	3.12	8.65	0.48	0.31	M4.5	7.26	0.18	11.44
LP 229-17	12.12		1.42	1.16	2.68	6.80	0.65	0.43	M3.5	7.25 ^S	1.63	9.89
2M1835+32	19.50				4.81	10.38	0.69	0.43	M8.5	5.67	0.02	13.12
Gl 721	0.58	-0.01	0.00	-0.01	0.00	0.58	0.00	0.00	A0 V	7.76	0.03	Vega
Gl 725 A	11.16	1.11	1.54	1.07	2.46	6.70	0.53	0.23	M3	3.53	0.02	9.19
Gl 725 B	11.94	1.14	1.59	1.12	2.55	7.23	0.52	0.23	M3.5	3.53	0.02	9.87
Gl 729	13.13		1.72	1.23	2.78	8.03	0.54	0.28	M3.5	2.97	0.02	10.78
Gl 752 A	10.27	1.16	1.50	1.03	2.32	5.82	0.62	0.22	M2.5	5.85	0.05	8.47
Gl 752 B	18.36		2.20	4.33	9.96	0.68	0.42	M8	5.85	0.05	12.80	VB 10

(continued)

Table A.1—(continued)

Name	M_V	$(U-B)$	$(B-V)$	$(V-R_C)$	M_K	$(J-H)$	$(H-K)$	Sp.	r	ϵ_r	M_{bol}	Comments
GJ 764	5.87	0.37	0.78	0.45	0.87	4.10	0.42	K0 V	5.76	0.02	5.54	σ Dra
GJ 768	2.22	0.08	0.22	0.14	0.27	1.78	0.08	A7 IV	5.13	0.02	2.32	α Aql
GJ 1245 A	15.12		1.90	1.59	3.60	8.60	0.52	0.37	M5.5	0.54	0.02	11.60
GJ 1245 B	16.02		1.98	1.66	3.73	9.15	0.50	0.39	M5.5	0.54	0.02	12.10
GJ 1245 C	18.16				10.15	0.66	0.40	M	4.70	0.10	12.80	
GJ 793	11.04	1.21	1.56	1.06	2.43	6.44	0.60	0.23	M2.5	7.96	0.07	9.11
LP 816-60	12.71		1.65			7.50	0.57	0.32	M4	5.49	0.11	10.0
GJ 809	9.26	1.24	1.49	0.92	1.99	5.40	0.65	0.19	M0.5	7.04	0.04	7.83
GJ 820 A	7.49	1.11	1.19	0.67	1.27	4.70	0.50	0.19	K5	3.50	0.01	6.75
GJ 820 B	8.31	1.22	1.37		1.62	4.98	0.65	0.14	K5	3.50	0.01	7.27
GJ 829 A	11.91	1.31	1.61	1.11	2.58	7.06	0.58	0.25	M3.5	6.74	0.08	9.80
GJ 829 B	11.91	1.31	1.61	1.11	2.58	7.06	0.58	0.25	M3.5	6.74	0.08	9.80
LHS 3799	13.92		1.84	1.41	3.22	7.98	0.60	0.32	M4.5	7.48	0.31	11.06
GJ 860 A	11.82	1.23	1.66	1.19	2.68	7.02	0.47	0.30	M3	4.03	0.02	9.56
GJ 860 B	13.27		1.80			7.62	0.62	0.07	M4	4.03	0.02	10.7
GJ 866 A	15.39	1.54	1.97	1.68	3.90	8.91	0.63	0.34	M5.5	3.45	0.05	11.7
GJ 866 B	15.94				4.20	8.96	0.53	0.34	M	3.45	0.05	11.9
GJ 866 C	17.43				4.65	9.31			M	3.45	0.05	12.5
GJ 873	11.76	1.10	1.58	1.19	2.69	6.70	0.59	0.23	M3.5	5.05	0.05	9.51
GJ 876	11.82	1.14	1.56	1.20	2.74	6.79	0.60	0.26	M4	4.70	0.04	9.51
GJ 880	9.47	1.19	1.48	0.99	2.11	5.35	0.63	0.21	M1.5	6.86	0.05	7.90
GJ 881	2.08	0.06	0.09	0.05	0.08	1.91	0.00	0.02	A3 V	7.69	0.05	1.75
GJ 892	6.49	0.88	1.00	0.57	1.05	4.11	0.48	0.08	K3 V	6.52	0.02	6.10
GJ 896 Aa	11.40	1.06	1.71	1.22	2.84	6.45	0.55	0.30	M3.5	6.34	0.10	9.00
GJ 896 Ab	12.99		1.65			7.57	0.79	0.13	M4.5	6.34	0.10	10.8
GJ 896 B	13.4					7.60			M4.5	6.34	0.11	11.0
GJ 1286	15.37		1.95	1.59	3.63	9.01	0.61	0.28	M5.5	7.24	0.18	11.81
GJ 905	14.85	1.48	1.90	1.52	3.45	8.45	0.63	0.32	M5	3.17	0.01	11.55
GJ 908	10.11	1.09	1.46	0.95	2.03	6.18	0.52	0.22	M1	5.94	0.05	8.63

^a \mathbf{M}_K inferred from $(B-V)/(V-K)$ relation; ^b \mathbf{M}_K inferred from M_V/M_K relation; ^b \mathbf{M}_K inferred from $(V-I)/(V-K)$ relation.

Table A.2. The southern 8-parsec sample.

