

High Energy Astrophysics

Third Edition

Providing students with an in-depth account of the astrophysics of high energy phenomena in the Universe, the third edition of this well-established textbook is ideal for advanced undergraduate and beginning graduate courses in high energy astrophysics.

Building on the concepts and techniques taught in standard undergraduate courses, this textbook provides the astronomical and astrophysical background for students to explore more advanced topics. Special emphasis is given to the underlying physical principles of high energy astrophysics, helping students understand the essential physics.

The third edition has been completely rewritten, consolidating the previous editions into one volume. It covers the most recent discoveries in areas such as gamma-ray bursts, ultra-high energy cosmic rays and ultra-high energy gamma rays. The topics have been rearranged and streamlined to make them more applicable to a wide range of different astrophysical problems.

Malcolm S. Longair is Emeritus Jacksonian Professor of Natural Philosophy and Director of Development at the Cavendish Laboratory, University of Cambridge. He has held many senior positions in physics and astronomy, and has served on and chaired many national and international committees, boards and panels, working with both NASA and the European Space Agency. He has received much recognition for his work, including a CBE in the millennium honours list for his services to astronomy and cosmology. He is a Fellow of the Royal Society of London, the Royal Society of Edinburgh, the Academia Lincei and the Istituto Veneto di Scienze, Arte e Literatura.

High Energy Astrophysics

Third Edition

MALCOLM S. LONGAIR

Emeritus Jacksonian Professor of Natural Philosophy,
Cavendish Laboratory,
University of Cambridge, Cambridge



CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521756181

© M. Longair 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN 978-0-521-75618-1 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate.

For Deborah

Contents

<i>Preface</i>	<i>page</i> xiii
<i>Acknowledgements</i>	xvii

Part I Astronomical background

1 High energy astrophysics – an introduction	3
1.1 High energy astrophysics and modern physics and astronomy	3
1.2 The sky in different astronomical wavebands	4
1.3 Optical waveband $3 \times 10^{14} \leq \nu \leq 10^{15}$ Hz; $1 \mu\text{m} \geq \lambda \geq 300$ nm	5
1.4 Infrared waveband $3 \times 10^{12} \leq \nu \leq 3 \times 10^{14}$ Hz; $100 \geq \lambda \geq 1 \mu\text{m}$	9
1.5 Millimetre and submillimetre waveband $30 \text{ GHz} \leq \nu \leq 3 \text{ THz}; 10 \geq \lambda \geq 0.1 \text{ mm}$	14
1.6 Radio waveband $3 \text{ MHz} \leq \nu \leq 30 \text{ GHz}; 100 \text{ m} \geq \lambda \geq 1 \text{ cm}$	17
1.7 Ultraviolet waveband $10^{15} \leq \nu \leq 3 \times 10^{16}$ Hz; $300 \geq \lambda \geq 10$ nm	21
1.8 X-ray waveband $3 \times 10^{16} \leq \nu \leq 3 \times 10^{19}$ Hz; $10 \geq \lambda \geq 0.01$ nm; $0.1 \leq E \leq 100$ keV	22
1.9 γ -ray waveband $\nu \geq 3 \times 10^{19}$ Hz; $\lambda \leq 0.01$ nm; $E \geq 100$ keV	25
1.10 Cosmic ray astrophysics	27
1.11 Other non-electromagnetic astronomies	32
1.12 Concluding remarks	34
2 The stars and stellar evolution	35
2.1 Introduction	35
2.2 Basic observations	35
2.3 Stellar structure	39
2.4 The equations of energy generation and energy transport	43
2.5 The equations of stellar structure	47
2.6 The Sun as a star	50
2.7 Evolution of high and low mass stars	59
2.8 Stellar evolution on the colour–magnitude diagram	68
2.9 Mass loss	70
2.10 Conclusion	75
3 The galaxies	77
3.1 Introduction	77
3.2 The Hubble sequence	78

3.3	The red and blue sequences	80
3.4	Further correlations among the properties of galaxies	86
3.5	The masses of galaxies	89
3.6	The luminosity function of galaxies	95
4	Clusters of galaxies	99
4.1	The morphologies of rich clusters of galaxies	99
4.2	Clusters of galaxies and isothermal gas spheres	102
4.3	The Coma Cluster of galaxies	106
4.4	Mass distribution of hot gas and dark matter in clusters	109
4.5	Cooling flows in clusters of galaxies	110
4.6	The Sunyaev–Zeldovich effect in hot intracluster gas	114
4.7	Gravitational lensing by galaxies and clusters of galaxies	116
4.8	Dark matter in galaxies and clusters of galaxies	123
Part II Physical processes		
5	Ionisation losses	131
5.1	Introduction	131
5.2	Ionisation losses – non-relativistic treatment	131
5.3	The relativistic case	136
5.4	Practical forms of the ionisation loss formulae	141
5.5	Ionisation losses of electrons	145
5.6	Nuclear emulsions, plastics and meteorites	146
5.7	Dynamical friction	151
6	Radiation of accelerated charged particles and bremsstrahlung of electrons	154
6.1	Introduction	154
6.2	The radiation of accelerated charged particles	154
6.3	Bremsstrahlung	163
6.4	Non-relativistic bremsstrahlung energy loss rate	166
6.5	Thermal bremsstrahlung	167
6.6	Relativistic bremsstrahlung	173
7	The dynamics of charged particles in magnetic fields	178
7.1	A uniform static magnetic field	178
7.2	A time-varying magnetic field	180
7.3	The scattering of charged particles by irregularities in the magnetic field	184
7.4	The scattering of high energy particles by Alfvén and hydromagnetic waves	187
7.5	The diffusion-loss equation for high energy particles	189
8	Synchrotron radiation	193
8.1	The total energy loss rate	193
8.2	Non-relativistic gyroradiation and cyclotron radiation	195

8.3	The spectrum of synchrotron radiation – physical arguments	198
8.4	The spectrum of synchrotron radiation – a fuller version	202
8.5	The synchrotron radiation of a power-law distribution of electron energies	212
8.6	The polarisation of synchrotron radiation	214
8.7	Synchrotron self-absorption	217
8.8	Useful numerical results	222
8.9	The radio emission of the Galaxy	224
9	Interactions of high energy photons	228
9.1	Photoelectric absorption	228
9.2	Thomson and Compton scattering	231
9.3	Inverse Compton scattering	237
9.4	Comptonisation	243
9.5	The Sunyaev–Zeldovich effect	257
9.6	Synchrotron–self-Compton radiation	260
9.7	Cherenkov radiation	264
9.8	Electron–positron pair production	270
9.9	Electron–photon cascades, electromagnetic showers and the detection of ultra-high energy γ -rays	272
9.10	Electron–positron annihilation and positron production mechanisms	275
10	Nuclear interactions	279
10.1	Nuclear interactions and high energy astrophysics	279
10.2	Spallation cross-sections	282
10.3	Nuclear emission lines	287
10.4	Cosmic rays in the atmosphere	292
11	Aspects of plasma physics and magnetohydrodynamics	298
11.1	Elementary concepts in plasma physics	298
11.2	Magnetic flux freezing	304
11.3	Shock waves	314
11.4	The Earth's magnetosphere	319
11.5	Magnetic buoyancy	321
11.6	Reconnection of magnetic lines of force	323
Part III High energy astrophysics in our Galaxy		
12	Interstellar gas and magnetic fields	333
12.1	The interstellar medium in the life cycle of stars	333
12.2	Diagnostic tools – neutral interstellar gas	333
12.3	Ionised interstellar gas	340
12.4	Interstellar dust	347
12.5	An overall picture of the interstellar gas	353

12.6	Star formation	361
12.7	The Galactic magnetic field	369
13	Dead stars	378
13.1	Supernovae	378
13.2	White dwarfs, neutron stars and the Chandrasekhar limit	394
13.3	White dwarfs	401
13.4	Neutron stars	401
13.5	The discovery of neutron stars	406
13.6	The galactic population of neutron stars	419
13.7	Thermal emission of neutron stars	421
13.8	Pulsar glitches	422
13.9	The pulsar magnetosphere	424
13.10	The radio and high energy emission of pulsars	427
13.11	Black holes	429
14	Accretion power in astrophysics	443
14.1	Introduction	443
14.2	Accretion—general considerations	443
14.3	Thin accretion discs	451
14.4	Thick discs and advective flows	461
14.5	Accretion in binary systems	464
14.6	Accreting binary systems	473
14.7	Black holes in X-ray binaries	486
14.8	Final thoughts	492
15	Cosmic rays	493
15.1	The energy spectra of cosmic ray protons and nuclei	493
15.2	The abundances of the elements in the cosmic rays	496
15.3	The isotropy and energy density of cosmic rays	502
15.4	Gamma ray observations of the Galaxy	503
15.5	The origin of the light elements in the cosmic rays	507
15.6	The confinement time of cosmic rays in the Galaxy and cosmic ray clocks	515
15.7	The confinement volume for cosmic rays	517
15.8	The Galactic halo	520
15.9	The highest energy cosmic rays and extensive air-showers	522
15.10	Observations of the highest energy cosmic rays	524
15.11	The isotropy of ultra-high energy cosmic rays	529
15.12	The Greisen–Kuzmin–Zatsepin (GKZ) cut-off	531
16	The origin of cosmic rays in our Galaxy	536
16.1	Introduction	536
16.2	Energy loss processes for high energy electrons	536

16.3	Diffusion-loss equation for high energy electrons	540
16.4	Supernova remnants as sources of high energy particles	545
16.5	The minimum energy requirements for synchrotron radiation	549
16.6	Supernova remnants as sources of high energy electrons	553
16.7	The evolution of supernova remnants	554
16.8	The adiabatic loss problem and the acceleration of high energy particles	556
17	The acceleration of high energy particles	561
17.1	General principles of acceleration	561
17.2	The acceleration of particles in solar flares	562
17.3	Fermi acceleration – original version	564
17.4	Diffusive shock acceleration in strong shock waves	568
17.5	Beyond the standard model	574
17.6	The highest energy cosmic rays	580
Part IV Extragalactic high energy astrophysics		
18	Active galaxies	585
18.1	Introduction	585
18.2	Radio galaxies and high energy astrophysics	585
18.3	The quasars	586
18.4	Seyfert galaxies	592
18.5	Blazars, superluminal sources and γ -ray sources	596
18.6	Low Ionisation Nuclear Emission Regions – LINERS	598
18.7	Ultra-Luminous Infrared Galaxies ULIRGs	598
18.8	X-ray surveys of active galaxies	600
18.9	Unification schemes for active galaxies	602
19	Black holes in the nuclei of galaxies	610
19.1	The properties of black holes	610
19.2	Elementary considerations	611
19.3	Dynamical evidence for supermassive black holes in galactic nuclei	613
19.4	The Soltan argument	623
19.5	Black holes and spheroid masses	625
19.6	X-ray observations of fluorescence lines in active galactic nuclei	626
19.7	The growth of black holes in the nuclei of galaxies	633
20	The vicinity of the black hole	637
20.1	The prime ingredients of active galactic nuclei	637
20.2	The continuum spectrum	637
20.3	The emission line regions – the overall picture	640
20.4	The narrow-line regions – the example of Cygnus A	641
20.5	The broad-line regions and reverberation mapping	646

20.6	The alignment effect and shock excitation of emission line regions	653
20.7	Accretion discs about supermassive black holes	656
21	Extragalactic radio sources	661
21.1	Extended radio sources – Fanaroff–Riley types	661
21.2	The astrophysics of FR2 radio sources	666
21.3	The FR1 radio sources	675
21.4	The microquasars	676
21.5	Jet physics	678
22	Compact extragalactic sources and superluminal motions	681
22.1	Compact radio sources	681
22.2	Superluminal motions	683
22.3	Relativistic beaming	686
22.4	The superluminal source population	693
22.5	Synchro-Compton radiation and the inverse Compton catastrophe	697
22.6	γ -ray sources in active galactic nuclei	699
22.7	γ -ray bursts	704
23	Cosmological aspects of high energy astrophysics	714
23.1	The cosmic evolution of galaxies and active galaxies	714
23.2	The essential theoretical tools	715
23.3	The evolution of non-thermal sources with cosmic epoch	720
23.4	The evolution of thermal sources with cosmic epoch	729
23.5	Mid- and far-infrared number counts	737
23.6	Submillimetre number counts	740
23.7	The global star-formation rate	743
23.8	The old red galaxies	746
23.9	Putting it all together	749
	Appendix Astronomical conventions and nomenclature	753
A.1	Galactic coordinates and projections of the celestial sphere onto a plane	753
A.2	Distances in astronomy	755
A.3	Masses in astronomy	759
A.4	Flux densities, luminosities, magnitudes and colours	760
A.5	Diffraction-limited telescopes	764
A.6	Interferometry and synthesis imaging	771
A.7	The sensitivities of astronomical detectors	774
A.8	Units and relativistic notation	779
	<i>Bibliography</i>	783
	<i>Name index</i>	825
	<i>Object index</i>	829
	<i>Index</i>	831

Preface

Ancient history

It was a challenge to write this third edition of *High Energy Astrophysics*. Writing the first edition was great fun and that rather slim volume reflected rather closely the lecturing style I adopted in presenting high energy astrophysics to final-year undergraduates in the period 1973–7. Although the material was updated when the manuscript was sent to the press in 1980, the book remained in essence a lecture course (Longair, 1981). The reception of the book was encouraging and in due course a second edition was needed. The subject had advanced so rapidly during the 1980s and early 1990s that the material could not be comfortably contained within one volume. The aim was originally to complete the task in two volumes, but by the time the Volumes 1 and 2 were completed, I had only reached the edge of our own Galaxy (Longair, 1997b,c).¹ Volume 3 was begun, but for various reasons, was not completed – the whole project was becoming somewhat unwieldy.

In the meantime, I completed three other major book-writing projects. The first of these was a new edition of *Theoretical Concepts in Physics* (Longair, 2003). Then, I completed *The Cosmic Century: A History of Astrophysics and Cosmology* (Longair, 2006). Finally, in 2008, the new edition of *Galaxy Formation* was published (Longair, 2008).

The new edition

Since the second edition of *High Energy Astrophysics*, many of the subject areas have changed out of all recognition and new areas of astrophysical research have been opened up, for example, ultra-high energy gamma-ray astronomy. The publication of *Theoretical Concepts in Physics*, *The Cosmic Century* and *Galaxy Formation* have made it feasible to condense the original plan of a three volume work into a single volume. In reorganising the material, some hard decisions had to be taken, but the convenience of including everything in one volume is worth the sacrifice of some of the material from the second edition. The principal decisions were as follows:

¹ The original volumes of the second edition were first published in 1992 (Volume 1) and 1994 (Volume 2). Major revisions and corrections were included in the 1997 reprints of both volumes. I regard the 1997 reissues as the definitive versions of the second edition.

- Much of the relevant historical material has been included in *The Cosmic Century* and so that material will not be repeated here. I make references to the appropriate sections of *The Cosmic Century* and other historical texts. I do this with considerable reluctance since the historical development of high energy astrophysics has influenced strongly the way in which the astrophysics has developed intellectually. History will not disappear completely, but it will not be as prominent as in the earlier editions.
- Much of the necessary material needed to obtain a modern view of galaxies and the large scale structure of the Universe is included in *Galaxy Formation*. In particular, there is no need to repeat much of the detailed discussion of galaxies and clusters, or the large scale structure and dynamics of the Universe. These topics are, however, central to many of the topics in this book and so summaries of the most important topics needed to understand the astronomical context of high energy astrophysics are provided in Part I.
- There was a strong emphasis upon the origin of cosmic rays in the first two editions. I still consider this to be excellent material, particularly in the area of ultra-high energy cosmic rays, but it has been somewhat abbreviated in the new edition.
- There was also a considerable amount of material on detectors and telescopes in the earlier edition. I believe this material is of the greatest interest and importance in understanding our ability of make observations in different wavebands. This aspect of the subject has been strongly moderated in the new edition. These are fascinating topics, but modern telescopes and detectors have become increasingly complex and sophisticated. Summaries of a number of important topics in the physics of astronomical detectors and telescopes are included as an appendix.
- In the second edition, I devoted some space to high energy astrophysics in the Solar System. This material has been abbreviated, but important topics such as the diffusion of energetic charged particles in the Solar Wind and the acceleration of charged particles in solar flares have been preserved.
- The opportunity has been taken to rationalise the presentation of the physical and astrophysical processes so that duplication of material is avoided.
- The writing has been very considerably tightened up so that the discussion is less discursive than in the earlier editions. Again, I regret the necessity of doing this since often these asides provide valuable physical insights for reader new to the subject.

The aims of the present edition are the same as the earlier editions. A very wide range of physical processes relevant for high energy astrophysics is discussed, the emphasis being strongly upon the understanding of the underlying physics. I aim to maintain the informal style of the earlier editions and have no hesitation about using the first person singular or expressing my personal opinion about the material under discussion. The emphasis is strongly upon physical principles and the discussion of general results rather than particular models which may have only ephemeral appeal.

As I learned during the writing of *The Cosmic Century*, physics and astrophysics have a symbiotic relation. On the one hand, the astrophysical sciences are concerned with the application of the laws of physics to phenomena on a large scale in the Universe. On the other hand, new laws of physics are discovered and tested through astronomical observations and their astrophysical interpretation. In these ways, the new astrophysics, of which

high energy astrophysics is one of the most important ingredients, is just as much a part of modern physics as laboratory physics.

Although there is limited scope for deviation from the central theme in this new edition, one of my original aims was to give the reader a feeling of what it is like to undertake research at the limits of present understanding. Astrophysics is fortunate in that many of the fundamental problems can be understood without a great deal of new physics or new physical concepts. Thus, the text may also be considered as an introduction to the way in which research is carried out in the astrophysical context.

Above all, however, this material is not only mind-stretching, but also great fun. I have no intention of inhibiting my enthusiasm and enormous enjoyment of the physics and astrophysics for its own sake.

Malcolm Longair
Cambridge and Venice
January 2010

Acknowledgements

There are many people whom it is a pleasure to thank for help and advice during the preparation of this volume. Just as the first edition was begun during a visit to the Osservatorio Astronomico di Arcetri in Florence in April 1980, so the second edition could not have been completed without the Regents' Fellowship of the Smithsonian Institution which I held at the Harvard-Smithsonian Astrophysical Observatory during the period April–June 1990. I am particularly grateful to Professors Irwin Shapiro and Giovanni Fazio for sponsoring this visit to Harvard during which time the final drafts of Chapters 1–10 of the first volume of the second edition were completed. During that period, I had particularly helpful discussions with Drs Eugene Avrett, George Rybicki, Giovanni Fazio, Margaret Geller and many others. I am particularly grateful to them for their advice.

Much of the preliminary rewriting was completed while I was at the Royal Observatory, Edinburgh. Among the many colleagues with whom I discussed the contents of this volume, I must single out Dr John Peacock who provided deep insights into many topics. In completing the final chapter on the high energy astrophysics of the Solar System, I greatly benefitted from the advice of Professors John Brown, Carole Jordan and Eric Priest. Not only did they point me in the correct directions but they also reviewed my first drafts of that chapter. I am especially grateful to them for this laborious task. Many colleagues made helpful suggestions about corrections and additions to the first edition, among whom Dr Roger Chevalier provided an especially useful list.

Coincidentally, the writing of the third edition began while I was a visitor at the Osservatorio Astronomico di Arcetri in Florence during the period April–June 2007. I thank Professor Francesco Palla and his colleagues for their hospitality during that visit. The catalogue of friends and colleagues who have continued to contribute to my understanding of high energy astrophysics and astrophysical cosmology since the publication of the second edition is enormous. Many of them are acknowledged in my recent books, but the list is so long that I would be bound to miss someone out. I acknowledge particular insights from my colleagues in the course of the book. Special thanks are due to Dr. David Green for his expert advice, not only on supernova remnants, but also on the more arcane idiosyncrasies of LaTeX.

To all of these friends and colleagues I make the usual disclaimer that any misrepresentation of the material presented in this book is entirely my responsibility and not theirs. Finally, I acknowledge the unfailing support and love of my family, Deborah, Mark and Sarah who have contributed much more than they will ever know to the completion of this book.



PART I

ASTRONOMICAL BACKGROUND

1.1 High energy astrophysics and modern physics and astronomy

The revolution in astronomy, astrophysics and cosmology since the end of the Second World War in 1945 has been driven by the opening up of the whole of the electromagnetic spectrum for astronomical observations. This revolution would not have been possible without the development of new techniques and technologies for making astronomical observation from the ground and from space. Hand in hand with these developments have been major advances in laboratory physics and the development of high speed computers. It is the combination of all these factors which has led to dramatic advances in the astrophysical and cosmological sciences.

Among the most important of the new disciplines is *high energy astrophysics*. I take this term to mean the astrophysics of high energy processes and their application in astrophysical and cosmological contexts. These processes, their application in astrophysics and how they lead to some of the most challenging problems of contemporary physics, are the subjects of this book. For example, we need to explain how the massive black holes present in the nuclei of active galaxies can be studied, how charged particles are accelerated to extremely high energies in astronomical environments, the origins of enormous fluxes of high energy particles and magnetic fields in active galaxies, the physical processes in the interiors and environments of neutron stars, the nature of the dark matter, the expected fluxes of gravitational waves in extreme astronomical environments, and so on. Thus, high energy astrophysics makes feasible the study of the properties of matter under physical conditions which cannot yet be reproduced in the laboratory. Indeed, in many cases, the problems can only be addressed in the astrophysical environment. The aim of this book is to set out the logical sequence of steps by which astrophysicists tackle these problems.

The aim of the astrophysical sciences is two-fold – the application of the laws of physics in the extreme physical conditions encountered in astronomical systems, and the discovery of new laws of physics from observation. This second aspect has a long and distinguished pedigree, as I have recounted in my book *The Cosmic Century* (Longair, 2006). We will encounter many new and exciting examples in the course of this exposition. Throughout the text, the emphasis will be upon those aspects of high energy astrophysics in which the astrophysical understanding is reasonably secure, and indicative of those areas where the astrophysics is still poorly understood.

The amount of material to be covered is enormous and so, to put some order into the presentation, the book is divided into four parts.

Part I The first part concerns the essential astronomical background needed to understand the context within which high energy astrophysical studies are carried out. If you already have a good grounding in astronomy and astrophysics, you may pass on to the subsequent parts. The first chapter introduces all the accessible astronomical wavebands and outlines the distinctive features of the astrophysical objects observed. There then follow chapters which summarise the essential features of stellar evolution, galaxies and clusters of galaxies, in order to understand the contexts within which high energy astrophysical phenomena are observed. Even studies such as the properties of galaxies have undergone a significant change of emphasis in the light of the evidence provided by very large surveys of galaxies, such as the Anglo-Australian Telescope 2dF Galaxy Survey and the Sloan Digital Sky Survey.

Part II Chapters 5–11 are principally concerned with the physical processes involved in the interactions and radiation of charged particles. The emphasis is upon a clear description of the physics of these processes. Generally, the simplest physical approach to understanding the processes is given first and then some of the more important of these are studied in more detail. Processes which dominate much of high energy astrophysics, such as bremsstrahlung, synchrotron radiation and inverse Compton scattering, merit such a more detailed treatment.

Part III Chapters 12–17 are principally concerned with high energy astrophysical processes in our Galaxy. A large suite of exotic objects is introduced, including white dwarfs, neutron stars, black holes and supernova explosions. The study of the origin of cosmic ray particles fits naturally into this discussion since these are the only samples of high energy particles originating in extreme astronomical environments which we can study directly within the Solar System. The acceleration of charged particles to high energies in Galactic environments provides clues to the much more extreme events which must take place in active galaxies.

Part IV Chapters 18–23 are devoted to extragalactic high energy astrophysics and involve some of the most extreme energetic phenomena in the Universe – the quasars, radio galaxies, TeV γ -ray sources, γ -ray bursts, and so on. The most extreme objects must involve physical processes originating close to supermassive black holes and what we observe is strongly influenced by relativistic aberration effects. In Chapter 23, some cosmological aspects of high energy astrophysics and the role that supermassive black holes may play in galaxy formation are described.

This is a very large programme and readers are encouraged to be selective in their use of the material and to customise it to their own requirements.

1.2 The sky in different astronomical wavebands

The dramatic change in perspective of astrophysical research over the last half century is conveniently illustrated by images of the celestial sphere in the different astronomical

wavebands now accessible to observation. These can be thought of as providing different temperature maps of the Universe according to Wien's displacement law,

$$\nu_{\max} = 10^{11}(T/\text{K}) \text{ Hz}; \quad \lambda_{\max} T = 3 \times 10^6 \text{ nm K}, \quad (1.1)$$

where the relations refer to the maximum intensity of a black-body, or Planck, spectrum, expressed either in frequency or wavelength units, of a body in thermodynamic equilibrium at temperature T . These relations are shown in Fig. 1.1a which includes the conventional labels of the different astronomical wavebands. In the optical waveband, for example, the typical temperatures of thermal sources of radiation are about 3000–10 000 K. Thermal sources in the X-ray waveband typically have temperatures of at least 10^7 – 10^8 K, while far-infrared observations provide images of the cold Universe, typical temperatures being about 30–100 K. Objects with a wider range of temperatures are observable in any given waveband because of the broad-band nature of the thermal radiation spectrum. Thermodynamically speaking, the above figures are only lower limits to the temperatures of sources which are observable in these wavebands. In the case of non-thermal sources of radiation, by which we mean radiation emitted by sources which do not possess a Maxwellian energy distribution of radiating particles, the effective temperature of the emitting particles can far exceed the above temperatures. This is particularly important for non-thermal sources such as Galactic and extragalactic radio sources, quasars and X- and γ -ray sources in which the continuum radiation is associated with the emission of ultra-relativistic electrons.

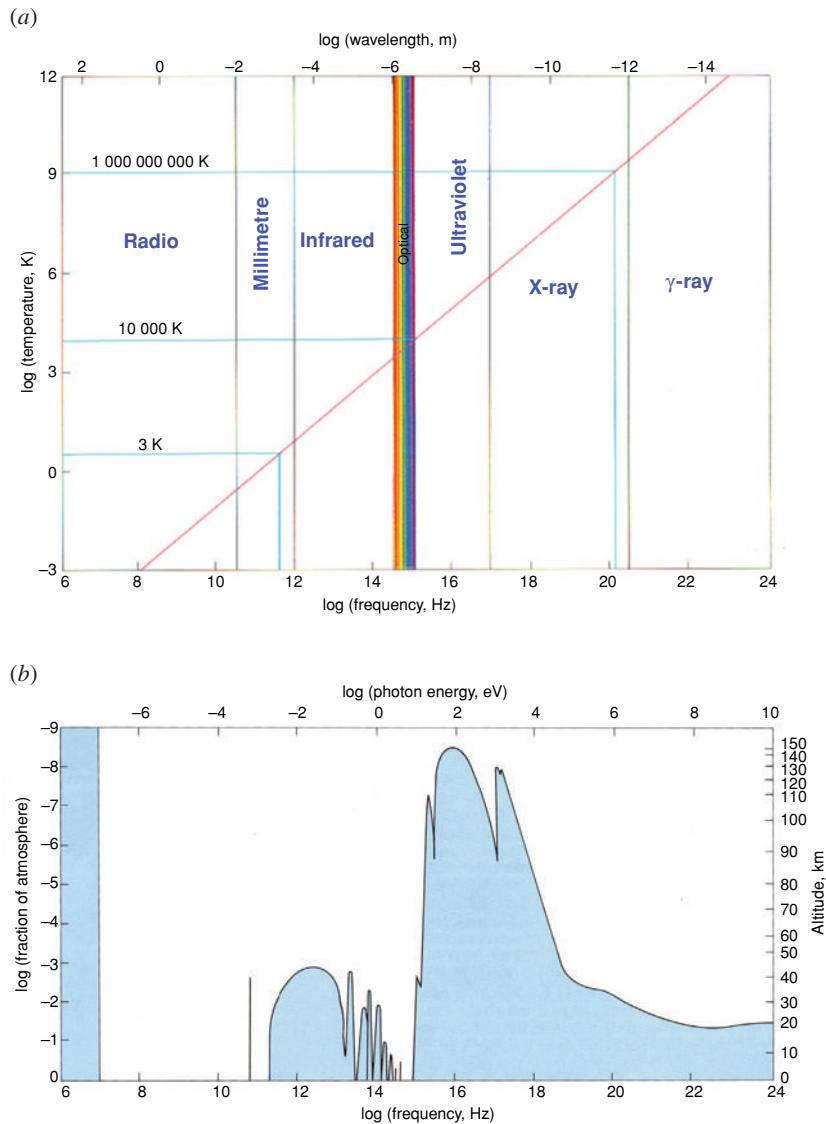
Astronomical observations can be made from ground-based observatories in the optical, near-infrared, millimetre and radio wavebands. Once space was opened up for astronomical observations in the late 1950s, it became possible to observe the sky in the mid- and far-infrared, ultraviolet and X- and γ -ray wavebands. The observability of the sky in different astronomical wavebands is illustrated in Fig. 1.1b, which shows the transparency of the atmosphere as a function of wavelength. In this representation, the solid line indicates how high a telescope has to be located above the surface of the Earth for the atmosphere to become transparent to radiation of different wavelengths. Let us first summarise the observational challenges and the nature of the objects which dominate all-sky images of these wavebands.¹

1.3 Optical waveband $3 \times 10^{14} \leq \nu \leq 10^{15}$ Hz;
 $1 \mu\text{m} \geq \lambda \geq 300 \text{ nm}$

1.3.1 Observing in the optical waveband

Until 1945, astronomy meant optical astronomy and Fig. 1.1a shows that this corresponds to studying the Universe in the rather narrow wavelength interval 300–800 nm, and hence to black-body temperatures in the range 3000–10 000 K. The wavelength range to which

¹ Many more details of the history of the different types of astronomy discussed in the succeeding sections of this chapter are included in my book *The Cosmic Century: A History of Astrophysics and Cosmology* (Longair, 2006).

**Fig. 1.1**

(a) The relation between the temperature of a black-body and the frequency ν (or wavelength λ) at which most of the energy is emitted (solid red line). The frequency (or wavelength) plotted is that corresponding to the maximum of a black-body at temperature T . Convenient expressions for this relation are: $\nu_{\max} = 10^{11}(T/\text{K}) \text{ Hz}$; $\lambda_{\max}T = 3 \times 10^6 \text{ nm K}$. The ranges of wavelength corresponding to the different wavebands – radio, millimetre, infrared, optical, ultraviolet, X- and γ -ray – are shown. (b) The transparency of the atmosphere for radiation of different wavelengths. The solid line shows the height above sea-level at which the atmosphere becomes transparent for radiation of different wavelengths (Giacconi *et al.*, 1968; Longair, 1988).

our eyes are sensitive is roughly 400–700 nm, corresponding to the blue and red ends of the optical spectrum, respectively. At the short wavelength end of the waveband, the atmosphere becomes opaque because of absorption by ozone in the upper atmosphere. The absorption sets in rather abruptly with decreasing wavelength so that observations from ground-based observatories at wavelengths less than about 320 nm are generally impossible. This has the beneficial effect of protecting us from the Sun's hard ultraviolet radiation. Most people derive their intuitive picture of the Universe from observations in the optical waveband.

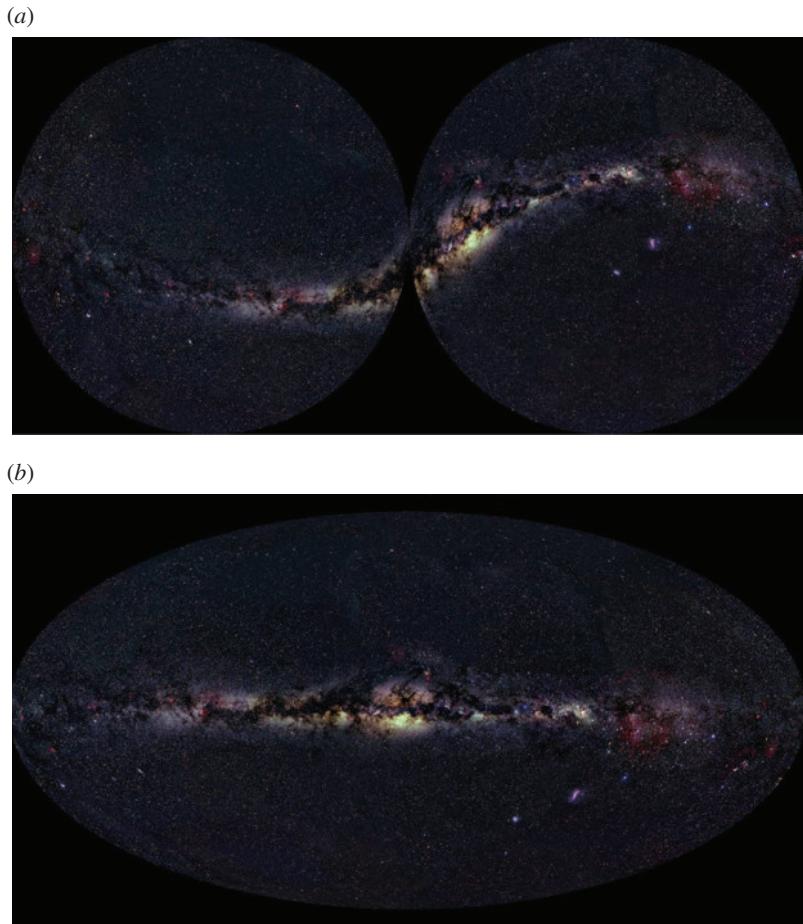
For most types of observation, photographic plates have now been replaced by electronic detectors such as charge-coupled devices (CCD) which have quantum efficiencies of about 80% at the red end of the optical spectrum (500–1000 nm). The band-gap of silicon corresponds to a limiting maximum wavelength of about 1 μm and so optical CCDs are more or less limited to the classical optical waveband. Nowadays, it is routine to observe with CCD arrays of, say, 2000×2000 picture elements (pixels) and greater. Mosaics of CCD arrays can be used to provide coverage of large areas of sky, as has been achieved in the Sloan Digital Sky Survey. The result has been a huge increase in the quantity and quality of the data which can be analysed astrophysically.

When it was commissioned in the late 1940s, the Palomar 200-inch telescope was an outstanding feat of optical-mechanical engineering and it dominated much of astrophysical and cosmological research for the subsequent 30 years. Five metres was regarded as the maximum feasible aperture because the telescope had to have sufficient stiffness to track and guide accurately over the entire celestial hemisphere. By the 1980s, it was realised that the route to larger aperture was to use the increasing power of computers to build lighter telescopes and then to restore the stiffness electronically by multiply-embedded computer control systems. In so doing, much improved performance has been achieved for telescopes in the 8–10 metre class. The incorporation of *adaptive optics* into the optical train of these telescopes has meant that they can now operate close to the diffraction limit. There are now plans for even larger telescopes, the challenge being to build them at affordable cost.

1.3.2 Optical all-sky images

Images of the northern and southern celestial hemispheres are shown in Fig. 1.2a. These are plotted in *equidistant azimuthal polar* or *zenith equidistant projections* and were reconstructed by Mellinger from 51 wide-angle photographs (Mellinger, 2007). The image of the northern celestial hemisphere on the left has the north celestial pole at declination $\delta = 90^\circ$ in the centre, while the celestial equator, $\delta = 0^\circ$, is the bounding circle around the edge of the picture.² Close inspection of the image shows a number of clearly recognisable constellations, for example, the Plough or Great Bear pointing towards the North Pole star, which is close to the centre of the image. The right-hand image shows the southern celestial hemisphere, centred on the southern celestial pole at $\delta = -90^\circ$. Because two images have been used to span the whole sky, the distortions are not too great, as shown in Fig. A.3a of Appendix A.1. In both diagrams, the Milky Way is clearly seen as a broad band of emission spanning both hemispheres. The Galactic Centre region lies in the southern

² For details of the coordinate systems and projections used in astronomy, see Appendix A.1.

**Fig. 1.2**

All-sky images of the celestial sphere in the optical waveband created by Dr. Axel Mellinger from 51 wide-angle images. The photographs were taken at observing sites in California, South Africa and Germany, image processed and joined together digitally. How these images were created is explained in his web site at <http://home.arcor-online.de/axel.mellinger/>. (a) The northern (left) and southern (right) celestial hemispheres are plotted in equidistant azimuthal polar or zenith equidistant projections. The Milky Way is the broad band of emission seen in both images and is much more prominent in the southern than in the northern skies. (b) The optical image of the whole sky in Galactic coordinates in a Hammer–Aitoff projection. The nearby dwarf companion galaxies to our own Galaxy, the Large and Small Magellanic Clouds, are seen in the southern Galactic hemisphere at about Galactic longitudes 290° and 310° , respectively. (Courtesy of Dr. Axel Mellinger.)

celestial hemisphere at $\delta \approx -29^\circ$ and much more of the Galactic plane can be observed from that hemisphere as compared with the northern hemisphere. The two bright galaxies close to the centre of the image of the Southern Galactic Hemisphere are the Large and Small Magellanic Clouds, our nearest neighbouring galaxies.

A *Hammer–Aitoff projection* of Mellinger's observations enables the complete 4π steradians of the celestial sphere to be projected onto a two-dimensional flat surface (Fig. 1.2b). This projection adopts a reasonable compromise between shape and scale distortions, the

magnitude of these being indicated in Fig. A.3b of Appendix A.1. Although equal areas are preserved, the geometric distortions become large towards the northern and southern Galactic poles. The north and south Galactic poles ($b = \pm 90^\circ$) are at the top and bottom of the image. The scale of Galactic longitude runs from 0° at the centre which is the direction of the Galactic Centre, through $+180^\circ$ at the left of the image, the anti-Centre direction, and then from $+180^\circ$ at the right of the image to 360° (or 0°) at the Centre. The Hammer–Aitoff projection is commonly used in the astronomical literature to display images of the whole sky, and the all-sky images in other astronomical wavebands discussed later in this chapter are presented in this projection.

The light seen in Fig. 1.2 is almost entirely the integrated light of stars. Some of the light of the Galaxy is due to hot diffuse gas, particularly the ionised gas observed in the vicinity of regions of star formation. One of the disadvantages of observing in the optical waveband is immediately apparent from Fig. 1.2. There are patchy dark features present in the image of the Milky Way and these are associated with *extinction* by interstellar dust grains. Tiny dust particles, typically about $1\text{ }\mu\text{m}$ in diameter, strongly scatter and absorb light rays, resulting in the patchy obscuration seen in Fig. 1.2b. Dust extinction complicates the interpretation of optical observations and corrections need to be made for it.

Optical observations are fundamental for astronomy because a significant fraction of the baryonic matter in the Universe is locked up in stars with masses within a factor of about 10 of that of the Sun and these emit a large fraction of their energy in the optical waveband. Since they have long lifetimes, they are the most readily observable objects in the Universe. The stars are assembled into galaxies and these are the basic building blocks of the Universe.

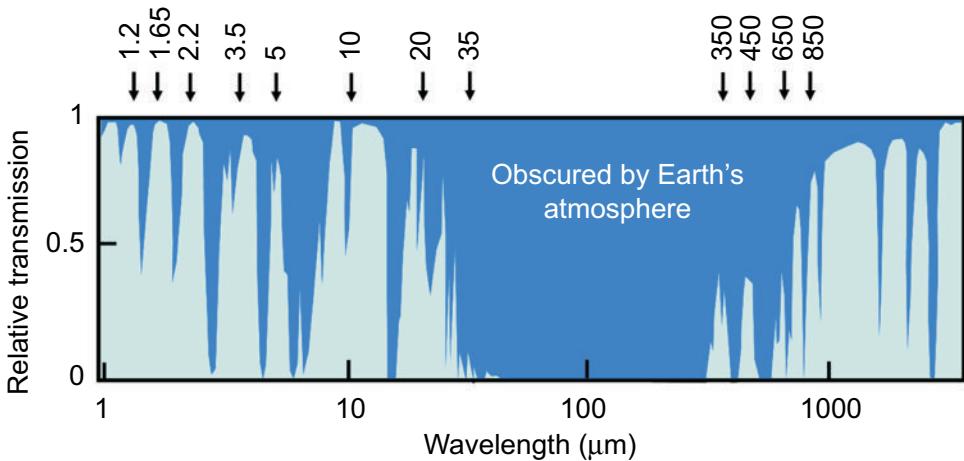
Many different types of high energy astrophysical object are present in our Galaxy, including supernovae, supernova remnants, white dwarfs, neutron stars, stellar-mass black holes and the supermassive black hole in the Galactic Centre. These are, however, often difficult to observe in the optical waveband, partly because they are intrinsically rather faint optically and also because of interstellar extinction. Optical observations are, however, crucial in identifying the sources of the radiation and understanding their roles in stellar evolution. A number of these compact stars are members of binary systems and the companion star can often be identified optically. This is of great importance in determining their distances and masses.³

$$\begin{aligned} \mathbf{1.4\ Infrared\ waveband\ } &3 \times 10^{12} \leq \nu \leq 3 \times 10^{14}\ \text{Hz}; \\ &100 \geq \lambda \geq 1\ \mu\text{m} \end{aligned}$$

1.4.1 Observing in the infrared waveband

The problem of dust extinction is a strong function of wavelength, the extinction coefficient α being roughly proportional to λ^{-1} , where α is defined by $I = I_0 e^{-\alpha r}$ and r is the distance

³ More details of methods of determining distances and masses are given in Appendices A.2 and A.3.

**Fig. 1.3**

The transmission of the atmosphere as a function of wavelength in the infrared ($1 \leq \lambda \leq 100 \mu\text{m}$) and submillimetre ($100 \leq \lambda \leq 1000 \mu\text{m}$) wavebands. The central wavelengths of the observable windows in these wavebands in microns are indicated by the numbers at the top of the diagram. The precipitable water vapour content of the atmosphere is assumed to be 1 mm. (After diagram, courtesy of the Royal Observatory, Edinburgh.)

of the source. Thus, the effects of extinction become rapidly much less important in the infrared as compared with the optical waveband. Infrared radiation suffers, however, from molecular absorption and scattering in the Earth's atmosphere, what is often referred to as *telluric absorption*, so that the sky can only be observed in certain wavelength windows. Figure 1.3 shows the transmission of the atmosphere in the waveband interval $1 \leq \lambda \leq 1000 \mu\text{m}$. The centres of the infrared windows in the wavelength range $1 \leq \lambda \leq 100 \mu\text{m}$ are at wavelengths of 1.2, 1.65, 2.2, 3.5, 5, 10, 20 and 35 μm and they are conventionally labelled the J, H, K, L, M, N, Q and Z infrared wavebands, respectively. The last two windows are only accessible from very high, dry sites and even observations at 10 μm are often difficult, except under the best observing conditions. Observations outside these windows have to be undertaken from balloons, high-flying aircraft or satellite observatories. There is thus a complementarity between the types of observation attempted from the ground and from above the Earth's atmosphere.

A distinctive problem to be overcome in infrared astronomy is that the telescope and the Earth's atmosphere are strong thermal emitters of infrared radiation. For example, the radiation of a black-body at room temperature, say 300 K, peaks at a wavelength of about 10 μm . Therefore, normally, the strength of the signal from an astronomical source is very much weaker than the background due to the telescope and the atmosphere at these wavelengths. For this reason, telescopes dedicated to thermal infrared observations, such as IRAS and the Spitzer Space Telescope, incorporate cooling of the telescope and the focal plane instrumentation to minimise the thermal background.

The infrared waveband is conveniently divided into *near* and *thermal* infrared wavelengths. The distinction is related to those parts of the waveband at which the observations are detector-noise limited (the near-infrared) and those in which the thermal background

radiation from the sky and the telescope are the dominant source of noise (the thermal infrared). The distinction thus depends upon the type of observation being undertaken. Broad-band observations at wavelengths longer than $3\text{ }\mu\text{m}$ are thermal background limited, whereas those at shorter wavelengths are normally detector-noise limited. In making observations in the thermal infrared waveband, the observer is almost always searching for very faint signals against an enormous thermal background. Detector technology for infrared wavelengths has made enormous strides over the last 20 years. Infrared detector arrays almost as large as the CCD detector arrays used in optical astronomy are now available and these have revolutionised essentially all areas of astronomy.

The observing strategy is therefore to observe the sky in those wavelength windows in which there is good atmospheric transparency from ground-based telescopes. This has the advantage that the observations can be made with 8–10 metre aperture telescopes and complex instrumentation. The wavebands which are inaccessible from the ground have to be observed from above the Earth’s atmosphere, preferably from satellite observatories. Necessarily, these are generally smaller than the ground-based telescopes and massive instrumentation cannot be accommodated. The Spitzer Infrared Space Telescope is a splendid example of the state-of-the-art in infrared space technology. In due course it will be superseded by the James Webb Space Telescope, which will be a 6.5 metre infrared-optimised space telescope.

1.4.2 Infrared all-sky images

Images of the whole sky in the near-infrared waveband have much reduced interstellar extinction by interstellar dust grains and the structure of our Galaxy can be clearly seen. Figures. 1.4*a* and *b* provide excellent examples of the structure of the Galaxy as observed in the 1.2–2.2 μm wavebands. Figure 1.4*a* is an all-sky image obtained by the DIRBE instrument of the Cosmic Background Explorer (COBE). This instrument scanned the sky in the J, H and K wavebands and these were combined to create the colour image seen in Fig. 1.4*a*. The disc and bulge of the Galaxy are clearly seen, as well as a thin dust absorption layer lying in the Galactic plane.

Figure 1.4*b* shows another approach to mapping the Galaxy using observations from the ground-based Two Micron All-Sky Survey (2MASS). This survey was carried out using two 1.3 metre dedicated infrared telescopes, one located at Mount Hopkins in Arizona and the other at the Cerro Tololo InterAmerican Observatory in Chile. Almost 300 million stars were catalogued. The image shown in Fig. 1.4*b* was created by plotting the positions of almost 100 million stars brighter than $K = 13.5$ from the 2MASS catalogue. This approach provides an even clearer image of the stellar distribution in the Galaxy. The Large and Small Magellanic Clouds are clearly visible in the southern Galactic hemisphere, as is the elongated central bulge of the Galaxy which has been interpreted as a bar in the central regions of the Galaxy. These images make the important point that, in the infrared waveband, interstellar dust becomes transparent and so it is possible to observe deep inside regions which are obscured at optical wavelengths. Among the most important of these are regions of star formation which are enshrouded in interstellar dust, and the very central regions of our own Galaxy. Observations of infrared stars very close to the Galactic Centre have provided wholly convincing evidence for a supermassive black hole with mass $M \approx 2.6 \times 10^6 M_\odot$.

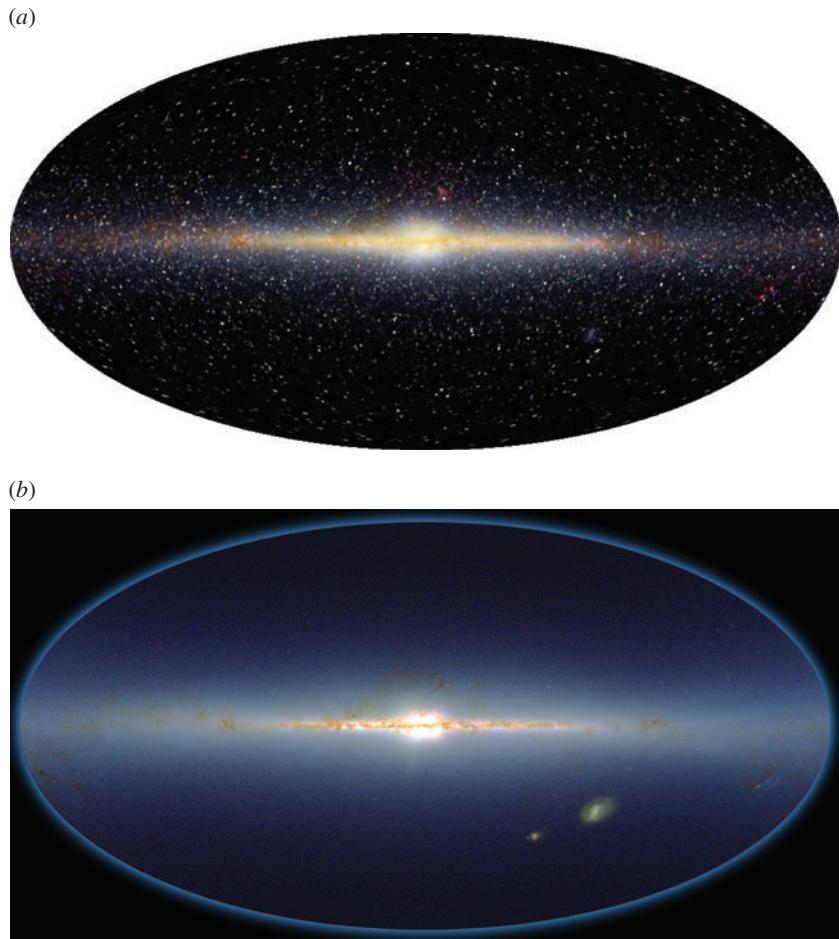
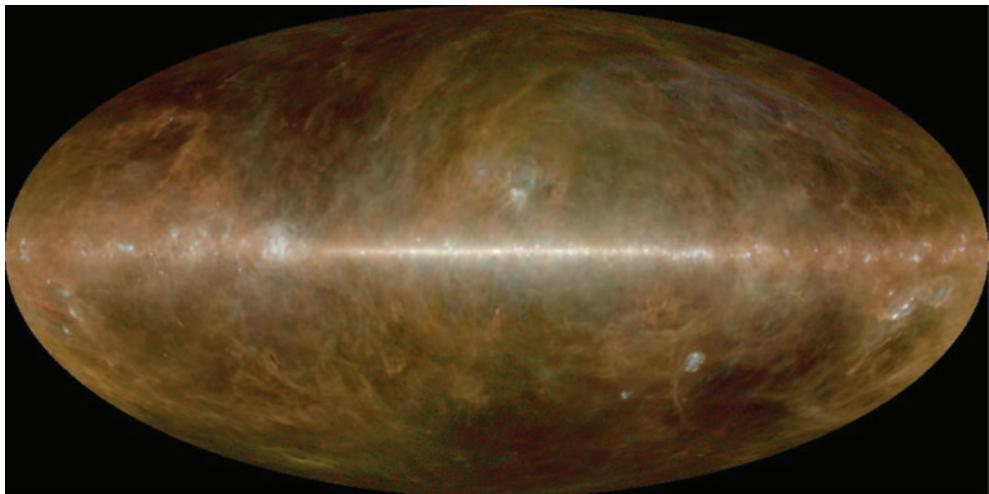


Fig. 1.4 Images of the celestial sphere in the near-infrared waveband. (a) A false-colour image of the near-infrared sky as observed by the DIRBE instrument of the Cosmic Background Explorer (COBE). Data at 1.25, 2.2 and 3.5 μm are colour-coded blue, green and red, respectively, in a Hammer–Aitoff projection. (Courtesy of NASA and the COBE Science Team.) (b) The structure of the Galaxy determined by the distribution of almost 100 million stars detected in the 2MASS sky survey. (Courtesy of the 2MASS Science Team and IPAC.)

Inspection of Fig. 1.2a shows that the typical temperatures of the objects which radiate in the $1\text{--}100\ \mu\text{m}$ waveband are $1000 > T > 10\ \text{K}$ and so Fig. 1.4 provides images of the cold Universe. Thus, cool stars, cool red giant envelopes and cold objects such as brown dwarfs can be observed directly in these wavebands. One of the most distinctive features of these wavebands is, however, the fact that, at wavelengths longer than about $3\ \mu\text{m}$, dust grains become strong emitters rather than absorbers of radiation. They emit more or less like little black-bodies at the temperature to which they are heated by the radiation they absorb. They do not radiate at shorter wavelengths because, if the grains were heated to temperatures greater than about $1000\ \text{K}$, they would evaporate.

**Fig. 1.5**

A composite image of the celestial sphere in the far-infrared waveband in a Hammer–Aitoff projection. The observations were made with the DIRBE instrument of the COBE satellite and were made at 60 μm (blue), 100 μm (green) and 240 μm (red). Zodiacal light due to sunlight scattered by interplanetary dust has been removed from this image. (Courtesy of Edward Wright and the COBE Science Team.)

The first complete survey of the far-infrared sky was carried out in 1983–4 by the Infrared Astronomical Satellite (IRAS) in four broad wavelength bands centred on 12, 25, 60 and 100 μm . It revealed intense far-infrared emission from regions of star formation in our own Galaxy and nearby galaxies as well as a host of new detections of stars, galaxies, active galaxies and quasars. Among the most important discoveries was a class of starburst galaxies which emit most of their radiation in the far-infrared waveband.

A more recent image of the far-infrared sky has been created from observations with the DIRBE instrument of COBE from all-sky maps made at 60, 100 and 240 μm (Fig. 1.5). The emission seen in this image is almost entirely the radiation of heated dust grains. Regions of star formation are particularly prominent features of the image. They can be seen forming a thin disc in the Galactic plane, as well as being present in the Magellanic Clouds, which are well known to be sites of active star formation. Intense emission associated with the Orion Molecular Cloud can be seen towards the right-hand edge of the image in the southern Galactic hemisphere. The Orion Nebula is of particular importance for studies of star formation since it is the region of massive star formation closest to the Earth. The colour coding of Fig. 1.5 is such that hot and cold dust have blue and red tinges, respectively. The bluish regions are mostly associated with discrete regions of active star formation, while the reddish clouds appear all over the image and extend to high Galactic latitudes. The latter clouds are often referred to as *infrared cirrus*.

The importance of these observations for high energy astrophysics is that they indicate where active regions of star formation are located. These are always associated with regions in which the interstellar gas densities are high and this is particularly important in studies

of the nuclei of active galaxies. The relation between star formation and high energy astrophysical activity is one of the more important and intriguing features of this study.

1.5 Millimetre and submillimetre waveband

$$30 \text{ GHz} \leqslant \nu \leqslant 3 \text{ THz}; 10 \geqslant \lambda \geqslant 0.1 \text{ mm}$$

1.5.1 Observing in the millimetre and submillimetre waveband

The millimetre and submillimetre wavebands are particularly rich astronomically. In addition to the extension of radio astronomical phenomena to shorter wavelengths, distinct features of these wavebands are the presence of a wealth of molecular lines in cool sources and the Cosmic Microwave Background Radiation. The transparency of the atmosphere varies dramatically with wavelength in this waveband (Fig. 1.3). At wavelengths less than about 1 mm, there are very strong absorption bands due to water vapour, carbon dioxide and other molecules in the Earth's atmosphere. The transparency of the atmosphere is particularly sensitive to the amount of water vapour in the atmosphere. To have a reasonable chance of making observations in the atmospheric windows at 850, 650, 450 and 350 μm , it is essential to observe from a high, dry site. Examples of such sites include: the Mauna Kea Observatory in Hawaii at 4200 m, where the James Clerk Maxwell Telescope (JCMT), the CalTech Submillimetre Observatory (CSO) and the Smithsonian SubMillimetre Array (SMA) are located; the Chajnantor plateau in the Atacama desert in Chile at 5100 m, the site of the Atacama Large Millimetre Array (ALMA); and the South Pole where the sub-zero temperatures ensure that there is very low precipitable atmospheric water vapour. At these sites, there is less than 1 mm of precipitable water vapour for considerable fractions of the time, enabling observations to be made in the shortest wavelength windows. To make observations in the other parts of the waveband, it is necessary to make observations from above the Earth's atmosphere, either from high-flying aircraft, such as the Kuiper Airborne Observatory, or from satellite observatories.

The receivers and detectors for the millimetre and submillimetre wavebands have developed dramatically over the last 10 years. Before that time, observations were made using single element detectors which were either heterodyne receivers, similar to those familiar in the radio waveband, or bolometers which measured the total incident power within a given waveband. In 1997, the first submillimetre camera, the SCUBA submillimetre bolometer array, was commissioned on the JCMT and has revolutionised studies in these wavebands. Arrays of heterodyne receivers are also now available which enable the spectral mapping of extended astronomical objects to be carried out.

1.5.2 Millimetre and submillimetre all-sky images

Millimetre and submillimetre all-sky images are dominated by the Cosmic Microwave Background Radiation which was discovered, more or less by chance, by Penzias and Wilson in 1965 (Penzias and Wilson, 1965). Figure 1.6a illustrates the stunning result that

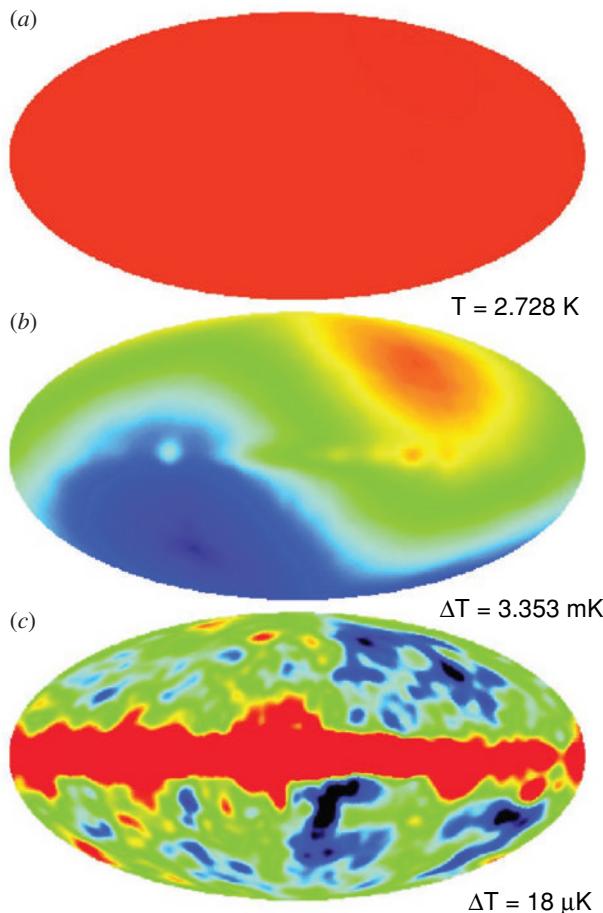
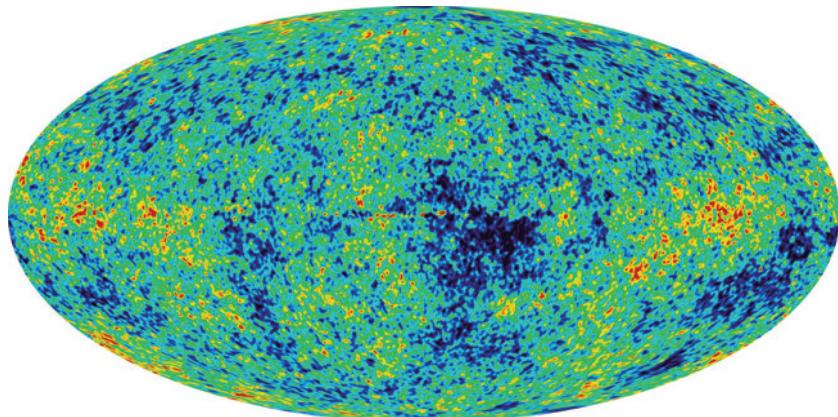


Fig. 1.6 Maps of the whole sky in Hammer–Aitoff projections in Galactic coordinates as observed at a wavelength of 5.7 mm (53 GHz) by the COBE satellite at different sensitivity levels. (a) The distribution of total intensity over the sky. (b) Once the uniform component is removed, a dipole component associated with the motion of the Earth through the isotropic background radiation is observed, as well as a weak signal from the Galactic plane. (c) Once the dipole component is removed, radiation from the plane of the Galaxy is seen as a bright band across the centre of the picture. The fluctuations seen at high Galactic latitudes are a combination of noise from the telescope and the instruments and a genuine cosmological signal. At high latitudes, an excess sky noise signal of cosmological origin amounts to $30 \pm 5 \mu\text{K}$ (Bennett *et al.*, 1996).

the Cosmic Microwave Background Radiation is extraordinarily uniform over the whole sky with a perfect black-body spectrum at a radiation temperature of 2.728 K. It is wholly convincing that this radiation is the cooled remnant of the hot early phases of the Big Bang.

At a sensitivity level of about one part in 1000 of the total intensity, large scale anisotropy of dipolar form is observed over the whole sky (Fig. 1.6b). The plane of our Galaxy can also be observed as a faint band of emission along the Galactic equator. The global dipole anisotropy is naturally attributed to aberration effects associated with the

**Fig. 1.7**

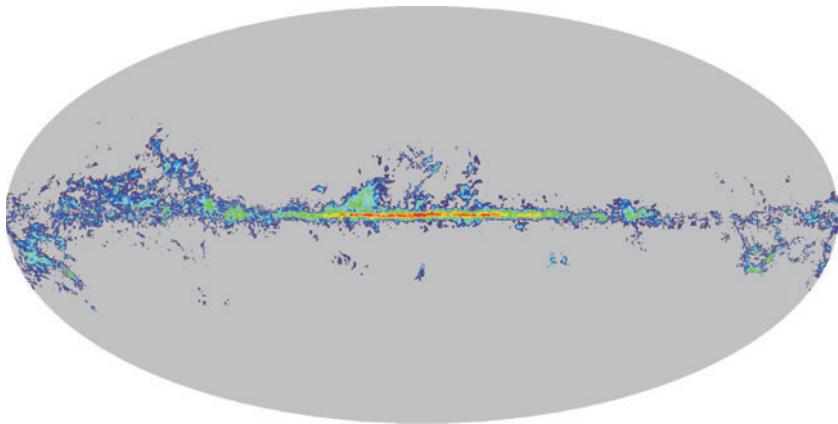
A map of the whole sky in Galactic coordinates as observed by the WMAP satellite at millimetre wavelengths (Bennett *et al.*, 2003). The angular resolution of the map is about 20 times higher than that of Fig. 1.6c. The emissions due to Galactic dust and synchrotron radiation have been subtracted from this map.

Earth's motion through an isotropic radiation field. Excluding regions close to the Galactic plane, the temperature distribution was found to have precisely the expected dipolar form, $T = T_0[1 + (v/c) \cos \theta]$, where θ is the angle with respect to the direction of maximum intensity and v is the Earth's velocity through the isotropic background radiation. The amplitude of the cosmic microwave dipole was 3.353 ± 0.024 mK (Bennett *et al.*, 1996). It was inferred that the Solar System is moving at about 350 km s^{-1} with respect to the frame of reference in which the radiation would be 100% isotropic.

On angular scales of 7° and greater, Bennett and his colleagues achieved sensitivity levels better than one part in 100 000 of the total intensity from analyses of the complete microwave dataset obtained over the four years of the COBE mission (Fig. 1.6c). At this sensitivity level, the radiation from the plane of the Galaxy is intense, but is confined to a broad strip lying along the Galactic equator. Away from the plane, there are significant intensity fluctuations of cosmological origin from beamwidth to beamwidth over the sky. These fluctuations are present at the level of only about 1 part in 100 000 of the total intensity. The detection of these fluctuations is a crucial result for understanding the origin of large scale structures in the Universe.⁴ It is interesting to compare the COBE map (Fig. 1.6c) with the more recent Wilkinson Microwave Anisotropy Probe (WMAP) observations made with about 20 times higher angular resolution (Bennett *et al.*, 2003) (Fig. 1.7). It can be seen that the same large scale features are present on both maps.

There is however, much more to the millimetre and submillimetre wavebands than just the Cosmic Microwave Background Radiation. The dust emission seen in Fig. 1.5 has a continuum spectrum with a strongly inverted spectrum, roughly $I_\nu \propto \nu^{3-4}$, and so contributes to the background radiation at submillimetre wavelengths. In addition, line emission of interstellar molecules is observed from the plane of the Galaxy and is particularly intense in regions of star formation. The commonest interstellar molecule is molecular

⁴ I have dealt *in extenso* with these observations and their interpretation in my book *Galaxy Formation* (Longair, 2008).

**Fig. 1.8**

A map of the whole sky in Galactic coordinates in the carbon dioxide molecule CO. (Courtesy of the LAMBDA program of GSFC of NASA.)

hydrogen but it has zero electric dipole moment and so is not observed in emission. The next most common molecule is carbon monoxide CO which has a strong electric dipole moment and is observed throughout the plane of the Galaxy, as can be seen in Fig. 1.8. The radiation is narrowly confined to the Galaxy plane in the hemisphere towards the Galactic Centre, but in the anti-Centre direction the distribution is somewhat broader. To the right of the image in the southern Galactic hemisphere, the giant molecular cloud associated with the Orion Complex can be seen, centred on the Orion Nebula.

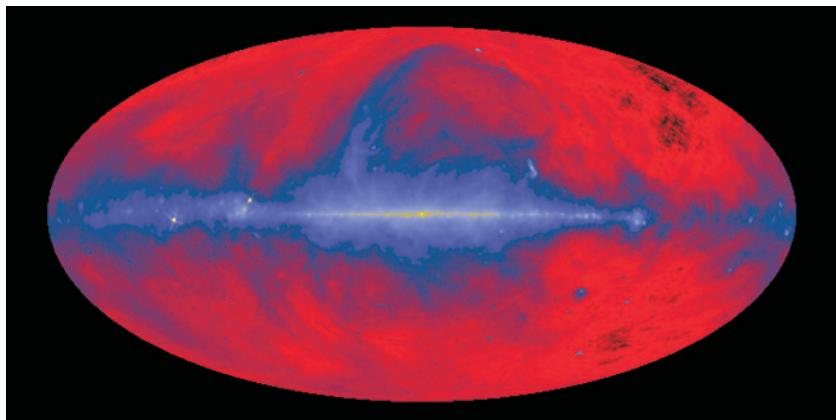
In addition, the continuum radiation of radio sources observed in the metre and centimetre radio wavebands is also present at millimetre wavelengths. These include the diffuse radio synchrotron and bremsstrahlung emission of the interstellar medium of our Galaxy and discrete Galactic and extragalactic radio sources, which are described in the next section. From the perspective of studies of the Cosmic Microwave Background Radiation, the dust and radio background components are regarded as interfering foregrounds which need to be carefully subtracted from the millimetre sky maps to reveal the underlying cosmological signals.

Observations in the millimetre and submillimetre wavebands impact high energy astrophysics in many different ways. Perhaps most significantly, the Cosmic Microwave Background provides an omnipresent radiation background from which high energy particles cannot escape.

1.6 Radio waveband $3 \text{ MHz} \leq \nu \leq 30 \text{ GHz}$; $100 \text{ m} \geq \lambda \geq 1 \text{ cm}$

1.6.1 Radio astronomy and the origin of high energy astrophysics

Radio waves of extraterrestrial origin were discovered by Jansky in the early 1930s but this caused little stir in the astronomical community. After the Second World War,

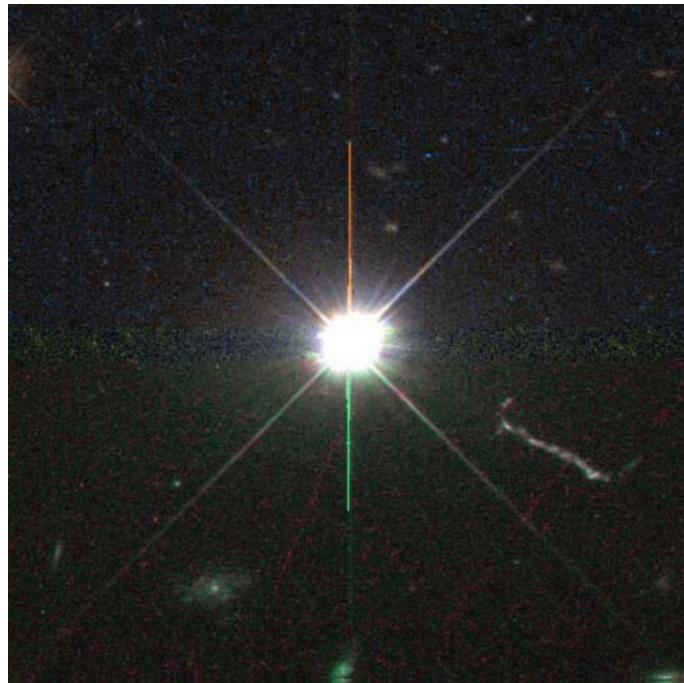
**Fig. 1.9**

Images of the celestial sphere at a radio frequency of 408 MHz in a Hammer–Aitoff projection. This image is dominated by the radio emission of relativistic electrons gyrating in the interstellar magnetic field, the process known as *synchrotron radiation*. The radiation is most intense in the plane of the Galaxy but it can be seen that there are extensive ‘loops’ and filaments of radio emission extending far out of the plane. (Courtesy of Max-Planck-Institut für Radioastronomie, Bonn.)

radio astronomy developed very rapidly as major advances were made in electronics, radio techniques and digital computers. Radio emission was discovered from a wide range of different astronomical objects. Some of the radio emission processes could be associated with phenomena observed at optical wavelengths, for example, the free–free or bremsstrahlung emission of hot electrons in regions of ionised hydrogen, but others were quite new. It was soon established that the radio emission of most sources was the synchrotron radiation of ultra-relativistic electrons spiralling in magnetic fields. Contrary to what might have been expected from Fig. 1.1a, the radio observations provided information about some of the very hottest, relativistic, plasmas in the Universe.

Two features of the radio observations were of particular significance. First, a number of the most massive galaxies known were found to be extremely powerful sources of radio waves. They were so powerful that it was easy to detect them as radio sources at cosmological distances. Estimates of the amount of energy necessary to power these radio sources showed that they must contain an energy in relativistic matter equivalent to a rest mass energy of about 100 million solar masses, that is, $10^8 M_\odot c^2 \approx 2 \times 10^{55}$ J. These galaxies had to be able to convert mass of this order into relativistic particle energy. The second key fact was that the radio emission did not generally originate from the galaxy itself but from huge radio lobes which extended far beyond the confines of the parent galaxy. In the 1960s and 1970s it was established that the sources of these vast energies were the active nuclei of the host galaxies and that the extended structures resulted from the expulsion of this energy from the nuclei in the form of jets of relativistic plasma. These discoveries revealed the presence of two major new components of the Universe, *relativistic plasma* and *magnetic fields*. These discoveries were the touchstone for the explosive growth of high energy and relativistic astrophysics over subsequent years.

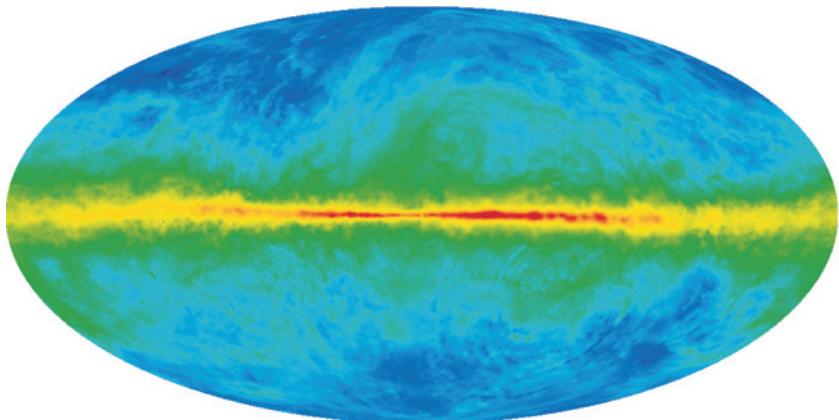
The study of these radio sources led to further discoveries. Amongst the earliest of these was the fact that supernova remnants are very powerful sources of synchrotron radio

**Fig. 1.10**

A Hubble Space Telescope image of the quasar 3C 273, showing the optical jet ejected from the quasar nucleus. The images at the bottom of the picture are galaxies at the same distance as the quasar. 3C 273 was the first quasar for which a redshift was measured, thanks to the presence of the redshifted Balmer series of hydrogen in emission in its optical spectrum (Schmidt, 1963). (Courtesy of NASA and the Space Telescope Science Institute.)

emission and so must be capable of accelerating charged particles to ultra-relativistic energies and creating strong magnetic fields. A remarkable outcome of the study of the extragalactic radio sources was the discovery of the *quasi-stellar radio sources*, or *quasars*, in the early 1960s. In these objects, the starlight of the galaxy is completely overwhelmed by the intense non-thermal optical radiation from the nucleus, in some cases, the optical luminosity being more than 1000 times greater than that of the parent galaxy (Fig. 1.10). These objects and their close relatives, the *BL-Lacertae* or *BL-Lac* objects, which were discovered in 1968, are among the most powerful energy sources known in the Universe.

But more was to follow. In 1967 Hewish and Bell constructed a low frequency radio array to study very short time-scale fluctuations imposed upon the intensities of compact radio sources by density fluctuations in the interplanetary plasma streaming out from the Sun, what is known as the *Solar Wind*. During the commissioning phase of the array, sources consisting entirely of pulsed radio emission with very stable periods of about 1 s were discovered, the *radio pulsars*. They were soon identified conclusively as rotating, magnetised *neutron stars* and thus provided the first definite proof of the existence of these highly compact stars in which the central densities are as high as $10^{18} \text{ kg m}^{-3}$. A key point from the perspective of relativistic astrophysics was the fact that solar mass objects had been discovered with radii only about a factor of 4 or 5 times greater than the Schwarzschild radius of solar mass black holes. Thus, in these compact objects, general

**Fig. 1.11**

A map of the whole sky in Galactic coordinates in the 21-cm line of neutral hydrogen. (Courtesy of the LAMBDA programme of GSFC of NASA.)

relativity is no longer a small correction term to the equations of motion – these objects provide laboratories for the study of matter in strong gravitational fields.

1.6.2 Neutral hydrogen and molecular line astronomy

One of the great predictions of modern astronomy was made during the Second World War by van de Hulst who, at the suggestion of Oort, calculated which emission and absorption lines of atoms, ions and molecules might be detectable from astronomical sources in the radio waveband. The most significant prediction was that neutral hydrogen should emit line radiation at a wavelength of about 21 cm because of the minute change in energy when the relative spins of the proton and electron in a hydrogen atom change. Although this is a highly forbidden transition with a spontaneous transition probability of only once every 12 million years, there is so much neutral hydrogen present in the Galaxy that it was predicted that it should be detectable. In 1951, the 21-cm line of neutral hydrogen was discovered by Ewen and Purcell and it has proved to be a very powerful tool for diagnosing not only the properties of the interstellar gas but also the dynamics of galaxies. The 21-cm line is generally so narrow that it provides an excellent measure of the velocity fields inside galaxies.

Figure 1.11 shows the distribution of neutral hydrogen in an all-sky projection in Galactic coordinates. The 21-cm emission from the plane of the Galaxy is confined to a rather thin layer, but in addition there are loops, high-velocity clouds and diffuse neutral hydrogen extending to high Galactic latitudes.

Molecules had been known to exist in the interstellar medium from observations of the absorption bands seen in the optical spectra of stars. The real significance of molecular line astronomy only became apparent, however, with the development of radio telescopes and line receivers operating in the centimetre and millimetre wavebands. In 1967, the hydroxyl radical, OH, was first detected by radio techniques in molecular lines at four frequencies in the range 1.6–1.7 GHz. This was a somewhat unexpected detection because the signals were

very strong indeed and variable in intensity. The brightness temperatures of the sources were greater than 10^9 K, indicating that some form of *maser* action must be overpopulating the upper energy levels of the transitions. The populations of the energy levels of the molecules must be far from equilibrium so that intensities far exceeding those expected from the thermodynamic temperature of the source region are observed.

Many more molecules were soon discovered, mostly through observation of the emission lines associated with rotational transitions in the centimetre, millimetre and submillimetre wavebands. Small molecules such as carbon monoxide radiate in the millimetre and submillimetre wavebands as was discussed in the last section, but larger linear molecules with up to 11 atoms radiate in the centimetre radio waveband. These studies led to the development of the discipline of *interstellar chemistry*. For the molecules to survive, it is essential that they should be shielded from the intense interstellar ultraviolet radiation field. It is therefore not surprising that molecules are found in large abundances in dusty star-formation regions in which they are protected from the interstellar flux of dissociating ultraviolet radiation.

1.6.3 Observing the radio sky

The pioneering radio astronomical observations were made at metre wavelengths but, as radio technology developed through the 1960s and 1970s, observations became possible at the shortest centimetre wavelengths. In addition to observations with single radio antennae, the principles of aperture synthesis were used to provide high angular resolution images by combining the signals in phase from large interferometer arrays. The state-of-the-art in high resolution imaging is provided by facilities such as the Very Large Array (VLA) in New Mexico and the Australia Telescope National Facility (ATNF). The use of very long baseline interferometry (VLBI) at centimetre and millimetre wavelengths can provide an angular resolution of milliarcseconds or better. These types of observation are of special importance for studies of the physics of those active galactic nuclei which are intense emitters in these wavebands.

At the low frequency end of this range, 1–10 MHz, observations of extraterrestrial sources become very difficult because of the reflection of radio waves by the plasma of the ionosphere. There are, however, certain favourable sites close to the auroral zones at which the sky can be observed. Even if the telescope is located above the ionosphere, however, observations at frequencies less than about 1 MHz become essentially impossible because of the same plasma reflection effects occurring in the interplanetary and interstellar plasma.

$$\begin{aligned} \textbf{1.7 Ultraviolet waveband } & 10^{15} \leq \nu \leq 3 \times 10^{16} \text{ Hz;} \\ & 300 \geq \lambda \geq 10 \text{ nm} \end{aligned}$$

The atmosphere is opaque to radiation in this waveband because of ozone and molecular absorption and so observations have to be carried out from above the atmosphere. The band

divides rather naturally into two regions. The region $300 \geq \lambda \geq 120$ nm can be studied using techniques similar to those used in the optical waveband. Ultraviolet spectrographs were flown on rockets in the mid-1960s and were followed by the series of orbiting astrophysical observatories, culminating in the launch of the International Ultraviolet Explorer (IUE) in 1978. As expected, a wide range of hot objects could be studied but perhaps of most importance was the fact that a wide range of the common elements could be observed because their strong resonance transitions fall in the ultraviolet spectral region. Active galactic nuclei are particularly strong emitters in the ultraviolet waveband because the non-thermal radiation observed in the optical waveband extends to far-ultraviolet wavelengths. This continuum radiation excites a wide range of ions and atoms which emit strong resonance lines in the ultraviolet waveband. These lines have proved to be particularly valuable diagnostic tools for the astrophysics of active galactic nuclei.

Observations at shorter ultraviolet wavelengths, $\lambda < 120$ nm, proved to be more difficult – these are referred to as the extreme ultraviolet (EUV) wavebands. There are two reasons for this. First, there is the problem of constructing an efficient telescope because most materials are strongly absorbant for normal incidence optics at wavelengths shorter than about 120 nm. One solution is to use grazing rather than normal incidence optics and then the ultraviolet radiation can be focussed in a similar manner to optical radiation. As a result, the telescopes look rather different from optical telescopes. Another problem is that at wavelengths shorter than 91.2 nm, the Lyman limit for hydrogen, it is expected that the interstellar gas becomes opaque because of photoelectric absorption by neutral hydrogen in the Lyman continuum. Fortunately, the distribution of neutral hydrogen is sufficiently clumpy for there to be ‘holes’ through the interstellar gas which enable the more distant Universe to be observed.

Surveys of the far-ultraviolet sky were carried out in the 1990s by the ROSAT Wide Field Camera which operated in the 60–210 eV energy band (6–20 nm) and the Extreme Ultraviolet Explorer (EUVE) of NASA which observed in the 6–74 nm waveband (Pye *et al.*, 1995; Christian, 2002). These surveys showed that the bright sources are remarkably uniformly distributed over the sky, but this is because these are mostly nearby objects in our own Galaxy. The majority population of the sources are hot white dwarfs, active and nearby late-type stars and cataclysmic variables. Along lines of sight in which the column density of neutral hydrogen is small, a total of 19 active galactic nuclei were observed by the ROSAT Wide Field Camera, eight being narrow-line Seyfert I galaxies, six broad-line radio Seyfert galaxies and five BL-Lac objects (Edelson *et al.*, 1999).

$$\textbf{1.8 X-ray waveband } 3 \times 10^{16} \leq \nu \leq 3 \times 10^{19} \text{ Hz}; \\ 10 \geq \lambda \geq 0.01 \text{ nm}; 0.1 \leq E \leq 100 \text{ keV}$$

1.8.1 Observing the X-ray sky

As in the case of the far-ultraviolet waveband, the atmosphere is opaque to X-rays because of photoelectric absorption by the atoms which make up the molecular gases of the atmosphere and so X-ray astronomy is wholly carried out from above the atmosphere (Fig. 1.1b). The

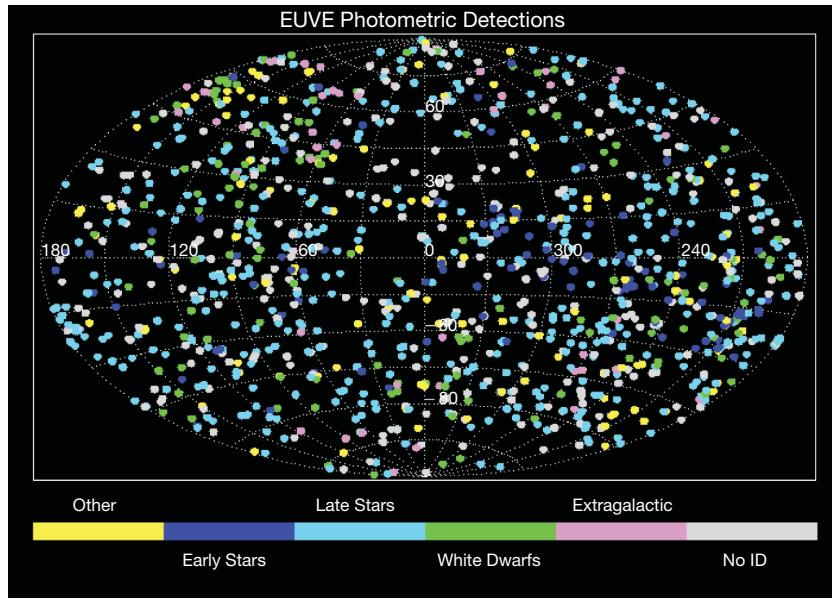


Fig. 1.12 An image of the celestial sphere in the extreme ultraviolet waveband, 6–74 nm, in a Hammer–Aitoff projection observed by the Extreme Ultraviolet Explorer (EUVE). Most of the 1200 point sources in the diagram are relatively nearby objects in our own Galaxy, as indicated by the colour coding at the bottom of the image (Christian, 2002).

detectors resemble those used in particle physics experiments – proportional counters and scintillation detectors are used as well as other devices such as CCDs in which the total energy deposited by the X-ray on entering the detector is measured. The photons are of such high energy that they behave like particles, and the telescopes for high energy X-rays are essentially collimators in which the resolution of the telescope is determined by the geometric design of the collimator. At low X-ray energies, $0.1 < E < 1$ keV, grazing incidence optics can be used to image the X-rays at the focal plane, but at higher energies the grazing incidence angles are so small that enormously long telescopes would be needed to focus the image.

Once rockets capable of lifting scientific payloads above the atmosphere became available, the exploration of the X-ray sky was possible but these provided only about five minutes of observation. This was enough, however, even in the first rocket flights of 1962 and 1963, to show that the X-ray sky was rich for astrophysical study. As in the case of the radio waveband, the sources which were first observed had not been predicted by astrophysicists. Amongst the earliest detections in the 1–10 keV waveband were the supernova remnant the Crab Nebula, the nearby radio galaxy M87, a number of stellar X-ray sources, which seemed to be highly variable, and the diffuse X-ray background radiation.

1.8.2 The X-ray sky

The full scope of X-ray astronomy became clear in the early 1970s with the launch of the first dedicated X-ray satellite, the UHURU satellite observatory, which mapped the X-ray sky and provided systematic monitoring of variable X-ray sources (Fig. 1.13a). The

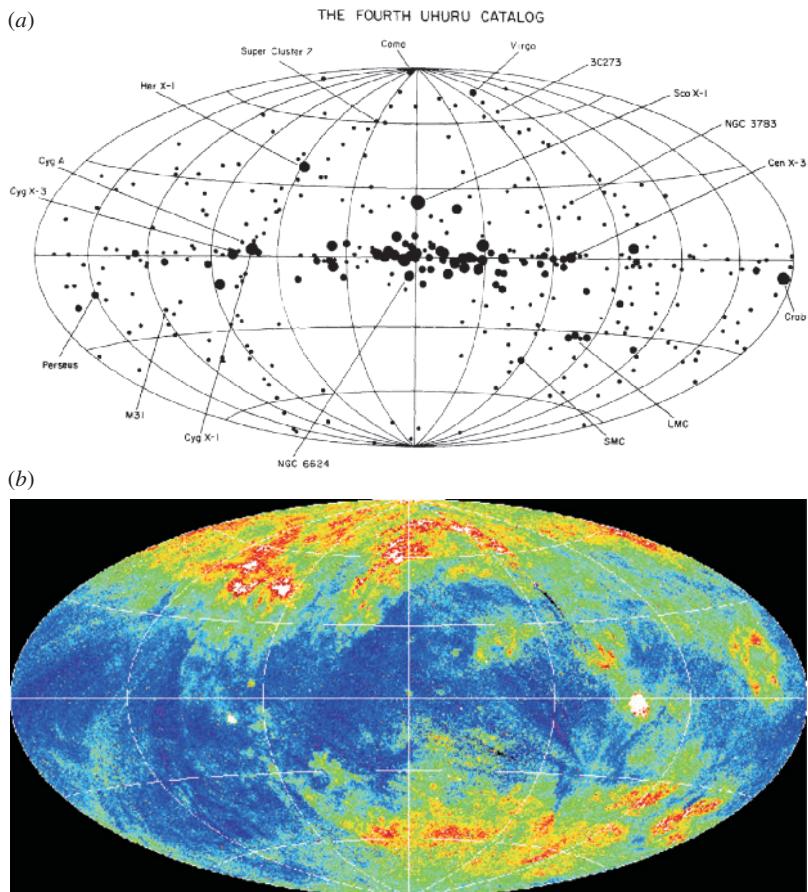


Fig. 1.13 (a) The UHURU map of the brightest X-ray sources in the 2–6 keV energy band. The identifications of a number of the brightest sources are indicated (Forman *et al.*, 1978). These include the quasar 3C 273, the Coma, Perseus and Virgo Clusters of galaxies, the radio galaxy Cygnus A, the low mass X-ray binary Sco X-1, the high mass binaries Cyg X-1 and Cyg X-3 and the supernova remnant the Crab Nebula. (b) The image of the celestial sphere in the softest X-ray energy band 0.25 keV derived from the ROSAT survey with the point sources removed. The colour coding is such that white is the greatest intensity and blue the lowest. At these soft X-ray energies, the intensity is anti-correlated with the distribution of neutral hydrogen (Fig. 1.11) because of photoelectric absorption by the interstellar gas. (Courtesy of the ROSAT project and the Max Planck Institute for Extraterrestrial Physics, Garching.)

variability of some of the Galactic sources was found to be due to the fact that the compact X-ray emitter is a member of an eclipsing binary star system. In a number of these cases, the X-ray binaries were found to contain ‘pulsating’ X-ray sources and these were soon identified with magnetised rotating neutron stars but, in the cases of the X-ray sources, the source of energy is the infall of matter transferred from the primary star, the process known as *accretion*. In the case of the pulsating X-ray sources, the inferred masses are consistent with their being neutron stars but, in a number of cases, the masses of the invisible secondaries exceed the upper limit for stable neutron stars. These objects must be

associated with stellar-mass *black holes* in binary systems and as such are objects of the greatest astrophysical interest.

In extragalactic astronomy, the nuclei of active galaxies were found to be intense and often variable X-ray sources, the emission processes taking place close to the Schwarzschild radius of a supermassive black hole. Other important classes of extragalactic X-ray sources are the clusters of galaxies. The mass of the cluster gives rise to a deep gravitational potential well in which the gas must be very hot if it is to form a stable extended atmosphere. The intense X-ray emission observed from the intracluster gas is the thermal free–free emission, or bremsstrahlung, at temperatures in the range 10^7 – 10^8 K. The high temperature of the gas is confirmed by the observation of very highly ionised iron lines from the intracluster gas. The characteristic of these cluster sources is that the thermal X-ray emission is extended and this provides a powerful means of identifying clusters of galaxies at large distances, as well as providing important tests of the theory of their formation.

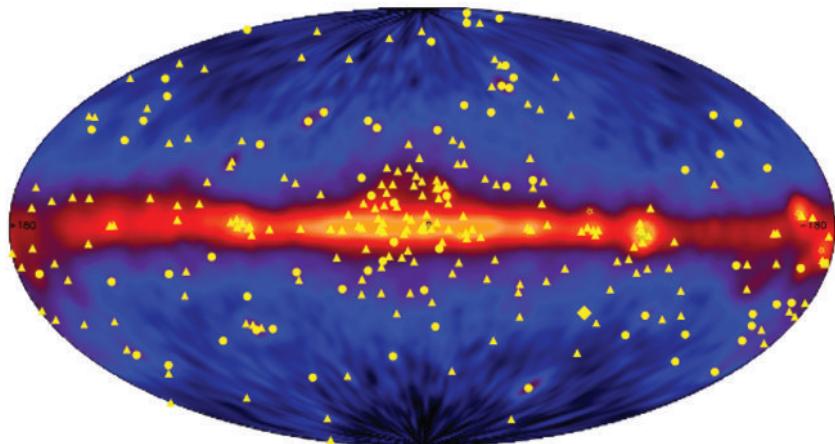
In 1978, the Einstein X-ray Observatory was launched. It provided the first high resolution images of many X-ray sources and made deep surveys of small areas of sky. Many different classes of astronomical object were detected as X-ray sources including regions of star formation and normal galaxies. Perhaps most significant of all was the fact that X-ray emission was detected from all types of star and not just from the binary sources in which there are special reasons why they should be strong X-ray sources.

Surveys of the whole sky were carried out in the X-ray waveband 0.25–2 keV by the ROSAT X-ray observatory during the 1990s, the final catalogues including over 100 000 X-ray sources. The image of the celestial sphere in the softest X-ray energy band, 0.25 keV, derived from the ROSAT survey with the point sources removed is shown in Fig. 1.13b. Regions of the greatest intensity are shown as white, while the lowest intensities are coloured blue. At these soft X-ray energies, the intensity is anticorrelated with the distribution of neutral hydrogen (Fig. 1.11) because of photoelectric absorption by the interstellar gas. At higher energies, the distribution of sources consists of a Galactic population of the types shown in Fig. 1.13a as well as an isotropic distribution of extragalactic sources, most of them being associated with active galactic nuclei.

The ROSAT mission was followed by two major observatory-class missions. The Chandra X-ray observatory of NASA was primarily a high resolution imaging telescope providing images with angular resolution $\theta \sim 0.5$ arcsec, comparable to the best images achieved by large ground-based optical telescopes. The second was the XMM-Newton X-ray Observatory of ESA which was primarily an X-ray spectroscopic mission with large collecting aperture to provide high sensitivity, high spectral resolution observations of all classes of X-ray source. It is no exaggeration to state that these telescopes have revolutionised the science of X-ray astrophysics.

1.9 γ -ray waveband $\nu \geqslant 3 \times 10^{19}$ Hz; $\lambda \leqslant 0.01$ nm;
 $E \geqslant 100$ keV

Photons with energies greater than about 100 keV are referred to as γ -rays. Except at the very highest energies, these studies have to be carried out from above the atmosphere.

**Fig. 1.14**

An image of the celestial sphere at γ -ray energies $\varepsilon \geq 100$ MeV in a Hammer–Aitoff projection from observations made by the EGRET instrument of the Compton Gamma-Ray Observatory (CGRO). The emission from the plane of the Galaxy consists of diffuse γ -ray emission from the interstellar gas, most of it associated with γ -rays produced by the decay of neutral pions, π^0 , generated in collisions between cosmic ray protons and nuclei and the interstellar gas. The yellow symbols show the distribution of discrete sources detected in the all-sky survey: circles are active galactic nuclei; five-point stars are pulsars; squares are solar flares; the diamond is the Large Magellanic Cloud; and the triangles are unidentified sources. (Courtesy of NASA and the EGRET science team.)

Between 100 keV and 1 MeV, photoelectric absorption is the dominant absorption mechanism but at higher energies Compton scattering and then electron–positron pair production become the principal absorption processes. The detectors used in γ -ray satellites are similar to those used in particle physics experiments but they have to be miniaturised so that they can be flown in orbit. At the very highest energies, $E \geq 10^{11}$ eV, γ -rays from extraterrestrial sources are so energetic that they initiate electromagnetic cascades in the upper atmosphere and the Cerenkov radiation of the ultra-relativistic electrons and positrons produced in these showers can be detected at ground level.

γ -ray emission from the plane of our Galaxy was first detected by the OSO III satellite in 1967. This was followed by the SAS-2 satellite which discovered the diffuse γ -ray background and by the COS-B satellite which provided a detailed map of the Galactic γ -ray emission and discovered about 25 discrete γ -ray sources. These included the pulsars in the Crab and Vela supernova remnants and the quasar 3C 273. A γ -ray map of the whole sky in the energy band $\varepsilon \geq 100$ MeV was obtained from observations with the EGRET instrument of the Compton Gamma-Ray Observatory (Fig. 1.14). The image of the sky is dominated by the intense γ -ray emission from the Galactic plane. At photon energies, $\varepsilon \geq 100$ MeV, the principal emission mechanism is the decay of neutral pions, π^0 , created in collisions between the nuclei of atoms and molecules of the interstellar gas and cosmic ray protons and nuclei. At lower energies, non-thermal processes, in particular inverse Compton scattering and bremsstrahlung, can make contributions to the background γ -ray emission.

At high Galactic latitudes, most of the discrete sources are associated with active galactic nuclei. In particular, the most intense and variable sources are associated with those radio

quasars which exhibit the phenomenon of superluminal motions. The variability is so rapid, on the time-scale of days or less, that relativistic beaming of the γ -rays is needed to account for their observed properties.

The first evidence of γ -ray line emission came from balloon observations in the early 1970s by the Rice University Group. In 1977 definitive observations of the electron–positron annihilation line at 511 keV in the direction of the Galactic Centre were made by balloon observations. Since then observations have also been made of the 1.809 MeV line of radioactive ^{26}Al by the HEAO-C satellite, this line also being detected from the direction of the Galactic Centre. These studies have been greatly advanced by observations by the INTEGRAL γ -ray observatory of ESA.