Name	M_V	$(U-B)$	$(B-V)$	$(V-R_C)$	M_K	$(J-H)$	$(H-K)$	Sp.	r	ϵ_r	M_{bol}	Comments
GJ 1	10.36	0.96	1.45	0.97	2.12	6.32	0.51	M1.5	4.36	0.02	6.78	
GJ 19	3.45	0.10	0.62	0.35	0.69	2.28	0.32	G2 IV	7.48	0.02	3.23	
LHS 1513	12.07	1.22	1.51	1.39	2.13	9.28	0.67	M3.5	7.73 ^s	1.16	9.9	
GJ 139	5.35	0.22	0.71	0.41	0.80	3.54	0.36	G5 V	6.06	0.02	5.05	
GJ 1061	15.20	1.52	1.90	1.59	3.58	8.80	0.53	M5.5	3.80 ^s	0.57	11.71	
LP 944-20	20.18				4.50	11.10	0.70	M9	4.97	0.10	13.63	BR10037-3535, BD
GJ 191	10.90	1.21	1.56	0.95	1.95	7.11	0.49	K7	3.92	0.01	9.50	Kapteyn's star
GJ 293	14.88	-0.17	0.66					DQ9	6.91	0.93		
LHS 288	15.64		1.82	1.51	3.61	9.47	0.23	M	4.49	0.27	12.10	
Denis1048	19.30			3.24	4.87	10.42	0.63	M9	4.03	0.04	13.27	brown dwarf
GJ 440	13.20	-0.63	0.19					DQ6	4.62	0.04		
GJ 480.1	12.82	1.39	1.73	1.17	2.61	7.99	0.51	M3.5	7.66	0.23		
LHS 337	13.33	1.25	1.69	1.91	2.98	7.98	0.42	M4.5	7.62 ^s	1.14		
GJ 551	15.49	1.37	1.85	1.63	3.62	8.80	0.58	M5.5	1.29	0.01	11.80	Proxima Cen
GJ 559 A	4.38	0.23	0.64	0.36	0.69	2.86 ^b		G2 V	1.35	0.01	4.50	α Cen
GJ 559 B	5.71	0.63	0.84	0.47	0.88	3.75 ^b		K0 V	1.35	0.01	6.0	4, B
GJ 588	10.45	1.11	1.51	1.07	2.39	5.92	0.61	M2.5	5.93	0.05		
GJ 667 A	7.02	0.82	1.04	0.64	1.20	4.61 ^b		K3 V	7.25	0.31	6.25	B
GJ 667 B	7.90					7.80 ^b		K5 V			6.6	B
GJ 667 C	10.75	1.15	1.57	0.90	2.12	6.54	0.55	M1.5			9.1	

(continued)

Table A.2—(continued)

Name	M_V	$(U-B)$	$(B-V)$	$(V-R_C)$	$(V-I_C)$	M_K	$(J-H)$	$(H-K)$	Sp.	r	ϵ_r	M_{bol}	Comments	
GJ 674	11.09	1.21	1.55	1.08	2.41	6.60	0.56	0.24	M2.5	4.54	0.03	9.4		
GJ 682	12.46	1.20	1.64	1.25	2.82	7.16	0.61	0.28	M3.5	5.05	0.06	10.05		
GJ 2130A	11.5		1.46			7.29	0.58	0.28	M	6.18	0.43	9.99		
GJ 2130B	12.8		1.43			7.63	0.60	0.27	M	6.18	0.43	11.42		
GJ 2130C	13.8					8.04	0.60	0.40	M	6.18	0.43	12.2		
GJ 693	11.95		1.63	1.16	2.56	6.83	0.56	0.25	M2	5.81	0.09	9.88		
SCR1845-6357	19.67			2.40	4.93	10.79	0.57	0.46	M8.5	3.5	0.3	12.9		
GJ 754	13.38	1.22	1.66	1.29	2.96	7.95	0.54	0.25	M4.5	5.90 ^s	0.43	10.80		
GJ 780	4.62	0.45	0.76	0.41	0.76	3.15	0.28	0.08	G8 V	6.11	0.02	4.30	δ Pav	
GJ 783 A	6.56	0.46	0.87	0.52	1.00	4.353			K3 V	6.05	0.03	6.10		
GJ 783 B	12.55					7.20			M3.5					
GJ 784	9.01	1.22	1.41	0.91	1.83	5.32	0.62	0.18	K7	6.20	0.04	7.75		
GJ 825	8.69	1.22	1.42	0.90	1.76	5.11	0.63	0.17	K7	3.51	0.01	7.50		
GJ 832	10.19	1.20	1.48	1.01	2.18	6.02	0.55	0.22	M1.5	4.94	0.03	8.55		
GJ 845A	6.89	0.99	1.05	0.62	1.14	4.43	0.55	0.08	K5 V	3.63	0.01	6.29	ϵ Indi	
GJ 845Ba						19.0	13.55	0.78	T1	3.63	0.01	16.5	ϵ Indi Ba	
GJ 845Bb						20.5	15.73	-0.04	-0.26	T6	3.63	0.01	17.5	ϵ Indi Bb
GJ 879	7.04	1.02	1.10	0.66	1.21	4.48	0.72	-0.03	K5 V	7.64	0.05	6.40		
GJ 887	9.77	1.18	1.50	0.98	2.02	5.90	0.72	0.11	M0.5	3.29	0.01	8.29		
GJ 915	13.59	-0.87	0.07			12.98	0.17	-0.02	DA5	7.80 ^s	0.55			

^a M_K inferred from (B-V)/(V-K) relation; ^b M_K inferred from M_V/M_K relation; ^c M_K inferred from (V-I)/(V-K) relation.

Index

2MASS (Two Micron All-Sky Survey) 28, 254

47 Tucanae 526

51 Pegasi 512, 513

α elements 177, 302, 309

α Persei (cluster) 205, 331, 392, 435

$\alpha\Omega$ dynamo 186

absolute magnitude 23

absorption coefficient 156

absorption line profile

Doppler profile 150

Lorentz profile 150, 161

microturbulent velocity 161

Voigt profile 151

activity

age/activity relations 216

Babcock model dynamo 187

Balmer line emission 81, 194, 196, 209, 213, 216, 222, 268

Ca II H&K lines 80, 194, 202

chromospheric emission 80, 195, 271

coronal emission 83

emission measure (EM) 199

flares 84, 206, 271

photometric effects 214

radio emission 222

rotation 223, 271

rotational dynamo 186

shell dynamo 187

star spots 196, 202, 273

turbulent dynamo 188

ultracool dwarfs 271

ultraviolet emission 198, 207, 226

X-ray emission 84, 201, 207, 222

AD Leo (Gl 388) 85, 192, 193, 200, 207, 211

adaptive optics 38

Air Force Geophysical Laboratory (AFGL) survey 28

alt-azimuth co-ordinate system 4

apparent magnitude 19, 23

array detectors 17

asymmetric drift 290

atmospheric refraction 4

atmospheric seeing 35

atmospheric transmission (terrestrial) 2

AU Mic (Gl 803) 62, 213, 508

autumnal equinox 6

β Pic 330, 506

Babcock model dynamo 187

Balmer line emission 81, 149, 196, 209, 213, 216, 272

- Barnard's star 69, 169, 466, 500
 baryonic matter 142
 Bessel spectral classification system 50
 binary stars
 activity 220
 and the luminosity function 366
 astrometric 408
 brown dwarf desert 520, 521
 companion star frequency (c.s.f.) 88
 eclipsing 412
 frequency 88, 137, 276, 277, 521
 mass determination 409
 mass ratios 91, 277, 520
 multiple star fraction (m.s.f) 88
 multiple systems 88
 search techniques 88, 241, 246
 speckle imaging 37
 spectroscopic, double-line 411
 spectroscopic, single-line 412
 unresolved systems 411
 white dwarf/red dwarf systems 241, 246, 376
 birthrate function 407
 black-body distribution 23, 153
 bolometric corrections 73, 259
 bolometric magnitude 23
 Boltzmann equation 148
 bound–bound process 148, 158
 bound-free process 149, 158
 broadband filter characteristics 20
 broadband photometric systems 20
 brown dwarfs
 and planets 239, 490
 as companions 241, 246, 248, 250, 257, 275
 atmospheric dust 55, 170, 178, 264
 binaries 137, 277
 definition 118, 120, 238
 evolution 120, 238
 formation 136
 in clusters 252
 infrared surveys 245, 254
 lithium test 55, 111, 239
 mass function 427, 435, 440, 443, 446
 models, Lyon group 122, 126, 266
 models, Tucson group 120, 122, 126
 temperatures 123, 234, 259
 variability 271
 BY Dra stars 203, 206, 224
- CCD (charge-coupled device) 17
 C dwarfs 56
 C giants 56
 Ca II H&K emission 80, 194, 202
 chemical abundance, stellar 154, 268
 chromosphere 80, 194
 circumstellar disks 135, 137–138, 438, 441, 495, 506, 515
 CM Dra (Gl 630.1A) 69, 133, 418
 CNO fusion cycle 108
 co-ordinate systems
 alt-azimuth 4
 ecliptic 6
 equatorial 5
 Galactic 8
 colour excess 25
 colour–magnitude diagrams 70, 261, 266
 convection 105, 111, 155, 186
 convergent point 380
 corona 80, 83, 194, 198
 coronal loops 199, 210
 coronagraph 250
 COROT 412, 528
 Coulomb interactions 110, 115, 119
 Cousins photometric system 20
 creation function 407
 CU Cnc (GJ 2069) 69, 133, 418
 curve of growth 151, 307
- dark current 17
 dark matter 88, 346, 428, 450, 457
Darwin mission 533
 DDO photometric system 305
 debris disk 506
 Deep Near-Infrared Survey (DENIS) 27, 254
 degeneracy 113, 118, 132

- detectors
 - array detectors 17
 - charge-coupled devices (CCDs) 17
 - infrared 18
 - photomultipliers 16
 - proportional counters 16
 - quantum efficiency (QE) 15
 - semiconductors 17
- deuterium fusion (deuterium burning) 110, 116
- differential emission measure (DEM) 199
- diffraction grating 33
- diffraction limit 35
- Digitised Sky Survey (DSS) 26
- distance modulus 23
- distance scale 311, 487
- Doppler line profile 150
- Drake equation 493
- dust
 - circumstellar 438, 441, 506
 - interstellar 24, 134, 304, 340, 436
 - in stellar/brown dwarf atmospheres 55, 170, 178, 264
- dynamo theory
 - $\alpha\Omega$ dynamo 186
 - Babcock model 187
 - turbulent dynamo 188
- ecliptic co-ordinate system 6
- Eddington limit 103
- effective temperature 24, 154, 268
- effective wavelength 20
- electromagnetic spectrum 1
- emission measure (EM) 199
- energy transport 104, 111
- equation of state 105, 113, 155
- equation of stellar statistics 340
- equatorial co-ordinate system 5
- equivalent width 151
- excitation potential 148
- extinction
 - terrestrial atmospheric 22
 - interstellar 24, 340, 375
- extrasolar planets
 - and brown dwarfs 239, 498, 513, 520
 - and M dwarfs 142, 517
 - astrometric detection 499
 - atmospheres 524
 - direct imaging 501, 533
 - Doppler detection 508
 - formation 495, 515, 526
 - habitability 494, 533
 - host star chemistry 524
 - hot jupiters 513, 515, 533
 - mass distribution 518
 - microlensing 505
 - multiple planets 517, 522
 - orbital eccentricity 523
 - orbital migration 515, 526
 - period distribution 523
 - photometric detection 502
 - pulsar planets 508
 - radii 523
 - semi-major axis distribution 520, 522
 - surveys 499
 - transits 503
- flare stars 80, 84
- flares, solar 210
- flares, stellar 84, 206
- flux 19
- flux conservation 155
- flux tubes 186, 210
- focal length 12
- focal ratio 13
- free-free process 158
- fusion reactions 108, 139
- G 29-38 241
- Galactic Bulge 314
- Galactic co-ordinate system 8
- Galactic Disk
 - abundance distribution 323
 - age 324
 - age-velocity relation 317
 - density distribution 320
 - formation 326