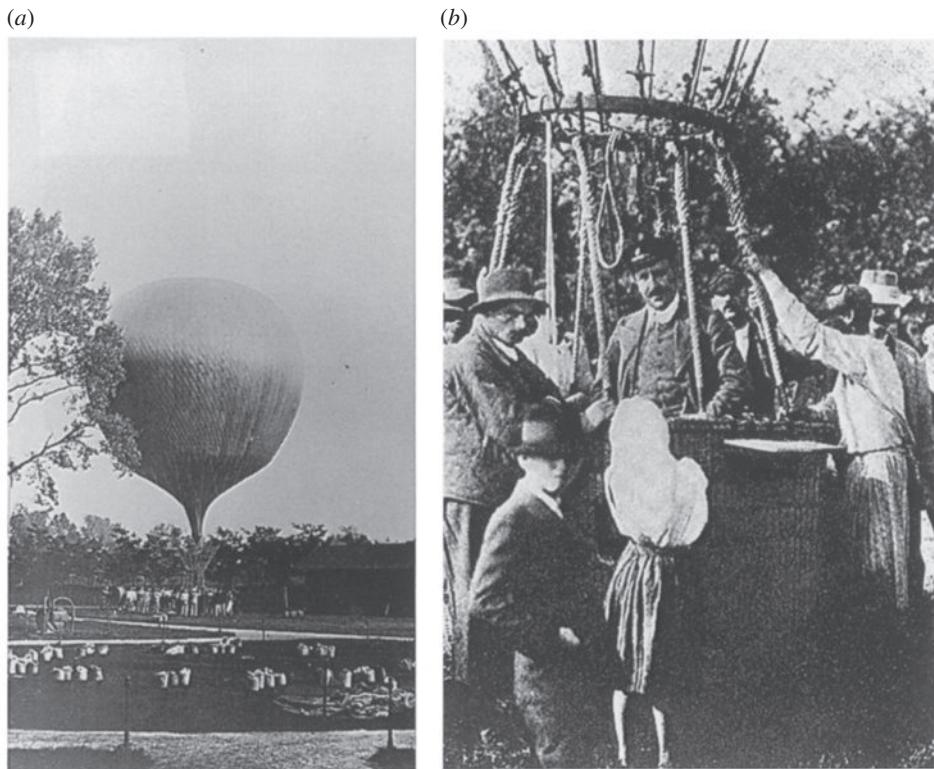
Another unexpected discovery was that of γ -ray bursts which were detected by the US Vela satellites and also by Soviet satellites. The Vela satellites were launched to monitor the sky in γ -rays to confirm compliance with the Nuclear Test Ban treaties. Bursts of γ -rays were discovered, but they proved to be of astronomical rather than terrestrial origin. The bursts last between 0.01 and 100 seconds and are uniformly distributed over the sky. Their nature as distant luminous extragalactic objects was established once it was realised that they have significant after-glowes at X-ray, optical and infrared wavelengths which enabled their positions to be determined accurately. The bursts are associated with extremely violent events involving stellar-mass objects in distant galaxies.

Very high energy γ -rays with $\epsilon \sim 10^{11-12}$ eV are detected by the optical Cherenkov radiation technique. γ -rays of these energies initiate electron–photon cascades in the upper atmosphere. The electrons are of such high energy that their velocities exceed the speed of light in air and consequently they emit optical Cherenkov radiation. The optical light emitted by these showers is detected at sea-level by telescope arrays. The introduction of multi-element detector arrays in the focal planes of the telescopes of the arrays, for example in the operation of the HESS array in Namibia, have revolutionised studies in these energy ranges. Among the more important observations have been images of the ultra-high energy γ -ray emission from supernova remnants, presumably associated with the high energy protons accelerated in their shells, and some relatively nearby active galactic nuclei which are of cosmological importance in setting upper limits to the extragalactic optical and infrared background radiation.

1.10 Cosmic ray astrophysics

1.10.1 A brief history of cosmic ray physics

The first hints that there is more to the Universe than stars, gas and dust came with the discovery of *cosmic rays*. The cosmic ray story began about 1900 when it was discovered that electroscopes discharged even if they were kept in the dark well away from sources of natural radioactivity. The big breakthrough came in 1912 and 1913 when first Hess and then Kolhörster made manned balloon ascents in which they measured the ionisation of the atmosphere with increasing altitude (Hess, 1912; Kolhörster, 1913) (Fig. 1.15). They

**Fig. 1.15**

The balloon flights of Victor Hess. (a) Preparation for one of his flights of 1911–12. (b) Hess after one of the successful balloon flights in which the increase in ionisation with altitude through the atmosphere was discovered (Sekido and Elliot, 1985).

found the startling result that the average ionisation increased with respect to the ionisation at sea-level above about 1.5 km (Table 1.1). This was clear evidence that the source of the ionising radiation must be located above the Earth's atmosphere.

In 1929, Skobeltsyn constructed a cloud chamber to study the properties of the electrons emitted in radioactive decays. Among the tracks, he noted some which were hardly deflected at all and which resembled electrons with energies greater than 15 MeV. He identified these with secondary electrons produced by the ‘Hess ultra γ -radiation’. These were the first pictures of the tracks of cosmic rays (Skobelzyn, 1929).

Also in 1929, the *Geiger–Müller detector* was invented which enabled individual cosmic rays to be detected and their arrival times determined very precisely (Geiger and Müller, 1928, 1929). In the same year, Bothe and Kolhörster carried one of the key experiments in cosmic ray physics in which they introduced the concept of *coincidence counting* to eliminate spurious background events (Bothe and Kolhörster, 1929). This coincidence technique is now standard practice in many different types of cosmic ray, X- and γ -ray experiments. By using two counters, one placed above the other, they found that simultaneous discharges of the two detectors occurred very frequently, even when a strong

Table 1.1 The variation of ionisation with altitude from the observations of Kolhörster (Kolhörster, 1913).

Altitude (km)	Difference between observed ionisation and that at sea-level ($\times 10^6$ ions m $^{-3}$)	Altitude (km)	Difference between observed ionisation and that at sea-level ($\times 10^6$ ions m $^{-3}$)
0	0	5	+16.9
1	-1.5	6	+28.7
2	+1.2	7	+44.2
3	+4.2	8	+61.3
4	+8.8	9	+80.4

absorber was placed between the detectors, indicating that charged particles of sufficient penetrating power to pass through both of them were common events. The inferred mass absorption coefficient agreed closely with that of the atmospheric attenuation of the cosmic radiation. They also showed that the flux of these particles could account for the observed intensity of cosmic rays at sea-level and that the energies of the particles had to be about 10^9 – 10^{10} eV.

The cloud chamber experiments showed that cosmic ray particles initiated showers of charged particles. Most of the high energy particles observed at the surface of the Earth are, in fact, secondary, tertiary or higher products of very high energy cosmic rays entering the top of the atmosphere. The full extent of some of these *extensive air showers* was established by Auger and his colleagues from observations with a number of separated detectors (Auger *et al.*, 1939). To their surprise, they found that the air showers could extend over dimensions greater than 100 metres on the ground and contained millions of ionising particles. The particles responsible for initiating the showers must have had energies exceeding 10^{15} eV at the top of the atmosphere. This was direct evidence for the acceleration of charged particles to extremely high energies in astronomical sources.

From the 1930s to the early 1950s, the cosmic radiation provided a natural source of very high energy particles which were energetic enough to penetrate into the nuclei of atoms. This was the principal technique by which new types of particles were discovered until the early 1950s. In 1930, Millikan and Anderson used an electromagnet 10 times stronger than that used by Skobeltsyn to study the tracks of particles passing through the cloud chamber. Anderson observed curved tracks identical to those of electrons but with positive rather than negative electric charge (Anderson, 1932). This discovery was confirmed by Blackett and Occhialini in 1933 using an automatic cloud chamber triggered when a cosmic ray passed through the volume of the chamber (Blackett and Occhialini, 1933). This discovery of the *positive electron* or *positron* coincided closely with Dirac's theory of the electron which had predicted its existence (Dirac, 1928a,b).

In 1936, Anderson and Neddermeyer used the cosmic ray technique to discover what they called *mesotrons*, particles with mass intermediate between that of the electron and the proton (Anderson and Neddermeyer, 1936). This discovery was more or less contemporaneous with Yukawa's prediction of the existence of an exchange particle which binds neutrons and protons together in the nucleus (Yukawa, 1935). In fact, the particles discovered by

Anderson and Neddermeyer, nowadays known as *muons*, were not the particles which bind nuclei together.

Similar experiments using nuclear emulsions were carried out immediately after the Second World War by Rochester and Butler who reported in 1947 the discovery of two cases of particle tracks in the form of ‘V’s with apparently no incoming particle (Rochester and Butler, 1947). Further examples of these *strange particles* were reported in the subsequent years and they are now referred to as *charged* and *neutral kaons* (K^+ , K^- , K^0). The culmination of these studies was the discovery of the *pion* (π) in 1947 using the nuclear emulsion technique – this was the particle predicted by Yukawa in 1935 (Lattes *et al.*, 1947).

By 1953, accelerator technology had developed to the point where energies comparable to those available in the cosmic rays could be produced in the laboratory with known energies and directed precisely onto the chosen target. After about 1953, the future of high energy physics lay in the accelerator laboratory rather than in the use of cosmic rays. Interest in cosmic rays shifted to the problems of their origin, chemical composition and their propagation in astrophysical environments from their sources to the Earth.

1.10.2 Cosmic ray astrophysics from space and from the ground

The astrophysical study of the origin and propagation of the cosmic ray particles had to await the 1960s when cosmic ray particle detectors were flown in satellites. These observations established many crucial facts about the primary particles present in the cosmic radiation. First of all, the energy spectra of the particles are of similar form to the typical spectrum of high energy particles inferred to be present in Galactic and extragalactic non-thermal radio sources. In the region of the energy spectrum which is unaffected by the propagation of the particles to the Earth through the Solar Wind ($E \geq 10^9$ eV), the energy spectra of the cosmic ray particles can be described by

$$N(E) dE = KE^{-x} dE \quad (1.2)$$

with $x \approx 2.5\text{--}2.7$ (Fig. 1.16). This relation is found to be applicable for protons, electrons and nuclei with energies in the range $10^9\text{--}10^{14}$ eV. The flux of cosmic ray particles can be related to the relativistic gas inferred to be present in the interstellar medium through two types of observation. First, the synchrotron radiation of ultra-relativistic electrons gyrating in the interstellar magnetic field is detected in the radio waveband. Secondly, the Galactic γ -ray emission at energies $E \gtrsim 100$ MeV is attributed to the decay of neutral pions π^0 created in collisions between interstellar high energy protons and nuclei and the nuclei of atoms, ions and molecules in the interstellar gas. The fact that these very different types of astronomy can be brought to bear successfully on these problems indicates that the cosmic ray particles observed at the top of the atmosphere sample the population of high energy particles pervading the whole interstellar medium of our Galaxy.

The chemical composition of the cosmic rays is similar to the abundances of the elements in the Sun with some important exceptions, particularly for the light elements lithium, beryllium and boron which are present with very high abundances in cosmic rays compared with their terrestrial values. These observations provide evidence about the chemical

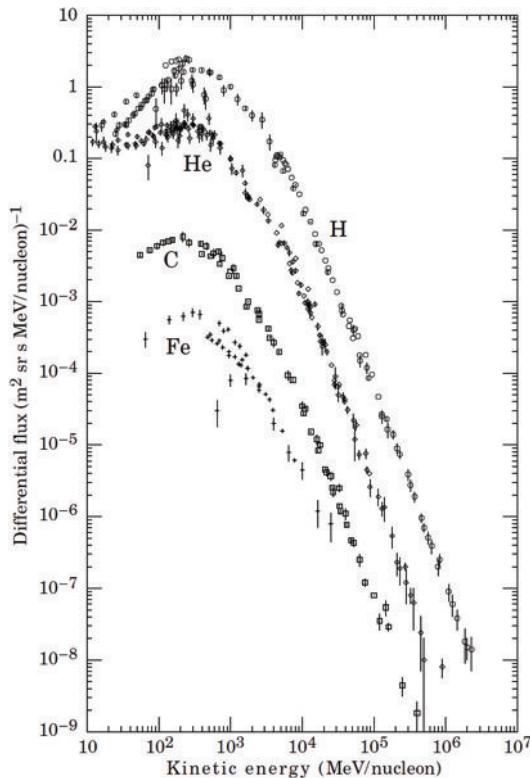


Fig. 1.16 The differential energy spectrum of cosmic rays as measured from above the Earth's atmosphere (Simpson, 1983). The solid line shows an estimate of the proton spectrum once allowance is made for the effects of solar modulation (see Sect. 7.3).

composition of the cosmic rays as they were accelerated in their sources and also about the modifications which must have taken place during propagation from their sources to the Earth. The importance of these observations for high energy astrophysics is that these are the only *particles* detected on Earth or in its vicinity which have traversed a considerable distance through the interstellar medium and which were accelerated in events such as supernovae in the relatively recent past, probably within the last 10^7 years.

At the very highest energies, cosmic rays are detected by large air shower arrays located on the surface of the Earth. The arrival rate of the most energetic particles is very low indeed but particles with energies up to about 10^{20} eV have been detected. One important puzzle was the origin of these extremely energetic particles. Until recently, their arrival directions seemed to be isotropic over the sky and, at these extreme energies, their trajectories should not be significantly influenced by the magnetic field in our own Galaxy. These problems have been largely resolved by the first observations by the huge Auger air-shower array in Argentina, which has improved sensitivity and angular resolution compared with previous experiments. Significant anisotropies have now been discovered in the arrival directions of the highest energy cosmic rays and a statistically significant association with nearby active

galaxies established. In addition, the expected cut-off in the spectrum above about 3×10^{19} eV due to interactions with photons of the Cosmic Microwave Background Radiation has been established. The acceleration mechanism for these particles is still uncertain.

1.11 Other non-electromagnetic astronomies

1.11.1 Neutrino astrophysics

The first triumph of neutrino astrophysics was the detection of neutrinos from the nuclear reactions which power the Sun. The neutrino signal detected by Davis and his colleagues at the *solar neutrino experiment* located in the Goldstake gold-mine in South Dakota amounted to only about a third of that predicted by the best solar models. The results of the Kamiokande experiment in Japan confirmed the deficit of neutrinos and showed that the detected neutrinos indeed originated in the Sun. During the 1990s, the GALLEX and SAGE experiments showed that the low energy neutrinos from the principal reaction of the main pp chain were present, but again at a somewhat lower level than expected. The solution to these discrepancies was the discovery of neutrino oscillations which not only showed that neutrinos have finite rest masses but also could account for the deficit of solar neutrinos. This picture has been confirmed in detail at the Sudbury Neutrino Observatory (SNO) which measured separately the contributions of the electron neutrinos and those of the muon and tau neutrinos.

The second key observation was the fortuitous detection of neutrinos from the explosion of the supernova SN 1987A in the Large Magellanic Cloud by the Kamiokande and IMB experiments. Only 20 neutrinos were detected altogether by these experiments in a 10 second interval. These neutrinos originated in the collapse of the central core of the blue supergiant star Sanduleak -69 202 to form a neutron star. These observations have provided insights into the physical processes by which the collapse of the core and the ejection of the stellar envelope took place.

These two spectacular results have encouraged the development of large neutrino detector arrays to observe the energetic neutrinos which are expected to accompany high energy phenomena in extreme astrophysical environments.

1.11.2 The search for gravitational waves

Einstein's general theory of relativity predicts the existence of gravitational waves, the gravitational counterparts of electromagnetic waves. Because of the weakness of the gravitational interaction, however, the sources have to be very luminous indeed if there is to be any chance of detecting them directly. The sources of the waves must involve very compact, indeed relativistic systems, and so there is no question but that they must involve high energy astrophysical processes. A great boost to the endeavours to detect the waves by direct observation was provided by the observed decay of the orbits of binary neutron star systems. The observed acceleration of their orbits match precisely the predictions of

gravitational radiation theory. The direct detection of gravitational waves remains, however, one of the most demanding challenges facing astronomical technologists.

The search for gravitational waves was begun by Weber in a pioneering set of experiments carried out in the 1960s. His first published results caused a sensation when he claimed to have found a positive detection of gravitational waves by correlating the signals from two gravitational wave detectors separated by a distance of 1000 km at the University of Maryland and the Argonne National Laboratory (Weber, 1969). In a subsequent paper, he reported that the signal originated from the general direction of the Galactic Centre (Weber, 1970). These results were received with considerable scepticism by the astronomical community since the reported fluxes far exceeded what even the most optimistic relativists would have predicted for the flux of gravitational waves originating anywhere in the Galaxy. As a result of Weber's claims, a major effort was made by experimentalists to reproduce his results and, in the end, these were not successful.

The challenge to the experimental community was how to detect the extremely tiny strains expected from sources of gravitational waves. The outcome was the approval of a number of major national and international experiments designed to detect the elusive gravitational waves. The LIGO project, an acronym for Laser Interferometer Gravitational-Wave Observatory, consists of two essentially identical interferometers each with 4-km baselines located at Livingston, Louisiana and Hanford near Richland, Washington. Similarly, the VIRGO project is a French-Italian collaboration to construct an interferometer with a 3-km baseline at a site near Pisa, Italy. The GEO600 experiment is a German-UK interferometer project with a 600 metre baseline, while the Japanese TAMA project is a 300-metre baseline interferometer located at Mitaka, near Tokyo. For all these projects, there was a long development programme to reach the sensitivities at which there is a good chance of detecting gravitational waves from celestial sources.

At the time of writing, all the gravitational wave observatories are entering their operational phases with more or less their design sensitivities. None of them have yet detected gravitational waves, but it will be no surprise if they are discovered in the next few years. The potential sources of detectable radiation include the collapse of stellar cores in supernova explosions, collisions and coalescences of neutron stars or black holes, rotations of neutron stars with deformed crusts, the continuous emission of very close binary neutron stars and black holes and primordial gravitational radiation created during the very earliest phases of our Universe.

1.11.3 Astroparticle physics

The term *astroparticle physics* is used to describe principally experiments to detect dark matter particles by laboratory experiments. The discipline has its roots in the realisation that our Galaxy possesses a dark matter halo and that it is unlikely to be made up of different forms of baryonic matter, such as low mass stars. It is entirely plausible that the dark matter consists of some form of particle as yet unrecognised in laboratory experiments. These dark matter particles might be the lightest supersymmetric partners of known types of particles, or some unknown type of massive neutrino.

Increasingly sensitive searches are being carried out in experiments such as the CDMS programme being carried out at the Soudan dark matter experiment. Thanks to an enormous and dedicated effort by many physicists, these experiments are now setting important limits to the cross-sections for the interaction of the dark matter particles with the material of the detectors.

1.12 Concluding remarks

The broad-brush *tour d'horizon* presented in this chapter summarises the enormous range of topics and disciplines involved in the study high energy astrophysical phenomena in our Universe. Over the succeeding chapters, we begin the long process of supporting the assertions of this chapter by a detailed analysis of the physical processes which need to be understood in order to put some coherence into this vast panorama. These are undoubtedly some of the most demanding and exciting areas of modern scientific endeavour.

2.1 Introduction

The theory of stellar structure and evolution is one of the most exact of the astrophysical sciences. It is inextricably involved in many of the topics needed to understand the role which high energy astrophysical processes play in the origin and evolution of stars and galaxies, providing, for example, evidence on their chemical abundances, the ages of the systems, and so on. The objective of this chapter is to provide a succinct summary of a number of the key results needed in the subsequent development of the story. Many of the equations and concepts will recur in different guises in the course of the exposition. There are many excellent books on these vast topics, my personal favorites being the books by Tayler, Karttunen and his colleagues, and by Kippenhahn and Weigert (Tayler, 1994; Karttunen *et al.*, 2007; Kippenhahn and Weigert, 1990). The last volume is a classic and is particularly strong on the physics of the stars.

2.2 Basic observations

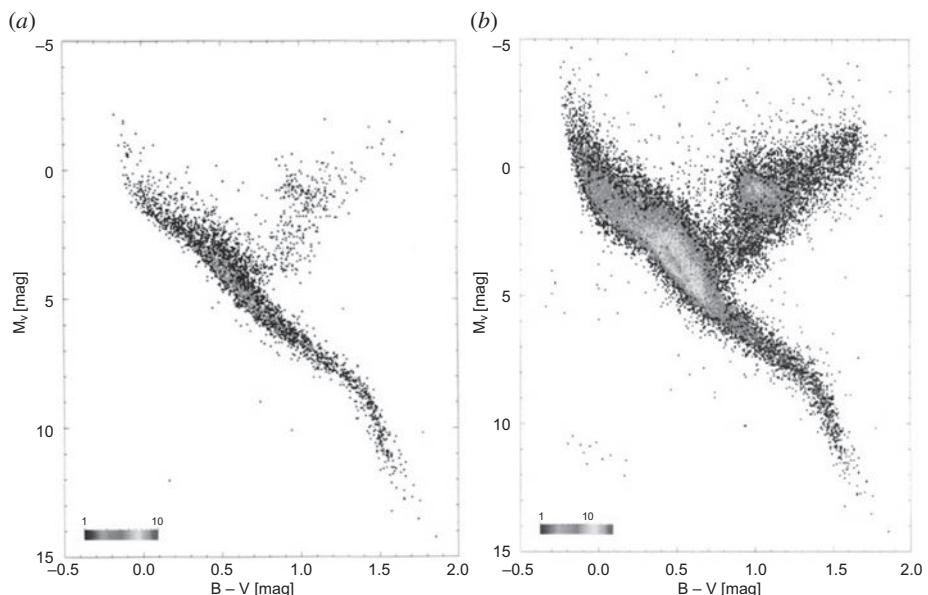
It is necessary to become familiar with some of the vocabulary of the study of the stars and the basic results of observation. These studies begin with measurements of the total amount of radiation emitted by a star, its luminosity L , and its surface temperature T . The spectra of stars are not black-bodies and so the *effective temperature* T_{eff} is introduced. It is defined to be the temperature of a black-body of the same radius as the star which would emit the same luminosity. Therefore, $L = 4\pi R^2 \sigma T_{\text{eff}}^4$, where σ is the Stefan–Boltzmann constant, $\sigma = 5.6705 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. For reference, values for the Sun are given in Table 2.1.

What makes the study of the structure and evolution of stars one of the most exact of the astrophysical sciences is the fact that, although a wide range of combinations of effective temperature and luminosity are found among the stars, most of them lie along certain well-defined loci or branches in the luminosity–temperature diagram (Fig. 2.1). As discussed in Appendix A, it is more convenient observationally to plot colour against luminosity.¹ Figure 2.1 is known as a *Hertzsprung–Russell*, or *H-R*, *diagram*, or, equivalently, as

¹ Summaries of astronomical measures of distance, mass, flux density, luminosity, apparent and absolute magnitude, colour, and so on, are given in Appendices A.1–A.4

Table 2.1 The properties of the Sun.

1 solar mass (M_{\odot})	= 1.989×10^{30} kg	$\approx 2 \times 10^{30}$ kg
1 solar radius (R_{\odot})	= 6.9598×10^8 m	$\approx 7 \times 10^8$ m
Luminosity of Sun (L_{\odot})	= 3.90×10^{26} W	$\approx 4 \times 10^{26}$ W
Effective temperature ($T_{\text{eff}\odot}$)	= 5780 K	≈ 5800 K
Absolute V magnitude M_V	= 4.83	
$B-V$ colour	= 0.63	

**Fig. 2.1**

The Hertzsprung–Russell or colour–magnitude diagram for nearby stars as determined by the *Hipparcos* astrometric satellite. (a) The H-R diagram for 4902 nearby stars for which distances are known to better than 5%. The abscissa is the $(B - V)$ colour of the star and the ordinate is the absolute magnitude in the V waveband. (b) The same diagram for 41 704 stars which have distances known to better than 20%. (From the Hipparcos and Tycho Catalogues, Vol. 1 (ed. M.A.C. Perryman), ESA SP-1200, 1997.)

a *colour–magnitude diagram*. The stars plotted in Fig. 2.1 constitute a random sample of stars in the solar neighbourhood in an apparent-magnitude limited sample. Most stars lie along a locus running from the bottom right to the top left of the H-R diagram and it is known as the *main sequence*. Notice the huge range of stellar luminosities compared with the range of temperatures, 20 absolute magnitudes corresponding to a range of 10^8 in luminosity. What distinguishes stars along the main sequence is their mass. The most massive stars lie at the top left of the main sequence and the lowest mass stars at the bottom right. For stars with masses in the range $1-10 M_{\odot}$, this relation can be written $L \propto M^{\alpha}$ where $\alpha \approx 3.5$. The exponent α is smaller for stars with masses greater than $10 M_{\odot}$ and also for stars less massive than the Sun. Our Sun lies about the middle of the sequence with $M_V = 4.83$ and $B - V = 0.63$.

Table 2.2 The Harvard spectral classification system.

Class	Class characteristics	Type	T_{eff}/K
O	Hot stars with He II absorption lines; strong ultraviolet continuum	O5	40 000
B	He I lines attain maximum strength; no He II lines; H developing later	B0	28 000
		B5	15 000
A	H lines attain maximum strength at A0, decreasing later; Ca II increasing	A0	9900
		A5	8500
F	Ca II stronger; Fe and other metal lines appear	F0	7400
		F5	6500
G	Ca II very strong; Fe and other metals strong; H weaker; solar type spectrum	G0	6030
		G5	5520
K	Neutral metallic lines dominate and CH CN bands developing; continuum weak in blue	K0	4900
		K5	4130
M	Very red; TiO ₂ bands developing strongly	M0	3480
		M5	2800
		M8	2400

Extending from about the location of the Sun towards the top right of the H-R diagram is the *giant branch*. Stars in this region of the diagram are much more luminous for a given colour compared with those on the main sequence and consequently, according to the Stefan–Boltzmann law, they must have very much larger radii. There is also a small cluster of stars lying to the bottom left of the H-R diagram below the main sequence. These are hot, blue, compact stars known as *white dwarfs*.

The spectra of the stars provide detailed information about their surface properties. In a remarkable pioneering analysis, Cannon and her colleagues at the Harvard Observatory ordered the spectra of stars into a continuous sequence on the basis of the presence or absence of different absorption lines in their spectra. The Harvard spectral sequence turned out to be a temperature sequence. The spectral types are still known by the designations used by the Harvard team and the names, properties and typical temperatures of the spectral types are summarised in Table 2.2. Finer subdivision can be made within each OBAFGKM class, the numbers 0 to 9 being included after each letter. Examples of modern spectra for different stellar types for main sequence stars are shown in Fig. 2.2 (Silva and Cornell, 1992). It is clear that the hot, blue O and B stars have spectra which peak in the ultraviolet waveband, while the cool red K stars have maxima towards the red end of the spectrum.

Other spectral features of the stellar spectra turned out to be sensitive to the luminosity of the star and so approximate luminosities can be estimated from these. The location of the different luminosity classes in the H-R diagram are indicated schematically in Fig. 2.3. This extension of the Harvard sequence is known as the Yerkes or MKK system and the names of the luminosity classes are listed in the figure caption of Fig. 2.3. Our Sun is a G2V star.

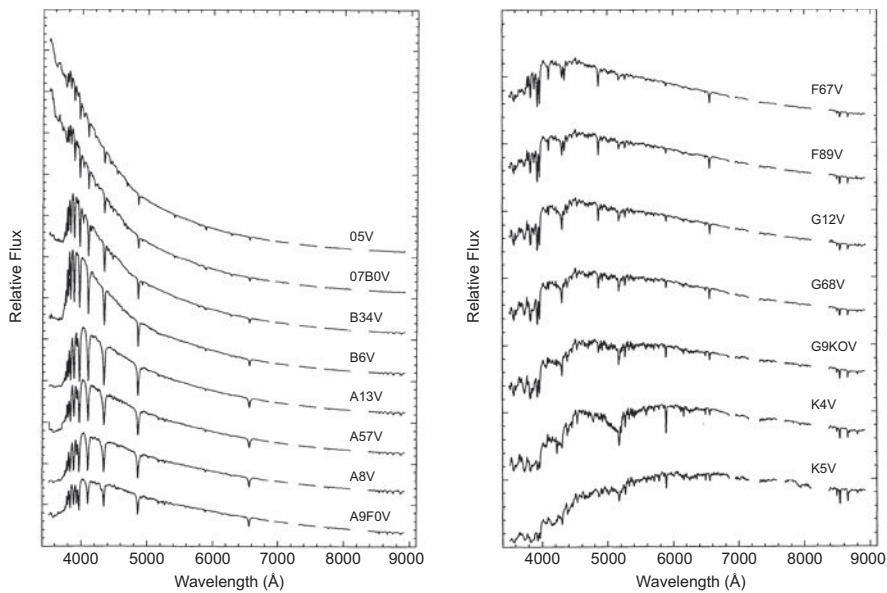


Fig. 2.2 Illustrating the spectra of different spectral types of main sequence stars from O to K (Silva and Cornell, 1992).

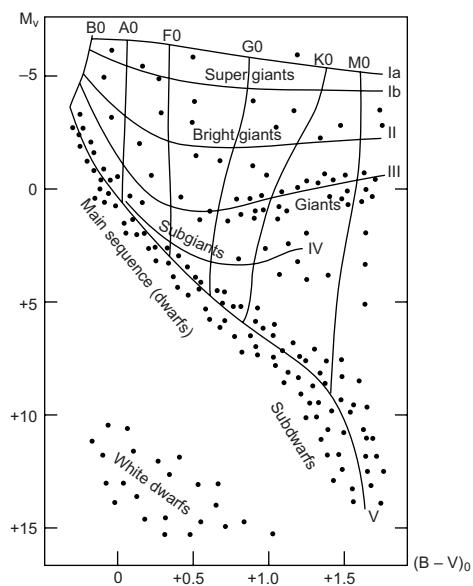
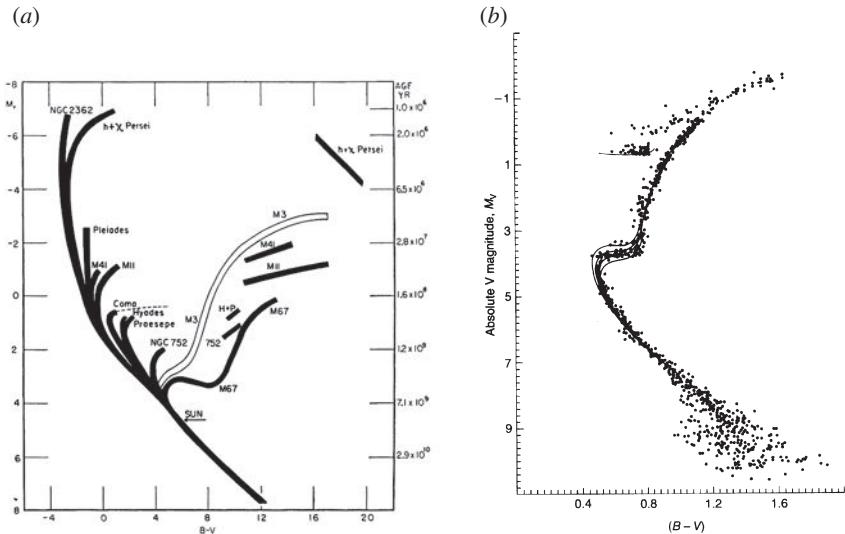


Fig. 2.3 Illustrating the loci of the different luminosity classes on the H-R diagram. The different luminosity classes are named as follows: I Supergiant stars, II Bright giants, III Giants, IV Sub-giants, V Main sequence, VI Subdwarfs, VII White dwarfs (after Schneider 2006).



(a) The H-R diagrams for star clusters of different ages. The youngest cluster is NGC 2362 and the oldest M67. The open line is for the globular cluster M3. The age scale on the right-hand vertical axis is in years (Sandage, 1957).
 (b) The Hertzsprung–Russell diagram for the old globular cluster 47 Tucanae. Note the appearance of the horizontal branch at absolute magnitude $M_V \approx 0.5$. The solid lines show the best fits to the data using theoretical models of the evolution of stars from the main sequence onto the giant branch due to Vanden Berg. The best-fit isochrones have ages in the range $1.2\text{--}1.4 \times 10^{10}$ years and the cluster is metal-rich relative to the other globular clusters, the metal abundance corresponding to about 20% of the solar value (Hesser *et al.*, 1987).

Clusters of stars are of special importance in understanding the evolution of the stars since it can be assumed that all the stars in a particular cluster have the same age. Therefore, the differences between the colour–magnitude diagrams are mostly due to the different ages of the clusters and the chemical compositions of the stars in the clusters. Examples of the H-R diagrams for a number of clusters of different ages are shown in Fig. 2.4a. A rough age scale for the main sequence termination point, which will be discussed below, is included on the right-hand vertical axis. The location of the Sun on the main sequence is indicated. An example of the H-R diagram for the old globular cluster 47 Tucanae (47 Tuc) is shown in Fig. 2.4b. There is a well developed giant branch and also a *horizontal branch* at absolute magnitude $M_V \approx 0.5$. The horizontal branch stars result from mass-loss processes during evolution on the giant branch.

2.3 Stellar structure

Stars are objects in which the force of gravity is balanced by the pressure gradient of the hot gas within the star. In all stable stars, this hydrostatic equilibrium is very precisely maintained, the source of energy to maintain the pressure gradient for stars on the main

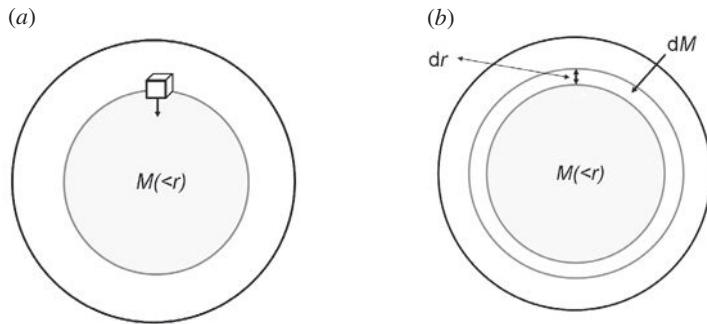


Fig. 2.5 (a) Illustrating the origin of the equation of hydrostatic support. (b) Illustrating the equation of conservation of mass.

sequence, the giant and horizontal branches being nuclear energy generation occurring in their centres. For stars like the Sun, the most common element is hydrogen and the next most abundant helium-4 (${}^4\text{He}$) with a cosmic abundance of about 24% by mass. The abundance of all the heavier elements, including species such as carbon, nitrogen, oxygen and iron, amount to only about 1–2% by mass of that of hydrogen – these are commonly referred to as the *metals*. In the centres of main sequence stars, the temperature is sufficiently high for hydrogen to be converted into helium, releasing in the process about 0.7% of the rest mass energy of the hydrogen, corresponding to the nuclear binding energy of helium.

Let us develop the equations of stellar structure which will be used in a variety of different contexts in the course of this study. To do this, we need the *four differential equations of stellar structure* as well as information about the *equation of state* of the stellar material. It is assumed that the stars evolve very slowly and so they can be taken to be quasi-static. In addition, we assume the stars are spherically symmetric, that is, there is no rotation and magnetic fields are unimportant. The equations are: (i) the equation of hydrostatic support, (ii) the law of conservation of mass, (iii) the equation of energy generation, and (iv) the equation of radiative transport.

2.3.1 The equations of hydrostatic support and mass conservation

Consider the forces acting on a little cube at radius r within the star (Fig. 2.5a). If its surface area is dA and thickness dr , the inward force of gravity is

$$F_{\text{gr}} = \frac{GmM(< r)}{r^2} = \frac{GM(< r)\varrho(r)dA dr}{r^2}. \quad (2.1)$$

This is resisted by the pressure forces acting on either side of the cube. In the plane-parallel approximation, the net outward pressure force is

$$F_p = dA[p(r) - p(r + dr)] = -dA dr \frac{dp}{dr}. \quad (2.2)$$

Balancing the forces (2.1) and (2.2),

$$\frac{dp}{dr} = -\frac{GM(< r)\varrho(r)}{r^2}. \quad (2.3)$$

This is the *equation of hydrostatic support*.

The mass between radii r and $r + dr$ is

$$M(r + dr) - M(r) = dM = 4\pi r^2 \varrho(r) dr , \quad (2.4)$$

and hence

$$\frac{dM}{dr} = 4\pi r^2 \varrho(r) . \quad (2.5)$$

This is the *equation of mass conservation*.

It is convenient to rewrite these equations with the mass $M = M(< r)$ as a variable in the radial direction. The first two equations of stellar structure then become

$$\frac{dp}{dr} = -\frac{GM\varrho}{r^2} ; \quad \frac{dM}{dr} = 4\pi r^2 \varrho . \quad (2.6)$$

We can already do useful things with these equations. Suppose there were no pressure support for the Sun. How long would it take to collapse to a very small size? In the absence of pressure support, the dynamics of the little cube would be

$$F_{\text{gr}} = m \frac{dv}{dt} = \frac{GmM_\odot}{r^2} \quad \text{or} \quad \frac{dv}{dt} = \frac{GM_\odot}{r^2} . \quad (2.7)$$

Integrating with respect to time,

$$\frac{1}{2} [v^2]_0^r = \left[\frac{GM_\odot}{r} \right]_{r_\odot}^r . \quad (2.8)$$

This is just the law of conservation of energy in a gravitational field. We can now estimate the infall speed when the Sun has reached half its present size, $v_{1/2} = (2GM_\odot/r_\odot)^{1/2}$. The collapse time is therefore roughly

$$t_c \sim \frac{r_\odot}{v_{1/2}} = \left(\frac{r_\odot^3}{2GM_\odot} \right)^{1/2} . \quad (2.9)$$

For the Sun, $t_c \sim 20$ minutes. This time-scale is often referred to as the *dynamical time-scale* for the star. It also represents the time it would take gravity to re-establish the quasi-static equilibrium status of the star.

Let us divide the two equations (2.6) by one another.

$$\frac{dp}{dM} = -\frac{GM}{4\pi r^4} . \quad (2.10)$$

Now integrate from the centre to the surface of the star.

$$-\int_0^{M_\odot} \frac{dp}{dM} dM = p_c - p_s = \int_0^{M_\odot} \frac{GM}{4\pi r^4} dM , \quad (2.11)$$

where the suffices c and s refer to the centre and surface of the star. We *underestimate* the value of the last integral if we set $r = r_\odot$ and so, setting $p_s = 0$, we find

$$p_c > \frac{GM_\odot^2}{8\pi r_\odot^4} = 4.5 \times 10^{13} \text{ N m}^{-2} = 4.5 \times 10^8 \text{ atmospheres} . \quad (2.12)$$

Thus, the gas in the centre of the Sun is at an extremely high pressure.

2.3.2 The virial theorem for stars

Next we can derive the *virial theorem* for stars – this is one of the key results of stellar astrophysics. Setting $V = (4\pi/3)r^3$, we reorganise equation (2.10) and integrate from the centre (c) to the surface (s) of the star:

$$4\pi r^3 dp = 3V dp = - \left(\frac{GM}{r} \right) dM; \quad \int_{p_c}^{p_s} 3V dp = - \int_0^{M_s} \left(\frac{GM}{r} \right) dM = \Omega. \quad (2.13)$$

The quantity Ω on the right-hand side of the second equation of (2.13) is the total gravitational potential energy of the star, noting that Ω is a *negative quantity*. Integrating the left-hand side by parts, we find

$$3 \int_{p_c}^{p_s} p dV + \Omega = 0. \quad (2.14)$$

Finally, we write dV in terms of the corresponding mass element dM , $dM = \rho dV$,

$$3 \int_0^{M_s} \left(\frac{p}{\rho} \right) dM + \Omega = 0, \quad (2.15)$$

where ρ is the density of the stellar material. This is the *virial theorem* for stars. Many important general results can be derived from the virial theorem.

Let us first work out the minimum temperature in the centre of the Sun. We obtain a lower bound to the gravitational potential energy $-\Omega$ if we set $r = r_\odot$

$$-\Omega = \int_0^{M_\odot} \left(\frac{GM}{r} \right) dM > \int_0^{M_\odot} \frac{GM dM}{r_\odot} = \frac{GM_\odot^2}{2r_\odot}. \quad (2.16)$$

If we assume the material of the Sun is a perfect gas, $p = \rho kT/m$, where m is the mean molecular weight of the particles. Therefore, the integral in (2.15) becomes

$$3 \int_0^{M_\odot} \left(\frac{p}{\rho} \right) dM = \frac{3k}{m} \int T dM = \frac{3k\bar{T} M_\odot}{m}, \quad (2.17)$$

where \bar{T} is the mass-weighted average temperature of the Sun. Finally, we use the inequality of (2.16) combined with the equalities (2.15) and (2.17) to write

$$-\Omega > \frac{GM_\odot^2}{2r_\odot}; \quad \bar{T} > \frac{GM_\odot m}{6kr_\odot}. \quad (2.18)$$

If the material of the Sun is assumed to be fully ionised hydrogen, its mean molecular weight is $m = (m_p + m_e)/2 \approx m_p/2$. Therefore, the minimum temperature is

$$\bar{T} > \frac{GM_\odot m_p}{12kr_\odot} = 2 \times 10^6 \text{ K}. \quad (2.19)$$

Thus, the central regions of the Sun must be very hot. Notice that this temperature is very much greater than that corresponding to the ionisation potential of hydrogen, $T = I_H/k = 1.6 \times 10^5 \text{ K}$, where $I_H = 13.6 \text{ eV}$ and so the gas is certainly very highly ionised.

We can now write the virial theorem in terms of the internal thermal energy per unit mass u . If γ is the ratio of specific heats and n_f the number of degrees of freedom,

$\gamma = (n_f + 2)/n_f$ and the internal energy density is

$$\text{internal energy density} = n_f \times \frac{1}{2}kT \times n = \frac{n k T}{(\gamma - 1)} = \frac{p}{(\gamma - 1)}, \quad (2.20)$$

where n is the number density of particles. Hence, the internal energy per unit mass is $u = p/(\gamma - 1)\varrho$. Therefore, the integral in (2.16) becomes

$$3 \int_0^{M_s} \frac{p}{\varrho} dM = 3 \int_0^{M_s} (\gamma - 1)u dM = 3(\gamma - 1)U, \quad (2.21)$$

where U is the total internal thermal energy of the star. For a monatomic gas, such as a fully ionised gas, $\gamma = 5/3$ and so

$$2U + \Omega = 0. \quad (2.22)$$

Thus, the magnitude of the gravitational potential energy is twice the internal thermal energy of the star.

The *Kelvin–Helmholz* or *thermal time-scale for stars* can be derived from the virial theorem. The magnitude of the gravitational potential energy is twice the internal thermal energy of the star. Therefore, we can work out how long it would take the Sun to radiate away all its internal thermal energy:

$$t_{\text{KH}} = \frac{U}{L_\odot} \sim \frac{GM_\odot^2}{r_\odot L} = 3 \times 10^7 \text{ years}, \quad (2.23)$$

where KH stands for Kelvin–Helmholtz, after two of the pioneers who first carried out this calculation. The *Kelvin–Helmholtz time-scale* is often referred to as the *thermal time-scale* of the star. Since the Earth is 4.6×10^9 years old, there must be an internal energy source in the Sun to keep it shining.

The *thermal paradox for stars* is the statement that, as stars radiate away their thermal energy, they heat up. The reason is that the total energy of the star is the sum of its thermal and gravitational potential energies, $E = U + \Omega$. But the virial theorem tells us that $U = -\Omega/2$ and so the total energy is

$$E = \frac{\Omega}{2} = -U, \quad (2.24)$$

a negative quantity. Thus, as the star loses energy, the total energy becomes more negative and so U must increase, in other words, the star becomes hotter. This non-intuitive result is entirely associated with the fact that the gravitational potential energy is a negative quantity.

2.4 The equations of energy generation and energy transport

The *third equation of stellar structure* describes the energy generation rate within the star. The energy generated within the star diffuses outwards and so the contribution to the outflow of energy from the shell of radius r and thickness dr is

$$dL = 4\pi r^2 \varrho \varepsilon dr, \quad (2.25)$$

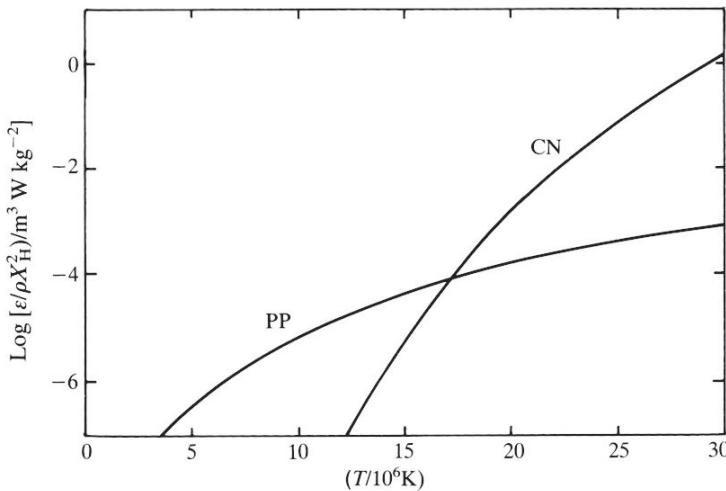


Fig. 2.6

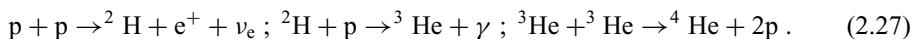
The overall nuclear energy generations for the p-p chain and the CNO cycle as a function of temperature (Tayler, 1994).

where ε is the energy generation rate per unit mass and is a function of the local temperature and density conditions. Notice that L is the rate of flow of energy, or the power, passing through the spherical surface at radius r . Hence the differential equation for L is

$$\frac{dL}{dr} = 4\pi r^2 \rho \varepsilon . \quad (2.26)$$

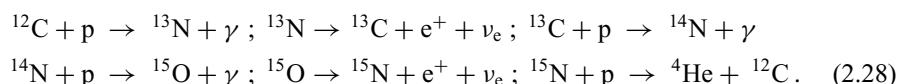
For main sequence stars, the source of energy is the nuclear conversion of hydrogen into helium and is a strong function of temperature. If the central temperature of the star is less than about 1.7×10^7 K, the *proton–proton (p-p) chain reaction* is the primary energy source for the star; if the temperature is greater than this value, the reaction cycle known as the *carbon–nitrogen–oxygen (CNO) cycle* is the dominant process (Fig. 2.6).

The principal reactions of the p-p chain involve the following nuclear processes:



The energy generation rate for the p-p chain can be described by $\varepsilon \propto \rho T^4$. The first interaction in the chain is a weak interaction which involves the formation of deuterium from two protons. The detection of the electron neutrinos produced in this reaction is a key test of the theory. Other important side-chains will be discussed later.

In the CNO cycle, helium is formed by the successive addition of protons to heavier nuclei which, when they become too massive for nuclear stability, decay by ejecting an α -particle and so create helium. Carbon acts as a catalyst for the formation of helium through the successive addition of protons accompanied by two β^+ decays in the second and fifth interactions in the cycle:



The energy generation rate for the CNO cycle can be described by $\varepsilon \propto \rho T^{17}$ and is the dominant process at high temperatures, $T > 1.7 \times 10^7$ K. The internal structure of the star depends crucially upon which of these processes is dominant.

The energy generation equations do not tell us the rate at which the energy passes through the sphere of radius r . For this, we need the equation of radiative transfer, the *fourth equation of stellar structure* which describes how energy is transported through the star. There are two principal mechanisms of energy transport, *radiation* and *convection*. If the temperature gradient in the star exceeds the adiabatic gradient, that is, it is superadiabatic, convective motions stabilise the energy transport so that the variation of temperature with pressure, or density, is limited to the adiabatic gradient. Specifically, the condition is

$$\frac{d \ln T}{d \ln p} \geq \frac{\gamma - 1}{\gamma}, \quad (2.29)$$

where γ is the ratio of specific heats of the material of the star. In practice, what is done is to work out the structure of the star and then test whether or not there are superadiabatic regions in which convective transport of energy takes place.

Radiative transport of energy is much more important than thermal conduction because the mean free path for photons, although small, is still very much greater than the mean free path for electrons and the photons diffuse at the speed of light. The standard form of the heat diffusion equation is $F = -\lambda dT/dr$, where F is the power per unit area parallel to the direction of the temperature gradient and λ is the heat diffusion coefficient. Therefore, the total rate of flow of energy through the spherical surface at radius r is $L = 4\pi r^2 F$.

In the radiative transport of energy within stars, the radiation is scattered many times, because of the very high density of the material and the large cross-section for scattering. Because of the very large numbers of scatterings, the radiation at any point inside the star is almost precisely isotropic and has a black-body spectrum at the local temperature of the material of the star. The diffusion of energy takes place through the very gradual decrease in temperature with increasing radius.

In astrophysical applications, the quantity known as the *opacity* κ of the stellar material is used rather than the heat diffusion coefficient. κ is defined as the fraction of the flux density of radiation which is absorbed or scattered per unit mass per unit path length. If the increment of flux density dF is intercepted by the material of the star on traversing a distance dr , κ is defined by

$$dF = -\kappa \rho F dr. \quad (2.30)$$

The spectrum of the radiation inside the stars is very close to a black-body spectrum at the local temperature and so the equation of radiative transfer can be written in a form which is directly related to local physical conditions in the star. The flux density decrease corresponds to a decrease in radiation pressure with radius through the star. The energy loss per second from the increment of path length dr is $-\kappa \rho F dr$ and hence the corresponding change in momentum per unit area per unit time, that is, the change of radiation pressure, is

$$dp = -\frac{\kappa \rho F}{c} dr. \quad (2.31)$$

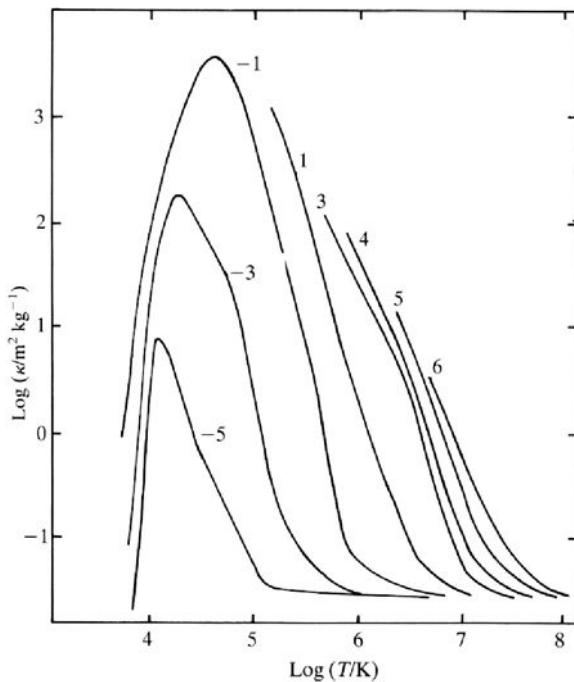


Fig. 2.7

The opacity of matter with the chemical composition of the Sun for different temperatures and densities (Tayler, 1994). The solid lines on the diagram show the opacity for different densities of the stellar material in units of $\log(\text{kg m}^{-3})$.

The radiation is locally black-body radiation at temperature T and so, according to the Stefan–Boltzmann law, $p = \frac{1}{3}aT^4$. Therefore,

$$\frac{dp}{dT} = \frac{dp}{dr} \frac{dr}{dT} = \frac{4}{3}aT^3. \quad (2.32)$$

But, from (2.31), we have derived an expression for dp/dr which involves the flux density of radiation F . Therefore,

$$F = -\frac{4}{3} \frac{acT^3}{\kappa\rho} \frac{dT}{dr}, \quad (2.33)$$

or, in terms of the luminosity passing through the sphere at radius r ,

$$L = -\frac{16\pi acr^2 T^3}{3\kappa\rho} \frac{dT}{dr}. \quad (2.34)$$

This is the *fourth equation of stellar structure*.

The opacity κ is a complex function of temperature and density because of the large number of processes which contribute to the absorption and re-emission of photons at different temperatures (Fig. 2.7). At the very highest temperatures, the plasma is fully ionised and the dominant scattering process is *Thomson scattering* for which the Thomson cross-section $\sigma_T = e^4/6\pi\epsilon_0^2 m_e^2 c^4 = 6.653 \times 10^{-29} \text{ m}^{-2}$ is independent of frequency. In the intermediate temperature range, the dominant processes are *free–free* or *bremsstrahlung*

Table 2.3 Approximate values of the quantities β and γ in the expression $\kappa \propto \varrho^\beta T^\gamma$ for the opacity of stellar material (Tayler, 1994).

Temperature	Temperature range (K)	Physical processes	β	γ
Low	$10^4\text{--}10^{4.5}$	Atomic and molecular absorption	0.5	4
Medium	$10^{4.5}\text{--}10^7$	Bound-free and free-free absorption	1	-3.5
High	$> 10^7$	Electron scattering	0	0

absorption and *bound-free absorption*. Summing over all the contributions at the different frequencies to the average opacity κ , the appropriate weighting is given by

$$\frac{1}{\kappa} = \frac{\pi}{acT^3} \int_0^\infty \frac{1}{\kappa_\nu} \frac{\partial B}{\partial T} d\nu , \quad (2.35)$$

which is known as the *Rosseland mean opacity*. The dependence of κ upon the temperature T and density ϱ of the plasma in the intermediate temperature range is therefore $\kappa \propto \varrho T^{-7/2}$. It is convenient to approximate the dependence of κ on density and temperature in different temperature ranges by power-law relations of the form $\kappa \propto \varrho^\beta T^\gamma$. The values quoted by Tayler are shown in Table 2.3.

2.5 The equations of stellar structure

The four equations of stellar structure are therefore:

$$\frac{dp}{dr} = -\frac{GM\varrho}{r^2} , \quad \text{hydrostatic equilibrium ,} \quad (2.36)$$

$$\frac{dM}{dr} = 4\pi r^2 \varrho , \quad \text{conservation of mass ,} \quad (2.37)$$

$$\frac{dL}{dr} = 4\pi r^2 \varrho \varepsilon , \quad \text{energy generation ,} \quad (2.38)$$

$$\frac{dT}{dr} = -\frac{3\kappa\varrho}{16\pi acr^2 T^3} L , \quad \text{energy transport .} \quad (2.39)$$

To create models of quasi-static stars, the equations need to be supplemented by the equation of state of the stellar material under different conditions of density and temperature and appropriate boundary conditions need to be satisfied at the surface of the star. Account needs to be taken of those regions of the star which are in convective rather than radiative equilibrium. Such stellar models have been the subject of a great deal of computer modelling since the 1960s when digital computers first became available to theoretical astrophysicists – these are now essential tools for studies of the astrophysics of the stars.

Some insight into the physics of stellar interiors can be derived from simplified stellar models, in particular, from the study of *homologous stellar models*. In these, it is assumed that the material of the star has the same composition at all radii and that the same properties of energy generation and transport apply throughout the star. Using the

power-law approximations for the dependence of the energy generation rate and opacity upon density and temperature given in the last section and adopting the equation of state of a perfect gas, $p = nkT$, the equations of stellar structure can be written so that the variation of quantities such as pressure, temperature and luminosity with radius follow relations which scale as different powers of the mass of the star. Tayler provides an excellent discussion of the procedures involved (Tayler, 1994).

A consequence of these simplified models is that they result in power-law relations for the dependence of different properties of the star upon mass.² For example, for stars like the Sun for which the p-p chain is the source of energy and the opacity can be described by $\beta = -3.5$ and $\gamma = 1$, we find

$$R \propto M^{1/13}; \quad L \propto M^{71/13} = M^{5.5}; \quad L \propto M^{71/13} \propto T_{\text{eff}}^{284/69} = T_{\text{eff}}^{4.1}, \quad (2.40)$$

where we have introduced the *effective temperature* T_{eff} defined by the relation $L = 4\pi R^2 a T_{\text{eff}}^4$. Similar calculations can be carried out for other combinations of expressions for the opacity and energy generation rates. For example, for very high mass stars, the CNO cycle is the more important energy generation process and the opacity is determined by Thomson scattering. Then,

$$R \propto M^{4/5}; \quad L \propto M^3; \quad L \propto T_{\text{eff}}^{60/7} = T_{\text{eff}}^{8.6}. \quad (2.41)$$

For a wide range of assumptions about the opacity of the stellar material and the energy generation rate, there is a power-law relation of the form $L \propto M^b$, where b lies in the range 3–5.5. In addition, there are very strong dependences of luminosity L upon the effective temperature T_{eff} in (2.40) and (2.41), which describe the *main sequence* in a theorist's luminosity–temperature diagram. As a result, the models can account for the huge range of luminosity associated with quite a narrow range of effective temperature.

In reality, the structure of the stars is much more complicated than that suggested by the homologous stellar models. We need to take account of the following factors:

- The assumption of homogeneity – inevitably stars become inhomogeneous as nuclear processes convert hydrogen into helium in their cores.
- The dependence of the properties of stars upon their chemical compositions.
- The effects of convection.
- The effects of radiation pressure.
- The detailed physics of nuclear reaction rates and stellar opacity.
- Proper boundary conditions at the surfaces of the stars.

To do justice to these topics, we need computer models for the structure and evolution of the stars.

An instructive example of the evolution of the structure of a $1.3 M_{\odot}$ star from detailed computations carried out by Kippenhahn and Weigert is shown in Fig. 2.8 (Kippenhahn and Weigert, 1990). Most of its lifetime is spent as a main sequence star, steadily burning hydrogen to helium in its central core which grows with time as the fuel in the core is

² This can be demonstrated by order-of-magnitude methods which are included as Appendix A3 of Chapter 3 of my book *The Cosmic Century* (Longair, 2006).

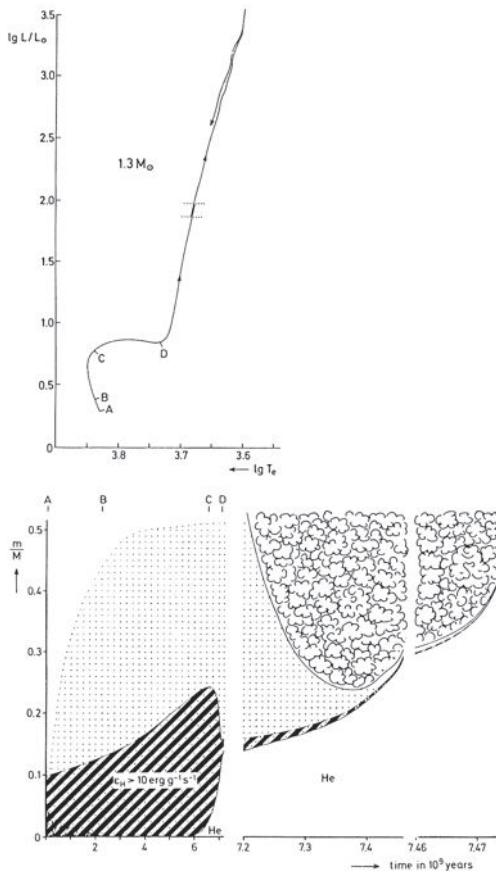


Fig. 2.8 The evolution of the internal structure of a $1.3 M_{\odot}$ star showing how it evolves from the main sequence to the giant branch. The scale on the ordinate is the fractional mass contained within a given radius. The letters A, B, C and D show the structure of the star and its corresponding location on the H-R diagram. Notice the changing time-scale along the abscissa which shows that the star spends most of its lifetime close to the main sequence. The main region of hydrogen burning is indicated by the hatched areas, while the ‘cloudy’ areas indicate regions in which convective energy transport takes place. The diagram illustrates the formation of the extensive outer convective zone as the star evolves up the giant branch (Kippenhahn and Weigert, 1990).

consumed. Once the star has settled onto the main sequence, its luminosity and effective temperature change very little until it moves off the main sequence when the core begins to contract and the red giant envelope expands. When the nuclear fuel in the central region is exhausted, an isothermal helium core is formed and hydrogen burning continues in a shell about it. Schönberg and Chandrasekhar showed that there do not exist stable stellar models in which the inert stellar core contains more than about 10% of the mass of the star (Schönberg and Chandrasekhar, 1942). The pressure at the base of the hydrogen-burning shell becomes too great and causes the inner regions to collapse. The key quantity is the ratio of the mean molecular weights μ in the core and the envelope – the fraction of the mass of the star in the core should not exceed $(\mu_{\text{core}}/\mu_{\text{env}})^2$, where the μ s are mean molecular

weights per electron. For a helium core surrounded by an envelope with normal cosmic abundances, the limit corresponds to about 10% of the mass of the star being in the core.

These considerations enable a simple estimate of the main-sequence lifetime of the Sun and stars to be made. The energy released in converting hydrogen into helium by either the p-p chain or the CNO cycle can be estimated from the mass deficit found by comparing the masses of hydrogen and helium nuclei. The fraction of the rest mass of the ingredients released in the nuclear interaction $4p \rightarrow {}^4He$ is

$$\frac{4m_p - m_{He}}{4m_p} = 0.007. \quad (2.42)$$

Since $m_p c^2 = 1$ GeV, roughly 7 MeV is liberated per hydrogen nucleus which is combined into helium-4. Stars move off the main sequence when the central 10% of their mass has been converted into helium and so the total energy released in this process is $E = 0.007(0.1 \times M)c^2$. Since the luminosity of the star is L , its main-sequence lifetime is

$$T_{MS} = \frac{E}{L} = \frac{0.007(0.1 \times M)c^2}{L}.$$

Inserting the values for the Sun, we find $T_\odot = 10^{10}$ years.

We can use this result to find the lifetimes of main sequence stars of different masses. If the mass-luminosity relation has the form $L \propto M^x$, where $x \sim 3.5$ for stars with $M \sim M_\odot$, then, by exactly the same argument, the lifetime of the star is

$$T(M) = 10^{10} \left(\frac{M}{M_\odot} \right)^{-(x-1)} \text{ years}. \quad (2.43)$$

2.6 The Sun as a star

Detailed computations indicate that the central temperature of the Sun is about 1.5×10^7 K and the region within which the p-p nuclear chain reactions take place occupies roughly the central 10% of the Sun by radius. Within the central 70% of the Sun by radius, energy is transported outwards by radiative diffusion. In the outer 30% of the Sun, which only contains a small fraction of the mass of the star, energy transfer is by convection and these convective motions are responsible for the remarkable forms activity observed on the Sun's surface (see Sect. 2.7.1).

Granted the outline of stellar structure discussed in Sect. 2.5, how well can the theory account for the observed properties of the Sun? Two important developments over the last 30 years have enabled the physics of the solar interior to be studied in remarkable detail. These are the measurement of the modes of oscillation of the Sun, the discipline known as *solar seismology* or *helioseismology*, and the detection of neutrinos released in the nuclear reactions taking place in the centre of the Sun. These are crucial topics for studies of stellar structure and evolution.

2.6.1 Helioseismology and the internal structure of the Sun

It is simplest to think of the Sun as a resonant sphere which, when perturbed, vibrates at frequencies corresponding to its normal modes of oscillation. The convective envelope of the Sun provides a natural source of excitation which can stimulate the Sun to resonate in these modes. In terrestrial seismology, the resonance modes of the Earth can be found by tracing the paths of sound waves inside the Earth and exactly the same procedure can be employed to study physical conditions inside the Sun. These studies are therefore referred to as *solar* or *helioseismology*.

There are two principal methods for measuring the solar oscillations, both of which are technically very challenging. In one approach, the brightness of the Sun is measured with very high precision so that variations as small as one part in 10^6 of the total intensity can be measured. In the other approach, very precise measurements of the Doppler shifts of the solar atmosphere are made—the techniques must be precise enough to measure velocity differences of about 1 m s^{-1} or less. Both approaches have now been successfully used to measure the resonant modes of the Sun, those which penetrate into its core being of particular interest for the study of physical conditions in the nuclear burning regions.

The theory of the modes of oscillation of the Sun is a beautiful example of the power of classical theoretical physics applied to an astrophysical problem, much of the pioneering analysis being contained in Lamb's classical text *Hydrodynamics* of 1932 (Lamb, 1932). The modes of oscillation of the Sun can be thought of as standing waves resulting from the interference of oppositely directed propagating waves. In the simplest approximation, the Sun can be considered to be spherically symmetric and so the natural representation of the perturbations is in terms of associated Legendre functions, similar to those used to describe the amplitudes of the wavefunctions of the hydrogen atom (Fig. 2.9b). Following Deubner and Gough, if ξ is the vertical component of the fluid displacement, the decomposition into normal modes can be written

$$\xi(r, \theta, \phi, t) = \Re \left[R(r) P_l^m(\cos \theta) \frac{\cos m\phi}{\sin} e^{i\omega t} \right] \quad (2.44)$$

where the separation of variables consists of the associated Legendre function $P_l^m(\cos \theta)$ describing the angular variation of the amplitude of the displacement and $R(r)$ are radial eigenfunctions. \Re indicates that the real part of the function should be taken (Deubner and Gough, 1984). The adopted terminology for the Sun is that l is called the *degree*, n the *order* and m the *azimuthal order* of a particular mode. The different wave modes probe to different depths in the Sun. For example, in Fig. 2.9a, the rays correspond to modes with frequency $3000 \mu\text{Hz}$ and in order of decreasing depth of penetration their degrees l are 0 (the straight ray passing through the centre), 2, 20, 25 and 75. These observations enable the speed of sound c_s to be determined throughout the Sun, where $c_s = (\gamma p / Q)^{1/2} \propto T^{1/2}$.

The modes of oscillation consist of two types, acoustic or *p*-modes, in which the restoring force is provided by pressure fluctuations, and gravity or *g*-modes, for which the restoring force is buoyancy. The modes of greatest interest for the study of the central regions of the Sun are the acoustic modes of small degree l since they probe into its central regions (Fig. 2.9a). For a mode of given degree, there are many different orders n which measure

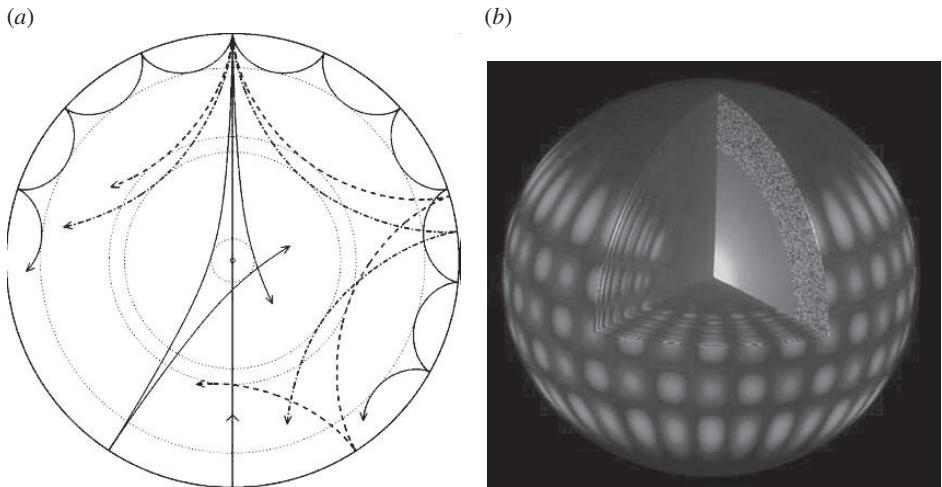


Fig. 2.9

(a) Propagation of sound waves through a cross-section of a solar model. The paths of rays are bent by the increase with depth of the sound speed until they reach the inner turning point indicated by the dotted circles, at which the waves undergo total internal reflection. At the surface, the waves are reflected because of the rapid decrease in density (Christensen-Dalsgaard, 2002). (b) A schematic diagram illustrating one of the normal modes of oscillation of the Sun.

the vertical component of the wavenumber. As in the hydrogen atom, n is related to the number of nodes in the solutions of the radial wave equation. Figure 2.9b shows a pictorial representation of a normal mode of oscillation of the Sun.

An example of the power spectrum of solar oscillations from the GOLF experiment of ESA's Solar and Heliospheric Observatory (SOHO) is shown in Fig. 2.10. The power spectrum shows low degree p -modes and there are two types of separation of the resonant frequencies. The ‘large’ separations, corresponding to frequency differences $\Delta\nu_0$ of about 60 µHz, correspond to modes of the same degree l but of order n differing by 1. There are also ‘small’ differences δ_{nl} associated with alternate resonances and these are associated with the difference in frequency between modes with ‘quantum numbers’ (n, l) and those with $(n - 1, l + 2)$. The physical significance of $\Delta\nu_0$ is that it is associated with the average sound speed throughout the Sun. For low values of l , the modes are identical in the outer regions of the Sun but differ in the central regions. Thus, the values of δ_{nl} are sensitive to physical conditions in the core of the Sun.

The spectrum of solar oscillations obtained by experiments such as the SOHO observatory is very rich. It provides unique information about the speed of sound, which depends upon detailed knowledge of the equation of state of matter in bulk at temperatures between 10^4 and 1.5×10^7 K, as a function of radius in the solar interior, as well as about its internal rotational velocity field. The power spectrum of the oscillations can be inverted and compared with the predictions of the standard solar models. The results of analysis of the SOHO data are shown in Fig. 2.11 which shows that the square of the speed of sound has been determined to better than 0.2% throughout most of the Sun. The biggest discrepancy occurs at the turbulent boundary between the inner radiative and outer convective zones, shown by the prominent outer shaded band in Fig. 2.11b. This turbulent layer at the base

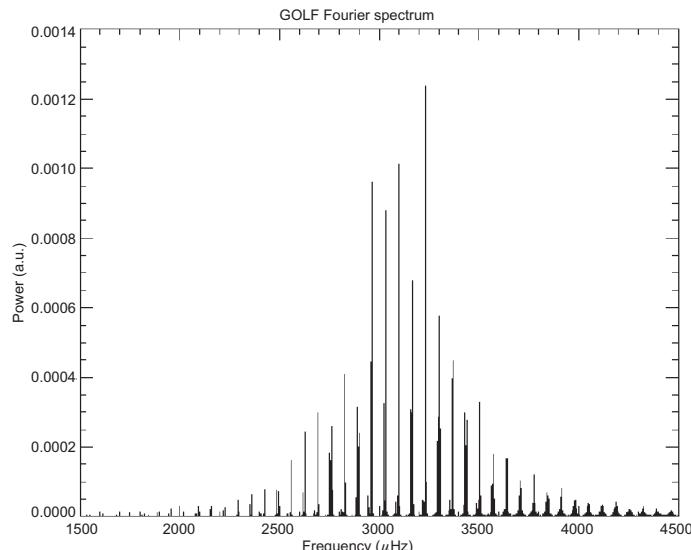


Fig. 2.10 The p -mode Fourier spectrum from the GOLF experiment of the ESA SOHO mission. These data are from a 690-day time series of calibrated velocity signal, which exhibits an excellent signal-to-noise ratio. In addition to the various l -modes, fine structure splittings of all the lines are present. (Courtesy of ESA and the SOHO science team.)

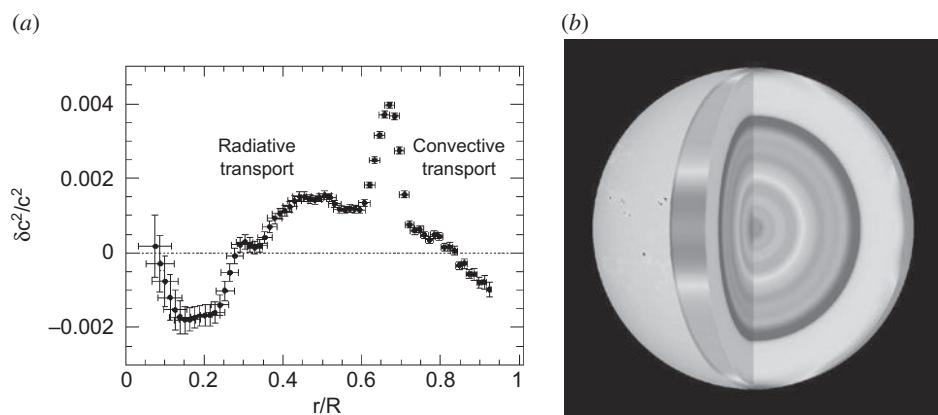
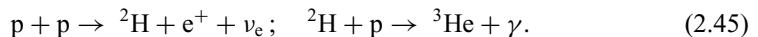


Fig. 2.11 (a) A comparison of the best-fitting standard model of the internal structure of the Sun and the results of observations of solar oscillations in terms of the fractional deviations of the square of the sound speed relative to that model. The agreement is better than 0.2% throughout most of the Sun. (b) A schematic diagram illustrating the same results shown in Fig. 2.11a in terms of the internal temperature of the Sun. The differently shaded bands indicate deviations from the standard solar model. The central temperature may be 0.1% cooler than the expected value of 15×10^6 K. (Courtesy of ESA and the SOHO science team.)

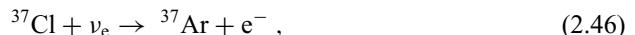
of the convective zone is believed to be the source of many of the features observed on the Sun's surface, including the dynamo which is responsible for maintaining the Sun's magnetic field.

2.6.2 Observations of solar neutrinos

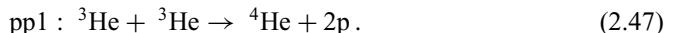
The p-p chain is the principal source of energy in the Sun, the first two reactions being:



The first reaction, in which deuterium is formed, is the principal source of solar neutrinos but they are of low energy, the maximum energy being 0.420 MeV. The reaction rate for this process has never been measured experimentally at the energies of interest for nucleosynthesis in the Sun and so the reaction rate is based upon theoretical estimates. It was originally hoped that these neutrinos could be detected by a chlorine detector. In what is essentially an inverse β decay process,



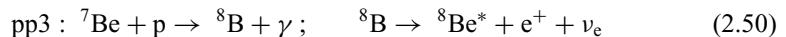
radioactive argon ${}^{37}\text{Ar}$ is created and the amount created can be measured from the number of radioactive decays of the argon nuclei. The threshold energy for the reaction is, however, 0.814 MeV, greater than the energy of the p-p chain neutrinos. There are three alternative routes which lead to the formation of helium-4 from helium-3. The most straightforward is the pp1 branch, which has already been discussed:



The other routes involve the formation of ${}^7\text{Be}$ as a first step



Then ${}^7\text{Be}$ can either interact with an electron (the pp2 branch) or a proton (the pp3 branch) to form two ${}^4\text{He}$ nuclei:



The pp1 chain is most important at low temperatures, $T < 10^7$ K, while the others are more important at higher temperatures. Notice that the pp2 and pp3 chains depend upon there being ${}^4\text{He}$ present to begin with but, since about 24% of the mass of baryonic matter in the Universe is expected to be in the form of ${}^4\text{He}$ as a result of primordial nucleosynthesis, there was already a considerable amount of helium present even in unprocessed stellar material. The electron neutrinos emitted in the decay of ${}^8\text{B}$ nuclei have maximum energy 14.06 MeV and so can be detected in a chlorine experiment.³

The famous *solar neutrino experiment* was carried out by Davis and his colleagues using a 100 000 gallon tank of perchloroethylene C_2Cl_4 located at the bottom of the Homestake

³ For many more details of these nuclear reactions and the experiments to detect solar neutrinos, *Neutrino Astrophysics* by J.N. Bahcall can be recommended (Bahcall, 1989).

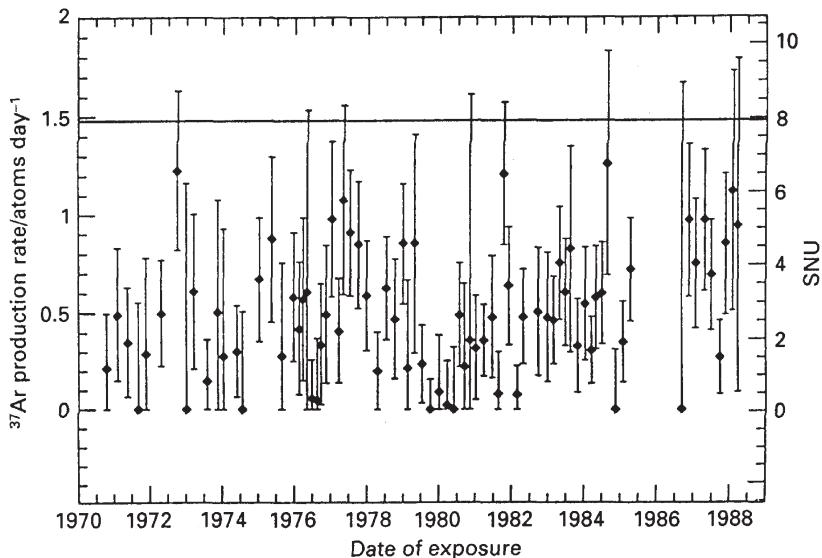


Fig. 2.12 The observed flux of solar neutrinos from the ^{37}Cl experiment carried out by Davis and his colleagues during the period 1970–88. The solid line at 8 SNU is the expectation of the standard solar model of Bahcall and Ulrich (Bahcall, 1989).

gold-mine in South Dakota. As the statistics improved over the years, a significant flux of neutrinos was detected but it corresponded to only about one-quarter of the flux predicted by the standard solar models (Fig. 2.12). This discrepancy is the famous *solar neutrino problem*. The results quoted by Bahcall in 1989 were:

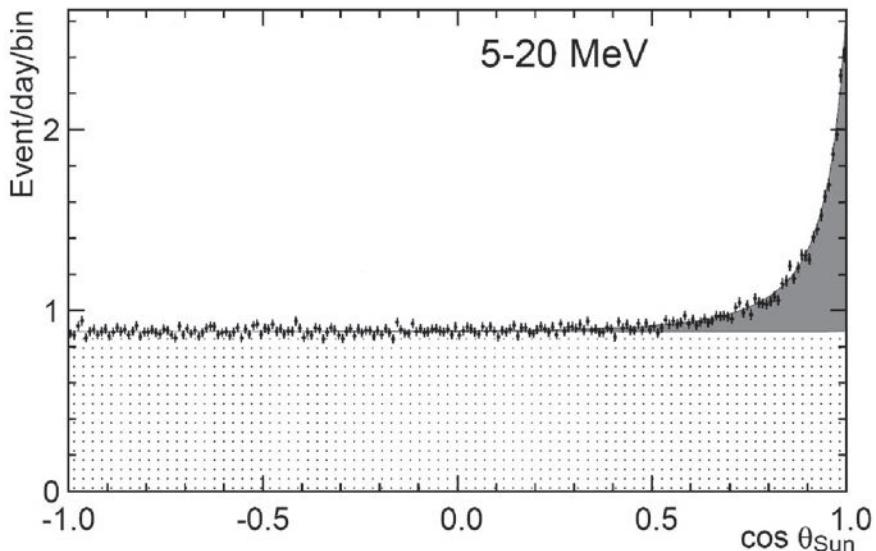
Observed flux of neutrinos:	2.1 ± 0.9 SNU
Predicted flux of neutrinos:	7.9 ± 2.6 SNU

where $1 \text{ SNU} = 1 \text{ Solar Neutrino Unit} = 10^{-36}$ absorptions per second per ^{37}Cl nucleus (Bahcall, 1989). The errors quoted are formal 3σ errors for both the observations and the predictions. The helioseismology observations were important in showing that this discrepancy cannot be due to uncertainties in the astrophysics of the internal structure of the Sun in the nuclear burning regions.

Confirmation that the flux of high energy neutrinos indeed originated within the Sun was provided by the Japanese Kamiokande II experiment (Hirata *et al.*, 1990). The high energy neutrinos scatter electrons which recoil with relativistic velocities. The Cherenkov detectors which lined the walls of the Kamiokande II experiment measured the direction of travel of the scattered electrons and thus the arrival directions of the neutrinos could be inferred. A significant excess flux of neutrinos coming from the direction of the Sun was discovered. The final results of the Kamiokande II experiment from 1036 days of observations from 1987 to 1995 were:

$$\text{Flux of neutrinos} = 2.56 \pm 0.16 \text{ (stat)} \pm 0.16 \text{ (syst)} , \quad (2.52)$$

where (stat) refers to the statistical errors and (syst) to the systematic errors.

**Fig. 2.13**

The angular distribution in $\cos \theta_{\text{Sun}}$ of solar neutrino event candidates from 1496 days of observation by the SuperKamiokande experiment. θ_{Sun} is the angle between the momentum vector of an electron and the direction of the Sun. The shaded area indicates the elastic scattering peak. The dotted area is the isotropic background of roughly 1 event day $^{-1}$ bin $^{-1}$ due to spallation products induced by cosmic ray muons, γ -rays from outside the detector and radioactivity in the water of the detector. The angular resolution of the detector system has been taken into account in calculating the expected distribution of arrival directions of the neutrinos from the Sun (Hosaka *et al.*, 2006).

The experiment was upgraded with an active volume of 32 000 tons of pure water and 11 200 photomultiplier tubes and renamed SuperKamiokande. The rate of detection of high energy neutrinos was greatly enhanced and, from 1258 days of observation, their flux was found to be:

$$\text{Flux of neutrinos} = 2.32 \pm 0.03 \text{ (stat)}^{+0.008}_{-0.007} \text{ (syst)} \quad (2.53)$$

(Fukuda *et al.*, 2001), in agreement with the earlier results and those of Davis. Figure 2.13 shows the distribution of arrival directions of the neutrinos with respect to the direction of the Sun in the final SuperKamiokande experiment, the background being due to natural radioactivity.

A key test of the solar models is the detection of the low energy neutrinos from the first interaction of the p-p chain which is directly related to the luminosity of the Sun. These much more plentiful low energy neutrinos can be detected using gallium as the detector material. The number of radioactive germanium nuclei created by the inverse β decay process by interactions of the electron neutrinos with gallium nuclei is a measure of the neutrino flux:



The international GALLEX and SAGE experiments each required about 30 tons of pure gallium to produce a significant result. The final result of the GALLEX experiment

(Hampel *et al.*, 1999) completed in 1999 in the Gran Sasso Laboratory in central Italy was

$$\text{Measured flux of neutrinos} = 77.5 \pm 6.2 \text{ SNU}, \quad (2.55)$$

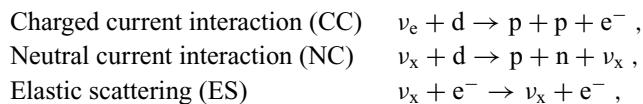
significantly less than the flux of 129^{+8}_{-6} SNU expected from the improved standard solar models of Bahcall and his colleagues (Bahcall *et al.*, 1997b). The result reported by the SAGE experiment, located at the Baksan Neutrino Observatory in the northern Caucasus mountains, was:

$$\text{Measured flux of neutrinos} = 70.9^{+5.3}_{-5.2} (\text{stat})^{+3.7}_{-3.2} (\text{syst}) \quad (2.56)$$

(Abdurashitov *et al.*, 2002, 2003).

There was a great deal of speculation about the solution of the solar neutrino problem. The favoured solution was that the deficit was associated with the phenomenon of *neutrino oscillations* in which the electron neutrinos can change type $\nu_e \rightarrow \nu_\mu$, $\nu_e \rightarrow \nu_\tau$, if the neutrinos have small but finite rest masses. In the case of electron neutrinos propagating in a vacuum, it would be expected that on average, only half the electron neutrinos emitted by the Sun would be detected as electron neutrinos, while the other half would have been transformed into ν_μ and ν_τ neutrinos. In fact, the exact fraction which are converted into ν_μ and ν_τ neutrinos can be altered from 50% as the neutrinos propagate through the material of the Sun as a result of the Mikheyev–Smirnov–Wolfenstein (MSW) effect (Mikheyev and Smirnov, 1985; Wolfenstein, 1978), as proposed by Bahcall and Bethe (Bahcall and Bethe, 1990).

The test of the neutrino oscillation picture is to measure the total flux of all types of neutrino emitted by the Sun, as well as the electron neutrinos. This has been achieved at the Sudbury Neutrino Observatory (SNO) located in Ontario, Canada. In this experiment, the detector material is 1000 tons of ultra-pure heavy water, D₂O. The great advantage of using heavy water as a detector is that the total flux of all three neutrino species can be measured and as well as the flux of electron neutrinos. Three different types of interaction of the incoming neutrinos with the material of the active volume of the detector are involved:



where ν_x refers to all three neutrino flavours, $x = e, \mu$ and τ . The key point is that the charged current (CC) interaction is sensitive only to electron neutrinos, while the neutral current (NC) reaction is sensitive to all three neutrino species. The elastic scattering (ES) is sensitive to all three flavours as well, but with considerably reduced sensitivity for μ and τ neutrinos (Ahmad *et al.*, 2002). The separation of the neutrino signal into different types is made by combining the directionality of the arrival directions of the neutrinos with their energy spectra, the energies of the neutrinos being estimated from the strength of the Cherenkov radiation signal associated with each event.

The data from 306.4 days of observation are shown in Figures 2.15*a* and *b*, below, which illustrate the different angular and energy dependencies of the three types of neutrino

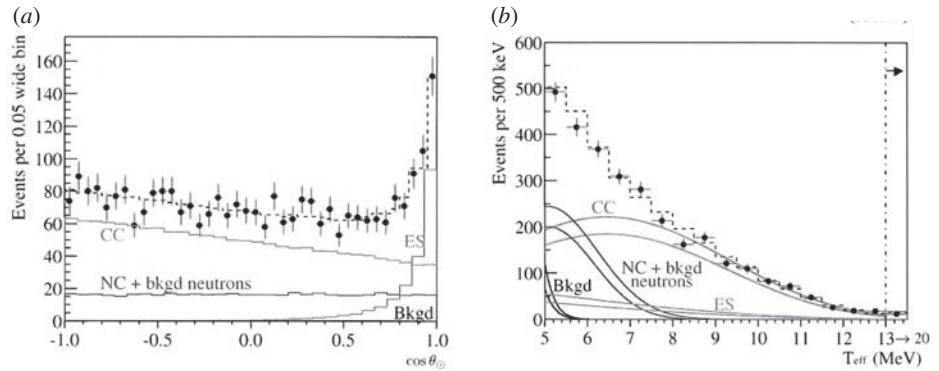


Fig. 2.14 (a) The distribution of $\cos \theta$ for neutrino events recorded by the SNO experiment. (b) The kinetic energy distribution of the neutrino events shown in (a). The histograms show the predicted distributions for elastic scattering (ES), charged current reactions (CC) and neutral current reactions (NC) from Monte Carlo simulations. All distributions are for events with kinetic energies greater than 5 MeV (Ahmad *et al.*, 2002).

interaction. The resulting estimates of the fluxes of electron and ($\mu + \tau$) neutrinos are shown in Fig. 2.14. The best estimates of the neutrino fluxes and their uncertainties quoted by the SNO consortium are as follows:

$$\begin{aligned}\phi(\nu_e) &= 1.76^{+0.05}_{-0.05} (\text{stat})^{+0.09}_{-0.09} (\text{syst}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}, \\ \phi(\nu_\mu + \nu_\tau) &= 3.41^{+0.45}_{-0.45} (\text{stat})^{+0.48}_{-0.45} (\text{syst}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}, \\ \phi(\nu_e + \nu_\mu + \nu_\tau) &= 5.09^{+0.44}_{-0.43} (\text{stat})^{+0.46}_{-0.43} (\text{syst}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}.\end{aligned}$$

These can be compared with the expectations of the standard solar models of Bahcall and his colleagues which are shown as the dashed band labelled ϕ_{SSM} in Fig. 2.15. It can be seen that the process of neutrino oscillations can completely resolve the solar neutrino problem. The next task is to reconcile the observed fluxes of low energy pp neutrinos with those of the higher energy ${}^8\text{B}$ neutrinos, but this is a non-trivial calculation which goes far beyond our present ambitions. Suffice to say that, when account is taken of the MSW effect in modifying the expectations of vacuum neutrino oscillations, the observed fluxes of all types of neutrinos can be reconciled with the standard solar model.

This is undoubtedly one of the most remarkable discoveries of modern astrophysics and demonstrates the role of astrophysics in making discoveries which strike right to the heart of fundamental physics. The same phenomenon of neutrino oscillations has now been observed in studies of μ neutrinos created in the upper atmosphere through the interaction of high energy cosmic rays with the nuclei of atoms in the Earth's atmosphere (Ashie *et al.*, 2005) and also by long baseline measurements involving neutrino detectors at different distances from the terrestrial neutrino sources (Eguchi *et al.*, 2003).

The reason for emphasising these solar experiments is that they give us confidence in the astrophysics used to describe the internal structure of the Sun and, by extension, the stars.

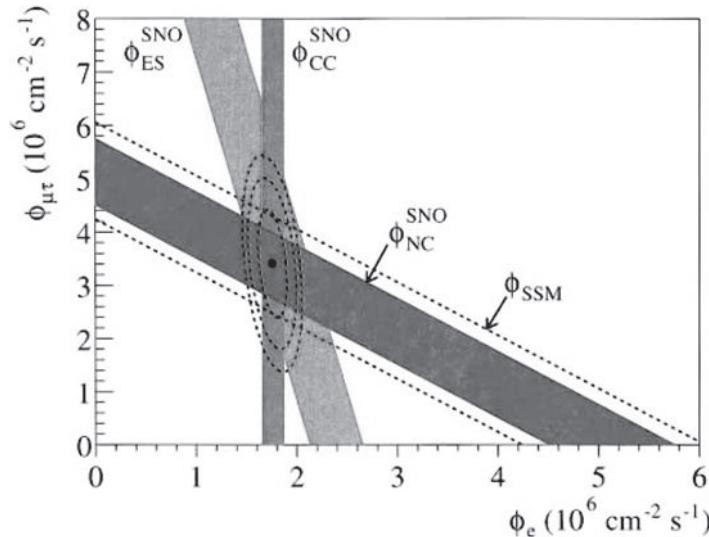


Fig. 2.15 The flux of neutrinos with energies $E_\nu > 5 \text{ MeV}$, the flux of electron neutrinos being plotted on the abscissa and combined flux of μ and τ neutrinos on the ordinate. The diagonal solid band shows the total neutrino flux and the dashed lines the 1σ uncertainties on the predicted total flux. The bands intersect at the best-fitted estimates for $\phi(\nu_e)$ and $\phi(\nu_\mu + \nu_\tau)$, consistent with neutrino flavour transformations (Ahmad *et al.*, 2002).

2.7 Evolution of high and low mass stars

Once a star has settled onto the main sequence, its luminosity changes very little until it begins to move off the main sequence when the helium core has mass about 10% of the mass of the star (Fig. 2.8). At this point, the hydrogen fuel in the core has been consumed and the core becomes *isothermal*, hydrogen burning now proceeding in a shell about the core. There are, however, important differences between the way in which low and high mass stars reach this point in their evolution which affects their subsequent evolution. First of all, we need to understand the importance of the *Hayashi track* on the Hertzsprung–Russell diagram.

2.7.1 The Hayashi track

In Hayashi's pioneering paper, the analysis concerned the stability of fully convective stars (Hayashi, 1961). The condition that a region of a star is in convective, rather than radiative, equilibrium is that the temperature gradient exceeds the adiabatic gradient of the stellar material. In this context, the term 'gradient' refers to the derivative of the temperature with respect to pressure, which is a monotonically increasing function of decreasing radius within the star. Conventionally, the temperature gradient is written for the case of the

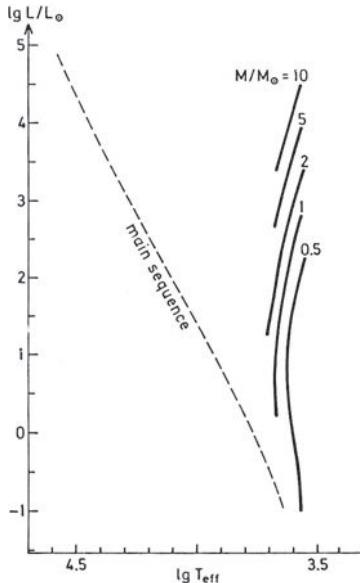


Fig. 2.16 Theoretical Hayashi tracks for fully convective stars of different masses presented by Kippenhahn and Weigert, after computations by Ezer and Cameron (Kippenhahn and Weigert, 1990)

radiative transport of energy as

$$\nabla_{\text{rad}} = \left(\frac{d \ln T}{d \ln p} \right)_{\text{rad}} \quad (2.57)$$

and depends upon the opacity of the stellar material. If the stellar material has ratio of specific heats γ , the adiabatic relation is $p \propto T^{\gamma/(\gamma-1)}$ and so $\nabla_{\text{ad}} = (\gamma - 1)/\gamma$. If the structure of the star is such that the temperature gradient exceeds this value, the material of the star becomes unstable and convection ensues. The simplest picture of what happens physically is that, when a ‘bubble’ of material is slightly compressed, it rises up the temperature gradient because of the buoyancy of the perturbed region. Convection transports energy more rapidly than radiation through the star and its internal structure reorganises itself under these convective motions until the temperature and pressure stratification satisfy the relation $\nabla_{\text{ad}} = (\gamma - 1)/\gamma$. Thus, for stars in which convection is maintained throughout the whole star, the temperature and pressure stratification is given almost exactly by the adiabatic gradient, since even a tiny departure to greater values of ∇_{rad} results in convective motions.

If the ratio of specific heats of an ideal gas is adopted, $\gamma = 5/3$, the models have polytropic index $n = (\gamma - 1)^{-1} = 3/2$. Hayashi then showed that there is an upper limit to a dimensionless parameter involving the mass, radius, temperature and pressure of the gas beyond which there exist no quasi-static solutions. This condition translates into steep loci on the Hertzsprung–Russell diagram, which are shown in Fig. 2.16 for stars of different mass. There are no quasi-static solutions to the right of these loci. A more detailed physical

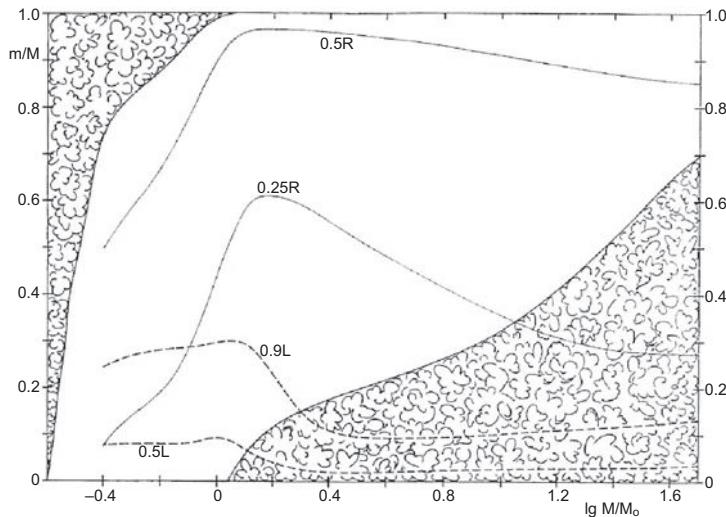
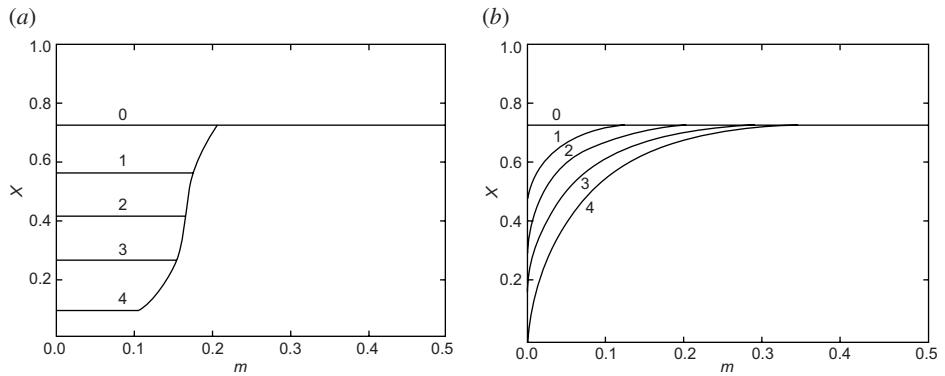


Fig. 2.17 Illustrating the convection zones in the interiors of main sequence stars (Kippenhahn and Weigert, 1990).