- Galactic Disk (*cont.*)
kinematics 290, 316
luminosity function 92, 364, 394
mass density, local 350, 450
mass, total 293, 316
missing mass 350, 450
moving groups 319
old disk 322
radial density distribution 320
rotation 290, 316
star formation history 327
thick disk 322, 324, 326
vertical density distribution 218, 321
- Galactic Halo
abundance distribution 302, 308
age 295, 308
density distribution 296, 298
ELS model 312
formation 302, 312
kinematics 290, 301
mass density, local 481
mass, total 296, 293
rotation 290, 301
satellite galaxy accretion 301, 313
SZ model 313
- gas polytropes 105
giant molecular clouds 134, 137, 317
GD 165B 52, 246
Gl 65B (UV Ceti) 80, 85
Gl 191 (Kapteyn's star) 69, 76, 193
Gl 278C (YY Gem) 69, 133, 169, 418
Gl 229 AB 55, 169, 250
Gl 388 (AD Leo) 85, 1920, 193, 200,
207, 211
Gl 406 (Wolf 359) 49, 68, 169, 245
Gl 630.1A (CM Dra) 69, 133, 418
Gl 644C (VB 8) 68, 169, 236, 245
Gl 699 (Barnard's star) 69, 169, 466, 500
Gl 752B (VB 10) 51, 80, 157, 169, 226,
245
Gl 803 (AU Mic) 62, 213, 508
Gl 866ABC 169
GJ 2069 (CU Cnc) 69, 133, 418
globular clusters 291, 295, 297, 298, 307,
309, 476
- Gould's Belt 329
grating equation 33
gravitational lensing 413
Gunn photometric system 19, 20
- H⁻ ion 77, 153, 175
H α emission 81, 149, 190, 194, 216, 225,
271
- Halo subdwarfs
and proper motion surveys 347, 459
colour-magnitude diagram 75, 297
luminosity function 468, 475
local mass density 487
mass function 481, 486
metallicity 76, 291, 296, 303
models 130, 174
photometric properties 77, 458
searches for 458
spectral classification 76, 463
spectroscopic bandstrengths 76, 175,
463
temperatures 177
- Harvard spectral classification system 47
- Hayashi tracks 137
- Herbig-Haro objects 135, 138
- heliocentric velocity 11
- Hertzsprung-Russell diagram 31, 48, 66,
126
- Hipparcos* satellite 10, 47, 389
- Hill sphere 496
- hot Jupiters 513, 515, 533
- HR 4796 506
- Hubble Space Telescope 49, 90, 411, 474,
477, 487, 496, 504, 515, 524
- Hyades 83, 84, 219, 252, 330, 384, 433
- hydrogen-burning limit 87, 108, 115, 118,
121, 124, 238
- hydrogen fusion 108, 139
- hydrostatic equilibrium 102, 104, 154
- IC 348 266, 329, 331, 439, 443
IC 2391 205, 330, 393
IC 2602 205, 331, 393

- infrared
 - detectors 18
 - photometry 20, 72, 79
 - radiation 2
 - spectroscopy 58, 62
 - surveys 26, 245, 254
- Infrared Astronomical Satellite (IRAS)
 - 28, 245
- initial mass function 405, 428, 443, 446
- instrumental magnitude 22
- interferometry 38, 530
- interstellar absorption 24, 134, 304, 340, 436
- ionisation 149, 159
- isoplanatic angle 39
- IRAS (Infrared Astronomical Satellite) 28
- James Webb Space Telescope (JWST) 281
- Jeans mass 136, 448
- Johnson photometric system 20
- Jupiter 132, 226, 496, 502, 512, 523
- Kapteyn's star (Gl 191) 69, 76, 458
- Kelu 1 52, 253, 265, 271
- Kelvin–Helmholtz timescale 103, 108
- Kepler mission 504, 528
- KHM spectral classification system 49
- Kuiper belt 502, 506
- λ Orionis 445
- L dwarfs
 - activity 225, 271
 - atmospheric dust 55, 129, 178, 264
 - bolometric corrections 73, 259
 - discovery 52, 246, 253, 254
 - flares 226, 271
 - luminosities 123, 259
 - luminosity function 372
 - masses 123, 429
 - mid-infrared observations 62
 - near-infrared observations 58, 258
 - photometric properties 72, 258
- radii 125, 132, 270
- radio emission 227, 271
- rotation 225, 271
- spectral classification 52
- spots 226, 273
- subdwarfs 468
- temperatures 180, 263, 268
- Lane's law 102
- Lane–Emden equations 105
- LHS 2065 52, 245
- LHS 2397a 224, 245
- LHS 2924 11, 52, 169, 224, 245
- line blanketing 159, 170
- lithium fusion (lithium burning) 110, 239
- Local Bubble 316, 329
- local thermodynamic equilibrium (LTE) 157
- Lorentz line profile 150, 161
- low-mass stars
 - atmospheres 147
 - core temperature 118, 120
 - entropy 117
 - evolution 139
 - formation 102, 134
 - Hertzsprung–Russell diagram 136
 - hydrogen-burning limit 87, 108, 115, 118, 121, 124, 239, 263
 - masses 408, 418
 - models, D'Antona & Mazzitelli 118, 126, 130
 - models, Lyon group 122, 126, 129, 264
 - models, Tucson group 120, 122, 126, 263
 - radii 69, 125, 132
 - surface gravities 154
 - temperatures 69, 164, 169, 180
- LP 944-20 229, 253
- L subdwarfs 468
- luminosity function
 - and binary stars 366
 - definition 92
 - Galactic Disk 92, 364, 394
 - Galactic Halo 468, 475
 - globular clusters 476
 - history 341