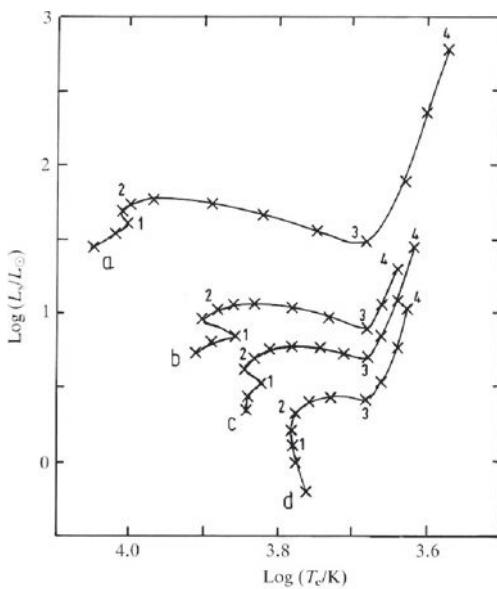
discussion of the structure of fully convective stars is given by Kippenhahn and Weigert 1990. These considerations are very important for *pre-main sequence evolution*, for the *internal structure of stars on the main sequence* and for the *red giant phase* of stellar evolution.

The results of applying the instability criterion to stars on the main sequence is illustrated in Fig. 2.17 in which it can be seen that for stars with mass greater than that of the Sun, their central regions are in convective equilibrium, whereas for stars with mass less than the Sun, the central regions are in radiative equilibrium. Notice that in the Sun, although only a very small fraction of the mass is in the outer convection zone, it corresponds to the outer 30% by radius (Fig. 2.12).

These differences are important in understanding the physics of the central nuclear burning regions of stars – these are in radiative equilibrium for stars less massive than the Sun and in convective equilibrium for more massive stars. Specifically, in *high mass stars*, the transport of energy by convection in the central regions results in unprocessed material being continually convected into the nuclear burning regions and so the hydrogen abundance decreases uniformly within these regions until the hydrogen is exhausted. In contrast, in *low mass stars*, the size of the hydrogen-burning zone increases gradually outwards with time until 10% of the mass is in a central helium core. Thus, the exhaustion of the fuel in the core is rather more gentle in the case of low mass stars as compared with those in convective equilibrium. These differences are illustrated in Fig. 2.18 which shows how the hydrogen is depleted in the central regions of high and low mass stars. X is the abundance of hydrogen by mass and $m = M(r)/ $M_{\odot}$$. In Fig. 2.19, the corresponding differences in the evolution of $1-2.5M_{\odot}$ stars from the main sequence to the red giant branch are illustrated.

**Fig. 2.18**

Illustrating the evolution of the mass fraction of helium as a function of the mass fraction $m = M(< r)/M_{\odot}$ within (a) high mass stars and (b) low mass stars (Tayler, 1994). X is the abundance of hydrogen by mass. The numbers 0 to 4 indicate the decrease in the mass fraction of hydrogen until there is less than 10% left in the very centre.

**Fig. 2.19**

The post-main sequence evolution of stars with masses of $2.25 M_{\odot}$, $1.5 M_{\odot}$, $1.25 M_{\odot}$ and M_{\odot} (from top to bottom) (Tayler, 1994).

2.7.2 High mass stars

For stars on the main sequence, the central temperature is roughly proportional to the mass of the star and so, in stars with mass $M \geq 1.7 M_{\odot}$, the CNO cycle dominates (Fig. 2.6). The evolution of the internal structure of a $5 M_{\odot}$ star is shown in Fig. 2.20 in a similar format to Fig. 2.8 for a $1.3 M_{\odot}$ star (Kippenhahn and Weigert, 1990). The heavily hatched

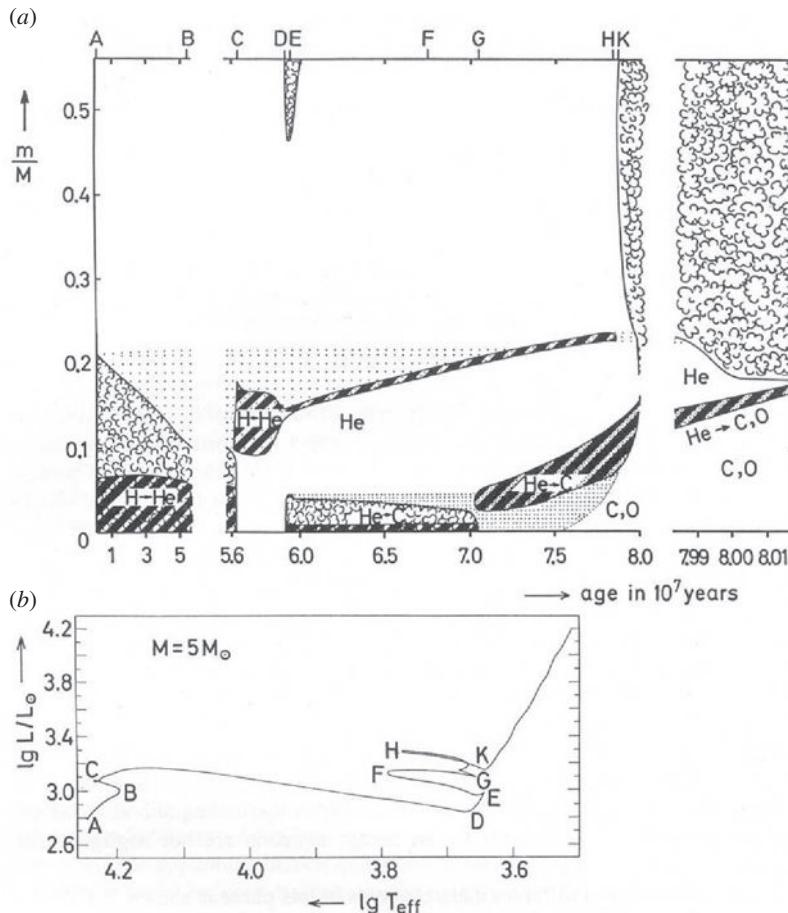


Fig. 2.20 The evolution of the internal structure of a star of $5 M_\odot$ of extreme Population I illustrating the synthesis of carbon and oxygen in the core of the star. The abscissa shows the age of the model star after the ignition of hydrogen in units of 10^7 years. Note the varying time-scale along the abscissa. The ordinate shows the radial coordinate in terms of the mass m within a given radius relative to M , the total mass of the star. The cloudy regions indicate convective zones. The corresponding positions of the star on the H-R diagram at each stage in its evolution are shown in the lower diagram (Kippenhahn and Weigert, 1990).

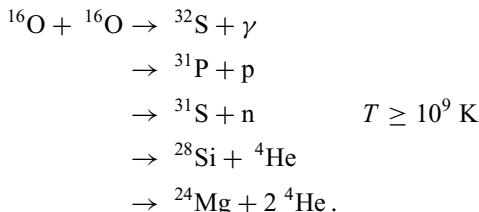
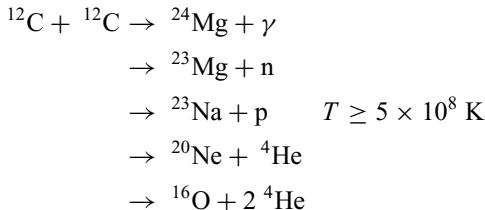
areas indicate the regions in which there is large nuclear energy production. The evolution proceeds as follows:

- At the point A, the star begins its lifetime on the main sequence. The convective core contains 21% of the mass of the star and nuclear burning takes place within the inner 7% by mass. During the first 5.6×10^7 years, the star remains at roughly the same location on the H-R diagram, evolving to the point B.
- By the point C, the central hydrogen fuel is exhausted and during the transition from C to D, an isothermal helium core is formed which begins to collapse, accompanied by the rapid expansion of the envelope to form a giant star. During the evolution from C

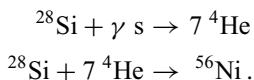
to D, hydrogen burning continues in a shell about the helium core. At the point D, the star arrives at the Hayashi track and then an outer convection zone is formed in the giant envelope.

- The continuing contraction of the central regions heats up the core until helium burning takes place at E. In the helium burning process, helium is converted into carbon ${}^3\text{He} \rightarrow {}^{12}\text{C}$ through the rare triple- α process. This is accompanied by an excursion to higher temperatures across the H-R diagram to F.
- Helium burning continues until the central helium abundance is reduced to zero and an isothermal ${}^{12}\text{C}$ core forms at G. Helium burning continues in a shell about the isothermal C,O core.
- Throughout the stages D to H, hydrogen shell burning continues to larger and larger radii, but at H hydrogen shell burning ends because the temperature in the envelope is too low.
- At K, the star develops a deep outer convection zone and subsequently moves almost vertically up the Hayashi track.

In yet more massive stars, post-main sequence evolution proceeds by successive core and shell burning to produce nuclei with higher and higher binding energies. For the most massive stars, the sequence continues with carbon and oxygen burning to produce silicon which can eventually be burned to create iron peak elements. These processes can be written



In the case of silicon burning, which begins at a temperature of about 2×10^9 K, the reactions proceed slightly differently because the high energy γ -rays remove protons and ${}^4\text{He}$ particles from the silicon nuclei and the heavier elements are synthesised by the addition of ${}^4\text{He}$ nuclei through reactions which can be schematically written



It is therefore expected that in the final stages of evolution of very massive stars, the star will take up an ‘onion-skin’ structure with a central core of iron peak elements and successive surrounding shells of silicon, carbon and oxygen, helium and hydrogen (Fig. 2.21).

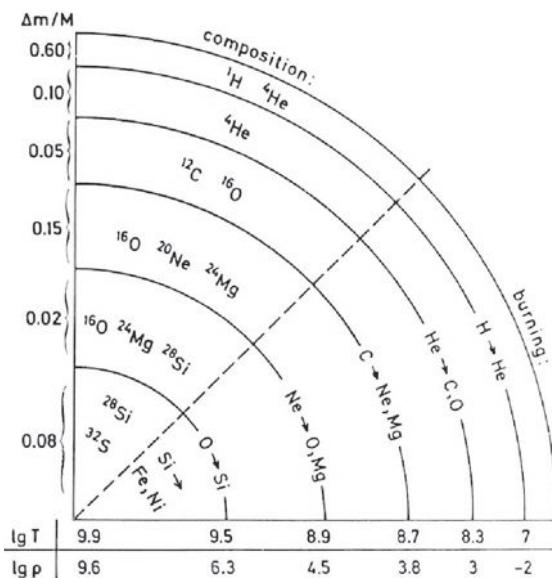
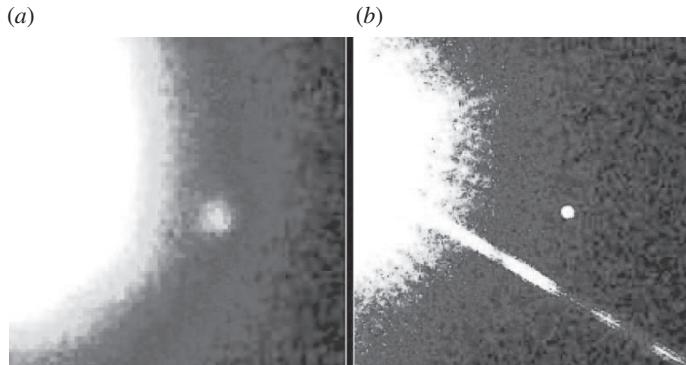


Fig. 2.21 A schematic illustration of the ‘onion-skin’ picture of the interior structure of a highly evolved $25 M_{\odot}$ star. Typical values of the mass, density (in g cm^{-3}) and temperature (in K) of the different shells are indicated along the axes (Kippenhahn and Weigert, 1990).

Iron is the most tightly bound of the chemical elements and therefore the process of nuclear burning to reach lower energy states cannot proceed beyond iron. To proceed further, two processes are important involving neutron reactions with iron peak elements. In these reactions, a neutron is absorbed and the subsequent products depend upon whether or not the nucleus formed has time to decay before the addition of further neutrons takes place. The case in which the decay occurs first is referred to as the *slow* or *s-process* and that in which several neutrons are added before β decay terminates the sequence is known as the *rapid* or *r-process*. The latter is likely to be important in the extreme conditions during explosive nucleosynthesis where very high densities and temperatures are attained and large fluxes of neutrons are produced by the inverse β decay process. This is believed to be the process which is responsible for the synthesis of neutron-rich species such as the heaviest isotopes of tin, ^{122}Sn and ^{124}Sn .

The products of the s-process are estimated by calculations in which iron, by far the most abundant of the elements heavier than oxygen, is irradiated by neutrons. The products are sensitive to the irradiation time but it has been shown that, if it is assumed that there is a range of irradiation times, the Solar System abundances of the elements heavier than iron can be accounted for. This theory has been particularly successful in accounting for the anomalously high abundances of heavy elements such as barium and zirconium and, in particular, for the unstable element technetium Tc, the longest lived isotope of which has a lifetime of only 2.6×10^6 years.

**Fig. 2.22**

(a) The discovery image of the faint brown dwarf companion to the solar-type star Gliese 229 obtained at the Palomar Observatory on 27 October 1994 (Nakajima *et al.*, 1995). (b) A confirmatory image taken by the Hubble Space Telescope on November 17 1995. (Courtesy of T. Nakajima, S. Kulkarni, S. Durrance and D. Golimowski, NASA, ESA and the HST Science Institute.)

2.7.3 Low mass stars

For stars less massive than the Sun, the central temperatures are lower and their luminosities correspondingly smaller. Such stars can be seen on the H-R diagrams of the 47 Tuc and Orion star clusters (Figs. 2.4b and 2.26 below). Eventually, at a low enough mass, the central temperature is not sufficiently high for the nuclear reactions of the p-p chain to take place, the corresponding mass being about $0.08 M_{\odot}$. Because of the strong dependence of luminosity upon mass, the lowest mass stars are very low luminosity objects and can only be detected nearby.

Objects with masses in the range $0.08 M_{\odot} > M > 0.01 M_{\odot}$ are referred to as *brown dwarfs* and these are very faint infrared objects. The first convincing example of a brown dwarf was discovered in 1995 by direct imaging of Gliese 229B, the faint companion of the nearby star Gliese 229 (Nakajima *et al.*, 1995) (Fig. 2.22). The spectrum of Gliese 229B displayed strong methane and water vapour absorption, similar to the spectrum of Jupiter. The surface temperature was less than 1000 K, too low for nuclear burning to take place in its core. Many candidates have since been found in near-infrared sky surveys, including the two Micron All-Sky Survey (2MASS) and the Sloan Digital Sky Survey. Numerous candidates have also been found in deep infrared surveys of nearby star-forming regions, such as the Pleiades, Orion and ρ Ophiuchus clusters.

Objects with masses in the range $0.01 M_{\odot} > M > 0.001 M_{\odot}$ are referred to as *exoplanets*, the mass of Jupiter being $0.001 M_{\odot}$. Extrasolar planets have been found by a number of methods. The most successful to date has been the Doppler technique of observing the wobbling of the host star about the barycentre of the planetary system because of the orbital motion of the planets. The first detection of a Jupiter-mass planet orbiting a normal star was made by this technique by Mayor and Queloz in 1995 (Mayor and Queloz, 1995). Their success can be attributed to the development of very stable spectrographs with very high spectral resolution. The amplitude of the motion of 51 Peg, a solar-type star, is very

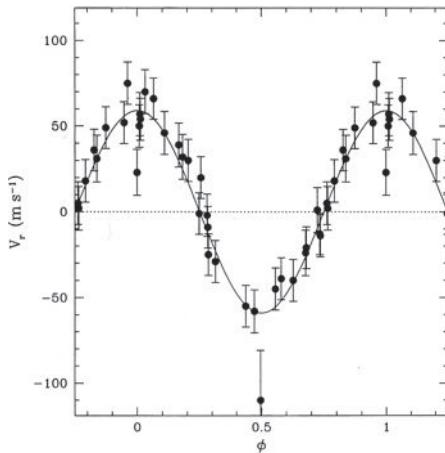


Fig. 2.23 The variation of the radial velocity of the star 51 Peg as a function of orbital phase. The period of the planet's orbit about the barycentre of the system is 4.231 days (Mayor and Queloz, 1995).

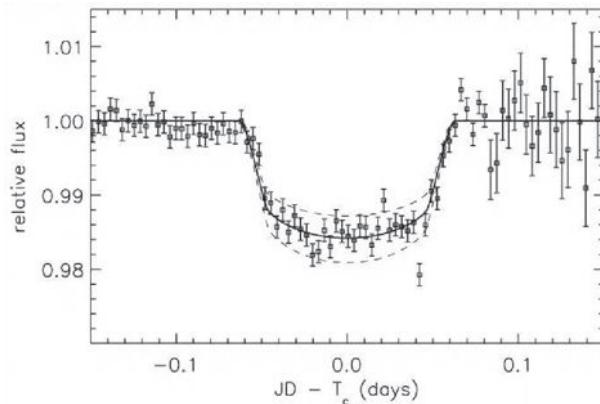


Fig. 2.24 The discovery record of the photometric time series for the star HD 209458 for September 9 and 16 1999 plotted as a function of time. The data have been averaged in 5 minute bins (Charbonneau *et al.*, 2000).

much greater than would be expected for a planetary system such as our own (Fig. 2.23). The period of the planet is only 4.231 days. Analysis of the orbital data have shown that the mass of the planet is at least 0.46 Jupiter masses and its semi-major axis only 0.052 AU. To date, most of the many extrasolar planets now known were discovered by the radial velocity technique, observational limitations generally restricting the discoveries to planets with masses of roughly that of Jupiter or greater.

A second successful technique has been to search for a small decrease in the flux of the star caused by the transit of the planet over the stellar disc. This occultation technique was first successfully used to detect a planet orbiting about the star HD 209458 (Charbonneau *et al.*, 2000). This star is a G0 V dwarf star, similar to the Sun, and so, assuming the stellar radius to be $1.1 R_\odot$ and its mass $1.1 M_\odot$, the eclipse data have been interpreted as being

due to the transit of a gaseous giant planet with radius 1.27 times the radius of Jupiter in an orbit with inclination of 87° .

These discoveries resulted in two major surprises, which have forced the revision of the theory of the formation of planetary systems. Firstly, a large fraction of Jupiter-like companions orbit about 100 times closer to their parent stars than in our Solar System. More than half the gaseous giant planets orbit within 1 AU of the host star and a significant fraction within 0.1 AU. Our Solar System seems to be the odd-man out. A favoured solution is that, since such gaseous giants could not have formed so close to the primary star, Jupiter-sized planets must have been formed much further away and then undergone *orbital migration* under the influence of tidal forces. The second surprise was that the orbits of many of the Jupiter-sized planets are *highly elliptical*. This poses problems for the standard picture of planet formation in which the planets are formed by accretion in a protoplanetary disc. Suggestions have included the proposals that they formed directly by gravitational condensation, rather than by accretion in a protoplanetary disc, or that their orbits may have been strongly perturbed by a companion star or maybe that a gravitational sling-shot mechanism ejected the planet into an elliptical orbit through an encounter with another planet.

2.8 Stellar evolution on the colour–magnitude diagram

The picture of stellar evolution developed so far needs to be further refined for precise comparison with observation. In a more complete exposition, we need to take account of the dependence of the location of the main sequence on the metallicity, or metal abundance, of the stars. It has been assumed that the perfect gas law holds good throughout the star, but in precise work, the equation of state needs to be determined for the local conditions of temperature, density and metallicity inside the star. The effects of electron degeneracy pressure upon the structure of the star have been neglected. Inside the Sun, the effects of degeneracy are not important. When the core of the star shrinks, however, its density increases and the gas can become degenerate. We will deal with fully degenerate stars for the cases of white dwarfs and neutron stars in due course.

Putting together these and many other effects, theoretical stellar evolution tracks can be plotted on what might be called a ‘theorist’s’ luminosity–effective temperature diagram (Fig. 2.25). Stars spend relatively long periods of time in the shaded areas and pass rapidly across unshaded areas and so, statistically, the shaded regions are the locations where stars are expected to be observed on an H–R diagram. These evolutionary tracks can provide a convincing explanation for the colour–magnitude diagrams of star clusters of different ages.

As explained in Sect. 2.5, the *main sequence termination point* is a robust measure of the age of a star cluster. These ages are derived from the expression (2.43) derived above, $T(M) = 10^{10} (M/M_\odot)^{-(x-1)}$ years, combined with appropriate luminosity–temperature and temperature–mass relations. For example, for solar mass stars, the homologous models

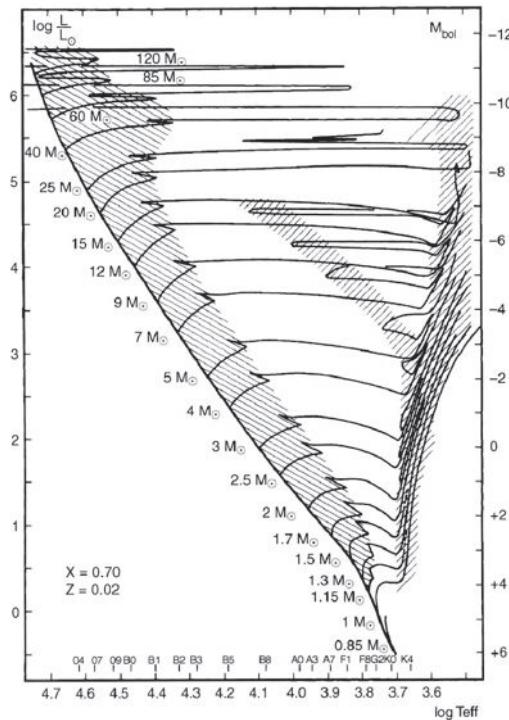


Fig. 2.25 Theoretical stellar evolution tracks on a ‘theorist’s’ luminosity–effective temperature diagram. Stars spend relatively long periods of time in the shaded areas of the diagram and pass rapidly across unshaded areas. X is the mass fraction of hydrogen, Y that of helium and Z that of elements heavier than helium, the ‘metals’ (Maeder and Meynet, 1989).

gave the results

$$L \propto T_{\text{eff}}^{4.1}; \quad T_{\text{eff}} \propto M^{69/52}. \quad (2.58)$$

The more massive the star the greater the rate at which it burns up its nuclear fuel. Thus, massive stars, say $20M_{\odot}$, have effective temperatures of about 10^5 K. They are luminous blue stars with ages only about 10^6 years. Even younger stars are observed in the most nearby massive star-forming region, the Orion star cluster (Fig. 2.26). The colour–magnitude diagram shows a main sequence extending to about $60 M_{\odot}$ (Hillenbrand, 1997). Many of these stars are still deeply embedded in the giant molecular cloud from which they were formed. The cluster has the morphology of an *open star cluster* which is dynamically young.

In contrast, the rich *globular cluster* 47 Tucanae is dynamically old with the distribution of stars strongly concentrated towards the centre. The colour–magnitude diagram has a main sequence which only extends to about the mass of the Sun and there is a very well-populated *giant branch*, as well as a *horizontal branch* (Hesser *et al.*, 1987) (Fig. 2.4b). The detailed study of the H–R diagram of the cluster by Hesser and his colleagues showed that the metallicity is only 20% of the solar value and the age of the cluster between about $(1.2\text{--}1.4) \times 10^{10}$ years. The other examples shown in Fig. 2.4a enable stellar evolution to

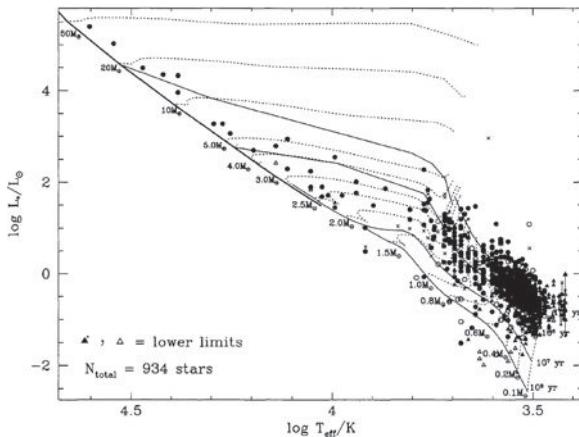


Fig. 2.26 The colour–magnitude diagram for the Orion star cluster. Superimposed on the diagram is the zero-age main sequence and pre-main sequence evolutionary tracks (Hillenbrand, 1997).

be studied in considerable detail under reasonably controlled astrophysical conditions. The *oldest globular clusters*, found in the halo of our Galaxy, have ages of about 1.4×10^{10} years, providing an estimate of the age of the Universe.

2.9 Mass loss

An important part of the story is the phenomenon of *mass loss* from the outer envelopes of stars. Stars lose mass from their surfaces throughout much of their lives. The Einstein X-ray Observatory established that essentially all classes of normal stars emit X-rays, the radiation generally originating from hot stellar coronae or stellar winds. Thus, coronae similar to that observed about our own Sun must be common to most classes of star. Such stellar coronae are believed to be heated by waves or shock waves originating in the convective layers close to the surface of the star and this energy is dissipated above the photosphere leading to strong heating of the lower density gas in the immediate vicinity of the Sun. The gas in the solar corona is heated to temperatures in excess of 10^5 K so that it is no longer bound to the Sun and a stellar wind, in our case the *Solar Wind*, is created. This may be termed *quiescent mass loss*. There are, however, much more violent forms of mass loss which are associated with the various evolutionary changes which stars undergo both when they are on the main sequence and after they have left it.

2.9.1 P-Cygni profiles and Wolf–Rayet stars

One of the most direct methods of observing mass loss is through the observation of *P-Cygni* profiles associated with the emission lines of hot stars (Fig. 2.27). In this type of profile, the emission line originates in the stellar atmosphere but the short wavelength side of the star

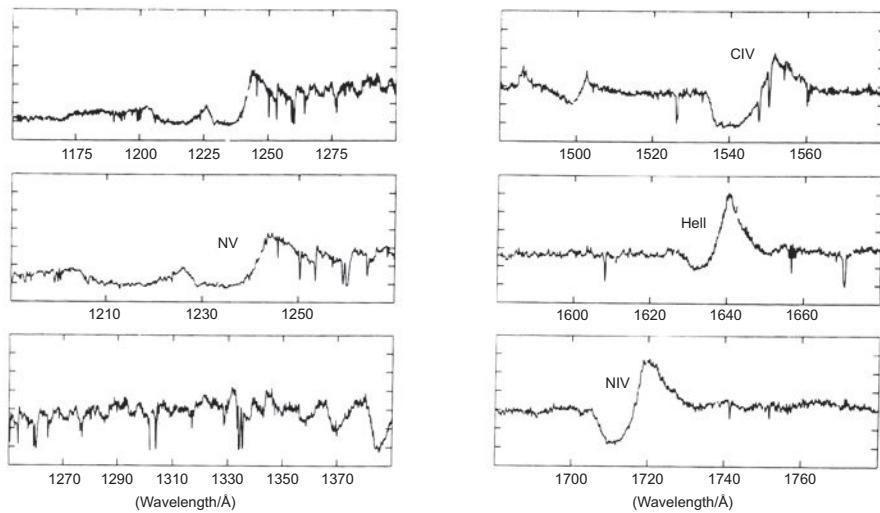


Fig. 2.27 Examples of the P-Cygni profiles of emission lines in the spectrum of the hot Wolf–Rayet star HD 93131. The outflow of gas in the form of a wind causes absorption of both the line and continuum radiation to the short wavelength side of the emission line. In this spectrum there are many strong emission lines and P-Cygni profiles are observed in the lines of N IV, N V, He II and C IV. (Willis *et al.*, 1986).

is strongly modified by absorption by the same types of ions responsible for the emission line in outflowing material along the line of sight towards the observer. The outflowing material absorbs not only the emission line radiation but also the underlying continuum of the star. Observations of this type were made with particular success in the ultraviolet waveband by the International Ultraviolet Explorer (IUE) because the resonance lines of many of the common elements fall in this waveband. In the example of the Wolf–Rayet star HD 93131 shown in Fig. 2.27, P-Cygni profiles are associated with the ions of N IV, N V, He II and C IV. As a result, mass loss rates have been determined for many classes of hot star.

In the evolutionary tracks shown in Fig. 2.25, it was assumed that mass loss is unimportant but it is now clear that, for the most luminous stars, mass loss plays a major role in their evolution. With increasing mass on the main sequence, radiation pressure becomes more and more important, until at high enough luminosities, the star would exceed the Eddington limiting luminosity (Sect. 13.2.2). Observational evidence and theoretical investigations indicate that stars with masses greater than about $60 M_{\odot}$ are subject to a radiation-driven pulsational instability which becomes nonlinear and ejects layers of gas from the surface of the star until its mass is reduced to about $60 M_{\odot}$. Many of the massive stars with masses up to this limiting value exhibit enormous mass loss rates, values as large as $10^{-4} M_{\odot} \text{ year}^{-1}$ being common. The extreme star Eta Carinae is estimated to have mass about $120 M_{\odot}$ and has a mass loss rate of about $10^{-2} M_{\odot} \text{ year}^{-1}$, as can be seen in the spectacular Hubble Space Telescope image of the large lobes associated with its bipolar outflow (Fig. 2.28). These stars lose mass at such a high rate that they lose their hydrogen envelopes during what would normally be their main sequence phase of evolution, exposing the helium cores

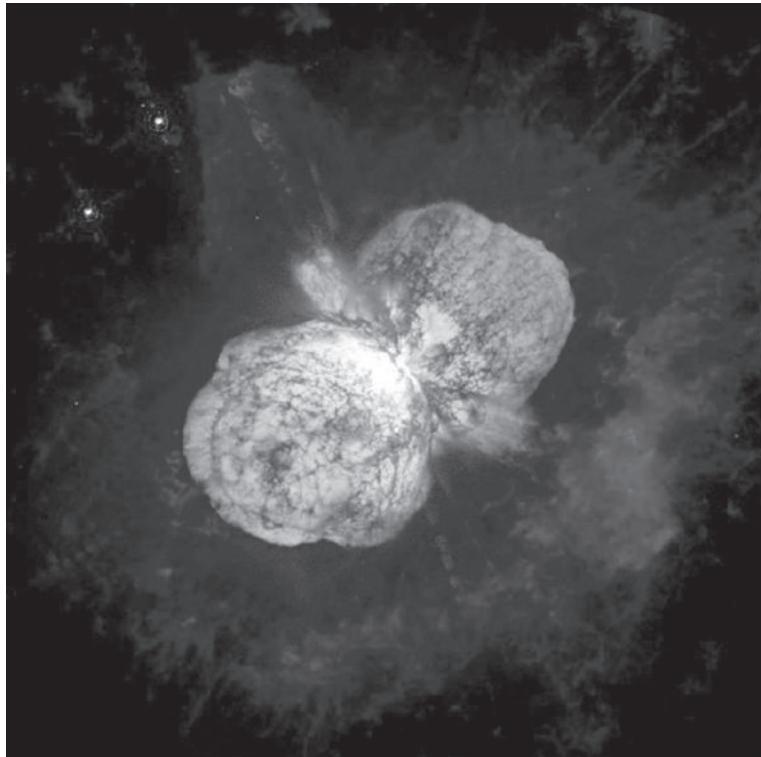


Fig. 2.28 A Hubble Space Telescope image of the ultra-luminous star Eta Carinae and the associated bipolar Homunculus Nebula. The bipolar structure was partly created in an eruption of Eta Carinae, which was observed in 1843. Eta Carinae is the bright star located at the point where the lobes of the Homunculus touch. (Courtesy of NASA, ESA and the Space Telescope Science Institute.)

created in their centres. In less extreme cases, they may evolve towards the red giant region and then suffer further mass loss from their surfaces. Mass loss of these forms is believed to be the origin of the class of star known as *Wolf–Rayet* stars which are massive helium stars with high abundances of carbon or nitrogen. Typical mass loss rates in these stars are about $3 \times 10^{-5} M_{\odot}$ year $^{-1}$. A number of the Wolf–Rayet stars are members of binary systems and so Roche lobe overflow may also be an important mass loss mechanism (see Sects. 13.4 and 13.5).

The Wolf–Rayet stars come in two main varieties, the WC stars which exhibit very strong carbon lines but no nitrogen, and the WN stars which have strong nitrogen lines but are deficient in carbon. It is likely that these differences reflect the different evolutionary status of the two types. The WN stars can be naturally associated with massive O stars in which the products of hydrogen burning through the CNO cycle are exposed due to the effects of strong mass loss from their surfaces. In contrast, the WC stars can be naturally associated with stars which have proceeded through to helium burning in their cores. The triple- α process takes place at a higher temperature than the CNO cycle and has the effect, not only of creating ^{12}C but also of destroying the nitrogen. Evidently, there must be considerable

mixing and mass loss to make the products of nuclear processing apparent in the stellar atmosphere. These stars may be important in explaining some of the abundance anomalies observed in the cosmic rays. There is convincing evidence that this type of mass loss must have been important in the evolution of the progenitor star of the supernova 1987A (see Sect. 12.1.2).

2.9.2 The horizontal branch

Evidence that mass loss must occur in stars with mass $M \sim M_{\odot}$ is provided by the evidence of *horizontal branch stars* which are observed on the H-R diagrams of globular clusters such as 47 Tucanae (Fig. 2.4b). These stars have high abundances of helium and models which account for their surface properties indicate that they have masses $M \approx 0.5M_{\odot}$. Further evidence is provided by the *RR-Lyrae* variable stars which are members of the horizontal branch population. They are only found in the region of the H-R diagram where the *instability strip* intersects the horizontal branch. Models which can account for the regular variability properties of RR-Lyrae stars indicate that their masses are also about $0.5 M_{\odot}$. Since the main sequence termination point for the oldest stars in the Galaxy has just reached one solar mass, the horizontal branch stars must have suffered highly significant lost mass from their outer layers.

There is a plausible explanation for such stars in the context of stellar evolution. When stars with masses less than $2 M_{\odot}$ consume all the hydrogen fuel in their cores, the inert helium core contracts and become degenerate. When the temperature in the core becomes sufficiently great for helium burning to begin, the core does not immediately expand because the pressure of the degenerate gas is independent of temperature and so a nuclear runaway situation develops in which the temperature continues to rise and the nuclear fusion rate increases exponentially – the *helium flash* is all over in a few hours and only terminates when the temperature has increased to such a high value that the degeneracy is relieved. The precise subsequent evolution is not certain, but most of the energy released in the helium flash is probably absorbed by the envelope, resulting in the partial ejection of the envelope on a dynamical time-scale. This is probably the process responsible for the formation of horizontal branch stars as part of the natural evolution of stars with masses $M \approx M_{\odot}$. Models of horizontal branch stars indicate that they then evolve back towards the tip of the giant branch.

2.9.3 Planetary nebulae

As the star moves towards the tip of the giant branch, it reaches the region occupied by long period variables and unstable stars. These are stars in the very final phases of evolution and there is a continuity in properties between the various classes of objects found in this region of the diagram. The long period variables and the OH/IR stars appear to form a continuous sequence with increasingly long oscillation periods leading ultimately to a region of the H-R diagram populated by unstable stars. For stars with masses in the range $2-10 M_{\odot}$, nuclear burning does not proceed beyond the formation of a degenerate oxygen–carbon core. Instabilities in the outer layers of the star result in the expulsion of the envelope of

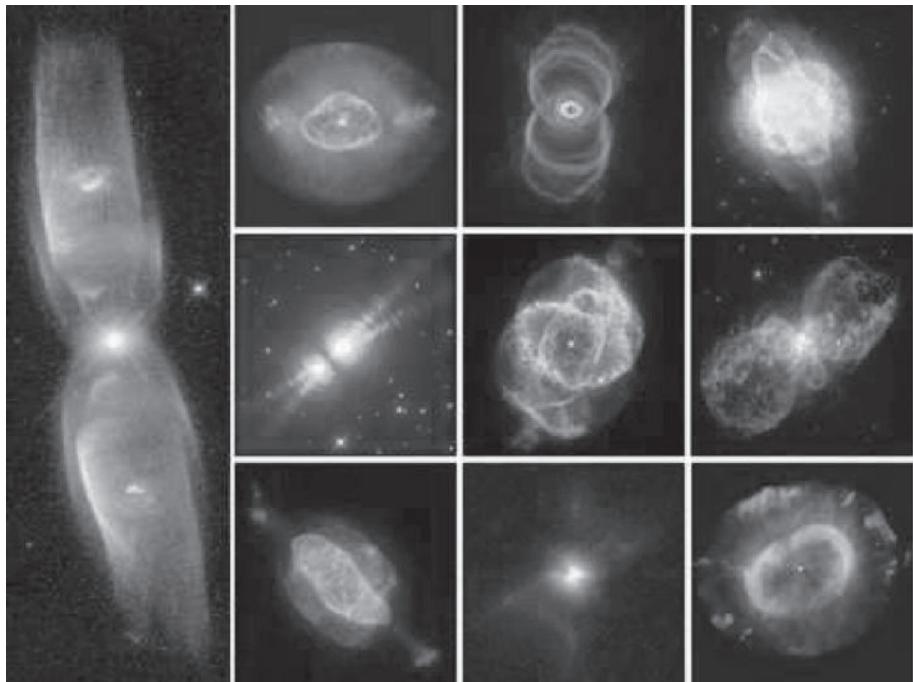
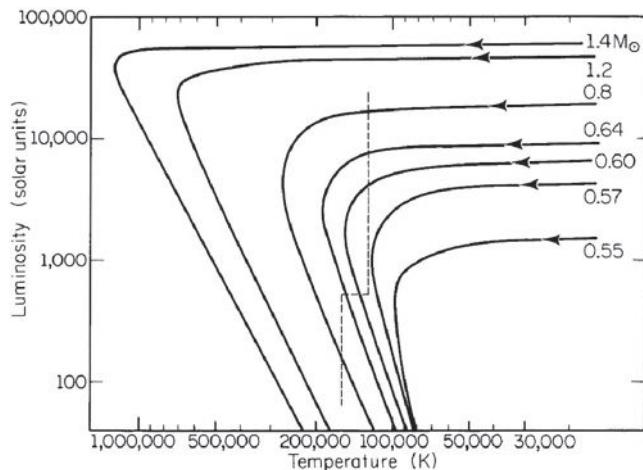


Fig. 2.29 A gallery of images of planetary nebulae observed by the Hubble Space Telescope. Many of the complexities of the images are associated with mass loss events in which the ejected material encounter the debris of previous mass loss events. (Courtesy of the NASA, ESA and the Space Telescope Science Institute.)

the giant star, leading to the formation of the *planetary nebula* phase of evolution. These roughly spherical shells of gas are observed moving outwards from the central star, the velocities being typical of the escape velocity from the surface of a star belonging to the giant branch – examples of their beautiful images are shown in Fig. 2.29. The wealth of complex structures is probably associated with a sequence of mass ejection events, subsequent expulsions of stellar material encountering the debris of past events. Dust shells about giant stars are also detected by their far-infrared emission, either in the wavebands accessible from the ground at 10 and 20 μm or from space infrared telescopes such as the Infrared Astronomical Satellite (IRAS). Dust particles condense in the cooling outflows from giant stars and these are then heated up by the stellar radiation from the giant star. The dust is heated to temperatures in the range 100–1000 K and this is readily detected as intense far-infrared radiation.

The central stars of planetary nebulae are observed to be very hot with surface temperatures which can exceed 100 000 K. Their optical spectra show little evidence for hydrogen, the lower mass remnants being essentially helium or carbon–oxygen stars, the implication being that most of the outer layers of the stars have already been expelled. These very hot compact stars follow a sequence on the H-R diagram which indicates that they end up as *white dwarf stars* (Fig. 2.30).

**Fig. 2.30**

The evolutionary tracks for the central stars of planetary nebulae. The evolutionary tracks shown on the diagram are theoretical tracks for the central stars of planetary nebulae which result in remnants with masses between 1.44 and $0.55 M_{\odot}$. The former were formed from asymptotic giant branch stars with mass $\sim 10 M_{\odot}$ and the latter from $0.8 M_{\odot}$ stars. Most planetary nebulae are observed to the right of the dashed lines (Kaler, 2001).

2.9.4 Overall mass loss rates

Summing over all forms of mass loss from stars in our own Galaxy, it is likely that about $1-10 M_{\odot}$ of material each year is returned to the interstellar medium. This means that the interstellar medium is constantly being replenished by stellar mass loss. Over the last 10^{10} years, it is likely that a considerable fraction of the baryonic mass of the Galaxy has been circulated through stellar interiors, providing a plausible explanation for the fact that the abundances of the elements in stars seem to have a fairly universal character. What we have not addressed in this section is how the observed abundances of the elements are created. Obviously, many of the mass loss processes described above involve the expulsion of the outer layers of the stars and newly synthesised elements in their cores are not available for enriching the interstellar gas unless there is considerable mixing. It is likely that *supernova explosions* are responsible for much of the chemical enrichment whilst the overall gaseous content of the interstellar gas is maintained by the somewhat more quiescent forms of mass loss described in this section.

2.10 Conclusion

This brief introduction of the ideas of stellar evolution is intended to provide the context for the study of high energy astrophysical phenomena in the subsequent chapters. Intentionally, we have not described two of the most important parts of the life cycle of stars – their birth

and death. The processes involved in the birth of stars are closely related to the study of the properties of the interstellar gas and these topics are the subject of Chap. 11.

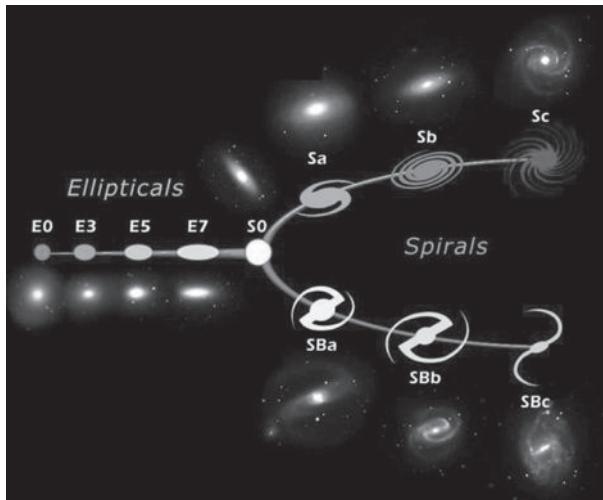
The death of stars is even more important since unquestionably this involves the formation of objects which are central to high energy astrophysics. We know rather precisely the types of objects which can be formed at the end of a star's lifetime. In all three types of 'dead star', there is no longer any nuclear generation of energy. In *white dwarfs*, internal pressure support is provided by electron degeneracy pressure and their masses are roughly the mass of the Sun or less. A second possible end point is as a *neutron star* in which internal pressure support is provided by neutron degeneracy pressure. These stars are very compact, having masses about the mass of the Sun and radii about 10 km. They have been found in two ways. In the first, they are the parent bodies of *radio pulsars* which are rotating, magnetised, neutron stars. In the second case, they are the compact 'invisible' secondary stars of *binary X-ray sources* in which the X-rays are produced by matter falling from the normal primary star onto the neutron star, the process known as *accretion*. As part of that study, we will study the evolution of stars in binary stellar systems. The third possibility is that the star collapses to a *black hole*. We will show in Chap. 12 that white dwarfs and neutron stars cannot have masses greater than about $3 M_{\odot}$ at most while, for greater masses, the only stable configuration is as a *black hole*. These objects play a central role in high energy astrophysics and will be studied in some detail.

3.1 Introduction

Galaxies are complex, many-body systems. Typically, a galaxy can consist of hundreds of millions or billions of stars, it can contain considerable quantities of interstellar gas and dust and can be subject to environmental influences through interactions with other galaxies and with the intergalactic gas. Star formation takes place in dense regions of the interstellar gas. To complicate matters further, it is certain that dark matter is present in galaxies and in clusters of galaxies and that its mass is considerably greater than the mass in baryonic matter. Consequently, the dynamics of galaxies are dominated by this invisible dark component, the nature of which is unknown.

Traditionally, galaxies have been classified by meticulous morphological studies of samples of bright galaxies. These morphological classification schemes had to encompass a vast amount of detail and this was reflected in Hubble's pioneering studies, as elaborated by de Vaucouleurs, Kormendy, Sandage, van den Bergh and others. The Hubble sequence of galaxies has real astrophysical significance because a number of physical properties are correlated with Hubble type. While the detailed study of individual galaxies was feasible for reasonably large samples, a different approach had to be adopted for massive surveys of galaxies such as the Anglo-Australian 2dF survey (AAT 2dF) and the Sloan Digital Sky Survey (SDSS) which have provided enormous quantitative databases for the studies of galaxies. As a result, classification schemes had to be based upon parameters which could be derived from computer analysis of the galaxy images and spectra. What this new approach loses in detail, it more than makes up for in huge statistics and in the objective nature of the classification procedures.

These recent developments have changed the complexion of the description of the properties of galaxies. While the new samples provide basic global information about the properties of galaxies, the old schemes describe many features which need to be incorporated into the understanding of the detailed evolution and internal dynamics of particular classes of galaxy. As a result, we need to develop in parallel both the traditional and more recent approaches to the study of galaxies. We will summarise briefly some of their more important properties, as well as elucidating aspects of the essential physics. The books *Galaxies in the Universe: an Introduction* by Sparke and Gallagher, *Galactic Astronomy* by Binney and Merrifield and *Galactic Dynamics* by Binney and Tremaine can be thoroughly recommended as much more thorough introductions to these topics (Sparke and Gallagher, 2000; Binney and Merrifield, 1998; Binney and Tremaine, 2008). I have given an extended

**Fig. 3.1**

The Hubble sequence of galaxies with images and sketches illustrating their defining characteristics (Kennicutt, 2006).

introduction to many aspects of galaxies in my book *Galaxy Formation*, which can be consulted for more details (Longair, 2008).

3.2 The Hubble sequence

Galaxies come in a wide variety of different *morphologies*. Some order was put into this diversity by Edwin Hubble in his pioneering studies of the properties of galaxies as extragalactic systems (Hubble, 1936). Hubble ordered the galaxies in what came to be known as the *Hubble sequence*, distinguishing those of elliptical appearance, the elliptical or E galaxies, from the normal S and barred SB spiral galaxies, as illustrated schematically by the ‘tuning fork’ diagram shown in Fig. 3.1. For elliptical galaxies, the number n after the E describes the ellipticity of the image, $n = 10 \times (a - b)/a$, where a and b are the major and minor axes of the galaxies. De Vaucouleurs argued convincingly that classes Sd and SBd should be included to the right of the sequence and that the irregular galaxies should be shown even further to the right. Hubble believed that the tuning fork diagram was an evolutionary sequence and so those to the left of the diagram, the ellipticals, are still often referred to as *early-type galaxies*, while those to the far right, the spirals and irregulars are often called *late-type galaxies*. The classic Hubble types shown in Fig. 3.1 and Fig. 3.2 mostly refer to luminous galaxies.

Figure 3.2 shows Hubble Space Telescope images of examples of these Hubble types. The ellipsoidal distribution of old stars in the giant elliptical galaxy M87 is shown in Fig. 3.2a. Several bright globular clusters associated with the smooth distribution of starlight can be seen. The image shows the remarkable non-thermal jet, seen also in the radio and X-ray wavebands, which originates in a massive black hole in the nucleus which is also an intense

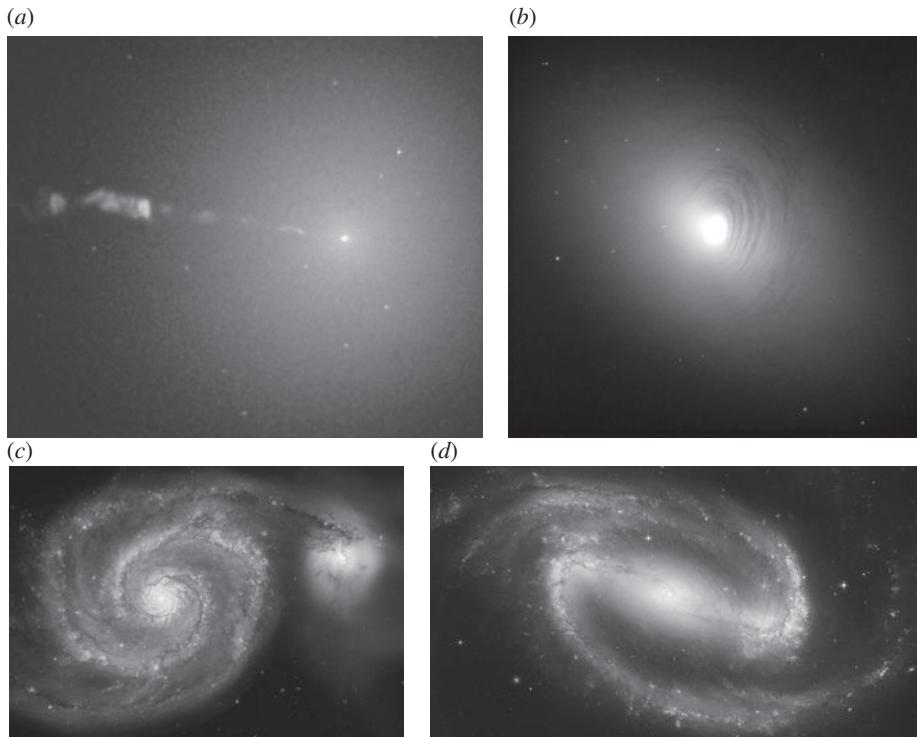
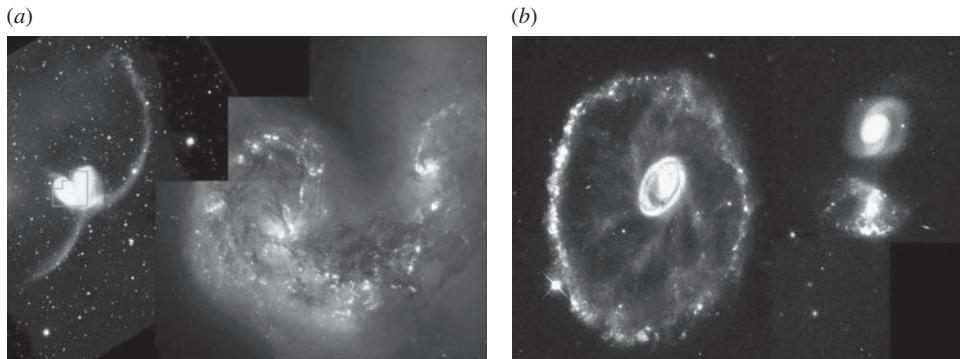


Fig. 3.2 Examples of luminous galaxies. (a) M87: giant elliptical galaxy. (b) NGC 2787: SB0 galaxy. (c) M51: spiral galaxy. (d) NGC 1300: barred spiral galaxy. (Courtesy of NASA, ESA and the Space Telescope Science Institute.)

source of non-thermal optical radiation. The lenticular (lens-like), or S0, galaxies can be thought of as spiral galaxies with the spiral arms removed. They have a clear bulge and disc structure, but at later stages along the S0 sequence, dust lanes are commonly found, as can be seen in the image of SB0 galaxy NGC2728 in Fig. 3.2b. Figure 3.2c is a beautiful image of the spiral galaxy M51 (or NGC 5194) with its nearby dusty companion NGC 5195. Intense regions of ongoing star formation, which define the spiral arms, are red because of the effects of dust extinction. The blue regions of the spiral arms are associated with recently formed stars which have escaped from their birth sites. Although apparently a symmetric galaxy, there are important deviations from symmetry induced by the close encounter with its nearby companion. The barred spiral galaxy NGC1300 shown in Fig. 3.2d displays prominent spiral arms emanating from the ends of the central bar. The arms are defined by populations of rather young blue stars which cannot have moved far from their birth places in giant molecular clouds.

The Hubble classification in its revised and extended form can encompass the forms of virtually all galaxies. A few galaxies, less than 1% at the present day, have very strange morphologies and these are referred to collectively as *peculiar galaxies*. Most of these strange morphologies are associated with strong gravitational interactions, or collisions, between galaxies. In Fig. 3.3a, the Antennae is interpreted as a collision between two

**Fig. 3.3**

Examples of peculiar galaxies. (a) The Antennae is a collision between two gas-rich spiral galaxies. (b) The Cartwheel is a ring galaxy which is interpreted as a collision between a gas-rich spiral and a nearby companion. (Courtesy of NASA, ESA and the Space Telescope Science Institute.)

gas-rich spiral galaxies. The collision between the interstellar gas clouds belonging to the two galaxies has given rise to a great deal of star formation, as indicated by the large number of blue star clusters and dense clouds of interstellar dust. The long tails associated with the colliding galaxies are attributed to the galaxies interacting in a prograde collision, meaning that the rotation axes of the galaxies are in the same sense as their axis of rotation about their common centre of mass (Toomre and Toomre, 1972). In other cases, the stellar component is in the form of a ring rather than a disc or spheroid. These are known as *ring galaxies*, an example being the Cartwheel galaxy shown in Fig. 3.3b. The remarkable ring structure is attributed to the passage of a companion galaxy through the central regions of a disc galaxy which causes a ‘tidal wave’ to propagate out through the disc, compressing the gas and giving rise to star formation in a ring.

3.3 The red and blue sequences

A number of important correlations exist between the physical properties of galaxies and their morphological types, details of which are described in the texts recommended in Sect. 3.1. It is convenient to illustrate these correlations using the results of analyses of the massive databases derived from the *AAT 2dF* and *Sloan Digital Sky Surveys* which contain about 225 000 and a million galaxies, respectively. Such huge samples necessitate the development of computer algorithms which provide a quantitative approach to the characteristics of galaxies. The outcome of these studies is that what are traditionally referred to as *early* and *late-type galaxies* are found to form two distinct sequences which are known as the *red* and *blue* sequences, or the *red sequence* and the *blue cloud*. In summary,

- the *red sequence* consists mostly of non-star-forming, high-mass spheroidal galaxies, or, more colloquially ‘old, red and dead’ galaxies;

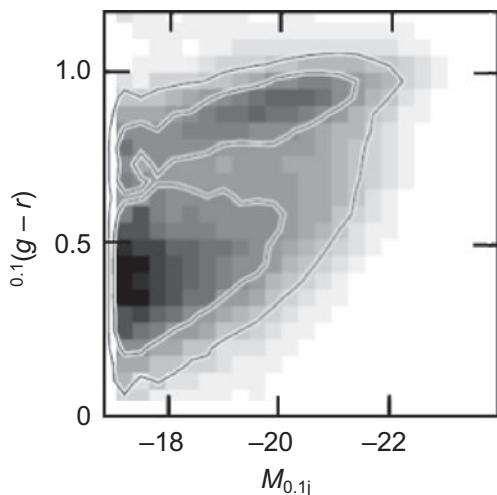


Fig. 3.4 Illustrating the bimodality in the distribution of the colour $^{0.1}(g - r)$ of galaxies as a function of optical absolute magnitude (Blanton *et al.*, 2003).

- the *blue sequence* or *blue cloud* consists mostly of star-forming, low-mass galaxies which are disc-dominated.

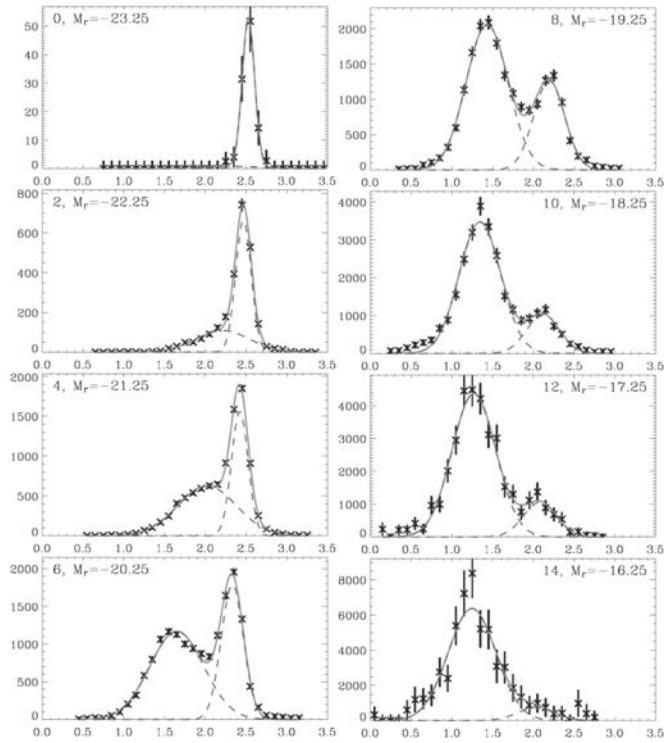
3.3.1 Colour and absolute magnitude

Perhaps the most striking diagram which illustrates the distinction between the two sequences is the plot of the colour of 144 000 galaxies from the SDSS catalogue against absolute magnitude M (Blanton *et al.*, 2003). In Fig. 3.4, the magnitudes are measured in the standard SDSS g and r filters which have mean wavelengths of 500 and 650 nm respectively. The superscript 0.1 refers to the mean redshift of the galaxies in the sample. Superimposed on the diagram are isodensity contours, most of the galaxies lying within the heavy white contours. The separation into two sequences is clearly defined, the oval region at the top of the diagram being the *red sequence* and the broader region towards the bottom right the *blue sequence*, or *blue cloud*.

Baldry and his colleagues have shown that the absolute magnitude distribution of galaxies in the red and blue sequences can be very well described by Gaussian distributions over the magnitude range $-23.5 \leq M_r \leq -15.75$ (Baldry *et al.*, 2004) (Fig. 3.5). The red galaxies are the most luminous, while the blue galaxies form the dominant population at low absolute magnitudes and this is reflected in the different luminosity functions for red and blue galaxies.

3.3.2 Sérsic index and colour

The same bimodality is present in the structural properties of the different types of galaxies. The pioneers of the studies of galaxies showed that the surface brightness distributions of the

**Fig. 3.5**

Illustrating the bimodality in the distribution of the colours of galaxies as a function of optical absolute magnitude for a sample of 66 848 galaxies selected from the Sloan Digital Sky Survey. The distributions of colours have been fitted by pairs of Gaussians. The data have been binned in intervals of 0.1 in the rest frame ($u - r$) colour. The galaxy distributions are binned in 0.5 magnitude intervals. Only half of the histograms presented by the authors are shown (Baldry *et al.*, 2004).

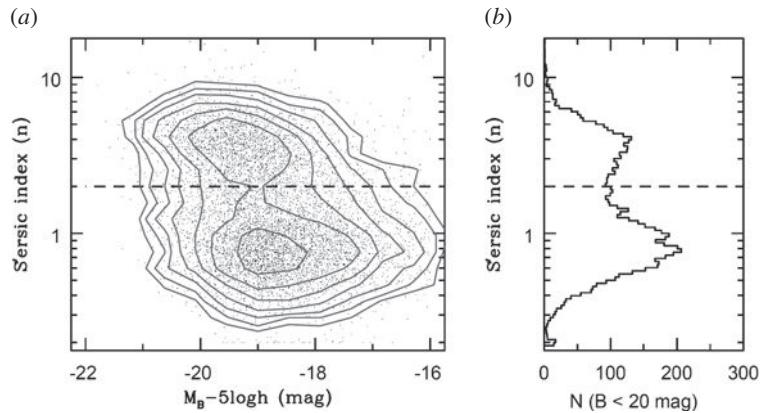
classic Hubble types can be decomposed into two components, a spheroid or bulge and a disc distribution which follow different variations with increasing radius. The light distribution in the disc is closely exponential $I(r) = I_0 \exp(-r/h)$, while that of the spheroid can be described by *de Vaucouleurs' law* which can be written

$$\log_{10} \left[\frac{I(r)}{I(r_e)} \right] = -3.3307 \left[\left(\frac{r}{r_e} \right)^{1/4} - 1 \right]. \quad (3.1)$$

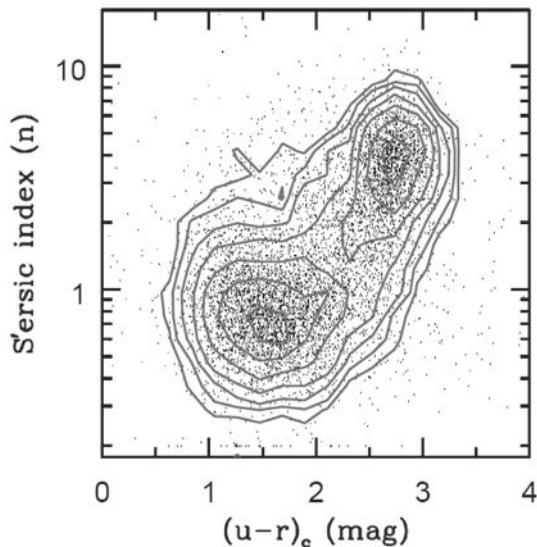
Sérsic proposed that both light distributions could be represented by the formula

$$\log_{10} \left[\frac{I(r)}{I(r_e)} \right] = -b_n \left[\left(\frac{r}{r_e} \right)^{1/n} - 1 \right], \quad (3.2)$$

where r_e is the radius within which half of the total light is emitted and b_n is a normalisation constant (Sérsic, 1968). The value $n = 4$ corresponds to *de Vaucouleurs' law* and describes the light distribution in elliptical galaxies and the bulge component of spiral and S0 galaxies. The value $n = 1$ corresponds to the exponential light distribution of disc galaxies. Values

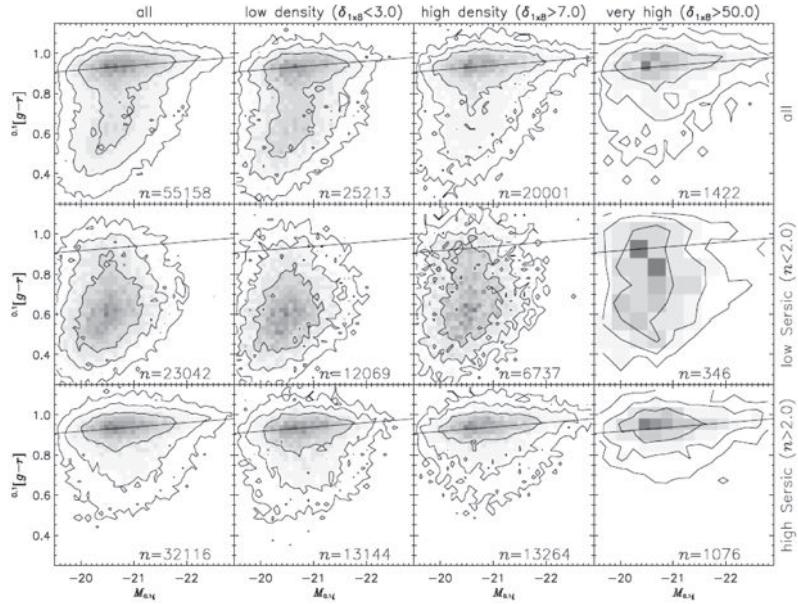
**Fig. 3.6**

(a) A plot of the observed value of the Sérsic index n as a function of the absolute blue magnitude in a sample of 10 095 galaxies from the Millennium Galaxy Catalogue. (b) The histogram showing the number of galaxies in equal logarithmic bins of Sérsic index n (Driver *et al.*, 2006).

**Fig. 3.7**

A plot of Sérsic index against colour for 10 095 galaxies selected from the Millennium Galaxy Catalogue (Driver *et al.*, 2006).

of the Sérsic index n have been determined for very large samples of galaxies from the Millennium Galaxy Catalogue and Fig. 3.6 shows that the light distribution in galaxies splits very beautifully into two populations, one centred on the value $n = 4$, corresponding to de Vaucouleurs' law, and the other on the value $n = 1$, corresponding to the exponential light distribution of disc galaxies when plotted against absolute magnitude (Driver *et al.*, 2006). This separation is even more pronounced in Fig. 3.7 in which the Sérsic index is

**Fig. 3.8**

Illustrating the bimodality of the distribution of the colours of galaxies as a function of the density of galaxies in which the galaxy is observed and as a function of their structures as parameterised by the Sérsic index n (Hogg *et al.*, 2004).

plotted against colour. In both Figs 3.6 and 3.7, the dividing line between the two sequences occurs at about $n = 2$.

3.3.3 The effect of the galaxy environment

Different types of galaxy are found in different galactic environments. For example, Dressler showed that elliptical galaxies are found with a much greater probability in rich clusters of galaxies, while spiral and irregular galaxies are found in much less dense galactic environments, including the general field (Dressler, 1980). The same distinction is found for the red and blue sequences as demonstrated by Hogg and his colleagues (Hogg *et al.*, 2004). Their sample consisted of 55 158 galaxies from the SDSS in the redshift interval $0.08 \leq z \leq 0.12$. The local galaxy density about any given galaxy was defined by the quantity $\delta_{1\times 8}$, meaning the *overdensity* about any galaxy in a cylindrical volume with transverse comoving radius $1 h^{-1}$ Mpc and comoving half-length along the line of sight of $8 h^{-1}$ Mpc. Thus, a galaxy in an environment with the average density of galaxies has $\delta_{1\times 8} = 0$. Values of $\delta_{1\times 8} \geq 50$ are found in the cores of rich clusters.

The top row of Fig. 3.8 shows contour plots of the number density of galaxies in the colour–absolute magnitude diagram of Fig. 3.4, but now shown separately for different overdensity environments, ranging from low excess number densities, $\delta_{1\times 8} \leq 3$, to very high density environments $\delta_{1\times 8} \geq 50$. These data quantify the statement that red galaxies are found preferentially in rich galaxy environments. The second and third rows further

split the sample of galaxies into those with Sérsic parameters greater and less than 2. These diagrams quantify the statement that red spheroidal galaxies are found in the richest cluster regions and these are avoided by the blue disc-like galaxies.

3.3.4 Mean stellar age and concentration index C

Another way of distinguishing the red and blue sequences is to use measures of the age of their stellar populations and the degree of concentration of the light towards their centres. Kauffmann and her colleagues used a sample of 122 808 galaxies from the SDSS to study the average age of their stellar populations using the amplitude of the Balmer break, or Balmer discontinuity, at 400 nm, $D_n(4000)$, and the Balmer absorption line index $H\delta_A$. The latter provides a measure of the strengths of the Balmer absorption lines which are particularly strong in galaxies which have undergone a recent burst of star formation (Kauffmann *et al.*, 2003). They showed that these indices provide good measures of the average star-formation activity in galaxies over the last 10^9 and $(1-10) \times 10^9$ years respectively.

The concentration index C is defined to be the ratio $C = (R90/R50)$, where $R90$ and $R50$ are the radii enclosing 90% and 50% of the Petrosian r-band luminosity of the galaxy. The concentration parameter C is strongly correlated with Hubble type, $C = 2.6$ separating the early from late-type galaxies. Those galaxies with concentration indices $C \geq 2.6$ are early-type galaxies, reflecting the fact that the light is more concentrated towards their centres.

$D_n(4000)$ and $H\delta_A$ are plotted against the concentration index C and the mean stellar mass density within the half-light radius μ_* in Fig. 3.9. Again, the galaxy population is divided into two distinct sequences. Kauffmann and her colleagues show that the dividing line between the two sequences occurs at a stellar mass $M \approx 3 \times 10^{10} M_\odot$. Lower mass galaxies have young stellar populations, low surface mass densities and the low concentration indices typical of discs. A significant fraction of the lowest mass galaxies have experienced recent starbursts. For stellar masses $M \geq 3 \times 10^{10} M_\odot$, the fraction of galaxies with old stellar populations increases rapidly. These also have the high surface mass densities and high concentration indices typical of spheroids or bulges.

3.3.5 The new perspective

The division of galaxies into members of the blue and red sequences corresponds to the division into early and late-type galaxies. To a good approximation, galaxies earlier than Sa in the Hubble sequence are members of the red sequence and later galaxies belong to the blue sequence. The relative number densities of galaxies of different types are now well established with large statistics. Bell and his colleagues have shown that, while the red sequence contains only 20% of the galaxies by number, these contribute 40% of the stellar luminosity density and 60% of the average stellar mass density at the present epoch (Bell *et al.*, 2003).

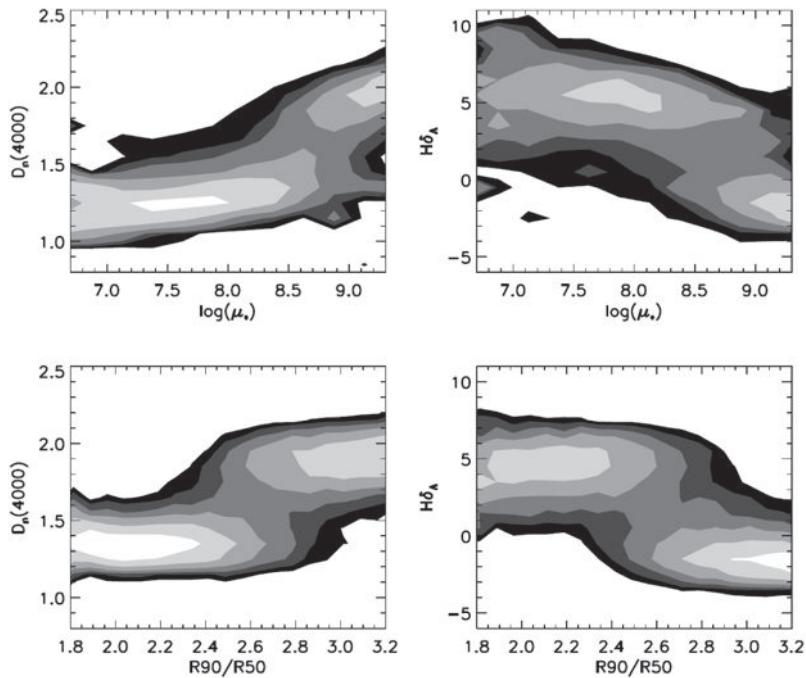


Fig. 3.9 Density distributions showing the trends of the stellar age indicators $D_n(4000)$ and $H\delta_A$ with concentration index $C = (R90/R50)$ and surface mass density μ_* (Kauffmann *et al.*, 2003).

3.4 Further correlations among the properties of galaxies

The correlations between the properties of galaxies summarised in Sect. 3.3 were derived from studies of huge samples of galaxies. Further important correlations have been derived from detailed studies of smaller samples.

3.4.1 Correlations along the Hubble sequence

What gives the Hubble classification physical significance is the fact that a number of physical properties are correlated with position along the sequence. Many of these were reviewed by Roberts and Haynes in an analysis of the properties of a large sample of bright galaxies selected primarily from the *Third Reference Catalogue of Bright Galaxies* (de Vaucouleurs *et al.*, 1991; Roberts and Haynes, 1994).

- *Neutral hydrogen.* There is a clear distinction between elliptical and spiral galaxies in that very rarely is neutral hydrogen observed in ellipticals whereas all spiral and late-type galaxies have significant gaseous masses. The upper limit to the mass of neutral hydrogen in elliptical galaxies corresponds to $M_{\text{HI}}/M_{\text{tot}} \leq 10^{-4}$. For spiral galaxies, the fractional mass of the galaxy in the form of neutral hydrogen ranges from about 0.01 for

Sa galaxies to about 0.15 for irregular galaxies, the increase being monotonic along the Hubble sequence.

- *Total surface density and surface density of neutral hydrogen.* These quantities change in opposite senses along the Hubble sequence. The total surface density, as determined by the total mass of the galaxy and its characteristic radius, decreases monotonically along the sequence, whereas the surface density of neutral hydrogen increases along the sequence.
- *Luminosity function of H II Regions.* In a pioneering study, Kennicutt and his colleagues determined the luminosity function of H II regions in different galaxy types (Kennicutt *et al.*, 1989). Normalising to the same fiducial mass, it was found that there is a much greater frequency of H II regions in late-type as compared with early-type galaxies and that the relation is monotonic along the sequence.

Roberts and Haynes pointed out that an obvious interpretation of these correlations and those discussed in Sect. 3.3 is that there are different rates of star formation in different types of galaxy. As they express it, the various correlations provide information about the past, current and future star-formation rates in galaxies. The correlation with colour along the sequence is related to the past star-formation history of the galaxy; the changes in the luminosity function of H II regions refer to star-formation rates at the present epoch; the large fraction of the mass of neutral hydrogen and its large surface density at late stages in the sequence show that these galaxies may continue to have high star-formation rates in the future.

3.4.2 The Tully–Fisher relation for spiral galaxies

In 1975, Tully and Fisher discovered that, for spiral galaxies, the widths of the profiles of the 21-cm line of neutral hydrogen, which is due to the rotational motion of the gas in their discs, are strongly correlated with their intrinsic luminosities, when corrected for the effects of inclination. They found the relation $L_B \propto \Delta V^\alpha$, where $\alpha = 2.5$ (Tully and Fisher, 1977).

The correlation was found to be much tighter in the infrared as compared with the blue waveband, because the luminosities of spiral galaxies in the blue waveband are significantly influenced by interstellar extinction within the galaxies themselves, whereas, in the infrared waveband the dust becomes transparent. What has come to be called the *infrared Tully–Fisher relation* $L_H \propto \Delta V^4$ is very tight indeed (Aaronson and Mould, 1983). Hence, measurement of the 21-cm velocity width of a spiral galaxy can be used to infer its absolute H magnitude and hence, by measuring its flux density in the H waveband, its distance can be estimated. This procedure has resulted in some of the best distance estimates for spiral galaxies and has been used in programmes to measure the value of Hubble’s constant.

3.4.3 Faber–Jackson relation and fundamental plane

Faber and Jackson found a strong correlation between luminosity L and central velocity dispersion σ of elliptical galaxies of the form $L \propto \sigma^x$ where $x \approx 4$ (Faber and Jackson, 1976). Thus, if the velocity dispersion σ is measured for an elliptical galaxy, its intrinsic

luminosity can be found from the *Faber–Jackson relation* and so, by measuring its observed flux density, its distance can be found.

A similar procedure involves the *fundamental plane* which lies in a three-dimensional space in which luminosity L is plotted against the central velocity dispersion σ and the mean surface brightness Σ_e within the half-light radius r_e , that is, $\Sigma_e = L(\leq r_e)/\pi r_e^2$. Dressler, Djorgovski and their colleagues found an even stronger correlation than the Faber–Jackson relation when surface brightness was included,

$$L \propto \sigma^{8/3} \Sigma_e^{-3/5} \quad (3.3)$$

(Dressler *et al.*, 1987; Djorgovski and Davis, 1987). Dressler and his colleagues found just as good a correlation if they introduced a new diameter D_n , which was defined as the circular diameter within which the total mean surface brightness of the galaxy exceeded a particular value. The surface brightness was chosen to be 20.75 B magnitudes arcsec $^{-2}$. The correlation found was $\sigma \propto D_n^{3/4}$, thus incorporating the dependence of both L and Σ_e into the new variable D_n .

The origin of these empirical correlations is not understood but they enable the distances of individual galaxies to be determined to about 25% and for clusters of galaxies to about 10%.

3.4.4 Mass–metallicity relation for galaxies

An important correlation for the astrophysics of galaxies is the relation between their luminosities, masses, colours and the abundances of the heavy elements, the last being referred to as their *metallicities*. In her pioneering studies, Faber showed that, for elliptical galaxies, there is a correlation between their luminosities and the strength of the magnesium absorption lines (Faber, 1973). In subsequent analyses, a similar relation was established over a wide range of luminosities and between the central velocity dispersion of the elliptical galaxy and the strength of the Mg₂ index (Bender *et al.*, 1993). They also showed that the Mg₂ index was strongly correlated with the ($B - V$) colours of the bulges of these galaxies and so the correlation referred to the properties of the galaxy as a whole.

A similar relation was found by Visvanathan and Sandage for elliptical galaxies in groups and clusters of galaxies in the sense that the more luminous the galaxy, the redder they were observed to be (Visvanathan and Sandage, 1977). The sense of the correlation was the same as that found by Faber and her colleagues since galaxies with greater metallicities have greater line blanketing in the blue and ultraviolet regions of the spectrum and hence are redder than their lower metallicity counterparts.

A similar correlation was first established for late-type and star-forming galaxies by Lequeux and his colleagues (Lequeux *et al.*, 1979). These pioneering studies involved determining the gas-phase metallicities of the galaxies and were followed by a number of studies which extended the luminosity–metallicity correlation to a range of 11 magnitudes in absolute luminosity and a factor of 100 in metallicity (Zaritsky *et al.*, 1994). These studies laid the foundation for the analyses of the huge databases of galaxies available from the Sloan Digital Sky Survey.

In the analysis of Tremonti and her colleagues, rather than using luminosity, they work directly with the stellar mass of the galaxy (Tremonti *et al.*, 2004). This approach has become

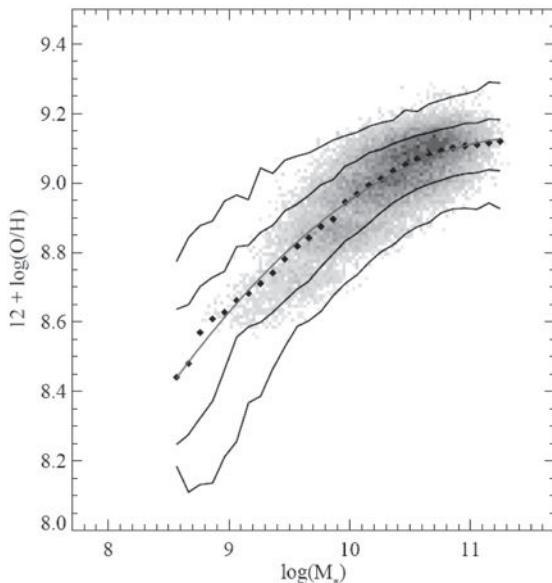


Fig. 3.10 The stellar mass–gas phase metallicity relation for 53 400 star-forming galaxies from the SDSS. The large black points represent the median in bins of 0.1 dex in mass which include at least 100 data points. The thin line through the data is a best-fitting smooth curve and the solid lines are the contours which enclose 68% and 95% of the data (Tremonti *et al.*, 2004).

feasible thanks to the development of efficient and reliable codes for determining the stellar and gaseous masses of galaxies from their optical spectra (Bruzual and Charlot, 2003; Charlot and Longhetti, 2001). It turns out that the correlation with stellar mass is stronger than that with luminosity. Figure 3.10 shows the strong correlation between metallicity and the total stellar mass of the galaxy of star-forming galaxies. These observations provide important constraints on the physics of the evolution of galaxies.

3.5 The masses of galaxies

The masses of galaxies can be measured using the *virial theorem* (2.22), which we have already encountered in the somewhat different context of the stability of stars under gravity (Sect. 2.3.1). This is such an important result that it is worthwhile rederiving it from purely dynamical arguments.

3.5.1 The virial theorem for galaxies and clusters

Suppose a system of particles (stars or galaxies), each of mass m_i , interact with each other only through their mutual forces of gravitational attraction. Then, the acceleration of the

i th particle due to all other particles can be written vectorially

$$\ddot{\mathbf{r}}_i = \sum_{j \neq i} \frac{Gm_j(\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|^3}. \quad (3.4)$$

Now, take the scalar product of both sides with $m_i \mathbf{r}_i$:

$$m_i(\mathbf{r}_i \cdot \ddot{\mathbf{r}}_i) = \sum_{j \neq i} Gm_i m_j \frac{\mathbf{r}_i \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|^3}. \quad (3.5)$$

Differentiating $(\mathbf{r}_i \cdot \mathbf{r}_i)$ with respect to time

$$\frac{d}{dt}(\mathbf{r}_i \cdot \mathbf{r}_i) = 2\dot{\mathbf{r}}_i \cdot \mathbf{r}_i, \quad (3.6)$$

and then, taking the next derivative,

$$\frac{1}{2} \frac{d^2}{dt^2}(\mathbf{r}_i^2) = \frac{d}{dt}(\dot{\mathbf{r}}_i \cdot \mathbf{r}_i) = (\ddot{\mathbf{r}}_i \cdot \mathbf{r}_i + \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i) = (\ddot{\mathbf{r}}_i \cdot \mathbf{r}_i + \dot{\mathbf{r}}_i^2). \quad (3.7)$$

Therefore, (3.4) can be rewritten

$$\frac{1}{2} \frac{d^2}{dt^2}(m_i \mathbf{r}_i^2) - m_i \dot{\mathbf{r}}_i^2 = \sum_{j \neq i} Gm_i m_j \frac{\mathbf{r}_i \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|^3}. \quad (3.8)$$

Now we sum over all the particles in the system,

$$\frac{1}{2} \frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i^2 - \sum_i m_i \dot{\mathbf{r}}_i^2 = \sum_i \sum_{j \neq i} Gm_i m_j \frac{\mathbf{r}_i \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|^3}. \quad (3.9)$$

The double sum on the right-hand side represents the sum over all the elements of a square $n \times n$ matrix with all the diagonal terms zero. If we sum the elements ij and ji of the matrix, we find

$$Gm_i m_j \left[\frac{\mathbf{r}_i \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|^3} + \frac{\mathbf{r}_j \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_j - \mathbf{r}_i|^3} \right] = -\frac{Gm_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (3.10)$$

Therefore,

$$\frac{1}{2} \frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i^2 - \sum_i m_i \dot{\mathbf{r}}_i^2 = -\frac{1}{2} \sum_{i,j} \frac{Gm_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (3.11)$$

where the factor $\frac{1}{2}$ on the right-hand side is included because the sum is over all elements of the array and so the sum of each pair would be counted twice.

Now, $\sum_i m_i \dot{\mathbf{r}}_i^2$ is twice the total kinetic energy, T , of all the particles in the system, that is,

$$T = \frac{1}{2} \sum_i m_i \dot{\mathbf{r}}_i^2. \quad (3.12)$$

The gravitational potential energy of the system is

$$U = -\frac{1}{2} \sum_{\substack{i,j \\ j \neq i}} \frac{G m_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (3.13)$$

Therefore,

$$\frac{1}{2} \frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i^2 = 2T - |U|. \quad (3.14)$$

If the system is in statistical equilibrium

$$\frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i^2 = 0, \quad (3.15)$$

and therefore

$$T = \frac{1}{2} |U|. \quad (3.16)$$

This is the equality known as the *virial theorem* in stellar dynamics. Notice that it is the same as the expression (2.22) which was derived adopting the equation of state of a perfect gas. Since the ratio of specific heat capacities for a perfect gas $\gamma = 5/3$ corresponds to counting only the degrees of freedom associated with the kinetic energy of motion of the particles, the equivalence of the two approaches is apparent.

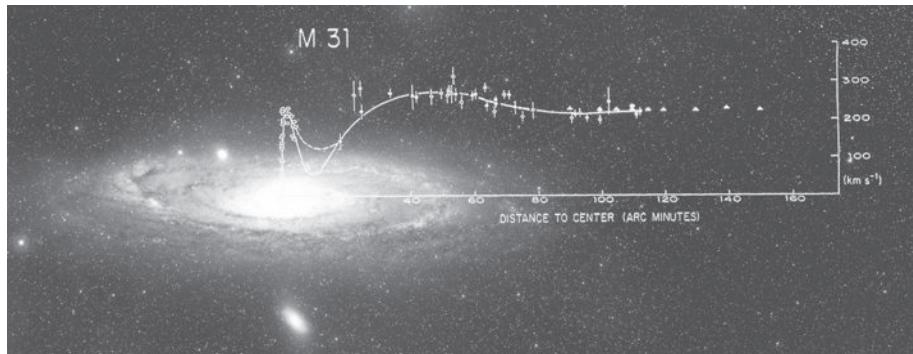
At no point in the above derivations have any assumptions been made about the orbits or velocity distributions of the particles. The velocities might be random, but the particles might also have highly elongated orbits about the centre of the galaxy. In the case of the discs of spiral galaxies, the velocity vectors of the stars are highly ordered and the mean rotational speed about the centre is much greater than the random velocities of the stars. The virial theorem applies to all cases provided the system is in dynamical equilibrium.

The application of the theorem to galaxies and clusters is not straightforward. Generally, only radial velocities can be measured from the Doppler shifts of the spectral lines. Assumptions also need to be made about the spatial and velocity distributions of stars in the galaxy or the galaxies in a cluster. If the velocity distribution is isotropic, the velocity dispersion is the same in the two perpendicular directions as along the line of sight and so $\langle v^2 \rangle = 3\langle v_{\parallel}^2 \rangle$, where v_{\parallel} is the radial velocity. If the velocity dispersion is independent of the masses of the stars or galaxies, the total kinetic energy is

$$T = \frac{1}{2} \sum_i m_i \dot{\mathbf{r}}_i^2 = \frac{3}{2} M \langle v_{\parallel}^2 \rangle, \quad (3.17)$$

where M is the total mass of the system. If the velocity dispersion varies with mass, $\langle v_{\parallel}^2 \rangle$ is a mass-weighted velocity dispersion. If the system is spherically symmetric, a suitably weighted mean separation R_{cl} can be estimated from the observed surface distribution of stars or galaxies and so the gravitational potential energy can be written $|U| = GM^2/R_{\text{cl}}$. The mass of the system is then

$$T = \frac{1}{2} |U|; \quad M = 3\langle v_{\parallel}^2 \rangle R_{\text{cl}}/G. \quad (3.18)$$

**Fig. 3.11**

The rotation curve for the nearby giant spiral galaxy M31, showing the flat rotation curve extending well beyond the optical image of the galaxy. (Courtesy of Dr. Vera Rubin.) The points beyond the optical image of the galaxy were obtained from radio observations of the 21-cm line of neutral hydrogen.

3.5.2 The rotation curves of spiral galaxies

The masses of spiral galaxies can be estimated from their *rotation curves*, the variation of the orbital, or rotational, speed $v_{\text{rot}}(r)$ about the centre of the galaxy with distance r from its centre. In a few galaxies, there is a well defined maximum in the rotation curve and the velocity of rotation decreases monotonically with increasing distance from the centre. In most cases, however, the rotational velocities in the outer regions of spiral galaxies are remarkably constant with increasing distance from the centre. Figure 3.11 shows that the flat rotation curve of our spiral neighbour M31, the Andromeda Nebula, extends far beyond the optical image of the galaxy and this is commonly found in spiral galaxies (Bosma, 1981).

Let us assume that the distribution of mass in the galaxy is spherically symmetric, so that we can write the mass within radius r as $M(\leq r)$. According to Gauss's law for gravity, for any spherically symmetric variation of mass with radius, we can find the radial acceleration at radius r by placing the mass within radius r , $M(\leq r)$, at the centre of the galaxy. Equating the centripetal acceleration to the gravitational acceleration,

$$\frac{GM(\leq r)}{r^2} = \frac{v_{\text{rot}}^2(r)}{r}; \quad M(\leq r) = \frac{v_{\text{rot}}^2(r)r}{G}. \quad (3.19)$$

For a point mass, $M(\leq r) = M_\odot$, and we recover Kepler's third law of planetary motion, the orbital period T being equal to $2\pi r/v_{\text{rot}} \propto r^{3/2}$. This result can also be written $v_{\text{rot}} \propto r^{-1/2}$ and is the variation of the circular rotational velocity expected in the outer regions of a galaxy if most of the mass is concentrated within the central regions.

If the rotation curve of the spiral galaxy is flat, $v_{\text{rot}} = \text{constant}$, $M(\leq r) \propto r$ and so the mass within radius r increases linearly with distance from the centre. This contrasts dramatically with the distribution of light in the discs, bulges and haloes of spiral galaxies which decrease exponentially with increasing distance from the centre. Consequently, *the local mass-to-luminosity ratio must increase in the outer regions of spiral galaxies*.

It is most convenient to quote the results in terms of mass-to-luminosity ratios relative to that of the Sun. For the visible parts of spiral galaxies, for which the rotation curves are well determined, mean mass-to-light ratios in the B waveband are in the range 1–10. This is similar to the value found in the solar neighbourhood; averaging over the masses and luminosities of the local stellar populations, a value of $M/L \approx 3$ is found. The M/L ratio must however increase to much larger values at large values of r . Values of $M/L \approx 10 - 20 M_\odot/L_\odot$ are found in the outer regions of spiral galaxies, similar to the values found for elliptical galaxies. These data provide crucial evidence for the presence of *dark matter in galaxies*.

There are theoretical reasons why spiral galaxies should possess dark haloes. Ostriker and Peebles showed that, without such a halo, a differentially rotating disc of stars is subject to a bar instability (Ostriker and Peebles, 1973). Their argument has been confirmed by subsequent computer simulations and suggests that dark haloes can stabilise the discs of spiral galaxies. Thus, although the initial assumption that the mass distribution in spiral galaxies should be spherically symmetric might have appeared to fly in the face of their disc-like properties, there are good reasons why the dominant contributor to the mass of these systems is a dark, roughly spherical halo.

3.5.3 The masses of elliptical galaxies

The virial theorem can also be used to estimate the masses of elliptical galaxies. Measurements of the Doppler broadening of the widths of stellar absorption lines in galaxies provide estimates of the velocity dispersion $\langle \Delta v_{\parallel}^2 \rangle$ of stars along the line of sight through the galaxy. Typical mass-to-luminosity ratios for elliptical galaxies found in this way lie in the range $10 - 20 M_\odot/L_\odot$.

The trouble with this argument is that it has been assumed that the velocity distribution of the stars in the elliptical galaxy is isotropic. In fact, there is compelling evidence that, in general, elliptical galaxies are triaxial systems, meaning that the velocity dispersions in the three orthogonal directions are different. It is not particularly surprising that this should be the case since the thermalisation time by gravitational encounters between stars for typical stellar systems is much longer than the age of the Universe. Therefore, although the system may well have reached a state of dynamical equilibrium, this does not necessarily mean that the velocity distribution has been randomised by collisions (see Sect. 5.6).

There is compelling observational evidence that elliptical galaxies are in fact triaxial systems. First of all, in many systems not only does the ellipticity of the isophotes of the surface brightness distribution vary with radius, but also the position angle of the major axis of the isophotes can change as well (Bertola and Galletta, 1979). A second piece of evidence is the observation that, in some ellipticals, rotation takes place along the minor as well as along the major axis (Bertola *et al.*, 1991). Thirdly, the flattening of the elliptical galaxies is too great to be explained by the rotation of an axisymmetric distribution of stars with an isotropic velocity distribution at each point within the galaxy (Davies *et al.*, 1983). Figure 3.12 shows the ellipticities ε of elliptical galaxies as a function of their rotational velocities v_m ; $\bar{\sigma}$ is the velocity dispersion of the stars in the galaxies. The open circles represent luminous elliptical galaxies, the filled circles lower luminosity ellipticals and the crosses the bulges of spiral galaxies. If the ellipticity were entirely due to rotation with an

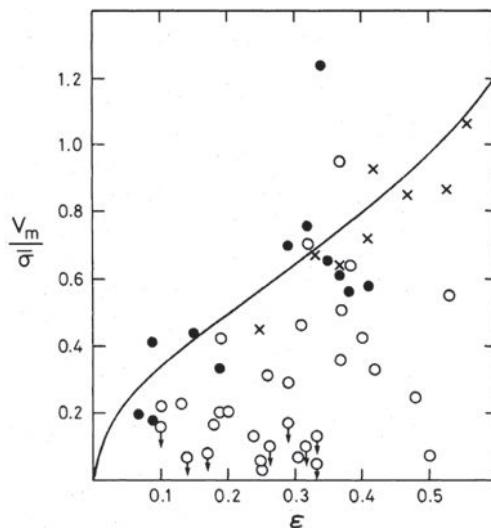


Fig. 3.12 A diagram showing the flattening of elliptical galaxies as a function of their rotational velocities. The open circles are luminous elliptical galaxies, the filled circles are lower luminosity ellipticals and the crosses are the bulges of spiral galaxies. If the ellipticity were entirely due to rotation with an isotropic stellar velocity distribution at each point, the galaxies would be expected to lie along the solid lines. This diagram shows that, at least for massive ellipticals, this simple picture of rotational flattening cannot be correct (Davies *et al.*, 1983).

isotropic stellar velocity distribution throughout the galaxy, the points would be expected to lie along the solid line. Figure 3.12 shows that massive elliptical galaxies are not rotating fast enough to account for the observed flattening. Thus, application of the virial theorem can be potentially misleading. Furthermore, these triaxial systems are stable. Schwarzschild showed that there exist stable triaxial configurations not dissimilar from those necessary to explain some of the internal dynamical properties of what appear on the surface to be simple ellipsoidal stellar distributions (Schwarzschild, 1979). His analysis showed that there exist stable orbits about the major and minor axes but not about the immediate axis of the triaxial figure.

Evidence that there must indeed be considerable amounts of dark matter in the haloes about two of the giant elliptical galaxies in the Virgo Cluster, M49 and M87, has been presented by Côté and his colleagues (Côté *et al.*, 2001, 2003). They measured the radial velocities of a large sample of globular clusters in the haloes of these galaxies, some of which can be seen in Fig. 3.2a, and so were able to extend the range of radii over which the velocity dispersion in these galaxies could be measured. Their measurements for M49 are shown by the filled circles at radii $R \geq 10$ kpc in Fig. 3.13. The points at radii less than 10 kpc show the velocity dispersion measured by other authors and it can be seen that the data are consistent with the velocity dispersion remaining remarkably constant out to radii up to 40 kpc from the centre. Various attempts to account for the variation of the velocity dispersion with radius are indicated by the different lines on the diagram in which it is assumed that the mass distribution follows the radial optical intensity distribution, but with various extreme assumptions about the anisotropy of the stellar velocity distribution.

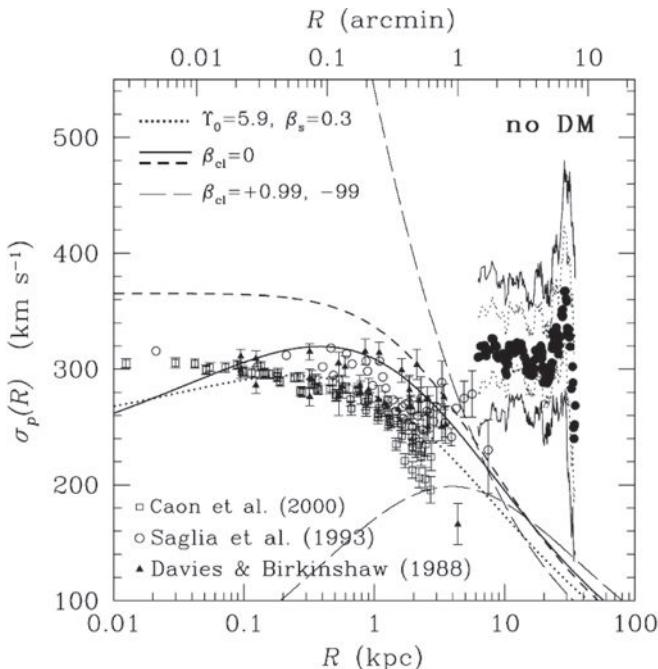


Fig. 3.13 The velocity dispersion of stars and globular clusters in the nearby giant elliptical galaxy M49 (NGC 4472). The data points at $R < 10$ kpc are obtained from the velocity width of the stellar absorption lines. The filled circles at radii $R > 10$ kpc are derived from the velocity dispersion of globular clusters. The dotted and solid lines bracketing these points show the one and two sigma ranges of their estimates of the velocity dispersion. Various models for velocity dispersion assuming that the mass follows the light are shown (Côté *et al.*, 2003).