- nearby stars 92, 364, 394
- open clusters 383, 386
- ultracool dwarfs 374
- Lutz–Kelker corrections 354

- M4 479, 484, 509
- M dwarfs
 - activity 79, 185
 - atmospheric dust 55, 129, 170, 178
 - bolometric corrections 73, 263
 - early observations 46
 - evolution 139
 - flares 206
 - luminosities 69, 123, 169
 - masses 87, 123, 169
 - mass loss 213
 - magnetic fields 189
 - metallicity 76, 174
 - mid-infrared observations 62
 - models, D’Antona & Mazzitelli 126
 - models, Lyon group 122, 126
 - models, NextGen 162, 164, 168, 171
 - models, Tucson group 120, 122, 126
 - near-infrared observations 58
 - photometric properties 70, 169, 213, 368, 371, 382
 - radii 69, 125, 132, 169
 - radio observations 62
 - rotation 205, 223
 - spectral classification 48
 - spectroscopic bandstrengths 76, 175, 214
 - spectrum synthesis 166
 - surface gravity 154, 169, 270
 - temperatures 69, 127, 164, 169
- M giants
 - dwarf/giant gravity indicators 64
 - spectral classification 65
- M subdwarfs
 - distance scale 487
 - hydrogen-burning limit 121
 - identification 75
 - metallicity 76, 465
 - models 130, 174
- old disk 78
- photometric properties 77, 458
- searches for 458
- spectral classification 76, 175, 463
- spectroscopic bandstrengths 76, 175, 463
- temperatures 177
- ultraool subdwarfs 466
- magnetic fields
 - dynamics 186
 - on G dwarf 190
 - on M dwarfs 191
 - on the Sun 188
 - Zeeman effect 189
- magnitude
 - absolute 23
 - apparent 19
 - bolometric 23
 - instrumental 22
 - measurement 21
 - systems 20
- Malmquist bias 357
- mass function
 - definition 94, 406
 - Galactic Disk 94, 416, 428
 - Galactic Halo 481
 - globular clusters 484
 - high-mass stars 406
 - initial mass function 405, 428, 443, 446
 - nearby stars 94, 426
 - open clusters 433
 - Salpeter 405
 - substellar 427, 435, 440, 443, 446
- mass loss 213
- mass–luminosity relationship 117, 124, 131, 404, 424
- Maxwell–Boltzmann velocity distribution 110
- metallicity, stellar 76, 154, 174, 303
- microlensing 413
- microturbulent velocity 161
- Mie scattering 24
- missing mass (Galactic Disk) 350, 450
- mixing-length approximation 104, 112, 155

- molecular dissociation 149
- Morgan–Keenan (MK) spectral classification system 49
- Mount Wilson spectral classification system 49
- multiple star systems 88
- narrowband photometry 20, 66, 304
- nearby star catalogues 361
- Neupert effect 211
- NGC 2420 384
- NGC 2516 219, 393
- NGC 6397 477, 480, 484
- nuclear fusion
 - deuterium burning 110, 116
 - CNO cycle 108
 - lithium burning 110, 239
 - proton–proton (PP) chain 108, 139
- nucleosynthesis 109, 139, 302, 309
- nutation 6
- old disk subdwarfs 78
- opacity
 - absorption coefficient 112, 155
 - continuum 158, 175
 - grey atmosphere 157
 - H^- 77, 153, 164, 175
 - H_2 (pressure induced) 77, 159, 175
 - line 158, 164
 - opacity distribution function 159
 - opacity sampling 159
 - scattering 158
 - Rosseland mean 108, 157, 162
- open clusters
 - α Persei 205, 392, 435
 - and activity 218
 - dynamical evolution 379
 - Hyades 83, 84, 219, 252, 330, 384, 433
 - IC 2391 205
 - IC 2602 205
 - luminosity functions 385
 - mass functions 433
 - membership criteria 375
- Pleiades 83, 84, 202, 205, 219, 225, 252, 388, 433
- Praesepe 291, 330, 433
- Orion Nebula Cluster 90, 137, 270, 331, 441, 443
- optical depth 155, 162
- Oschin Schmidt telescope 25
- Palomar Observatory Sky Survey (POSS I, POSS II) 25, 254
- parallax
 - photometric 71, 350, 365
 - spectroscopic 66, 373
 - trigonometric 8
- Parenago’s discontinuity 317, 406
- Pauli exclusion principle 113
- PC0025+0447 224, 225, 227, 273
- perfect gas law 107, 111, 155
- photoelectric effect 15
- photography 15
- photometers 16
- photometric systems
 - Cousins 20
 - DDO 305
 - Gunn 19–20
 - Johnson 20
 - Strömgren 304
 - Washington 305
- photomultiplier 16
- photosphere
 - dust formation 55, 178, 264
 - gravity 154, 268, 269
 - metallicity 154, 268
 - opacity 155, 158
 - spectral line formation 150
 - temperature 154, 156, 169, 180, 267
- Planck formula 23
- plate scale 12
- Pleiades 83, 84, 137, 202, 205, 219, 225, 252, 329, 388, 433
- point-spread function 36
- POSS I, POSS II (Palomar Observatory Sky Survey) 25, 254
- Poynting–Robertson effect 233