Even models in which the globular clusters are on radial orbits cannot account for the independence of the line-of-sight velocity dispersion out to 40 kpc. Côté and his colleagues concluded that these data provide evidence that the velocity dispersion is isotropic and that there must be dark matter haloes about these galaxies. The fact that the velocity dispersion remains constant out to large radii has exactly the same explanation as the flatness of the rotation curves of spiral galaxies, expression (3.18). The mass within radius R must increase proportional to R .

3.6 The luminosity function of galaxies

The frequency with which galaxies of different intrinsic luminosities L are found in space is described by the *luminosity function* of galaxies, $\phi(L) dL$, which is defined to be the space density of galaxies with intrinsic luminosities in the range L to $L + dL$. The luminosity function of galaxies derived from a sample of 221 414 galaxies observed in the 2dF galaxy survey is shown in Fig. 3.14, which also shows the separation of the function into those for

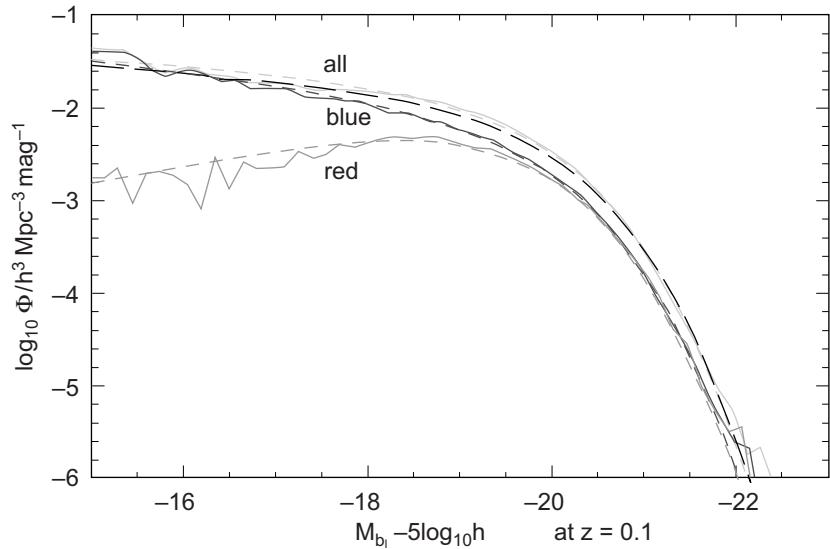


Fig. 3.14 The luminosity function of galaxies derived from a sample of 221 414 galaxies observed in the 2dF galaxy survey. The overall luminosity function and those of the red and blue galaxies in the sample have been fitted by Schechter luminosity functions (Cole *et al.*, 2005).

red and blue galaxies (Cole *et al.*, 2005). The lines show best-fits to a luminosity function of the form

$$\phi(x) dx = \phi^* x^\alpha e^{-x} dx , \quad (3.20)$$

where $x = L/L^*$ and L^* characterises the ‘break’ in the luminosity function. This form of function is known as a *Schechter luminosity function* and consists of a power law with slope α and a high luminosity exponential cut-off at luminosities greater than the ‘break’ luminosity L^* .

It is traditional in optical astronomy to write the luminosity function in terms of astronomical magnitudes rather than luminosities and then the simplicity of the Schechter function is somewhat spoiled:

$$\Phi(M) dM = \frac{2}{5} \phi^* \ln 10 \{ \text{dex}[0.4(M^* - M)] \}^{\alpha+1} \times \exp\{-\text{dex}[0.4(M^* - M)]\} dM , \quad (3.21)$$

where M^* is the absolute magnitude corresponding to the luminosity L^* . We have used the notation $\text{dex } y$ to mean 10^y . The values of the parameters for the 2dF galaxy survey, which was carried out in the b_J waveband, were $\alpha = -1.18$, $M^* = -19.52 + 5\log_{10}h$ and $\phi^* = 0.0156 h^3 \text{ Mpc}^{-3}$. These values are not so different from the values derived by Felten in his heroic analysis of the luminosity function in the B waveband: $\alpha = -1.25$, $M^* = -20.05 + 5\log_{10}h$ and $\phi^* = 0.012 h^3 \text{ Mpc}^{-3}$ (Felten, 1985). Notice that, as expected from the histograms of Fig. 3.5, the luminosity function is dominated by the blue galaxies at low luminosities.

3.6.1 The luminosity density of starlight in the Universe

An important calculation is the integrated luminosity of all the galaxies within a given volume of space, in other words, the luminosity density of the radiation due to starlight in the Universe. The luminosity density is

$$\varepsilon_{B(0)} = \int_0^\infty L \phi(L) dL = \phi^* L^* \int_0^\infty x^{a+1} e^{-x} dx = \phi^* L^* \Gamma(a+2), \quad (3.22)$$

where Γ is the gamma function. Using the values determined by Felten for the field luminosity function quoted above,

$$\varepsilon_{B(0)} = 1.8 \times 10^8 h L_\odot \text{ Mpc}^{-3}. \quad (3.23)$$

The value found from the SDSS luminosity function (Blanton *et al.*, 2003) in the $^{0.1}r$ waveband is

$$(1.84 \pm 0.04) \times 10^8 h L_\odot \text{ Mpc}^{-3}. \quad (3.24)$$

This result is consistent with other estimates of the luminosity density, for example from the Two-Degree Field Galaxy Redshift Survey and the Millennium Galaxy Catalogue.

3.6.2 The mass-to-luminosity ratio for the Universe

A useful reference value for cosmological studies is the average mass-to-luminosity ratio for the Universe, if it is assumed to have the critical cosmological density, $\varrho_c = 3H_0^2/8\pi G = 2.0 \times 10^{-26} h^2 \text{ kg m}^{-3}$. In terms of solar units, the mass-to-luminosity ratio would be

$$\frac{\varrho_c}{\varepsilon_B} = \left(\frac{M}{L} \right)_B = 1600 h \left(\frac{M_\odot}{L_\odot} \right)_B. \quad (3.25)$$

Although there is some variation about this estimate, its importance lies in the fact that it is significantly greater than the typical mass-to-luminosity ratios of galaxies and clusters of galaxies, even when account is taken of the dark matter which must be present. This result indicates that the mass present in galaxies and clusters of galaxies is not sufficient to close the Universe.

3.6.3 Useful statistics about galaxies

It is convenient to have available values for the mean space density and luminosity of galaxies. If $a = -1.25$, $\langle L \rangle = 1.25 L^* = 1.55 \times 10^{10} h^{-2} L_\odot$. Adopting the mean luminosity density of the Universe, the typical number density of galaxies is

$$\bar{n} = \varepsilon_{B(0)}/\langle L \rangle = 10^{-2} h^3 \text{ Mpc}^{-3}. \quad (3.26)$$

Thus, the typical galaxies which contribute most of the integrated light of galaxies are separated by a distance of about $5h^{-1}$ Mpc, if they were uniformly distributed in space. Galaxies such as our own and M31 have luminosities $L_{\text{Gal}}(\text{B}) \approx 10^{10} L_\odot$.

These data enable limits to be placed upon the average mass density in stars at the present epoch. Adopting a typical mass-to-luminosity ratio for the *visible* parts of galaxies

of $M/L \approx 3$ the *density parameter in stars* $\Omega_* h = \rho_*/\rho_c$ at the present epoch would be $\Omega_* h = 2 \times 10^{-3}$. A very much more careful analysis using the combined SDSS and *Two Micron All-Sky Survey* (2MASS) catalogues of galaxies provides an upper limit to the stellar mass density in the local Universe (Bell *et al.*, 2003):

$$\Omega_* h = \rho_*/\rho_c = (2 \pm 0.6) \times 10^{-3}. \quad (3.27)$$

This value can be compared with the concordance value of the mean baryonic mass density in the Universe which can be derived independently from primordial nucleosynthesis arguments and from analysis of the power spectrum of fluctuations in the Cosmic Microwave Background Radiation, $\Omega_{\text{baryon}} h^2 = 0.0223$ (Longair, 2008). Thus, there is much more baryonic matter in the Universe than would be inferred from the light of galaxies. Most of it must be in forms which are not detectable as starlight.

Associations of galaxies range from pairs and small groups, through giant clusters containing over a thousand galaxies, to the vast structures on scales much greater than clusters such as the vast ‘walls’ and voids observed in the distribution of galaxies. Clustering occurs on all scales and very few galaxies can be considered truly isolated. Rich clusters of galaxies are of particular interest because they are the largest gravitationally bound systems in the Universe. The gravitational potential of the cluster is defined by the distribution of dark matter, the mass of which greatly exceeds that of the baryonic matter, such as that contained in the stars in galaxies and the associated interstellar gas and the intracluster gas. The deep gravitational potential wells of clusters can be observed directly through the bremsstrahlung X-ray emission of hot intracluster gas which forms a hydrostatic atmosphere within the cluster. The hot gas can also be detected through the decrements which it causes in the Cosmic Microwave Background Radiation as a result of the Sunyaev–Zeldovich effect. Gravitational lensing has proved to be a very powerful tool for defining the large scale distribution of dark matter in clusters, as well as in individual galaxies within them. Interactions of galaxies with each other and with the intergalactic medium in the cluster can be studied and radio source events can strongly perturb the distribution of hot gas.

Clusters of galaxies, therefore, provide laboratories for studying many different aspects of galactic evolution and the role of high energy astrophysical phenomena within rather well-defined astronomical environments.

4.1 The morphologies of rich clusters of galaxies

Rich clusters of galaxies are of particular importance in this study. Much of the pioneering effort was carried out by Abell, who was one of the principal observers for the 48-inch Schmidt Telescope Palomar Observatory Sky Survey. While the plates were being taken, he systematically catalogued the rich clusters of galaxies appearing on the plates, the word ‘rich’ meaning that there was no doubt as to the reality of the clusters (Abell, 1958). The cluster Abell 2218 and the nearby Virgo Cluster of galaxies are shown in Fig. 4.1. A corresponding catalogue for the southern hemisphere was created with the completion of the ESO-SERC Southern Sky Survey (Abell *et al.*, 1989). In both cases, the clusters were discovered by visual inspection of the Sky Survey plates. Crucial to the success of Abell’s programme was his adherence to a strict set of criteria for the inclusion of clusters



Fig. 4.1 (a) Abell 2218, a rich regular cluster. There is a supergiant cD galaxy in the centre. The image also shows a number of arcs which are the gravitationally lensed images of very distant background galaxies. (Courtesy NASA, ESA and the Space Telescope Science Institute.) (b) The nearby Virgo Cluster of galaxies is classified as an irregular cluster.

in the catalogue. These included richness, compactness and distance criteria,¹ which have proved to be remarkably robust when compared with more recent algorithmic approaches to cluster classification, for example, using the digital data from the Sloan Digital Sky Survey (Bahcall *et al.*, 2003b).

The combined sample of rich clusters is complete to a distance of about $600h^{-1}$ Mpc, corresponding to redshift $z = 0.2$ and there is good agreement between number densities of rich clusters in the northern and southern hemispheres. The space density of Abell clusters

¹ Many more details of these criteria, the statistics of clusters of different richness and many other properties of clusters of galaxies are included in Chapter 4 of *Galaxy Formation* (Longair, 2008).

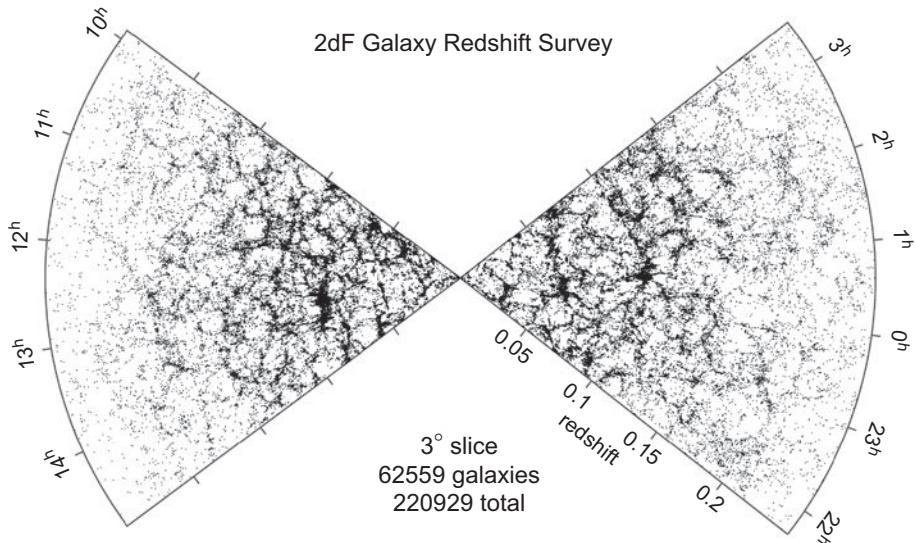


Fig. 4.2 The cosmic web as defined by the AAT 2dF survey of galaxies. The distribution of galaxies is complete out to redshifts $z \approx 0.2$ and contains 62 559 galaxies within a 3° wedge on the sky in both the northern and southern galactic hemispheres (Colless *et al.*, 2001). (Image courtesy of the 2dF Galaxy Redshift Survey team.)

with richness classes $R \geq 1$ is

$$N_{\text{cl}}(R \geq 1) \approx 10^{-5} h^3 \text{ Mpc}^{-3}. \quad (4.1)$$

Therefore, the typical distance between centres of rich clusters, if they were uniformly distributed in space, would be $\sim 50h^{-1}$ Mpc. This figure can be compared with the space density of ‘mean galaxies’ of $10^{-2} h^3 \text{ Mpc}^{-3}$ and their typical separations of $5h^{-1}$ Mpc (see Sect. 3.6.3).

Abell clusters themselves are strongly correlated in space, both with each other and with the distribution of galaxies in general. These associations were originally described in terms of the superclustering of galaxies by Abell and Zwicky. Some impression of the relation between the rich clusters of galaxies and the general distribution of galaxies in the Universe can be obtained from the ‘cone diagram’ of the distribution of galaxies obtained from the AAT 2dF galaxy survey (Fig. 4.2). In this image, the positions of each galaxy within a wedge of angle 4° are plotted as a function of redshift. If the distribution of galaxies in space were uniform, the points would be uniformly scattered over the region within which the sample is complete, in this case, out to redshifts $z \approx 0.2$. On the contrary, Fig. 4.2 shows that their distribution is highly inhomogeneous with the galaxies concentrated into sheets or filaments with huge holes or voids in between, the largest voids being about $50h^{-1}$ Mpc in diameter. This ‘sponge-like’ distribution of galaxies is often referred to as the *cosmic web*. The rich clusters are generally found in the densest regions of the cosmic web, for example, where the giant walls intersect. These features of the galaxy distribution can be quantified in terms of cross-correlation functions between the distribution of clusters and galaxies in general (Bahcall *et al.*, 2003a).

Table 4.1 The properties of rich clusters of galaxies (Bahcall, 1977).

Property/Class	Regular	Intermediate	Irregular
Bautz–Morgan type	I, I-II, II	(II), II-III	(II-III), III
Galaxy content	Elliptical/S0 rich	Spiral-poor	Spiral-rich
E : S0 : S ratio	3 : 4 : 2	1 : 4 : 2	1 : 2 : 3
Symmetry	Spherical	Intermediate	Irregular shape
Central concentration	High	Moderate	Very little
Central profile	Steep gradient	Intermediate	Flat gradient
Mass segregation	Marginal evidence for $m - m(1) < 2$	Marginal evidence for $m - m(1) < 2$	No segregation
Examples	Abell 2199, Coma	Abell 194, 539	Virgo, Abell 1228

Rich clusters of galaxies can be broadly classified as regular, intermediate and irregular. In order to refine the morphological description of clusters, various classification schemes have been proposed to describe different aspects of their properties. These include:

Bautz–Morgan types I, II, III In type I clusters there is a dominant central galaxy, often a cD galaxy, which is much brighter than the next brightest cluster galaxies. In type III, there is no dominant galaxy.

Galaxy content The types of galaxy in a cluster can be characterised by the relative number of elliptical, S0 and spiral galaxies. These were described by Oemler as elliptical/S0 rich, spiral-poor, spiral-rich (Oemler, 1974).

Symmetry The shapes of the clusters can be described as spherical, intermediate or irregular.

Central concentration of the galaxy distribution This is described as high, moderate or very little.

Central profile The radial gradient of the number density of galaxies can be described as steep, intermediate or flat

Mass segregation In some clusters, the most massive galaxies are located preferentially towards the centre; in others there is little or no mass segregation as a function of radius.

Table 4.1 shows that there are clear correlations between the properties of regular, intermediate and irregular clusters and the above characteristics. The different properties of clusters largely reflect whether or not they have had time to evolve to a quasi-static density distribution, in other words, whether they are relatively young or old dynamically.

4.2 Clusters of galaxies and isothermal gas spheres

In regular clusters, the space density of galaxies increases towards their central *cores*. Outside the core, the space density of galaxies decreases steadily until it disappears into the

background of unrelated objects. It turns out that the spatial distribution of galaxies in such clusters can be modelled by the distribution of mass in an *isothermal gas sphere*. The term *isothermal* means that the temperature, or mean kinetic energy of the particles, is constant throughout the cluster. In the case of clusters, this means that the velocity distribution of the galaxies is Maxwellian with the same mean kinetic energy per galaxy throughout the cluster. If all the galaxies had the same mass, the velocity dispersion would be the same at all locations within the cluster. Although the galaxies in regular clusters have certainly had time to virialise, that is, to come into dynamical equilibrium according to the virial theorem, it takes very much longer for energy exchange by gravitational encounters between galaxies to establish a Maxwellian distribution of velocities. Nonetheless, let us work out the density distribution of an isothermal gas sphere as a reference model for comparison with the observations.

We begin with the equations of hydrostatic support and mass conservation (2.6), which are repeated here for convenience:

$$\frac{dp}{dr} = -\frac{GM\varrho}{r^2}; \quad \frac{dM}{dr} = 4\pi r^2 \varrho. \quad (4.2)$$

Reordering the first equation of (4.2) and differentiating with respect to r ,

$$\begin{aligned} \frac{r^2}{\varrho} \frac{dp}{dr} &= -GM, & \frac{d}{dr} \left(\frac{r^2}{\varrho} \frac{dp}{dr} \right) &= -G \frac{dM}{dr}, \\ \frac{d}{dr} \left(\frac{r^2}{\varrho} \frac{dp}{dr} \right) + 4\pi Gr^2 \varrho &= 0. \end{aligned} \quad (4.3)$$

Equation (4.3) is known as the *Lane–Emden equation*. The pressure p and the density ϱ are related by the perfect gas law at all radii r , $p = \varrho kT/\mu$ and $\frac{3}{2}kT = \frac{1}{2}\mu\langle v^2 \rangle$, where μ is the mass of an atom, molecule or galaxy and $\langle v^2 \rangle$ their mean square velocity. Therefore, substituting for p ,

$$\frac{d}{dr} \left(\frac{r^2}{\varrho} \frac{d\varrho}{dr} \right) + \frac{4\pi G\mu}{kT} r^2 \varrho = 0. \quad (4.4)$$

Equation (4.4) is a nonlinear differential equation and, in general, is solved numerically. There is, however, a useful analytic solution for large values of r . If $\varrho(r)$ is expressed as a power series in r , $\varrho(r) = \sum A_n r^{-n}$, there is a solution for large r with $n = 2$,

$$\varrho(r) = \frac{2}{Ar^2} \quad \text{where} \quad A = \left(\frac{4\pi G\mu}{kT} \right). \quad (4.5)$$

This mass distribution has the regrettable property that the total mass of the cluster diverges at large values of r ,

$$\int_0^\infty 4\pi r^2 \varrho(r) dr = \int_0^\infty \frac{8\pi}{A} dr \rightarrow \infty. \quad (4.6)$$

There are, however, reasons why there should be a cut-off to the distribution at large radii. First of all, at very large distances, the particle densities become so low that the mean free path between collisions is very long. The thermalisation time-scales consequently become greater than the time-scale of the system. The radius at which this occurs is known as *Smoluchowski's envelope*. Secondly, in astrophysical systems, the outermost stars or

Table 4.2 The density distribution $y(x)$ and the projected density distribution $N(q)$ for an isothermal gas sphere.

x, q	$y(x)$	$N(q)$	x, q	$y(x)$	$N(q)$
0	1.0	1.0	12	0.0151	0.0839
0.5	0.9597	0.9782	14	0.0104	0.0694
1.0	0.8529	0.9013	16	0.0075	0.0591
1.5	0.7129	0.8025	20	0.0045	0.0457
2	0.5714	0.6955	30	0.0019	0.0313
3	0.3454	0.5033	40	0.0010	0.0229
4	0.2079	0.3643	50	0.0007	0.0188
5	0.1297	0.2748	100	1.75×10^{-4}	0.0101
6	0.0849	0.2143	200	5.08×10^{-5}	0.0053
7	0.0583	0.1724	300	2.32×10^{-5}	0.0036
8	0.0418	0.1420	500	8.40×10^{-6}	0.0021
9	0.0311	0.1209	1000	2.0×10^{-6}	0.0010
10	0.0238	0.1050			

galaxies are stripped from the cluster by tidal interactions with neighbouring systems. Therefore, if clusters are modelled by isothermal gas spheres, it is perfectly permissible to introduce a cut-off at some suitable large *tidal radius* r_t , resulting in a finite total mass.

It is convenient to rewrite (4.4) in dimensionless form by writing $\varrho = \varrho_0 y$, where ϱ_0 is the central mass density, and introducing a *structural index* or *structural length* α , where α is defined by the relation

$$\alpha = \frac{1}{(A\varrho_0)^{1/2}} . \quad (4.7)$$

Distances from the centre can then be measured in terms of the dimensionless distance $x = r/\alpha$. Then, (4.4) becomes

$$\frac{d}{dx} \left[x^2 \frac{d(\log y)}{dx} \right] + x^2 y = 0 . \quad (4.8)$$

Two versions of the solution of (4.8) are listed in Table 4.2. In column 2, the solution of y as a function of distance x is given; in the third column, the projected distribution onto a plane is given, this being the observed distribution of a cluster of stars or galaxies on the sky. If q is the projected distance from the centre of the cluster, the surface density $N(q)$ is related to $y(x)$ by the integral

$$N(q) = 2 \int_q^\infty \frac{y(x)x}{(x^2 - q^2)^{1/2}} dx . \quad (4.9)$$

Inspection of Table 4.2 shows that α is a measure of the size of the core of the cluster. Fitting the projected distribution $N(q)$ to the distribution of stars or galaxies in a cluster, a *core radius* can be defined as that radius at which the projected density falls to half the central value. The value $N(q) = 1/2$ is found at $q = 3$ and so $R_{1/2} = 3\alpha$ is a convenient measure of the core radius of the cluster.

Having measured $R_{1/2}$, the central mass density of the cluster can be found if the velocity dispersion of the galaxies in this region is also known. From Maxwell's equipartition theorem, $\frac{1}{2}\mu\langle v^2 \rangle = \frac{3}{2}kT$ and therefore, from the definition of α ,

$$\alpha^2 = \frac{1}{A\varrho_0} = \frac{kT}{4\pi G\mu\varrho_0} = \frac{\langle v^2 \rangle}{12\pi G\varrho_0}. \quad (4.10)$$

Observationally, we can only measure the radial component of the galaxies' velocities v_{\parallel} . Assuming the velocity distribution of the galaxies in the cluster is isotropic,

$$\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle = 3\langle v_{\parallel}^2 \rangle. \quad (4.11)$$

Expressing the central density ϱ_0 in terms of $R_{1/2}$ and $\langle v_{\parallel}^2 \rangle$,

$$\varrho_0 = \frac{9\langle v_{\parallel}^2 \rangle}{4\pi G R_{1/2}^2}. \quad (4.12)$$

Thus, assuming the central density of a cluster can be represented by an isothermal gas sphere, we can find its central mass density by measuring $\langle v_{\parallel}^2 \rangle$ and $R_{1/2}$.

Improved versions of the isothermal sphere model were evaluated by King and these have been found to provide good fits to the light distributions of globular clusters, galaxies and regular clusters of galaxies (King, 1966, 1981). The models were derived from studies of solutions of the Fokker–Planck equation for the distribution function $f(v, r)$ of the stars in a cluster under the condition that there should be no particles present with velocities which enable them to escape from the cluster. This might occur for two reasons. Either the stars have velocities which exceed the escape velocity from the cluster, or the stars travel to distances greater than the tidal radius of the cluster when they are lost from the cluster. In either case, the cluster can be modelled as a *truncated isothermal gas sphere* in which none of the stars can have velocities exceeding some value v_e . This is implemented by truncating the Maxwell velocity distribution at this velocity which in turn results in models with finite tidal radii r_t . The luminosity profiles, equivalent to $N(q)$ for such clusters, are shown in Fig. 4.3, the models being parameterised by the quantity $\log r_t/r_c$, the logarithm of the ratio of the tidal and core radii. In the limit $r_t/r_c \rightarrow \infty$, the models become isothermal gas spheres.

According to Bahcall, the observed distribution of galaxies in regular clusters can be described by truncated isothermal distributions $N(r)$ of the form

$$N(r) = N_0[f(r) - C], \quad (4.13)$$

where $f(r)$ is the projected isothermal distribution normalised to $f(r) = 1$ at $r = 0$ and C is a constant which reduces the value of $N(r)$ to zero at some radius R_h such that $f(R_h) = C$ (Bahcall, 1977). For regular clusters core radii lie in the range $R_{1/2} = 150\text{--}400$ kpc, the Coma Cluster having $R_{1/2} = 220$ kpc. Bahcall found that there is a relatively small dispersion in the values of C required to provide a satisfactory fit to the profiles of many regular clusters, typically the value of C corresponding to about 1.5% of the isothermal central density.

Other density distributions have been proposed to describe the space density distribution of galaxies in clusters. These include de Vaucouleurs' law for elliptical galaxies (3.1) as

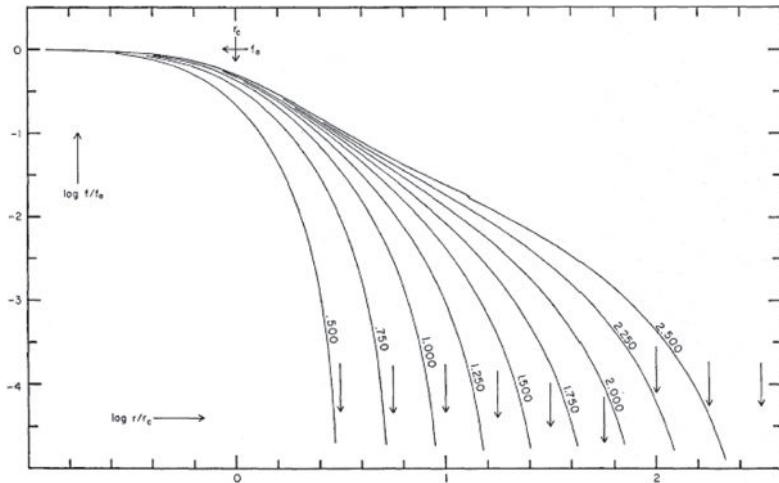


Fig. 4.3 King models for the distribution of stars in globular clusters, galaxies or clusters of galaxies (King, 1966, 1981). The curves show the projected distribution of stars or galaxies, equivalent to $N(q)$ in Table 4.2, and are parameterised by the quantity $\log(r_t/r_c)$ where r_t is the tidal radius and r_c the core radius. The arrows indicate $\log r_t$.

well as other possibilities such as the Plummer model which is derived from a gravitational potential with a core radius b of the form

$$\phi = -\frac{GM}{(r^2 + b^2)^{1/2}}, \quad (4.14)$$

where M is the total mass of the system. Using Poisson's law for gravity in spherical polar coordinates,

$$\nabla^2\phi = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) = 4\pi G \varrho, \quad (4.15)$$

the density distribution is found to be

$$\varrho(r) = \frac{3M}{4\pi b^3} \left(1 + \frac{r^2}{b^2} \right)^{-5/2}. \quad (4.16)$$

Binney and Tremaine discuss these and other possible distributions (Binney and Tremaine, 2008).

4.3 The Coma Cluster of galaxies

Let us apply these concepts to the Coma Cluster of galaxies, Abell 1656. The Coma Cluster is a rich regular cluster at redshift $z = 0.0231$ for which a large amount of data is available on the radial velocities of the galaxies and their projected number density as a function of radius. The surface density distribution of galaxies in the cluster and the variation of their

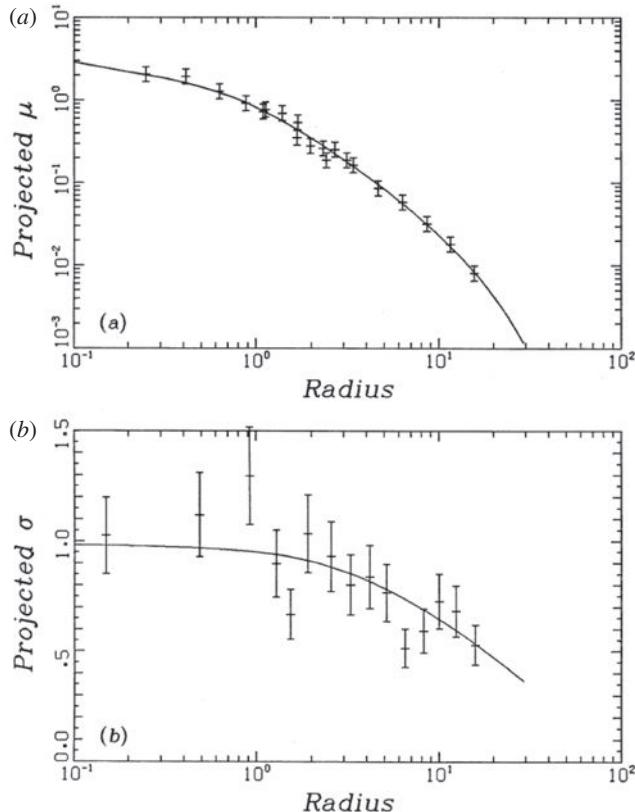
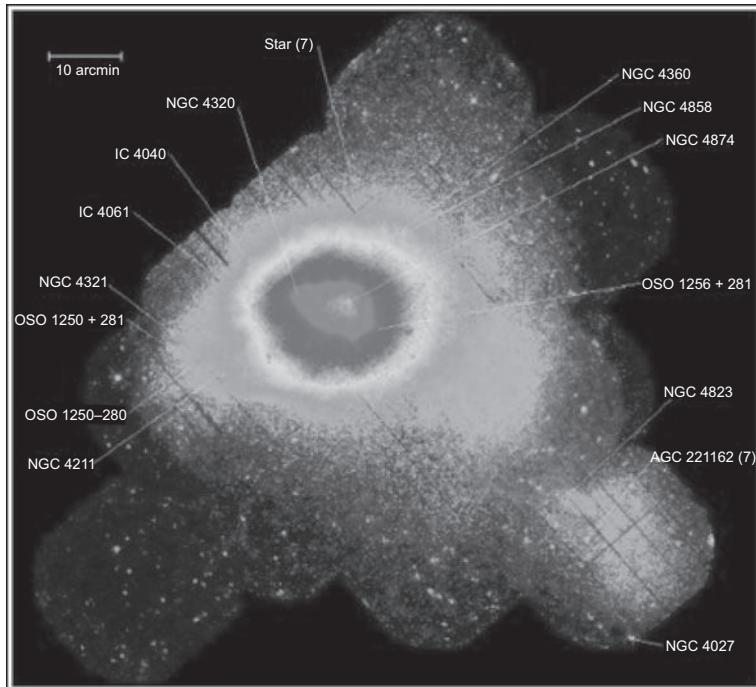


Fig. 4.4 (a) The surface density profile for the distribution of galaxies in the Coma Cluster according to Kent and Gunn. (b) The projected velocity dispersion as a function of radius for galaxies in the Coma Cluster (Kent and Gunn, 1982).

velocity dispersion with radius are shown in Fig. 4.4 from the analysis of Kent and Gunn, who obtained radial velocities for about 300 cluster members (Kent and Gunn, 1982). The projected surface density of galaxies is satisfactorily described by a King profile with tidal radius $r_t = 16h^{-1}$ Mpc. The assumption that the cluster has attained a relaxed, bound equilibrium configuration is confirmed from an estimate of the *crossing time* of a typical galaxy in the cluster. The crossing time is defined to be $t_{\text{cr}} = R/\langle v^2 \rangle^{1/2}$ where R is the size of the cluster and $t_{\text{cr}} = R/\langle v^2 \rangle^{1/2}$ is the root mean square velocity of galaxies in the cluster. For the Coma Cluster, $\langle v^2 \rangle^{1/2} \approx 10^3$ km s $^{-1}$ and $R \approx 2$ Mpc and so the crossing time is about 2×10^9 years, roughly a tenth the age of the Universe. Therefore, the cluster must be gravitationally bound.

These data were further analysed by Merritt who assumed first of all that the overall mass distribution in the cluster follows the galaxy distribution, that is, the mass-to-luminosity ratio is a constant throughout the cluster, and that the velocity distribution is everywhere isotropic (Merritt, 1987). The mass of the Coma Cluster was found to be $1.79 \times 10^{15} h^{-1} M_\odot$, assuming that the cluster extends to $16h^{-1}$ Mpc. The mass within a radius $1h^{-1}$ Mpc of the cluster centre is $6.1 \times 10^{14} M_\odot$ and the mass-to-blue luminosity ratio in the cluster core

**Fig. 4.5**

An X-ray image of the Coma Cluster of galaxies obtained by the XMM-Newton Observatory, showing the X-ray emission associated with the main body of the Coma Cluster and the smaller cluster associated with NGC 4839. (Courtesy of the Max Planck Institute for Extraterrestrial Physics and ESA.)

about $350h M_{\odot}/L_{\odot}$. The population of galaxies in the central region of the Coma Cluster is dominated by elliptical and S0 galaxies, for which the typical mass-to-luminosity ratios are about $10-20 M_{\odot}/L_{\odot}$. This discrepancy of about a factor of 20 between the mass associated with the visible parts of galaxies and the total mass is attributed to the presence of *dark matter* in the cluster.

This result is subject to the same concerns which were discussed in the context of estimating the masses of elliptical galaxies (Sect. 3.5.3). In his careful analysis, Merritt concluded that, even making extreme assumptions about the anisotropy of the velocity distribution of the galaxies in the cluster, the inferred mass-to-luminosity ratio only varied from about 0.4 to at least three times the reference value of $350h M_{\odot}/L_{\odot}$, while the mass-to-luminosity ratio within the core of the cluster, $r \leq 1 h^{-1} \text{ Mpc}$, was always very close to this value. There can be no doubt that the dynamics of the cluster are dominated by dark matter.

More recently, it has been shown that the Coma Cluster is probably not quite the quiescent regular cluster it appears to be. Colless and Dunn added 243 more radial velocities to the sample, bringing the total number of cluster members with radial velocities to 450 (Colless and Dunn, 1996). They found that, in addition to the main body of the cluster, there is a distinct subcluster, the brightest member of which is NGC 4839. The main cluster has mass $0.9 \times 10^{15} h^{-1} M_{\odot}$, while the less massive cluster has mass $0.6 \times 10^{14} h^{-1} M_{\odot}$. These clusters are clearly seen in the XMM-Newton X-ray image of the Coma Cluster (Fig. 4.5).

The masses derived from the X-ray observations agree with those derived by Colless and Dunn, who inferred that the subcluster is in the process of merging with the main body of the Coma Cluster.

4.4 Mass distribution of hot gas and dark matter in clusters

The X-ray image of the Coma Cluster (Fig. 4.5) demonstrates the power of X-ray astronomy in the study of clusters of galaxies. Intense X-ray emission is a common feature of rich clusters of galaxies, the emission being the bremsstrahlung of hot intracluster gas, as inferred from the extended nature of the emission and from the detection of the highly ionised iron line Fe xxvi in their X-ray spectra (Mitchell *et al.*, 1976). These X-ray observations provide a very powerful probe of the gravitational potential of the cluster enabling the distribution of both the hot gas and the total gravitating mass to be determined (Fabricant *et al.*, 1980).

The cluster is assumed to be spherically symmetric and the gas in hydrostatic equilibrium within the gravitational potential defined by the total mass distribution in the cluster, that is, by the sum of the visible and dark matter as well as the intracluster gaseous mass. If p is the pressure of the gas and ϱ its density, both of which vary with position within the cluster, the requirement of hydrostatic equilibrium is again (2.3):

$$\frac{dp}{dr} = -\frac{GM(\leq r)\varrho}{r^2}. \quad (4.17)$$

The pressure is related to the local gas density ϱ and temperature T by the perfect gas law $p = \varrho kT / \mu m_H$, where m_H is the mass of the hydrogen atom and μ is the mean molecular weight of the gas. For a fully ionised gas with the standard cosmic abundance of the elements, a suitable value is $\mu = 0.6$. Differentiating the perfect gas law with respect to r and substituting into (4.17),

$$\frac{\varrho kT}{\mu m_H} \left(\frac{1}{\varrho} \frac{d\varrho}{dr} + \frac{1}{T} \frac{dT}{dr} \right) = -\frac{GM(\leq r)\varrho}{r^2}. \quad (4.18)$$

Reorganising (4.18),

$$M(\leq r) = -\frac{kTr^2}{G\mu m_H} \left[\frac{d(\log \varrho)}{dr} + \frac{d(\log T)}{dr} \right]. \quad (4.19)$$

Thus, the overall mass distribution within the cluster can be determined if the variation of the gas density and temperature with radius are known. Assuming the cluster is spherically symmetric, these can be derived from high sensitivity X-ray intensity and spectral observations. A suitable form for the bremsstrahlung spectral emissivity of a plasma, which will be derived in Sect. 6.5, is

$$\kappa_\nu = \frac{1}{3\pi^2} \frac{Z^2 e^6}{\varepsilon_0^3 c^3 m_e^2} \left(\frac{m_e}{kT} \right)^{1/2} g(\nu, T) N N_e \exp \left(-\frac{h\nu}{kT} \right), \quad (4.20)$$

where N_e and N are the number densities of electrons and nuclei, respectively, Z is the charge of the nuclei and $g(\nu, T)$ is the Gaunt factor, which can be approximated by

$$g(\nu, T) = \frac{\sqrt{3}}{\pi} \ln \left(\frac{kT}{hv} \right). \quad (4.21)$$

The spectrum of thermal bremsstrahlung is roughly flat up to X-ray energies $\varepsilon = hv \sim kT$, above which it cuts off exponentially. Thus, by making precise spectral measurements, it is possible to determine the temperature of the gas from the location of the spectral cut-off and the column density of the hot gas from the X-ray surface brightness. The spectral emissivity has to be integrated along the line of sight through the cluster. Performing this integration and converting it into an intensity, the observed surface brightness at projected radius a from the cluster centre is

$$I_\nu(a) = \frac{1}{2\pi} \int_a^\infty \frac{\kappa_\nu(r)r}{(r^2 - a^2)^{1/2}} dr, \quad (4.22)$$

assuming spherical symmetry. Cavaliere noted that this is an Abel integral which can be inverted to find the emissivity of the gas as a function of radius (Cavaliere, 1980),

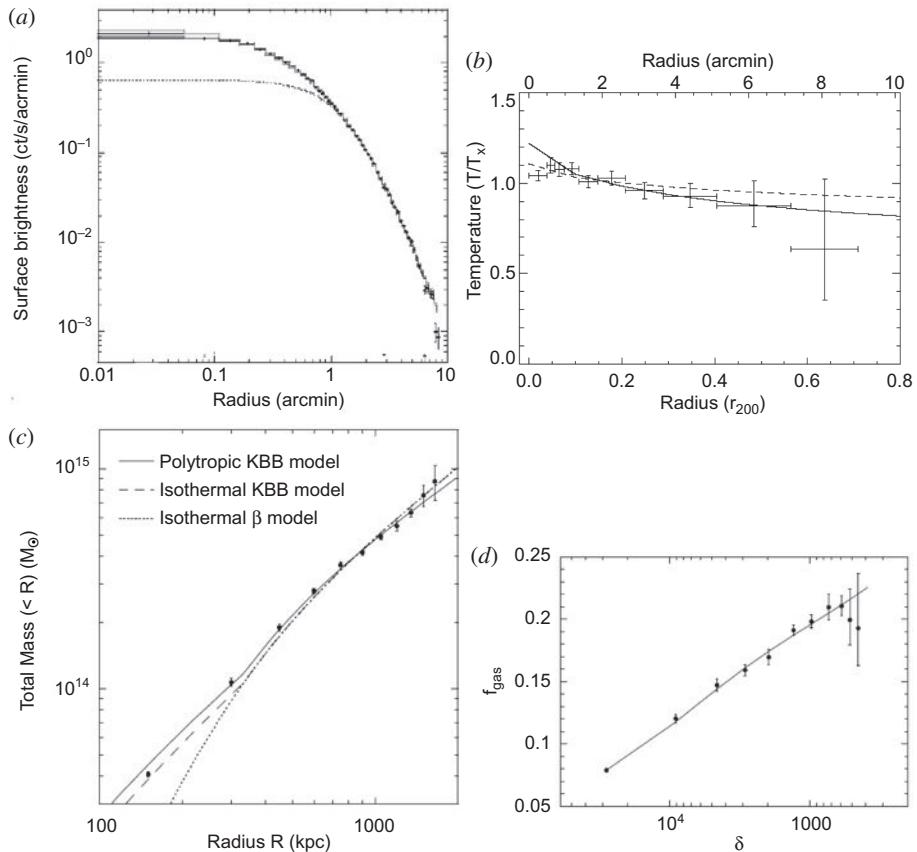
$$\kappa_\nu(r) = \frac{4}{r} \frac{d}{dr} \int_r^\infty \frac{I_\nu(a)a}{(a^2 - r^2)^{1/2}} da. \quad (4.23)$$

A beautiful example of the combined use of X-ray imaging and spectroscopy is provided by XMM-Newton X-ray Observatory observations of the rich cluster Abell 1413 by Pratt and Arnaud (2002). The observations included spatially resolved X-ray spectroscopy and so the projected temperature variation with radius in the cluster could be determined. First, the average X-ray surface brightness distribution as a function of radius was fitted by an empirical model (Fig. 4.6a). Then, the projected average temperature of the gas was estimated in annuli at different radial distances from the centre of the cluster (Fig. 4.6b). These were deprojected and the variation of the total mass within radius r derived using (4.19) (Fig. 4.6c). Finally, the ratio of gas density to total density as a function of radius, or in the case of Fig. 4.6d, the overdensity relative to the critical cosmological density, could be found.

These data are typical of what is found in rich clusters of galaxies. The dominant form of mass is the dark matter, the nature of which is unknown. About 20% of the mass is in the form of hot intergalactic gas and this is typically about five times the mass in the visible parts of galaxies. The spectroscopic observations also enable the mass of iron in the intracluster medium to be determined and this is typically found to be between about 20 and 50% of the solar value, indicating that the intergalactic gas has been enriched by the products of stellar nucleosynthesis.

4.5 Cooling flows in clusters of galaxies

If the density of the hot intracluster gas is sufficiently high, the gas may cool over cosmological time-scales. At high enough temperatures, the principal radiation loss mechanism

**Fig. 4.6**

Illustrating the determination of the physical properties of the cluster A1413 from X-ray imaging and spectroscopy by the XMM-Newton X-ray Observatory. (a) The X-ray brightness distribution as a function of distance from the centre of the cluster. (b) The projected radial distribution of the temperature of the gas. (c) The integrated mass distribution as a function of distance from the centre. (d) The fraction of gas density to total mass density f_{gas} within the cluster as a function of overdensity δ relative to the critical cosmological density (Pratt and Arnaud, 2002).

for the gas is the same thermal bremsstrahlung process which is responsible for the X-ray emission. The total energy loss rate per unit volume is

$$-\left(\frac{dE}{dt}\right) = 1.435 \times 10^{-40} Z^2 T^{\frac{1}{2}} \bar{g} N N_e \quad \text{W m}^{-3}, \quad (4.24)$$

where Z is the charge of the ions, N and N_e are the number densities of ions and electrons, respectively, and \bar{g} is a mean Gaunt factor which has value roughly 1 – we assume $Z = 1$ and $N = N_e$ (see Sect. 6.5). The thermal energy density of the fully ionised plasma is $\varepsilon = 3NkT$ and so the characteristic cooling time for the gas is

$$t_{\text{cool}} = \frac{3NkT}{|dE/dt|} = 10^{10} \frac{T^{1/2}}{N} \quad \text{years}, \quad (4.25)$$

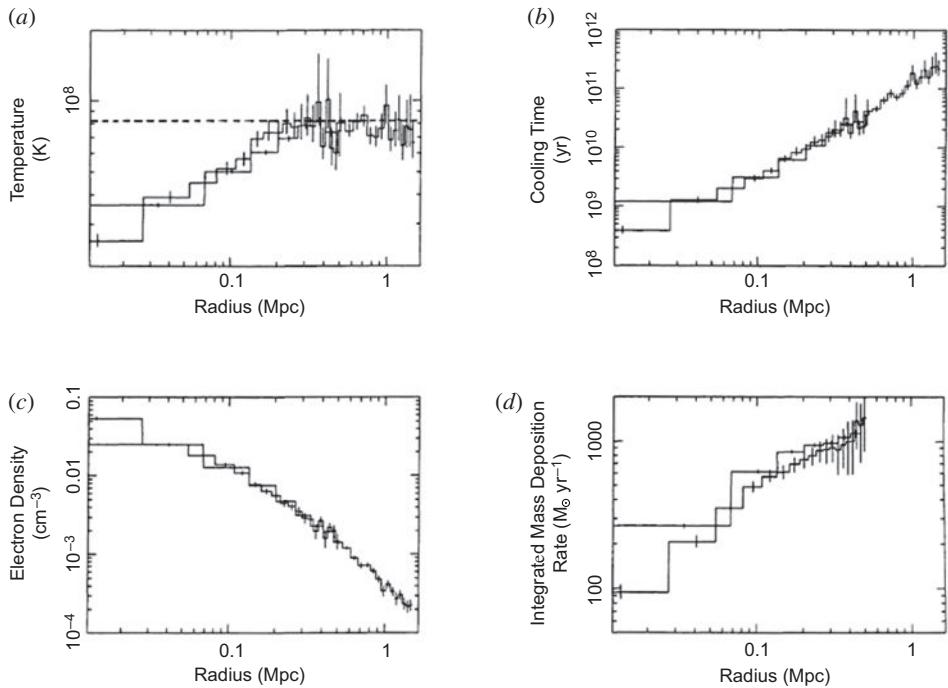
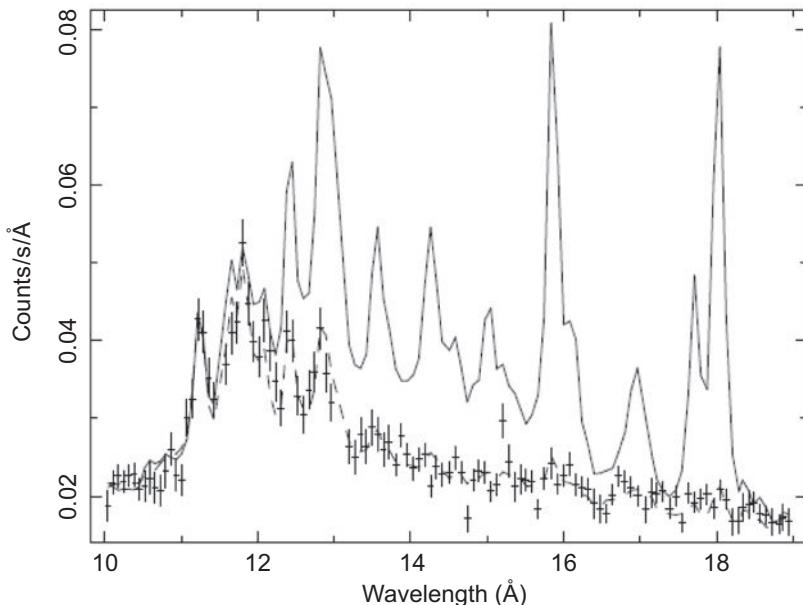


Fig. 4.7 The properties of the intracluster gas in the cluster Abell 478 obtained by deprojecting images taken by the ROSAT X-ray Observatory (White *et al.*, 1994). The cooling time of the gas is less than 10^{10} years within a radius of 200 kpc (Fabian, 1994).

where the temperature is measured in kelvins and the number density of ions or electrons in particles m⁻³. Thus, if the typical temperature of the gas is 10^7 – 10^8 K, the cooling time is less than 10^{10} years if the electron density is greater than about 3×10^3 – 10^4 m⁻³. These conditions are found in the central regions of many clusters which are intense X-ray emitters. As a result, the central regions of these hot gas clouds cool and, to preserve pressure balance, the gas density increases resulting in the formation of a *cooling flow*.

An example of the cooling flow in the cluster Abell 478 is illustrated by the diagrams shown in Fig. 4.7. The ROSAT observations were deprojected to determine mean values of the density and temperature of the gas as a function of radial distance from the centre. Figure 4.7a shows that the temperature decreases towards the central regions, while the electron density increases to values greater than 10^4 m⁻³ at the very centre (Fig. 4.7c). At a radius of 200 kpc, the electron temperature is $T = 7 \times 10^7$ K and the electron density $N_e = 8 \times 10^3$ m⁻³. Inserting these values into (4.25), the cooling time is 10^{10} years (Fig. 4.7b). Outside this radius, the temperature of the gas is constant.

As a result, matter drifts slowly in through the surface at radius $r_{\text{cool}} \approx 200$ kpc, at which the cooling time of the gas is equal to the age of the cluster. The X-ray luminosity of the

**Fig. 4.8**

Comparison of the observed high resolution X-ray spectrum of the cluster of galaxies Sérsic 159–03 observed by the ESA XMM-Newton satellite with the predicted spectrum of a standard cooling flow model without heating. The strong lower excitation lines from ions such as Fe xvii are absent, indicating the lack of cool gas in the cluster (de Plaa *et al.*, 2005).

cooling flow results from the internal energy of each element of the gas as well as the work done as it drifts slowly in towards the central regions whilst maintaining hydrostatic equilibrium. For Abell 478, the cooling flow results in a mass inflow rate of about 600–800 $M_{\odot} \text{ y}^{-1}$ (Fig. 4.7d) and so over a period of 10^{10} years, such cooling flows can contribute significantly to the baryonic mass in the central regions of the cluster. According to Fabian, about half of the clusters detected by the Einstein X-ray Observatory have high central X-ray surface brightnesses and cooling times less than 10^{10} years (Fabian, 1994). Abell 478 has a particularly massive flow, more typically, the mass flow rates being about 100 – $300 M_{\odot} \text{ y}^{-1}$.

This cannot be the whole story, however, since X-ray spectroscopic observations of the cores of clusters have shown that there is an absence of cool gas which would be expected if there were no other energy sources. This is vividly demonstrated by observations of the cluster Sérsic 159–03 which has a cool core (de Plaa *et al.*, 2005). The X-ray spectrum of the cluster is shown in Fig. 4.8, the solid line indicating the wealth of X-ray emission lines expected according to standard models of cooling flows. The observed spectrum differs dramatically from the expectations of the cooling flow models, because of the absence of strong lines associated with ions such as Fe xvii. This lack of cool gas is a feature of many of the cooling flows observed in rich clusters of galaxies (Kaastra *et al.*, 2004). The inference is that there must be some mechanism for reheating the cooling gas.

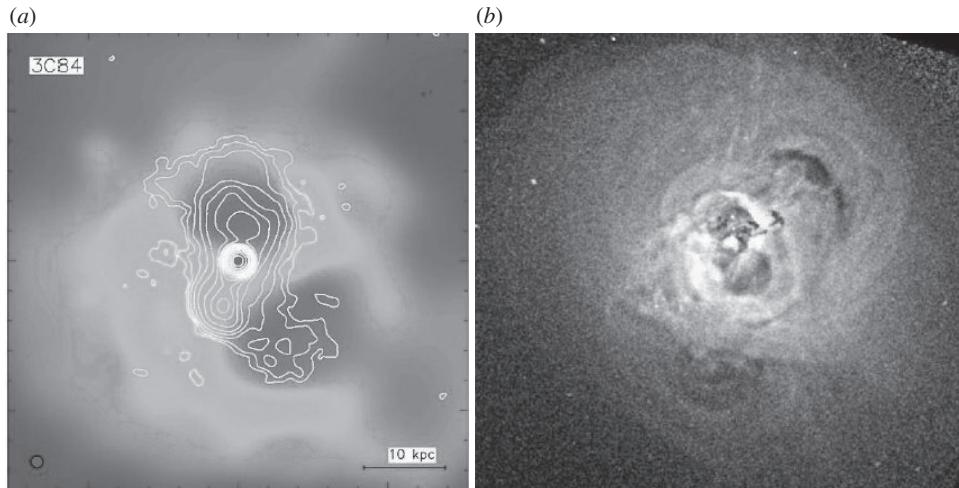


Fig. 4.9 The central regions of the Perseus Cluster of galaxies observed by the Chandra X-ray Observatory. (a) The central regions of the cluster showing the cavities evacuated by the radio lobes which are shown by the white contour lines (Fabian *et al.*, 2000). (b) An unsharp-mask image of the central regions of the cluster showing the various features caused by the expanding radio lobes. Many of the features are interpreted as sound waves caused by the weak shock wave associated with the expansion of the radio lobes (Fabian *et al.*, 2006).

Many models have been proposed to resolve this problem, some of these being discussed by Kaastra and his colleagues (Kaastra *et al.*, 2004). A highly suggestive set of observations made by the Chandra X-ray Observatory have indicated that the cooling gas in the central regions of a number of clusters is perturbed by the presence of radio lobes associated with recent radio source events. In the central region of the Perseus Cluster of galaxies, for example, buoyant lobes of relativistic plasma have pushed back the intracluster gas, forming ‘holes’ in the X-ray brightness distribution (Fig. 4.9a) (Fabian *et al.*, 2000). In a very long X-ray exposure with the Chandra X-ray Observatory, Fabian and his colleagues identified what they interpret as isothermal sound waves produced by the weak shock waves associated with the expanding lobes (Fig. 4.9b). They showed that the energy injected into the intracluster gas by these sound waves can balance the radiative cooling of the flow (Fabian *et al.*, 2006).

4.6 The Sunyaev–Zeldovich effect in hot intracluster gas

A quite different way of studying hot gas in clusters of galaxies is through observation of decrements in the intensity of the Cosmic Microwave Background Radiation in the centimetre waveband associated with the Sunyaev–Zeldovich effect. As the photons of the background radiation pass through the gas cloud, a few of them suffer Compton scattering by the hot electrons. As discussed in Sect. 9.5, although to first order the photons are just as likely to gain as lose energy in these scatterings, to second order there is a net statistical

gain of energy and so the spectrum of the Cosmic Microwave Background Radiation is shifted to slightly higher energies. As a result, there is expected to be a decrease in the intensity of the background radiation in the Rayleigh–Jeans region of the spectrum, that is, at energies $h\nu \ll kT_r$, while in the Wien region, $h\nu \gg kT_r$, there should be a slight excess. T_r is the temperature of the background radiation.

The magnitude of the distortion is determined by the *Compton scattering optical depth* y through the region of hot gas,

$$y = \int \left(\frac{kT_e}{m_e c^2} \right) \sigma_T N_e dl . \quad (4.26)$$

The resulting decrement in the Rayleigh–Jeans region of the spectrum is

$$\frac{\Delta I_\nu}{I_\nu} = -2y . \quad (4.27)$$

Thus, the magnitude of the decrement along any line of sight through the cluster provides a measure of the quantity $\int N_e T_e dl$, in other words, the integral of the pressure of the hot gas along the line of sight. For typical parameters of the hot intracluster gas, the predicted decrement amounts to $\Delta I/I \sim 10^{-4}$. The spectral signature of the effect is quite distinctive over the peak of the spectrum of the Cosmic Microwave Background Radiation (Fig. 9.13) and has been worked out in detail by Challinor and Lasenby (Challinor and Lasenby, 1998). This form of distortion has been measured in 15 Abell clusters in the SuZIE experiment carried out at the CalTech Submillimetre Observatory on Mauna Kea (Fig. 9.14) (Benson *et al.*, 2004).

An important feature of the Sunyaev–Zeldovich effect is that, if the hot gas clouds have the same properties at all redshifts, the observed decrement is independent of redshift since the Compton scattering results in only a *fractional* change in the temperature of the background radiation. This prediction is beautifully illustrated by the maps of decrements in the Cosmic Microwave Background Radiation obtained by the OVRO and BIMA millimetre arrays which span a range of redshift from 0.1 to 0.8 (Carlstrom *et al.*, 2000) (Fig. 4.10). These clusters were all known to be X-ray sources and there is good agreement between the sizes of the X-ray images and the Sunyaev–Zeldovich decrements.

The combination of the Sunyaev–Zeldovich and thermal bremsstrahlung observations of the intracluster gas enable the dimensions of the hot gas cloud to be determined independent of knowledge of the redshift of the cluster. In simple terms, the Sunyaev–Zeldovich effect determines the quantity $N_e T_e L$, where L is the dimension of the volume of hot gas. The bremsstrahlung emission of the cluster determines the quantity $L^3 N_e^2 T^{1/2}$. The temperature T can be estimated from the shape of the bremsstrahlung spectrum and so N_e can be eliminated between these two relations, enabling an estimate of L to be found. By measuring the angular size θ of the emitting volume, the distance of the cluster can be found from $D = L/\theta$. Once the redshift of the cluster has been measured, Hubble's constant can be estimated (Appendix A.2). This is one of the more promising physical methods of estimating Hubble's constant without the necessity of using a hierarchy of distance indicators.

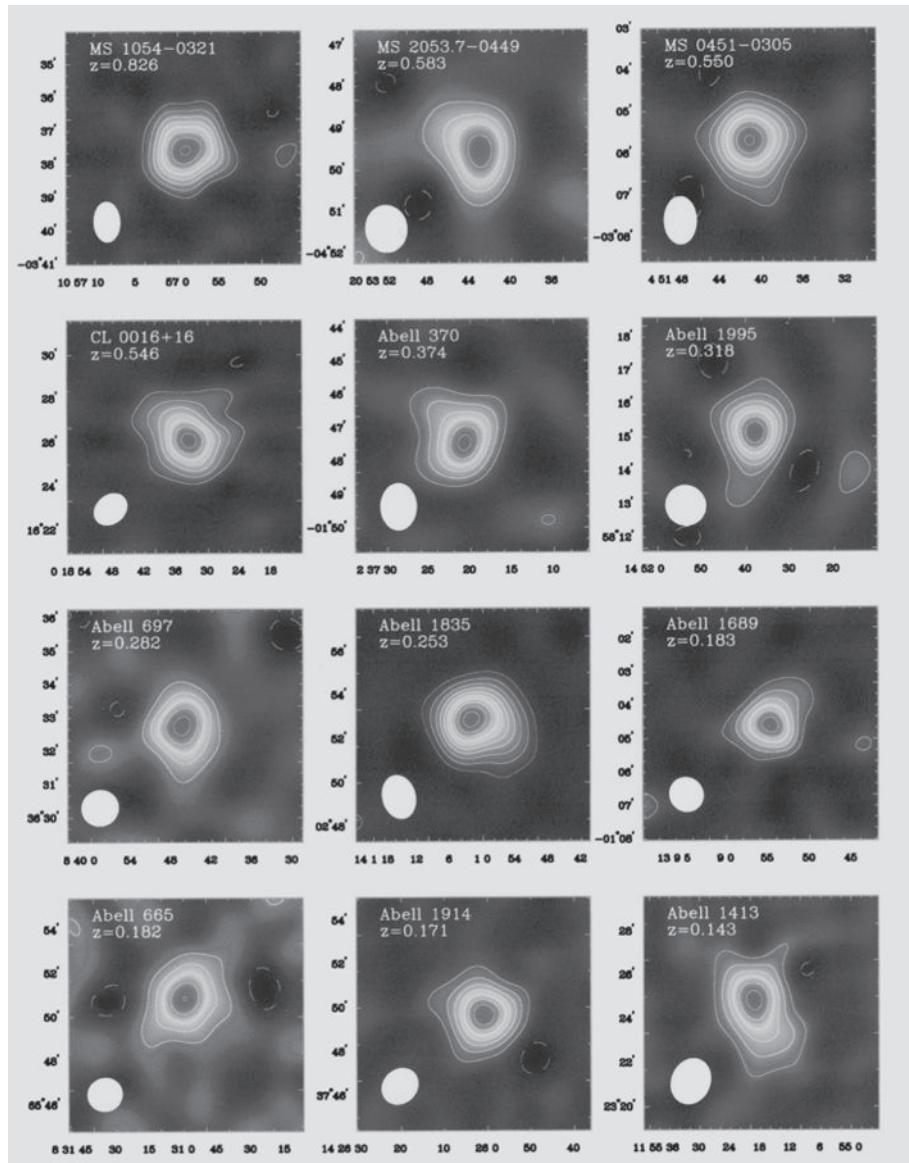


Fig. 4.10 Images of the Sunyaev–Zeldovich decrement in 12 distant clusters with redshifts in the range 0.14–0.89 (Carlstrom *et al.*, 2000). Each of the images is plotted on the same intensity scale. The data were taken with the OVRO and BIMA millimetre arrays. The filled ellipse at the bottom left of each image shows the full-width half-maximum of the effective resolution used in reconstructing the images.

4.7 Gravitational lensing by galaxies and clusters of galaxies

A beautiful method for determining the mass distribution in galaxies and clusters of galaxies has been provided by the observation of gravitationally lensed images of background

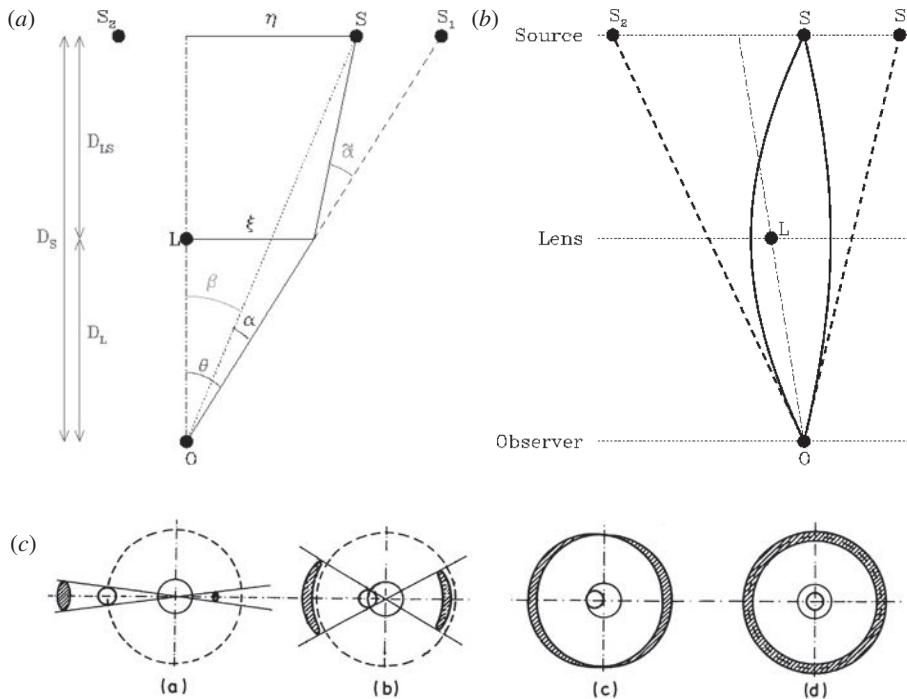


Fig. 4.11

(a) Illustrating the geometry of the deflection of light by a deflector, or lens, of mass M (Wambsganss, 1998).
 (b) Illustrating the two light paths from the source to the observer for a point mass (Wambsganss, 1998).
 (c) Illustrating the changes of the appearance of a compact background source as it passes behind a point mass. The dashed circles correspond to the Einstein radius. When the lens and the background source are precisely aligned, an Einstein ring is formed with radius equal to the Einstein radius θ_E .

galaxies. In the case of clusters of galaxies, these consist of spectacular arcs about the central core of the cluster (Fig. 4.1a) as well as distorted images of background galaxies caused by the individual galaxies in the cluster.

Many of the most important results can be derived from the formula for the gravitational deflection of light rays by the Sun, first derived by Einstein in his great paper of 1915 on the general theory of relativity (Einstein, 1915). He showed that the deflection of light by a point mass M due to the bending of space-time amounts to precisely twice that predicted by a Newtonian calculation,

$$\tilde{\alpha} = \frac{4GM}{\xi c^2}, \quad (4.28)$$

where ξ is the ‘collision parameter’ (Fig. 4.11a). The angles in Fig. 4.11a have been exaggerated to illustrate the geometry of the deflection. For the very small deflections involved in the gravitational lens effect, ξ is almost exactly the distance of closest approach of the light ray to the deflector.

Chwolson in 1924 and Einstein in 1936 realised that, if a background star were precisely aligned with a deflecting point object, the gravitational deflection of the light rays would result in a circular ring, centred upon the deflector (Fig. 4.11c) (Chwolson, 1924; Einstein,

1936). It is a straightforward calculation to work out the radius of what came to be known as an ‘Einstein ring’. In Fig. 4.11a, the distance of the background source is D_S and that of the deflector, or lens, D_L , the distance between them being D_{LS} . Suppose the observed angular radius of the Einstein ring is θ_E . Then, for a point source on-axis, since all the angles are small,

$$\theta_E = \tilde{\alpha} \left(\frac{D_{LS}}{D_S} \right) = \frac{4GM}{\xi c^2} \left(\frac{D_{LS}}{D_S} \right), \quad (4.29)$$

where $\tilde{\alpha}$ is the deflection given by (4.28). Since $\xi = \theta_E D_L$,

$$\theta_E^2 = \frac{4GM}{c^2} \left(\frac{D_{LS}}{D_S D_L} \right) = \frac{4GM}{c^2} \frac{1}{D}, \quad (4.30)$$

where $D = (D_S D_L / D_{LS})$. Thus, the *Einstein angle* θ_E , the angle subtended by the Einstein ring at the observer, is given by the relation

$$\theta_E = \left(\frac{4GM}{c^2} \right)^{1/2} \frac{1}{D^{1/2}}. \quad (4.31)$$

The above relation is also correct if the sources are at cosmological distances, provided the *Ds* are *angular diameter distances* (Blandford and Narayan, 1992).²

Expressing the mass of the deflector in terms of solar masses M_\odot and the distance D in Gpc ($= 10^9$ pc $= 3.056 \times 10^{25}$ m),

$$\theta_E = 3 \times 10^{-6} \left(\frac{M}{M_\odot} \right)^{1/2} \frac{1}{D_{\text{Gpc}}^{1/2}} \text{ arcsec}. \quad (4.32)$$

Thus, clusters of galaxies with masses $M \sim 10^{15} M_\odot$ at cosmological distances $D \sim c/H_0$ can result in Einstein rings with angular radii tens of arcseconds. Beautiful examples of partial Einstein rings about the centre of the cluster Abell 2218 have been observed with the Hubble Space Telescope by Kneib, Ellis and their colleagues (Fig. 4.1a). The ellipticity and the incompleteness of the rings reflect the facts that the gravitational potential of the cluster is not precisely spherically symmetric and that the background galaxy and the cluster are not perfectly aligned.

This is just the beginning of a remarkable story concerning the ability of strong and weak gravitational lensing to provide key astrophysical and cosmological information about the distribution of dark matter in the Universe. For many more details, the very accessible review by Wambsganss and the comprehensive discussion of all aspects of gravitational lensing presented in the volume *Gravitational Lensing: Strong, Weak and Micro* by Schneider, Kochanek and Wambsganss can be thoroughly recommended (Wambsganss, 1998; Schneider *et al.*, 2006). Let us consider one important development of the above results for extended deflectors.

² For more details, see Sects 5.5.3 and 7.5 of *Galaxy Formation* (Longair, 2008).

The simplest generalisation of the above result is to lenses with an axially symmetric mass distribution along the line of sight. In that case, the deflection is given by the expression

$$\tilde{\alpha} = \frac{4GM(\leq \xi)}{\xi c^2}, \quad (4.33)$$

where $M(\leq \xi)$ is the total projected mass within the radius ξ at the lens, a result corresponding to Gauss's theorem for Newtonian gravity.

The necessary condition for the formation of a gravitationally lensed image about an object of mass M and radius R can be derived from this result. For simplicity, suppose the lens is a uniform disc of radius R and mass M . Then, using the result (4.33), the deflection for rays grazing the edge of the disc is

$$\tilde{\alpha} = \frac{4GM(< R)}{Rc^2} = \frac{4\pi G\Sigma}{c^2} R, \quad (4.34)$$

where we have introduced the surface density of the lens as $\Sigma = M/\pi R^2$. The deflection measured by the observer at the origin is

$$\alpha(\theta) = \frac{D_{LS}}{D_S} \tilde{\alpha} = \frac{D_{LS}}{D_S} \frac{4\pi G\Sigma}{c^2} R. \quad (4.35)$$

Let us now introduce a *critical surface density* defined by

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_S}{D_{LS} D_L} = \frac{c^2}{4\pi G} \frac{1}{D}. \quad (4.36)$$

Then,

$$\alpha(\theta) = \frac{\Sigma}{\Sigma_{\text{crit}}} \frac{R}{D_L} = \frac{\Sigma}{\Sigma_{\text{crit}}} \theta. \quad (4.37)$$

Thus, if the surface density of the deflector is of the same order as the critical surface density, multiple images can be observed. In terms of the critical cosmological density, $\varrho_c = 3H_0^2/8\pi G = 3H_0^2/8\pi G = 2 \times 10^{-26} h^2 \text{ kg m}^{-3}$,

$$\Sigma_{\text{crit}} \sim \varrho_c \frac{c^2}{H_0^2} \frac{1}{D}. \quad (4.38)$$

If the sources are at cosmological distances $D \sim c/H_0$, the critical surface density is

$$\Sigma_{\text{crit}} \sim \varrho_c \frac{c}{H_0}. \quad (4.39)$$

Thus, for sources at cosmological distances, the critical surface density is roughly $2h \text{ kg m}^{-2}$.

Let us apply the result (4.33) to the case of an isothermal gas sphere, which provides a good description of the mass distribution in clusters of galaxies. We consider the simple analytic solution (4.5), which has the unpleasant features of being singular at the origin and of having infinite mass when integrated to an infinite distance, but these are unimportant for the present analysis, which is often referred to as the case of a *singular isothermal sphere*. Assuming that the velocity distribution is isotropic and that $\langle v_{\parallel}^2 \rangle$ is the observed velocity

dispersion along the line of sight,

$$\varrho(r) = \frac{2}{Ar^2} \quad \text{where} \quad A = \frac{4\pi G\mu}{kT} = \frac{4\pi G}{\langle v_{\parallel}^2 \rangle}. \quad (4.40)$$

The surface density $\Sigma(\xi)$, at projected distance ξ , is found by integrating along the line of sight, say, in the z -direction

$$\Sigma(\xi) = 2 \int_0^\infty \varrho(r) dz = 2 \int_0^{\pi/2} \varrho(r) \xi \sec^2 \theta d\theta = \frac{\langle v_{\parallel}^2 \rangle}{\pi G} \frac{1}{\xi} \int_0^{\pi/2} d\theta = \frac{\langle v_{\parallel}^2 \rangle}{2G} \frac{1}{\xi}. \quad (4.41)$$

Therefore, the total mass within the distance ξ perpendicular to the line of sight at the deflector is

$$\int_0^\xi \Sigma(\xi) 2\pi \xi d\xi = \frac{\pi \langle v_{\parallel}^2 \rangle \xi}{G}. \quad (4.42)$$

The gravitational deflection of the light rays is therefore

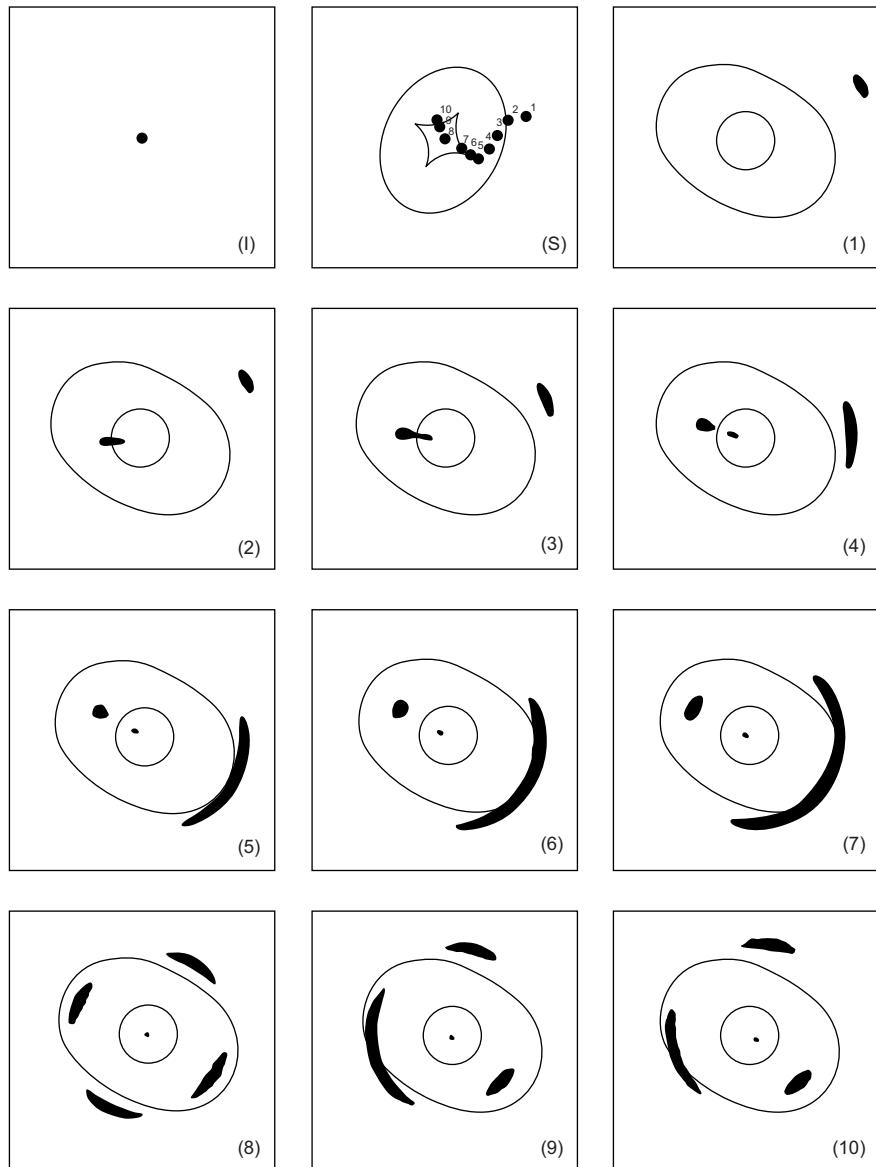
$$\tilde{\alpha} = \frac{4GM(<\xi)}{\xi c^2} = \frac{4\pi \langle v_{\parallel}^2 \rangle}{c^2}. \quad (4.43)$$

This is the remarkable result we have been seeking. For a singular isothermal sphere, the gravitational deflection is *independent* of the distance at which the light rays pass by the lens. We can therefore find the Einstein radius θ_E directly from (4.29)

$$\theta_E = \frac{4\pi \langle v_{\parallel}^2 \rangle}{c^2} \frac{D_{LS}}{D_S} = 28.8 \langle v_{\parallel}^2 \rangle \frac{D_{LS}}{D_S} \text{ arcsec}, \quad (4.44)$$

where $\langle v_{\parallel}^2 \rangle$ is the observed velocity dispersion of the galaxies in the cluster measured in units of 10^3 km s $^{-1}$. Fort and Mellier note that this is a rather robust expression for estimating the masses of clusters of galaxies (Fort and Mellier, 1994). They find that for a variety of plausible mass distributions the estimates agree to within about 10%.

Strong lensing of background sources only occurs if they lie within the Einstein angle θ_E of the axis of the lens. An excellent discussion of the shapes and intensities of the gravitationally distorted images of background sources for more general mass distributions is given by Fort and Mellier (1994). The gravitational lensing is not true lensing in the sense of geometric optics but rather the light rays come together to form caustics and cusps. Figure 4.12 shows the types of images expected for gravitational lensing by an ellipsoidal gravitational potential. The background source is shown in panel (I) and, in the second panel labelled (S), different positions of the background source with respect to the critical inner and outer caustic lines associated with the gravitational lens are shown. These are lines along which the lensed intensity of the image is infinite. The images labelled (1) to (10) show the observed images of the background source when it is located at the positions labelled on the second panel (S). The numbers and shapes of the images depend upon the location of the source with respect to the caustic surfaces. It can be seen that the predicted images resemble the arcs seen in Fig. 4.1a. For clusters of galaxies, the inferred masses are in good agreement with the values obtained by measuring the velocity dispersion of the cluster galaxies and with the X-ray method of measuring total masses.

**Fig. 4.12**

The gravitational distortions of a background source (Panel I) when it is located at different positions with respect to the axis of the gravitational lens. In this example, the lens is an ellipsoidal non-singular squeezed isothermal sphere. The 10 positions of the source with respect to the critical inner and outer caustics are shown in the panel (S). The panels labelled (1) to (10) show the shapes of the images of the lensed source (Kneib, 1993). Note the shapes of the images when the source crosses the critical caustics. Positions (6) and (7) correspond to cusp catastrophes and position (9) to a fold catastrophe (Fort and Mellier, 1994).

Gravitational lensing probes directly the total mass distribution, independent of the distribution of baryonic matter and so can be used to address a number of key astrophysical questions. For example,

- What is the distribution of mass in the dark matter haloes of galaxies and clusters?
- What are the tidal radii of the mass distributions for galaxies, both in the general field and in the cores of clusters?
- What is the bias parameter for galaxies, meaning the ratio between the clustering amplitudes for the baryonic and dark matter?
- Is there structure in the distribution of dark matter within galaxies and clusters, or is it smooth?

Strong lensing effects such as those illustrated in Fig. 4.1a enable the mass distribution to be determined on the scale of the inner caustic surfaces but, in addition, weak lensing can be detected statistically to much larger radii. As can be seen from panels 1, 2 and 3 of Fig. 4.12, the gravitationally lensed images are predicted to be stretched tangentially to the line joining the lens to the background galaxy. Therefore, by measuring the orientations of the images of large numbers of background galaxies, the effects of weak gravitational lensing can be distinguished statistically from the intrinsic ellipticities of galaxies.

As Schneider emphasises in his review, galaxy–galaxy imaging may well provide the best constraints statistically on the dimensions of dark matter haloes (Schneider *et al.*, 2006). A good example of what has been achieved is provided by the Red-Sequence Cluster Survey which involved $\sim 1.2 \times 10^5$ lensing galaxies and $\sim 1.5 \times 10^6$ fainter background galaxies in an area of 45.5 square degrees (Hoekstra *et al.*, 2004). The lensing galaxies had median redshift $z \approx 0.35$ and the background galaxies $z \approx 0.53$. These data showed that the dark matter haloes were somewhat rounder than the light distribution of the galaxies. Interestingly, the analysis of the shear data on larger angular scales provided evidence for truncation of the isothermal density distribution at a radius of $(185 \pm 30) h^{-1}$ kpc, one of the few direct estimates of the scale of the dark matter haloes.

A good example of the power of this technique is the determination of the mass distribution in a sample of 22 early-type galaxies which were imaged by the Advanced Camera for Surveys (ACS) of the Hubble Space Telescope (Gavazzi *et al.*, 2007). In the central regions, the mass distributions were determined by optical spectroscopy and by strong gravitational lensing. In the outer regions, the statistical weak gravitational lensing technique enabled the mass profile to be determined out to about 300 kpc. Gavazzi and his colleagues found that the total mass density profile was consistent with that of an isothermal sphere, $\rho \propto r^{-2}$, over two decades in radius, $(3 - 300) h^{-1}$ kpc, despite the fact that the inner regions are dominated by baryonic matter whilst the outer regions are dominated by dark matter. They found that the average stellar mass-to-light ratio was $M_*/L_V = 4.48 \pm 0.46h M_\odot/L_\odot$ while the overall average virial mass-to-light ratio was $M_{\text{vir}}/L_V = 246^{+101}_{-87} h M_\odot/L_\odot$.

4.8 Dark matter in galaxies and clusters of galaxies

The unknown nature of the dark matter which is the dominant form of gravitating mass in the outer regions of large galaxies, in clusters of galaxies and other large scale systems is one of the greatest problems of astrophysics and cosmology. It is convenient to consider separately the possibilities that the dark matter is *baryonic* or *non-baryonic*.

4.8.1 Baryonic dark matter

By *baryonic matter*, we mean ordinary matter composed of protons, neutrons and electrons and for convenience we will include black holes in this discussion. Certain forms of baryonic matter are very difficult to detect because they are very weak emitters of electromagnetic radiation. Examples of such weak emitters are brown dwarf stars with masses $M \leq 0.08 M_{\odot}$, in which the central temperatures are not hot enough to burn hydrogen into helium (Sect. 2.7.3). Although brown dwarfs are estimated to be about twice as common as stars with masses $M \geq 0.08 M_{\odot}$, they contribute very little to the mass density in baryonic matter as compared with normal stars because of their low masses. The consensus of opinion is that brown dwarfs could only make a very small contribution to the dark matter.

A strong limit to the total amount of baryonic matter in the Universe is provided by considerations of primordial nucleosynthesis. The standard Big Bang model is remarkably successful in accounting for the observed abundances of light elements such as helium-4, helium-3, deuterium and lithium-7 though the process of *primordial nucleosynthesis*. An important consequence of that success story is that the primordial abundances of the light elements, particularly of deuterium and helium-3, are sensitive tracers of the mean baryon density of the Universe. Steigman finds a best estimate of the mean baryon density of the Universe of $\Omega_B = 0.0455$ assuming $h = 0.7$, compared with a mean density of matter in the Universe of $\Omega_0 \approx 0.3$ (Steigman, 2004). Thus, ordinary baryonic matter is only about one tenth of the total mass density of the Universe, most of which must therefore be in some non-baryonic form.

Black holes are another possibility for the dark matter. The supermassive black holes in the nuclei of galaxies have masses which are typically only about 0.1% of the mass of the bulges of their host galaxies and so they contribute negligibly to the mass density of the Universe. There might, however, be an invisible intergalactic population of massive black holes. Limits to the number density of such black holes can be set in certain mass ranges from studies of the numbers of gravitationally lensed images observed in large samples of extragalactic radio sources. In their VLA survey of a very large sample of such sources, Hewitt and her colleagues set an upper limit to the number density of black holes with masses in the range $10^{10} \leq M \leq 10^{12} M_{\odot}$ of $\Omega_{BH} \ll 1$ (Hewitt *et al.*, 1987). The same technique can be used to study the mass density of black holes in the mass range $10^6 \leq M \leq 10^8 M_{\odot}$. Wilkinson and his colleagues searched a sample of 300 compact radio sources studied by VLBI techniques for examples of multiple gravitationally lensed images

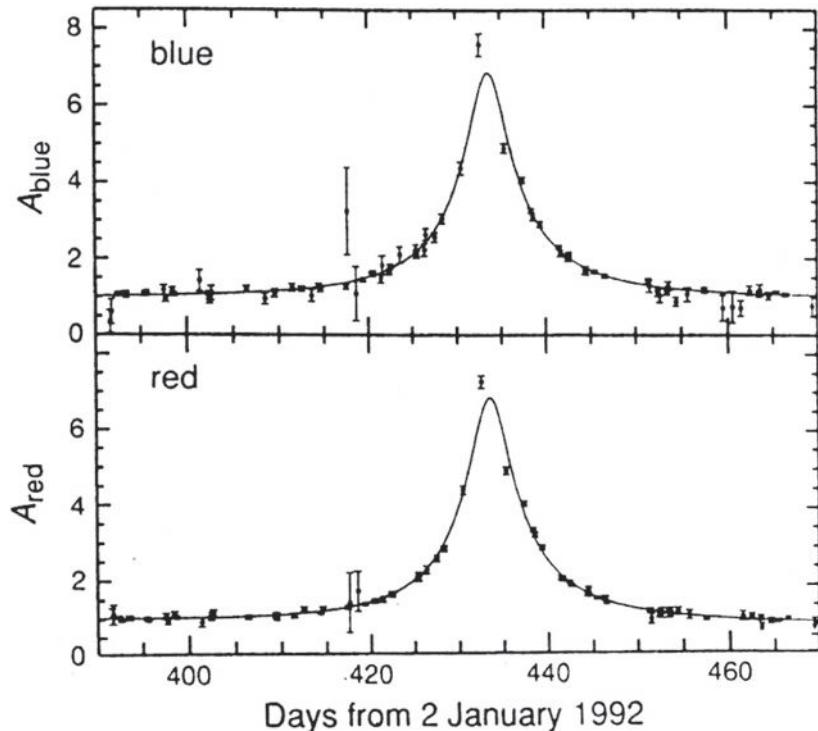


Fig. 4.13 The gravitational microlensing event recorded by the MACHO project in February and March 1993. The horizontal axis shows the date in days measured from day zero on 2 January 1992. The vertical axis shows the amplification of the brightness of the lensed star relative to the unlensed intensity in the blue and red wavebands. The solid lines show the expected variations of brightness of a lensed star with time. The same characteristic light curve is observed in both wavebands, as expected for a gravitational microlensing event (Alcock *et al.*, 1993b).

but none was found. The upper limit to the cosmological mass density of these black holes corresponded to less than 1% of the critical cosmological density (Wilkinson *et al.*, 2001).

An impressive approach to setting limits to the contribution which discrete low mass objects, collectively known as MAssive Compact Halo Objects, or MACHOs, could make to the dark matter in the halo of our own Galaxy, has been the search for gravitational microlensing signatures of such objects as they pass in front of background stars. The MACHOs include low mass stars, white dwarfs, brown dwarfs, planets and black holes. These lensing events are very rare and so very large numbers of background stars have to be monitored. This technique is sensitive to MACHOs with a very wide range of masses, from 10^{-7} to $100 M_{\odot}$. In addition, the expected light curve of such gravitational lensing events has a characteristic light curve which is independent of wavelength. The time-scale of the brightening is roughly the time it takes the MACHO to cross the Einstein radius of the dark deflector. The first example of such a microlensing event was discovered in October 1993 (Fig. 4.13), the mass of the invisible lensing object being estimated to lie in the range $0.03 < M < 0.5 M_{\odot}$ (Alcock *et al.*, 1993a).

By the end of the MACHO project, 13 definite and four possible events were observed in the direction of the Large Magellanic Cloud, significantly greater than the 2–4 detections expected from known types of star (Alcock *et al.*, 2000). The best statistical estimates suggest that the mean mass of these MACHOs is between 0.15 and $0.9 M_{\odot}$. The statistics are consistent with MACHOs making up about 20% of the necessary halo mass. Somewhat fewer microlensing events were detected in the EROS project which found that less than 25% of the mass of the standard dark matter halo could consist of dark objects with masses in the range 2×10^{-7} to $1 M_{\odot}$ at the 95% confidence level (Afonso *et al.*, 2003). The consensus view is that MACHOs alone cannot account for all the dark matter in the halo of our Galaxy and so some form of non-baryonic matter must make up the difference.

4.8.2 Non-baryonic dark matter

The general consensus is that the dark matter is most likely to be in some non-baryonic form and so is of the greatest interest for particle physicists. Three of the most popular possibilities are axions, neutrinos with finite rest mass and Weakly Interacting Massive Particles, or WIMPs.

Axions The smallest mass candidates are the *axions* which were invented by particle theorists in order to ‘save quantum chromodynamics from strong CP violation’. If they exist, they must have been created when the thermal temperature of the Universe was about 10^{12} K but they were out of equilibrium and never acquired thermal velocities – they remained ‘cold’. Their rest mass energies are expected to lie in the range 10^{-2} – 10^{-5} eV. The role of such particles in cosmology and galaxy formation is discussed by Efstathiou (1990) and by Kolb and Turner (1990).

Neutrinos with finite rest mass A second possibility is that the three known types of neutrino have finite rest masses. Laboratory tritium β -decay experiments have provided an upper limit to the rest mass of the electron antineutrino of $m_{\nu} \leq 2$ eV (Weinheimer, 2001), although the particle data book suggests a conservative upper limit of 3 eV (see <http://www-pdg.lbl.gov/pdg.html>). The discovery of neutrino oscillations has provided a measurement of the mass difference between the μ and τ neutrinos of $\Delta m^2 \sim 3 \times 10^{-3}$ (Eguchi *et al.*, 2003; Aliu *et al.*, 2005). Thus, although their masses are not measured directly, they probably have masses of the order of 0.1 eV. The reason that these values are of interest is that neutrinos of rest mass of about 10–20 eV would be enough to provide the critical cosmological density. Taking $h = 0.7$, if the neutrino rest mass were about 15 eV and there were six neutrino species, the electron, muon and tau neutrinos and their antiparticles, the known types of neutrino could close the Universe. However, if the mass of the neutrinos is of the order 0.1 eV, they certainly could not account for the amount of dark matter present in the Universe.

WIMPs A third possibility is that the dark matter is in some form of *Weakly Interacting Massive Particle*, or WIMP. This might be the gravitino, the supersymmetric partner of the graviton, or the photino, the supersymmetric partner of the photon, or some form of as yet unknown massive neutrino-like particle. There is the real possibility that clues will be found from experiments to be carried out in the TeV energy range with the Large Hadron Collider (LHC) and the next generation International Linear Collider (ILC). According

to generic arguments given by Trodden, physics beyond the standard model of particles physics is essential and almost any model involves new particles at the TeV scale (Trodden, 2006).

4.8.3 Astrophysical and experimental limits

Useful astrophysical limits can be set to the number densities of different types of neutrino-like particles in the outer regions of giant galaxies and in clusters of galaxies. The WIMPs and massive neutrinos are collisionless fermions and therefore there are constraints on the phase space density of these particles, which translate into a lower limit to their masses. Let us give a simple derivation of this result. More details of this calculation are given by Tremaine and Gunn, who provide a tighter constraint on the masses of these hypothetical particles (Tremaine and Gunn, 1979).

Neutrino-like particles are fermions and are subject to the Pauli exclusion principle according to which there is a maximum number of particle states in phase space for a given momentum p_{\max} . The elementary phase volume is \hbar^3 and, recalling that there can be two particles of opposite spin per state, the maximum number of particles with momenta up to p_{\max} is

$$N \leq 2 \frac{g}{\hbar^3} \frac{4\pi}{3} p_{\max}^3 , \quad (4.45)$$

per unit volume, where g is the statistical weight of the neutrino species. If there is more than one neutrino species present, this number is multiplied by N_ν . Bound gravitating systems such as galaxies and clusters of galaxies are subject to the virial theorem (Sect. 3.5.1) and so, if σ is the root-mean-square velocity dispersion of the objects which bind the system, $\sigma^2 = GM/R$. Therefore the maximum velocity which particles within the system can have is the escape velocity from the cluster, $v_{\max} = (2GM/R)^{1/2} = \sqrt{2}\sigma$. The neutrino-like particles bind the system and so its total mass is $M = NN_\nu m_\nu$, where m_ν is the rest mass of the particle. We therefore find the following lower limit to the rest mass of the neutrinos in terms of observable quantities:

$$m_\nu^4 \geq \left(\frac{9\pi}{8\sqrt{2}g} \right) \frac{\hbar^3}{N_\nu G \sigma R^2} ; \quad m_\nu \geq \frac{1.5}{(N_\nu \sigma_3 R_{\text{Mpc}}^2)^{1/4}} \text{ eV} , \quad (4.46)$$

where the velocity dispersion σ_3 is measured in units of 10^3 km s^{-1} and R is measured in Mpc. For clusters of galaxies, typical values are $\sigma = 1000 \text{ km s}^{-1}$ and $R = 1 \text{ Mpc}$. If there are six neutrino species, namely, electron, muon, tau neutrinos and their antiparticles, $N_\nu = 6$ and then $m_\nu \geq 0.9 \text{ eV}$ would be required to bind the clusters, greater than the laboratory upper limit to the mass of the electron antineutrino.

There is a further constraint on the possible masses of WIMPs. Studies of the decay of the W^\pm and Z^0 bosons at CERN have shown that the width of the decay spectrum is consistent with there being only three neutrino species with rest mass energies less than about 40 GeV. Therefore, if the dark matter is in some form of ultra-weakly interacting particle, its rest mass energy must be greater than 40 GeV.

Another important constraint is that, if the masses of the particles were greater than 15 eV and they are as common as neutrinos and photons, as expected in the standard Big Bang model, the present density of the Universe would exceed the critical mass density ρ_c . Therefore there would have to be some suppression mechanism to ensure that, if $m \geq 40$ GeV, these particles are very much less common than the photons and electrons neutrinos at the present day.³

The search for evidence for different types of dark matter particles has developed into one of the major areas of *astroparticle physics*. An important class of experiments involves the search for weakly interacting particles with masses $m \geq 1$ GeV, which could make up the dark halo of our Galaxy. In order to form a bound dark halo about our Galaxy, the particles would have to have velocity dispersion $\langle v^2 \rangle^{1/2} \sim 230$ km s⁻¹ and their total mass is known. Therefore, the number of WIMPs passing through a terrestrial laboratory each day is a straightforward calculation. The challenge is to detect the very small number of events expected because of the very small cross-section for the interaction of WIMPs with the nuclei of atoms.

A good example of the quality of the data now available is provided by the results of the Cryogenic Dark Matter Search (CDMS) at the Soudan Underground Laboratory in Minnesota, USA. The CDMS experiment has set a 90% confidence upper limit to the spin-independent WIMP–nucleon interaction cross-section at its most sensitive mass of 60 GeV/c² of $\sigma_w \leq 1.6 \times 10^{-47}$ m² (Akerib *et al.*, 2006). This cross-section can be compared with the weak interaction cross-section for neutrino–electron scattering, $\sigma = 3 \times 10^{-49}(E/m_e c^2)$ m². Already the CDMS result constrains the predictions of supersymmetric models of particle physics. The sensitivity of these experiments should increase by successive orders of magnitude through the different phases of the SuperCDMS proposal.

³ More details of suppression mechanisms are given in Sect. 10.6 of *Galaxy Formation* (Longair, 2008).