- Praesepe 330, 391, 433
 precession 6
 probability plots 318
 proper motion
 measurement 10
 surveys 93, 253, 344, 370, 459
 proplyds (protoplanetary disks) 138, 441, 506
 proportional counters 16
 proton–proton chain (P–P cycle) 108, 139, 239
 protoplanetary disks (proplyds) 138, 441, 506
 protostars 135, 137
 pulsars 493
 pulsar planets 508
- quantum efficiency (QE) 15
- ρ Ophiuchus cluster 270, 331, 455
 radiative core 112, 115
 radiative transfer equation 155
 radio observations 62, 271
 Rayleigh scattering 2, 24
 Rayleigh–Jeans approximation 24
 reduced proper motion 344, 462
 relativistic degeneracy 113
 Rossby number 223
 rotation 203, 223, 272
 rotational dynamo 186
 RR Lyrae stars 291, 298, 470
 Russell–Vogt theorem 74, 297
- σ Orionis 445
 Saha equation 149
 Salpeter mass function 405
 Schmidt telescope 14
 secular acceleration 501
 shell dynamo 187
 signal-to-noise ratio 22
 SIM Planetquest 416, 420, 532
 Sloan Digital Sky Survey (SDSS) 29, 254, 256
- Solar cycle 202
 Solar flares 210
 Solar nebula 135, 495
 Solar System
 formation 495
 models 496
 source function 156
 speckle interferometry 37
 spectral classification
 Bessel M dwarf system 50
 Harvard system 47
 KHM M dwarf system 49
 L dwarf classification 53
 luminosity class 48
 Morgan–Keenan (MK) system 49
 Mount Wilson system 49
 T dwarf classification 61
 spectral resolution 34
 spectrograph 31
 spectroscopic indices 66, 77
 spectrum synthesis 166
 spherical trigonometry 7
 spiral arms 329
 Spitzer conductivity 197
 Spitzer telescope 28, 62, 259, 281
 Stark effect 211
 star formation 102, 134, 447
 star spots 194, 202, 273, 514
 Stefan’s law 24
 stellar flares 206, 271
 stellar populations 290
 stellar pulsations 513
 Strehl ratio 39
 Strömgren photometric system 304
 subdwarfs
 hydrogen-burning limit 121
 identification 75
 metallicity 75
 models 130, 174
 old disk 78
 photometric properties 77
 spectral classification 76
 temperatures 177

- substellar mass function 427, 435, 440, 443, 446
- surveys, infrared
2MASS 28
AFGL 28
DENIS 27
IRAS 28
TMSS 27
- surveys, optical
Digitised Sky Survey (DSS) 26
Palomar (POSS) 25
SDSS 29
UK Schmidt (UKST) 26
- surveys, X-ray
Einstein 30
ROSAT 30
- Sun**
bolometric magnitude 23
luminosity 23
magnetic field 186
Solar cycle 202
Solar flares 210
sunspots 186
temperature 24
sunspots 202
supernovae 309, 326
- T dwarfs**
bolometric corrections 73, 259
discovery 55, 250, 256
luminosities 123, 258, 261
masses 423
photometric properties 72, 258
radii 125, 132
spectral classification 61
subdwarfs 460
temperatures 239, 263, 268
- T Tauri stars 65, 90, 137, 506, 516
- Taurus cloud 90, 137, 316, 329, 445
- telescopes
angular resolution 12
Cassegrain 13
diffraction limit 12, 35
focal ratio 13
- focal length 12
plate scale 12
Ritchey–Chrétien 14
Schmidt 14
X-ray 15–16
- Terrestrial Planet Finder (TPF) 533
- thermal equilibrium 102, 104
- thermodynamic equilibrium 148
- TMSS (Two Micron Sky Survey) 27, 245
- transition object 121, 123, 239
- transition region 80, 194, 196, 222
- turbulent dynamo 188
- TW Hydriæ association 83, 84, 278, 279, 330, 506
- Two Micron All-Sky Survey (2MASS) 28, 254
- Two Micron Sky Survey (TMSS) 27, 245
- UK Schmidt telescope (UKST) 26
- ultraviolet
excess 303
flares 207
observations 86, 198, 226
radiation 3
- UV Ceti (Gl 65B) 80, 85
- Vatican conference (1958) 292, 315, 319
- VB 8 (Gl 644C) 68, 169, 240, 245
- VB 10 (Gl 752B) 51, 80, 157, 169, 226, 245
- Vega 9, 139
- velocity dispersion
and age 217
disk 317
halo 301
Schwarzschild ellipsoid 12, 301, 317, 460
thick disk 317
- vernal equinox 6
- vertex deviation 12, 316
- Voigt profile 151
- V/V_{\max} statistic 362, 470

- Washington photometric system 305
white dwarfs 241, 246, 325, 327, 376
Wien's law 24
Wide-field Infrared Satellite Explorer
(WISE) 281
Wolf 359 (Gl 406) 49, 68, 169, 245
- X-ray telescopes 15, 16
X-rays
observations 84, 201, 207, 221, 271
radiation 3
- surveys 29
telescopes 15
- YY Gem (Gl 278C) 69, 133, 418
- Zebra effect 206
Zeeman effect 189
zenith distance 4
zodiacal dust 506
ZZ Ceti stars 241, 243, 246