PART II

PHYSICAL PROCESSES

The second part of this book is concerned with elementary physical processes involved in studies of high energy phenomena in the Universe. There are many excellent books which discuss this material at various levels of sophistication. Those which I have found most helpful are Jackson's *Classical Electrodynamics* (Jackson, 1999), *Radiation Processes in Astrophysics* by Rybicki and Lightman (1979) and *Electromagnetic Processes* by Gould (2005). Zombeck's *Handbook of Space Astronomy and Astrophysics* (Zombeck, 2006) contains a very useful compendium of relevant data.

My intention is to emphasise the underlying physical principles involved in these processes so that the functional forms of the equations have an intuitive significance. I will build up each discussion gently, often deriving approximate results which give physical insight before deriving, or quoting, the results of more complete calculations. I will treat the key processes of synchrotron radiation and inverse Compton scattering in some detail.

In the various calculations and derivations, I use Système International (SI) units, which have been officially adopted by almost all countries in the world. According to the Wikipedia web site (2008), 'Three nations have not officially adopted the International System of Units as their primary or sole system of measurement: Liberia, the Union of Myanmar (Burma) and the United States.' I hope those readers whose nations have not yet adopted the SI system of units will bear with me, for the sake of the majority who have. Unfortunately, many of the diagrams appearing in the literature are presented in a variety of non-SI units and the reader will have to make the translations between units. This is unlikely to pose any serious problem. Where practical, I will provide appropriate translations.

5.1 Introduction

When high energy particles pass through a solid, liquid or gas, they can cause considerable wreckage to the constituent atoms, molecules and nuclei. Specifically, they cause:

- (i) the ionisation and excitation of the atoms and molecules of the material. In the process of ionisation, electrons are torn off atoms by the electrostatic forces between the charged high energy particle and the electrons. This is not only a source of ionisation but also a source of heating of the material because of the transfer of kinetic energy to the electrons;
- (ii) the destruction of crystal structures and molecular chains;
- (iii) nuclear interactions between the high energy particles and the nuclei of the atoms of the material.

In this chapter we will be principally concerned with the first of these processes, *ionisation losses*, which are important in a number of different contexts. They influence the propagation of high energy particles under cosmic conditions and the associated energy losses provide an effective mechanism for heating the interstellar gas, for example, in giant molecular clouds. Equally important is the use of the ionisation losses of high energy particles in particle detectors – these provide a means of identifying the properties of the particles as well as providing a measure of their incident fluxes upon the detector.

There is a pedagogical reason for beginning with ionisation losses. From the astrophysical perspective, ionisation losses provide an example of the procedures which have to be followed in working out the various ways in which high energy particles interact with matter. We will show how the results can be adapted to apparently quite different physical problems – for example, to the destruction of crystal structures and molecular chains and to gravitational interactions between stars. These are intended to provide insight into the wide applicability of the techniques and concepts introduced in this chapter.

5.2 Ionisation losses – non-relativistic treatment

Consider first the collision of a high energy proton or nucleus with a stationary electron. Only a very small fraction of the kinetic energy of the high energy particle is transferred

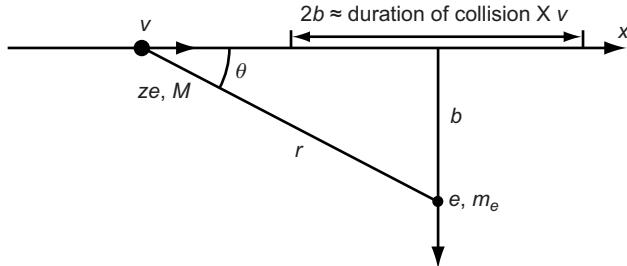


Fig. 5.1

The geometry of the collision of a high energy particle with a stationary electron, illustrating the definition of the collision parameter b .

to the electron as can be appreciated from the case of a head-on collision of a high energy particle of mass M and velocity v with an electron of mass m_e . Taking the particles to be solid spheres, it is a simple calculation to show that the maximum velocity acquired by the electron in a non-relativistic collision is $[2M/(M + m_e)]v$. Recalling that $m_e \ll M$, this is approximately $2v$. Therefore, the loss of kinetic energy of the high energy particle is less than $\frac{1}{2}m_e(2v)^2 = 2m_e v^2$ and its fractional kinetic energy loss is less than $\frac{1}{2}m_e(2v)^2 / \frac{1}{2}Mv^2 = 4m_e/M$. Since $M \gg m_e$, the fractional loss of energy per collision is very small. Therefore, in real collisions in which the interaction is mediated by the electrostatic fields of the particles, the incident high energy particle is essentially undeviated. All that happens is that the electrons of the medium receive a small momentum impulse through the electrostatic attraction or repulsion of the high energy particle.

We begin with a non-relativistic treatment in which the high energy particle is assumed to move so fast that its trajectory is undeviated and the electron remains stationary during the interaction (Fig. 5.1). The charge of the high energy particle is ze and its mass M ; b , the distance of closest approach of the particle to the electron, is called the *collision parameter*. The total *momentum impulse* given to the electron in this encounter is $\int F dt$. By symmetry, the forces parallel to the line of flight of the high energy particle cancel out and therefore we need only work out the component of force perpendicular to the line of flight. Then,

$$F_{\perp} = \frac{ze^2}{4\pi\epsilon_0 r^2} \sin\theta ; \quad dt = \frac{dx}{v} . \quad (5.1)$$

Changing variables to the angle θ shown in Fig. 5.1, $b/x = \tan\theta$, $r = b/\sin\theta$ and therefore $dx = (-b/\sin^2\theta)d\theta$; v is effectively constant and therefore the momentum impulse is

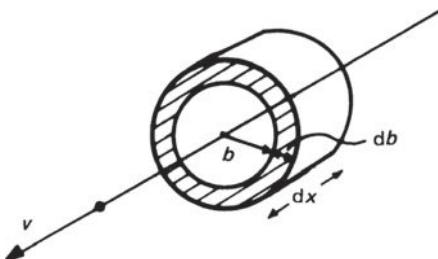
$$\int_{-\infty}^{\infty} F_{\perp} dt = - \int_0^{\pi} \frac{ze^2}{4\pi\epsilon_0 b^2} \sin^2\theta \frac{b \sin\theta}{v \sin^2\theta} d\theta = - \frac{ze^2}{4\pi\epsilon_0 bv} \int_0^{\pi} \sin\theta d\theta . \quad (5.2)$$

Therefore,

$$\text{momentum impulse } p = \frac{ze^2}{2\pi\epsilon_0 bv} . \quad (5.3)$$

The kinetic energy transferred to the electron is

$$\frac{p^2}{2m_e} = \frac{z^2 e^4}{8\pi^2 \epsilon_0^2 b^2 v^2 m_e} = \text{energy loss by high energy particle} . \quad (5.4)$$

**Fig. 5.2**

Illustrating the cylindrical volume within which collisions with collision parameters b to $b + db$ take place in the distance increment dx .

We now need to find the *average energy loss per unit path length* and so we work out the number of encounters with collision parameters in the range b to $b + db$ and integrate over collision parameters. From the geometry of Fig. 5.2, the total energy loss of the high energy particle, $-dE$, in length dx is:

$$\begin{aligned} & (\text{number of electrons in volume } 2\pi b db dx) \times (\text{energy loss per interaction}) \\ &= \frac{z^2 e^4 N_e}{8\pi^2 \varepsilon_0^2 v^2 m_e} \times \int_{b_{\min}}^{b_{\max}} \frac{2\pi b}{b^2} db dx , \end{aligned} \quad (5.5)$$

where N_e is the number density, or concentration, of electrons. Notice that the limits b_{\max} and b_{\min} to the range of collision parameters have been included in this integral. Integrating,

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi \varepsilon_0^2 v^2 m_e} \ln \left(\frac{b_{\max}}{b_{\min}} \right) . \quad (5.6)$$

Notice how the logarithmic dependence upon b_{\max}/b_{\min} comes about. The closer the encounter, the greater the momentum impulse, $p \propto b^{-2}$. However, there are more electrons at large distances ($\propto b db$) and hence, on integrating, we obtain only a logarithmic dependence of the energy loss upon the range of collision parameters. We will encounter the same phenomenon in the case of bremsstrahlung (Sect. 6.4) and in working out the conductivity of a plasma (Sect. 11.1). You may well ask, ‘Why introduce the limits b_{\max} and b_{\min} , rather than work out the answer properly?’ The reason is that the proper sum is significantly more complicated and would take account of the acceleration of the electron by the high energy particle and include a quantum mechanical treatment of the interaction. Our approximate methods give remarkably good answers, however, because the limits b_{\max} and b_{\min} only appear inside the logarithm and hence need not be known very precisely.

5.2.1 Upper limit b_{\max}

An upper limit to the range of integration over collision parameters, corresponding to the smallest energy transfer, occurs when the duration of the collision is of the same order as the period of the electron in its orbit in the atom. Then, the interaction is no longer impulsive. In the limit in which the duration of the collision is much greater than the period of the orbit,

the electron feels a slowly varying weak field and, in terms of the dynamics of particles, to be discussed later, it ‘conserves its motion adiabatically’ during the perturbation and no ionisation takes place. What do we mean by the *duration* of the collision? The energy transfer to the electron can be derived as follows. If we take the time during which the particle experiences a strong interaction with the electron to be $\tau = 2b/v$ (Fig. 5.1) and multiply by the electrostatic force at the distance of closest approach b , then

$$F = ze^2/4\pi\varepsilon_0 b^2; \quad \text{momentum impulse } p = F\tau = \frac{ze^2}{2\pi\varepsilon_0 bv}. \quad (5.7)$$

This is the same answer as (5.3). In other words, we can think of the encounter as lasting a time $\tau = 2b/v$. If the collision time is the same as the orbital period of the electron, we obtain an order of magnitude estimate for b_{\max} . Hence,

$$2b_{\max}/v \approx 1/\nu_0, \quad (5.8)$$

where ν_0 is the orbital frequency of the electron. Writing $\omega_0 = 2\pi\nu_0$,

$$b_{\max} \approx \frac{v}{2\nu_0} = \frac{\pi v}{\omega_0}. \quad (5.9)$$

5.2.2 Lower limit b_{\min}

There are two possibilities for b_{\min} :

- (i) According to classical physics, the closest distance of approach corresponds to that collision parameter at which the electrostatic potential energy of the interaction of the high energy particle and the electron is equal to the maximum possible energy transfer which, according to our first calculation, is $2m_e v^2$. Thus,

$$ze^2/4\pi\varepsilon_0 b_{\min} \approx 2m_e v^2; \quad b_{\min} = ze^2/8\pi\varepsilon_0 m_e v^2. \quad (5.10)$$

We can show that, if this amount of energy were transferred during the interaction, the electron would move a distance of order b_{\min} during the encounter and so the assumption on which the calculation is based breaks down. To demonstrate this, the average velocity of the electron perpendicular to the line of flight of the high energy particle during the encounter is p/m_e . Therefore, the distance moved in the collision time $\tau = 2b/v$ is $(p/m_e) \times (2b/v) = ze^2/4\pi\varepsilon_0 m_e v^2$, which is of the same order of magnitude as b_{\min} .

- (ii) A second possible value of b_{\min} is associated with the fact that we ought to have carried out a quantum mechanical calculation to describe close encounters between the atomic system and the high energy particle. The maximum velocity acquired by the electron in the encounter is $\Delta v \approx 2v$ and hence its change in momentum is $\Delta p = 2m_e v$. There is therefore a corresponding uncertainty in the position Δx according to the Heisenberg

uncertainty principle, $\Delta x \approx \hbar/2m_e v$. Therefore,

$$b_{\min} = \hbar/2m_e v. \quad (5.11)$$

If this turns out to be the appropriate value of b_{\min} , a quantum calculation should have been carried out. Granted this defect in our calculation, the value of b_{\min} still tells us the smallest meaningful value of b for the purposes of our integration.

We choose whichever of these values of b_{\min} is the *larger* for the physical conditions of the problem. The ratio of possible values of b_{\min} is:

$$\frac{b_{\min}(\text{quantum})}{b_{\min}(\text{classical})} = \frac{\hbar}{2m_e v} \frac{8\pi\varepsilon_0 m_e v^2}{ze^2} = \frac{4\pi\varepsilon_0 v \hbar}{ze^2} = \frac{1}{z\alpha} \left(\frac{v}{c}\right) = \frac{137}{z} \left(\frac{v}{c}\right), \quad (5.12)$$

where $\alpha = e^2/4\pi\varepsilon_0 c \hbar \approx 1/137$ is the fine structure constant. Thus, if the high energy particles have $v/c \gtrsim 0.01$, the quantum limit should be used. The expression (5.6) also applies for ionisation losses involving non-relativistic particles interacting with cold matter, for example, the gas in a giant molecular cloud. In this case, the typical velocities of the particles can be less than $0.01c$ and so the classical limit should be used.

In the high velocity, non-relativistic limit, the loss rate per unit path length (5.6) becomes

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi\varepsilon_0^2 v^2 m_e} \ln\left(\frac{2\pi m_e v^2}{\hbar\omega_0}\right). \quad (5.13)$$

The angular frequency ω_0 of the electron in its orbit can be expressed in terms of its atomic binding energy. For the Bohr model of the atom, ω_0 is the orbital angular frequency of the electron in its ground state and the binding energy, or ionisation potential I , is $I = \frac{1}{2}\hbar\omega_0$. Therefore,

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi\varepsilon_0^2 v^2 m_e} \ln\left(\pi \frac{m_e v^2}{I}\right). \quad (5.14)$$

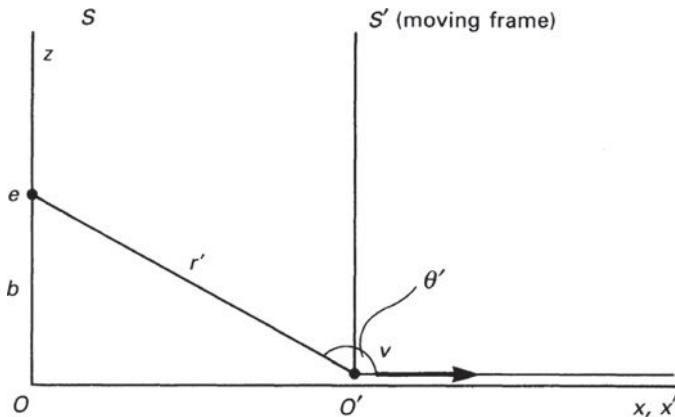
In practice, I should be some properly weighted mean over all states of the electrons in the atom, that is, we should write \bar{I} not I . The value of \bar{I} takes account of the fact that there are electrons in many different energy levels in the atoms of the medium which can be ejected by the high energy particle. The value of \bar{I} cannot be calculated exactly except for the simplest atoms and has to be found by experiment. Conventionally, the loss rate is written,

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi\varepsilon_0^2 v^2 m_e} \ln\left(\frac{m_e v^2}{\bar{I}}\right), \quad (5.15)$$

where we recognise $2m_e v^2$ as an old friend, the maximum kinetic energy E_{\max} which can be transferred to the electron.

Another way of obtaining the same result is to work out the energy spectrum of the ejected electrons. It is left as an exercise to the reader to show that the energy spectrum per unit path length is of power-law form:

$$N(E) dE = \frac{z^2 e^4 N_e}{8\pi\varepsilon_0^2 v^2 m_e} \frac{dE}{E^2}. \quad (5.16)$$

**Fig. 5.3**

The reference frames S and S' in standard configuration used in evaluating the strength of the electric field of a relativistic charged particle at time $t > 0$.

Integration over all energies from \bar{I} to E_{\max} gives the same logarithmic term, $\ln(E_{\max}/\bar{I})$, derived above.

Inspection of formula (5.15) shows that the ionisation loss rate is independent of the mass of the high energy particle. If we measure the loss rate per unit path length, $-dE/dx$, we obtain information about $(z/v)^2$. Notice also that the ionisation losses are proportional to m_e^{-1} and therefore ‘ionisation’ losses due to electrostatic interactions of the high energy particles with protons and nuclei can be safely neglected.

5.3 The relativistic case

The extension of the above analysis to the case of a highly relativistic high energy particle is straightforward. The electron is again accelerated by the electric field of the relativistic particle and so the next step is to work out how the inverse square law of electrostatics is modified when the source of the field is moving relativistically. This is an important calculation and will reappear a number of times in the course of the exposition.

5.3.1 The relativistic transformation of an inverse square law Coulomb field

We orient the reference frames S and S' in standard configuration with the high energy particle moving along the positive x -axis and the electron located at a distance b along the z -axis in S (Fig. 5.3). The coordinate systems are set up so that $t = t' = 0$ and $x = x' = 0$ when the high energy particle is at its distance of closest approach in S . At time t , the particle is located at x in S . In S' , the coordinates of the electron (or its displacement four-vector) are $[ct', -vt', 0, b]$ (see Appendix A.4.2). Furthermore, in S' the electric field E

of the particle is spherically symmetric about the origin $0'$ and hence, at the electron,

$$\begin{aligned} E_{x'} &= \frac{ze}{4\pi\epsilon_0 r'^2} \cos\theta' = -\frac{ze}{4\pi\epsilon_0 r'^3} \frac{x'}{r'}, \\ E_{z'} &= \frac{ze}{4\pi\epsilon_0 r'^2} \sin\theta' = \frac{ze}{4\pi\epsilon_0 r'^3} \frac{b}{r'}, \end{aligned}$$

where $r'^2 = (vt')^2 + b^2$ and θ' is the angle between the positive x -axis and the direction of the electron in S' . We now relate time measured by the stationary observer on the electron in S to that measured by the observer moving with high energy particle,

$$ct' = \gamma \left(ct - \frac{vx}{c} \right). \quad (5.17)$$

But, by our choice of coordinates, $x = 0$ for the electron in S and hence $t' = \gamma t$. Therefore,

$$\begin{aligned} E_{x'} &= -\frac{ze(\gamma vt)}{4\pi\epsilon_0 [b^2 + (\gamma vt)^2]^{1/2}}, \\ E_{z'} &= \frac{zeb}{4\pi\epsilon_0 [b^2 + (\gamma vt)^2]^{1/2}}. \end{aligned}$$

Notice that we have expressed the field in S' in terms of coordinates in S . The inverse Lorentz transforms for the electric field strength E and the magnetic flux density B from S' to S are:

$$\left. \begin{aligned} E_x &= E_{x'} & B_x &= B_{x'}, \\ E_y &= \gamma(E_{y'} + vB_{z'}) & B_y &= \gamma \left(B_{y'} - \frac{v}{c^2} E_{z'} \right), \\ E_z &= \gamma(E_{z'} + vB_{y'}) & B_z &= \gamma \left(B_{z'} + \frac{v}{c^2} E_{y'} \right). \end{aligned} \right\}$$

Since $B_{x'} = B_{y'} = B_{z'} = 0$ in S' , we find

$$\left. \begin{aligned} E_x &= -\frac{\gamma zevt}{4\pi\epsilon_0 [b^2 + (\gamma vt)^2]^{3/2}} & B_x &= 0, \\ E_y &= 0 & B_y &= -\frac{\gamma zevb}{4\pi\epsilon_0 c^2 [b^2 + (\gamma vt)^2]^{3/2}}, \\ E_z &= \frac{\gamma zeb}{4\pi\epsilon_0 [b^2 + (\gamma vt)^2]^{3/2}} & B_z &= 0. \end{aligned} \right\} \quad (5.18)$$

Notice that $B_y = -(v/c^2)E_z$.

The expressions (5.18) for the electric field strength E and the magnetic flux density B associated with a relativistically moving charge are rather useful. In the non-relativistic limit, $v/c \ll 1$, the expressions for the electric field revert to the standard form of Coulomb's law as would be expected. When the particle is relativistic, however, the electric field at the electron is much enhanced but it is experienced by the electron for a much shorter time. Figure 5.4, taken from Jackson's exposition, illustrates the differences between the non-relativistic and relativistic cases (Jackson, 1999). At its distance of closest approach, $x = 0, t = 0$, E_z is greater in the relativistic case by a factor γ as compared with the low

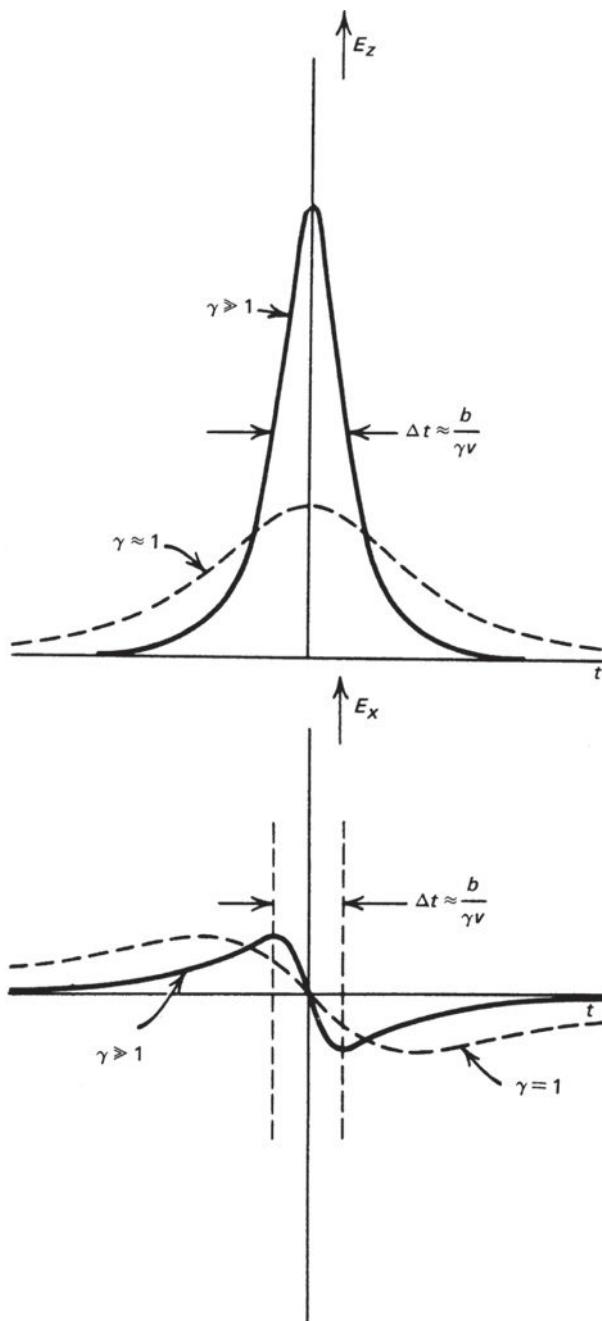


Fig. 5.4 The electric fields E_x and E_z of a relativistically moving charged particle as observed from the laboratory frame of reference S. The cases of a non-relativistic particle, $\gamma = 1$ (dashed line) and a relativistic particle, $\gamma \gg 1$ (solid line), are compared (Jackson, 1999).

velocity case, whereas the half-width of the pulse E_z , or the collision time, is shorter by a factor of $1/\gamma$. The magnitude of the E_x component is smaller by a factor of $1/\gamma$ compared with the E_z component. In the ultra-relativistic limit, $v \rightarrow c$, the pulse looks very like an electromagnetic wave, with $|E_z| = c|B_y|$ propagating in the positive x -direction.

5.3.2 Relativistic ionisation losses

Because of the symmetry of the E_x field about $t = 0$, there is no net momentum impulse imparted to the electron in the x -direction. There is, however, a net momentum impulse associated with the E_z field, namely,

$$\int F_z dt = \int_{-\infty}^{\infty} e E_z dt = \frac{ze^2\gamma b}{4\pi\varepsilon_0} \int_{-\infty}^{\infty} \frac{dt}{[b^2 + (\gamma vt)^2]^{3/2}}. \quad (5.19)$$

Changing variables to $q = \gamma vt/b$,

$$\int_{-\infty}^{\infty} F_z dt = \frac{ze^2\gamma b}{4\pi\varepsilon_0} \frac{2}{\gamma vb^2} \int_0^{\infty} \frac{dq}{(1+q^2)^{3/2}} = \frac{ze^2}{2\pi\varepsilon_0 vb}, \quad (5.20)$$

exactly the same as expression (5.3). This should not be unexpected because the argument given in Sect. 5.2 indicates that it is the product of E_z and the collision time which determines the magnitude of the momentum impulse – E_z increases by a factor γ while τ decreases by the same factor.

The integration over collision parameters proceeds as in the non-relativistic case and so all we need worry about are the values of b_{\max} and b_{\min} to include inside the logarithmic term. The correct form may be found either by asking how the values of b_{\max} and b_{\min} change in the relativistic case, or by making a relativistic generalisation of the logarithmic form $\ln(E_{\max}/\bar{I})$, when the high energy particle is relativistic.

In the first approach, b_{\max} is greater by a factor γ because the duration of the impulse is shorter by this factor. In the case of b_{\min} , the transverse momentum of the electron is greater by a factor γ and hence, because of the Heisenberg uncertainty principle,

$$\Delta x \approx b_{\min} = \frac{\hbar}{\Delta p} \propto \gamma^{-1}. \quad (5.21)$$

Thus, we expect the logarithmic term to have the form $\ln(2\gamma^2 m_e v^2 / \bar{I})$. The second approach is a useful exercise in relativity.

5.3.3 Relativistic collision between a high energy particle and a stationary electron

The momentum four-vectors of the high energy particle and the electron in the laboratory frame of reference are (see Appendix A.8.2, equation A.44);

$$\begin{aligned} \text{high energy particle} & [\gamma M, \gamma M \mathbf{v}] = [\gamma M, \gamma M v, 0, 0], \\ \text{electron} & [m_e, 0, 0, 0]. \end{aligned}$$

We transform both four-vectors into a frame of reference moving at velocity V_F , for which the Lorentz factor is $\gamma_F = (1 - V_F^2/c^2)^{-1/2}$ and $V_F \parallel \mathbf{v}$. Therefore, the relativistic

three-momenta are:

$$\begin{array}{ll} \text{high energy particle} & (\gamma M v)' = \gamma_F(\gamma M v - V_F \gamma M), \\ \text{electron} & p'_e = \gamma_F(0 - V_F m_e). \end{array}$$

In the centre of momentum frame $(\gamma M v)' + p'_e = 0$ and hence,

$$V_F = \frac{\gamma M v}{m_e + \gamma M}. \quad (5.22)$$

In this frame of reference, the relativistic three-momentum of the electron is $-\gamma_F V_F m_e$, that is, the particle is travelling in the negative x' -direction. The maximum energy exchange is obtained if the electron is sent back along the positive x' -direction following the collision. Since the collision is elastic, its three-momentum is $+\gamma_F V_F m_e$ and the zeroth component of the four-vector, the total energy, is unchanged in the centre of momentum frame of reference. Now we transform the four-momentum $[\gamma_F m_e, \gamma_F V_F m_e, 0, 0]$ back into the laboratory frame of reference. Transforming the zeroth component of the momentum four-vector using the inverse Lorentz transformation, we have

$$(\gamma m_e)_{\text{in } S} = \gamma_F \left(\gamma_F m_e + \frac{V_F}{c^2} \gamma_F V_F m_e \right). \quad (5.23)$$

Therefore, the total energy in S is $\gamma_F^2 m_e c^2 (1 + V_F^2/c^2)$. Correspondingly, the maximum kinetic energy of the electron is

$$\gamma_F^2 m_e c^2 (1 + V_F^2/c^2) - m_e c^2 = 2(V_F^2/c^2) \gamma_F^2 m_e c^2.$$

Now, $m_e \ll \gamma M$ and hence $V_F \approx v$; $\gamma_F \approx \gamma$. In the ultra-relativistic limit, the maximum energy transfer to the electron is

$$E_{\max} = 2\gamma^2 m_e v^2. \quad (5.24)$$

If we use this expression for E_{\max} , we recover the same logarithmic factor as before,

$$\ln(2\gamma^2 m_e v^2 / \bar{I}). \quad (5.25)$$

5.3.4 The Bethe–Bloch formula

The exact result derived from relativistic quantum theory is given by the Bethe–Bloch formula

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi \epsilon_0^2 m_e v^2} \left[\ln \left(\frac{2\gamma^2 m_e v^2}{\bar{I}} \right) - v^2/c^2 \right]. \quad (5.26)$$

We have succeeded in deriving this formula except for the final factor $-v^2/c^2$ which is always small. As discussed earlier, \bar{I} is treated as a parameter to be fitted to laboratory experimental data.

According to the Bethe–Bloch formula, the energy loss rate depends only upon the velocity of the particle and its charge. The dependence of the loss rate upon the kinetic energy of the particle is shown schematically in Fig. 5.5. For velocities $v \ll c$, or kinetic

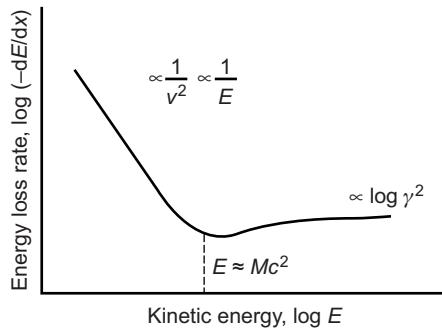


Fig. 5.5

A schematic representation of the energy loss rate due to ionisation losses.

energies $E \ll Mc^2$, the ionisation loss rate decreases as v^{-2} or E_{kin}^{-1} . At kinetic energies $E \gg Mc^2$, the loss rate increases only logarithmically with increasing energy, as $\ln \gamma^2$ according to our analysis. For kinetic energies $E_{\text{kin}} \sim Mc^2$, there is a minimum loss rate.

These results are found to be satisfactory for not-too-relativistic high energy particles in not-too-dense materials. For very high energies and dense media, the Bethe–Bloch formula overestimates the losses of the highest energy particles. The reason for this is that it has been assumed that the energy transfers to the electrons are added incoherently, that is, we assumed that there is no net reaction of the electrons back on the field of the high energy particle, which is equivalent to saying that the polarisation of the medium has been neglected. So far the interactions have been assumed to take place in free space and this holds good for interactions which do not extend to many atomic diameters. For highly relativistic particles, however, the upper limit to the range of collision parameters is $\gamma v / 4v_0$ and we cannot neglect collective effects for the most energetic particles. Jackson splits up the range of collision parameters at a value b_0 into near and distant encounters and then treats the distant ones as if they took place in a medium having a refractive index ϵ (Jackson, 1999):

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi \epsilon_0^2 m_e v^2} \left[\ln \left(\frac{\gamma m_e v}{\hbar} b_0 \right) + \ln \left(\frac{b(\gamma, \epsilon)}{b_0} \right) - \frac{v^2}{c^2} \right]. \quad (5.27)$$

Since b_0 appears in both logarithms, it is not too important to use an exact value for it. This phenomenon is known as the *density effect* and was first discussed by Fermi. Jackson shows that, in the extreme relativistic limit, the second term in square brackets is $\ln(1.123 c / b_0 \omega_p)$, where ω_p is the plasma frequency, $\omega_p = (N_e e^2 / \epsilon_0 m_e)^{1/2}$. To recover the previous formula, Jackson shows that the term should be replaced by $\ln(1.123 \gamma c / b_0 \omega)$, where $\omega = \bar{I}/\hbar$.

5.4 Practical forms of the ionisation loss formulae

The energy loss formulae do not involve explicitly the mass of the high energy particle but only its velocity v , or equivalently its Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$ and its charge z . The mass of the high energy particle can be written $M \approx N_{\text{nucl}} m_{\text{nucl}}$, where N_{nucl} is the

number of nucleons in the nucleus and m_{nucl} is the average nucleon mass, which is roughly that of the proton or neutron, that is, $m_{\text{nucl}} = (m_p + m_n)/2 \approx m_p \approx m_n$. Therefore, since the kinetic energy of the particle is $(\gamma - 1)Mc^2$, the kinetic energy per nucleon is

$$(\gamma - 1)Mc^2/N_{\text{nucl}} = (\gamma - 1)m_{\text{nucl}}c^2. \quad (5.28)$$

Thus, if we have some way of measuring the charge z of the particle, the ionisation losses measure its *kinetic energy per nucleon*.

Suppose the atomic number of the medium through which the high energy particle passes is Z and the number density of atoms is N . Then, $N_e = NZ$ and so

$$-\frac{dE}{dx} = \frac{z^2 e^4 N Z}{4\pi \epsilon_0^2 m_e v^2} \left[\ln \left(\frac{2\gamma^2 m_e v^2}{\bar{I}} \right) - v^2/c^2 \right] = z^2 N Z f(v). \quad (5.29)$$

dE/dx is often referred to as the *stopping power* of the material. It can also be expressed, not in terms of length, but in terms of the total mass per unit cross-section traversed by the particle. Thus, if a particle travels a distance x through material of density ρ , it is said to have traversed ρx kg m⁻² of the material. Then, writing $\rho x = \xi$,

$$-\frac{dE}{d\xi} = z^2 f(v) \frac{NZ}{\rho} = z^2 f(v) \frac{Z}{m}, \quad (5.30)$$

where m is the mass of a nucleus of the material. The benefit of expressing the losses in this way is that Z/m is rather insensitive to Z for all the stable elements. For light elements Z/m is $(1/2 m_{\text{nucl}})$ while for uranium, it decreases to about $(1/2.4 m_{\text{nucl}})$. Thus, the variation of the energy loss rate from element to element is mostly due to variations in \bar{I} .

The energy loss rate, expressed as $-(dE/d\xi)/z^2$, for high energy particles passing through different materials is shown in Fig. 5.6, which is taken from Chapter 27, *Passage of particles through matter*, of *The Review of Particle Physics* (Amsler *et al.*, 2008). In this presentation, the relativistic momentum, proportional to $\gamma(v/c)$, is plotted on the ordinate, rather than the kinetic energy per nucleon. Although the diagrams are plotted for singly charged high energy particles, such as protons, muons and pions, the curves can be scaled as z^2 for nucleons of different charges. Despite the wide range of values of \bar{I} for those materials, the curves lie remarkably close together because the mean ionisation potential only appears inside the logarithm in the expression (5.29). If we measure simultaneously the energy loss $dE/d\xi$ and the momentum, or kinetic energy per nucleon of the particle, we define a single point on these loss rate diagrams and the only remaining variable is the charge z . Since the loss rate increases as z^2 , the loss rate at a given kinetic energy is a sensitive measure of z .

Another useful feature of these curves is that the minimum ionisation loss rate occurs at Lorentz factors $\gamma \approx 2$, corresponding to kinetic energies $E \approx Mc^2$. A good approximation is that the *minimum ionisation loss rate* for any species in any medium is roughly

$$-\frac{dE}{d\xi} = 0.2z^2 \text{ MeV} (\text{kg m}^{-2})^{-1} = 2z^2 \text{ MeV} (\text{g cm}^{-2})^{-1}. \quad (5.31)$$

If this ionisation loss rate is measured, we can be sure that the particle is relativistic.

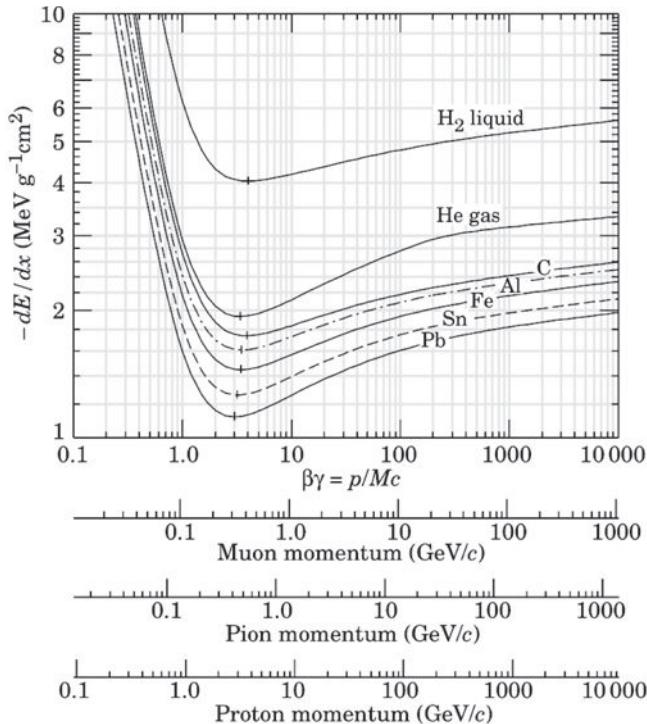


Fig. 5.6 Mean energy loss rates in liquid (bubble chamber) hydrogen, gaseous helium, carbon, aluminium, iron, tin and lead (Amsler *et al.*, 2008).

One way of estimating the total initial energy of the particle is to measure how far it travels through the medium before it is brought to rest. This distance is called the *range* R of the particle and is found by integrating the energy loss rate from the particle's initial energy E_0 until it is brought to rest:

$$R = \int_0^{E_0} \frac{dE}{(dE/dx)} . \quad (5.32)$$

This calculation breaks down at the very smallest kinetic energies but the particle travels only a very short distance once its kinetic energy falls below that at which our calculation is valid. As before, $-dE/d\xi = z^2 f(v) (Z/m)$ where Z/m is roughly constant and so

$$R = \frac{m}{Z z^2} \int_0^{E_0} \frac{dE}{f(v)} . \quad (5.33)$$

Now

$$E = (\gamma - 1)Mc^2 ; \quad dE = d(\gamma Mc^2) = Mv\gamma^3 dv , \quad (5.34)$$

and so

$$\frac{Rz^2}{M} = \frac{m}{Z} \int_0^{E_0} \frac{v\gamma^3 dv}{f(v)} , \quad (5.35)$$

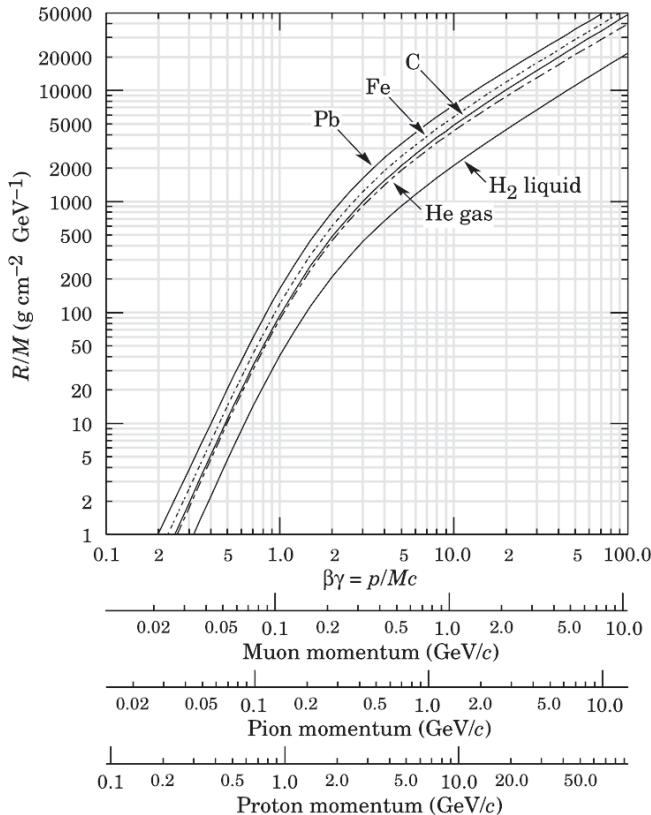


Fig. 5.7

Range of singly charged particles in liquid (bubble chamber) hydrogen, helium gas, carbon, iron, and lead. For example: for a K^+ whose momentum is $700 \text{ MeV}/c$, $\gamma v = 1.42$. For lead, we find $R/M = 396 \text{ g cm}^{-2} \text{GeV}^{-1}$, and so the range is 195 g cm^{-2} (Amsler *et al.*, 2008).

which is a function of only v_0 , γ_0 or the initial kinetic energy per nucleon of the particle. Thus, if different types of high energy particle are projected into a material, the range gives information about the initial kinetic energy per nucleon, the charge z and the mass M of the particle. This integral has been evaluated in Chapter 27, *Passage of particles through matter*, of *The Review of Particle Physics* (Amsler *et al.*, 2008) with the results shown in Fig. 5.7. These computations show how insensitive the range R , expressed as Rz^2/M , is to the material into which the particle is injected.

The process of ionisation energy loss is statistical in nature since the high energy particle makes random encounters with the electrons of the atoms of the material. There is therefore a spread in the ranges of identical high energy particles which enter the material with the same kinetic energies because some particles make more encounters than others, a phenomenon known as *straggling* which imposes a fundamental limit to the accuracy with which the initial kinetic energy can be measured. For particles of a given kinetic energy, an approximately Gaussian distribution of path lengths is expected.

What happens to the energy that is deposited in the material? A trail of ions is left behind and those electrons that are sufficiently energetic ionise further atoms of the material. For a given energy loss rate, a mean number of ion-electron pairs is produced, which is almost independent of the material. The observed values are that one ion-electron pair in air is produced for every 34 eV, in hydrogen for every 36 eV and in argon for every 26 eV. Thus, measuring the number of ion pairs produced in the material in the length dx enables the energy loss dE deposited in the material to be found.

Ionisation losses are important astrophysically in the heating and ionisation of cold, dense molecular clouds in the interstellar medium. Inside giant molecular clouds, a great deal of interstellar chemistry takes place despite the low temperature of the gas, $T \approx 10\text{--}50$ K. At these low temperatures, the gas should be completely neutral. The clouds are, however, permeated by the interstellar flux of high energy particles and their ionisation losses can ionise and heat the material of the clouds. This is believed to be the process responsible for the production of the low levels of ionisation present in molecular clouds. Estimating the ionisation rate due to the interstellar flux of high energy particles is not straightforward because it depends upon the spectrum of the particles at low energies and upon their ability to penetrate into cold clouds. The ionisation losses of protons find medical applications in cancer therapy. Figure 5.6 and equation (5.26) show that most of the energy loss of the proton occurs when the particle becomes non-relativistic. By selecting carefully the energy of the protons, the energy loss rate can be tuned to deposit most of the protons' energy at a certain path length through the body, targeting cancerous cells and leaving the healthy overlying tissue intact.

5.5 Ionisation losses of electrons

There are two important differences between the ionisation losses of electrons and those of protons and nuclei discussed above. First, the interacting particles, the high energy electron and the 'thermal' electrons, are identical, and second the electrons suffer much larger deviations in each collision than the high energy protons and nuclei, which remained effectively undeviated in the electrostatic encounters with cold electrons. The net result is, however, not so different from what was found before. The formula for the ionisation losses of an electron with total energy $\gamma m_e c^2$ is as follows (Enge, 1966):

$$-\frac{dE}{dx} = \frac{e^4 N_e}{8\pi \epsilon_0^2 m_e v^2} \left[\ln \frac{\gamma m_e v^2 E_{\max}}{2\bar{I}^2} - \left(\frac{2}{\gamma} - \frac{1}{\gamma^2} \right) \ln 2 + \frac{1}{\gamma^2} + \frac{1}{8} \left(1 - \frac{1}{\gamma} \right)^2 \right], \quad (5.36)$$

where N_e is the number density of ambient electrons and E_{\max} is the maximum kinetic energy which can be transferred to an electron in a single interaction. It is left as an exercise to carry out an exact version of the calculation performed in Sect. 5.3.3 and show that the

maximum kinetic energy transfer is

$$E_{\max} = \frac{2\gamma^2 M^2 m_e v^2}{m_e^2 + M^2 + 2\gamma m_e M}, \quad (5.37)$$

where M is the rest mass of the fast-moving particle, v its velocity and γ the corresponding Lorentz factor. In the case of electron–electron collisions, $M = m_e$ and E_{\max} takes the value

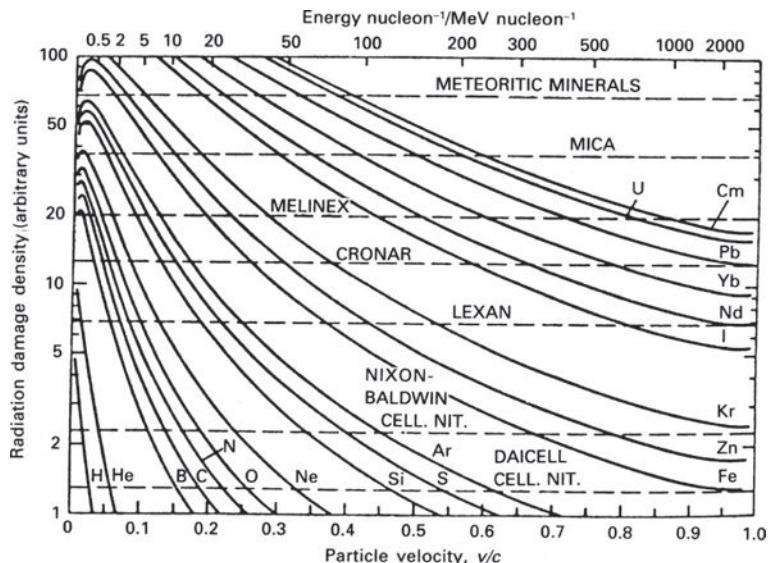
$$E_{\max} = \frac{\gamma^2 m_e v^2}{1 + \gamma}. \quad (5.38)$$

The resulting ionisation loss formula is of similar form to that given in Sect. 5.4 as may be observed by setting $z = 1$ in the loss rate (5.26). Differences are found when the loss rates are compared for protons and electrons *of the same kinetic energy*. The loss rate of the protons is then *greater* than that of the electrons, until the particles become relativistic. The physical reason for this is that a proton of the same kinetic energy as an electron moves more slowly past the electrons in the atom and hence there is a larger momentum impulse acting on the electrons. When both the proton and the electron are relativistic, however, they move past the stationary electrons at the speed of light resulting in the same momentum impulse.

5.6 Nuclear emulsions, plastics and meteorites

Two applications of the ionisation loss formula for protons and nuclei should be noted. The first is of largely historical interest and concerns the use of *nuclear emulsions*, which were direct descendants of the photographic emulsions used by Röntgen in the discovery of X-rays and by Becquerel in the discovery of radioactivity. Nuclear emulsions were designed to be sensitive to the electrons liberated by the ionisation losses of charged particles, rather than to X-rays and α -, β - and γ -rays. The emulsions consisted of a high concentration of silver bromide crystals, AgBr, embedded in a matrix of gelatin. When a high energy particle entered the emulsion, its ionisation losses resulted in a stream of electrons along its path. These electrons activated the silver bromide crystals and thus rendered them developable. During ‘development’, the activated grains were converted into grains of silver whilst the rest of the emulsion became transparent so that the track of the particle was revealed as a trail of developed grains – the number of silver grains was proportional to the energy loss rate per unit path length. The use of nuclear emulsions attained a high degree of sophistication during the 1940s and 1950s and resulted in the discovery of many short-lived particles (see Sect. 1.10.1).

Another way in which high energy particles make their presence known is through the *radiation damage* which they cause in materials. Above a certain threshold ionisation rate, the damage is permanent and these tracks can be revealed because the damaged areas have much higher chemical reactivity than undamaged areas. Therefore, by careful etching, the path of the particle can be identified without dissolving away all the material. In a good

**Fig. 5.8**

The radiation damage density, or ‘ionisation rate’ J , as a function of velocity for different incident nuclei. Approximate thresholds at which permanent tracks are formed in various materials and minerals are indicated by dashed lines (Price and Fleischer, 1971).

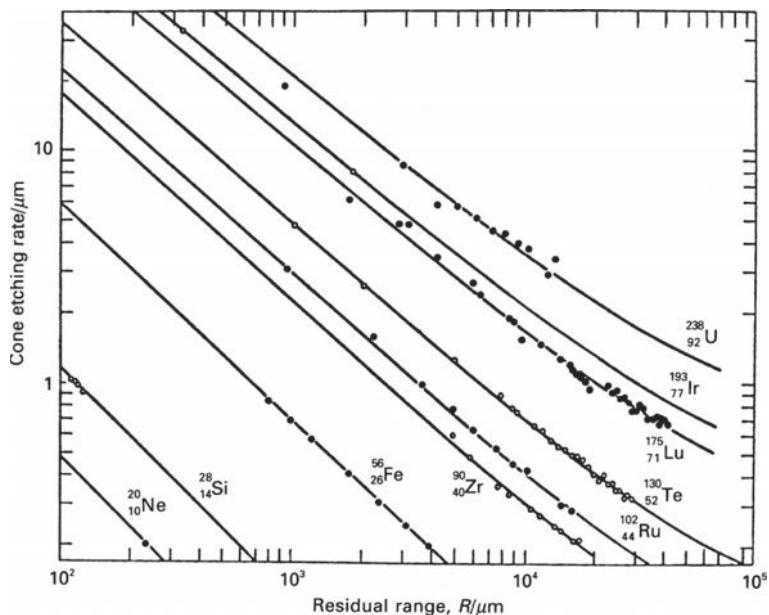
detector, the material suffers as much damage as possible by the incident particle, polymers being best for this purpose because they are long, complicated molecules and so can be disrupted and wrecked in the most interesting ways – displaced atoms, broken molecular chains, free radicals, and so on (Reedy *et al.*, 1983).

Empirically, it is found that the *radiation damage density* J can be described by a formula similar to the ionisation loss formula,

$$J = a \frac{Z^2}{v^2} \left[\ln(\gamma^2 v^2) - \frac{v^2}{c^2} + K - \delta \left(\frac{v}{c} \right) \right]. \quad (5.39)$$

The constants are now parameters to be fitted to the experimentally observed radiation damage density. Figure 5.8 shows the radiation damage rates for a wide range of different materials, from the minerals found in meteorites, through mica, Lexan polycarbonate to daicellulose nitrate, one of the most sensitive materials. For Lexan polycarbonate, for example, relativistic nuclei heavier than iodine can be detected, but only iron nuclei with velocities less than about $0.4c$ register permanent tracks.

The results of a balloon flight of 1969 are shown in Fig. 5.9. The experiment consisted of a large stack of plastics and emulsions flown for 80 hours at altitude. Seven nuclei with charges greater than iron were detected. It can be seen that some very heavy elements survived the journey through interstellar space and that one of them may well have been a uranium nucleus. On the Apollo space missions up to Apollo 17, plastic sheets were exposed on the Moon’s surface. When the astronauts from Apollo 12 brought back the camera from the Surveyor satellite, which had landed on the Moon’s surface two years earlier, etchable tracks were found in the filters of the camera.

**Fig. 5.9**

Studies of very heavy nuclei using the method of radiation damage density in plastics. The neon and silicon data are averages of measurements of many tracks from accelerator calibrations. The iron data represent the spread in measurement of about 50 stopping nuclei. The data points for the six extremely heavy nuclei have etch rates measured at many positions along their trajectories in a large stack of Lexan polycarbonate (Price and Fleischer, 1971).

Figure 5.8 shows that *meteoric materials* are sensitive to cosmic rays heavier than about iron and similar analyses can be made of samples of lunar rocks which have been exposed to cosmic rays. The study of meteorites is an enormous subject and provides many crucial clues about the early history of the Solar System. Meteorites are interplanetary rocks which reach the surface of the Earth without being completely vaporised by ablation in the Earth's atmosphere. The material of the meteorites is as old as the Solar System, that is, about 4.6×10^9 years old. It is inferred that the parent bodies of the meteorites formed in the very early Solar System and it is probable that the asteroids, which form the broad asteroid belt between Mars and Jupiter, are the meteoritic parent bodies. Meteorites are formed by fragmentation of these asteroids, probably in collisions between asteroidal bodies. When the meteorites are broken off from their parent bodies, they are exposed to the flux of high energy particles within the Solar System.

The meteorites contain crystals which behave in the same way as the plastic materials described above in that, when they are bombarded with high energy particles, etchable tracks are created within the body of the crystals. Although the volume of the crystals in the meteorites is very small, the exposure times to the cosmic rays can be very long and hence they provide information about the average cosmic ray flux over very long time intervals. Etching techniques are used to reveal the fossil tracks of cosmic rays, the etchant seeping through very fine faults in the crystals which are then rendered visible by silvering. The example presented in Fig. 5.10a shows a meteoritic sample and Fig. 5.10b one from a

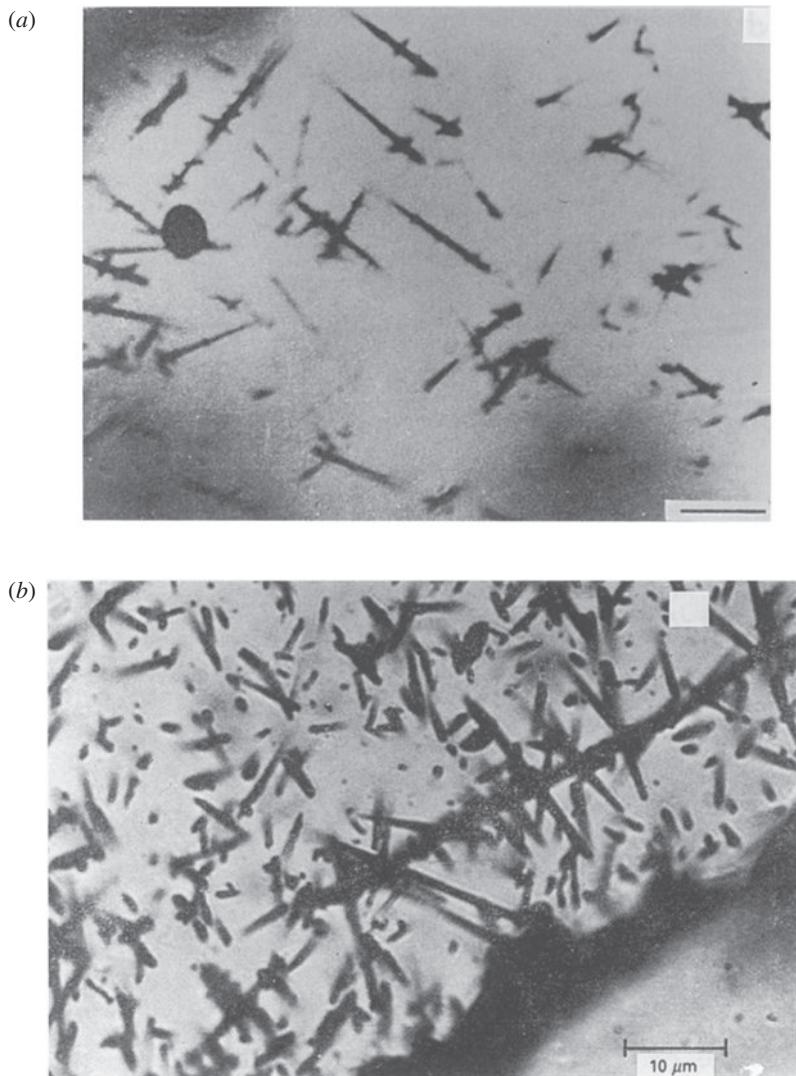


Fig. 5.10 Photomicrographs of tracks of heavy elements in meteoritic and lunar samples. (a) A typical example of the tracks seen in meteoritic crystals. Most of these tracks are iron nuclei (Caffee *et al.*, 1988). (b) Tracks in lunar feldspar from lunar rock 14310 show large numbers of iron tracks, as well as one of a much heavier nucleus (Lal, 1972).

sample of lunar rock brought back by the Apollo 14 astronauts. The latter contains many short tracks due to iron nuclei but there are also much longer tracks associated with elements with atomic numbers greater than that of iron. The particles responsible for forming these tracks may be either Galactic cosmic rays or high energy particles accelerated in solar flares. The distinction between these two types of cosmic rays is that the solar cosmic rays are generally of very much lower energy than the Galactic cosmic rays, very few indeed being observed with energies greater than 1 GeV. Consequently, they penetrate less than a few millimetres beneath the surface of the meteorite. In contrast, the Galactic cosmic rays

Table 5.1 Radioactive nuclides created by spallation in meteorites (Reedy *et al.*, 1983).

Radionuclide	Half-life (years)	Main targets	Particles
³ H	12.323	O, Mg, Si	GCR, SCR
¹⁰ Be	1.6×10^6	O, Mg, Si, (N)	GCR
¹⁴ C	5730	O, Mg, Si, (N)	GCR, SCR
²² Na	2.602	Mg, Al, Si	SCR, GCR
²⁶ Al	7.16×10^5	Al, Si, (Ar)	SCR GCR
³² Si	105	(Ar)	GCR
³⁶ Cl	3.0×10^5	Ca, Fe, (Ar)	GCR
³⁷ Ar	35.0 days	Ca, Fe	GCR, SCR
³⁹ Ar	269	K, Ca, Fe	GCR
⁴⁰ K	1.28×10^9	Fe	GCR
⁴⁶ Sc	83.82 days	Fe, Ti	GCR
⁴⁸ V	15.97 days	Fe, Ti	GCR, SCR
⁵³ Mn	3.7×10^6	Fe	SCR, GCR
⁵⁴ Mn	312.2 days	Fe	SCR, GCR
⁵⁵ Fe	2.7	Fe	SCR, GCR
⁵⁶ Co	78.76 days	Fe	SCR
⁵⁹ Ni	7.6×10^4	Fe, Ni	GCR, SCR
⁶⁰ Co	5.272	Co, Ni	GCR
⁸¹ Kr	2.1×10^5	Sr, Zr	GCR, SCR
¹²⁹ I	1.6×10^7	Te, Ba, La, Ce	GCR

have very much higher energies and can penetrate much more deeply into the meteorite. The tracks detected at depths greater than 1 cm into the meteorite are certainly of Galactic origin.¹

A second way in which the cosmic rays provide crucial information is through the spallation products which they produce in the material of the meteorite – we will have much more to say about spallation, the process of chipping nucleons from heavy nuclei by collisions with cosmic rays, in Chap. 10. The spallation products produced by high energy cosmic rays are not only lighter elements, as indicated in Table 5.1, but also neutrons which can interact with the nuclei of the minerals to produce rare isotopes which are then trapped inside the meteorite. Important examples of stable nuclei produced as cosmogenic nuclides include rare isotopes such as ³He, ²¹Ne and ³⁸Ar. The abundances of the stable elements continue to increase linearly in abundance with time, if the interplanetary flux of cosmic rays is constant. Wasson, for example, quotes rates of formation of ³He and ²¹Ne of $2 \times 10^{-17} \rho$ and $3.5 \times 10^{-18} \rho$ particles per year respectively, where ρ is the density of the material of the meteorite in kilograms per cubic metre, assuming the present intensity of the interstellar flux of cosmic rays (Wasson, 1985).

¹ Recent examples of the use of meteorites as tools for studying the early Solar System through cosmic ray bombardment are given in the review by Eugster and his colleagues (Eugster *et al.*, 2006)

The spallation process in meteorites also accounts for the observation of isotopes with short half-lives, such as tritium ${}^3\text{H}$, ${}^{14}\text{C}$ and ${}^{10}\text{Be}$, their half-lives being 12.5, 5.6×10^3 and 2.5×10^6 years, respectively, as well as a host of rarer radioactivities. Table 5.1 shows a list of cosmic ray induced radionuclides, which have been measured in terrestrial and extraterrestrial matter (Reedy *et al.*, 1983). This table includes the principal target nuclei as well as an indication of the source of the high energy particles which are responsible for their formation, GCR meaning Galactic cosmic rays and SCR solar cosmic rays.

These two techniques can be used to provide estimates of the exposure ages of the meteorites to the cosmic rays. Many of the meteorites must have fragmented from their parent bodies more than about 10^7 years ago and there is an age distribution which extends up to 10^9 years and more. These studies show that the cosmic ray flux must have been within about 50% of its present value over the last 10^9 years (Reedy *et al.*, 1983). A literal interpretation of the results suggests that over the last 10^7 years, the flux of cosmic rays has been about 50% greater than it was during the preceding 10^9 years. Thus, it seems that our Solar System has been bombarded by roughly the same flux of cosmic rays for the last billion years.

5.7 Dynamical friction

Having analysed ionisation losses, it is straightforward to adapt the results for gravitational rather than electrostatic interactions. In the gravitational case, the deceleration of a fast-moving star by gravitational interactions with other stars is referred to as *dynamical friction* and is the process by which a stellar system establishes a thermal distribution of velocities by energy exchange. The following arguments, developed by my colleague Rashid Sunyaev and me some years ago, are in no sense original but they show how helpful working by physical analogy can be.

By analogy with the analysis of Sect. 5.2, we consider the interaction of a massive, fast-moving star with a cluster of stars. The star transfers kinetic energy to the other stars in the cluster and so loses energy. The difference between the electrostatic and gravitational cases is that gravity is very much weaker than the electrostatic force. The same type of formula for the loss of kinetic energy of the massive star as that derived in Sect. 5.2 is, however, expected. To convert from the electrostatic to the gravitational case, the forms of the inverse square laws of electrostatics and gravitation can be compared:

$$F = \frac{(ze)e}{4\pi\varepsilon_0 r^2}; \quad F = \frac{GMm}{r^2}. \quad (5.40)$$

We therefore replace $(ze)e/4\pi\varepsilon_0$ by GMm , where M is the mass of the fast-moving star and m is the mass of each of the swarm of less massive stars. We make the following identifications:

$$ze/(4\pi\varepsilon_0)^{1/2} \equiv G^{1/2}M; \quad e/(4\pi\varepsilon_0)^{1/2} \equiv G^{1/2}m. \quad (5.41)$$

If the number density of particles is N , the energy loss rate due to gravitational interactions can be found directly from (5.6),

$$-\frac{dE}{dx} = \frac{4\pi G^2 M^2 m N}{v^2} \ln\left(\frac{b_{\max}}{b_{\min}}\right). \quad (5.42)$$

This relation can be written in terms of the mass density $\rho = Nm$ through which the particle moves:

$$-\frac{dE}{dx} = \frac{4\pi G^2 M^2 \rho}{v^2} \ln\left(\frac{b_{\max}}{b_{\min}}\right). \quad (5.43)$$

This is the energy loss rate due to the force of dynamical friction acting upon the massive star.

We can therefore define a loss-time τ during which the massive particle loses its initial kinetic energy $E = \frac{1}{2}Mv^2$ in transferring energy to the light particles,

$$\tau = \frac{E}{(-dE/dt)} = \frac{\frac{1}{2}Mv^2}{v(-dE/dx)} = \frac{v^3}{8\pi G^2 M m N \ln(b_{\max}/b_{\min})}. \quad (5.44)$$

The loss-time τ is closely related to the *gravitational relaxation time* τ_r of a star in the cluster, meaning the time it takes to change the energy of a typical star in the cluster by roughly a factor of 2 due to random gravitational encounters with other stars. This is also roughly the time to establish equipartition of kinetic energy with the other stars in the cluster and so to set up a Maxwellian velocity distribution. A much more complete analysis is needed to describe the interaction of particles of the same mass which are all in motion. The expression for the gravitational relaxation time τ_r is

$$\tau_r = \frac{3\sqrt{2}}{32\pi} \frac{v^3}{G^2 m^2 N \ln(b_{\max}/b_{\min})} \quad (5.45)$$

(Spitzer and Hart, 1971). The similarity of this relation with the one we derived above may be observed by setting $M = m$ in (5.44).

Let us apply this result to a cluster of stars which has yet to come into *thermal equilibrium* through their mutual gravitational interactions. There are N_c stars in the cluster which has radius R . A natural upper bound to the range of collision parameters, b_{\max} , is the radius of the cluster, since there will not be gravitational interactions at greater distances. As before, a lower limit is set by the requirement that the particles cannot exchange more than their kinetic energies:

$$\frac{1}{2}m v^2 \approx \frac{Gm^2}{b_{\min}}; \quad b_{\min} \approx \frac{2Gm}{v^2}. \quad (5.46)$$

Therefore,

$$\frac{b_{\max}}{b_{\min}} \approx \frac{Rv^2}{2Gm}. \quad (5.47)$$

The virial theorem states that, in dynamical equilibrium, the total kinetic energy of the particles in the cluster is half the gravitational potential energy (Sect. 3.5.1). Hence,

$$U = 2T; \quad \frac{1}{2} \frac{GM_c^2}{R} \approx N_c m v^2, \quad (5.48)$$

where the mass of the cluster M_c is $N_c m$. Therefore, $GM_c^2 \approx 2RN_c m v^2$ and so from (5.47),

$$\frac{b_{\max}}{b_{\min}} \approx \frac{N_c}{4}. \quad (5.49)$$

Thus, the gravitational relaxation time can be written

$$\tau_r = \frac{3\sqrt{2}v^3}{32\pi G^2 m^2 N \ln(N_c/4)}. \quad (5.50)$$

To apply this result to star clusters, it is convenient to relate the relaxation time τ_r to the crossing time of a typical star in the cluster, $\tau_{\text{cr}} = R/v$. Noting that $4\pi R^3 N/3 = N_c$ and using the virial theorem in the form (5.48), we find

$$\tau_r = \frac{\sqrt{2}}{32} \frac{N_c}{\ln(N_c/4)} \tau_{\text{cr}}. \quad (5.51)$$

Binney and Tremaine (2008) quote a similar expression

$$\tau_r = 0.1 \frac{N_c}{\ln N_c} \tau_{\text{cr}}. \quad (5.52)$$

Let us apply these results to globular clusters and galaxies. Typical parameters for a globular star cluster are: $R = 10$ pc, $M = 0.3M_\odot$, $v = 8$ km s $^{-1}$, $N_c = 10^6$ – these figures are self-consistent according to the virial theorem. The crossing time is then about 10^6 years and the relaxation time of the order of 10^{10} years. Therefore, there is time for the stars to develop into a relaxed bound system, particularly when account is taken of the fact that globular clusters are strongly centrally concentrated – in the central regions, the relaxation time is much less than that of the cluster as a whole. For galaxies with 10^{11} stars and crossing times of the order 10^8 years, there is certainly not time for the stars to be thermalised according to (5.52) – rather, the stars behave like a collisionless fluid and their dynamics are determined by the mean gravitational potential due to the galaxy as a whole.

Although the above analysis applies for stellar objects, let us apply the same calculation to the galaxies in a cluster of galaxies, recognising that now the ‘particles’ are extended objects. Values consistent with the virial theorem would be $R = 2.5$ Mpc, $N = 1000$, $M = 10^{11} M_\odot$ and $v = 10^3$ km s $^{-1}$. The crossing time would then be of the order of 10^9 years and the gravitational relaxation time τ_r about 10^{11} years. Thus, in general, the galaxies in a cluster will not have come into equipartition, although they must have attained gravitational equilibrium according to the virial theorem. Regular clusters are, however, centrally concentrated and the most massive galaxies, $M \approx 10^{13} M_\odot$, have relaxation times with the lighter members and with each other which are much shorter than the above estimate. Indeed, the most massive galaxies can relax in less than 10^{10} years and this can account for the observation that the most massive galaxies in regular clusters are found in their centres, having transferred their kinetic energy to the lighter members.

Radiation of accelerated charged particles and bremsstrahlung of electrons

6.1 Introduction

Bremsstrahlung, or *free-free emission*, appears in many different guises in astrophysics. Applications include the radio emission of compact regions of ionised hydrogen at temperature $T \approx 10^4$ K, the X-ray emission of binary X-ray sources at $T \approx 10^7$ K and the diffuse X-ray emission of intergalactic gas in clusters of galaxies, which may be as hot as $T \approx 10^8$ K. It is also an important loss mechanism for relativistic cosmic ray electrons. Before proceeding to the analysis of the bremsstrahlung of electrons, we need to establish a number of general results concerning the electromagnetic radiation of accelerated charged particles and its spectrum. These results will be of wide applicability to the many radiation processes studied in this book.

6.2 The radiation of accelerated charged particles

6.2.1 Relativistic invariants

Gould has provided an excellent introduction to the use of relativistic invariants in the study of electromagnetic processes (Gould, 2005). We will develop a number of these in the course of this exposition. The first of these is the transformation of the energy loss rate by electromagnetic radiation as observed in different inertial frames of reference, that is, how dE/dt changes from one inertial frame of reference to another.

In fact, dE/dt is a Lorentz invariant between inertial frames of reference. The simplest way of obtaining this result is to note that the energy dE emitted in the form of radiation in the time dt is the zeroth component of the momentum four-vector $[dE/c, d\mathbf{p}]$ and $c dt$ is the zeroth component of the displacement four-vector $[c dt, d\mathbf{r}]$.¹ Therefore, both the energy dE and the time interval dt transform in the same way between inertial frames of reference and so their ratio dE/dt is also an invariant. To express this result in another way, the momentum and displacement four-vectors are parallel four-vectors and so transform in the same way between inertial frames of reference.

¹ For the relativistic notation and conventions used throughout this book, see Appendix A.8.2.

This result can also be appreciated from the following argument. In the moving instantaneous rest frame of an accelerated charged particle, the total energy loss dE' has dipole symmetry and so is emitted with zero net momentum (see Sect. 6.2.2 below). Therefore, its four-momentum can be written $[dE'/c, 0]$. This radiation is emitted in the interval of time dt' , which is the zeroth component of the displacement four-vector $[c dt', 0]$. Using the inverse Lorentz transforms to relate dE' and $c dt'$ to dE and $c dt$, we find

$$dE = \gamma dE'; \quad dt = \gamma dt', \quad (6.1)$$

and hence

$$dE/dt = dE'/dt'. \quad (6.2)$$

6.2.2 The radiation of an accelerated charged particle – J. J. Thomson's treatment

The expressions for the properties of the electromagnetic radiation of accelerated charged particles are central to the understanding of radiation processes in high energy astrophysics and so two versions are presented. The normal derivation proceeds from Maxwell's equations and involves writing down the retarded potentials for the electric and magnetic fields at some distant point \mathbf{r} from the accelerated charge (see Sect. 6.2.3). It is, however, instructive to begin with a remarkable argument due to J. J. Thomson which indicates very clearly the origins of the radiation of an accelerated charged particle and the polarisation properties of the radiation. This argument was given by Thomson in his derivation of the formula for the Thomson scattering cross-section σ_T in the context of the scattering of X-rays by electrons (Thomson, 1906).

Consider a charge q stationary at the origin O of some inertial frame of reference S at time $t = 0$. Suppose the charge suffers a small acceleration to velocity Δv in the short interval of time Δt . Thomson visualised the resulting field distribution in terms of the electric field lines attached to the accelerated charge. After time t , we can distinguish between the field configuration inside and outside a sphere of radius $r = ct$ centred on the origin of S , recalling that electromagnetic disturbances are propagated at the speed of light in free space (Fig. 6.1a). Outside the sphere, the field lines do not yet know that the charge has moved away from the origin because information cannot travel faster than the speed of light and therefore they are radial, centred on O . Inside this sphere, the field lines are radial about the origin of the frame of reference which is centred on the moving charge. Between these two regions, there is a thin shell of thickness $c\Delta t$ in which we have to join up corresponding electric field lines (see Fig. 6.1a). Geometrically, it is clear that there must be a component of the electric field in the circumferential direction in this shell, that is, in the \mathbf{i}_θ -direction. This 'pulse' of electromagnetic field is propagated away from the charge at the speed of light and consequently represents an energy loss from the accelerated charged particle.

Let us work out the strength of the electric field in the pulse. We assume that the increment in velocity Δv is very small, that is, $\Delta v \ll c$, and therefore it is safe to assume that the field lines are radial not only at $t = 0$ but also at time t in the frame of reference S . There will, in

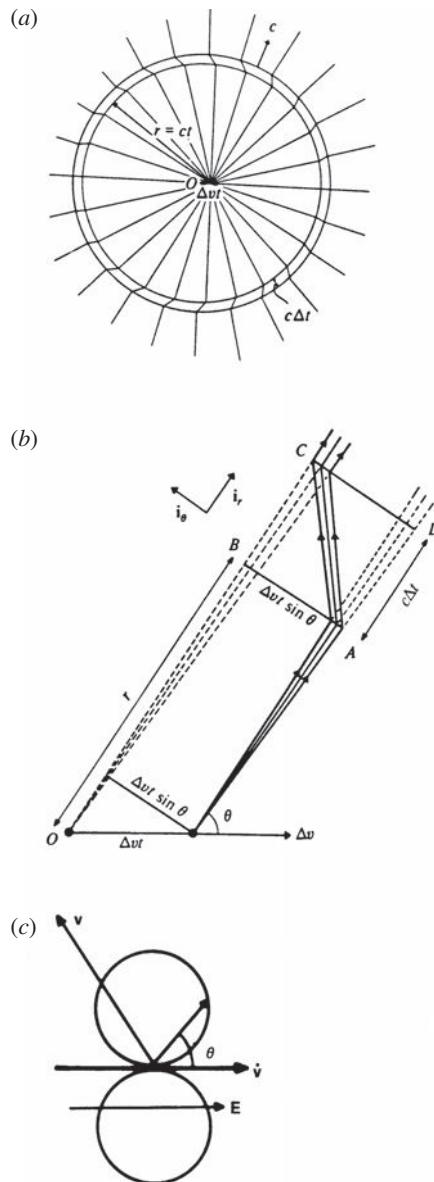


Fig. 6.1 (a) Illustrating J.J. Thomson's method of evaluating the radiation of an accelerated charged particle. The diagram shows schematically the configuration of electric field lines at time t due to a charge accelerated to a velocity Δv in time Δt at $t = 0$. (b) An expanded version of part of (a) used to evaluate the strength of the azimuthal component E_θ of the electric field due to the acceleration of the electron. (c) The polar diagram of the radiation field E_θ emitted by an accelerated electron, showing the magnitude of the electric field strength as a function of polar angle θ with respect to the instantaneous acceleration vector \boldsymbol{a} . Note that the radiation properties of the charged particle in its instantaneous rest frame are independent of the velocity vector \mathbf{v} , which in general need not be parallel to \boldsymbol{a} , as illustrated in the diagram. The polar diagram $E_\theta \propto \sin \theta$ corresponds to circular lobes with respect to the acceleration vector (Longair, 2003).

fact, be small aberration effects associated with the velocity Δv , but these are second-order compared with the gross effects we are discussing. We may therefore consider a small cone of field lines at an angle θ with respect to the acceleration vector of the charge at $t = 0$ and a similar one at the later time t when the charge is moving at a constant velocity Δv (Fig. 6.1b). We now join up electric field lines between the two cones through the thin shell of thickness $c\Delta t$ as shown in the diagram. The strength of the E_θ -component of the field is given by the number of field lines per unit area in the i_θ -direction. From the geometry of Fig. 6.1(b), which exaggerates the discontinuities in the field lines, the E_θ component is given by the relative sizes of the sides of the rectangle $ABCD$, that is,

$$\frac{E_\theta}{E_r} = \frac{\Delta v t \sin \theta}{c \Delta t}. \quad (6.3)$$

But, E_r is given by Coulomb's law,

$$E_r = \frac{q}{4\pi\epsilon_0 r^2}, \quad \text{where } r = ct.$$

Therefore

$$E_\theta = \frac{q(\Delta v/\Delta t) \sin \theta}{4\pi\epsilon_0 c^2 r}.$$

$\Delta v/\Delta t$ is the acceleration $|\mathbf{a}|$ of the charge and hence

$$E_\theta = \frac{q|\mathbf{a}| \sin \theta}{4\pi\epsilon_0 c^2 r}. \quad (6.4)$$

Notice that the radial component of the field decreases as r^{-2} , according to Coulomb's law, but the tangential component decreases only as r^{-1} , because in the shell, as t increases, the field lines become more and more stretched in the E_θ -direction, as can be appreciated from (6.3). Alternatively, we can write $q\mathbf{a} = \ddot{\mathbf{p}}$, where \mathbf{p} is the electric dipole moment of the charge with respect to some origin, and hence

$$E_\theta = \frac{|\ddot{\mathbf{p}}| \sin \theta}{4\pi\epsilon_0 c^2 r}. \quad (6.5)$$

This electric field component represents a pulse of electromagnetic radiation, and hence the rate of energy flow per unit area per second at distance r is given by the magnitude of the Poynting vector $S = |\mathbf{E} \times \mathbf{H}| = E^2/Z_0$, where $Z_0 = (\mu_0/\epsilon_0)^{1/2}$ is the impedance of free space. The rate of energy flow through the area $r^2 d\Omega$ subtended by solid angle $d\Omega$ at angle θ and at distance r from the charge is therefore

$$Sr^2 d\Omega = - \left(\frac{dE}{dt} \right) d\Omega = \frac{|\ddot{\mathbf{p}}|^2 \sin^2 \theta}{16\pi^2 Z_0 \epsilon_0^2 c^4 r^2} r^2 d\Omega = \frac{|\ddot{\mathbf{p}}|^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3} d\Omega. \quad (6.6)$$

To find the total radiation rate $-dE/dt$, we integrate over the solid angle. Because of the symmetry of the emitted intensity with respect to the acceleration vector, we can integrate over the solid angle defined by the circular strip between the angles θ and $\theta + d\theta$, $d\Omega = 2\pi \sin \theta d\theta$:

$$-\left(\frac{dE}{dt} \right) = \int_0^\pi \frac{|\ddot{\mathbf{p}}|^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3} 2\pi \sin \theta d\theta. \quad (6.7)$$

We find the key result

$$-\left(\frac{dE}{dt}\right) = \frac{|\ddot{\mathbf{p}}|^2}{6\pi\varepsilon_0 c^3} = \frac{q^2 |\mathbf{a}|^2}{6\pi\varepsilon_0 c^3}. \quad (6.8)$$

This result is sometimes referred to as *Larmor's formula* – precisely the same result comes out of the full theory. These formulae embody the three essential properties of the radiation of an accelerated charged particle.

- (i) The total radiation rate is given by Larmor's formula (6.8). Notice that, in this formula, the acceleration is the *proper acceleration* of the charged particle in the relativistic sense and that the radiation loss rate is that measured in the *instantaneous rest frame* of the particle.
- (ii) The *polar diagram* of the radiation is of *dipolar* form, that is, the electric field strength varies as $\sin\theta$ and the power radiated per unit solid angle varies as $\sin^2\theta$ where θ is the angle with respect to the acceleration vector of the particle (Fig. 6.1c). Notice that there is no radiation along the acceleration vector and the field strength is greatest at right angles to it.
- (iii) The radiation is *polarised*, the electric field vector, as measured by a distant observer, lying in the direction of the acceleration vector of the particle as projected onto the sphere at distance \mathbf{r} from the charged particle, that is, in the direction of the polar angle unit vector \mathbf{i}_θ (see Fig. 6.1b).

These are very useful rules which enable us to understand the radiation properties of particles in many different astrophysical situations. It is important to remember that these rules are applicable in the *instantaneous rest frame* of the particle and we have to look carefully at what an external observer sees if the particle is moving at a relativistic velocity.

6.2.3 The radiation of an accelerated charged particle – from Maxwell's equations

The standard analysis begins with Maxwell's equations in free space:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (6.9a)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}, \quad (6.9b)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (6.9c)$$

$$\nabla \cdot \mathbf{E} = \rho_e / \varepsilon_0. \quad (6.9d)$$

We introduce the *scalar* and *vector potentials*, ϕ and \mathbf{A} respectively, in order to simplify the evaluation of the vector fields \mathbf{E} and \mathbf{B} at distance \mathbf{r} from the accelerated charge through the definitions

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (6.10a)$$

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi. \quad (6.10b)$$

The reason for this is that the fields \mathbf{E} and \mathbf{B} are the components of a four-tensor. It is therefore much easier to work in terms of the four-vector potential $[\phi/c, \mathbf{A}]$ and then take the derivatives (6.10) to find \mathbf{E} and \mathbf{B} . Substituting for \mathbf{E} and \mathbf{B} in (6.9b),

$$\nabla \times (\nabla \times \mathbf{A}) = \mu_0 \mathbf{J} - \frac{1}{c^2} \frac{\partial}{\partial t} \left(\frac{\partial \mathbf{A}}{\partial t} + \nabla \phi \right). \quad (6.11)$$

We recall that

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (6.12)$$

and therefore, substituting and interchanging the order of the time and spatial derivatives,

$$\begin{aligned} \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c^2} \frac{\partial}{\partial t} (\nabla \phi), \\ \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= -\mu_0 \mathbf{J} + \nabla \left[\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right]. \end{aligned} \quad (6.13)$$

Making the same substitutions for \mathbf{E} and \mathbf{B} into (6.9d),

$$\nabla \cdot \left(-\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi \right) = \frac{\rho_e}{\epsilon_0},$$

and so, interchanging the order of differentiation,

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) + \nabla^2 \phi = -\frac{\rho_e}{\epsilon_0}.$$

Now add $-(1/c^2)(\partial^2 \phi / \partial t^2)$ to both sides of the equation and we obtain

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho_e}{\epsilon_0} - \frac{\partial}{\partial t} \left[\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right]. \quad (6.14)$$

The equations (6.13) and (6.14) have remarkably similar forms and, if we were able to set the quantities in the square brackets of each equation equal to zero, we would obtain two simple inhomogeneous wave equations for \mathbf{A} and ϕ separately. Fortunately, we are able to do this because there is considerable freedom in the definition of the vector potential \mathbf{A} . In classical electrodynamics, \mathbf{A} only appears as the quantity which, when curled, results in the magnetic field \mathbf{B} which is what we measure in the laboratory. We can always add to \mathbf{A} the gradient of any scalar quantity and it will be guaranteed to disappear upon curling. If we write $\mathbf{A}' = \mathbf{A} + \text{grad } \chi$, then we know from (6.10a) that the value of \mathbf{B} will be unchanged. What about \mathbf{E} ? Substituting for \mathbf{A} in (6.10b),

$$\mathbf{E} = -\frac{\partial \mathbf{A}'}{\partial t} - \nabla(\phi - \dot{\chi}).$$

Thus, we need to replace ϕ by $\phi' = \phi - \dot{\chi}$. Therefore, we can express the condition that $\nabla \cdot \mathbf{A}' + (1/c^2)(\partial \phi' / \partial t)$ should vanish as follows:

$$\begin{aligned} \nabla \cdot (\mathbf{A}' - \nabla \chi) + \frac{1}{c^2} \frac{\partial}{\partial t} (\phi' + \dot{\chi}) &= 0, \\ \nabla \cdot \mathbf{A}' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} &= \nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2}. \end{aligned} \quad (6.15)$$

Thus, provided we can find a suitable function χ which satisfies (6.15), we obtain the following pair of equations separately for \mathbf{A} and ϕ :

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J}, \quad (6.16a)$$

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho_e}{\epsilon_0}. \quad (6.16b)$$

In fact, it turns out that it is possible to obtain these equations with the more restrictive requirement

$$\nabla^2 \chi - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t^2} = 0.$$

This procedure is known as *selecting the gauge* and this particular choice is known as the *Lorentz gauge* (Jackson, 1999).

Equations (6.16) have standard forms of solution:²

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} d^3 r', \quad (6.17a)$$

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho_e(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \quad (6.17b)$$

The point at which the fields are measured is \mathbf{r} and the integration is over the electric current and charge distributions throughout space. The terms in $|\mathbf{r} - \mathbf{r}'|/c$ take account of the fact that the current and charge distributions should be evaluated at *retarded times*. We now make a number of simplifications to obtain the results we are seeking. First of all, in the case of an accelerated charged particle, the integral of the product of the current density \mathbf{J} and the volume element $d^3 r'$ is just the product of its charge times its velocity,

$$\mathbf{J}\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) d^3 r' = q \mathbf{v} \delta(\mathbf{r}),$$

where $\delta(\mathbf{r})$ is the Dirac delta function. The expression for the vector potential is therefore

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{q \mathbf{v}}{r}. \quad (6.18)$$

We now take the time derivative of \mathbf{A} in order to find \mathbf{E} ,

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = -\frac{\mu_0}{4\pi} \frac{q \ddot{\mathbf{r}}}{r} = -\frac{q \ddot{\mathbf{r}}}{4\pi \epsilon_0 c^2 r}.$$

This is exactly the same expression for \mathbf{E} as (6.4) derived in Sect. 6.2.2 and so we need not repeat the rest of the argument which results in (6.8). Notice, however, that the integrals (6.17) are much more powerful tools than those used in that section. I leave as exercises to the reader the demonstration that the solutions represent outgoing electromagnetic waves from the accelerated charge and also that the \mathbf{E} and \mathbf{B} fields are orthogonal to each other and to the radial direction of propagation of the wave from the origin in the far field limit.

² I have given a simple derivation of these solutions in *Theoretical Concepts in Physics* (Longair, 2003).

Another important point is that these results are correct provided the velocities of the charges are small. A more complete analysis results in the following expressions for the field potentials which are valid for all velocities – the *Liénard–Wiechert potentials*:

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \left[\frac{q\mathbf{v}}{1 - (\mathbf{v} \cdot \mathbf{n})/c} \right]_{\text{ret}} ; \quad \phi(\mathbf{r}, t) = \frac{1}{4\pi \epsilon_0 r} \left[\frac{q}{1 - (\mathbf{v} \cdot \mathbf{n})/c} \right]_{\text{ret}}, \quad (6.19)$$

where \mathbf{n} is the unit vector in the direction of the point of observation from the moving charge. In both cases, the potentials are evaluated at retarded times relative to the location of the observer. The reason for drawing attention to these more general potentials is that the terms in the denominators, $1 - (\mathbf{v} \cdot \mathbf{n})/c$, will reappear on a number of occasions in our treatment of charges and sources of radiation moving at high velocities. For example, in the case of a particle moving towards the point of observation at a velocity close to that of light, it represents the fact that the particle almost catches up with the radiation it emits.

6.2.4 The radiation losses of accelerated charged particles moving at relativistic velocities

We often have to deal with accelerated high energy particles moving at relativistic velocities. We can adapt the results already obtained to many of these problems. It is assumed that, in the particle's instantaneous rest frame, the acceleration of the particle is small and this is normally the case. We need the following general results: first, the norm of the acceleration four-vector is an invariant in any inertial frame of reference and, second, the acceleration four-vector of the particle, \mathbf{A} , not to be confused with the vector potential \mathbf{A} of the last section, can be written

$$\mathbf{A} = \gamma \left[c \frac{\partial \gamma}{\partial t}, \frac{\partial(\gamma \mathbf{v})}{\partial t} \right] = \left[\left(\frac{\mathbf{v} \cdot \mathbf{a}}{c^2} \right) \gamma^4 c, \gamma^2 \mathbf{a} + \left(\frac{\mathbf{v} \cdot \mathbf{a}}{c^2} \right) \gamma^4 \mathbf{v} \right], \quad (6.20)$$

where the acceleration $\mathbf{a} = \ddot{\mathbf{r}}$ and the velocity of the particle $\mathbf{v} = \dot{\mathbf{r}}$ are measured in the observer's frame of reference S . In the instantaneous rest frame of the particle, S' , the acceleration four-vector is $[0, \mathbf{a}_0]$, where $\mathbf{a}_0 = (\ddot{\mathbf{r}})_0$ is the *proper acceleration* of the particle. We now equate the norms of the four-vectors in the reference frames S and S' :

$$-\mathbf{a}_0^2 = c^2 \gamma^8 (\mathbf{v} \cdot \mathbf{a}/c^2)^2 - [\gamma^2 \mathbf{a} + (\mathbf{v} \cdot \mathbf{a}/c^2) \gamma^4 \mathbf{v}]^2. \quad (6.21)$$

After a little straightforward algebra, we find

$$\mathbf{a}_0^2 = \gamma^4 [\mathbf{a}^2 + \gamma^2 (\mathbf{v} \cdot \mathbf{a})/c^2]. \quad (6.22)$$

Now, the radiation rate (dE/dt) is a Lorentz invariant (Sect. 6.2.1) and therefore

$$\left(\frac{dE}{dt} \right)_S = \left(\frac{dE'}{dt'} \right)_{S'} = \frac{q^2 |\mathbf{a}_0|^2}{6\pi \epsilon_0 c^3} = \frac{q^2 \gamma^4}{6\pi \epsilon_0 c^3} \left[\mathbf{a}^2 + \gamma^2 \left(\frac{\mathbf{v} \cdot \mathbf{a}}{c} \right)^2 \right]. \quad (6.23)$$

Notice that all the quantities \mathbf{a} , \mathbf{v} and γ are measured in S . This is a useful formula. Let us rewrite it in a slightly different form by resolving the acceleration of the particle into components parallel $a_{||}$ and perpendicular a_{\perp} to the velocity vector \mathbf{v} , that is,

$$\mathbf{a} = a_{||} \mathbf{i}_{||} + a_{\perp} \mathbf{i}_{\perp} \quad \text{and} \quad |\mathbf{a}|^2 = |a_{||}|^2 + |a_{\perp}|^2.$$

Therefore,

$$\begin{aligned} \mathbf{a}^2 + \gamma^2(\mathbf{v} \cdot \mathbf{a}/c)^2 &= |a_{\parallel}|^2 + |a_{\perp}|^2 + \gamma^2(v a_{\parallel}/c)^2, \\ &= |a_{\perp}|^2 + |a_{\parallel}|^2(1 + \gamma^2 v^2/c), \\ &= |a_{\perp}|^2 + |a_{\parallel}|^2 \gamma^2. \end{aligned} \quad (6.24)$$

Therefore, the loss rate can also be written,

$$\left(\frac{dE}{dt} \right)_S = \frac{q^2 \gamma^4}{6\pi \varepsilon_0 c^3} (|a_{\perp}|^2 + \gamma^2 |a_{\parallel}|^2). \quad (6.25)$$

These results will prove useful in the subsequent development.

6.2.5 Parseval's theorem and the spectral distribution of the radiation of an accelerated electron

The final tool we need before tackling bremsstrahlung is the decomposition of the radiation field of the electron into its spectral components. Parseval's theorem provides an elegant procedure for relating the kinematic history of the particle to its radiation spectrum.

We introduce the Fourier transform of the acceleration of the particle through the Fourier transform pair:

$$\dot{\mathbf{v}}(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \dot{\mathbf{v}}(\omega) \exp(-i\omega t) d\omega, \quad (6.26)$$

$$\dot{\mathbf{v}}(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \dot{\mathbf{v}}(t) \exp(i\omega t) dt. \quad (6.27)$$

According to Parseval's theorem, $\dot{\mathbf{v}}(\omega)$ and $\dot{\mathbf{v}}(t)$ are related by the following integral:

$$\int_{-\infty}^{\infty} |\dot{\mathbf{v}}(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |\dot{\mathbf{v}}(t)|^2 dt. \quad (6.28)$$

This is proved in all textbooks on Fourier analysis. We can therefore apply this relation to the energy radiated by a particle which has an acceleration history $\dot{\mathbf{v}}(t)$:

$$\int_{-\infty}^{\infty} \frac{dE}{dt} dt = \int_{-\infty}^{\infty} \frac{e^2}{6\pi \varepsilon_0 c^3} |\dot{\mathbf{v}}(t)|^2 dt = \int_{-\infty}^{\infty} \frac{e^2}{6\pi \varepsilon_0 c^3} |\dot{\mathbf{v}}(\omega)|^2 d\omega. \quad (6.29)$$

Now, what we really want is $\int_0^{\infty} \dots d\omega$ rather than $\int_{-\infty}^{\infty} \dots d\omega$. Since the acceleration is a real function, there is another theorem in Fourier analysis which tells us that

$$\int_0^{\infty} |\dot{\mathbf{v}}(\omega)|^2 d\omega = \int_{-\infty}^0 |\dot{\mathbf{v}}(\omega)|^2 d\omega,$$

and hence we find

$$\text{total emitted radiation} = \int_0^{\infty} I(\omega) d\omega = \int_0^{\infty} \frac{e^2}{3\pi \varepsilon_0 c^3} |\dot{\mathbf{v}}(\omega)|^2 d\omega.$$

Therefore

$$I(\omega) = \frac{e^2}{3\pi \varepsilon_0 c^3} |\dot{\mathbf{v}}(\omega)|^2. \quad (6.30)$$

This is the total energy per unit bandwidth emitted throughout the period during which the particle is accelerated. For a distribution of particles, this result must be integrated over all the particles contributing to the radiation at frequency ω .

6.3 Bremsstrahlung

In the 1930s, Carl Anderson found that the ionisation loss rate given by the Bethe–Bloch formula (5.26) underestimates the energy loss rate for relativistic electrons. The additional energy loss mechanism was associated with the radiation of electromagnetic waves because of the acceleration of the electron in the electrostatic field of the nucleus. This radiation, first noted by Nikola Tesla in the 1880s in a different context, was called ‘braking radiation’ or, in German, *bremsstrahlung*. The process is identical to that known as *free–free emission* in the language of atomic physics, in the sense that the radiation corresponds to transitions between unbound states of the electron in the field of the nucleus. In 1934, computations of the spectrum of non-relativistic and relativistic bremsstrahlung were carried out by Bethe and Heitler (1934). More recently, detailed analyses appropriate for astrophysical applications have been presented by Koch and Motz (1959) and Blumenthal and Gould (1970).

We adopt here a classical approach, to which quantum mechanical parts are added as appropriate. The quantum mechanical treatment is beyond the scope of this book but is very important in deriving the photon distribution expected in the case of high energy interactions. We have already derived the expression for the acceleration of an electron in the electrostatic field of a high energy proton or nucleus (Sect. 5.3.1). Now the roles of the particles are interchanged – the electron moves at a high velocity past the stationary nucleus but, by symmetry, the field experienced by the electron in its rest frame is exactly the same as before. To work out the spectrum of the radiation emitted in such electrostatic encounters, we first take the Fourier transform of the acceleration of the electron and then use the expression (6.30) to determine the radiation spectrum. We then integrate this result over all collision parameters, just as in the case of ionisation losses, and use suitable limits for the collision parameters b_{\max} and b_{\min} . In the case in which the electron is moving relativistically, we transform the result back into the laboratory frame of reference.

Both the relativistic and non-relativistic calculations begin in the same way. The electrostatic accelerations of the electron in its rest frame parallel and perpendicular to its direction of motion, a_{\parallel} , and a_{\perp} , given by (5.18), are

$$\left. \begin{aligned} a_{\parallel} &= \dot{\mathbf{v}}_x = -\frac{eE_x}{m_e} = \frac{\gamma Ze^2 v t}{4\pi \epsilon_0 m_e [b^2 + (\gamma v t)^2]^{3/2}}, \\ a_{\perp} &= \dot{\mathbf{v}}_z = -\frac{eE_z}{m_e} = \frac{\gamma Ze^2 b}{4\pi \epsilon_0 m_e [b^2 + (\gamma v t)^2]^{3/2}}, \end{aligned} \right\} \quad (6.31)$$

where Ze is the charge of the nucleus.

We now take the Fourier transforms of the accelerations (6.31). On this occasion, we work out the calculation in some detail so that it can be seen how approximate methods

give similar results.

$$\dot{v}_x(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \frac{\gamma Ze^2 vt}{4\pi \epsilon_0 m_e [b^2 + (\gamma vt)^2]^{3/2}} \exp(i\omega t) dt , \quad (6.32a)$$

$$\dot{v}_z(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \frac{\gamma Ze^2 b}{4\pi \epsilon_0 m_e [b^2 + (\gamma vt)^2]^{3/2}} \exp(i\omega t) dt . \quad (6.32b)$$

Changing variables to $x = \gamma vt/b$,

$$\begin{aligned} \dot{v}_x(\omega) &= \frac{1}{(2\pi)^{1/2}} \frac{Ze^2}{4\pi \epsilon_0 m_e} \frac{1}{\gamma bv} \int_{-\infty}^{\infty} \frac{x}{(1+x^2)^{3/2}} \exp\left(i\frac{\omega b}{\gamma v} x\right) dx , \\ &= \frac{1}{(2\pi)^{1/2}} \frac{Ze^2}{4\pi \epsilon_0 m_e} \frac{1}{\gamma bv} I_1(y) , \end{aligned} \quad (6.33a)$$

$$\begin{aligned} \dot{v}_z(\omega) &= \frac{1}{(2\pi)^{1/2}} \frac{Ze^2}{4\pi \epsilon_0 m_e} \frac{1}{bv} \int_{-\infty}^{\infty} \frac{1}{(1+x^2)^{3/2}} \exp\left(i\frac{\omega b}{\gamma v} x\right) dx , \\ &= \frac{1}{(2\pi)^{1/2}} \frac{Ze^2}{4\pi \epsilon_0 m_e} \frac{1}{bv} I_2(y) , \end{aligned} \quad (6.33b)$$

where $y = \omega b / \gamma v$. The integrals $I_1(y)$ and $I_2(y)$ are

$$I_1(y) = 2iyK_0(y) \quad I_2(y) = 2yK_1(y) ,$$

where K_0 and K_1 are modified Bessel functions of order zero and one (Gradshteyn and Ryzhik, 1980; Abramovitz and Stegun, 1965). The radiation spectrum of the electron in an encounter with a charged nucleus with collision parameter b is therefore

$$\begin{aligned} I(\omega) &= \frac{e^2}{3\pi \epsilon_0 c^3} [|a_{\parallel}(\omega)|^2 + |a_{\perp}(\omega)|^2] , \\ &= \frac{e^2}{3\pi \epsilon_0 c^3} \frac{1}{2\pi} \left(\frac{Ze^2}{4\pi \epsilon_0 m_e bv} \right)^2 \left[\frac{1}{\gamma^2} I_1^2(y) + I_2^2(y) \right] , \\ &= \frac{Z^2 e^6}{24\pi^4 \epsilon_0^3 c^3 m_e^2 v^2} \frac{\omega^2}{\gamma^2 v^2} \left[\frac{1}{\gamma^2} K_0^2\left(\frac{\omega b}{\gamma v}\right) + K_1^2\left(\frac{\omega b}{\gamma v}\right) \right] . \end{aligned} \quad (6.34)$$

The radiation spectrum, displaying separately the terms arising from the accelerations parallel and perpendicular to the direction of motion of the electron, is shown in Fig. 6.2 (Jackson, 1999). The impulse perpendicular to the direction of travel contributes the greater intensity, even in the non-relativistic case, $\gamma = 1$. In addition, this component results in significant radiation at low frequencies. When the particle is relativistic, the intensity due to acceleration along the trajectory of the particle is decreased by a factor of γ^{-2} relative to the non-relativistic case. Thus, the dominant contribution to the radiation spectrum results from the momentum impulse perpendicular to the line of flight of the electron.

It is instructive to study the asymptotic limits of $K_0(y)$ and $K_1(y)$. These are:

$$y \ll 1 \quad K_0(y) = -\ln y; \quad K_1(y) = 1/y ,$$

$$y \gg 1 \quad K_0(y) = K_1(y) = (\pi/2y)^{1/2} \exp(-y) .$$

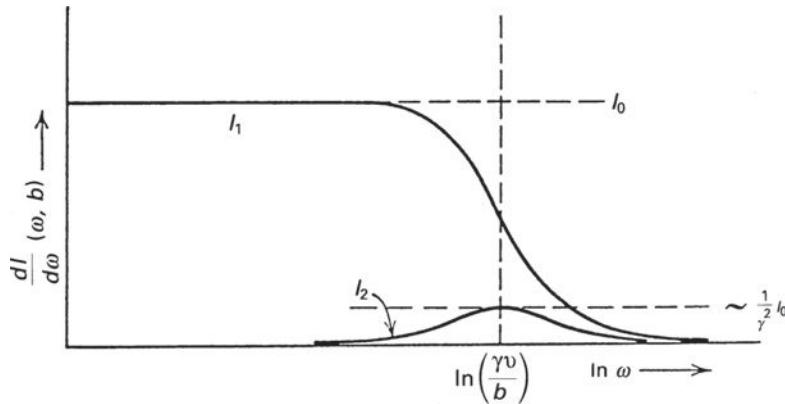


Fig. 6.2 The spectrum of bremsstrahlung resulting from the acceleration of the electron parallel and perpendicular to its initial direction of motion (Jackson, 1999).

At high frequencies, there is an exponential cut-off in the radiation spectrum

$$I(\omega) = \frac{Z^2 e^6}{48\pi^3 \epsilon_0^3 c^3 m_e^2 v^2} \frac{\omega}{\gamma v b} \left[\frac{1}{\gamma^2} + 1 \right] \exp\left(-\frac{2\omega b}{\gamma v}\right). \quad (6.35)$$

Note the origin of this cut-off. The duration of the relativistic collision is roughly $\tau = 2b/\gamma v$ (see Fig. 5.4). Thus, the dominant Fourier component of the radiation spectrum corresponds to frequencies $v \approx 1/\tau = \gamma v/2b$ and hence to $\omega \approx \pi v \gamma / b$, that is, to order of magnitude, $\omega b / \gamma v \approx 1$. The exponential cut-off means that there is little power emitted at frequencies greater than $\omega \approx \gamma v / b$.

The low frequency spectrum has the form

$$I(\omega) = \frac{Z^2 e^6}{24\pi^4 \epsilon_0^3 c^3 m_e^2 v^2 b^2} \left[1 + \frac{1}{\gamma^2} \left(\frac{\omega b}{\gamma v} \right)^2 \ln^2 \left(\frac{\omega b}{\gamma v} \right) \right]. \quad (6.36)$$

In the limit $\omega b / \gamma v \ll 1$, the second term in square brackets can be neglected and hence a good approximation for the low frequency intensity spectrum is

$$I(\omega) = \frac{Z^2 e^6}{24\pi^4 \epsilon_0^3 c^3 m_e^2 b^2 v^2} = K. \quad (6.37)$$

As noted above, the low frequency spectrum is almost entirely due to the momentum impulse perpendicular to the direction of travel of the electron. We could have guessed that the low frequency spectrum of the emission would be flat because, so far as these frequencies are concerned, the momentum impulse is a delta function, that is, the duration of the collision is very much less than the period of the waves. The Fourier transform of a delta function is a flat spectrum $I(\omega) = \text{constant}$. To a good approximation, the low frequency spectrum is flat up to frequency $\omega = \gamma v / b$ above which the spectrum falls off exponentially. Note also that, once again, the factor γ has disappeared from the intensity spectrum (6.37), even in the relativistic case. We recall that the momentum impulse is the same in the relativistic and non-relativistic cases as was demonstrated by the expression (5.20).

Finally, we integrate over all collision parameters which contribute to the radiation at frequency ω . So far, we have performed a completely general analysis in the rest frame of the electron. If the electron is moving relativistically, the number density of nuclei it observes is enhanced by a factor γ because of relativistic length contraction. Hence, in the moving frame of the electron, $N' = \gamma N$ where N is the space density of nuclei in the laboratory frame of reference. The number of encounters per second is $N'v$ and since all parameters are now measured in the rest frame of the electron, we add superscript dashes to all the relevant parameters. The radiation spectrum in the frame of the electron is therefore

$$I(\omega') = \int_{b'_{\min}}^{b'_{\max}} 2\pi b' \gamma N v K db' = \frac{Z^2 e^6 \gamma N}{12\pi^3 \epsilon_0^3 c^3 m_e^2} \frac{1}{v} \ln \left(\frac{b'_{\max}}{b'_{\min}} \right). \quad (6.38)$$

6.4 Non-relativistic bremsstrahlung energy loss rate

First of all, we evaluate the total energy loss rate by bremsstrahlung of a high energy but non-relativistic electron. We can therefore set $\gamma = 1$, drop the dashes on b_{\max} and b_{\min} and neglect relativistic correction factors. Then, the low frequency radiation spectrum (6.38) becomes

$$I(\omega) = \frac{Z^2 e^6 N}{12\pi^3 \epsilon_0^3 c^3 m_e^2} \frac{1}{v} \ln \Lambda, \quad (6.39)$$

where $\Lambda = (b_{\max}/b_{\min})$. Again, we have to make the correct choice of limiting collision parameters b_{\max} and b_{\min} . For b_{\max} , we integrate out to those values of b for which $\omega b/v = 1$. For larger values of b , the radiation at frequency ω lies on the exponential tail of the spectrum and makes a negligible contribution to the intensity (see Fig. 6.2). For b_{\min} , we have the same options described in Sect. 5.2.2 – at low velocities, $v \leq (Z/137)c$, we use the classical limit, $b_{\min} = Ze^2/8\pi\epsilon_0 m_e v^2$ (expression (5.10)). This would be appropriate for the bremsstrahlung of a region of ionised hydrogen at $T \approx 10^4$ K. At high velocities, $v \geq (Z/137)c$, the quantum restriction, $b_{\min} \approx \hbar/2m_e v$ (expression (5.11)), should be used and this is the appropriate limit to describe, for example, the X-ray bremsstrahlung of hot intergalactic gas in clusters of galaxies. Thus, the choices are

$$\Lambda = \frac{8\pi\epsilon_0 m_e v^3}{Ze^2 \omega} \quad \text{for low velocities ,} \quad (6.40a)$$

$$\Lambda = \frac{2m_e v^2}{\hbar\omega} \quad \text{for high velocities .} \quad (6.40b)$$

Notice that we have simplified the algebra by restricting the analysis to the flat, low frequency part of the radiation spectrum. There is, as usual, a cut-off at high frequencies corresponding to b_{\min} .

It is interesting to compare our result with the full answer derived by Bethe and Heitler who carried out a full quantum mechanical treatment of the radiation process (Bethe and Heitler, 1934; Carron, 2007). The electron cannot give up more than its total kinetic energy in the radiation process and so no photons are radiated with energies greater than

$\varepsilon = \hbar\omega = \frac{1}{2}m_e v^2$. In the same notation as above, the intensity of radiation of a single electron of energy $E = \frac{1}{2}m_e v^2$ in the non-relativistic limit is

$$I(\omega) = \frac{8}{3} Z^2 \alpha \hbar r_e^2 \frac{m_e c^2}{E} v N \ln \left[\frac{1 + (1 - \varepsilon/E)^{1/2}}{1 - (1 - \varepsilon/E)^{1/2}} \right], \quad (6.41)$$

where $\alpha = e^2/4\pi\hbar\varepsilon_0 c \approx 1/137$ is the fine structure constant and $r_e = e^2/4\pi\varepsilon_0 m_e c^2$ is the classical electron radius.³ The term in front of the logarithm is exactly the same as that in (6.39). In addition, in the limit of low energies $\varepsilon \ll E$, the term inside the logarithm reduces to $4E/\varepsilon$, exactly the same as (6.40b).

To find the total energy loss rate of a high energy particle, we integrate (6.39) over all frequencies. In practice, this means integrating from 0 to ω_{\max} where ω_{\max} corresponds to the cut-off, $b_{\min} \approx \hbar/2m_e v$. This is approximately

$$\omega_{\max} = \frac{2\pi}{\tau} \sim \frac{2\pi v}{b_{\min}} \approx \frac{4\pi m_e v^2}{\hbar}, \quad (6.42)$$

that is, to order of magnitude, $\hbar\omega \sim \frac{1}{2}m_e v^2$. This is the kinetic energy of the electron and is the maximum amount of energy which can be lost in a single encounter with the nucleus. We should therefore integrate (6.39) from $\omega = 0$ to $\omega_{\max} \approx m_e v^2/2\hbar$. Hence,

$$\begin{aligned} - \left(\frac{dE}{dt} \right)_{\text{brems}} &\approx \int_0^{\omega_{\max}} \frac{Z^2 e^6 N}{12\pi^3 \varepsilon_0^3 c^3 m_e^2} \frac{1}{v} \ln \Lambda d\omega \\ &\approx \frac{Z^2 e^6 N v}{24\pi^3 \varepsilon_0^3 c^3 m_e \hbar} \ln \Lambda \\ &= (\text{constant}) Z^2 N v. \end{aligned} \quad (6.43)$$

The total energy loss rate of the electron is proportional to v , that is, to the square root of the kinetic energy E : $-dE/dt \propto E^{1/2}$. This is in contrast to the case of relativistic bremsstrahlung losses discussed in Sect. 6.6 (see equation (6.69)). In practical applications of this formula, it is necessary to integrate over the energy distribution of the particles. For example, the energy spectrum of the electrons may well be of Maxwellian or of power-law form, $N(E) dE \propto E^{-x} dE$.

6.5 Thermal bremsstrahlung

6.5.1 Spectral emissivity of thermal bremsstrahlung

To work out the spectrum of bremsstrahlung of a thermal plasma at temperature T , the expressions for the spectrum of radiation of a single particle (6.39) should be integrated

³ Notice that (6.41) contains explicitly the constant \hbar because we have worked in terms of the energy radiated per unit angular frequency, while the Bethe–Bloch formula is normally quoted per unit energy interval. This \hbar cancels with that in the fine structure constant α to leave an expression for the intensity (6.39) which is independent of \hbar , as it must since it was derived by purely classical arguments.

over the collision parameters and then over a Maxwellian distribution of electron velocities

$$N_e(v) dv = 4\pi N_e \left(\frac{m_e}{2\pi kT} \right)^{3/2} v^2 \exp \left(-\frac{m_e v^2}{2kT} \right) dv. \quad (6.44)$$

The algebra becomes somewhat cumbersome at this stage. We can find the correct order-of-magnitude answer if we write $\frac{1}{2}m_e v^2 = \frac{3}{2}kT$ in (6.39). Then, an approximate expression for the spectral emissivity of a plasma of electron density N_e in the low frequency limit is

$$I(\omega) \approx \frac{Z^2 e^6 N N_e}{12\sqrt{3}\pi^3 \epsilon_0^3 c^3 m_e^2} \left(\frac{m_e}{kT} \right)^{1/2} g(\omega, T), \quad (6.45)$$

where $g(\omega, T)$ is known as a *Gaunt factor*. Note that the low frequency spectrum is more or less independent of frequency, the only dependence upon ω being the slowly varying function in the Gaunt factor. At high frequencies the spectrum of thermal bremsstrahlung cuts off exponentially as $\exp(-\hbar\omega/kT)$, reflecting the exponential decrease in the population of electrons in the high energy tail of a Maxwellian distribution. Finally, the total energy loss rate of the plasma may be found by integrating the spectral emissivity over all frequencies. Because of the exponential cut-off, the correct functional form is obtained by integrating (6.45) from 0 to $\omega = kT/\hbar$, that is,

$$-\frac{dE}{dt} = (\text{constant}) Z^2 T^{1/2} \bar{g} N N_e. \quad (6.46)$$

Detailed calculations give the following results, in terms of the frequency v rather than the angular frequency ω . The spectral emissivity of the plasma is

$$\begin{aligned} \kappa_v &= \frac{1}{3\pi^2} \left(\frac{\pi}{6} \right)^{1/2} \frac{Z^2 e^6}{\epsilon_0^3 c^3 m_e^2} \left(\frac{m_e}{kT} \right)^{1/2} g(v, T) N N_e \exp \left(-\frac{hv}{kT} \right) \\ &= 6.8 \times 10^{-51} Z^2 T^{-1/2} N N_e g(v, T) \exp(-hv/kT) \text{ W m}^{-3} \text{ Hz}^{-1}, \end{aligned} \quad (6.47)$$

where the number densities of electrons N_e and of nuclei N are in particles per cubic metre. At frequencies $hv \ll kT$, the Gaunt factor has only a logarithmic dependence on frequency. Suitable forms at radio and X-ray wavelengths are:

$$\text{Radio: } g(v, T) = \frac{\sqrt{3}}{2\pi} \left[\ln \left(\frac{128\epsilon_0^2 k^3 T^3}{m_e e^4 v^2 Z^2} \right) - \gamma^{1/2} \right], \quad (6.48a)$$

$$\text{X-ray: } g(v, T) = \frac{\sqrt{3}}{\pi} \ln \left(\frac{kT}{hv} \right), \quad (6.48b)$$

where $\gamma = 0.577\dots$ is Euler's constant. The functional forms of both logarithmic terms in (6.48a,b) can be readily derived from the corresponding expressions (6.40a,b). For frequencies $hv/kT \gg 1$, $g(v, T)$ is approximately $(hv/kT)^{1/2}$.

The total loss rate of the plasma is

$$-\left(\frac{dE}{dt} \right)_{\text{brems}} = 1.435 \times 10^{-40} Z^2 T^{1/2} \bar{g} N N_e \text{ W m}^{-3}. \quad (6.49)$$

Detailed calculations show that the frequency averaged value of the Gaunt factor \bar{g} lies in the range 1.1–1.5 and thus, to a good approximation, we can write $\bar{g} = 1.2$. The subject

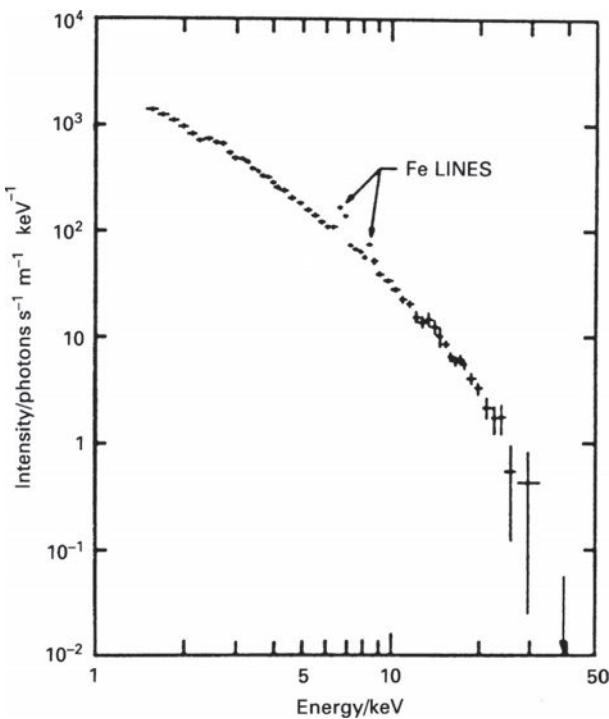


Fig. 6.3 The X-ray spectrum of the Perseus Cluster of galaxies observed by the HEAO-A2 instrument. The continuum emission can be accounted for by the thermal bremsstrahlung of hot intracluster gas at a temperature corresponding to $kT = 6.5 \text{ keV}$, that is, $T = 7.5 \times 10^7 \text{ K}$. The thermal nature of the radiation is confirmed by the observation of the Ly α and Ly β emission lines of highly ionised iron, Fe^{+25} , at energies of 6.7 and 7.9 keV, respectively. The ionisation potential of Fe^{+24} is 8.825 keV and hence the gas must be very hot. Note also the cluster of unresolved lines of highly ionised silicon, sulphur, calcium and argon in the energy range 1.8–4 keV (Mushotzky, 1980).

of suitable Gaunt factors for use in the thermal bremsstrahlung formulae is large and complex. A compilation of useful results is given by Karzas and Latter (1961) and a more recent survey for a wide range of astrophysical conditions by Sutherland (1998).

Figure 6.3 shows the spectrum of the intergalactic gas in the Perseus Cluster of galaxies as observed in the X-ray waveband by the HEAO-A2 experiment.⁴ The derived temperature of the emitting gas is $T = 7.5 \times 10^7 \text{ K}$. Confirmation of this high temperature is provided by the observation of lines of almost fully ionised iron, Fe xxvi, at 6.7 and 7.9 keV which are seen in Fig. 6.3. Since the gas is collisionally excited, the electron temperature of the hot gas must lie in the range $10^7\text{--}10^8 \text{ K}$. The interpretation of the diffuse X-ray emission from the cluster as the bremsstrahlung of hot gas enables the mass of intergalactic gas in the cluster to be estimated as well as providing an astrophysical tool for measuring the mass of the cluster as a whole (Sect. 4.4).

⁴ Note that it is common practice in X- and γ -ray astronomy to quote spectra in terms of the number of photons per unit energy interval rather than intensity and so a flat intensity spectrum, $I(\nu) d\nu \propto \nu^0 d\nu$, corresponds to a photon number intensity $N(\varepsilon) d\varepsilon \propto \varepsilon^{-1} d\varepsilon$, where $\varepsilon = h\nu$.

6.5.2 Thermal bremsstrahlung absorption

It is instructive to work out the coefficient for thermal bremsstrahlung absorption corresponding to the emissivity κ_ν . The resulting spectrum is the signature of compact regions of ionised hydrogen in the radio waveband. We begin with the general procedure for relating emission and absorption coefficients.

We first write down the transfer equation for radiation in terms of the intensity of radiation I_ν , that is, the radiant energy passing per second through unit area at normal incidence per steradian per unit bandwidth. In traversing dx , the decrease in intensity is $\chi_\nu I_\nu dx$ where χ_ν is the absorption coefficient. The increase in intensity in the same distance increment is $\kappa_\nu dx/4\pi$, where κ_ν is the emissivity of the plasma, meaning the power emitted per unit volume per unit bandwidth. Therefore, the transfer equation is

$$\frac{dI_\nu}{dx} = -\chi_\nu I_\nu + \frac{\kappa_\nu}{4\pi}. \quad (6.50)$$

In thermodynamic equilibrium at temperature T , dI_ν/dx is zero, the bremsstrahlung emission being exactly balanced by absorption by the same physical process, the *principle of detailed balance*. In thermodynamic equilibrium, the spectrum has black-body form and so

$$\chi_\nu I_\nu = \kappa_\nu / 4\pi, \quad (6.51)$$

where I_ν is the Planck spectrum of black-body radiation at temperature T ,

$$I_\nu(T) = \frac{2h\nu^3}{c^2} \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]^{-1} \quad \text{or} \quad I(\omega) = \frac{\hbar\omega^3}{\pi^2 c^2} \left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1 \right]^{-1}, \quad (6.52)$$

where $I(\omega)$ is the intensity integrated over 4π steradians per unit angular frequency ω . Substituting into (6.51),

$$\chi_\nu(T) = \frac{\kappa_\nu c^2}{8\pi h\nu^3} \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right].$$

The absorption coefficient for thermal bremsstrahlung is therefore

$$\chi_\nu = (\text{constant}) \frac{NN_e T^{-1/2}}{\nu^3} g(\nu, T) \left[1 - \exp\left(-\frac{h\nu}{kT}\right) \right]. \quad (6.53)$$

At high frequencies, $h\nu \gg kT$, the absorption coefficient has functional dependence

$$\chi_\nu \propto NN_e T^{-1/2} \nu^{-3} g(\nu, T). \quad (6.54)$$

At low frequencies $h\nu \ll kT$, expanding the exponential term for small values of $h\nu/kT$, we find $1 - \exp(-h\nu/kT) = h\nu/kT$ and hence

$$\chi_\nu \propto NN_e T^{-3/2} \nu^{-2} g(\nu, T). \quad (6.55)$$

Let us derive the same results in terms of the Einstein coefficients for spontaneous and stimulated emission and stimulated absorption. The definitions of these quantities for transitions between the upper energy level 2 and the lower energy level 1 are:

- A_{21} = transition probability per unit time for spontaneous emission,
- $B_{21}I(\omega)$ = transition probability for induced or stimulated emission per unit time,
- $B_{12}I(\omega)$ = transition probability for stimulated absorption per unit time,

where $I(\omega)$ is now the intensity of radiation integrated over 4π steradians per unit angular frequency and the angular frequency ω corresponds to the energy difference $\hbar\omega = E_2 - E_1$ between the upper and lower states. If N_2 and N_1 are the populations of the states 2 and 1, respectively, the condition for thermodynamic equilibrium is that the sum of the spontaneous and induced emission should balance the number of induced absorptions,

$$N_2 A_{21} + N_2 B_{21}I(\omega) = N_1 B_{12}I(\omega). \quad (6.56)$$

Solving for $I(\omega)$,

$$I(\omega) = \frac{A_{21}/B_{21}}{\frac{N_1}{N_2} \frac{B_{12}}{B_{21}} - 1}. \quad (6.57)$$

In thermodynamic equilibrium, N_1/N_2 is given by the Boltzmann relation

$$\frac{N_1}{N_2} = \frac{g_1}{g_2} \exp\left(\frac{\hbar\omega}{kT}\right),$$

where g_1 and g_2 are the statistical weights of levels 1 and 2. Therefore,

$$I(\omega) = \frac{A_{21}/B_{21}}{\frac{g_1}{g_2} \frac{B_{12}}{B_{21}} \exp\left(\frac{\hbar\omega}{kT}\right) - 1}. \quad (6.58)$$

This expression must correspond to the Planck function (6.52) written in terms of $I(\omega)$ and hence, comparing coefficients,

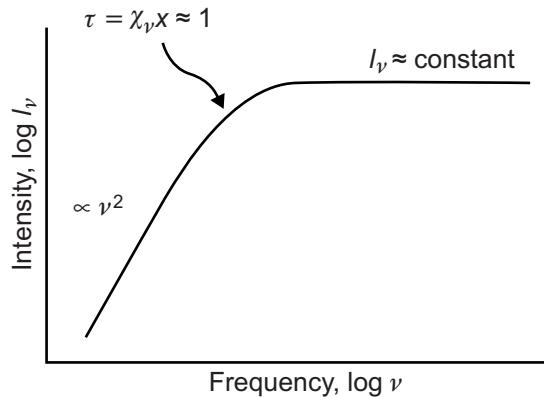
$$g_1 B_{12} = g_2 B_{21}; \quad A_{21} = \frac{\hbar\omega^3}{\pi^2 c^2} B_{21}. \quad (6.59)$$

This analysis, first given by Einstein in 1916, results in the relations between the elementary processes of emission and absorption. In terms of elementary atomic processes, the emissivity of the plasma is

$$\kappa(\omega) = \hbar\omega N_2 A_{21}. \quad (6.60)$$

In the transfer equation for radiation corresponding to (6.50), we include the terms for absorption and stimulated emission

$$\begin{aligned} \frac{dI(\omega)}{dx} &= \hbar\omega N_2 A_{21} - N_1 B_{12} \hbar\omega I(\omega) + N_2 B_{21} \hbar\omega I(\omega) \\ &= \kappa(\omega) - \hbar\omega I(\omega)(N_1 B_{12} - N_2 B_{21}). \end{aligned} \quad (6.61)$$

**Fig. 6.4**

The spectrum of thermal bremsstrahlung at low radio frequencies at which self-absorption becomes important. This is the characteristic spectrum of the compact regions of ionised hydrogen found in regions of star formation.

Thus,

$$\chi_\nu = \hbar\omega (N_1 B_{12} - N_2 B_{21}) = \hbar\omega N_1 B_{12} \left(1 - \frac{N_2}{N_1} \frac{B_{21}}{B_{12}}\right) = \hbar\omega N_1 B_{12} \left(1 - \frac{N_2 g_1}{N_1 g_2}\right). \quad (6.62)$$

If the matter is in thermal equilibrium, but not necessarily with the radiation,

$$N_2/N_1 = (g_2/g_1) \exp(-h\nu/kT),$$

and therefore

$$\chi_\nu = \hbar\omega N_1 B_{12} [1 - \exp(-\hbar\nu/kT)]. \quad (6.63)$$

This result is formally identical to (6.53). The last term in square brackets is derived from the stimulated emission term B_{21} . Thus, the absorption coefficient $\chi'_\nu = \hbar\omega N_1 B_{12}$ is referred to as the absorption coefficient for bremsstrahlung *uncorrected for stimulated emission*, whereas (6.58) is referred to as the absorption coefficient for bremsstrahlung *taking account of stimulated emission*.

Different forms of the absorption coefficients are encountered in different astronomical applications. Stellar astrophysicists normally use (6.54) which is directly related to the opacity of the stellar material for photon diffusion – these astronomers are interested in the opacity of the medium for photons having energies $\hbar\omega \approx kT$. On the other hand, radio astronomers always deal with very low energy photons, $\hbar\omega \ll kT$, and they use the formula (6.55).

Let us apply (6.55) to the spectrum of a compact region of ionised hydrogen as observed at radio wavelengths. The optical depth of the medium τ is defined to be

$$\tau = \int \chi_\nu dx = (\text{constant}) \int N_e^2 T^{-3/2} \nu^{-2} dx. \quad (6.64)$$

Integrating the transfer equation (6.50) along a column with uniform electron density,

$$\int_0^{I_0} \frac{dI_\nu}{(\kappa_\nu/4\pi - \chi_\nu I_\nu)} = \int_0^x dx .$$

Assuming there is no background radiation is present, $I_\nu = 0$ at $x = 0$. Integrating,

$$I_\nu = \frac{\kappa_\nu}{4\pi \chi_\nu} [1 - \exp(-\chi_\nu x)] . \quad (6.65)$$

This formula makes sense. If $\tau = \chi_\nu x \ll 1$,

$$I_\nu = \frac{\kappa_\nu}{4\pi \chi_\nu} (\chi_\nu x) = \frac{\kappa_\nu x}{4\pi} . \quad (6.66)$$

If $\tau = \chi_\nu x \gg 1$,

$$\begin{aligned} I_\nu &= \frac{\kappa_\nu}{4\pi \chi_\nu} = \frac{2hv^3}{c^2} \left[\exp\left(\frac{hv}{kT}\right) - 1 \right]^{-1}, \\ &= \frac{2kT}{c^2} v^2 \quad \text{if } hv \ll kT . \end{aligned} \quad (6.67)$$

Thus, the spectrum of the compact region of ionised hydrogen has a characteristic shape with $I_\nu = \text{constant}$ if $\tau \ll 1$ and $I_\nu \propto v^2$ if $\tau \gg 1$, corresponding to the Rayleigh–Jeans tail of a black-body distribution at temperature T . This form of spectrum is found in the compact H II regions close to regions of star formation. The temperature of the region may be estimated from the intensity of radiation in the Rayleigh–Jeans region of the spectrum and a mean, temperature-weighted value of N_e found from the point at which the region becomes optically thick.

6.6 Relativistic bremsstrahlung

We begin with (6.38) for the spectrum of relativistic bremsstrahlung in the frame of the moving electron. We need appropriate values for the collision parameters b'_{\max} and b'_{\min} . Since these collision parameters are linear dimensions perpendicular to the line of flight of the electron, they take the same values in S and S' .

At first sight it would seem that the value of b_{\min} should be the same as before, $b_{\min} = \hbar/\gamma m_e v$. We are now dealing, however, with the *radiation* of the accelerated electron and it should radiate *coherently*. If the electron has ‘size’ Δx and the duration of the impulse is shorter than the electron’s travel time across Δx , the different bits of the ‘probability distribution’ of the electron experience the momentum impulse at different times and so the radiation of the electron is not coherent. Therefore the duration of the impulse Δt must be at least as long as the travel time $\Delta x/v$ across the electron, that is, $\Delta t \geq \Delta x/v$. Therefore,

$$\frac{b}{\gamma v} \geq \frac{\hbar}{\gamma m_e v \cdot v} \quad \text{and hence} \quad b_{\min} = \frac{\hbar}{m_e v} . \quad (6.68)$$

There is now no Lorentz factor γ in the denominator of the minimum collision parameter. This result can also be understood from a different perspective. In the rest frame of the

electron, the value of b_{\min} in (6.68) corresponds to a collision time $\tau \sim 2b_{\min}/v$ and so to an angular frequency $\omega' \sim 2\pi/\tau \sim m_e v^2/\hbar$. This corresponds to a photon energy $\hbar\omega' \sim m_e v^2$. Transforming to the external frame, $\hbar\omega = \gamma m_e v^2$. This is exactly the condition that the electron gives up its total kinetic energy to the photon in a collision. Notice that exactly the same physical argument was applied to the derivation of the non-relativistic value of b_{\min} for ionisation losses (see Sect. 5.5.2).

An important case is that in which the relativistic electron interacts with neutral matter, in which case the electron is shielded from the nucleus by the electron clouds of atoms, unless the collision parameter is small. We can find a suitable estimate of b_{\max} by considering, for example, the Fermi–Thomas model of the atom (Leighton, 1959). The electrostatic field of the nucleus can be written approximately as

$$V(r) = \frac{Ze^2}{4\pi\epsilon_0 r} \exp\left(-\frac{r}{a}\right), \quad (6.69)$$

where

$$a = 1.4 a_0 Z^{-1/3} \quad \text{and} \quad a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} = 0.53 \times 10^{-10} \text{ m},$$

and a_0 is the Bohr radius of the hydrogen atom. Thus, for neutral atoms, a suitable value for b_{\max} is $b_{\max} = 1.4 a_0 Z^{-1/3}$.

In the ultra-relativistic limit, $\gamma \rightarrow \infty$, (6.38) therefore becomes

$$I(\omega') = \frac{Z^2 e^6 \gamma N}{12\pi^3 \epsilon_0^3 c^3 m_e^2 v} \ln\left(\frac{1.4 a_0 m_e v}{Z^{1/3} \hbar}\right). \quad (6.70)$$

We now transform this spectrum into the laboratory reference frame. We have already shown in Sect. 6.2.1 that dE/dt is a relativistic invariant. In the present case, $I(\omega')$ has the dimensions of energy per unit time per unit bandwidth. Thus, we need only ask how $\Delta\omega$ transforms between frames. It is simplest to note that ω transforms in the same way as E and hence, as shown in Sect. 6.2.1,

$$\Delta\omega = \gamma \Delta\omega',$$

that is, the bandwidth increases by a factor γ in S . Therefore in S , the intensity per unit bandwidth is smaller by a factor γ ,

$$I(\omega) = \frac{Z^2 e^6 N}{12\pi^3 \epsilon_0^3 c^3 m_e^2 v} \ln\left(\frac{192 v}{Z^{1/3} c}\right). \quad (6.71)$$

The intensity spectrum is independent of frequency up to energy $\hbar\omega = (\gamma - 1)m_e c^2$, which corresponds to the electron giving up all its kinetic energy in a single collision. The total energy loss rate is found by integrating over frequency,

$$-\frac{dE}{dt} = \int_0^{E/\hbar} I(\omega) d\omega. \quad (6.72)$$

Since $v \approx c$,

$$-\frac{dE}{dt} = \frac{Z^2 e^6 N E}{12\pi^3 \epsilon_0^3 c^4 \hbar} \ln\left(\frac{192}{Z^{1/3}}\right). \quad (6.73)$$

We can compare this with the formula derived by Bethe and Heitler from the full relativistic quantum treatment (Bethe and Heitler, 1934),

$$-\frac{dE}{dt} = \frac{Z(Z+1.3)e^6 N}{16\pi^3 \epsilon_0^3 m_e^2 c^4 \hbar} E \left[\ln \left(\frac{183}{Z^{1/3}} \right) + \frac{1}{8} \right]. \quad (6.74)$$

Thus, although we have had to make a number of approximations, we have come remarkably close to the correct answer. The term $(Z + 1.3)$ takes account of electron-electron interactions between the high energy electron and those bound to the atoms of the ambient material. Notice that, in contrast to the non-relativistic case (6.43), the relativistic bremsstrahlung energy loss rate is proportional to the energy of the electron. Many more details of appropriate bremsstrahlung formulae for different materials in different energy ranges are included in the chapter *Passage of particles through matter* in *The Review of Particle Properties* (Amsler *et al.*, 2008).

The cases of relativistic bremsstrahlung for a partially or fully ionised plasma have been treated by Koch and Motz (1959) and Blumenthal and Gould (1970). A useful compilation of results and references which can be applied to relativistic bremsstrahlung in diffuse astrophysical plasmas is provided by Strong and his colleagues in the appendix to their paper (Strong *et al.*, 2000).

The relativistic bremsstrahlung energy loss rate $-dE/dt$ is proportional to E , resulting in the exponential loss of energy by the electron. A radiation length X_0 can therefore be defined over which the electron loses a fraction $(1 - 1/e)$ of its energy, $-dE/dx = E/X_0$. As in Sect. 5.4, it is convenient to describe this length in terms of the number of kilograms per metre squared traversed by the electron, $\xi_0 = \rho X_0$. In the ultrarelativistic limit

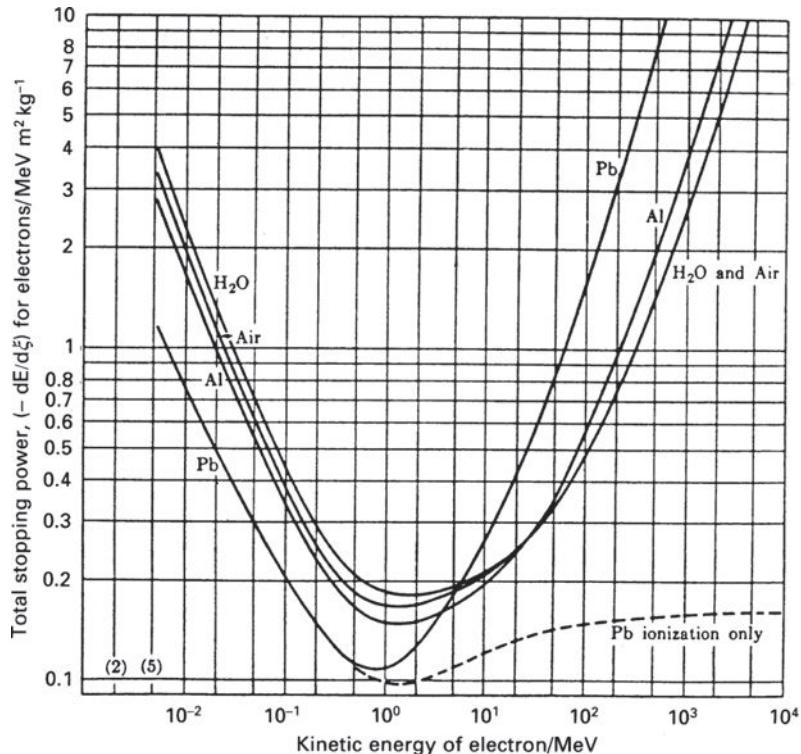
$$-\frac{dE}{d\xi} = -\frac{dE}{dt} \frac{1}{\rho c} = \frac{E}{\rho X_0} = \frac{E}{\xi_0}. \quad (6.75)$$

It is also convenient to express the radiation length ξ_0 in terms of the atomic mass M_A of the atoms of the material. If N_0 is Avogadro's number, $N = N_0 \rho / M_A$. We recall that M_A grams of any substance contain N_0 particles. According to the article *Passage of particles through matter* in *The Review of Particle Properties* (Amsler *et al.*, 2008), the following expression for ξ_0 provides an accurate fit to the data to a few percent:

$$\xi_0 = \frac{7164 M_A}{Z(Z+1) \ln(287/\sqrt{Z})} \text{ kg m}^{-2}. \quad (6.76)$$

The form of the total energy loss rate $-(dE/d\xi)$, or the *total stopping power*, for different materials is illustrated in Fig. 6.5. Below about 1 MeV, at which the electron becomes non-relativistic, ionisation losses remain the dominant loss mechanism, but for greater energies, relativistic bremsstrahlung losses rapidly become dominant. A *critical energy* E_c can be defined as that energy at which bremsstrahlung losses are equal to ionisation losses. For hydrogen, air and lead, the values of E_c are 340, 83 and 6.9 MeV, respectively. The radiation lengths for these materials are:

hydrogen	$\xi_0 = 580 \text{ kg m}^{-2}$	$X_0 = 6.7 \text{ km}$,
air	$\xi_0 = 365 \text{ kg m}^{-2}$	$X_0 = 280 \text{ m}$,
lead	$\xi_0 = 58 \text{ kg m}^{-2}$	$X_0 = 5.6 \text{ mm}$.

**Fig. 6.5**

The total stopping power for electrons in air, water, aluminium and lead. At energies less than 1 MeV, the dominant loss mechanism is ionisation losses. At higher energies, the dominant loss process is bremsstrahlung. For comparison, the contribution from ionisation losses for electrons in lead is also shown as a dashed line (Enge, 1966).

The value for air is of particular interest because the total depth of the atmosphere is about $10\,000 \text{ kg m}^{-2}$. Therefore, cosmic ray electrons must suffer catastrophic bremsstrahlung losses when they enter the atmosphere.

An important way of expressing the radiation spectrum is in terms of the photon number flux density. Let us rewrite the spectrum as a flux density of photons $N(\omega) d\omega$ in the energy interval $\hbar\omega$ to $\hbar(\omega + d\omega)$. Then

$$I(\omega) d\omega = N(\omega) \hbar\omega d\omega. \quad (6.77)$$

Therefore,

$$N(\omega) \propto 1/\omega \quad \text{up to energy} \quad \hbar\omega = (\gamma - 1)m_e c^2.$$

This means that the photon flux density diverges at zero frequency. As indicated in Fig. 6.2, however, the intensity of radiation remains finite at zero frequency. The important point is that, although the likelihood of an energetic photon being emitted is small, when it is emitted, it takes away with it a significant fraction of the energy of the electron. The spectrum of bremsstrahlung plotted in terms of the flux of photons per unit frequency interval is shown schematically on a linear frequency scale in Fig. 6.6 – this shows the

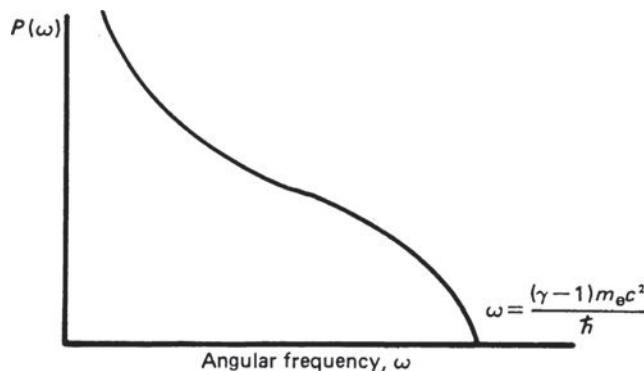


Fig. 6.6 The probability per unit bandwidth of the emission of a photon by bremsstrahlung as a function of angular frequency of the emitted photon plotted on linear intensity and frequency scales.

probability distribution of energy packets being emitted. On average, we expect one or two very energetic photons to be emitted in each radiation length. Thus, a very high energy cosmic ray electron deposits most of its energy into one or two high energy photons within a very short distance of entering the atmosphere.

Relativistic bremsstrahlung is likely to be of importance astrophysically. Wherever there are relativistic electrons with energy E , they can interact with atoms and molecules to generate photons with frequencies up to $\nu = E/h$, their average energy being about $(1/3)E$. In Sect. 17.3 it will be shown that a power-law electron energy spectrum of the form $N(E) \propto E^{-x}$ results in an intensity spectrum of γ -rays of exactly the same power-law form, $N_\gamma(\varepsilon) \propto \varepsilon^{-x}$, provided the intensity is measured in terms of the flux density of photons $\text{m}^{-2} \text{s}^{-1} \text{MeV}^{-1} \text{sr}^{-1}$. This process may well be important in understanding the low energy γ -ray emission of the interstellar medium (Strong *et al.*, 2000).

The dynamics of charged particles in magnetic fields

Magnetic fields are present everywhere in astrophysical environments (see Sect. 12.4) and so the dynamics of charged particles are strongly influenced by the Lorentz force, $\mathbf{F} = ze(\mathbf{v} \times \mathbf{B})$, which they inevitably experience. This has many consequences for high energy astrophysics. Charged particles move in spiral paths about magnetic field lines, tying them to the magnetic field distribution. Any net streaming motion of the charged particles along magnetic field lines is, however, limited by plasma instabilities and by scattering in pitch angle by small-scale irregularities in the magnetic field. As a result, charged particles can cross field lines. Relativistic electrons radiate cyclotron and synchrotron radiation because of their spiral motion and these emissions provide tracers of the distribution of high energy particles and magnetic fields in galaxies. These topics are crucial in the study of the dynamics of high energy particles in magnetic fields, which are major themes in Parts III and IV of this study.

7.1 A uniform static magnetic field

We begin with the simplest case of the motion of a particle of rest mass m_0 , charge ze and velocity \mathbf{v} , corresponding to a Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$, in a uniform, static magnetic field \mathbf{B} . The equation of motion is

$$\frac{d}{dt}(\gamma m_0 \mathbf{v}) = ze(\mathbf{v} \times \mathbf{B}). \quad (7.1)$$

The left-hand side of this equation can be expanded as follows:

$$m_0 \frac{d}{dt}(\gamma \mathbf{v}) = m_0 \gamma \frac{d\mathbf{v}}{dt} + m_0 \gamma^3 \mathbf{v} \frac{(\mathbf{v} \cdot \mathbf{a})}{c^2},$$

because the Lorentz factor γ should be written more properly as $\gamma = (1 - \mathbf{v} \cdot \mathbf{v}/c^2)^{-1/2}$. In a magnetic field, the three-acceleration $\mathbf{a} = d\mathbf{v}/dt$ is always perpendicular to \mathbf{v} and consequently $\mathbf{v} \cdot \mathbf{a} = 0$. As a result,

$$\gamma m_0 \frac{d\mathbf{v}}{dt} = ze(\mathbf{v} \times \mathbf{B}). \quad (7.2)$$

Now split \mathbf{v} into components parallel and perpendicular to the uniform magnetic field, v_{\parallel} and v_{\perp} , respectively (Fig. 7.1). The *pitch angle* θ of the particle's orbit is shown in Fig. 7.1 and is defined by $\tan \theta = v_{\parallel}/v_{\perp}$, that is, the angle between the vectors \mathbf{v} and \mathbf{B} . Since v_{\parallel} is parallel to \mathbf{B} , (7.2) shows that there is no change in v_{\parallel} , $v_{\parallel} = \text{constant}$. The acceleration of

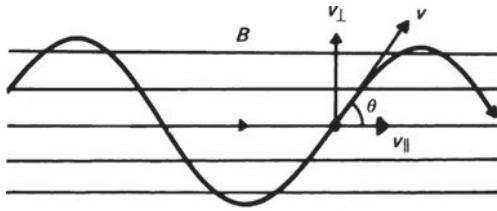


Fig. 7.1

Illustrating the dynamics of a charged particle in a uniform magnetic field.

the charged particle perpendicular to the magnetic field direction \mathbf{B} and to v_{\perp} is

$$\gamma m_0 \frac{dv}{dt} = ze v_{\perp} |\mathbf{B}| (\mathbf{i}_v \times \mathbf{i}_B) = ze |\mathbf{v}| |\mathbf{B}| \sin \theta (\mathbf{i}_v \times \mathbf{i}_B),$$

where \mathbf{i}_v and \mathbf{i}_B are unit vectors in the directions of \mathbf{v} and \mathbf{B} , respectively.

Thus, the particle's acceleration vector is perpendicular to the plane containing both the instantaneous velocity vector \mathbf{v} and the direction of the magnetic field \mathbf{B} . Because the magnetic field is uniform, this constant acceleration perpendicular to the instantaneous velocity vector results in circular motion about the magnetic field direction. Equating this acceleration to the centripetal acceleration,

$$\frac{v_{\perp}^2}{r} = \frac{ze|\mathbf{v}||\mathbf{B}|\sin\theta}{\gamma m_0}, \quad \text{that is,} \quad r = \frac{\gamma m_0 |\mathbf{v}| \sin \theta}{ze |\mathbf{B}|}. \quad (7.3)$$

Thus, the motion of the particle consists of a constant velocity along the magnetic field direction and circular motion with radius r about it, that is, a *spiral path* with *constant pitch angle* θ . The radius r is known as the *gyroradius* or *cyclotron radius* of the particle. Its angular frequency ω_g about the magnetic field direction is known as the *angular cyclotron frequency* or *angular gyrofrequency*,

$$\omega_g = \frac{v_{\perp}}{r} = \frac{ze|\mathbf{B}|}{\gamma m_0}. \quad (7.4)$$

The corresponding *gyrofrequency* ν_g , that is, the number of times per second that the particle gyrates about the magnetic field direction, is

$$\nu_g = \frac{\omega_g}{2\pi} = \frac{ze|\mathbf{B}|}{2\pi\gamma m_0}. \quad (7.5)$$

For a non-relativistic particle, $\gamma = 1$ and hence $\nu_g = ze|\mathbf{B}|/2\pi m_0$. A useful figure to remember is the non-relativistic gyrofrequency of an electron, $\nu_g = e|\mathbf{B}|/2\pi m_e = 28 \text{ GHz T}^{-1}$, where the magnetic field strength is measured in tesla, T, or $\nu_g = 2.8 \text{ MHz G}^{-1}$ if the magnetic flux density is measured in gauss, G.

In this simple case, the axis of the particle's trajectory is parallel to the magnetic field direction and is known as the *guiding centre* of the particle's motion, that is, it is the mean direction of translation of the particle about which the gyration takes place. In more complicated magnetic field configurations, it is convenient to work in terms of the *guiding centre motion* of the charged particle and this determines the general drift of particles in the field. Examples of this are discussed in the next section.

Table 7.1 The properties of protons, carbon and iron nuclei having Lorentz factors $\gamma = 2$ and 100.

	Proton	Carbon nucleus		Iron nucleus	
Lorentz factor, γ	2	100	2	100	2
Velocity, v	$(\sqrt{3}/2)c$	$0.99995c$	$(\sqrt{3}/2)c$	$0.99995c$	$(\sqrt{3}/2)c$
Mass number, A	1	1	12	12	56
Atomic number, z	1	1	6	6	26
Rest mass energy, mc^2	1 GeV	1 GeV	12 GeV	12 GeV	56 GeV
Total energy, γmc^2	2 GeV	100 GeV	24 GeV	1200 GeV	112 GeV
Kinetic energy, $(\gamma - 1)mc^2$	1 GeV	99 GeV	12 GeV	1188 GeV	56 GeV
Kinetic energy per nucleon	1 GeV	99 GeV	1 GeV	99 GeV	1 GeV
Momentum, $pc = (\gamma m v)c^\dagger$	$\sqrt{3}$ GeV	99.995 GeV	20.8 GeV	1199.9 GeV	96.99 GeV
Rigidity, pc/ze	$\sqrt{3}$ GV	99.995 GV	$2\sqrt{3}$ GV	199.99 GV	3.73 GV
					215.4 GV

† To obtain the dimensions of GeV, the momentum has been multiplied by c , the velocity of light.

Let us rewrite the expression for the radius of the particle's path in the following form

$$r = \frac{\gamma m_0 v \sin \theta}{ze} = \left(\frac{pc}{ze} \right) \frac{\sin \theta}{|\mathbf{B}|c}, \quad (7.6)$$

where $p = \gamma m_0 |v|$ is the relativistic three-momentum of the particle. Thus, if we inject particles with the same value of pc/ze into a magnetic field \mathbf{B} at the same pitch angle θ , they have exactly the same dynamical behaviour. By extension, this result remains true for any magnetic field configuration. The quantity pc/ze is called the *rigidity* or *magnetic rigidity* of the particle. Since pc has the dimensions of energy and e the dimensions of charge, pc/ze has the dimensions of volts – a useful unit for high energy particles is gigavolts (GV). In cosmic rays studies, the energies of cosmic rays are often quoted in terms of their rigidities rather than their energies per nucleon. It is useful to compare the energies, momenta and rigidities of protons, carbon and iron nuclei with Lorentz factors $\gamma = 2$ and 100, as shown in Table 7.1.

7.2 A time-varying magnetic field

In the magnetic field configuration shown in Fig. 7.1, the charged particle moves in a spiral path with constant radius and pitch angle. In reality, the magnetic field distribution can change with time and with spatial position. We consider the case in which the magnetic flux density \mathbf{B} varies slowly with time, by which we mean that the fractional change in the magnetic field strength $\Delta B/B$ changes very little in a single orbital period $T = v_g^{-1}$. Let us first consider the non-relativistic version of the problem of the motion of a charged particles in a varying magnetic field adopting an approach which highlights the essential physics.

7.2.1 Physical approach to the non-relativistic case

A charged particle gyrating about its guiding centre in a magnetic field is equivalent to a current loop. The equivalent current is the rate at which charge passes a particular point in the loop per second, $i = zev_{\perp}/2\pi r$. The area of the loop is $A = \pi r^2$ and so the magnetic moment μ of the current loop is

$$\mu = iA = \frac{zev \sin \theta}{2\pi r} \pi r^2 = \frac{zev_{\perp}}{2} r .$$

In the non-relativistic limit, $r = m_0 v_{\perp} / zeB$, and therefore

$$\mu = \frac{m_0 v_{\perp}^2}{2B} = \frac{w_{\perp}}{B} , \quad (7.7)$$

where w_{\perp} is the kinetic energy of the particle in the direction perpendicular to the guiding centre.

Now suppose there is a small change ΔB in the magnetic flux density B in one orbit. Then, an electromotive force \mathcal{E} is induced in the loop because of the changing magnetic field and the particle in its orbit is accelerated. The work done on the charged particle per orbit by the electromotive force is

$$ze\mathcal{E} = ze\pi r^2 \frac{dB}{dt} = ze\pi r^2 \frac{\Delta B}{\Delta T} ,$$

where $\Delta T = 2\pi r / v_{\perp}$ is the period of one orbit. Therefore, the change in kinetic energy of the particle in one orbit is

$$\Delta w_{\perp} = \frac{zer v_{\perp}}{2} \Delta B = \frac{m_0 v_{\perp}^2}{2B} \Delta B = \frac{w_{\perp}}{B} \Delta B .$$

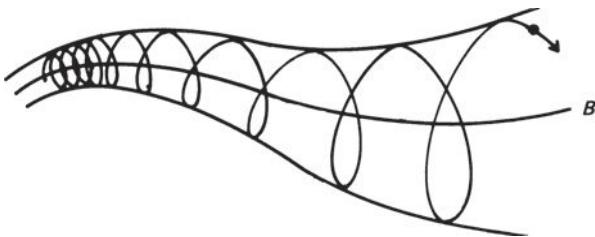
The corresponding change in the magnetic moment of the current loop is

$$\Delta\mu = \Delta \left(\frac{w_{\perp}}{B} \right) = \frac{\Delta w_{\perp}}{B} - \frac{w_{\perp} \Delta B}{B^2} = \frac{\Delta w_{\perp}}{B} - \frac{\Delta w_{\perp}}{B} = 0 , \quad (7.8)$$

that is, the magnetic moment of the particle is an *invariant* provided the field is slowly varying. There are other ways of expressing this important result. As illustrated by (7.8), $\Delta\mu = 0$ is equivalent to $\Delta(w_{\perp}/B) = 0$. Since $w_{\perp} = p_{\perp}^2/2m_0$, this is the same as

$$\Delta(p_{\perp}^2/B) = 0 . \quad (7.9)$$

This result accounts for the phenomenon of *magnetic mirroring*. If the particle moves into a region of converging magnetic field lines, the magnetic flux density B increases and therefore the perpendicular kinetic energy of the particle w_{\perp} must also increase. However, the kinetic energy of the particle is constant because no work is done by a static magnetic field and therefore the increase in p_{\perp}^2 must take place at the expense of the parallel component of the particle's motion w_{\parallel} . Now w_{\parallel} goes to zero at the point at which $w_{\perp} = w$ and so the particle is reflected back along the magnetic field configuration (Fig. 7.2). This phenomenon accounts for the trapping of charged particles in the Earth's radiation belts since they are reflected in the converging field lines as they approach the Earth's magnetic poles.

**Fig. 7.2**

The dynamics of a charged particle in a slowly varying magnetic field illustrating how the particle's guiding centre follows the mean magnetic field direction. The radius of curvature of the particle's path is such that a constant magnetic flux is enclosed by its orbit.

Since $r = m_0 v_{\perp} / zeB$ and $p_{\perp}^2 = (zeB)^2$, $\Delta(p_{\perp}^2/B) = 0$ also implies

$$\Delta(Br^2) = 0. \quad (7.10)$$

Thus, the particle follows the guiding centre in such a way that the number of field lines within the particle's orbit is a constant, as illustrated in Fig. 7.2.

The expressions (7.9) and (7.10) are referred to as the *first adiabatic invariant* of the particle's motion in a magnetic field and can be derived from the *principle of adiabatic invariance*. This is the best way of deriving the relativistic generalisations of these formulae which are:

$$\left. \begin{aligned} \Delta(Br^2) &= 0 & r &= \gamma m_0 v_{\perp} / zeB, \\ \Delta(p_{\perp}^2/B) &= 0 & p_{\perp} &= \gamma m_0 v_{\perp}, \\ \Delta(\gamma\mu) &= 0 & \mu &= \gamma m_0 v_{\perp}^2 / 2B. \end{aligned} \right\} \quad (7.11)$$

7.2.2 Adiabatic invariant approach

According to the Lagrangian formulation of classical dynamics, if q_i and p_i are the canonical coordinates and momenta, for each coordinate that is periodic, the action integral $J = \oint p_i dq_i$ is a constant for a given mechanical system with specified initial conditions (Jackson, 1999). If the properties of the system change slowly compared with the period of oscillation, the action integral J is an invariant. Such a change is called an *adiabatic change* – this is exactly what is needed to investigate the dynamics of a charged particle moving in a slowly varying magnetic field.

The components of velocity and position perpendicular to the magnetic field direction are both periodic. The action integral is therefore

$$J = \oint \mathbf{P}_{\perp} \cdot d\mathbf{l}, \quad (7.12)$$

where \mathbf{P}_{\perp} is the canonical momentum of the particle perpendicular to the magnetic field direction and $d\mathbf{l}$ is the line element along the circular path of the particle. For a charged particle in a magnetic field, the canonical momentum perpendicular to the field is

$$\mathbf{P}_{\perp} = \mathbf{p}_{\perp} + e\mathbf{A},$$

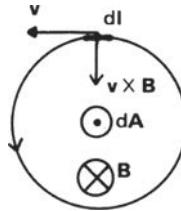


Fig. 7.3

Illustrating how to find the sign of the increment of magnetic flux in evaluating the action integral J .

where \mathbf{p}_\perp is the relativistic three-momentum of the particle perpendicular to \mathbf{B} , and \mathbf{A} is the vector potential of the magnetic field, $\mathbf{B} = \nabla \times \mathbf{A}$. Therefore,

$$\begin{aligned} J &= \oint_C \mathbf{P}_\perp \cdot d\mathbf{l} = \oint_C \mathbf{p}_\perp \cdot d\mathbf{l} + e \int_C \mathbf{A} \cdot d\mathbf{l} , \\ &= \oint_C \gamma m_0 \mathbf{v}_\perp \cdot d\mathbf{l} + e \int_S \mathbf{B} \cdot d\mathbf{S} , \\ &= 2\pi r \gamma m_0 v_\perp + e \int_S \mathbf{B} \cdot d\mathbf{S} , \end{aligned} \quad (7.13)$$

where $d\mathbf{S}$ is the element of area contained within the contour C associated with the line integral $\oint d\mathbf{l}$.

Let us study the vector relations between $d\mathbf{l}$, $d\mathbf{B}$ and $d\mathbf{S}$. If $d\mathbf{B}$ is directed into the paper and \mathbf{v} has the direction shown in Fig. 7.3, the Lorentz force $(\mathbf{v} \times \mathbf{B})$ for a positively charged particle results in circular motion as shown. The vector area $d\mathbf{S}$ consequently points out of the paper, that is, in the opposite direction to \mathbf{B} . Thus, the second term in equation (7.13) is negative. Therefore, since $\omega = v_\perp/r$,

$$J = 2\pi r^2 \gamma m_0 \omega - e\pi r^2 B .$$

But the angular gyrofrequency is $\omega = eB/\gamma m_0$ and hence

$$J = e\pi r^2 B = eAB ,$$

where A is the area swept out by the particle. According to the above rule for adiabatic invariants, J is a constant for slowly varying changes in \mathbf{B} , that is,

$$\Delta(\pi r^2 B) = 0 . \quad (7.14)$$

This is the same result quoted in equation (7.11) and the other invariants follow immediately from the relations $r = \gamma m_0 v_\perp/eB$, $p_\perp = \gamma m_0 v_\perp$ and $\mu = \gamma m_0 v_\perp^2/2B$.

We could go on and work out the behaviour in more complicated cases – what happens when the particles are in regions where there is a magnetic field gradient, what is the effect of a gravitational field, and so on? However, the point will now be clear that individual particles are tied to magnetic field lines and it takes a great deal to make them move across. Northrop's monograph *Adiabatic Motion of Charged Particles* provides an introduction to these more advanced topics (Northrop, 1963).

To anticipate the considerations of Sect. 10.5, these results for individual charged particles are closely related to those involved in magnetic flux freezing. These are, however,

separate problems, although the treatments I give make them look rather similar. The flux freezing argument is a magnetohydrodynamic process in which we treat the plasma as a perfectly conducting fluid. The treatment using individual particles is a microscopic approach and, to make the two approaches equivalent, it has to be shown that the equations of magnetohydrodynamics can be derived from the microscopic equations of motion. This is far from trivial (Clemmow and Dougherty, 1969).

7.3 The scattering of charged particles by irregularities in the magnetic field

According to the analysis of the last section, charged particles move in such a way that they enclose the same field lines, so long as the field is slowly varying. There are, however, bound to be irregularities in the magnetic field and these have the effect of scattering the particles in pitch angle. If these scatterings are random, the result is a uniform distribution of pitch angles. This is an assumption we will make on a number of occasions and there are good physical reasons for it.

A good example of irregularities in a large scale magnetic field is the case of the magnetic fluctuations in the interplanetary magnetic field. Direct measurements of these were made by the Mariner 4 space probe which went on to take the first pictures of the Martian surface. The magnetic flux density was measured continuously throughout the flight from the Earth to Mars. The magnitude of the magnetic irregularities as a function of physical scale was described by applying Parseval's theorem (Sect. 6.2.5) to find the power spectrum of the fluctuations in the magnetic field:

$$\int_{-\infty}^{\infty} B^2(t) dt = \int_{-\infty}^{\infty} B^2(\omega) d\omega , \quad (7.15)$$

where $B(\omega)$ is the Fourier transform of the measured magnetic flux density with time, $B(t)$. The power spectrum $B^2(v)$ shown in Fig. 7.4, measured as the noise power per unit frequency interval, shows that most of the power is in fluctuations on the scale of about 10^9 m.

If the particles have gyroradii much smaller than the scale of the fluctuations in the magnetic field, the trajectories of the particles follow their guiding centres and changes in pitch angle result from conserving their adiabatic invariants (Sect. 7.2). In the opposite limit in which the particles have gyroradii much greater than the scale of the fluctuations, the particles do not 'feel' the fine structure in the field but move in orbits determined by the mean magnetic field which is much greater in magnitude than the fluctuating component (Fig. 7.5a). Thus, it is only in the case in which the fluctuations have the same scale as the gyroradii of the particles that there is significant scattering. Figure 7.5b illustrates how a significant change in the pitch angle of the particle can occur in a single gyroradius. The scattering of the particles by the random superposition of these fluctuations leads to stochastic changes in the pitch angles of the particles.

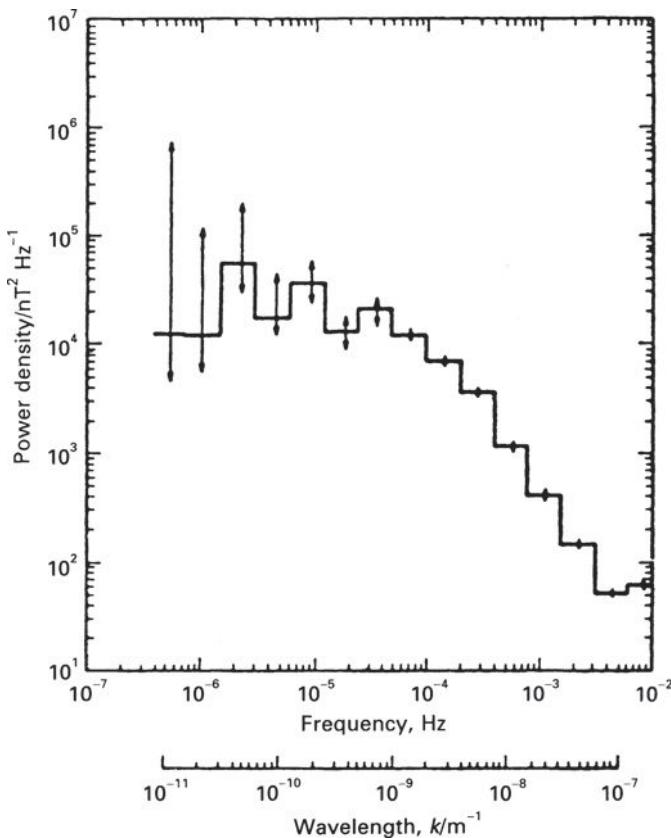


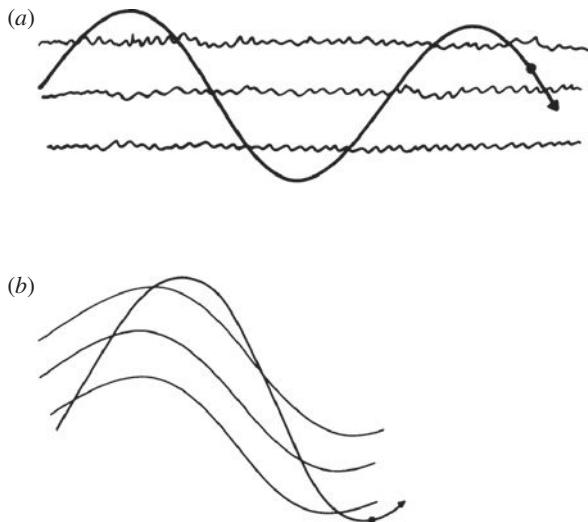
Fig. 7.4

The power spectrum of the magnetic field energy density per unit frequency interval as measured by the magnetometers on board the Mariner 4 spacecraft. The strength of the magnetic field is measured in nanoteslas, $1 \text{ nT} = 10^{-9}$ (Jokipii, 1973).

Let us work out the magnetic rigidity R at which we would expect the magnetic fluctuations to be important in scattering high energy particles in the case of the interplanetary medium. We recall that we showed in Sect. 7.2 that particles of different charges and masses but the same magnetic rigidities have the same dynamics in any magnetic field distribution. The gyroradius of the particle in terms of its magnetic rigidity is

$$r_g = \left(\frac{pc}{ze} \right) \frac{1}{Bc} = \frac{R}{Bc}, \quad (7.16)$$

where we have assumed that the pitch angle is $\theta = \pi/2$. We equate this gyroradius to the wavelength at which there is most power in the power spectrum of magnetic irregularities. Taking $r = \lambda_c = 2 \times 10^9 \text{ m}$ and a mean interplanetary magnetic flux density $B = 3 \text{ nT}$, $R = 2 \text{ GV}$. This rigidity is remarkably similar to that at which the spectra of cosmic ray protons and nuclei become strongly influenced by *solar modulation*, that is, modifications of the spectrum of the cosmic rays because of the influence of the outflowing Solar Wind seen in Fig. 1.16. The figures in Table 7.1 show that cosmic rays with kinetic energies about

**Fig. 7.5**

Illustrating the dynamics of a charged particle in a magnetic field, (a) when the irregularities in the magnetic field are on a scale much smaller than the gyroradius of the particle's orbit; (b) when they are of the same order of magnitude.

1 GeV per nucleon all have magnetic rigidities of a few GV. The evaluation of diffusion coefficients for high energy particles given the spectrum of magnetic irregularities has been carried out by Jokipii (1973). Let us carry out order-of-magnitude calculations to work out the diffusion coefficient for the particles subject to random pitch angle scattering.

The important assumption is that the magnetic field irregularities are random. The power spectrum of the magnetic field strength describes how much energy there is in each Fourier component of the field and it is implicit in this procedure that the phases of the waves are assumed to be random. What this means physically is that the particle ‘feels’ the influence of a particular field component for about one wavelength before it encounters another wave with random phase relative to the last wave. The model of the diffusion process is therefore that the particle experiences any given wave for about one wavelength before it is scattered by another wave of random phase.

In a single wavelength, the average inclination of the field lines from the mean field direction due to magnetic irregularities is $\phi \approx B_1/B_0$, where B_0 is the mean magnetic flux density and B_1 is the amplitude of the random component. Therefore, the pitch angles of particles with gyroradii $r_g \approx \lambda$ change by about this amount per wavelength. The guiding centre is therefore displaced by a distance $r \approx \phi r_g$ and this represents diffusion of the particles across the magnetic field lines as well as a change in their pitch angles.

In the next wavelength, the particle meets another wave of roughly the same energy density but the change in pitch angle is now random with respect to the previous wave and so the particle is scattered randomly in pitch angle. Therefore, to be scattered randomly through 1 radian, the particle has to be scattered N times, where $N^{1/2}\phi = 1$. The distance for scattering through 1 radian is thus $\lambda_{sc} \approx N\lambda \approx Nr_g \approx r_g\phi^{-2}$. This is the effective mean free path for pitch angle scattering of a particle diffusing along the magnetic field. In this

distance, the pitch angle of the particle has been changed by a large factor – the particle loses all memory of its initial pitch angle in this distance.

We now combine this result with the spectrum of irregularities in the interplanetary magnetic field to work out the mean free path as a function of magnetic rigidity R . Since $\lambda_{\text{sc}} \approx r_g \phi^{-2} \approx r_g (B_1/B_0)^2$, we need the energy density in the fluctuating component of the magnetic field on the scale λ . The power spectrum is given per unit frequency and so the energy density in the fluctuating magnetic field on scale $\lambda = v/V$ is $B_1^2(v)v/2\mu_0$. Therefore,

$$\lambda_{\text{sc}} \approx \frac{r_g}{\phi^2} = r_g \frac{B_0^2}{B_1^2(v)v}. \quad (7.17)$$

Our results are similar to those obtained in Jokipii's detailed calculations, including, to order of magnitude, the values of the numerical constants (Jokipii, 1973). These concepts have been applied successfully to the scattering of high energy particles in the Solar Wind and the modulation of the spectra of cosmic ray protons and nuclei. Similar considerations can be applied to the diffusion of particles in the interstellar medium, although information about the spectrum of fluctuations on the relevant scales is not available.

7.4 The scattering of high energy particles by Alfvén and hydromagnetic waves

Suppose a uniform magnetic field is embedded in a partially ionised plasma and a flux of high energy particles propagates along the magnetic field direction at a high streaming velocity. What is the interaction between the flux of high energy particles and the magnetoactive plasma?

The results of these investigations are as follows. If the plasma is fully ionised, the high energy particles resonate with irregularities in the magnetic field and are scattered in pitch angle, exactly as described in Sect. 7.3. In addition, magnetic fluctuations are generated by Alfvén and hydromagnetic waves which grow in amplitude under the influence of the streaming motions so that, even if there were no magnetic irregularities to begin with, they are generated by the streaming of the high energy particles. The full theory of the growth of the waves is non-trivial and we make no attempt to do it justice here. Wentzel and Cesarsky provide excellent reviews of these aspects of plasma physics (Wentzel, 1974; Cesarsky, 1980). We can understand the underlying physics using arguments similar to those developed in Sect. 7.3.

If the perturbation in the magnetic flux density is B_1 , pitch angle scattering results in changing the pitch angle of the particles by about 90° after a mean free path $\lambda_{\text{sc}} \approx r_g/\phi^2$, where $\phi = B_1/B_0$; the corresponding diffusion coefficient is $D = (1/3) v \lambda_{\text{sc}}$. This mechanism converts streaming motion into a random distribution of pitch angles over a distance λ_{sc} . Complications arise for two reasons. First, the waves with which the particles resonate are Alfvén and hydromagnetic waves, which are the characteristic low-frequency ‘sound’ waves found in a magnetised plasma. The circularly polarised hydromagnetic waves

are particularly important because they can resonate with the spiral motion of the charged particles. Second, the strength of the perturbed component B_1 is due to the streaming of the particles themselves. Physically, the forward momentum of the beam is transferred to the waves which must grow as a result.

The growth rate Γ of the instability can be derived from the simple physical picture described above, the equation for the exponential increase in the energy density U of the Alfvén waves being $U = U_0 \exp \Gamma t$. For simplicity, we consider the high energy particles to be protons. First, we convert the expression for the mean free path of a high energy proton into a time-scale τ_s for scattering through 90° ,

$$\tau_s = \frac{\lambda_{SC}}{v} = \left(\frac{r_g}{v} \right) \left(\frac{B_0}{B_1} \right)^2. \quad (7.18)$$

The energy density in Alfvén or hydromagnetic waves is the energy density in the perturbing magnetic field B_1 , $U_A = B_1^2/2\mu_0$ and the Alfvén speed is $v_A = B_0/(\mu_0\rho)^{1/2}$, where $\rho = N_p m_p$ is the mass density of the fully ionised plasma. Making these substitutions,

$$\tau_s \approx \left(\frac{r_g}{v} \right) \left(\frac{v_A^2 N_p m_p}{U_A} \right). \quad (7.19)$$

To find the rate of momentum transfer to the waves, it is simplest to work in terms of their *momentum density*. For all types of wave motion, the momentum density is $P_{\text{wave}} = U_{\text{wave}}/v$, where P_{wave} and U_{wave} are the energy and momentum densities, respectively, and v is the speed of the waves. In the present case, the speed of the waves is the Alfvén speed and so

$$\frac{dP_{\text{wave}}}{dt} = \frac{d}{dt} \left(\frac{U_{\text{wave}}}{v_A} \right). \quad (7.20)$$

This is equal to the rate at which momentum is lost from the streaming relativistic particles. The momentum supplied to unit volume over the time-scale τ_s is $EN(E)v/c^2$, where E is the energy of those protons which are resonant with the Alfvén waves, that is, $r_g(E) \sim \lambda_A$, and $N(E)$ is their number density. Therefore, the equation for the growth rate of the momentum of the waves is

$$\frac{1}{v_A} \frac{dU_{\text{wave}}}{dt} = \frac{EN(E)v}{\tau_s c^2}. \quad (7.21)$$

Substituting for τ_s , we find

$$\frac{dU_{\text{wave}}}{dt} = \frac{EN(E)v^2}{r_g v_A N_p m_p c^2} U_{\text{wave}}. \quad (7.22)$$

From (7.22), we find the growth rate of the waves,

$$\Gamma = \frac{EN(E)v^2}{r_g v_A N_p m_p c^2}. \quad (7.23)$$

We now write the gyroradius of the protons in terms of their velocity v and total energy E , $r_g = (Ev/eBc^2)$. Then,

$$\Gamma = \frac{eB}{m_p} \frac{N(E)}{N_p} \left(\frac{v}{v_A} \right). \quad (7.24)$$

But $\omega_g = (eB/m_p)$ is the non-relativistic angular gyrofrequency of the proton and so we obtain the final answer,

$$\Gamma = \omega_g \frac{N(E)}{N_p} \left(\frac{v}{v_A} \right). \quad (7.25)$$

This result is of similar form to that given by Cesarsky for the typical growth rate of the instability (Cesarsky, 1980):

$$\Gamma(k) = \omega_g \frac{N(\geq E)}{N_p} \left(-1 + \frac{|v|}{v_A} \right), \quad (7.26)$$

where $N(\geq E)$ means all those particles with energies greater than or equal to that energy E which resonates with the wave. The result is that the instability develops until the streaming velocity of the high energy particles is reduced to the Alfvén velocity, $v_A = B_0/(\mu_0 \rho)^{1/2}$.

Applying this result to the interstellar gas, if the density of the ionised component is $N = 10^5 \text{ m}^{-3}$ and the magnetic flux density $B_0 = 3 \times 10^{-10} \text{ T}$, then $v_A = 2 \times 10^4 \text{ m s}^{-1}$. This mechanism therefore provides a means of preventing the streaming of cosmic rays along the magnetic field lines and, at the same time, isotropising the particle distribution in pitch angle.

These results apply for the case of a fully ionised plasma. They are somewhat modified if there are neutral particles in the interstellar medium since these can lead to damping of the Alfvén waves. The instability is only effective if the waves produced by it are not damped before they have time to grow to significant amplitude. The presence of neutral particles in the interstellar plasma can abstract energy from the Alfvén waves by neutral-ion collisions, in a time short compared with the growth time. The significance of the neutral particles is that they provide a mechanism for removing kinetic energy from the waves, whereas ionised particles are constrained to oscillate with the waves. The damping rate for the waves is given by Kulsrud and Pearce for temperatures $T = 10^3$ and 10^4 K , $\Gamma^* = \Gamma_0 N_H = (3.3 \text{ and } 8.4) \times 10^{-9} N_H \text{ s}^{-1}$, respectively, where N_H is the number density of neutral hydrogen atoms (Kulsrud and Pearce, 1969).

7.5 The diffusion-loss equation for high energy particles

The considerations of Sects 7.3 and 7.4 suggest that, because of random scattering by irregularities in the magnetic field, either associated with fluctuations in the field or with the growth of instabilities due to the streaming motions of the particles, high energy charged particles can be considered to diffuse from their sources through the interstellar medium. A scalar diffusion coefficient D can therefore be used to describe their motion. As the particles diffuse, they are subject to various energy gains and losses, nuclei may suffer spallation which results in their transformation into lighter nuclei, and so on. A useful tool for studying the effects of such phenomena on the spectrum of the particles is the partial differential equation which describes the energy spectrum at different points in the interstellar medium in the presence of energy losses and with the continuous supply of

fresh particles from sources. We give two derivations of the *diffusion-loss equation* for high energy particles, which will find numerous applications throughout this text, both for nuclei and electrons, as well as providing a convenient way of deriving the predicted spectrum of accelerated particles.

7.5.1 Elementary approach

Consider an elementary volume dV into which particles are injected at a rate $Q(E, t) dV$. The particles within dV are subject to energy gains and losses which we write

$$-\frac{dE}{dt} = b(E), \quad (7.27)$$

where, if $b(E)$ is positive, the particles lose energy. Consider first the change in the energy spectrum of the particles $N(E) dE$ due to the energy losses $b(E)$ in the absence of injection of particles. At time t , the number of particles in the energy range E to $E + \Delta E$ is $N(E) \Delta E$. At a later time $t + \Delta t$, these particles are replaced by those that had energies in the range E' to $E' + \Delta E'$ at time t , where

$$E' = E + b(E) \Delta t \quad \text{and} \quad E' + \Delta E' = (E + \Delta E) + b(E + \Delta E) \Delta t. \quad (7.28)$$

Performing a Taylor expansion for small values of ΔE and subtracting,

$$\Delta E' = \Delta E + \frac{db(E)}{dE} \Delta E \Delta t. \quad (7.29)$$

Therefore, the change in $N(E) \Delta E$ in the time interval Δt is

$$\Delta N(E) \Delta E = -N(E, t) \Delta E + N[E + b(E) \Delta t, t] \Delta E'. \quad (7.30)$$

Performing another Taylor expansion for small $b(E) \Delta t$ and substituting for $\Delta E'$, we obtain

$$\Delta N(E) \Delta E = \frac{dN(E)}{dE} b(E) \Delta E \Delta t + N(E) \frac{db(E)}{dE} \Delta E \Delta t, \quad (7.31)$$

that is,

$$\frac{dN(E)}{dt} = \frac{d}{dE} [b(E)N(E)]. \quad (7.32)$$

This equation describes the time evolution of the particle spectrum in the elementary volume dV subject only to energy gains and losses. We may now add other terms to this transfer equation. If particles are injected at a rate $Q(E, t)$ per unit volume,

$$\frac{dN(E)}{dt} = \frac{d}{dE} [b(E)N(E)] + Q(E, t). \quad (7.33)$$

Particles enter and leave the volume dV by diffusion and this process depends upon the gradient of particle density $N(E)$. Adopting a scalar diffusion coefficient D ,

$$\frac{dN(E)}{dt} = \frac{d}{dE} [b(E)N(E)] + Q(E, t) + D \nabla^2 N(E). \quad (7.34)$$

This is the *diffusion-loss equation* for the time evolution of the energy spectrum of the particles.

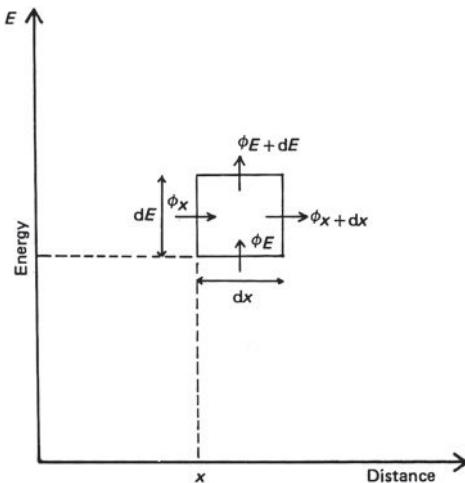


Fig. 7.6

A coordinate space diagram of energy against spatial coordinates used in deriving the diffusion-loss equation.

7.5.2 The coordinate space approach

A neater approach is to introduce a *coordinate space diagram* in which energy is plotted along the ordinate and spatial coordinates along the abscissa (Fig. 7.6). The fluxes ϕ of particles through different surfaces in the coordinate space are shown. If we consider the little rectangle, particles move in the x -direction by diffusion and in the y -direction by energy gains or losses. The number of particles in the distance increment dx and energy increment E to $E + dE$ is $N(E, x, t) dE dx$. Therefore, the rate of change of particle density in the box in coordinate space is

$$\begin{aligned} \frac{d}{dt} N(E, x, t) dE dx &= [\phi_x(E, x, t) - \phi_{x+dx}(E, x + dx, t)] dE \\ &\quad + [\phi_E(E, x, t) - \phi_{E+dE}(E + dE, x, t)] dx \\ &\quad + Q(E, x, t) dE dx , \end{aligned} \quad (7.35)$$

where $Q(E, x, t)$ is the rate of injection of particles per unit volume of coordinate space. Performing a Taylor expansion and simplifying the notation,

$$\frac{dN}{dt} = -\frac{\partial \phi_x}{\partial x} - \frac{\partial \phi_E}{\partial E} + Q . \quad (7.36)$$

ϕ_x is the flux of particles through the energy interval dE at the point x in space and hence, by definition,

$$\phi_x = -D \frac{\partial N}{\partial x} \quad \text{and so} \quad \frac{dN}{dt} = D \frac{\partial^2 N}{\partial x^2} - \frac{\partial \phi_E}{\partial E} + Q . \quad (7.37)$$

We can generalise (7.37) to three dimensions,

$$\frac{dN}{dt} = D \nabla^2 N - \frac{\partial \phi_E}{\partial E} + Q , \quad (7.38)$$

where ϕ_E is the flux of particles through dx which have energies in the range E to $E + dE$ at some time interval dt . If $-dE/dt = b(E)$ is the loss rate of particles of energy E , then the number passing through E in unit time is

$$N(E) \frac{dE}{dt} = \phi_E = -b(E)N(E). \quad (7.39)$$

Therefore we obtain

$$\frac{dN}{dt} = D \nabla^2 N + \frac{\partial}{\partial E} [b(E)N(E)] + Q(E), \quad (7.40)$$

as before.

We can add other terms to this equation, for example, to include terms describing spallation gains and losses, catastrophic loss of particles, radioactive decay, and so on. For example, in the case of the propagation of cosmic ray nuclei, (7.40) can be used to include the effects of spallation gains and losses. The diffusion loss equation for the species i becomes

$$\frac{\partial N_i}{\partial t} = D \nabla^2 N_i + \frac{\partial}{\partial E} [b(E)N_i] + Q_i - \frac{N_i}{\tau_i} + \sum_{j>i} \frac{P_{ji}}{\tau_j} N_j. \quad (7.41)$$

where N_i is the number density of nuclei of species i and is a function of energy, that is, we should write $N_i(E)$. The last two terms describe the effects of spallation gains and losses. τ_i and τ_j are the spallation lifetimes of particles of species i and j . The spallation of all species with $j > i$ results in contributions to N_i as indicated by the sum in the last term of (7.41). P_{ji} is the probability that, in an inelastic collision involving the destruction of the nucleus j , the species i is created.

Another important extension is to the statistical acceleration of particles by random collisions. The procedure starting from the Fokker–Planck equation involves the diffusion of particles in momentum or phase space and is described by Blandford and Eichler (1987). The resulting diffusion-loss equation can be written in terms of differentials with respect to energy since the particle distribution is assumed to be isotropic in real and momentum space

$$\frac{\partial N}{\partial t} = D \nabla^2 N + \frac{\partial}{\partial E} [b(E)N] + Q + \frac{1}{2} \frac{\partial^2}{\partial E^2} [d(E)N], \quad (7.42)$$

where $d(E) = \overline{(\Delta E)^2}$ is the mean square energy change of the particles per unit time (Ginzburg and Syrovatskii, 1964). We will use this expression in the study of the acceleration of charged particles and, in a slightly different guise, in the interpretation of the Kompaneets equation (Sect. 9.4.3).

The synchrotron radiation of ultra-relativistic electrons dominates much of high energy astrophysics. The radiation, which was first observed in early betatron experiments,¹ is the emission of high energy electrons gyrating in a magnetic field and is the process responsible for the radio emission of our Galaxy, of supernova remnants and extragalactic radio sources. It is also the origin of the non-thermal continuum optical emission of the Crab Nebula and quite possibly of the optical and X-ray continuum emission of quasars. The term *non-thermal emission* is frequently used in high energy astrophysics and is conventionally taken to mean the continuum radiation of a distribution of particles with a non-Maxwellian energy spectrum. Continuum emission is often referred to as ‘non-thermal’ if its spectrum cannot be accounted for by the spectrum of thermal bremsstrahlung or black-body radiation.

It is a major undertaking to work out all the detailed properties of synchrotron radiation. For more complete treatments, the enthusiast is referred to the books by Bekefi (1966), by Pacholczyk (1970) and by Rybicki and Lightman (1979), and to the review articles by Ginzburg and Syrovatskii (1965, 1969). Many of the most important results can, however, be derived by simple physical arguments (Scheuer, 1966). First of all, let us work out the total energy loss rate.

8.1 The total energy loss rate

Most of the essential tools have already been developed in Sects 6.2 and 7.1. To recapitulate the results of Sect. 7.1, in a uniform magnetic field, a high energy electron moves in a spiral path at a constant pitch angle α .² Its velocity along the field lines is constant whilst it gyrates about the magnetic field direction at the relativistic gyrofrequency $v_g = eB/2\pi\gamma m_e = 28\gamma^{-1}$ GHz T⁻¹, where γ is the Lorentz factor of the electron $\gamma = (1 - v^2/c^2)^{-1/2}$ (Fig. 8.1a). The electron is therefore accelerated towards the guiding centre of its orbit and its radiation rate can be derived from the results of Sect. 6.2.4. From (6.25), the radiation loss rate of a charged particle q with accelerations a_\perp and a_\parallel as measured in the laboratory frame of reference is

$$-\left(\frac{dE}{dt}\right)_{\text{rad}} = \frac{q^2\gamma^4}{6\pi\epsilon_0 c^3} [|a_\perp|^2 + \gamma^2 |a_\parallel|^2]. \quad (8.1)$$

¹ For more details, see *The Cosmic Century* (Longair, 2006).

² In this chapter, α is the pitch angle of the electron rather than θ , which is reserved for integrating over angular coordinates.

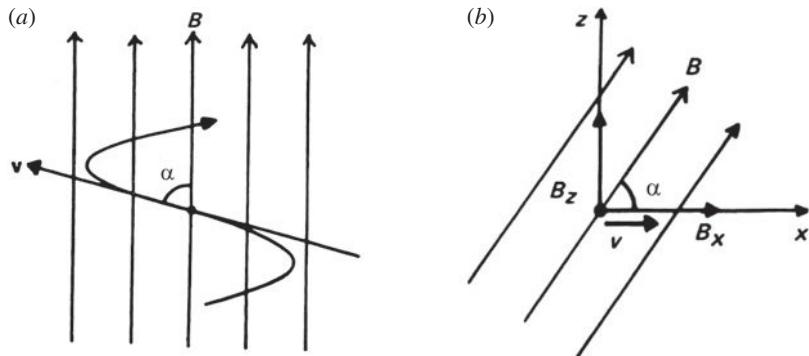


Fig. 8.1 The coordinates used in working out the total radiation rate due to synchrotron radiation.

The acceleration is always perpendicular to the velocity vector of the particle and hence from (7.3), $a_{\perp} = evB \sin \alpha / \gamma m_e$ and $a_{\parallel} = 0$. Therefore, the total radiation loss rate of the electron is

$$\begin{aligned} -\left(\frac{dE}{dt}\right) &= \frac{\gamma^4 e^2}{6\pi\epsilon_0 c^3} |a_{\perp}|^2 = \frac{\gamma^4 e^2}{6\pi\epsilon_0 c^3} \frac{e^2 v^2 B^2 \sin^2 \alpha}{\gamma^2 m_e^2} \\ &= \frac{e^4 B^2}{6\pi\epsilon_0 c m_e^2} \frac{v^2}{c^2} \gamma^2 \sin^2 \alpha . \end{aligned} \quad (8.2)$$

Another pleasant way of arriving at the same result is to start from the fact that, in the instantaneous rest frame of the electron, the acceleration of the particle is small and therefore in that frame we can use the non-relativistic expression for the radiation rate. Let us choose the coordinate system shown in Fig. 8.1b in which the instantaneous direction of motion of the electron in the laboratory frame, the frame in which \mathbf{B} is fixed, is taken to be the positive x -axis. Then, to find the force acting on the particle, we transform the field quantities into the instantaneous rest frame of the electron using the standard relativistic transformations for the magnetic field strength (see Sect. 5.3.1). In S' , the force on the electron is

$$\mathbf{F}' = m_e \dot{\mathbf{v}}' = e(\mathbf{E}' + \mathbf{v}' \times \mathbf{B}') = e\mathbf{E}' , \quad (8.3)$$

since the particle is instantaneously at rest in S' , $\mathbf{v}' = 0$. Therefore, in transforming the magnetic flux density \mathbf{B} into S' , we need only consider the transformed components of the electric field \mathbf{E}' .

$$\begin{aligned} E'_x &= E_x , & E'_x &= 0 , \\ E'_y &= \gamma(E_y - vB_z) , & \text{and hence} & E'_y &= -v\gamma B_z , \\ E'_z &= \gamma(E_z + vB_y) , & & E'_z &= 0 . \end{aligned}$$

Therefore

$$\dot{\mathbf{v}}' = -\frac{e\gamma v B \sin \alpha}{m_e} . \quad (8.4)$$

Consequently, in the rest frame of the electron, the loss rate by radiation is

$$-\left(\frac{dE}{dt}\right)' = \frac{e^2 |\dot{\mathbf{v}}'|^2}{6\pi\epsilon_0 c^3} = \frac{e^4 \gamma^2 B^2 v^2 \sin^2 \alpha}{6\pi\epsilon_0 c^3 m_e^2}. \quad (8.5)$$

Since $(dE/dt)'$ is a Lorentz invariant (Sect. 6.2.1), we recover (8.2).

Let us rewrite (8.5) in the following way,

$$-\left(\frac{dE}{dt}\right) = 2 \left(\frac{e^4}{6\pi\epsilon_0^2 c^4 m_e^2} \right) \left(\frac{v}{c} \right)^2 c \frac{B^2}{2\mu_0} \gamma^2 \sin^2 \alpha. \quad (8.6)$$

where we have used the relation $c^2 = (\mu_0\epsilon_0)^{-1}$. The quantity in the first set of round brackets on the right-hand side of this expression is the Thomson cross-section σ_T . Therefore,

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \left(\frac{v}{c} \right)^2 \gamma^2 \sin^2 \alpha, \quad (8.7)$$

where $U_{\text{mag}} = B^2/2\mu_0$ is the energy density of the magnetic field. In the ultra-relativistic limit, $v \rightarrow c$, the total loss rate is

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \gamma^2 \sin^2 \alpha. \quad (8.8)$$

These results apply for electrons with pitch angle α . As discussed in Sects 7.3 and 7.4, the pitch angle distribution is likely to be randomised either by irregularities in the magnetic field distribution or by streaming instabilities. As a result, the distribution of pitch angles for a population of high energy electrons is expected to be isotropic. In addition, during its lifetime, any high energy electron is randomly scattered in pitch angle and so, by averaging over pitch angle, an expression for its average energy loss rate is obtained. Averaging over an isotropic distribution of pitch angles $p(\alpha) d\alpha = \frac{1}{2} \sin \alpha d\alpha$, we find the average energy loss rate,

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \gamma^2 \left(\frac{v}{c} \right)^2 \frac{1}{2} \int_0^\pi \sin^3 \alpha d\alpha = \frac{4}{3} \sigma_T c U_{\text{mag}} \left(\frac{v}{c} \right)^2 \gamma^2. \quad (8.9)$$

8.2 Non-relativistic gyroradiation and cyclotron radiation

We consider first the case of non-relativistic gyroradiation in which case $v \ll c$ and $\gamma = 1$. The expression for the loss rate of the electron is then

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \left(\frac{v}{c} \right)^2 \sin^2 \alpha = \frac{2\sigma_T}{c} U_{\text{mag}} v_\perp^2, \quad (8.10)$$

and the radiation is emitted at the non-relativistic gyrofrequency of the electron $\nu_g = eB/2\pi m_e$.

The polarisation properties of gyroradiation are quite distinctive. In the non-relativistic case, there are no beaming effects and what is observed by the distant observer can be derived from the rules given in Sect. 6.2.2. When the magnetic field is perpendicular to

the line of sight, *linearly polarised radiation* is observed because the acceleration vector performs simple harmonic motion in a plane perpendicular to the magnetic field direction. The electric field strength varies sinusoidally at the gyrofrequency as the dipole distribution of radiation sweeps past the observer. When the magnetic field direction is parallel to the line of sight, the acceleration vector is continuously changing direction as the electron moves in a circular orbit about the magnetic field lines and therefore the radiation is observed to be 100% *circularly polarised*. When observed at an arbitrary angle θ to the magnetic field direction, the radiation is observed to be *elliptically polarised*, the ratio of axes of the polarisation ellipse being $\cos \theta$.

In the case of mildly relativistic cyclotron radiation, the beaming of the radiation cannot be neglected. Even for slowly moving electrons, $v \ll c$, not all the radiation is emitted at the gyrofrequency because of small aberration effects which slightly distort the observed angular distribution of the intensity from a $\cos^2 \theta$ law. The observed polar diagram of the radiation may be decomposed by Fourier analysis into a sum of equivalent dipoles radiating at harmonics of the relativistic gyrofrequency, $v_r = v_g/\gamma$. These harmonics have frequencies

$$v_l = \frac{lv_r}{\left(1 - \frac{v_{||}}{c} \cos \theta\right)}, \quad (8.11)$$

where l takes integral values, $l = 1, 2, 3, \dots$, the fundamental gyrofrequency having $l = 1$. The factor $[1 - (v_{||}/c) \cos \theta]$ in the denominator takes account of the Doppler shift of the radiation of the electron due to its translational motion along the field lines $v_{||}$, projected onto the line of sight to the observer. In the limit $lv/c \ll 1$, the total power emitted in a given harmonic for the case $v_{||} = 0$ is

$$-\left(\frac{dE}{dt}\right)_l = \frac{2\pi e^2 v_g^2}{\epsilon_0 c} \frac{(l+1)l^{2l+1}}{(2l+1)!} \left(\frac{v}{c}\right)^2. \quad (8.12)$$

Hence, to order of magnitude,

$$\left(\frac{dE}{dt}\right)_{l+1} / \left(\frac{dE}{dt}\right)_l \approx \left(\frac{v}{c}\right)^2. \quad (8.13)$$

Thus, the energy radiated in high harmonics is small when the particle is non-relativistic. Notice that the loss rate (8.12) reduces to (8.10) for $l = 1$.