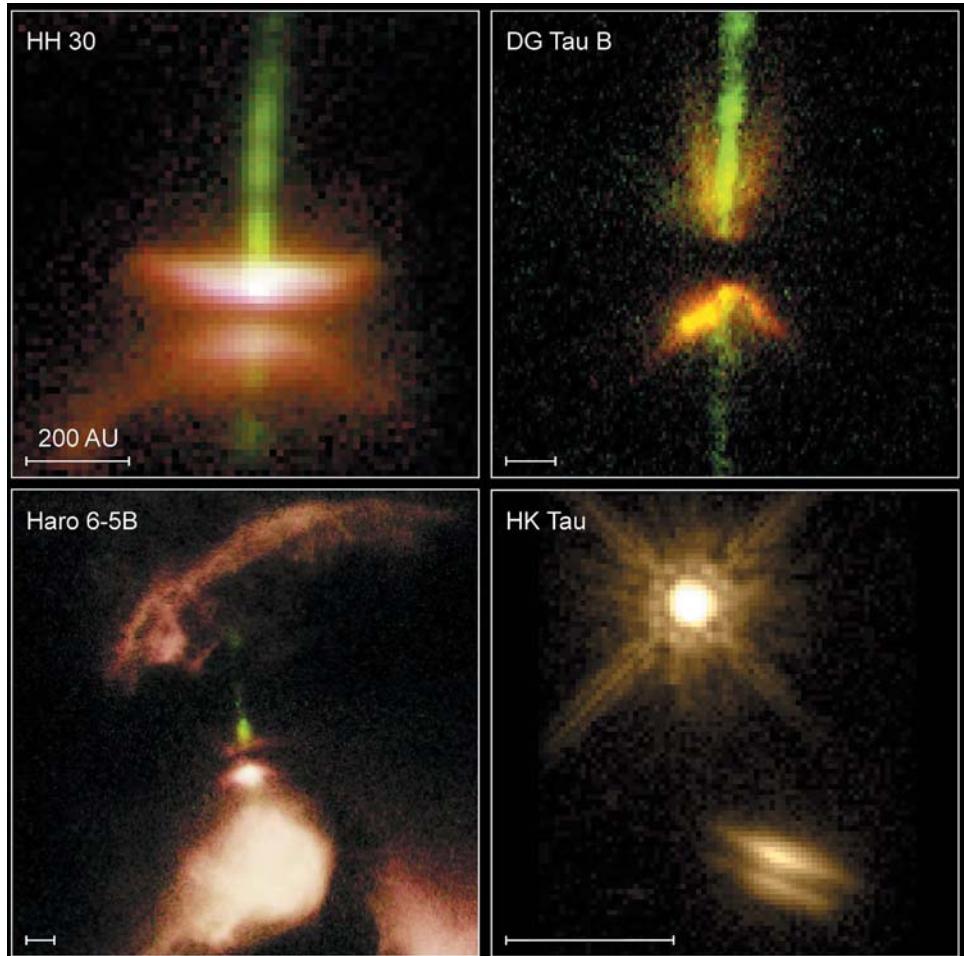


Figure 3.18. Hubble Space Telescope images of gaseous jets from Herbig–Haro objects. The protostellar disk is evident as the dark lane in HH 30.
Courtesy of STScI.

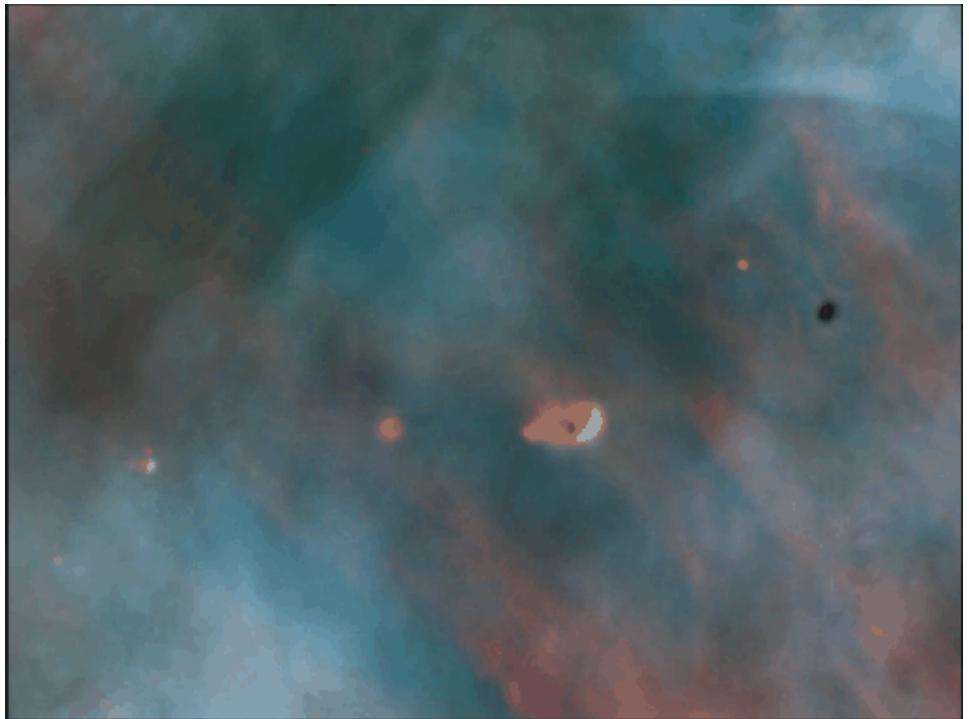


Figure 3.19. A protoplanetary disk, or ‘proplyd’, set against the Orion Nebula.
Courtesy of STScI.