When the electrons become significantly relativistic, the energy radiated in the higher harmonics becomes important. The Doppler and aberration effects result in a spread of emitted frequencies associated with the different pitch angles of an electron of total energy $E = \gamma mc^2$. The result is broadening of the emission line of a given harmonic and, for high harmonics, the lines become so broadened that the emission spectrum is continuous rather than consisting of a series of discrete harmonics. The results of calculations of the cyclotron radiation for a mildly relativistic plasma having $kT_e/m_ec^2 = 0.1$, corresponding to $\gamma = 1.1$ and $v/c \approx 0.4$, are shown in Fig. 8.2 (Bekefi, 1966). The spectra of the first 20 harmonics are shown as well as the total emission spectrum found by summing the spectra of the individual harmonics. One way of thinking about the spectrum of synchrotron radiation

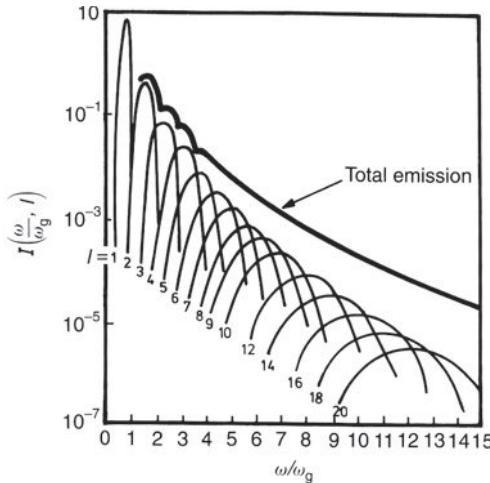


Fig. 8.2 The spectrum of emission of the first 20 harmonics of mildly relativistic cyclotron radiation for an electrons with $v = 0.4c$ (Bekefi, 1966).

is to consider it to be the relativistic limit of the process illustrated in Fig. 8.2 – all the harmonics are washed out and a smooth continuum spectrum is observed. Just as in the case of gyroradiation, the harmonics of cyclotron radiation are elliptically polarised.

Cyclotron absorption features in the energy range 10–100 keV have been observed in a number of accreting pulsars which are X-ray sources (Coburn *et al.*, 2006). The first example was discovered in the X-ray binary system Her X-1 and the broad absorption feature observed about 35 keV has been clearly detected in observations with the INTEGRAL γ -ray observatory (Klochkov *et al.*, 2008). The inferred magnetic flux densities for these sources lie in the range $(1 - 3) \times 10^8$ T, similar to the strong magnetic fields inferred from the spin down rates of radio pulsars.

Circularly polarised optical emission is observed in the eclipsing magnetic binary stars known as *AM Herculis binaries* or *polars*, circular polarisation percentages as large as 40% being observed. In these systems, a red dwarf star orbits a white dwarf with a very strong magnetic field. Accretion of matter from the surface of the red dwarf onto the magnetic poles of the white dwarf results in the heating of the matter to temperatures in excess of 10^7 K. Thus, in addition to radiating X-rays, these objects are strong sources of cyclotron radiation. Fields of order 2000 T have been found in these objects and hence the fundamental gyrofrequency is expected to correspond to a wavelength of about 5 μm . In the X-ray source EXO 033319–2554.2 (Fig. 8.3), the separate harmonics have been observed. The frequency spacing between harmonics enabled an estimate of 5600 T for the magnetic field strength to be made. In addition, observations of the variation of the circular polarisation with orbital phase enable the geometry of the magnetic field configuration to be determined.

An example of cyclotron features observed in absorption was made by Bignami and his colleagues in a very long X-ray observation of the isolated neutron star 1E1207.4–5209 by the XMM-Newton X-ray Observatory (Bignami *et al.*, 2003). The high sensitivity X-ray

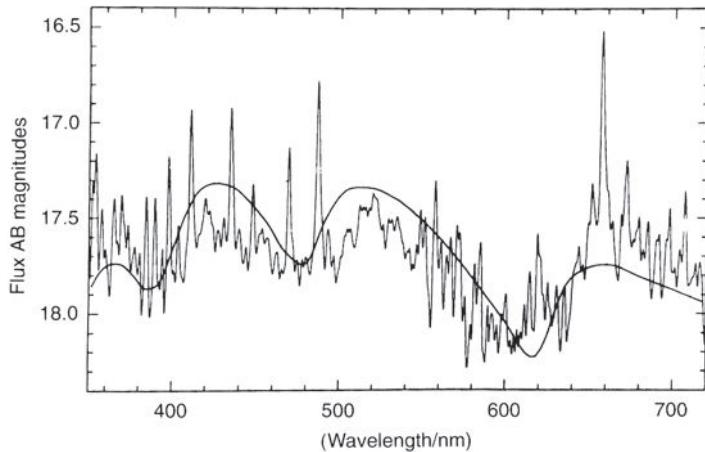


Fig. 8.3 A broad-band spectrum of the AM Herculis object EXO 033319–2554.2 which is a soft X-ray source. The presence of a strong magnetic field is inferred from the observation of strongly circularly polarised emission. The solid line shows a best fit of the cyclotron emission spectrum to the broad cyclotron harmonics at 420, 520 and 655 nm. The inferred strength of the magnetic field is 5600 T (Ferrario *et al.*, 1989).

spectral observations show three distinct features, regularly spaced at 0.7, 1.4 and 2.1 keV, once a smooth continuum spectrum has been subtracted from the total X-ray spectrum (Fig. 8.4). These features vary in strength at different phases of the rotation of the neutron star, the strongest absorption occurring at minimum intensity, as illustrated by the inset in Fig. 8.4. These features are interpreted as the fundamental and first two harmonics of cyclotron resonant absorption in the atmosphere of the neutron star. The inferred magnetic flux density in the absorbing region is found to be 8×10^6 T.

8.3 The spectrum of synchrotron radiation – physical arguments

The next step is to work out the spectrum of synchrotron radiation, an exercise which requires considerable effort. Let us therefore first analyse some basic features of radiation mechanisms involving relativistic electrons which will prove helpful in understanding the exact results.

One of the general features of the radiation of relativistic electrons is that the radiation is *beamed* in the direction of motion of the electron. This is primarily associated with the effects of relativistic aberration between the instantaneous rest frame of the electron and the observer's frame of reference. In addition, we need to consider carefully the time development of the radiation detected by the distant observer.

Consider first an electron gyrating about the magnetic field direction at a pitch angle of 90° . The electron is accelerated towards its guiding centre, that is, radially inwards, and in its instantaneous rest frame emits dipole radiation with respect to the acceleration

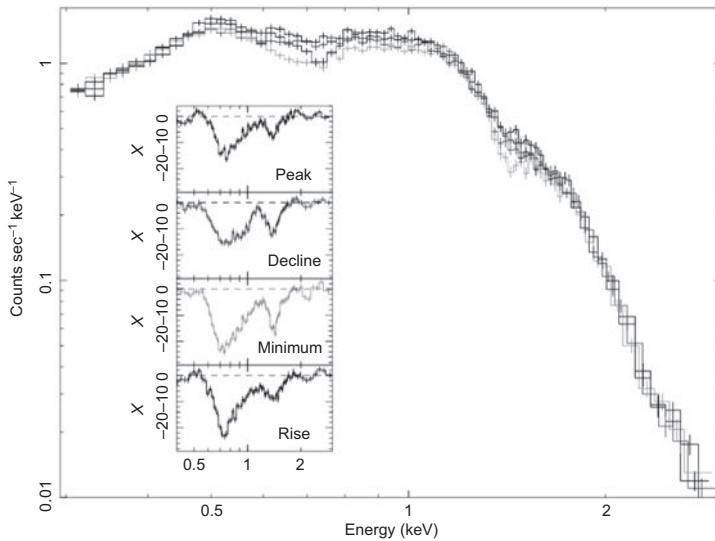


Fig. 8.4 Comparison of four X-ray spectra of the isolated neutron star 1E1207.4–5209 at four different phases of the star's rotation. The absorption lines are at their minimum (black points) at the maximum of the X-ray light curve while the absorption lines are more important at the minimum of the light curve (light grey points). The four panels in the inset show the residuals of the phase dependent spectra once a two-black-body fit to the continuum spectrum has been subtracted. Absorption features are observed in all four spectra at X-ray energies 0.7, 1.4 and 2.1 keV (Bignami *et al.*, 2003).

vector, as illustrated in Fig. 8.5a. We can therefore work out the radiation pattern in the laboratory frame of reference by applying the aberration formulae with the results illustrated schematically in Fig. 8.5b. As discussed in Sect. 5.2.2, the angular distribution of the intensity of radiation with respect to the acceleration vector in the instantaneous rest frame S' is $I_\nu \propto \sin^2 \theta' = \cos^2 \phi'$, where $\phi' = 90^\circ - \theta'$. The aberration formulae between the two frames are:

$$\sin \phi = \frac{1}{\gamma} \frac{\sin \phi'}{1 + (v/c) \cos \phi'}; \quad \cos \phi = \frac{\cos \phi' + v/c}{1 + (v/c) \cos \phi'}. \quad (8.14)$$

To illustrate the beaming of the radiation, consider the angles $\phi' = \pm\pi/4$, at which the intensity of radiation falls to half its maximum value, which occurs at $\phi' = \pi/2$ in the instantaneous rest frame of the electron. The corresponding angles in the laboratory frame of reference are

$$\sin \phi \approx \phi \approx \pm 1/\gamma, \quad (8.15)$$

recalling that $\gamma \gg 1$. Thus, the radiation emitted within $-\pi/4 < \phi' < \pi/4$ is beamed in the direction of motion of the electron within the angles $-1/\gamma < \phi < 1/\gamma$. In the observer's frame S , the dipole beam pattern is very strongly elongated in the direction of motion of the electron (Fig. 8.5b). When this elongated beam pattern sweeps past the observer, a pulse of radiation is observed every time the electron's velocity vector lies within an angle of about

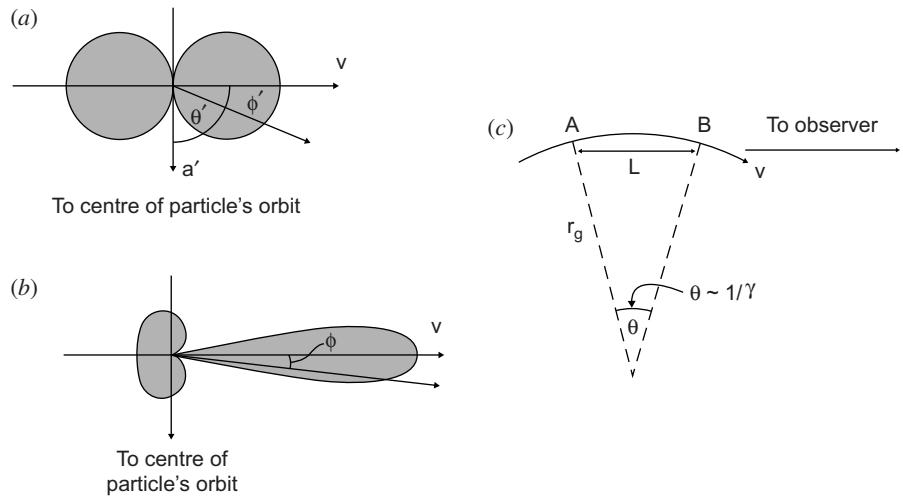


Fig. 8.5 Illustrating the relativistic beaming effects associated with synchrotron radiation. (a) The polar diagram of dipole radiation of the electron in its instantaneous rest frame. (b) The polar diagram of the radiation transformed into the laboratory frame of reference. (c) The geometry of the path of the electron during the time when the beamed radiation is observed by the distant observer.

$\pm 1/\gamma$ to the line of sight to the observer. The spectrum of the radiation received by the distant observer is the Fourier transform of this pulse, once the effects of the time delay of the radiation are taken into account. This analysis illustrates why the observed frequency of the radiation is very much greater than the gyrofrequency.

Significant radiation is only observed by a distant observer from about $1/\gamma$ radians of the electron's orbit but the observed duration of the pulse is less than $1/\gamma$ times the period of the orbit because radiation emitted at the trailing edge of the pulse almost catches up with the radiation emitted at the leading edge. Let us illustrate this key result by a simple calculation carried out entirely in the laboratory frame of reference S which concerns the time of arrival of the signals at the distant observer. The segment of the electron's orbit from which significant radiation is received by the distant observer is shown in Fig. 8.5c. Consider an observer located at a distance R from the point A. The radiation from A reaches the observer at time R/c . The radiation emitted from B takes place at time L/v later and it then travels a distance $(R - L)$ at the speed of light to reach the observer. The trailing edge of the pulse therefore arrives at the observer at a time $L/v + (R - L)/c$. The duration of the pulse as measured by the observer is therefore

$$\Delta t = \left[\frac{L}{v} + \frac{(R - L)}{c} \right] - \frac{R}{c} = \frac{L}{v} \left[1 - \frac{v}{c} \right]. \quad (8.16)$$

The observed duration of the pulse is much less than the time interval L/v , which might have been expected. Only if light propagated at an infinite velocity would the duration of the pulse be L/v . The intriguing point about this analysis is that the factor $1 - (v/c)$ is exactly the same factor which appears in the Liénard–Weichert potentials (6.19) and which takes account of the fact that the source of radiation is moving towards the observer. The

relativistic electron almost catches up with the radiation emitted at A since $v \approx c$, but not quite. We can rewrite (8.16) using the fact that

$$\frac{L}{v} = \frac{r_g \theta}{v} \approx \frac{1}{\gamma \omega_r} = \frac{1}{\omega_g}, \quad (8.17)$$

where ω_g is the non-relativistic angular gyrofrequency and $\omega_r = \omega_g/\gamma$ the relativistic angular gyrofrequency. We can also rewrite $(1 - v/c)$ as

$$\left(1 - \frac{v}{c}\right) = \frac{[1 - (v/c)][1 + (v/c)]}{[1 + (v/c)]} = \frac{(1 - v^2/c^2)}{1 + (v/c)} \approx \frac{1}{2\gamma^2}, \quad (8.18)$$

since $v \approx c$. Therefore, the observed duration of the pulse is

$$\Delta t \approx \frac{1}{2\gamma^2 \omega_g}. \quad (8.19)$$

This means that the duration of the pulse as observed by a distant observer in the laboratory frame of reference is roughly $1/\gamma^2$ times shorter than the non-relativistic gyroperiod $T_g = 2\pi/\omega_g$. The maximum Fourier component of the spectral decomposition of the observed pulse of radiation is expected to correspond to a frequency $v \sim \Delta t^{-1}$, that is,

$$v \sim \Delta t^{-1} \sim \gamma^2 \nu_g, \quad (8.20)$$

where ν_g is the non-relativistic gyrofrequency. This result is similar to the expression for the *critical frequency* for synchrotron radiation which will appear in the more complete analysis.

In the above calculation, it has been assumed that the electron moves in a circle about the magnetic field lines at pitch angle $\alpha = 90^\circ$. The same calculation can be performed for any pitch angle with the result

$$v \sim \gamma^2 \nu_g \sin \alpha. \quad (8.21)$$

The reason for performing this simple exercise in detail is that the beaming of the radiation of ultra-relativistic electrons is a very general property and does not depend upon the nature of the force causing the acceleration. The observed frequency of the beamed radiation can also be written

$$v \approx \gamma^2 \nu_g = \gamma^3 \nu_r = \frac{\gamma^3 v}{2\pi r_g}, \quad (8.22)$$

where ν_r is the relativistic gyrofrequency and r_g is the radius of the electron's orbit. In general, we may interpret r_g as the *instantaneous radius of curvature* of the electron's trajectory and v/r_g is the angular frequency associated with it. This result enables us to work out the frequency at which most of the radiation is emitted, provided we know the radius of curvature. The frequency of the observed radiation is roughly γ^3 times the angular frequency v/r where r is the instantaneous radius of curvature of the electron's trajectory. This result is important in the study of *curvature radiation* which has important applications in the emission of radiation from the magnetic poles of pulsars (Sect. 13.3).

For order of magnitude calculations, it is sufficient to know that the total energy loss rate of the relativistic electron is exactly given by (8.9) and that most of the radiation is emitted at a frequency $\nu \sim \gamma^2 \nu_g$, where ν_g is the non-relativistic gyrofrequency.

8.4 The spectrum of synchrotron radiation – a fuller version

I am not aware of any particularly simple way of deriving the spectral distribution of synchrotron radiation. The analysis given below follows closely the presentation of Rybicki and Lightman and proceeds by the following steps (Rybicki and Lightman, 1979):

- (i) Write down the expression for the energy emitted per unit bandwidth for an arbitrarily moving electron,
- (ii) Select a suitable set of coordinates in which to work out the field components radiated by the electron spiralling in a magnetic field,
- (iii) Then battle away at the algebra to obtain the spectral distribution of the field components.

8.4.1 The spectrum of radiation of an arbitrarily moving electron

We begin with the generalisation of the formulae for the radiation of an accelerated charge moving at a relativistic velocity. Repeating (6.19), the Liénard–Weichert potentials are:

$$A(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \left[\frac{q\mathbf{v}}{1 - \frac{\mathbf{v} \cdot \mathbf{n}}{c}} \right]_{\text{ret}} ; \quad \phi(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0 r} \left[\frac{q}{1 - \frac{\mathbf{v} \cdot \mathbf{n}}{c}} \right]_{\text{ret}} . \quad (8.23)$$

The differences as compared with the expression for a slowly moving charge (6.18) are the presence of the Doppler shift factor $[1 - (\mathbf{v} \cdot \mathbf{n})/c]$ in the denominator and the explicit recognition that retarded quantities have to be used to work out the fields at the observer. Let us write $\kappa = [1 - (\mathbf{v} \cdot \mathbf{n})/c]$.

These potentials lead to the expression for the relation between the acceleration and the spectral energy distribution of the radiation of an arbitrarily moving electron. We repeat here the expression for the radiation spectrum of the electron when there is no net motion (6.29), writing out explicitly the Fourier transform of the acceleration.

$$I(\omega) = \frac{e^2}{6\pi^2\epsilon_0 c^3} \left| \int_{-\infty}^{\infty} \hat{\mathbf{v}}(t) \exp(i\omega t) dt \right|^2 . \quad (8.24)$$

The corresponding result for the case of a moving electron can be written

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2}{16\pi^3\epsilon_0 c} \left| \int_{-\infty}^{\infty} \left\{ \mathbf{n} \times \left[\left(\mathbf{n} - \frac{\mathbf{v}}{c} \right) \times \frac{\dot{\mathbf{v}}}{c} \right] \kappa^{-3} \right\}_{\text{ret}} \exp(i\omega t) dt \right|^2 , \quad (8.25)$$

where the angular dependence of the emitted radiation has been preserved (Rybicki and Lightman, 1979; Jackson, 1999). The vector \mathbf{n} is the unit vector from the electron to the

point of observation, $\mathbf{n} = \mathbf{R}/|R|$. Integrating (8.25) over solid angle $d\Omega = 2\pi \sin \theta d\theta$ in the non-relativistic limit gives (8.24). The key differences between (8.24) and (8.25) are the inclusion of the Doppler shift factor κ^3 in the denominator and the fact that the expression in square brackets has to be evaluated at retarded time t' where $t' = t - R(t')/c$.

The next step is to manipulate (8.25) into a more manageable form. First of all, we change the integration from an integral over dt to one over dt' . Since $t' = t - R(t')/c$, we differentiate both sides, noting that the unit vector \mathbf{n} points towards the observer:

$$dt' = dt - \frac{1}{c} \frac{dR(t')}{dt'} dt'; \quad dt = dt' \left(1 - \frac{\mathbf{n} \cdot \mathbf{v}}{c}\right) = \kappa dt'. \quad (8.26)$$

A further simplification is to write the distance to the electron $R(t') = |\mathbf{r}| - \mathbf{n} \cdot \mathbf{r}_0(t')$, where $\mathbf{r}_0(t')$ is the position vector which describes the position of the electron relative to an origin at \mathbf{r} . Note that, in all our calculations, $\mathbf{r}_0(t') \ll \mathbf{r}$. Therefore, (8.25) becomes

$$\begin{aligned} \frac{dI(\omega)}{d\Omega} = & \frac{e^2}{16\pi^3\epsilon_0 c} \left| \int_{-\infty}^{\infty} \mathbf{n} \times \left[\left(\mathbf{n} - \frac{\mathbf{v}(t')}{c} \right) \times \frac{\dot{\mathbf{v}}(t')}{c} \right] \right. \\ & \left. \times \kappa^{-2} \exp \left[i\omega \left(t' - \frac{\mathbf{n} \cdot \mathbf{r}_0(t')}{c} \right) \right] dt' \right|^2. \end{aligned} \quad (8.27)$$

The next step is to simplify the vector triple product inside the integral using the pleasant identity

$$\mathbf{n} \times \left[\left(\mathbf{n} - \frac{\mathbf{v}}{c} \right) \times \frac{\dot{\mathbf{v}}}{c} \right] \kappa^{-2} = \frac{d}{dt'} \left\{ \kappa^{-1} \left[\mathbf{n} \times \left(\mathbf{n} \times \frac{\mathbf{v}}{c} \right) \right] \right\}. \quad (8.28)$$

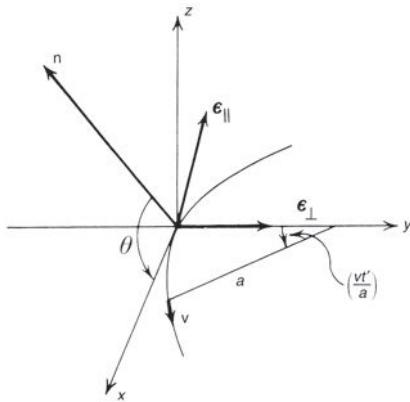
This is found by differentiating $\kappa^{-1}[\mathbf{n} \times (\mathbf{n} \times (\mathbf{v}/c))]$ with respect to t' and then using the vector triple product rule $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$. Substituting (8.28) into (8.27) and integrating by parts,

$$\frac{dI(\omega)}{d\Omega} = \frac{e^2 \omega^2}{16\pi^3\epsilon_0 c} \left| \int_{-\infty}^{\infty} \mathbf{n} \times \left(\mathbf{n} \times \frac{\mathbf{v}}{c} \right) \exp \left[i\omega \left(t' - \frac{\mathbf{n} \cdot \mathbf{r}_0(t')}{c} \right) \right] dt' \right|^2. \quad (8.29)$$

Notice that, by using the identity (8.28), we have apparently eliminated the acceleration of the charge – now only the dynamics of the electron appear in (8.29).

8.4.2 The system of coordinates

We now choose the most convenient set of coordinates for evaluating the integrals in (8.29). The electron spirals about the magnetic field lines at angular frequency $\omega_r = eB/\gamma m_e$ and at pitch angle α with respect to the magnetic field direction. At any time the orbit has a certain radius of curvature a and we take the instantaneous plane of its orbit to be the $x-y$ plane. We simplify the calculations considerably if we take the x -axis to have its origin at the point where the velocity vector \mathbf{v} of the electron lies in the $x-z$ plane which includes the observer and the y -axis to be the direction of the instantaneous radius vector \mathbf{a} of the electron at that time (Fig. 8.6). Thus, the unit vector \mathbf{n} pointing from the origin of the system of coordinates to the observer lies in the $x-z$ plane. Since \mathbf{v} is tangential to the orbit of the electron at $x = y = 0$, the vector \mathbf{n} is parallel to the magnetic field direction

**Fig. 8.6**

The geometry for evaluating the intensity and polarisation properties of synchrotron radiation. At $t = 0$, the electron velocity \mathbf{v} is instantaneously along the x -axis and a is the radius of curvature of the trajectory (Rybicki and Lightman, 1979). The unit vector \mathbf{n} points from the electron to the distant observer and lies in the x - z plane.

as seen in projection by the distant observer. This enables us to define another orthogonal set of coordinates with the same origin as x , y , z with the unit vector ϵ_{\parallel} lying in the plane containing \mathbf{n} and the magnetic field direction and the unit vector ϵ_{\perp} lying along the y -axis so that $\epsilon_{\parallel} = \mathbf{n} \times \epsilon_{\perp}$. The unit vectors ϵ_{\parallel} and ϵ_{\perp} therefore form the natural system of coordinates for describing the observed polarisation of the radiation, the \parallel and \perp symbols referring to components parallel and perpendicular to the magnetic field direction, as seen in projection by the observer.

8.4.3 The algebra

We first deal separately with the vector triple product and the exponent in the integral (8.29). To evaluate the vector triple product, we write down the coordinates of the electron in the $(\mathbf{n}, \epsilon_{\parallel}, \epsilon_{\perp})$ coordinate system, taking $x = y = z = 0$ as the point at which $t' = 0$. Therefore, after time t' , the electron has moved a distance vt' round the orbit corresponding to the angle $\varphi = vt'/a$ where a is the radius of curvature of the electron's orbit. From the geometry of Fig. 8.6,

$$\mathbf{v} = |\mathbf{v}| \left[i_x \cos \left(\frac{vt'}{a} \right) + \epsilon_{\perp} \sin \left(\frac{vt'}{a} \right) \right]. \quad (8.30)$$

We now decompose this velocity into components in the $(\mathbf{n}, \epsilon_{\parallel}, \epsilon_{\perp})$ coordinate system.

$$\mathbf{v} = |\mathbf{v}| \left[\epsilon_{\perp} \sin \left(\frac{vt'}{a} \right) + \mathbf{n} \cos \theta \cos \left(\frac{vt'}{a} \right) - \epsilon_{\parallel} \sin \theta \cos \left(\frac{vt'}{a} \right) \right], \quad (8.31)$$

where θ is the angle between the unit vector \mathbf{n} which points towards the observer and the x - y plane. Finally, we take the vector product $\mathbf{n} \times (\mathbf{n} \times \mathbf{v})$ recalling that $\epsilon_{\parallel} = \mathbf{n} \times \epsilon_{\perp}$ and $\epsilon_{\perp} = -\mathbf{n} \times \epsilon_{\parallel}$.

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{v}) = |\mathbf{v}| \left[-\sin \left(\frac{vt'}{a} \right) \epsilon_{\perp} + \sin \theta \cos \left(\frac{vt'}{a} \right) \epsilon_{\parallel} \right]. \quad (8.32)$$

Thus, the vector triple product $\mathbf{n} \times (\mathbf{n} \times \mathbf{v})$ reduces to the sum of vectors in the directions parallel and perpendicular to the magnetic field as seen in projection by the observer.

Next, we evaluate the term in the exponent in square brackets, $[t' - \mathbf{n} \cdot \mathbf{r}_0(t')/c]$ in (8.29). Again we refer to Fig. 8.6 to evaluate $\mathbf{r}_0(t')$, the position vector of the electron in its orbit. From the geometry of Fig. 8.6,

$$\mathbf{r}_0(t') = 2a \sin\left(\frac{vt'}{2a}\right) \left[\epsilon_{\perp} \sin\left(\frac{vt'}{2a}\right) + \mathbf{n} \cos\theta \cos\left(\frac{vt'}{2a}\right) - \epsilon_{\parallel} \sin\theta \cos\left(\frac{vt'}{2a}\right) \right]. \quad (8.33)$$

Then, substituting for $\mathbf{r}_0(t')$ into $[t' - \mathbf{n} \cdot \mathbf{r}_0(t')/c]$, we find

$$\left[t' - \frac{\mathbf{n} \cdot \mathbf{r}_0(t')}{c} \right] = t' - \frac{a}{c} \cos\theta \sin\left(\frac{vt'}{a}\right). \quad (8.34)$$

We now investigate the main contributions to the integral (8.29). The greatest contributions come from the smallest values of $[t' - \mathbf{n} \cdot \mathbf{r}_0(t')/c]$ since, if this quantity were large, there would be many ‘oscillations’ in the integral and these would average out to a very small value. Furthermore, we know from our physical analysis of synchrotron radiation in Sect. 8.3 that most of the radiation is strongly beamed in the direction of motion of the electron. Therefore, the principal contributions to the spectral distribution of the radiation are from small values of θ and correspondingly small values of vt'/a , as can be appreciated from the geometry of Fig. 8.6. Therefore, expanding (8.34) to third order in the small quantities θ and vt'/a ,

$$t' - \frac{\mathbf{n} \cdot \mathbf{r}_0(t')}{c} = t' \left(1 - \frac{v}{c} \right) + \frac{v}{c} \frac{\theta^2}{2} t' + \frac{v^3}{6ca^2} t'^3. \quad (8.35)$$

Since $v \approx c$ and $\gamma \gg 1$, we use (8.18) to write $(1 - v/c) = 1/2\gamma^2$ and hence,

$$\begin{aligned} t' - \frac{\mathbf{n} \cdot \mathbf{r}_0(t')}{c} &= \frac{1}{2\gamma^2} \left[t' \left(1 + \gamma^2 \frac{v}{c} \theta^2 \right) + \frac{v^3 \gamma^2 t'^3}{3ca^2} \right] \\ &= \frac{1}{2\gamma^2} \left[t'(1 + \gamma^2 \theta^2) + \frac{c^2 \gamma^2 t'^3}{3a^2} \right], \end{aligned} \quad (8.36)$$

where we have set $v = c$ in the last relation.

We next make the same small angle approximations for $\mathbf{n} \times (\mathbf{n} \times \mathbf{v}/c)$ and find

$$\mathbf{n} \times \left(\mathbf{n} \times \frac{\mathbf{v}}{c} \right) = \frac{|\mathbf{v}|}{c} \left[-\sin\left(\frac{vt'}{a}\right) \epsilon_{\perp} + \sin\theta \cos\left(\frac{vt'}{a}\right) \epsilon_{\parallel} \right] \approx \left(-\frac{vt'}{a} \epsilon_{\perp} + \theta \epsilon_{\parallel} \right). \quad (8.37)$$

We can now write down the integrals for the intensities in the ϵ_{\perp} and ϵ_{\parallel} directions by substituting (8.36) and (8.37) into (8.29):

$$\frac{dI_{\perp}(\omega)}{d\Omega} = \frac{e^2 \omega^2}{16\pi^3 \epsilon_0 c} \left| \int_{-\infty}^{\infty} \frac{vt'}{a} \exp \left\{ \frac{i\omega}{2\gamma^2} \left[t'(1 + \gamma^2 \theta^2) + \frac{c^2 \gamma^2}{3a^2} t'^3 \right] \right\} dt' \right|^2, \quad (8.38)$$

$$\frac{dI_{\parallel}(\omega)}{d\Omega} = \frac{e^2 \omega^2 \theta^2}{16\pi^3 \epsilon_0 c} \left| \int_{-\infty}^{\infty} \exp \left\{ \frac{i\omega}{2\gamma^2} \left[t'(1 + \gamma^2 \theta^2) + \frac{c^2 \gamma^2}{3a^2} t'^3 \right] \right\} dt' \right|^2. \quad (8.39)$$

We are almost there. Because most of the power emitted by the electron is contained within small values of θ , corresponding to small values of t' , there is little error in taking the limits of the integrals to be from $-\infty$ to $+\infty$. We make the following changes of variable to reduce the integrals to standard forms:

$$\theta_\gamma^2 = (1 + \gamma^2 \theta^2); \quad y = \gamma c t' / a \theta_\gamma; \quad \eta = \omega a \theta_\gamma^3 / 3 c \gamma^3. \quad (8.40)$$

Then

$$\frac{dI_\perp(\omega)}{d\Omega} = \frac{e^2 \omega^2}{16\pi^3 \epsilon_0 c} \left(\frac{a \theta_\gamma^2}{c \gamma^2} \right)^2 \left| \int_{-\infty}^{\infty} y \exp \left[i \eta \frac{3}{2} \left(y + \frac{y^3}{3} \right) \right] dy \right|^2, \quad (8.41)$$

$$\frac{dI_\parallel(\omega)}{d\Omega} = \frac{e^2 \omega^2 \theta^2}{16\pi^3 \epsilon_0 c} \left(\frac{a \theta_\gamma}{c \gamma} \right)^2 \left| \int_{-\infty}^{\infty} \exp \left[i \eta \frac{3}{2} \left(y + \frac{y^3}{3} \right) \right] dy \right|^2. \quad (8.42)$$

The integrals can be expressed in terms of modified Bessel functions using the following relations which can be derived from relations 10.4.22 to 10.4.32 presented by Abramovitz and Stegun (1965):

$$\int_0^\infty \cos \left[\frac{3\eta}{2} \left(x + \frac{1}{3} x^3 \right) \right] dx = \frac{1}{\sqrt{3}} K_{1/3}(\eta), \quad (8.43)$$

$$\int_0^\infty x \sin \left[\frac{3\eta}{2} \left(x + \frac{1}{3} x^3 \right) \right] dx = \frac{1}{\sqrt{3}} K_{2/3}(\eta), \quad (8.44)$$

where $K_{2/3}$ and $K_{1/3}$ are modified Bessel functions of orders 2/3 and 1/3, respectively. We use the symmetry of the integrands to find the following expressions for the integrals (8.41) and (8.42):

$$\frac{dI_\perp(\omega)}{d\Omega} = \frac{e^2 \omega^2}{12\pi^3 \epsilon_0 c} \left(\frac{a \theta_\gamma^2}{c \gamma^2} \right)^2 K_{2/3}^2(\eta), \quad (8.45)$$

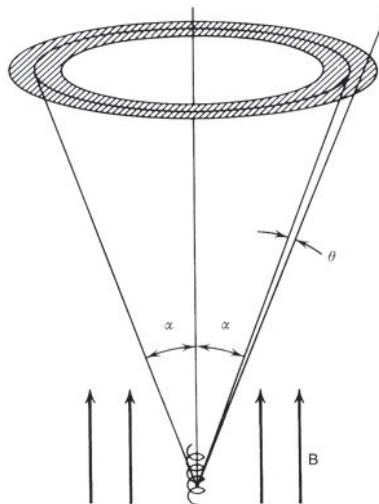
$$\frac{dI_\parallel(\omega)}{d\Omega} = \frac{e^2 \omega^2 \theta^2}{12\pi^3 \epsilon_0 c} \left(\frac{a \theta_\gamma}{c \gamma} \right)^2 K_{1/3}^2(\eta). \quad (8.46)$$

The final step is to integrate over the angle θ . Since most of the radiation is emitted within a very small angle θ with respect to the pitch angle of the electron, it can be assumed that, over one period of gyration of the electron about the magnetic field direction, the angle over which the integral is to be taken is $2\pi \sin \alpha d\theta$ because the element of solid angle varies very little over $d\theta$, whilst the radiation pattern is a strong function of θ (Fig. 8.7). We make little error in taking the limits of the integrals over θ to be $\pm\infty$ because all the power is concentrated in the angle $d\theta$ about the pitch angle α . Therefore, the integrals can be written:

$$I_\perp(\omega) = \frac{e^2 \omega^2 a^2 \sin \alpha}{6\pi^2 \epsilon_0 c^3 \gamma^4} \int_{-\infty}^{\infty} \theta_\gamma^4 K_{2/3}^2(\eta) d\theta, \quad (8.47)$$

$$I_\parallel(\omega) = \frac{e^2 \omega^2 a^2 \sin \alpha}{6\pi^2 \epsilon_0 c^3 \gamma^2} \int_{-\infty}^{\infty} \theta_\gamma^2 \theta^2 K_{1/3}^2(\eta) d\theta. \quad (8.48)$$

These integrals have been evaluated by Westfold (1959) and by Le Roux (1961). The following relations may be found from Westfold's paper by comparing his equations (23)

**Fig. 8.7**

Synchrotron emission from an electron with pitch angle α . The radiation is confined to the shaded solid angle (Rybicki and Lightman, 1979).

and (25):

$$\int_{-\infty}^{\infty} \theta_{\gamma}^4 K_{2/3}^2 \left(\frac{x}{2} \theta_{\gamma}^3 \right) d\theta = \frac{\pi}{\sqrt{3}\gamma x} \left[\int_x^{\infty} K_{5/3}(z) dz + K_{2/3}(x) \right], \quad (8.49)$$

$$\int_{-\infty}^{\infty} \gamma^2 \theta^2 \theta_{\gamma}^2 K_{1/3}^2 \left(\frac{x}{2} \theta_{\gamma}^3 \right) d\theta = \frac{\pi}{\sqrt{3}\gamma x} \left[\int_x^{\infty} K_{5/3}(z) dz - K_{2/3}(x) \right]. \quad (8.50)$$

It will be recalled that $\theta_{\gamma} = (1 + \gamma^2 \theta^2)^{1/2}$ and $x = 2\omega a / 3c\gamma^3$. It is traditional to write

$$F(x) = x \int_x^{\infty} K_{5/3}(z) dz; \quad G(x) = x K_{2/3}(x). \quad (8.51)$$

Then, using the expression $a = 3c\gamma^3 x / 2\omega$ to eliminate a from (8.47) and (8.48), we find

$$I_{\perp}(\omega) = \frac{\sqrt{3}e^2\gamma \sin \alpha}{8\pi\epsilon_0 c} [F(x) + G(x)] \quad (8.52)$$

$$I_{\parallel}(\omega) = \frac{\sqrt{3}e^2\gamma \sin \alpha}{8\pi\epsilon_0 c} [F(x) - G(x)]. \quad (8.53)$$

8.4.4 The results

After the labour of the last few pages, we present the results of these calculations in the form of formulae, tables and graphs. First of all, we introduce the *critical angular frequency* ω_c defined by $\omega_c = 3c\gamma^3 / 2a$ so that $x = \omega/\omega_c = v/v_c$. We recall that a is the radius of curvature of the electron's spiral orbit. At any instant, the plane of the electron's orbit is inclined at a pitch angle α to the magnetic field. Therefore, with respect to the guiding

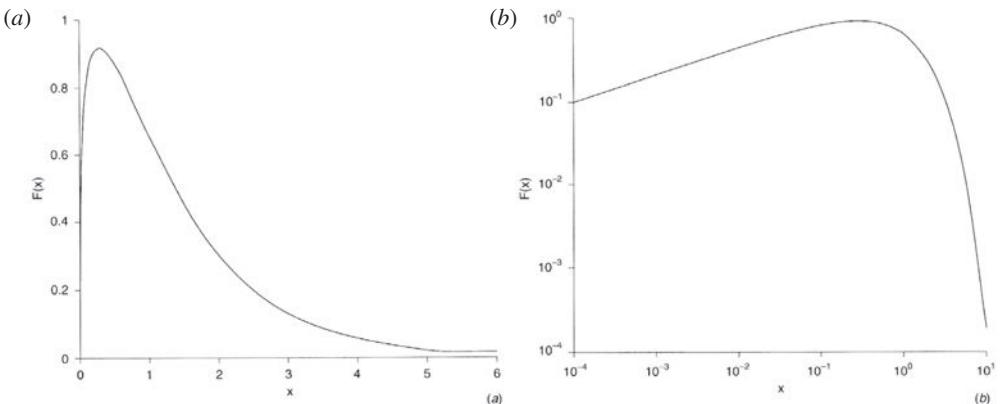


Fig. 8.8 The spectrum of the synchrotron radiation of a single electron shown (a) with linear axes; (b) with logarithmic axes. The function is plotted in terms of $x = \omega/\omega_c = v/v_c$ where ω_c is the critical angular frequency $\omega_c = 2\pi v_c = (3/2)(c/v)\gamma^2\omega_g \sin \alpha$ where α is the pitch angle of the electron and ω_g is the non-relativistic gyrofrequency, $\omega_g = eB/m_e$.

centre of the electron's trajectory, the radius of curvature is $a = v/(\omega_r \sin \alpha)$ and hence

$$\omega_c = 2\pi v_c = \frac{3}{2} \left(\frac{c}{v} \right) \gamma^3 \omega_r \sin \alpha , \quad (8.54)$$

or, taking the limit $v \rightarrow c$ and rewriting the expression in terms of the *non-relativistic gyrofrequency* $v_g = eB/2\pi m_e = 28 \text{ GHz T}^{-1}$,

$$v_c = \frac{3}{2} \gamma^2 v_g \sin \alpha . \quad (8.55)$$

This is a key result and is remarkably similar to that derived in Sect. 8.3 for the frequency at which most of the radiation is emitted, $v \approx \gamma^2 v_g$.

In integrating over $2\pi \sin \theta d\theta$ in (8.47) and (8.48), (8.52) and (8.53) represent the energy emitted in the two orthogonal polarisations during one period of the electron in its orbit, that is, in a time $T_r = v_r^{-1} = 2\pi\gamma m_e/eB$. Therefore, the emissivities of the electron in the two polarisations are

$$j_{\perp}(\omega) = \frac{I_{\perp}(\omega)}{T_r} = \frac{\sqrt{3}e^3 B \sin \alpha}{16\pi^2 \epsilon_0 c m_e} [F(x) + G(x)] , \quad (8.56)$$

$$j_{\parallel}(\omega) = \frac{I_{\parallel}(\omega)}{T_r} = \frac{\sqrt{3}e^3 B \sin \alpha}{16\pi^2 \epsilon_0 c m_e} [F(x) - G(x)] . \quad (8.57)$$

The total emissivity of a single electron by synchrotron radiation is the sum of $j_{\perp}(\omega)$ and $j_{\parallel}(\omega)$:

$$j(\omega) = j_{\perp}(\omega) + j_{\parallel}(\omega) = \frac{\sqrt{3}e^3 B \sin \alpha}{8\pi^2 \epsilon_0 c m_e} F(x) . \quad (8.58)$$

This is the spectral emissivity of a single electron by synchrotron radiation in the ultra-relativistic limit. It is shown graphically in Fig. 8.8 in linear and logarithmic forms and

Table 8.1 The synchrotron radiation spectrum $F(x)$ of a single ultra-relativistic electron where $x = \omega/\omega_c = v/v_c$ and $\omega_c = 2\pi v_c = (3/2)(c/v)\gamma^2\omega_g \sin \alpha$ where ω_g is the non-relativistic gyrofrequency, $\omega_g = eB/m_e$ (see (8.51) and (8.58)).

x	$F(x)$	x	$F(x)$
1.0×10^{-4}	0.0996	5.0×10^{-1}	0.872
1.0×10^{-3}	0.213	8.0×10^{-1}	0.742
1.0×10^{-2}	0.445	1	0.655
3.0×10^{-2}	0.613	2	0.301
1.0×10^{-1}	0.818	3	0.130
2.0×10^{-1}	0.904	5	2.14×10^{-2}
2.8×10^{-1}	0.918	10	1.92×10^{-4}
3.0×10^{-1}	0.918		

the function $F(x)$ is given in tabular form in Table 8.1. The features of the spectrum are similar to those deduced by the physical arguments given in Sect. 8.3. The spectrum has a broad maximum, $\Delta\nu/\nu \sim 1$, centred roughly at the frequency $\nu \approx v_c$ – the maximum of the emission spectrum in fact has value $\nu_{\max} = 0.29v_c$. The spectrum is smooth and continuous and use is made of this feature in large synchrotron radiation facilities to generate a precisely defined, high intensity, continuum spectrum at infrared, optical, ultraviolet and X-ray wavelengths.

Let us investigate various features of the emission spectrum. First of all, let us take the integral of the emission spectrum over all frequencies to ensure that we have obtained the correct expression for the total energy loss rate:

$$\begin{aligned} -\frac{dE}{dt} &= \int_0^\infty j(\omega) d\omega = \frac{\sqrt{3}e^3 B \omega_c \sin \alpha}{8\pi^2 \epsilon_0 c m_e} \int_0^\infty F(x) dx \\ &= \left(\frac{9\sqrt{3}}{4\pi} \right) \left(\frac{e^2}{6\pi \epsilon_0^2 c^4 m_e^2} \right) c \frac{B^2}{2\mu_0} \gamma^2 \sin^2 \alpha \int_0^\infty F(x) dx \\ &= \sigma_T c U_{\text{mag}} \gamma^2 \sin^2 \alpha \left(\frac{9\sqrt{3}}{4\pi} \right) \int_0^\infty F(x) dx . \end{aligned} \quad (8.59)$$

The integrals presented by Rybicki and Lightman can be used to evaluate (8.59) (Rybicki and Lightman, 1979):

$$\int_0^\infty x^\mu F(x) dx = \frac{2^{\mu+1}}{(\mu+2)} \Gamma\left(\frac{\mu}{2} + \frac{7}{3}\right) \Gamma\left(\frac{\mu}{2} + \frac{2}{3}\right) , \quad (8.60)$$

$$\int_0^\infty x^\mu G(x) dx = 2^\mu \Gamma\left(\frac{\mu}{2} + \frac{4}{3}\right) \Gamma\left(\frac{\mu}{2} + \frac{2}{3}\right) . \quad (8.61)$$

Setting $\mu = 0$ in (8.60) and using the recurrence relations for Γ -functions given by Abramovitz and Stegun (1965),

$$\frac{9\sqrt{3}}{4\pi} \int_0^\infty F(x) dx = \frac{9\sqrt{3}}{4\pi} \Gamma\left(\frac{7}{3}\right) \Gamma\left(\frac{2}{3}\right) = 2 , \quad (8.62)$$

and so

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \gamma^2 \sin^2 \alpha. \quad (8.63)$$

This is exactly the result (8.8) for the total energy loss rate.

Next, the asymptotic expressions for the emissivity of the electron in the high and low frequency limits can be found from the asymptotic expressions for the function $F(x)$ quoted by Rybicki and Lightman:

$$F(x) = \frac{4\pi}{\sqrt{3}\Gamma(1/3)} \left(\frac{x}{2}\right)^{1/3} \quad x \ll 1, \quad (8.64)$$

$$F(x) = \left(\frac{\pi}{2}\right)^{1/2} x^{1/2} \exp(-x) \quad dx \gg 1. \quad (8.65)$$

The high frequency emissivity of the electron is therefore given by an expression of the form

$$j(v) \propto v^{1/2} \exp(-v/v_c), \quad (8.66)$$

which is dominated by the exponential cut-off at frequencies $v \gg v_c$. There is very little power at frequencies $v > v_c$ because there is very little structure in the polar diagram of the radiation emitted by the electron at angles $\theta \ll \gamma^{-1}$.

At low frequencies, $v \ll v_c$, the spectrum is

$$\begin{aligned} j(\omega) &= \frac{\sqrt{3}e^3 B \sin \alpha}{8\pi^2 \epsilon_0 c m_e} \frac{4\pi}{\sqrt{3}\Gamma(1/3)} \left(\frac{\omega}{2\omega_c}\right)^{1/3} \\ &= \frac{e^2}{3^{1/3}\Gamma(1/3)2\pi\epsilon_0 c} \left(\frac{eB \sin \alpha}{\gamma m_e}\right)^{2/3} \omega^{1/3}, \end{aligned} \quad (8.67)$$

that is, the emissivity is proportional to $v^{1/3}$.

Scheuer has presented a pleasant argument to explain the origin of this dependence (Scheuer, 1966). The expression (8.23) for the vector potential \mathbf{A} determines the intensity of the radiation field. Let us take the limit of small angles to the line of sight to the observer:

$$\begin{aligned} \mathbf{A} &= \frac{\mu_0}{4\pi r} \frac{e\mathbf{v}}{\left[1 - \frac{v}{c} \cos \theta\right]} = \frac{\mu_0}{4\pi r} \frac{e\mathbf{v}}{\left[1 - \frac{v}{c} \left(1 - \frac{\theta^2}{2}\right)\right]} \\ &= \frac{\mu_0}{4\pi r} \frac{e\mathbf{v}}{\left[\left(1 - \frac{v}{c}\right) + \frac{v\theta^2}{2c}\right]} \\ &= \frac{\mu_0}{2\pi r} \frac{e\mathbf{v}}{\left(\frac{1}{\gamma^2} + \theta^2\right)}, \end{aligned} \quad (8.68)$$

where we have used the relation $(1 - (v/c)) \approx 1/2\gamma^2$ (8.19) and set $v = c$. The radiation is strongly beamed in the forward direction, $\theta \ll \gamma^{-1}$, and is emitted at angular frequencies $\omega \sim \omega_c$. This result is associated with the fact that the electron is moving at a velocity very

close to that of light and so the first term in the denominator of (8.68) is dominant,

$$\mathbf{A} \approx \frac{\mu_0 e \gamma^2 \mathbf{v}}{2\pi r} . \quad (8.69)$$

At angles $\theta \gg \gamma^{-1}$, corresponding to Fourier components with frequencies less than v_c , the magnitude of the vector potential is determined by the angle θ rather than by how close the velocity of the electron is to that of light, that is,

$$\mathbf{A} \approx \frac{\mu_0 e \mathbf{v}}{2\pi r \theta^2} . \quad (8.70)$$

Thus, the low frequency region of the spectrum should not depend upon the precise value of the Lorentz factor γ . Another way of expressing this result is that the intensity of emission should be independent of the rest mass of the electrons responsible for the radiation. Let us therefore rewrite the expression for the total energy loss rate of synchrotron radiation in terms of the relativistic gyrofrequency of the electron ω_r and the critical frequency $v_c = (3/2)\gamma^3 v_r \sin \alpha$. Because of the exponential cut-off to the emissivity at frequencies greater than the critical frequency, the total energy loss rate of the electron can be found by integrating the spectrum from $\nu = 0$ to the critical frequency,

$$-\frac{dE}{dt} = \int_0^{\omega_c} j(\omega) d\omega = 2\sigma_T c U_{\text{mag}} \gamma^2 \sin^2 \alpha , \quad (8.71)$$

and so

$$-\frac{dE}{dt} = 2 \left(\frac{e^4}{6\pi\epsilon_0^2 c^4 m_e^2} \right) c \frac{B^2}{2\mu_0} \gamma^2 \sin^2 \alpha = \frac{e^4 c^3 B^2}{6\pi\epsilon_0} \frac{\gamma^4}{E^2} \sin^2 \alpha , \quad (8.72)$$

where $E = \gamma m_e c^2$ is the total energy of the electron. Now,

$$\omega_r = \frac{eB}{\gamma m_e} = \frac{eBc^2}{E} , \quad (8.73)$$

and hence

$$-\frac{dE}{dt} = \frac{e^2 \omega_r^2 \sin^2 \alpha}{6\pi\epsilon_0 c} \gamma^4 . \quad (8.74)$$

Substituting for γ^4 , we find

$$-\frac{dE}{dt} = \int_0^{\omega_c} j(\omega) d\omega = \left(\frac{2}{3} \right)^{4/3} \frac{e^2 (\omega_r \sin \alpha)^{2/3}}{6\pi\epsilon_0 c} \omega_c^{4/3} , \quad (8.75)$$

which depends only upon the angular gyrofrequency ω_r and ω_c . The angular gyrofrequency depends only upon the total energy of the electron rather than its mass since $\omega_r = eBc^2/E$. Therefore, we can differentiate the expression (8.75) and find

$$j(\omega) = \left(\frac{2}{3} \right)^{4/3} \frac{2e^2 (\omega_r \sin \alpha)^{2/3}}{9\pi\epsilon_0 c} \omega^{1/3} . \quad (8.76)$$

This is of exactly the same form as found above from the exact analysis, apart from a slightly different numerical constant.

8.5 The synchrotron radiation of a power-law distribution of electron energies

The next calculation is to evaluate the radiation spectrum for a distribution of electron energies. The energy spectra of cosmic rays and cosmic ray electrons can be approximated by power-law distributions and the spectra of non-thermal sources can often be represented by power-law spectra. Let us therefore work out the emission spectrum for a power-law distribution of electron energies, $N(E) dE = \kappa E^{-p} dE$, where $N(E) dE$ is the number density of electrons in the energy interval E to $E + dE$. Let us first give a simple physical picture of the origin of results, before working out the answer in more detail.

8.5.1 Physical arguments

We make use of the fact that the spectrum of synchrotron radiation is quite sharply peaked near the critical frequency ν_c (Fig. 8.8), certainly much narrower than the breadth of the power-law electron energy spectrum. In a simple approximation, it can therefore be assumed that an electron of energy E radiates away its energy at the critical frequency ν_c , which can be approximated by

$$\nu \approx \nu_c \approx \gamma^2 \nu_g = \left(\frac{E}{m_e c^2} \right)^2 \nu_g; \quad \nu_g = \frac{eB}{2\pi m_e}. \quad (8.77)$$

Therefore, the energy radiated in the frequency range ν to $\nu + d\nu$ can be attributed to electrons with energies in the range E to $E + dE$ and so

$$J(\nu) d\nu = \left(-\frac{dE}{dt} \right) N(E) dE. \quad (8.78)$$

The quantities on the right-hand side of (8.78) are:

$$E = \gamma m_e c^2 = \left(\frac{\nu}{\nu_g} \right)^{1/2} m_e c^2; \quad dE = \frac{m_e c^2}{2\nu_g^{1/2}} \nu^{-1/2} d\nu \quad (8.79)$$

$$-\left(\frac{dE}{dt} \right) = \frac{4}{3} \sigma_T c \left(\frac{E}{m_e c^2} \right)^2 \frac{B^2}{2\mu_0}. \quad (8.80)$$

Substituting into (8.78), the emissivity is expressed in terms of κ , B , ν and fundamental constants:

$$J(\nu) = (\text{constants}) \kappa B^{(p+1)/2} \nu^{-(p-1)/2}. \quad (8.81)$$

Thus, the emitted spectrum, written as $J(\nu) \propto \nu^{-a}$, where a is known as the *spectral index*, is determined by the slope of the electron energy spectrum p , rather than by the shape of the emission spectrum of a single electron. The quadratic nature of the relation between emitted frequency and the energy of the electron accounts for the difference in slopes of the emission spectrum and the electron energy spectrum, $a = (p - 1)/2$. The emissivity also depends upon the combination of quantities $\kappa B^{(p+1)/2} \propto \kappa B^{a+1}$.

8.5.2 The full analysis

We consider first a power-law distribution of electron energies at a fixed pitch angle α . We have to integrate the contributions of electrons of different energies to the intensity at angular frequency ω , or equivalently, at fixed $x = \omega/\omega_c$. Thus, at a particular frequency, we integrate over the low frequency tail of $F(x)$ for high energy electrons and over the exponential cut-off for low energy electrons. Recalling that

$$x = \frac{\omega}{\omega_c} = \frac{\omega}{(3/2)\gamma^2 \omega_g \sin \alpha} = \frac{2\omega m_e^2 c^4}{3E^2 \omega_g \sin \alpha} = \frac{A}{E^2}, \quad (8.82)$$

the emissivity per unit volume is

$$J(\omega) = \int_0^\infty j(x) \kappa E^{-p} dE. \quad (8.83)$$

From (8.82),

$$E = (A/x)^{1/2}; \quad dE = -\frac{1}{2} A^{1/2} x^{-3/2} dx, \quad (8.84)$$

and so

$$J(\omega) = \frac{\kappa}{2A^{(p-1)/2}} \int_0^\infty j(x) x^{(p-3)/2} dx = \frac{\sqrt{3}e^3 B \kappa \sin \alpha}{16\pi^2 \epsilon_0 c m_e A^{(p-1)/2}} \int_0^\infty F(x) x^{(p-3)/2} dx. \quad (8.85)$$

We can now use the integral (8.60) with $\mu = (p-3)/2$ to evaluate the integral (8.85):

$$J(\omega) = \frac{\sqrt{3}e^3 B \kappa \sin \alpha}{8\pi^2 \epsilon_0 c m_e (p+1)} \left(\frac{\omega m_e^3 c^4}{3eB \sin \alpha} \right)^{-(p-1)/2} \Gamma \left(\frac{p}{4} + \frac{19}{12} \right) \Gamma \left(\frac{p}{4} - \frac{1}{12} \right). \quad (8.86)$$

To complete the analysis we integrate over the pitch angle α . The emissivity of the electron at a particular frequency ω depends strongly upon α as shown by the relations (8.82) and (8.86). As we have discussed above, the distribution of pitch angles is likely to be isotropic and so the probability distribution of α is $\frac{1}{2} \sin \alpha d\alpha$. Using the result,

$$\frac{1}{2} \int_0^\pi \sin^{(p+3)/2} \alpha d\alpha = \frac{\sqrt{\pi}}{2} \Gamma \left(\frac{p+5}{4} \right) / \Gamma \left(\frac{p+7}{4} \right), \quad (8.87)$$

the emission per unit volume is

$$J(\omega) = \frac{\sqrt{3}e^3 B \kappa}{16\pi^2 \epsilon_0 c m_e (p+1)} \left(\frac{\omega m_e^3 c^4}{3eB} \right)^{-(p-1)/2} \times \frac{\sqrt{\pi} \Gamma \left(\frac{p}{4} + \frac{19}{12} \right) \Gamma \left(\frac{p}{4} - \frac{1}{12} \right) \Gamma \left(\frac{p}{4} + \frac{5}{4} \right)}{\Gamma \left(\frac{p}{4} + \frac{7}{4} \right)}. \quad (8.88)$$

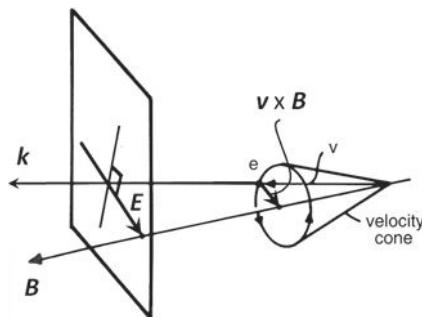


Fig. 8.9 Illustrating the geometry of the velocity cone of an ultra-relativistic electron and the polarisation of the received radiation.

We observe that the key dependences for the emissivity,

$$J(v) \propto \kappa B^{(p+1)/2} v^{-(p-1)/2} = \kappa B^{a+1} v^{-a}, \quad (8.89)$$

are the same as those which were derived by cruder methods in Sect. 8.5.1.

8.6 The polarisation of synchrotron radiation

As discussed in Sect. 8.2, the radiation of a non-relativistic electron is circularly polarised when viewed along the direction of the magnetic field lines; in general, when viewed at any angle, the radiation is elliptically polarised. In the case of relativistic electrons, however, significant radiation is only observed if the trajectory of the electron lies within an angle $1/\gamma$ of the line of sight. To understand the polarisation properties of synchrotron radiation, it is helpful to introduce the concept of *the velocity cone*, which is the cone described by the velocity vector v of the electron as it spirals about the magnetic field. The axis of the cone is the magnetic field direction and the velocity vector precesses about this direction at the relativistic gyrofrequency.

Consider first the case of those electrons with velocity cones lying precisely along the line of sight to the observer (Fig. 8.9). At the instant the electron points directly to the observer, its acceleration vector, a , is in the direction $v \times B$. The observed radiation is linearly polarised parallel to the direction $v \times B$ in the plane perpendicular to the wave vector k as indicated by the vectors k and E in Fig. 8.9. The E vector is perpendicular to the projection of B onto the plane of the sky. In fact, as we have shown in Sect. 8.4, there is also a component parallel to the magnetic field direction associated with the radiation observed when the electron is not precisely pointing towards the observer within the cone of opening angle $1/\gamma$. The radiation from a single electron is elliptically polarised because the component parallel to the field has a different time dependence within each pulse as compared with that of the perpendicular component. This is reflected in the fact that the frequency spectra of the two polarisations of synchrotron radiation are different (Fig. 8.10).

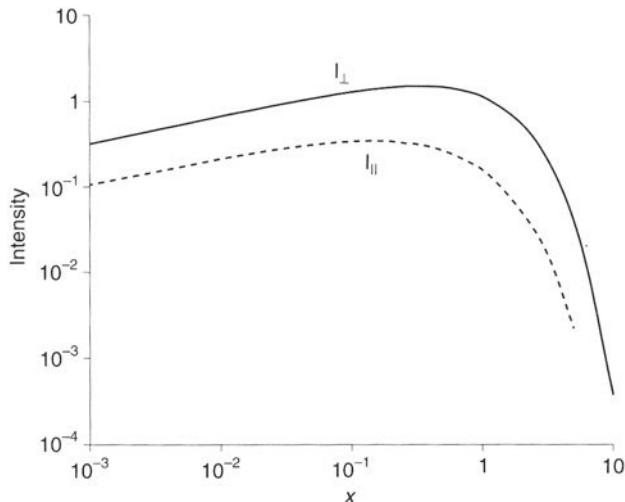


Fig. 8.10 The intensity spectra of the two polarisations I_{\perp} (solid line) and I_{\parallel} (dashed line) of the synchrotron radiation of a single high energy electron.

When there is a distribution of pitch angles, however, all the electrons with velocity cones within the angle $1/\gamma$ of the line of sight contribute to the intensity measured by the observer. These contributions are elliptically polarised in opposite senses on either side of the velocity cone. The total net polarisation is found by integrating over all electrons which contribute to the intensity and, because the angle $1/\gamma$ on either side of the line of sight is very small when the electron is ultra-relativistic, the components of elliptical polarisation parallel to the projection of \mathbf{B} cancel out and the resultant polarisation is linear. This means that we obtain the correct expression for the linearly polarised component of the radiation if we take averages of the j_{\parallel} and j_{\perp} components and neglect their time variation through the pulse.

Exact results for the linear polarisation of synchrotron radiation can be found from the formulae derived above. Consider first the emission of a single electron and work out the total amount of energy in each polarisation. From (8.56) and (8.57), we find

$$\frac{I_{\perp}}{I_{\parallel}} = \frac{\int_0^{\infty} [F(x) + G(x)] dx}{\int_0^{\infty} [F(x) - G(x)] dx}. \quad (8.90)$$

Using (18.60) and (8.61) with $\mu = 0$,

$$\frac{I_{\perp}}{I_{\parallel}} = \frac{\Gamma\left(\frac{7}{3}\right)\Gamma\left(\frac{2}{3}\right) + \Gamma\left(\frac{4}{3}\right)\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{7}{3}\right)\Gamma\left(\frac{2}{3}\right) - \Gamma\left(\frac{4}{3}\right)\Gamma\left(\frac{2}{3}\right)}. \quad (8.91)$$

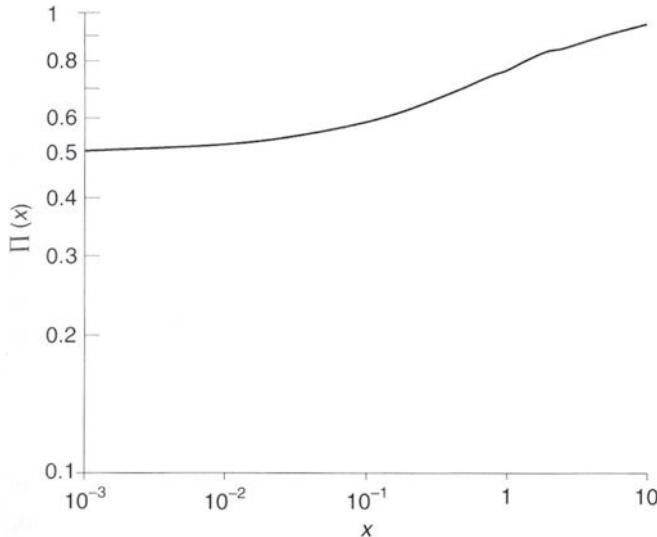


Fig. 8.11 The polarisation Π of the synchrotron radiation of a single electron as a function of frequency.

Since $\Gamma(n+1) = n\Gamma(n)$,

$$\frac{I_{\perp}}{I_{\parallel}} = \frac{\frac{4}{3} + 1}{\frac{4}{3} - 1} = 7. \quad (8.92)$$

Thus, the energy liberated in the two polarisations by a single electron is exactly in the ratio 7:1, a result derived at an early stage in his analysis by Le Roux (1961).

We have already derived the formulae necessary for working out the fractional polarisation as a function of frequency for a single electron. The fractional polarisation is defined to be

$$\Pi = \frac{I_{\perp}(\omega) - I_{\parallel}(\omega)}{I_{\perp}(\omega) + I_{\parallel}(\omega)}. \quad (8.93)$$

Inserting the expressions for the emissivities in the two polarisations given by the expressions (8.56) and (8.57), we find

$$\Pi(\omega) = \frac{G(x)}{F(x)}. \quad (8.94)$$

This function is displayed in Fig. 8.11.

The most useful result is the percentage polarisation at frequency ω for a power-law distribution of electron energies. If the electrons have energy spectrum $N(E) = \kappa E^{-p} dE$, we integrate over all energies which contribute to the intensity observed at frequency ω . Performing the same type of calculation as in Sect. 8.4, the fractional polarisation is

$$\Pi = \frac{\int_0^{\infty} G(x)x^{(p-3)/2} dx}{\int_0^{\infty} F(x)x^{(p-3)/2} dx}. \quad (8.95)$$

Using again the expressions (8.60), (8.61) and the relation $\Gamma(n+1) = n\Gamma(n)$, we find

$$\Pi = \frac{p+1}{4} \frac{\Gamma\left(\frac{p}{4} + \frac{7}{12}\right)}{\Gamma\left(\frac{p}{4} + \frac{19}{12}\right)} = \frac{p+1}{4\left(\frac{p}{4} + \frac{7}{12}\right)} = \frac{p+1}{p+\frac{7}{3}}. \quad (8.96)$$

Thus, for a typical value of the exponent of the energy spectrum of the electrons, $p = 2.5$, the fractional polarisation of synchrotron radiation is expected to be about 72%. Consequently, the synchrotron radiation of ultra-relativistic electrons in a uniform magnetic field is expected to be highly polarised.

If the electrons do not have extreme values of γ , some circular polarisation is expected because of the inexact cancellation of the elliptically polarised components on either side of the velocity cone. There are two reasons for this. Firstly, the numbers of electrons on either side of the velocity cone are different simply because of the $\sin\alpha$ factor in the expression for the solid angle contained within $d\alpha$, $d\Omega = \frac{1}{2}\sin\alpha d\alpha$. Secondly, within the cone $\theta \sim 1/\gamma$ the electrons which radiate with smaller values of α must have larger energies to radiate at frequency ω because the frequency at which most of the radiation is emitted is $\omega = \gamma^2\omega_g \sin\alpha$. Because $N(E) = \kappa E^{-p}$, different numbers of electrons radiate at frequency ω on either side of the velocity cone. These two effects mean that the cancellation of the elliptical polarisation is not exact, particularly if the values of γ are not so large. These somewhat lengthy calculations have been carried out by Legg and Westfold (1968) and Ginzburg *et al.* (1968). To order of magnitude, the fractional circular polarisation amounts to about γ^{-1} of the linear polarisation and the effect is therefore quite small. Circular polarisation has been detected from a number of compact sources of radio emission at about the 1% level and these provide independent information about the energies of the emitting electrons.

8.7 Synchrotron self-absorption

According to the principle of detailed balance, to every emission process there is a corresponding absorption process – in the case of synchrotron radiation, this is known as *synchrotron self-absorption*. Let us give a simple order-of-magnitude calculation of the basic physics of the process before working out the absorption coefficient properly.

8.7.1 Physical arguments

Suppose a source of synchrotron radiation has a power-law spectrum, $S_\nu \propto \nu^{-a}$, where the spectral index is $a = (p - 1)/2$. If the source has the same physical size at all frequencies, its *brightness temperature*, $T_b = (\lambda^2/2k)(S_\nu/\Omega)$, is proportional to $\nu^{-(2+a)}$, where S_ν is its flux density and Ω is the solid angle the source subtends at the observer (see Appendix A.7.2). We recall that the *brightness temperature* T_b is defined using the expression for the

intensity I_ν of black-body radiation

$$I_\nu = \frac{S_\nu}{\Omega} = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT_b) - 1} \approx \frac{2kT_b}{\lambda^2}, \quad (8.97)$$

in the Rayleigh–Jeans limit. T_b is a lower limit to the temperature of the region because thermodynamically no region can emit incoherent radiation with intensity greater than that of a black-body at its thermodynamic temperature. Typically, the spectra of radio sources have $a \approx 1$ and so, at low enough frequencies, the brightness temperature of the radiation may approach the ‘thermal’ temperature of the radiating electrons. When this occurs, self-absorption effects are expected to be important.

We derived the expressions for the synchrotron radiation spectrum of a power-law energy distribution of relativistic electrons, $N(E) dE = \kappa E^{-p} dE$ in Sect. 8.4. This energy spectrum is *not* a thermal equilibrium spectrum, which for relativistic electrons would be a *relativistic Maxwellian distribution*. The concept of temperature can still be used, however, for electrons of a particular energy E for the following reasons. Firstly, the spectrum of the radiation emitted by electrons of energy E is peaked about the critical frequency $\nu \approx \nu_c$ and so the emission and absorption processes at frequency ν are associated with electrons of roughly the same energy. Second, the characteristic time-scale for the relativistic electron gas to relax to an equilibrium spectrum is very long indeed under typical cosmic conditions because the electron number densities are very low and all interaction times with matter are very long. Therefore, we can associate a temperature T_e with electrons of a given energy through the relativistic formula which relates electron energy to temperature

$$\gamma m_e c^2 = 3kT_e. \quad (8.98)$$

This result follows from the fact that the ratio of specific heat capacities γ_{SH} is $4/3$ for a relativistic gas. The internal thermal energy density of a gas is $u = NkT/(\gamma_{SH} - 1)$, where N is the number density of electrons. Setting $\gamma_{SH} = 5/3$ we obtain the classical result $E = \frac{3}{2}kT_e$ and, setting $\gamma_{SH} = 4/3$, we obtain the expression (8.98) for the mean energy per electron.

As a result, the *effective temperature* T_e of the electrons now becomes a function of their energy. Since $\gamma \approx (\nu/\nu_g)^{1/2}$,

$$T_e \approx (m_e c^2 / 3k)(\nu/\nu_g)^{1/2}. \quad (8.99)$$

For a self-absorbed source, the brightness temperature of the radiation must be equal to the effective kinetic temperature of the emitting electrons, $T_b = T_e$, and therefore, in the Rayleigh–Jeans limit,

$$S_\nu = \frac{2kT_e}{\lambda^2} \Omega = \frac{2m_e}{3\nu_g^{1/2}} \Omega \nu^{5/2} \propto \frac{\theta^2 \nu^{5/2}}{B^{1/2}}, \quad (8.100)$$

where Ω is the solid angle subtended by the source $\Omega \approx \theta^2$ and θ is the angular size of the source.

This calculation illustrates the physical original of the steep low-frequency spectrum expected in sources in which synchrotron self-absorption is important, $S_\nu \propto \nu^{5/2}$. It does not follow the Rayleigh–Jeans law because the effective kinetic temperature of the electrons

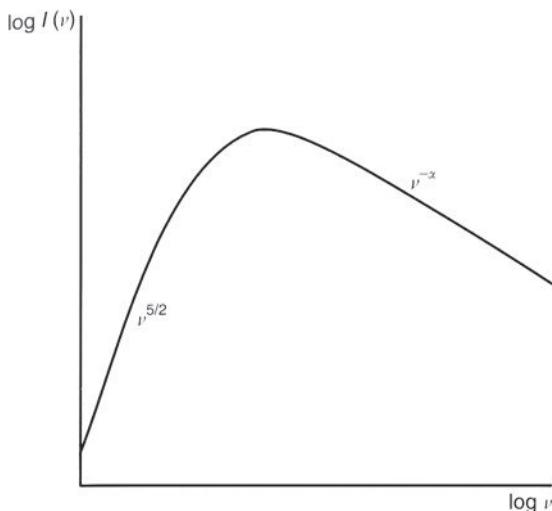


Fig. 8.12 The spectrum of a source of synchrotron radiation which exhibits the phenomenon of synchrotron self-absorption.

varies with frequency. Note also that the spectral form $S_v \propto v^{5/2}$ is independent of the spectrum of the emitting electrons so long as the magnetic field is uniform. The typical spectrum of a self-absorbed radio source is shown in Fig. 8.12.

Spectra of roughly this form are found at radio, centimetre and millimetre wavelengths in the nuclei of active galaxies and quasars. An important aspect of these observations is that they provide unambiguous evidence for the presence of relativistic electrons in the source regions. A typical set of parameters for such sources are that their angular sizes, as measured by very long baseline interferometry, are about 1 milliarcsec and their flux densities about 1 Jy at a wavelength of 6 cm. Then, the brightness temperature of the source is $T_b \approx 10^{10}$ K, a lower limit to the effective temperature of the electrons. Since $m_e c^2 / 3k = 2 \times 10^9$ K, it follows that the emitting electrons are relativistic.

8.7.2 The absorption coefficient for synchrotron self-absorption

The simplest way of working out the absorption coefficient for synchrotron self-absorption is to regard the emission of a photon of energy $h\nu$ as originating in a two-level system in which the electron makes a transition from a state with energy E and momentum \mathbf{p} (level 2) to one with energy $E' = E - dE$ and momentum $\mathbf{p}' = \mathbf{p} - d\mathbf{p}$ (level 1). We have already worked out classically the emission coefficient for this process (8.58) and hence the spontaneous transition probability which describes the rate of emission of photons in the frequency interval ν to $\nu + d\nu$ is

$$A_{21} = \frac{j(\nu, E)}{h\nu} \quad \text{photons Hz}^{-1} \text{ s}^{-1}, \quad (8.101)$$

where $j(\nu)$ is now the emissivity per unit frequency interval rather than per unit angular frequency, that is, $j(\nu, E) = 2\pi j(\omega, E)$. This expression contains no information about the

directional properties of the radiation. We will work out the absorption coefficient assuming the radiation is emitted isotropically which would be the case if the magnetic field in the source region were chaotic. There are complexities in a more complete calculation which are discussed by Ginzburg and Syrovatskii (1969).

The Einstein coefficients for absorption and spontaneous and induced emission (6.59) are

$$A_{21} = \frac{2h\nu^3}{c^2} B_{12} = \frac{2h\nu^3}{c^2} B_{21}. \quad (8.102)$$

These coefficients are defined in terms of the number density $n(\mathbf{p})$ of electrons per unit volume of phase space $d^3 p$, rather than per unit energy interval. The absorption coefficient is then given by the expression involving the Einstein coefficients but now for pairs of states separated in momentum by $d\mathbf{p} = (h\nu/c) \mathbf{i}_k$. For a particular pair of states, the absorption coefficient according to (6.62) is

$$\chi_\nu = \frac{h\nu}{4\pi} [n(\mathbf{p} - \hbar\mathbf{k})B_{12} d^3 p - n(\mathbf{p})B_{21} d^3 p]. \quad (8.103)$$

Making a Taylor expansion for small values of $h\nu/c$,

$$n(\mathbf{p} - \hbar\mathbf{k}) = n(\mathbf{p}) - \frac{h\nu}{c} \frac{dn}{dp} \quad \text{and so} \quad \chi_\nu = -\frac{h^2\nu^2}{4\pi c} B_{12} \frac{dn}{dp} d^3 p. \quad (8.104)$$

This result is integrated over all possible pairs of electron momenta which could be involved in the absorption process. Assuming an isotropic electron distribution in momentum space,

$$\begin{aligned} \chi_\nu &= - \int_0^\infty \frac{h^2\nu^2}{4\pi c} B_{12} \frac{dn}{dp} 4\pi p^2 dp = -\frac{hc}{2\nu} \int_0^\infty A_{21} \frac{dn}{dp} p^2 dp \\ &= -\frac{c}{2\nu^2} \int_0^\infty j(\nu, E) \frac{dn}{dp} p^2 dp. \end{aligned} \quad (8.105)$$

Now convert the electron momentum spectrum into an electron energy spectrum

$$p = E/c; \quad dp = dE/c. \quad (8.106)$$

Therefore,

$$4\pi p^2 n(p) dp = N(E) dE; \quad n(p) = \frac{c^3}{4\pi} \frac{N(E)}{E^2}. \quad (8.107)$$

and so the absorption coefficient χ_ν becomes

$$\chi_\nu = -\frac{c^2}{8\pi\nu^2} \int_0^\infty j(\nu, E) \frac{d}{dE} \left(\frac{N(E)}{E^2} \right) E^2 dE. \quad (8.108)$$

For a power-law distribution of electron energies, $N(E) = \kappa E^{-p}$,

$$\chi_\nu = \frac{(p+2)\kappa c^2}{8\pi\nu^2} \int_0^\infty j(\nu, E) E^{-(p+1)} dE. \quad (8.109)$$

Inserting the expression for $j(\nu)$ from (8.58),

$$\chi_\nu = \frac{\sqrt{3}e^3 B c \kappa \sin \alpha}{32\pi^2 \epsilon_0 m_e \nu^2} (p+2) \int_0^\infty F(x) E^{-(p+1)} dx . \quad (8.110)$$

Using the integral (8.60), we find

$$\int_0^\infty F(x) E^{-(p+1)} dx = \frac{1}{(p+2)} \left(\frac{A}{2}\right)^{-p/2} \Gamma\left(\frac{3p+22}{12}\right) \Gamma\left(\frac{3p+2}{12}\right) , \quad (8.111)$$

where we have set $x = \nu/\nu_c = \nu / (\frac{3}{2}\gamma^2 v_g \sin \alpha) = A/E^2$. Thus, the expression for the absorption coefficient is

$$\chi_\nu = \frac{\sqrt{3}e^3 \kappa c}{32\pi^2 \epsilon_0 m_e} \left(\frac{3e}{2\pi m_e^3 c^4}\right)^{p/2} \Gamma\left(\frac{3p+22}{12}\right) \Gamma\left(\frac{3p+2}{12}\right) (B \sin \alpha)^{(p+2)/2} \nu^{-(p+4)/2} . \quad (8.112)$$

For a randomly oriented magnetic field, we average over a random distribution of angles α , $p(\alpha) d\alpha = \frac{1}{2} \sin \alpha d\alpha$, and hence have to evaluate

$$\int_0^\infty \frac{1}{2} \sin \alpha \sin^{(p+2)/2} \alpha d\alpha = \frac{\sqrt{\pi}}{2} \Gamma\left(\frac{p+6}{4}\right) / \Gamma\left(\frac{p+8}{4}\right) . \quad (8.113)$$

Therefore, the absorption coefficient for synchrotron radiation in a randomly oriented magnetic field is

$$\chi_\nu = \frac{\sqrt{3\pi} e^3 \kappa B^{(p+2)/2} c}{64\pi^2 \epsilon_0 m_e} \left(\frac{3e}{2\pi m_e^3 c^4}\right)^{p/2} \frac{\Gamma\left(\frac{3p+22}{12}\right) \Gamma\left(\frac{3p+2}{12}\right) \Gamma\left(\frac{p+6}{4}\right)}{\Gamma\left(\frac{p+8}{4}\right)} \nu^{-(p+4)/2} . \quad (8.114)$$

Let us now apply this result to the emission spectrum of a region of thickness l . The transfer equation for radiation (6.50) is

$$\frac{dI_\nu}{dx} = -\chi_\nu I_\nu + \frac{J(\nu)}{4\pi} . \quad (8.115)$$

The solution is

$$I_\nu = \frac{J(\nu)}{4\pi \chi_\nu} [1 - e^{-\chi_\nu l}] . \quad (8.116)$$

If the source is optically thin, $\chi(\nu)l \ll 1$,

$$I_\nu = \frac{J(\nu)l}{4\pi} . \quad (8.117)$$

If the source is optically thick, $\chi(\nu)l \gg 1$,

$$I_\nu = \frac{J(\nu)}{4\pi \chi_\nu} . \quad (8.118)$$

The quantity $J(\nu)/4\pi\chi_\nu$ is often referred to as the *source function*. Substituting (8.114) for the absorption coefficient $\chi(\nu)$ and (18.88) for J_ν into (8.118), we find

$$I_\nu = (\text{constant}) \frac{m_e v^{5/2}}{\nu_g^{1/2}}, \quad (8.119)$$

where the constant is a number of order unity which involves numerous gamma functions. This is the same dependence as was found from our physical arguments in (8.100).

In a more complete analysis, we would work out separately the absorption coefficients in the two polarisations which are found to be different (Ginzburg and Syrovatskii, 1969). In the optically thick region, the electric vector of the emitted radiation is parallel, rather than perpendicular, to the magnetic field direction and the degree of polarisation is

$$\Pi = \left| \frac{I_\perp - I_\parallel}{I_\perp + I_\parallel} \right| = \frac{3}{6p + 13}, \quad (8.120)$$

for a uniform field.

8.8 Useful numerical results

It is convenient to have at hand a set of numerical results for the various relations derived in the preceding sections. The total energy loss rate by synchrotron radiation is

$$-\left(\frac{dE}{dt} \right) = 2\sigma_T c U_{\text{mag}} \gamma^2 \left(\frac{v}{c} \right)^2 \sin^2 \theta, \quad (8.121)$$

and can be written

$$-\left(\frac{dE}{dt} \right) = 1.587 \times 10^{-14} B^2 \gamma^2 \left(\frac{v}{c} \right)^2 \sin^2 \theta \quad \text{W} \quad (8.122)$$

where the units of magnetic flux density B are tesla and γ is the Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$. When averaged over an isotropic distribution of pitch angles θ , the result is

$$-\left(\frac{dE}{dt} \right) = \frac{4}{3} \sigma_T c U_{\text{mag}} \gamma^2 \left(\frac{v}{c} \right)^2, \quad (8.123)$$

which can be written

$$-\left(\frac{dE}{dt} \right) = 1.058 \times 10^{-14} B^2 \gamma^2 \left(\frac{v}{c} \right)^2 \quad \text{W}. \quad (8.124)$$

The emission spectrum of a single electron is

$$j(\nu) = 2\pi j(\omega) = \frac{\sqrt{3}e^3 B \sin \alpha}{4\pi \epsilon_0 c m_e} F \left(\frac{\nu}{\nu_c} \right), \quad (8.125)$$

Table 8.2 Constants for use with the synchrotron radiation formulae.

p	$a(p)$	$b(p)$
1	2.056	0.397
1.5	0.909	0.314
2	0.529	0.269
2.5	0.359	0.244
3	0.269	0.233
3.5	0.217	0.230
4	0.186	0.236
4.5	0.167	0.248
5	0.157	0.268

which becomes

$$j(v) = 2.344 \times 10^{-25} B \sin \alpha F\left(\frac{v}{v_c}\right) \text{ W Hz}^{-1}, \quad (8.126)$$

where again B is expressed in tesla and the function $F(v/v_c)$ is given in Table 8.1. The critical frequency v_c is given by

$$v_c = \left(\frac{3}{2}\right) \gamma^2 \frac{eB}{2\pi m_e} = 4.199 \times 10^{10} \gamma^2 B \text{ Hz}, \quad (8.127)$$

where B is measured in tesla.

The radiation spectrum of a power-law electron energy distribution $N(E) = \kappa E^{-p}$ in the case of a random magnetic field is

$$J(v) = 2\pi J(\omega) = \frac{\sqrt{3}e^3 B \kappa}{4\pi \epsilon_0 c m_e} \left(\frac{3eB}{2\pi v m_e^3 c^4}\right)^{(p-1)/2} a(p), \quad (8.128)$$

where

$$a(p) = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{p}{4} + \frac{19}{12}\right) \Gamma\left(\frac{p}{4} - \frac{1}{12}\right) \Gamma\left(\frac{p}{4} + \frac{5}{4}\right)}{(p+1) \Gamma\left(\frac{p}{4} + \frac{7}{4}\right)}. \quad (8.129)$$

In SI units, this becomes

$$J(v) = 2.344 \times 10^{-25} a(p) B^{(p+1)/2} \kappa \left(\frac{1.253 \times 10^{37}}{v}\right)^{(p-1)/2} \text{ W m}^{-3} \text{ Hz}^{-1}. \quad (8.130)$$

The constant $a(p)$ depends upon the energy spectral index p , and appropriate values of $a(p)$ are given in Table 8.2. This relation is only useful for those who wish to write the energy of the electrons in joules, that is, the energy spectrum $N(E)$ represents the number

density of electrons per joule. This is highly non-standard. If the energies of the electrons are measured in GeV and the units of $N(E)$ are electrons m^{-3} GeV^{-1} , the result is

$$J(\nu) = 2.344 \times 10^{-25} a(p) B^{(p+1)/2} \kappa' \left(\frac{3.217 \times 10^{17}}{\nu} \right)^{(p-1)/2} \text{W m}^{-3} \text{Hz}^{-1}. \quad (8.131)$$

Finally, the absorption coefficient χ_ν for a random magnetic field is

$$\chi_\nu = \frac{\sqrt{3}e^3 c}{8\pi^2 \epsilon_0 m_e} \kappa B^{(p+2)/2} \left(\frac{3e}{2\pi m_e^3 c^4} \right)^{p/2} b(p) \nu^{-(p+4)/2}, \quad (8.132)$$

$$b(p) = \frac{\sqrt{\pi}}{8} \frac{\Gamma\left(\frac{3p+22}{12}\right) \Gamma\left(\frac{3p+2}{12}\right) \Gamma\left(\frac{p+6}{4}\right)}{\Gamma\left(\frac{p+8}{4}\right)}. \quad (8.133)$$

In SI units, the value of χ_ν is

$$\chi_\nu = 3.354 \times 10^{-9} \kappa B^{(p+2)/2} (3.54 \times 10^{18})^p b(p) \nu^{-(p+4)/2} \text{m}^{-1}, \quad (8.134)$$

where the constant $b(p)$ depends upon the exponent p as listed in Table 8.2. In this version, the energies of the electrons are expressed in joules. If, instead, $N(E)$ is expressed in electrons m^{-3} GeV^{-1} , the expression becomes

$$\chi_\nu = 20.9 \kappa' B^{(p+2)/2} (5.67 \times 10^9)^p b(p) \nu^{-(p+4)/2} \text{m}^{-1}. \quad (8.135)$$

8.9 The radio emission of the Galaxy

The theory of synchrotron radiation in its astrophysical context can be tested by studying the intensity and spectrum of the Galactic radio emission. The radio map of the sky at a frequency of 408 MHz is shown in Fig. 1.9 where it can be seen that there is a ‘radio disc’ similar, in general terms, to the optical disc of the Galaxy. In addition, there are various ‘loops’ which extend out of the Galactic plane, the most prominent being the feature known as the North Polar Spur which originates at $l = 30^\circ$ and extends toward the Galactic north pole.

The determination of the Galactic radio spectrum and the radio emissivity of the interstellar medium are difficult observational problems because the Galactic radio emission extends over the whole sky and so, even in directions far away from that in which the telescope is pointing, some radiation creeps into the receiver through far-out side-lobes of the telescope beam. The best observations of the background spectrum are made with geometrically scaled aerials so that the reception pattern is identical at different wavelengths.

The spectra of the Galactic radio emission in the direction of the north Galactic pole and in the anti-Centre direction are shown in Fig. 8.13. At frequencies less than about 200 MHz, the spectrum can be described by a power law of the form $I(\nu) \propto \nu^{-0.4}$; at frequencies greater than about 400 MHz, the spectrum steepens, the spectral index being about 0.8–0.9

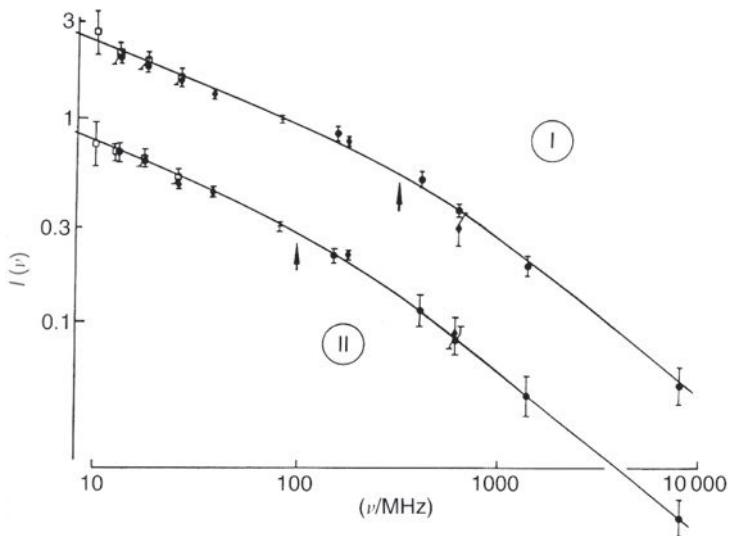


Fig. 8.13 The spectrum of the Galactic radio emission. Region I corresponds to the anti-Centre direction at high Galactic latitudes while region II corresponds to the interarm region (Webster, 1971, 1974).

(Webster, 1971). This spectrum can be compared with the predicted spectrum if the energy spectrum of the cosmic ray electrons observed at the top of the atmosphere is assumed to be representative of the local interstellar medium.

At energies greater than 10 GeV, at which the effects of solar modulation should not be significant, the electron spectrum can be well represented by a power law of differential form

$$dN = N(E) dE = 700 E^{-3.3} dE \quad \text{electrons m}^{-2} \text{s}^{-1} \text{sr}^{-1} \quad (8.136)$$

where the energy E is measured in GeV (Webber, 1983). Converting this spectrum into number density of electrons,

$$dn = n(E) dE = \frac{4\pi}{c} \frac{dN}{dE} = 2.9 \times 10^{-5} E^{-3.3} dE \quad \text{electrons m}^{-3}. \quad (8.137)$$

Let us assume that this spectrum is representative of that of ultra-relativistic electrons in local interstellar space.

Electrons of energy $E = \gamma m_e c^2$ radiate most of their energy at a frequency $\nu \approx 28\gamma^2 B$ GHz where B is measured in tesla. Let us suppose that the average local magnetic flux density in the Galaxy is $B = 3 \times 10^{-10} x$ T. Then, 10 GeV electrons radiate most of their energy at a frequency $\nu \approx 3.2x$ GHz. Unfortunately, the frequency range over which the electron energy spectrum is free of the effects of solar modulation is just outside the range over which the Galactic radio spectrum has been accurately measured.

The next problem is to work out the local synchrotron emissivity of the interstellar medium. There are two alternatives. One approach is to estimate the local thickness of the Galactic disc of radio emission. The problem here is that there are uncertainties about

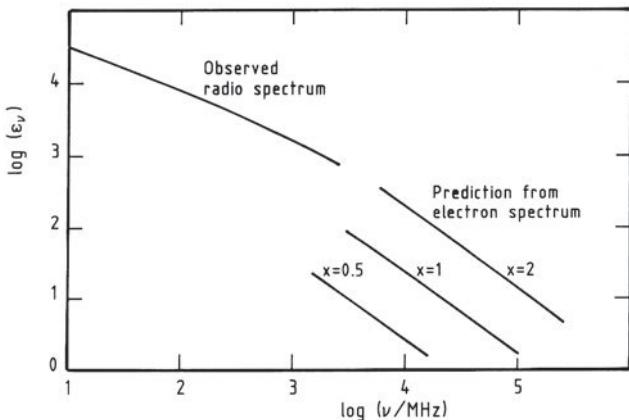


Fig. 8.14

Comparison of the observed radio emissivity of the interstellar medium with that expected from the local electron energy spectrum for different values of the magnetic field strength. The radio emissivity is shown in relative units. The adopted radio emissivity at 10 MHz is $3 \times 10^{-39} \text{ W m}^{-3} \text{ Hz}^{-1}$.

the exact thickness of the radio disc in our vicinity in the Galaxy. The intensity of the Galactic radiation in the direction of the Galactic pole at a frequency of 10 MHz is $10^{-20} \text{ W m}^2 \text{ sr}^{-1} \text{ Hz}^{-1}$ (Webber, 1983). If the half-thickness of the disc is taken to be 1 kpc, the corresponding volume emissivity is $4.2 \times 10^{-39} \text{ W m}^{-3} \text{ Hz}^{-1}$. A second approach is to make observations at very low radio frequencies at which regions of ionised hydrogen of large angular size are optically thick because of thermal bremsstrahlung absorption. Then the radio emission in the direction of the opaque cloud must originate in the interstellar medium between the cloud and the Earth. Caswell analysed his 10 MHz map of the Galactic radio emission in the direction of such clouds and found an average brightness temperature of $T_b = 240 \text{ K pc}^{-1}$ at 10 MHz, corresponding to a volume emissivity of $3 \times 10^{-39} \text{ W m}^{-3} \text{ Hz}^{-1}$ (Caswell, 1976). We will adopt this value and the Galactic radio spectrum has been normalised to it in Fig. 8.14.

We can now enter the synchrotron radiation formula (8.130) with the electron energy spectrum (8.137) so that $\kappa' = 2.9 \times 10^{-5} \text{ electrons m}^{-3} \text{ GeV}^{-(1-p)}$ and $p = 3.3$ for which $a(p) = 0.238$ (see Table 8.2). In Fig. 8.14, the predicted spectrum has been evaluated for magnetic field strengths $B = 0.15, 0.3$ and 0.6 nT , that is, $x = 0.5, 1$ and 2 .

The predicted spectrum of the radio emission joins smoothly onto the observed spectrum of the Galactic radio emission, provided it is assumed that the magnetic field strength is high, $B = 6 \times 10^{-10} \text{ T}$. The mean value of the magnetic field strength required to achieve this agreement is larger than the typical values assumed for the average interstellar magnetic field as derived from pulsar rotation measures – these are found to lie in the range $(1.5-3) \times 10^{-10} \text{ T}$. There are various possible explanations for this discrepancy. It might be that the Earth is located within a region of low relativistic electron density relative to the general interstellar medium. Also, the intensity of the Galactic radio emission depends upon the magnetic flux density as $B^{(p+1)/2} \propto B^{2.14}$ and hence, if the relativistic electron density were uniform, the intensity of emission along the line of sight is weighted as $\int B^{2.14} dl$,

whereas the magnetic field strength derived from pulsar rotation measures is weighted as $\int B_{\parallel} N_e dl / \int N_e dl$. Thus, the intensity of synchrotron radiation gives greater weight to regions of high magnetic field. Nonetheless, it is encouraging that the observed intensity is within roughly a factor of 2 of what might be reasonably expected, given the difficulty of establishing exact values for the local relativistic electron spectrum and radio emissivity, and thus one can assume with some confidence that the Galactic radio emission is indeed synchrotron radiation.

The three main processes involved in the interaction of high energy photons with atoms, nuclei and electrons are photoelectric absorption, Compton scattering and electron–positron pair production. These processes are important not only in the study of high energy astrophysical phenomena in a wide variety of different circumstances but also in the detection of high energy particles and photons. For example, photoelectric absorption is observed in the spectra of most X-ray sources at energies $\varepsilon \lesssim 1$ keV. Thomson and Compton scattering appear in a myriad of guises from the processes occurring in stellar interiors, to the spectra of binary X-ray sources, and inverse Compton scattering figures prominently in sources in which there are intense radiation fields and high energy electrons. Pair production is bound to occur wherever there are significant fluxes of high energy γ -rays – evidence for the production of positrons by this process is provided by the detection of the 511 keV electron–positron annihilation line in our own Galaxy.