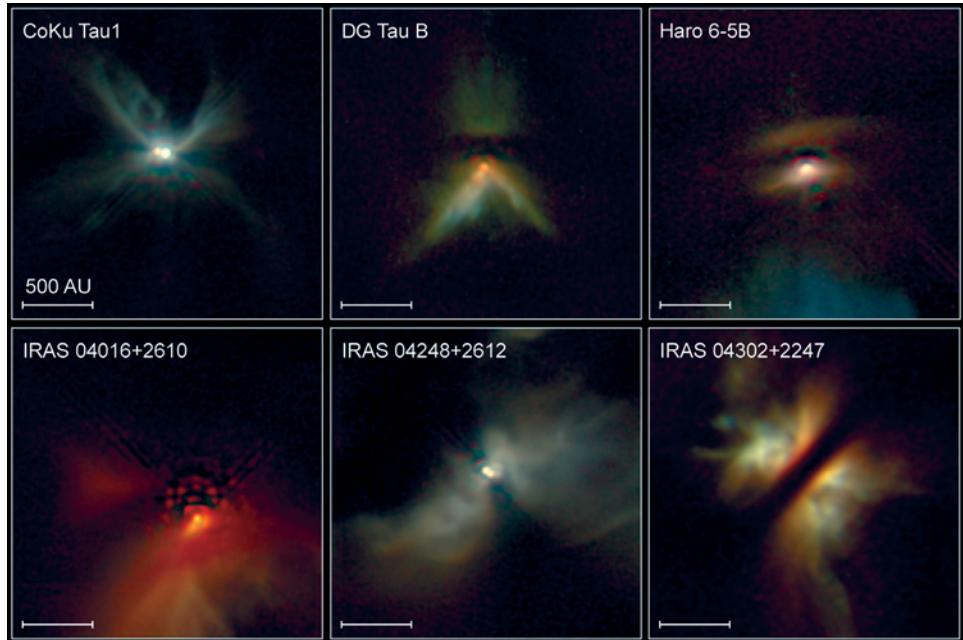


Figure 3.20. Hubble Space Telescope images of disks around young stars.

Courtesy of STScI.

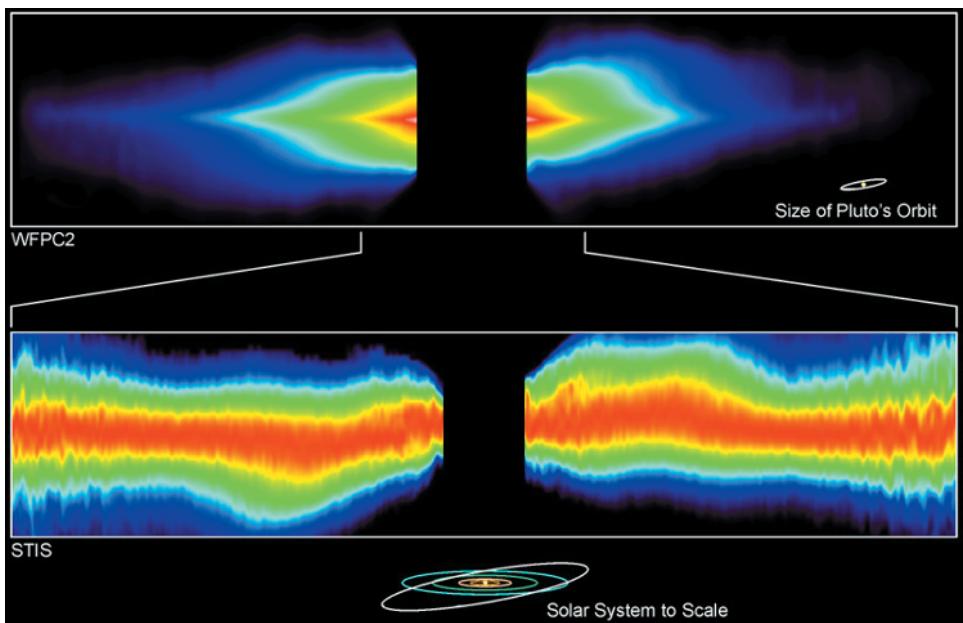


Figure 11.5. The debris disk in β Pic: the lower panel is an expanded representation of the innermost regions of the disk. In both cases, the solar system is shown to provide a reference scale.

Courtesy of J. Krist et al and STScI.

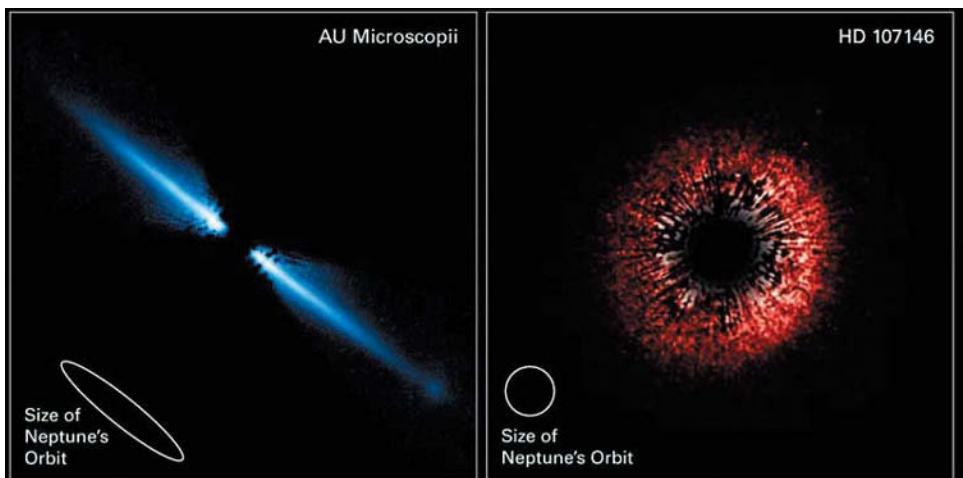


Figure 11.6. Hubble Space Telescope coronagraphic observations of the edge-on disk around AU Mic and the face-on disk around HD 107146.

Courtesy of J. Krist et al and STScI.

Printing: Mercedes-Druck, Berlin
Binding: Stein + Lehmann, Berlin