9.1 Photoelectric absorption

At low photon energies, $\hbar\omega \ll m_e c^2$, the dominant process by which photons interact with matter is photoelectric, or bound–free, absorption and is one of the principal sources of opacity in stellar interiors. We are principally interested here in the process in somewhat more rarefied plasmas. If the energies of the incident photons $\varepsilon = \hbar\omega$ are greater than the energy of the X-ray atomic energy level E_1 , an electron can be ejected from that level, the remaining energy ($\hbar\omega - E_1$) being carried away as the kinetic energy of the ejected electron, the *photoelectric effect*. The photon energy at which $\hbar\omega = E_1$ corresponds to an *absorption edge* in the spectrum of the radiation because ejection of electrons from this energy level is impossible if the photons are of lower energy. For photons with higher energies, the cross-section for photoelectric absorption from this level decreases as roughly ω^{-3} . Examples of the absorption cross-sections for a number of common elements are shown in Fig. 9.1 and the X-ray atomic energy levels of atoms up to iron are listed in Table 9.1.

The evaluation of these cross-sections is one of the standard calculations in the quantum theory of radiation (Heitler, 1954). For example, the analytic solution for the absorption cross-section for photons with energies $\hbar\omega \gg E_1$ and $\hbar\omega \ll m_e c^2$ due to the ejection of electrons from the K-shells of atoms, that is, from the 1s level, is

$$\sigma_K = 4\sqrt{2}\sigma_T\alpha^4 Z^5 \left(\frac{m_e c^2}{\hbar\omega}\right)^{7/2} = \frac{e^{12} m_e^{3/2} Z^5}{192\sqrt{2}\pi^5 \epsilon_0^6 \hbar^4 c} \left(\frac{1}{\hbar\omega}\right)^{7/2}, \quad (9.1)$$

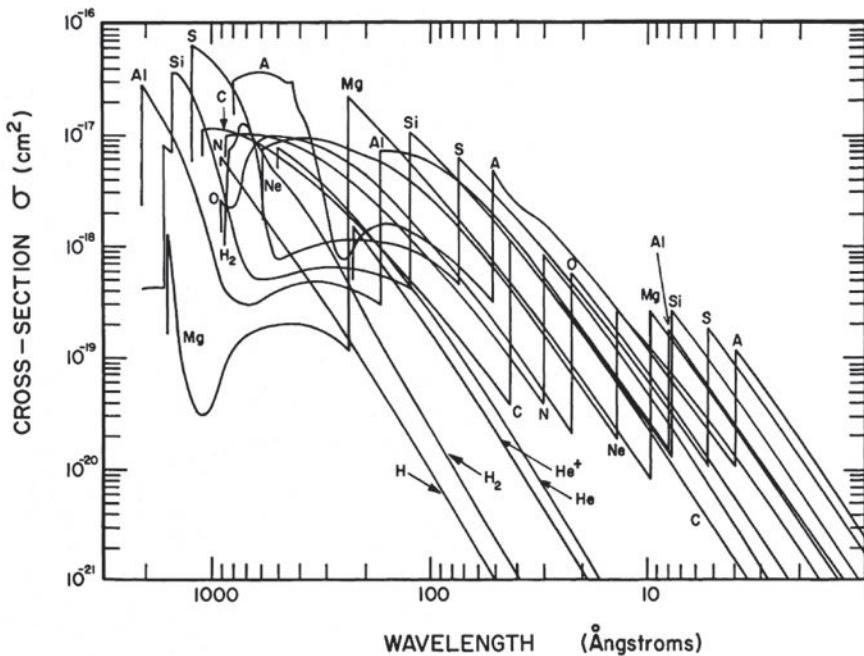


Fig. 9.1 Photoabsorption cross-sections of the abundant elements in the interstellar medium as a function of wavelength (Cruddace *et al.*, 1974).

where $\alpha = e^2/4\pi\epsilon_0\hbar c$ is the fine structure constant and $\sigma_T = 8\pi r_e^2/3 = e^2/6\pi\epsilon_0^2 m_e^2 c^4$ the Thomson cross-section. This cross-section takes account of the fact that there are 2 K-shell electrons in all elements except hydrogen, both 1s electrons contributing to the opacity of the material. The absorption cross-section has a strong dependence upon the atomic number Z and so, although heavy elements are very much less abundant than hydrogen, the combination of the ω^{-3} dependence and the fifth-power dependence upon Z means that quite rare elements can make significant contributions to the total absorption cross-section at ultraviolet and X-ray energies. More detailed calculations of these cross-sections with appropriate Gaunt factors are given by Karzas and Latter (1961).

These data enable the X-ray absorption coefficient for interstellar matter to be determined. Absorption cross-sections of the forms shown in Fig. 9.1 are summed, weighted by the cosmic abundance of the different elements,

$$\sigma_e(\varepsilon) = \frac{1}{n_H} \sum_i n_i \sigma_i(\varepsilon). \quad (9.2)$$

In this computation, the K-edges, corresponding to the ejection of electrons from the 1s shell of the atom or ion, provide the dominant source of opacity. The resulting total absorption coefficient for X-rays, assuming the standard cosmic abundances of the chemical elements, is shown in Fig. 9.2, the K-edges of different elements being indicated. In low resolution X-ray spectral studies, these edges cannot be resolved individually as distinct features and a useful linear interpolation formula for the X-ray absorption coefficient, σ_e , and the

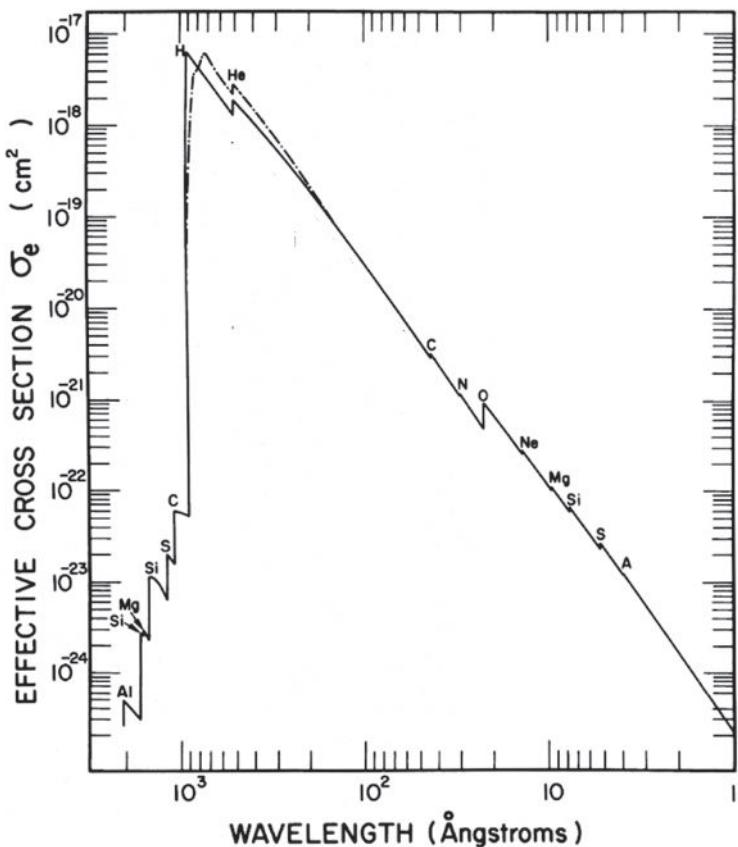
Table 9.1 The X-ray atomic energy levels for elements up to iron (Bearden and Burr, 1967).

Element	Energies in electron-volts (eV) X-ray term						
	K	L I	L II	L III	M I	M II,III	M IV,V
Hydrogen	13.598						
Helium	24.587						
Lithium	54.75						
Beryllium	111.0						
Boron	188.0		4.7				
Carbon	283.8		6.4				
Nitrogen	401.6		9.2				
Oxygen	532.0	23.7	7.1				
Fluorine	685.4	31	8.6				
Neon	866.9	45	18.3				
Sodium	1072.1	63.3	31.1				
Magnesium	1305.0	89.4	51.4				
Aluminium	1559.6	117.7	73.1				
Silicon	1838.9	148.7	99.2				
Phosphorus	2145.5	189.3	132.2				
Sulphur	2472.0	229.2	164.8				
Chlorine	2822.4	270.2	201.6	200.0	17.5	6.8	
Argon	3202.9	320	247.3	245.2	25.3	12.4	
Potassium	3607.4	377.1	296.3	293.6	33.9	17.8	
Calcium	4038.1	437.8	350.0	346.4	43.7	25.4	
Scandium	4492.8	500.4	406.7	402.2	53.8	32.3	6.6
Titanium	4966.4	563.7	461.5	455.5	60.3	34.6	3.7
Vanadium	5465.1	628.2	520.5	512.9	66.5	37.8	2.2
Chromium	5989.2	694.6	583.7	574.5	74.1	42.5	2.3
Manganese	6539.0	769.0	651.4	640.3	83.9	48.6	3.3
Iron	7112.0	846.1	721.1	708.1	92.9	54.0	3.6

corresponding optical depth, τ_e is

$$\tau_e(\hbar\omega) = \int \sigma_e N_H dl = 2 \times 10^{-26} \left(\frac{\hbar\omega}{1 \text{ keV}} \right)^{-8/3} \int N_H dl , \quad (9.3)$$

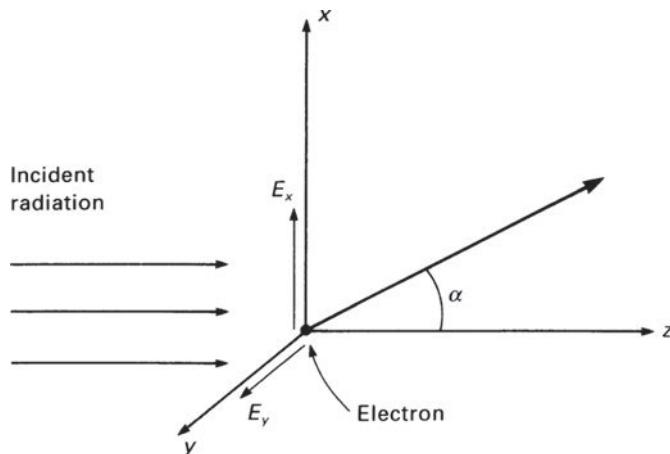
where the *column depth* $\int N_H dl$ is expressed in particles per square metre and N_H is the number density of hydrogen atoms in particles per cubic metre. For example, if the interstellar gas density were 10^6 hydrogen atoms m^{-3} , the optical depth of the medium is roughly unity for a path length of 1 kpc at 1 keV. Thus, the spectra of many X-ray sources turn over at about 1 keV because of interstellar photoelectric absorption. Because of the steep energy dependence of τ_e , photoelectric absorption is only important at energies $\hbar\omega \gg 1 \text{ keV}$ for sources with large column densities of matter between the source and the observer.

**Fig. 9.2**

The effective absorption cross-section per hydrogen atom for interstellar gas with typical cosmic abundances of the chemical elements. The solid line is for the gaseous component of the interstellar medium; the dot-dashed line includes molecular hydrogen. The discontinuities in the absorption cross-section as a function of energy are associated with the K-shell absorption edges of the elements indicated. The optical depth of the medium is $\tau_e = \int \sigma_e(\varepsilon) N_H d\varepsilon$ where N_H is the number density of hydrogen atoms (Cruddace *et al.*, 1974). Note that the cross-section is presented in units of cm^2 . For reference, $1 \text{ \AA} \equiv 12.4 \text{ keV}$ and $100 \text{ \AA} \equiv 0.124 \text{ keV}$.

9.2 Thomson and Compton scattering

In 1923, Compton discovered that the wavelength of hard X-ray radiation increases when it is scattered by stationary electrons (Compton, 1923). This was definitive proof of Einstein's quantum picture of the nature of light according to which it may be considered to possess both wave-like and particle-like properties (Einstein, 1905). In the Compton scattering process, the incoming high energy photons collide with stationary electrons and transfer some of their energy and momentum to the electrons. Consequently, the scattered photons have less energies and momenta than before the collisions. Since the energy and momentum of the photon is proportional to frequency, $E = \hbar\omega$ and $\mathbf{p} = (\hbar\omega/c)\mathbf{i}_k$, where \mathbf{i}_k is the unit

**Fig. 9.3**

Illustrating the geometry of the Thomson scattering of a beam of radiation by a free electron.

vector in the direction of travel of the photon, the loss of energy of the photon corresponds to an increase in its wavelength. We begin with the simpler process of *Thomson scattering* in which the photons, or electromagnetic waves, are scattered without change of energy.

9.2.1 Thomson scattering

Thomson first published the formula for what is now called the Thomson cross-section in 1906 (Thomson, 1906) and used his result to show that the number of electrons in each atom is of the same order as the element's atomic number. He used Larmor's formula, which we derived using Thomson's methods in Sect. 6.2.2.

We can carry out a completely classical analysis of the scattering of an unpolarised parallel beam of radiation through an angle α by a stationary electron using the radiation formula (6.6). It is assumed that the incident beam propagates in the positive z -direction (Fig. 9.3) and, without loss of generality, we can arrange the geometry of the scattering to be such that the scattering angle α lies in the x - z plane. The electric field strength of the unpolarised incident field is resolved into components of equal intensity with electric vectors in the orthogonal i_x and i_y directions (Fig. 9.3). The electric fields experienced by the electron in the x and y directions, $E_x = E_{x0} \exp(i\omega t)$ and $E_y = E_{y0} \exp(i\omega t)$, respectively, cause the electron to oscillate and the accelerations in these directions are

$$\ddot{r}_x = eE_x/m_e; \quad \ddot{r}_y = eE_y/m_e. \quad (9.4)$$

We can therefore enter these accelerations into the radiation formula (6.6), which describes the angular dependence of the emitted intensity upon the polar angle θ .

Treating first the x -acceleration, (6.6) can be used with the substitution $\alpha = \pi/2 - \theta$. The intensity of radiation scattered through angle θ into the solid angle $d\Omega$ is then

$$-\left(\frac{dE}{dt}\right)_x d\Omega = \frac{e^2 |\ddot{r}_x|^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3} d\Omega = \frac{e^4 |E_x|^2}{16\pi^2 m_e^2 \epsilon_0 c^3} \cos^2 \alpha d\Omega. \quad (9.5)$$

Taking time averages of E_x^2 , $\overline{E_x^2} = E_{x0}^2/2$. We sum over all waves contributing to the E_x -component of radiation and express the result in terms of the incident energy per unit area upon the electron. The latter is given by Poynting's theorem, $S_x = (\mathbf{E} \times \mathbf{H}) = c\epsilon_0 E_x^2 \mathbf{i}_z$. Since the radiation is incoherent, we sum over all the time-averaged waves to find that the total intensity in the direction α from the x -component of the acceleration is $S_x = \sum_i c\epsilon_0 E_{x0}^2/2$, and so

$$-\left(\frac{dE}{dt}\right)_x d\Omega = \frac{e^4 \cos^2 \alpha}{16\pi^2 m_e^2 \epsilon_0 c^3} \sum_i \overline{E_x^2} d\Omega = \frac{e^4 \cos^2 \alpha}{16\pi^2 m_e^2 \epsilon_0^2 c^4} S_x d\Omega. \quad (9.6)$$

Next consider scattering of the E_y -component of the incident field. From the geometry of Fig. 9.3, the radiation in the $x-z$ plane due to the acceleration of the electron in the y -direction corresponds to scattering through $\theta = 90^\circ$ and therefore the scattered intensity in the α -direction is

$$-\left(\frac{dE}{dt}\right)_y d\Omega = \frac{e^4}{16\pi^2 m_e^2 \epsilon_0^2 c^4} S_y d\Omega. \quad (9.7)$$

The total scattered radiation into $d\Omega$ is found by adding the intensities of the two independent field components,

$$-\left(\frac{dE}{dt}\right) d\Omega = \frac{e^4}{16\pi^2 m_e^2 \epsilon_0^2 c^4} (1 + \cos^2 \alpha) \frac{S}{2} d\Omega, \quad (9.8)$$

where $S = S_x + S_y$ and we recall that $S_x = S_y$ for unpolarised radiation. We now express the scattered intensity in terms of a differential scattering cross-section $d\sigma_T$ in direction α by the following relation,

$$\frac{d\sigma_T(\alpha)}{d\Omega} = \frac{\text{energy radiated per unit time per unit solid angle}}{\text{incident energy per unit time per unit area}}. \quad (9.9)$$

Since the total incident energy per unit time per unit area is S , the differential cross-section for Thomson scattering is

$$d\sigma_T = \frac{e^4}{16\pi^2 \epsilon_0^2 m_e^2 c^4} \frac{(1 + \cos^2 \alpha)}{2} d\Omega = \frac{3\sigma_T}{16\pi} (1 + \cos^2 \alpha) d\Omega, \quad (9.10)$$

which can be expressed in terms of the classical electron radius $r_e = e^2/4\pi\epsilon_0 m_e c^2$,

$$d\sigma_T = \frac{r_e^2}{2} (1 + \cos^2 \alpha) d\Omega. \quad (9.11)$$

To find the total cross-section for scattering, we integrate over all solid angles,

$$\sigma_T = \int_0^\pi \frac{r_e^2}{2} (1 + \cos^2 \alpha) 2\pi \sin \alpha d\alpha = \frac{8\pi}{3} r_e^2 = \frac{e^4}{6\pi \epsilon_0^2 m_e^2 c^4} = 6.653 \times 10^{-29} \text{ m}^2. \quad (9.12)$$

This is Thomson's famous result for the total cross-section for scattering of electromagnetic waves by stationary free electrons. It will reappear in many different guises in the course of the exposition. Let us note some of the important features of Thomson scattering.

- (i) The scattering is symmetric with respect to the scattering angle α . Thus, as much radiation is scattered in the backward as in the forward direction.
- (ii) The scattering cross-section for 100% polarised emission can be found by integrating the scattered intensity (9.5) over all angles,

$$-\left(\frac{dE}{dt}\right)_x = \frac{e^2 |\ddot{r}_x|^2}{16\pi^2 \epsilon_0 c^3} \int \sin^2 \theta \, 2\pi \sin \theta \, d\theta = \left(\frac{e^4}{6\pi \epsilon_0^2 m_e^2 c^4}\right) S_x = \sigma_T S_x . \quad (9.13)$$

We find the same total cross-section for scattering as before. This should not be surprising because it does not matter how the electron is forced to oscillate. For incoherent radiation, the energy radiated is proportional to the sum of the incident intensities of the radiation field and so the only important quantity so far as the electron is concerned is the total intensity of radiation incident upon it. It does not matter how anisotropic the incident radiation field is. One convenient way of expressing this result is to write the formula for the scattered radiation in terms of the energy density of radiation u_{rad} at the electron

$$u_{\text{rad}} = \sum_i u_i = \sum_i S_i / c , \quad (9.14)$$

and hence

$$-\left(\frac{dE}{dt}\right) = \sigma_T c u_{\text{rad}} . \quad (9.15)$$

- (iii) One distinctive feature of Thomson scattering is that the scattered radiation is polarised, even if the incident beam of radiation is unpolarised. This can be seen intuitively from Fig. 9.3 because all the E -vectors of the unpolarised beam lie in the $x-y$ plane. Therefore, when the electron is observed precisely in the $x-y$ plane, the scattered radiation is 100% polarised. On the other hand, if we look along the z -direction, we observe unpolarised radiation. If the *degree of polarisation* is defined as

$$\Pi = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} , \quad (9.16)$$

the fractional polarisation of the radiation is

$$\Pi = \frac{1 - \cos^2 \alpha}{1 + \cos^2 \alpha} . \quad (9.17)$$

This is therefore a means of producing polarised radiation from an initially unpolarised beam.

- (iv) Thomson scattering is one of the most important processes which impedes the escape of photons from any region. If the number density of photons of frequency ν is N , the rate at which energy is scattered out of the beam is

$$-\frac{d(Nh\nu)}{dt} = \sigma_T c N h \nu .$$

There is no change of energy of the photons in the scattering process and so, if there are N_e electrons per unit volume, the number density of photons decreases exponentially

with distance

$$-\frac{dN}{dt} = \sigma_T N_e N, \quad -dN/dx = \sigma_T N_e N, \quad N = N_0 \exp(-\int \alpha_T N_e dx). \quad (9.18)$$

We can express this by stating that the *optical depth* τ_T of the medium for Thomson scattering is

$$\tau = \int \sigma_T N_e dx. \quad (9.19)$$

In this process, the photons are scattered in random directions and so they perform a random walk, each step corresponding to the *mean free path* λ_T of the photon through the electron gas, where $\lambda_T = (\sigma_T N_e)^{-1}$. Thus, there is a very real sense in which the Thomson cross-section is the physical cross-section of an electron for the scattering of electromagnetic waves.

9.2.2 Compton scattering

In Thomson scattering, there is no change in the frequency of the radiation. This remains a good approximation provided the energy of the photon is much less than the rest mass energy of the electron, $\hbar\omega \ll m_e c^2$. In general, as long as the energy of the photon is less than $m_e c^2$ in the centre of momentum frame of reference, the scattering may be treated as Thomson scattering, as in our treatment of inverse Compton scattering in Sect. 9.3.3. There are, however, many important cases in which the frequency change associated with the collision between the electron and the photon cannot be neglected. Let us establish some of the more important general results.

Suppose the electron moves with velocity v through the laboratory frame of reference S . Let us use four-vectors to find an elegant solution for the change in energy of the scattered photons. The momentum four-vectors of the electron and the photon before and after the collision are as follows:

	Before	After
Electron	$\mathbf{P} = [\gamma m_e c, \gamma m_e v]$	$\mathbf{P}' = [\gamma' m_e c, \gamma' m_e v']$
Photon	$\mathbf{K} = \left[\frac{\hbar\omega}{c}, \frac{\hbar\omega}{c} \mathbf{i}_k \right]$	$\mathbf{K}' = \left[\frac{\hbar\omega'}{c}, \frac{\hbar\omega'}{c} \mathbf{i}_{k'} \right]$

The collision conserves four-momentum and hence

$$\mathbf{P} + \mathbf{K} = \mathbf{P}' + \mathbf{K}'. \quad (9.20)$$

Now, square both sides of this four-vector equation and use the properties of the norms of the momentum four-vectors of the electron and the photon:

$$\mathbf{P} \cdot \mathbf{P} = \mathbf{P}' \cdot \mathbf{P}' = m_e^2 c^2 \quad \text{and} \quad \mathbf{K} \cdot \mathbf{K} = \mathbf{K}' \cdot \mathbf{K}' = 0. \quad (9.21)$$

Therefore,

$$\begin{aligned} (\mathbf{P} + \mathbf{K})^2 &= (\mathbf{P}' + \mathbf{K}')^2, \\ \mathbf{P} \cdot \mathbf{P} + 2\mathbf{P} \cdot \mathbf{K} + \mathbf{K} \cdot \mathbf{K} &= \mathbf{P}' \cdot \mathbf{P}' + 2\mathbf{P}' \cdot \mathbf{K}' + \mathbf{K}' \cdot \mathbf{K}', \\ \mathbf{P} \cdot \mathbf{K} &= \mathbf{P}' \cdot \mathbf{K}'. \end{aligned} \quad (9.22)$$

Now multiply (9.20) by \mathbf{K}' and use the equality (9.22).

$$\begin{aligned} \mathbf{P} \cdot \mathbf{K}' + \mathbf{K} \cdot \mathbf{K}' &= \mathbf{P}' \cdot \mathbf{K}' + \mathbf{K}' \cdot \mathbf{K}', \\ \mathbf{P} \cdot \mathbf{K}' + \mathbf{K} \cdot \mathbf{K}' &= \mathbf{P} \cdot \mathbf{K}. \end{aligned} \quad (9.23)$$

This is the four-vector equation we seek. Let us reduce it to somewhat more familiar form by multiplying out the four-vector products. The scattering angle is given by $\mathbf{i}_k \cdot \mathbf{i}_{k'} = \cos \alpha$. The angle between the incoming photon and the velocity vector of the electron is θ and the angle between them after the collision is θ' . Then, $\cos \theta = \mathbf{i}_k \cdot \mathbf{v}/|\mathbf{v}|$ and $\cos \theta' = \mathbf{i}_{k'} \cdot \mathbf{v}'/|\mathbf{v}'|$. After a little algebra,

$$\frac{\omega'}{\omega} = \frac{1 - (v/c)\cos \theta}{1 - (v/c)\cos \theta' + (\hbar\omega/\gamma m_e c^2)(1 - \cos \alpha)}. \quad (9.24)$$

In the traditional argument, the *Compton effect* is described in terms of the increase in wavelength of the photon on scattering from a stationary electron, that is, for the case $v = 0, \gamma = 1$,

$$\frac{\omega'}{\omega} = \frac{1}{1 + (\hbar\omega/m_e c^2)(1 - \cos \alpha)}; \quad \frac{\Delta\lambda}{\lambda} = \frac{\lambda' - \lambda}{\lambda} = \frac{\hbar\omega}{m_e c^2}(1 - \cos \alpha). \quad (9.25)$$

This effect of ‘cooling’ the radiation and transferring the energy to the electron is sometimes called the *recoil effect*. Note, however, that (9.24) also shows more generally how energy can be exchanged between the electron and the radiation field. In the limit $\hbar\omega \ll \gamma m_e c^2$, the change in frequency of the photon is

$$\frac{\omega' - \omega}{\omega} = \frac{\Delta\omega}{\omega} = \frac{v}{c} \frac{(\cos \theta - \cos \theta')}{[1 - (v/c)\cos \theta']} . \quad (9.26)$$

Thus, to first order, the frequency changes are $\sim v/c$. Also to first order, if the angles θ and θ' are randomly distributed, a photon is just as likely to decrease as increase its energy. It can be shown that there is no net increase in energy of the photons to first order in v/c and it is only in second order, that is, to order v^2/c^2 , that there is a net energy change.

The Thomson cross-section is only adequate for cases in which the electron moves with velocity $v \ll c$ or if the photon has energy $\hbar\omega \ll m_e c^2$ in the centre of momentum frame of reference. If a photon of energy $\hbar\omega$ collides with a stationary electron, according to the analysis of Sect. 5.3.3, the centre of momentum frame moves at velocity

$$\frac{v}{c} = \frac{\hbar\omega}{m_e c^2 + \hbar\omega} . \quad (9.27)$$

Therefore, if the photons have energy $\hbar\omega \gtrsim m_e c^2$, we must use the proper quantum relativistic cross-section for scattering. Another case which can often arise is if the photons are of low energy $\hbar\omega \ll m_e c^2$ but the electron moves ultra-relativistically with $\gamma \gg 1$. Then,

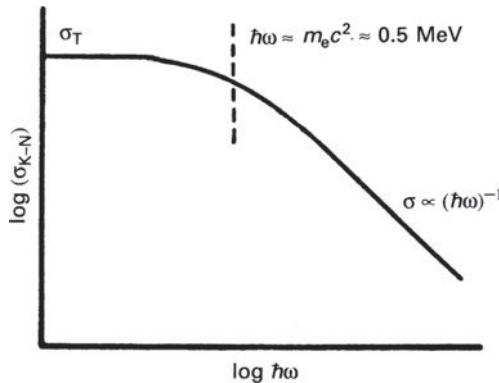


Fig. 9.4

A schematic diagram showing the dependence of the Klein–Nishina cross-section upon photon energy.

the centre of momentum frame moves with a velocity close to that of the electron and in this frame the energy of the photon is $\gamma \hbar\omega$. If $\gamma \hbar\omega \sim m_e c^2$, the quantum relativistic cross-section has to be used.

The relevant total cross-section is the Klein–Nishina formula:

$$\sigma_{K-N} = \pi r_e^2 \frac{1}{x} \left\{ \left[1 - \frac{2(x+1)}{x^2} \right] \ln(2x+1) + \frac{1}{2} + \frac{4}{x} - \frac{1}{2(2x+1)^2} \right\}, \quad (9.28)$$

where $x = \hbar\omega/m_e c^2$ and $r_e = e^2/4\pi\epsilon_0 m_e c^2$ is the classical electron radius. For low energy photons, $x \ll 1$, this expression reduces to

$$\sigma_{K-N} = \frac{8\pi}{3} r_e^2 (1 - 2x) = \sigma_T (1 - 2x) \approx \sigma_T. \quad (9.29)$$

In the ultra-relativistic limit, $\gamma \gg 1$, the Klein–Nishina cross-section becomes

$$\sigma_{K-N} = \pi r_e^2 \frac{1}{x} \left(\ln 2x + \frac{1}{2} \right), \quad (9.30)$$

so that the cross-section decreases roughly as x^{-1} at the highest energies (Fig. 9.4). If the atom has Z electrons, the total cross-section per atom is $Z\sigma_{K-N}$. Note that scattering by nuclei can be neglected because they cause very much less scattering than electrons, roughly by a factor of $(m_e/m_N)^2$, where m_N is the mass of the nucleus.

9.3 Inverse Compton scattering

In inverse Compton scattering, ultra-relativistic electrons scatter low energy photons to high energies so that the photons gain energy at the expense of the kinetic energy of the electrons. The process is called *inverse Compton scattering* because the electrons lose energy rather than the photons. We consider the case in which the energy of the photon in the centre of momentum frame of reference is much less than $m_e c^2$ and consequently the Thomson scattering cross-section can be used to describe the probability of scattering.

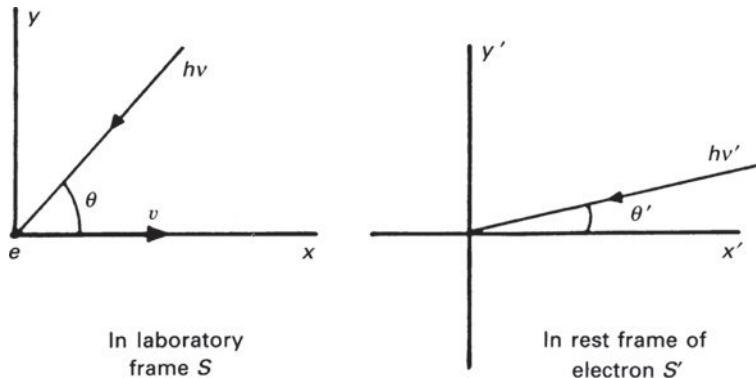


Fig. 9.5

The geometry of inverse Compton scattering in the laboratory frame of reference S and that in which the electron is at rest S' .

Many of the most important results can be worked out using simple arguments (Blumenthal and Gould, 1970; Rybicki and Lightman, 1979). The geometry of inverse Compton scattering is illustrated in Fig. 9.5 which depicts the collision between a photon and a relativistic electron as seen in the laboratory frame of reference S and in the rest frame of the electron S' . In the case in which $\gamma \hbar\omega \ll m_e c^2$, the centre of momentum frame of reference is very closely that of the relativistic electron. If the energy of the photon is $\hbar\omega$ and the angle of incidence θ in S , its energy in the frame S' is

$$\hbar\omega' = \gamma \hbar\omega [1 + (v/c) \cos \theta], \quad (9.31)$$

according to the relativistic Doppler shift formula. The angle of incidence θ' in the frame S' is related to θ in S by the aberration formulae

$$\sin \theta' = \frac{\sin \theta}{\gamma [1 + (v/c) \cos \theta]}; \quad \cos \theta' = \frac{\cos \theta + v/c}{[1 + (v/c) \cos \theta]}. \quad (9.32)$$

Provided $\hbar\omega' \ll m_e c^2$, the Compton interaction in the rest frame of the electron is Thomson scattering and hence the energy loss rate of the electron in S' is the rate at which energy is reradiated by the electron. According to (9.15), this loss rate is

$$-\left(\frac{dE}{dt}\right)' = \sigma_T c u'_{\text{rad}}, \quad (9.33)$$

where u'_{rad} is the energy density of radiation in the rest frame of the electron. As shown in Sect. 9.2.1, it is of no importance whether or not the radiation is isotropic – the free electron accelerates in response to any incident field. Therefore, our strategy is to work out the energy density u'_{rad} in the frame S' of the electron and then to use expression (9.15) to find $(dE/dt)'$. Using the result obtained in Sect. 6.3.1, this is also the loss rate (dE/dt) in the frame S .

We give two derivations of the key result. In the first method, we consider the rate of arrival of photons at the origin of the moving frame S' . Suppose the number density of photons in a parallel beam of radiation incident at angle θ to the x -axis is N . Then, the energy density of these photons in S is $N\hbar\omega$ and the flux density of photons incident upon

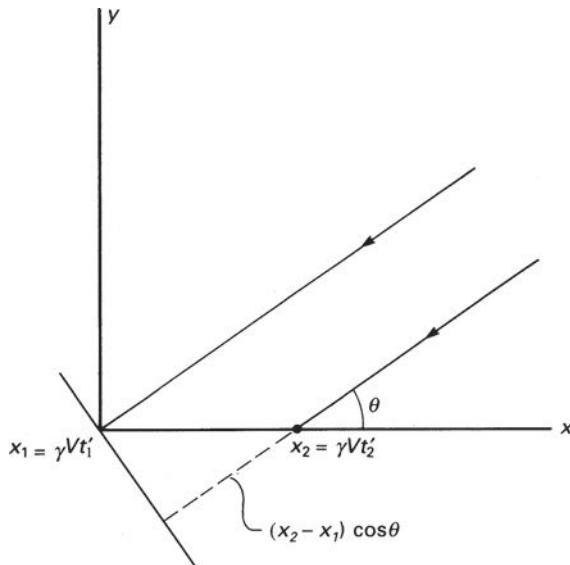


Fig. 9.6

Illustrating the rate of arrival of photons at the observer in the laboratory frame of reference (see text).

a stationary electron in S is $u_{\text{rad}}c = N\hbar\omega c$. To work out the flux density of the beam as observed in the frame of reference of the stationary electron S' , we need two things: the energy of each photon in S' and the rate of arrival of photons at the electron. The first of these is given by (9.31). To find the second factor, consider two photons which arrive at the origin of S' at times t'_1 and t'_2 at the angle θ' to the x' -axis. The coordinates of these events in S' are

$$[ct'_1, 0, 0, 0] \quad \text{and} \quad [ct'_2, 0, 0, 0].$$

The coordinates of these events in S are

$$[ct_1, x_1, 0, 0] = [\gamma ct'_1, \gamma Vt'_1, 0, 0] \quad \text{and} \quad [ct_2, x_2, 0, 0] = [\gamma ct'_2, \gamma Vt'_2, 0, 0],$$

respectively, where we have used the inverse Lorentz transformations,

$$ct = \gamma \left(ct' + \frac{Vx'}{c} \right) \quad \text{and} \quad x = \gamma (x' + Vt') . \quad (9.34)$$

This calculation makes the important point that the photons in the beam propagate along parallel but separate trajectories at an angle θ to the x -axis in S , as illustrated in Fig. 9.6. From the geometry of Fig. 9.6, it is apparent that the time difference when the photons arrive at a plane perpendicular to their direction of propagation in S is

$$\Delta t = t_2 + \frac{(x_2 - x_1)}{c} \cos \theta - t_1 = (t'_2 - t'_1)\gamma[1 + (v/c)\cos \theta] , \quad (9.35)$$

that is, the time interval between the arrival of photons from the direction θ' is shorter by a factor $\gamma[1 + (v/c)\cos \theta]$ in S' than it is in S . Thus, the rate of arrival of photons and correspondingly the number density of photons is greater by this factor $\gamma[1 + (v/c)\cos \theta]$ in S' as compared with that in S . Comparison with (9.31) shows that this is exactly the

same factor by which the energy of the photon has increased. Thus, as observed in S' , the energy density of the beam is

$$u'_{\text{rad}} = [\gamma(1 + (v/c)\cos\theta)]^2 u_{\text{rad}} . \quad (9.36)$$

The second way of deriving the result (9.36) is somewhat more elegant. It uses the fact that the four-volume $dt dx dy dz$ is invariant between any pair of inertial frames of reference. For reference frames in standard configuration, $dy = dy'$ and $dz = dz'$. Therefore, we need only consider the transformation of the differential product $dt dx$. According to the standard procedure for relating differential areas in different coordinate systems,

$$dt dx = \begin{vmatrix} \frac{\partial t}{\partial t'} & \frac{\partial x}{\partial t'} \\ \frac{\partial t}{\partial x'} & \frac{\partial x}{\partial x'} \end{vmatrix} dt' dx' , \quad (9.37)$$

where the determinant is the Jacobian of the transformation between the frames S and S' . It is straightforward to use the inverse Lorentz transformations (9.34) to show that the value of the determinant in (9.37) is unity. Therefore, the four-volume element $dt dx dy dz$ is an invariant between inertial frames of reference.

We now combine this result with other invariants to create new invariant relations between inertial frames of reference. Consider the number density of particles of energy E , $n(E)$, moving at velocity v at an angle θ to the x -axis, as illustrated in Fig. 9.6. The number of photons in the differential three-volume $dN(E) = n(E) dx dy dz$ is an invariant between inertial frames, where $n(E)$ is the number density of photons of energy E . Consequently, because the four-volume element $dt dx dy dz$ is an invariant between inertial frames of reference, so also is $n(E)/dt$. But, as was shown in Sect. 6.2.1, dt and E transform in the same way between reference frames and so $n(E)/E$ is also an invariant between inertial frames. The change in energy of the photons between the frames S and S' is given by (9.31) and so the number density of photons increases by the same factor. We therefore recover the result (9.36) somewhat more economically.

The procedure of the last two paragraphs provides a powerful tool for creating many useful relativistic invariants. For example, the differential momentum four-vector is $\mathbf{P} = [dE/c, dp_x, dp_y, dp_z]$ and so $dE dp_x dp_y dp_z$ is an invariant volume in four-momentum space. We will return to this result in the discussion of occupation numbers in phase space in the context of Comptonisation in Sect. 9.4.

Returning to (9.36), it is now a simple calculation to work out the energy density of radiation observed by the electron in its rest frame. It is assumed that the radiation field is isotropic in S and therefore the contribution to u'_{rad} from the solid angle $d\Omega$ in S is

$$du'_{\text{rad}} = u_{\text{rad}} \gamma^2 [1 + (v/c)\cos\theta]^2 d\Omega = u_{\text{rad}} \gamma^2 [1 + (v/c)\cos\theta]^2 \frac{1}{2} \sin\theta d\theta . \quad (9.38)$$

Integrating over solid angle,

$$u'_{\text{rad}} = u_{\text{rad}} \int_0^\pi \gamma^2 [1 + (v/c)\cos\theta]^2 \frac{1}{2} \sin\theta d\theta = \frac{4}{3} u_{\text{rad}} \left(\gamma^2 - \frac{1}{4} \right) . \quad (9.39)$$

Substituting into (9.33) and using the result (6.2) that $(dE/dt) = (dE/dt)'$,

$$\frac{dE}{dt} = \frac{4}{3}\sigma_T cu_{\text{rad}} \left(\gamma^2 - \frac{1}{4} \right). \quad (9.40)$$

This is the energy gained by the photon field due to the scattering of the low energy photons. We have therefore to subtract the initial energy of the low-energy photons to find the total energy gain of the photon field in S . The rate at which energy is removed from the low energy photon field is $\sigma_T cu_{\text{rad}}$ and therefore, subtracting,

$$\frac{dE}{dt} = \frac{4}{3}\sigma_T cu_{\text{rad}} \left(\gamma^2 - \frac{1}{4} \right) - \sigma_T cu_{\text{rad}} = \frac{4}{3}\sigma_T cu_{\text{rad}}(\gamma^2 - 1).$$

Using the identity $(\gamma^2 - 1) = (v^2/c^2)\gamma^2$, the loss rate in its final form is

$$\left(\frac{dE}{dt} \right)_{\text{IC}} = \frac{4}{3}\sigma_T cu_{\text{rad}} \left(\frac{v^2}{c^2} \right) \gamma^2. \quad (9.41)$$

This is the result we have been seeking. It is exact so long as $\gamma \hbar\omega \ll m_e c^2$.

Notice the remarkable similarity of the result (9.41) to the expression (8.9) for the mean energy loss rate of the ultra-relativistic electron by synchrotron radiation

$$\left(\frac{dE}{dt} \right)_{\text{sync}} = \frac{4}{3}\sigma_T cu_{\text{mag}} \left(\frac{v^2}{c^2} \right) \gamma^2. \quad (9.42)$$

The reason for this is that the energy loss rate depends upon the electric field which accelerates the electron in its rest frame and it does not matter what the origin of that field is. In the case of synchrotron radiation, the electric field is the $(\mathbf{v} \times \mathbf{B})$ field due to motion of the electron through the magnetic field whereas, in the case of inverse Compton scattering, it is the sum of the electric fields of the electromagnetic waves incident upon the electron. In the latter case, the sum of the squares of the electric field strengths appears in the formulae for incoherent radiation and so the energies of the waves add linearly (see Sect. 9.2.1). Another way of expressing this similarity between the loss processes is to consider synchrotron radiation to be the scattering of ‘virtual photons’ observed by the electron as it gyrates about the magnetic field (Jackson, 1999).

The similarity of the synchrotron and inverse Compton scattering processes means that we can use the results of Sect. 8.5.1 to work out the spectrum of radiation produced by a power-law distribution of electron energies. The spectral index of the scattered radiation is $a = (p - 1)/2$, where p is the spectral index of the electron energy spectrum. Notice that this relation is true for the intensity of radiation measured in $\text{W m}^{-2} \text{ Hz}^{-1}$. In terms of photon flux density, the spectral index would be one power of frequency, or energy, steeper $a_{\text{ph}} = (p + 1)/2$

The next step is to determine the spectrum of the scattered radiation. This is a somewhat lengthy, but straightforward, calculation. Because of the extreme effects of aberration, the photons which interact with the electron in the frame S' propagate in the negative direction along the x' -axis, the spectrum of the incident radiation being found from (9.36). They are then scattered in the moving frame with the probability distribution given by the differential Thomson cross-section (9.8). The spectrum of the scattered radiation is then

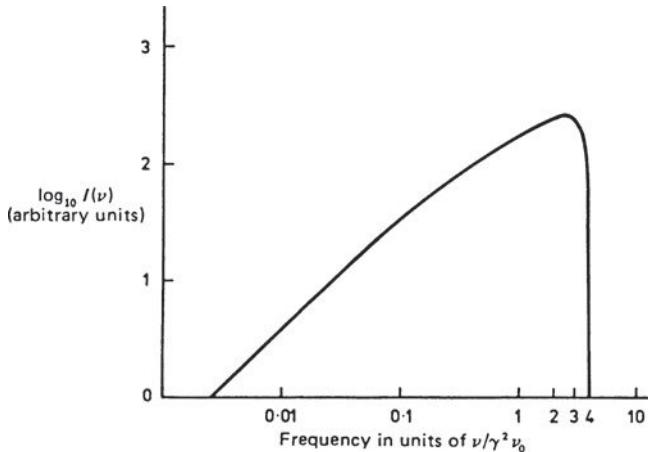


Fig. 9.7

The emission spectrum of inverse Compton scattering; ν_0 is the frequency of the unscattered radiation (Blumenthal and Gould, 1970).

transformed back into the laboratory frame of reference. The result of these calculations is given by Blumenthal and Gould for an incident isotropic photon field at a single frequency ν_0 (Blumenthal and Gould, 1970). The spectral emissivity $I(\nu)$ may be written

$$I(\nu) d\nu = \frac{3\sigma_{TC}}{16\gamma^4} \frac{N(\nu_0)}{\nu_0^2} \nu \left[2\nu \ln \left(\frac{\nu}{4\gamma^2\nu_0} \right) + \nu + 4\gamma^2\nu_0 - \frac{\nu^2}{2\gamma^2\nu_0} \right] d\nu, \quad (9.43)$$

where the isotropic radiation field in the laboratory frame of reference S is assumed to be monochromatic with frequency ν_0 ; $N(\nu_0)$ is the number density of photons. This spectrum is shown in Fig. 9.7. At low frequencies, the term in square brackets in (9.42) is a constant and hence the scattered radiation has a spectrum of the form $I(\nu) \propto \nu$.

It is an easy calculation to show that the maximum energy which the photon can acquire corresponds to a head-on collision in which the photon is sent back along its original path. The maximum energy of the photon is

$$(\hbar\omega)_{\max} = \hbar\omega\gamma^2(1+v/c)^2 \approx 4\gamma^2\hbar\omega_0. \quad (9.44)$$

Another important result can be derived from (9.41), the total energy loss rate of the electron. The number of photons scattered per unit time is $\sigma_T c u_{\text{rad}} / \hbar\omega_0$ and hence the average energy of the scattered photons is

$$\hbar\overline{\omega} = \frac{4}{3}\gamma^2 \left(\frac{v}{c} \right)^2 \hbar\omega_0 \approx \frac{4}{3}\gamma^2\hbar\omega_0. \quad (9.45)$$

This result gives substance to the hand-waving argument that the photon gains typically one factor of γ in transforming into S' and then gains another on transforming back into S .

The general result that the frequency of photons scattered by ultra-relativistic electrons is $\nu \sim \gamma^2\nu_0$ is of profound importance in high energy astrophysics. There are certainly electrons with Lorentz factors $\gamma \sim 100-1000$ in various types of astronomical source and consequently they scatter any low energy photons to very much higher energies. To give some examples, consider radio, infrared and optical photons scattered by electrons with

$\gamma = 1000$. The scattered radiation has average frequency (or energy) roughly 10^6 times that of the incoming photons. Radio photons with $\nu_0 = 10^9$ Hz become ultraviolet photons with $\nu = 10^{15}$ Hz ($\lambda = 300$ nm); far-infrared photons with $\nu_0 = 3 \times 10^{12}$ Hz, typical of the photons seen in galaxies which are powerful far-infrared emitters, produce X-rays with frequency 3×10^{18} Hz, that is, about 10 keV; optical photons with $\nu_0 = 4 \times 10^{14}$ Hz become γ -rays with frequency 4×10^{20} Hz, that is, about 1.6 MeV. It is apparent that the inverse Compton scattering process is an effective means of creating very high energy photons. It also becomes an inevitable drain of energy for high energy electrons whenever they pass through a region in which there is a large energy density of radiation.

9.4 Comptonisation

The calculations carried out in Sect. 9.2.2 demonstrate how energy can be interchanged between photons and electrons by Compton scattering in particular limiting cases. If the evolution of the spectrum of the source is dominated by Compton scattering, the process is often referred to as *Comptonisation*. This enormous subject is considered in much more detail by Pozdnyakov *et al.* (1983); Liedahl (1999) and Rybicki and Lightman (1979).

9.4.1 The basic physics of Comptonisation

The requirement that the evolution of the spectrum be determined by Compton scattering means that the plasma must be rarefied so that other radiation processes such as bremsstrahlung do not contribute additional photons to the system. In addition, the effects of Comptonisation are important if the plasma is very hot because then the exchange of energy per collision is greater. Examples of sources in which such conditions are found include the hot gas in the vicinity of binary X-ray sources, the hot plasmas in the nuclei of active galaxies, the hot intergalactic gas in clusters of galaxies and the early evolution of the hot primordial plasma.

Let us build up a simple picture of the Comptonisation process. We restrict the discussion to the non-relativistic regime in which $kT_e \ll m_e c^2$ and $\varepsilon = \hbar\omega \ll m_e c^2$ and so the Thomson cross-section can be used for interactions between radiation and the electrons. The expression for the energy transferred to stationary electrons from the photon field (9.25) can be written in terms of the fractional change of energy of the photon per collision in the limit $\hbar\omega \ll m_e c^2$,

$$\frac{\Delta\varepsilon}{\varepsilon} = \frac{\hbar\omega}{m_e c^2} (1 - \cos \alpha). \quad (9.46)$$

In the frame of reference of the electron, the scattering is Thomson scattering and so the probability distribution of the scattered photons is symmetrical about their incident directions. Therefore, when averages are taken over the scattering angle α , opposite values

of $\cos \alpha$ cancel out and the average energy increase of the electron is

$$\left\langle \frac{\Delta \varepsilon}{\varepsilon} \right\rangle = \frac{\hbar \omega}{m_e c^2}. \quad (9.47)$$

This is the recoil effect discussed in Sect. 9.2.2. In the opposite limit in which energy is transferred from the electrons to the photon field, we can adopt the low energy limit of the energy loss rate of high energy electrons by inverse Compton scattering. The derivation of (9.41) is correct for all values of the Lorentz factor γ and hence incorporates the effects of aberration and Doppler scattering, even if these effects are small. The low energy limit of (9.41) is

$$\frac{dE}{dt} = \frac{4}{3} \sigma_T c u_{\text{rad}} \left(\frac{v}{c} \right)^2. \quad (9.48)$$

The number of photons scattered per second is $\sigma_T N_{\text{phot}} c = \sigma_T u_{\text{rad}} c / \hbar \omega$, and so the average energy gain by the photons per Compton collision is

$$\left\langle \frac{\Delta \varepsilon}{\varepsilon} \right\rangle = \frac{4}{3} \left(\frac{v}{c} \right)^2. \quad (9.49)$$

The average energy gain per collision is second order in v/c because the first-order effects cancel out. The net increase in energy is statistical because implicitly, in deriving (9.41), we integrated over all angles of scattering.

If the electrons have a thermal distribution of velocities at temperature T_e , $\frac{1}{2} m_e \langle v^2 \rangle = \frac{3}{2} k T_e$ and hence

$$\frac{\Delta \varepsilon}{\varepsilon} = \frac{4kT_e}{m_e c^2}. \quad (9.50)$$

As a result, the equation describing the average energy change of the photon per collision is

$$\frac{\Delta \varepsilon}{\varepsilon} = \frac{4kT_e - \hbar \omega}{m_e c^2}. \quad (9.51)$$

There is therefore no energy transfer if $\hbar \omega = 4kT_e$. If $4kT_e > \hbar \omega$, energy is transferred to the photons whilst if $\hbar \omega > 4kT_e$ energy is transferred to the electrons.

In the case in which the electrons are hotter than the photons, the fractional increase in energy is $4kT_e/m_e c^2$ per collision and hence we need to evaluate the number of collisions which the photon makes with electrons before they escape from the scattering region. If the region has electron density N_e and size l , the optical depth for Thomson scattering is

$$\tau_e = N_e \sigma_T l. \quad (9.52)$$

If $\tau_e \gg 1$, the photons undergo a random walk in escaping from the region and so the photon travels a distance $l \approx N^{1/2} \lambda_e$ in N scatterings where $\lambda_e = (N_e \sigma_T)^{-1}$ is the mean free path of the photon. Therefore, in the limit $\tau_e \gg 1$, which is necessary to alter significantly the energy of the photon, the number of scatterings is $N = (l/\lambda_e)^2 = \tau_e^2$. If $\tau_e \ll 1$, the number of scatterings is τ_e and hence the condition for a significant distortion of the photon

spectrum by Compton scattering is $4y \gtrsim 1$, where

$$y = \frac{kT_e}{m_e c^2} \max(\tau_e, \tau_e^2) . \quad (9.53)$$

y is referred to as the *Compton optical depth*. Normally, the condition for Comptonisation to change significantly the spectrum of the photons is

$$y = \frac{kT_e}{m_e c^2} \tau_e^2 \gtrsim \frac{1}{4} . \quad (9.54)$$

Let us investigate how repeated scatterings change the energy of the photons. The analysis of Liedahl is rather pleasant (Liedahl, 1999). First of all, we convert (9.51) into a differential equation for the rate of change of the energy of the photons. If N is the number of scatterings,

$$\frac{d\varepsilon}{dN} = \frac{4kT_e - \hbar\omega}{m_e c^2} \varepsilon = A\varepsilon - \frac{\varepsilon^2}{m_e c^2} , \quad (9.55)$$

where $A = 4kT_e/m_e c^2$. Setting $x = \hbar\omega/m_e c^2 = \varepsilon/m_e c^2$, we find

$$\frac{dx}{dN} = Ax - x^2 . \quad (9.56)$$

It is straightforward to find the definite integral of this equation for a photon with initial energy ε_0 , or equivalently $x_0 = \varepsilon_0/m_e c^2$,

$$\frac{x}{A-x} = \frac{x_0}{A-x_0} e^{AN} . \quad (9.57)$$

Initially, the photon has energy $\varepsilon_0 \ll 4kT_e$ and therefore $A \gg x_0$. Hence,

$$\frac{x}{A-x} = \frac{x_0}{A} e^{AN} . \quad (9.58)$$

Furthermore, the Compton optical depth of the medium is $y = (kT_e/m_e c^2)N = AN/4$ and so, solving (9.58) for x , we find

$$\varepsilon = \varepsilon_0 \frac{e^{4y}}{1 + \frac{\hbar\omega_0}{4kT} e^{4y}} . \quad (9.59)$$

Thus, when $(\hbar\omega_0/4kT)e^{4y}$ is small, the energy of the photon increases exponentially as $\varepsilon = \varepsilon_0 e^{4y}$. However, when the Comptonisation is strong, $(\hbar\omega_0/4kT)e^{4y} \gg 1$, the energy of the photons saturate at $\varepsilon = 4kT$, as expected from (9.51).

We can now work out the number of scatterings in order to approach saturation. The saturation energy is $\hbar\omega = \varepsilon = 4kT_e$ and so let us work out the number of scatterings to attain the energy $\varepsilon/2 = 2kT_e$. Inserting this value into (9.59) and recalling that $y = (kT_e/m_e c^2)N$, we find

$$N = \frac{m_e c^2}{4kT_e} \ln \left(\frac{4kT_e}{\hbar\omega_0} \right) . \quad (9.60)$$

In the case in which the Thomson optical depth of the cloud τ_e is much greater than unity, $N \sim \tau_e^2$ and so

$$\tau_e = \left[\frac{m_e c^2}{4 k T_e} \ln \left(\frac{4 k T_e}{\hbar \omega_0} \right) \right]^{1/2}. \quad (9.61)$$

Liedahl gives the example of injecting optical photons with energy $\hbar \omega_0 = 4$ eV into a gas with temperature $k T_e = 10$ keV. Then, there would have to be $N \approx 100$ scatterings for the photons to approach a Comptonised X-ray spectrum. If this were associated with random scattering within a cloud of optical depth τ_e , there would have to be roughly τ_e^2 scatterings and so the Thomson optical depth of the region would have to be $\tau \approx 10$. The corresponding value of the Compton optical depth is $y = (k T_e / m_e c^2) N \approx 2$.

If the Compton optical depth y of the medium is very much greater than unity, the photon distribution approaches its equilibrium form entirely under Compton scattering. Photons are bosons and, consequently, the equilibrium spectrum is given in general by the Bose–Einstein distribution, the energy density of which is

$$u_v dv = \frac{8\pi h v^3}{c^3} \left[\exp \left(\frac{h v}{k T} + \mu \right) - 1 \right]^{-1} dv, \quad (9.62)$$

where μ is the *chemical potential*. In the case of the Planck spectrum, $\mu = 0$ and the number and energy densities of the photons are uniquely defined by a single parameter, the thermal equilibrium temperature of the matter and radiation T . If there is a mismatch between the number density of photons and the energy density of radiation, the equilibrium spectrum is the Bose–Einstein distribution with a finite chemical potential μ . The forms of these spectra are shown in Fig. 9.8 for different values of the chemical potential μ . In the limiting case $\mu \gg 1$, the spectrum is the Wien distribution reduced by the factor $\exp(-\mu)$,

$$u_v = \exp(-\mu) \frac{8\pi h v^3}{c^3} \exp \left(-\frac{h v}{k T} \right). \quad (9.63)$$

The average energy of the photons is

$$\langle \hbar \omega \rangle = k T_e \frac{\int_0^\infty x^3 \exp(-x) dx}{\int_0^\infty x^2 \exp(-x) dx} = 3k T_e, \quad (9.64)$$

exactly the same result derived by Einstein in his great paper of 1905 in which he introduced the concept of light quanta (Longair, 2003).

9.4.2 Pedagogical interlude – occupation number

We now need the equation which describes how the spectrum of radiation evolves towards the Bose-Einstein distribution. In the non-relativistic limit, this equation is known as the *Kompaneets equation*, which is discussed in Sect. 9.4.3. It is written in terms of the *occupation number* of photons in phase space, because we need to include both spontaneous and induced processes in the calculation. Let us compare this approach with that involving the coefficients of emission and absorption of radiation.

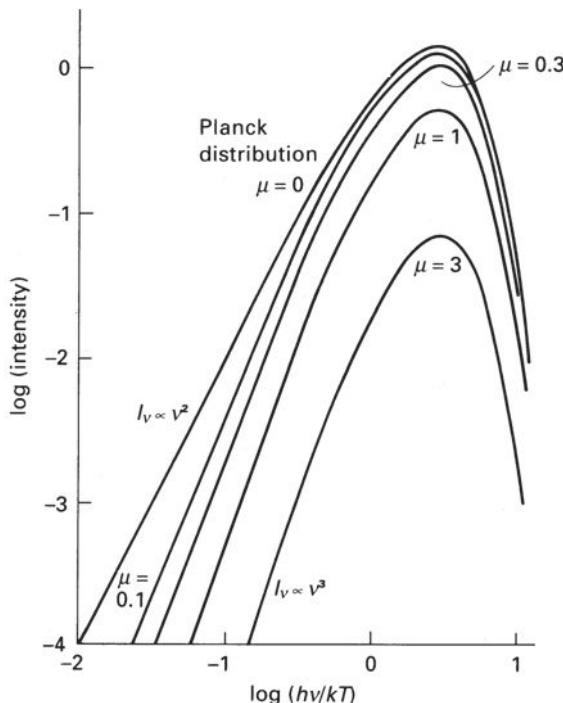


Fig. 9.8

Illustrating the intensity spectra of Bose–Einstein distributions with different values of the dimensionless chemical potential μ . The distribution with $\mu = 0$ is the Planck function. At energies $h\nu \gg \mu kT$ the distributions are similar to a Wien distribution but with intensity reduced by a factor $\exp(-\mu)$. At energies $h\nu \ll \mu kT$, the intensity spectrum is $I_\nu \propto \nu^3$. In general, for large values of μ , the distribution follows closely that of a Wien distribution with intensity reduced by a factor $\exp(-\mu)$.

My favorite reference for understanding the basic physics of spontaneous and induced processes is the beautiful discussion by Feynman in Chap. 4 of Volume III of his *Lectures on Physics* (Feynman *et al.*, 1965). In Sect. 4.4, he enunciates the key rule for the emission and absorption of photons, which are spin-1 bosons.

The probability that an atom will emit a photon into a particular final state is increased by a factor $(n + 1)$ if there are already n photons in that state.

Notice that the statement is made in terms of probabilities rather than quantum mechanical amplitudes – in the latter case, the amplitude would be increased by a factor $\sqrt{n + 1}$. We will use probabilities in our analysis. n will turn out to be the *occupation number*. To derive the Planck spectrum, consider an atom which can be in two states, an upper state 2 with energy $\hbar\omega$ greater than the lower state 1. N_1 is the number of atoms in the lower state and N_2 the number in the upper state. In thermodynamic equilibrium, the ratio of the numbers of atoms in these states is given by the Boltzmann relation,

$$\frac{N_2}{N_1} = \exp(-\Delta E / kT) = \exp(-\hbar\omega / kT), \quad (9.65)$$

where $\Delta E = \hbar\omega$ and the statistical weights g_2 and g_1 are assumed to be the same. When a photon of energy $\hbar\omega$ is absorbed, the atom is excited from state 1 to state 2 and, when a photon of the same energy is emitted from state 2, the atom de-excites from state 2 to state 1. In thermodynamic equilibrium, the rates for the emission and absorption of photons between the two levels must be exactly balanced. These rates are proportional to the product of the probability of the events occurring and the number of atoms present in the appropriate state. Suppose \bar{n} is the average number of photons in a given state in the phase space of the photons with energy $\hbar\omega$. Then, the absorption rate of these photons by the atoms in the state 1 is $N_1 \bar{n} p_{12}$, where p_{12} is the probability that the photon will be absorbed by an atom in state 1, which is then excited to state 2. According to the rule enunciated above by Feynman, the rate of emission of photons when the atom de-excites from state 2 to state 1 is $N_2(\bar{n} + 1)p_{21}$. At the quantum mechanical level, the probabilities p_{12} and p_{21} are equal. This is because the matrix element for, say, the p_{12} transition is the complex conjugate of the transition p_{21} and, since the probabilities depend upon the square of the magnitude of the matrix elements, they must be equal. This is called the *principle of jump rate symmetry*. Therefore,

$$N_1 \bar{n} = N_2(\bar{n} + 1). \quad (9.66)$$

Solving for \bar{n} and using (9.65),

$$\bar{n} = \frac{1}{e^{\hbar\omega/kT} - 1}. \quad (9.67)$$

The elementary volume of phase space for photons is $(2\pi)^3$. There are two independent polarisations for each state and hence the number of states in the phase space volume d^3k is $2 d^3k/(2\pi)^3$. If the photon distribution is isotropic, the photons which lie in the frequency interval ν to $\nu + d\nu$ have wavevectors k which lie in a spherical shell of radius k and thickness dk and so volume $d^3k = 4\pi k^2 dk$. Therefore, the number of states in this volume of photon phase space is

$$\frac{8\pi k^2 dk}{(2\pi)^3} = \frac{8\pi \nu^2 d\nu}{c^3} = \frac{\omega^2 d\omega}{c^3 \pi^2}. \quad (9.68)$$

To complete the calculation, the energy density of radiation is the product of the energy of each photon, the volume of phase space in which the photons have energies in the interval $\hbar\omega$ to $\hbar(\omega + d\omega)$ and the occupation number of each state,

$$u(\nu) d\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1} d\nu \quad \text{or} \quad u(\omega) d\omega = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/kT} - 1} d\omega. \quad (9.69)$$

We have recovered the Planck spectrum.

For our present purposes, the important relation is the general expression for the occupation number of the photons in phase space. If the energy density of isotropic radiation in the frequency interval ν to $\nu + d\nu$ is $u_\nu d\nu$, the number density of photons is $u_\nu d\nu/h\nu$ and the mean occupation number, $\overline{n(\nu)}$ or $\overline{n(\omega)}$, is

$$\overline{n(\nu)} = \frac{u_\nu c^3}{8\pi h\nu^3} \quad \text{or} \quad \overline{n(\omega)} = \frac{u(\omega) \pi^2 c^3}{\hbar\omega^3}. \quad (9.70)$$

There are particularly simple expressions for the occupation number of photons for the Bose–Einstein and Planck distributions,

$$\begin{aligned} \text{Bose–Einstein} \quad n(\nu) &= [\exp(x + \mu) - 1]^{-1}; \\ \text{Planck} \quad n(\nu) &= [\exp x - 1]^{-1}, \end{aligned} \quad (9.71)$$

where $x = h\nu/kT = \hbar\omega/kT$. The occupation number $n(\nu)$ determines when it is necessary to include stimulated emission terms in the expressions for interactions of photons. If $n > 1$, then the effects of stimulated emission cannot be neglected. For black-body radiation, this means in the Rayleigh–Jeans region of the spectrum, $h\nu \ll kT$.

Let us now rewrite the transfer equation for radiation in terms of occupation numbers. For the case of isotropic radiation, equation (6.61) can be written

$$\frac{dI(\omega)}{dx} = \hbar\omega N_2 A_{21} - N_1 B_{12} \hbar\omega I(\omega) + N_2 B_{21} \hbar\omega I(\omega). \quad (9.72)$$

We recall that the spontaneous emission coefficient is $\kappa_\nu = \hbar\omega N_2 A_{21}$. Notice that this equation is written in terms of the intensity of radiation integrated over 4π steradians per unit angular frequency and so is exactly equivalent to the analysis using the Einstein coefficients in Sect. 6.5.2. We now use the relations between the Einstein coefficients,

$$B_{12} = B_{21}; \quad A_{21} = \frac{\hbar\omega^3}{\pi^2 c^2} B_{21}, \quad (9.73)$$

to rewrite the transfer equation as

$$\frac{dI(\omega)}{dx} = \hbar\omega N_2 \frac{\hbar\omega^3}{\pi^2 c^2} B_{21} - N_1 B_{12} \hbar\omega I(\omega) + N_2 B_{21} \hbar\omega I(\omega). \quad (9.74)$$

We now rewrite the transfer equation in terms of occupation numbers using

$$n(\omega) = \frac{I(\omega) \pi^2 c^2}{\hbar\omega^3}, \quad (9.75)$$

so that

$$\frac{dn(\omega)}{dx} = \hbar\omega B_{21} \{-N_1 n(\omega) + N_2 [1 + n(\omega)]\}. \quad (9.76)$$

This is the equation we have been seeking. Notice how the rule described by Feynman comes naturally out of an analysis of Einstein's coefficients for spontaneous and stimulated emission. It also illustrates how the transfer equation for radiation can be written in a remarkably compact form using occupation numbers, including both spontaneous emission and simulated emission and absorption. To make this clearer, let us simplify the notation of (9.76) to be similar to that used in the next section,

$$\frac{dn}{dx} = \hbar\omega B_{21} [-N_1 n + N_2 (1 + n)]. \quad (9.77)$$

The three terms on the right-hand side in square brackets represent stimulated absorption with the minus sign and spontaneous emission and stimulated emission with the plus sign respectively. It will be noticed that the right-hand side of (9.77) is identical with (9.66) when the left-hand side is set equal to zero in thermal equilibrium.

The occupation number \bar{n} is a Lorentz invariant. This is most easily understood by considering the invariant volume of the differential momentum four-vector for photons,

$$d\mathbf{P} = [dE, dp_1, dp_2, dp_3] \equiv \hbar [d\omega, dk_1, dk_2, dk_3] = \hbar d\mathbf{K}. \quad (9.78)$$

By exactly the same argument as given in Sect. 9.3 for the invariant volume $d\tau dx dy dz$, it follows that the volume element in four-dimensional momentum space $d\omega dk_1 dk_2 dk_3$ is invariant under Lorentz transformation. The number of photons in the element of four-dimensional phase space N is an invariant number and so

$$\bar{n} = \frac{N}{d\omega d^3k} \quad (9.79)$$

is a Lorentz invariant, recalling that \bar{n} is defined per unit angular frequency per unit volume of k -space.

9.4.3 The Kompaneets equation

The Kompaneets equation for the evolution of the occupation number n under Compton scattering is named after the Soviet physicist Aleksander Solomonovich Kompaneets who published its derivation in 1956 (Kompaneets, 1956). In fact, the equation had been derived in the late 1940s by the combined efforts of Kompaneets, Landau, Gel'fand and Dyakov under the direction of Zeldovich as part of the Soviet atomic and hydrogen bomb programme.

The derivation of the Kompaneets equation is non-trivial since it has to take account of the interchange of energy between the photons and electrons and also include induced effects which become important when the occupation number n is large. The derivations outlined by Rybicki and Lightman and by Liedahl give an excellent impression of what is involved (Rybicki and Lightman, 1979; Liedahl, 1999).

Provided the fractional changes of energy per Compton interaction are small, the Boltzmann equation can be used to describe the evolution of the photon occupation number,

$$\frac{\partial n(\omega)}{\partial t} = c \int d^3p d\Omega \frac{d\sigma}{d\Omega} [f(\mathbf{p}')n(\omega')(1+n(\omega)) - f(\mathbf{p})n(\omega)(1+n(\omega'))]. \quad (9.80)$$

The first term in square brackets within the integral describes the increase in the occupation number due to photon scattering from frequency ω' to ω . Notice that this term includes the factor $(1+n(\omega))$ which takes account of the fact that the photons are bosons and so, as discussed in Sect. 9.4.2, there is an increased probability of scattering by this factor if the occupation number of the final state is already $n(\omega)$. The second term in square brackets describes the loss of photons of frequency ω by Compton scattering from ω to ω' , again the stimulated term $(1+n(\omega'))$ being included to take account of induced scattering.

In deriving the Kompaneets equation, it is assumed that the electron distribution remains Maxwellian at temperature T , so that

$$f(p) = \frac{N_e}{(2\pi m_e k T)^{3/2}} e^{-p^2/2m_e k T}. \quad (9.81)$$

The differential cross-section for Thomson scattering is given by (9.11). The change of angular frequency of the photon is given by (9.24) in the non-relativistic limit in which

the frequency changes are small. Liedahl provides a clear description of the approximations involved and the tricks which can be used to derive the final form of the equation, which is

$$\frac{\partial n}{\partial y} = \frac{1}{x^2} \frac{\partial}{\partial x} \left[x^4 \left(n + n^2 + \frac{\partial n}{\partial x} \right) \right] \quad (9.82)$$

where dy is the increment of Compton optical depth, $dy = (kT_e/m_e c^2) \sigma_T N_e c dt$, and $x = \hbar\omega/kT_e$.

Let us analyse the meanings of the various terms in (9.82), following the presentations of Liedahl (1999) and Blandford (1990). In the process of Comptonisation, the total number of photons is conserved, although their energies are changed by Compton scattering. Consequently, the conservation equation for the total number of photons is

$$\frac{d}{dt} \int_0^\infty \omega^2 n(\omega) d\omega = 0 \quad \text{or} \quad \int_0^\infty x^2 \frac{dn(x)}{dt} dx = 0, \quad (9.83)$$

and so the evolution of the photon spectrum corresponds to the conservation of a photon ‘fluid’ in phase space. There is therefore a continuity equation describing the flow of photons in phase space which can be written

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J} = 0, \quad (9.84)$$

where \mathbf{J} is the ‘current’ of photons in phase space. Now, the present analysis assumes that the distribution of photons is isotropic in phase space and so we need the divergence in spherical polar coordinates with only the radial x -component of the divergence present. It follows that

$$\frac{\partial n}{\partial t} = -\nabla \cdot \mathbf{J} = -\frac{1}{x^2} \frac{\partial}{\partial x} [x^2 J(x)], \quad (9.85)$$

where $J(x)$ is a scalar function of ‘radius’ x . This equation can be compared with the Kompaneets equation (9.82). It follows that

$$J(x) = -N_e \sigma_T c \frac{kT_e}{m_e c^2} x^2 \left(n + n^2 + \frac{\partial n}{\partial x} \right). \quad (9.86)$$

This equation enables us to understand the meanings of the various terms in the Kompaneets equation. Consider first the term

$$J(x) = -N_e \sigma_T c \frac{kT_e}{m_e c^2} x^2 n. \quad (9.87)$$

This term corresponds to the recoil effect described by (9.47),

$$\frac{d\omega}{\omega} = -\frac{\hbar\omega}{m_e c^2} \quad \text{or} \quad \frac{dx}{x} = -x \frac{kT_e}{m_e c^2}. \quad (9.88)$$

Just as the current or flux of particles in real space is $J = Nv$, so the current associated with the drift of photons in phase space because of the recoil effect is $J(x) = n dx/dt$. The rate of change of x is given by the number of scatterings per second times the average

change in x per scattering,

$$\frac{dx}{dt} = -x^2 \frac{kT_e}{m_e c^2} \times N_e \sigma_{TC} . \quad (9.89)$$

Therefore, the flux of photons $J(x)$ is

$$J(x) = n \frac{dx}{dt} = -N_e \sigma_{TC} \frac{kT_e}{m_e c^2} x^2 n , \quad (9.90)$$

exactly the same as (9.87). This analysis immediately enables us to understand the term in n^2 in (9.86). The term $n + n^2 = n(1 + n)$ and so the n^2 term takes account of the effects of induced scattering when the occupation number n is greater than unity.

The third term in (9.86) corresponds to the statistical increase of energy of the photons by Compton scattering. As discussed in Sect. 9.2.2, the ‘heating’ of the photon gas by the hotter electrons is a statistical phenomenon which corresponds to the diffusion of the photons in phase space. The equation governing this process is of the same form as that encountered in the diffusion-loss equation (7.42) which includes the statistical acceleration term. The corresponding term in the case of photon diffusion is that the current $J(x)$ in momentum space is described by a diffusion coefficient D_x so that $J(x) = D_x \partial n / \partial x$. In the case of the isotropic diffusion of photons in phase space, the transfer equation is

$$\frac{\partial n}{\partial t} = -\frac{1}{x^2} \frac{\partial J(x)}{\partial x} . \quad (9.91)$$

The diffusion coefficient D_x in this case is the mean square change in x per unit time, $\langle(\Delta x)^2\rangle$, the same as is found in the stochastic acceleration of particles according to the diffusion loss equation (7.42). In the present case, the change of energy of the photon is given by the non-relativistic limit of (9.31), $\hbar\omega' = \hbar\omega(1 + (v/c) \cos\theta)$, or $\Delta x = x(v/c) \cos\theta$. Averaging the mean square energy change over 4π steradians,

$$\langle(\Delta x)^2\rangle = x^2 \frac{v^2}{c^2} \int \cos^2\theta \times \frac{1}{2} \sin\theta d\theta = \frac{1}{3} x^2 \frac{v^2}{c^2} . \quad (9.92)$$

Setting $\frac{3}{2}kT_e = \frac{1}{2}m_e v^2$, we find the variance of Δx in a single scattering,

$$\langle(\Delta x)^2\rangle = x^2 \frac{kT_e}{m_e c^2} . \quad (9.93)$$

There are $N_e \sigma_{TC}$ scatterings per unit time and so, since the variance per unit time is the sum of the separate variances, we find

$$D(x) = \langle(\Delta x)^2\rangle = N_e \sigma_{TC} x^2 \left(\frac{kT_e}{m_e c^2} \right) . \quad (9.94)$$

Therefore the diffusion current has the form

$$J(x) = -N_e \sigma_{TC} x^2 \left(\frac{kT_e}{m_e c^2} \right) \frac{\partial n}{\partial x} , \quad (9.95)$$

exactly the same as the third term in the Kompaneets equation (9.86).

It is interesting to note that the Kompaneets equation has the same formal content as (9.51), which refers to the energy of an individual photon. The importance of the

Kompaneets equation is that it describes the evolution of the spectrum of the photon field in phase space and necessarily includes induced processes.

Generally, the solutions of the equation have to be found numerically, but there are a number of cases in which analytic solutions can be found. First of all, it is a useful exercise to show that the right-hand side of (9.82) is zero for a Bose–Einstein distribution for which the occupation number is $n = [\exp(x + \mu) - 1]^{-1}$. This solution also includes the case of the Planck spectrum for which $\mu = 0$. These solutions are found when there is time for the system to come to equilibrium under Compton scattering in the limit of very large values of y .

Pozdnyakov and his colleagues provide examples of the spectra of X-ray sources for increasing values of the Compton optical depth y (Fig. 9.9) (Pozdnyakov *et al.*, 1983). In these examples, the input photons are of very low energy and the electrons have temperature $kT_e = 25$ keV. The Thomson scattering optical depth τ takes values between 3 and 10, so that the Comptonisation process does not reach saturation, although the beginnings of the formation of the Wien peak are seen at the largest optical depths. At smaller optical depths, the spectrum mimics very closely a power-law spectrum up to energies $h\nu \approx kT_e$, above which a roughly exponential cut-off is found. It is helpful to illustrate how these features come about.

Following the presentation of Liedahl, it is convenient to modify the Kompaneets equation to take account of the source of low energy photons and the diffusion of photons out of the source region. Both terms can be taken to have similar forms to those found in the diffusion-loss equation (7.42). The rate of production of soft photons $Q(x)$ can be written $Q(x) = Q_0(x)$ for photons with values of $x \leq x_0$ and $Q(x) = 0$ for $x > x_0$, so that there are no initial photons with values of $x > x_0$. The *escape time* from the source region is determined by the optical depth of the region to Thomson scattering τ_{es} . As discussed in Sect. 9.4.1, the number of scatterings is given by the greater of τ_{es} or τ_{es}^2 , or in terms of the Compton optical depth, by $y_{\text{es}} = (kT_e/m_e c^2) \max(\tau_{\text{es}}, \tau_{\text{es}}^2)$. Therefore the modified Kompaneets equation can be written

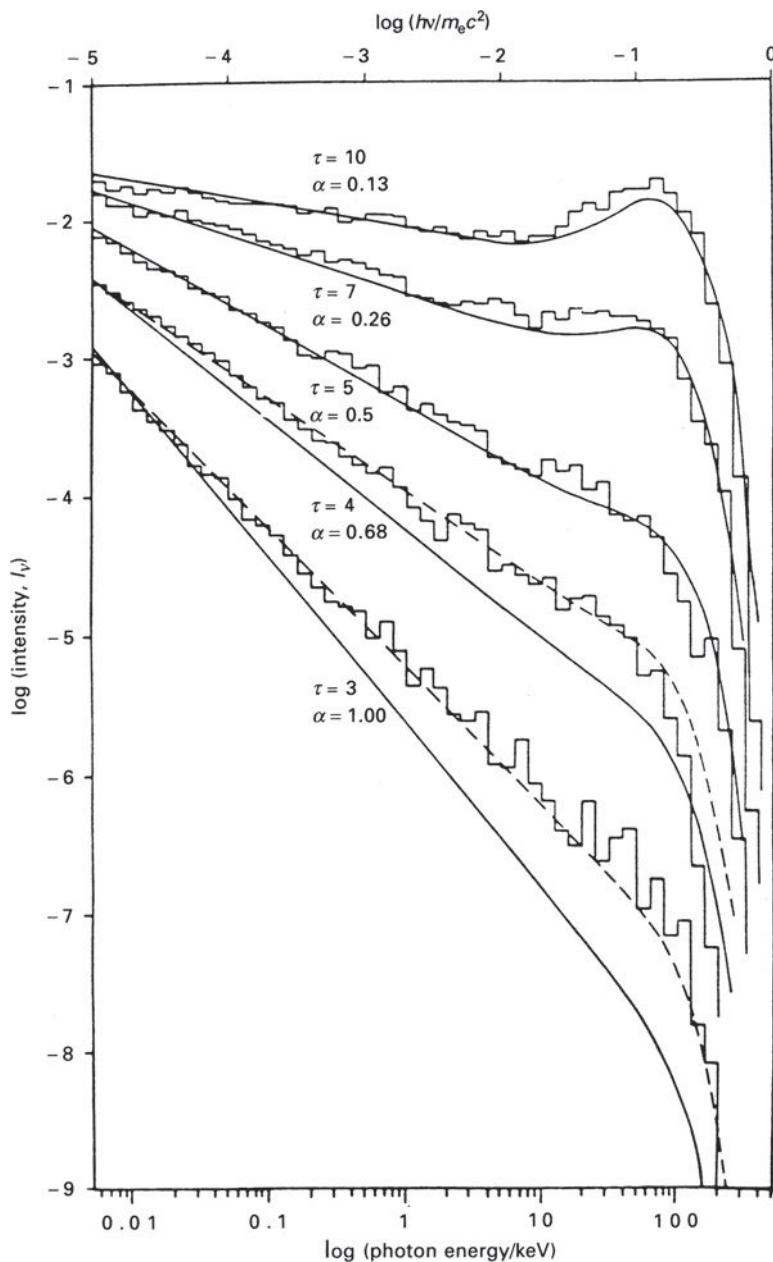
$$\frac{\partial n}{\partial y} = \frac{1}{x^2} \frac{\partial}{\partial x} \left[x^4 \left(n + n^2 + \frac{\partial n}{\partial x} \right) \right] + Q(x) - \frac{n}{y_{\text{es}}} , \quad (9.96)$$

where the source term $Q(x)$ is defined as the number of photons per unit element of phase space per unit Compton optical depth.

We are interested in steady-state solutions for photon energies $x \gg x_0$ and so $\partial n / \partial y = 0$ and $Q(x) = 0$. For our present purposes, we are interested in cases in which the photon occupation number n is very small and so we can neglect the induced Compton scattering terms in n^2 . With these simplifications, the modified Kompaneets equation becomes

$$y_{\text{es}} \frac{\partial}{\partial x} \left[x^4 \left(n + \frac{\partial n}{\partial x} \right) \right] - nx^2 = 0 . \quad (9.97)$$

Let us first consider the case of very large values of $x = \hbar\omega/kT_e \gg 1$. Then, since the occupation number n is very small, the term nx^2 in (7.97) is very much less than the first

**Fig. 9.9**

The Comptonisation of low frequency photons in a spherical plasma cloud having $kT_e = 25$ keV. τ is the Thomson scattering optical depth $\tau = N_e \sigma_T$ and α is the spectral index. The solid curves are analytic solutions of the Kompaneets equation using the parameters given by the relations (9.102) and (9.103) (Pozdnyakov *et al.*, 1983); the results of Monte Carlo simulations of the Compton scattering process are shown by the histograms and there is generally good agreement with the analytic solutions. A slightly better fit to the Monte Carlo calculations is found for the cases $\tau = 3$ and $\tau = 4$ if the analytic formula is fitted to the spectral index α found from the Monte Carlo simulations (dashed curve). These computations illustrate the development of the Wien peak for large values of the optical depth τ at energies $h\nu \approx kT_e$.

two terms. In this approximation, the first integral of (7.97) is

$$\left(\frac{\partial n}{\partial x} + n \right) = \frac{\text{constant}}{x^4}. \quad (9.98)$$

For large values of x , the right-hand side tends to zero and so the solution for n tends to $n = e^{-x}$. Converting the occupation number to an intensity of radiation using (9.75), we find

$$I(\omega) d\omega = \frac{\hbar\omega^3}{\pi^2 c^2} n(\omega) d\omega = \frac{\hbar\omega^3 e^{-\hbar\omega/kT_e}}{\pi^2 c^2} d\omega. \quad (9.99)$$

This is *Wien's law* in the limit $\hbar\omega \gg kT_e$ and accounts for the exponential cut-off of the Comptonised spectrum at these high energies.

For small values of x , the diffusion of photons in phase space results in heating of the photon gas, and the recoil effect, which results in a loss of energy of the photons, can be neglected. In this case, the Kompaneets equation becomes

$$y_{\text{es}} \frac{\partial}{\partial x} \left[x^4 \left(\frac{\partial n}{\partial x} \right) \right] - nx^2 = 0. \quad (9.100)$$

Inspection of (9.100) shows that power-law solutions of the form $n(x) = A x^m$ can be found. It is straightforward to find the value of m ,

$$m = -\frac{3}{2} \pm \left(\frac{9}{4} + \frac{1}{y_{\text{es}}} \right)^{1/2}, \quad (9.101)$$

and hence the intensity spectrum has the form $I(\omega) \propto \omega^{3+m}$. The positive root of (9.101) is appropriate if $y_{\text{es}} \gg 1$ and the negative root if $y_{\text{es}} \ll 1$. Thus, if the Compton optical depth y_{es} is very large, the value of m is zero and then we recover the Wien spectrum $I(\omega) \propto \omega^3$ in the limit $\hbar\omega \ll kT_e$. If $y_{\text{es}} \ll 1$, a power-law spectrum is obtained with an exponential cut-off at high energies, the solutions having to be joined together numerically. This is an intriguing example of a power-law spectrum being created through 'thermal' processes rather than being ascribed to some 'non-thermal' radiation mechanism involving ultra-relativistic electrons.

The above calculation is the simplest example of the formation of a power-law spectrum by purely thermal processes. The predicted power-law index is sensitive to the geometry of the source. For example, Pozdnyakov and his colleagues derived an improved version of the above calculation in which the predicted spectral index is

$$m = -\frac{3}{2} - \left(\frac{9}{4} + \gamma \right)^{1/2}, \quad (9.102)$$

where, for spherical geometry,

$$\gamma = \frac{\pi^2}{3} \frac{m_e c^2}{\left(\tau + \frac{2}{3} \right)^2 k T_e}, \quad (9.103)$$

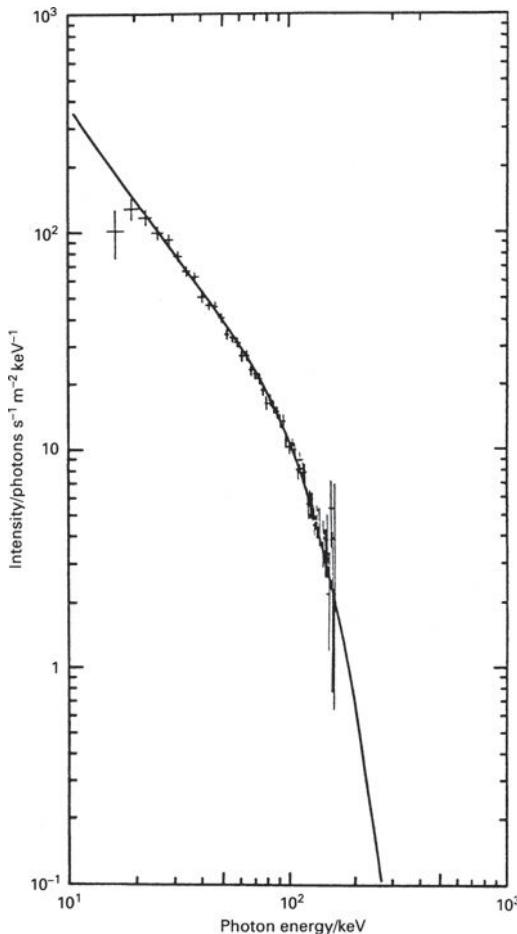


Fig. 9.10 The hard X-ray spectrum of the Galactic X-ray source Cygnus X-1 observed in a balloon flight of the Max Planck Institute for Extraterrestrial Physics on 20 September 1977 compared with the analytic solution of the Kompaneets equation with parameters $\tau_0 = 5$, $kT_e = 27$ keV (Sunyaev and Titarchuk, 1980).

where τ is the Thomson optical depth from the centre to the edge of the cloud. For a disc geometry,

$$\gamma = \frac{\pi^2}{12} \frac{m_e c^2}{(\tau + \frac{2}{3})^2 k T_e}, \quad (9.104)$$

where now τ is the Thomson optical depth from the centre to the surface of the disc. The theoretical curves shown by the solid lines in Fig. 9.9 have been obtained using the relations (9.102) and (9.103).

In a number of hard X-ray sources, a characteristic power-law spectrum with a high energy exponential cut-off is observed. Pozdnyakov and his colleagues have fitted the hard X-ray spectrum of the source Cygnus X-1 by such a form of spectrum with the parameters given in the caption of Fig. 9.10. Similar spectra have been observed for a number of

hard X-ray sources which contain neutron stars or black holes from observations by the XMM-Newton X-ray observatory and the INTEGRAL γ -ray observatory.

9.5 The Sunyaev–Zeldovich effect

An important application of the Kompaneets equation concerns spectral distortions of the Cosmic Microwave Background Radiation if the radiation traverses extensive regions of hot ionised gas with electron temperature T_e much greater than the radiation temperature T_{rad} . Compton scattering leads to distortions of the thermal spectrum of the background radiation if there are no additional sources of photons to match the number required for a Planck spectrum. Such conditions can occur in the pre- and post-recombination phases of the standard Big Bang. There are two convenient ways of describing the degree to which the observed spectrum differs from that of a perfect black-body, both of them discussed by Zeldovich and Sunyaev in the late 1960s (for details of their work, see Sunyaev and Zeldovich (1980)).

If there were injection of thermal energy into the intergalactic gas prior to the epoch of recombination at $z \approx 1000$ and the number of photons was conserved, the spectrum would relax to an equilibrium Bose–Einstein intensity spectrum with a finite dimensionless chemical potential μ ,

$$I_\nu = \frac{2h\nu^3}{c^2} \left[\exp\left(\frac{h\nu}{kT_r} + \mu\right) - 1 \right]^{-1}, \quad (9.105)$$

as discussed in Sect. 9.4.3. Such an injection of energy might have been associated with matter–antimatter annihilation or with the dissipation of primordial fluctuations and turbulence.

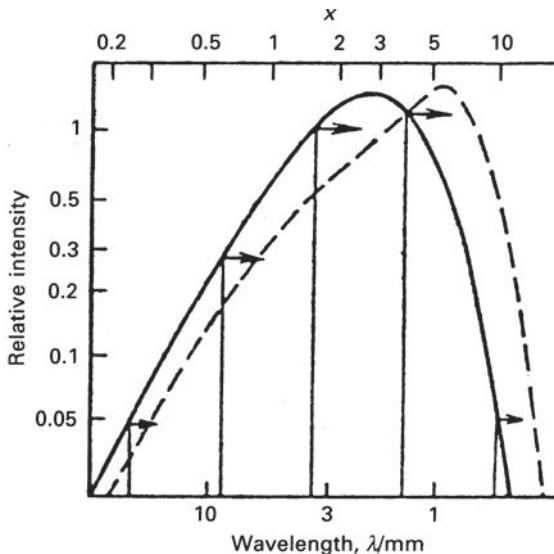
If the heating took place after the epoch of recombination, there would not be time to set up the equilibrium distribution and the predicted spectrum is found by solving the Kompaneets equation without the terms describing the cooling of the photons, that is,

$$\frac{\partial n}{\partial y} = \frac{1}{x^2} \frac{\partial}{\partial x} \left(x^4 \frac{\partial n}{\partial x} \right). \quad (9.106)$$

Assuming the distortions are small, Zeldovich and Sunyaev inserted the trial solution $n = (e^x - 1)^{-1}$ into the right-hand side of (9.106). It is straightforward to show that

$$\frac{\Delta n}{n} = \frac{\Delta I(\omega)}{I(\omega)} = y \frac{x e^x}{e^x - 1} \left(x \frac{e^x + 1}{e^x - 1} - 4 \right), \quad (9.107)$$

where the Compton optical depth is $y = \int (kT_e/m_e c^2) \sigma_T N_e dl$ (Zeldovich and Sunyaev, 1969). The effect of Compton scattering is to shift the spectrum to higher energies with the result that the intensity of radiation in the Rayleigh–Jeans region of the spectrum, $x \ll 1$,

**Fig. 9.11**

Illustrating the Compton scattering of a Planck distribution by hot electrons in the case in which the Compton optical depth is $y = \int (kT_e/m_e c^2) \sigma_T N_e dl = 0.15$. The intensity decreases in the Rayleigh–Jeans region of the spectrum and increases in the Wien region (Sunyaev, 1980).

decreases while that at $x \gg 1$ increases – the change-over occurs at $x = 4$ (Fig. 9.11). Expanding (9.107) for small values of x , the fractional decrease in intensity is

$$\frac{\Delta I(\omega)}{I(\omega)} = -2y. \quad (9.108)$$

In this process, the total energy in the radiation spectrum increases as the photons gain energy from the hot electrons. The increase of energy in the background radiation can be found from (9.59) for small values of y . Therefore, the increase in energy density of the background radiation is

$$\frac{\Delta \varepsilon_r}{\varepsilon_r} = e^{4y}. \quad (9.109)$$

The net result is that there is more energy in the background radiation than would be predicted from the measured temperature in the Rayleigh–Jeans region of the spectrum. Another way of expressing this result is to use the fact that

$$\frac{dI(\omega)}{I(\omega)} = \frac{dT_{RJ}}{T_{RJ}} = -2y, \quad (9.110)$$

and so $T_{RJ} = e^{-2y} T_0$. Consequently, if the radiation temperature of the background radiation is measured to be T_{RJ} , the total energy density is predicted to be $\varepsilon = a T_{RJ}^4 e^{12y}$.

The precision with which the observed spectrum of the Cosmic Microwave Background Radiation fits a perfect black-body spectrum therefore sets strong upper limits to the values of μ and y . The precise spectral measurements made by the FIRAS instrument of the

Cosmic Background Explorer result in the following limits (Page, 1997):

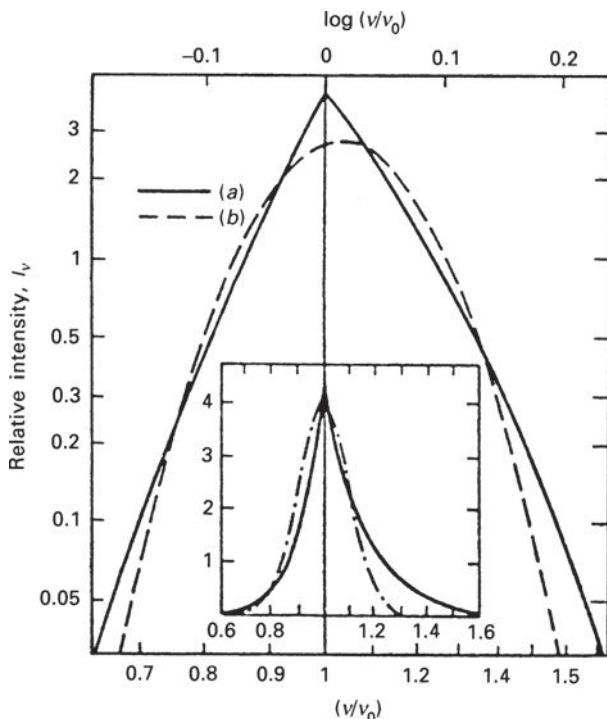
$$|y| \leq 1.5 \times 10^{-5}, \quad |\mu| \leq 10^{-4}. \quad (9.111)$$

These limits are of astrophysical importance in the study of the physics of the intergalactic gas, as well as constraining the amount of star and metal formation which could have taken place in young galaxies.

The importance of the Sunyaev–Zeldovich effect in causing decrements in the Cosmic Microwave Background Radiation because of the presence of hot gas in rich clusters of galaxies has already been discussed in Sect. 4.6 and illustrated by the remarkable radio maps of clusters with redshifts up to $z \approx 1$ in Fig. 4.10. These observations were made in the Rayleigh–Jeans region of the spectrum of the Cosmic Microwave Background Radiation and so provide a measure of the quantity $y = \int (kT_e/m_e c^2) \sigma_T N_e dl$. In conjunction with observations of the bremsstrahlung emission of the hot intracluster gas, the physical parameters of the hot gas cloud can be determined and, as explained in Sect. 4.6, enable estimates of Hubble’s constant to be made.

It is useful to make a pedagogical remark about the origin of the result $dI(\omega)/I(\omega) = -2y$. The result (9.50) shows that the average increase in energy of the photons in the Compton scattering process is $\Delta\varepsilon/\varepsilon = 4kT_e/m_ec^2$ and so naive application of this energy change results in the wrong answer for the amplitude of the decrement in that, since $I(\omega) \propto \omega^2$, an intensity decrement of $-8y$ would be expected. The reason for this discrepancy is the statistical nature of the Compton scattering process. Figure 9.12 shows the probability distribution of scattered photons in a single Compton scattering (Sunyaev, 1980). The average increase in energy is of order $(v/c)^2$, that is, second order in v/c , compared with the breadth of the wings of the scattering function which are of order v/c . Therefore, in addition to the increase in energy due to the second-order effect in $(v/c)^2$, we also have to take account of the scattering of photons by first-order Compton scatterings. In the Rayleigh–Jeans limit, in which the spectrum is $I_\omega \propto \omega^2$, there are more photons scattered down in energy to frequency ω than are scattered up from lower frequencies. These Doppler scatterings increase the intensity at frequency ω by an increment $+6y$ so that the net decrement is $-2y$, as given by the Kompaneets equation. This digression illustrates the power of the Kompaneets equation in automatically taking account of the statistical aspects of the diffusion of photons in phase space.

The spectral signature of the Sunyaev–Zeldovich effect has a distinctive form over the peak of the spectrum of the Cosmic Microwave Background Radiation and is given in the first-order approximation by (9.107). The exact shape of this function has been the subject of a number of studies which take full account of special relativistic effects and expand the Kompaneets equation to higher orders in $\partial n/\partial x$. The results of numerical solutions of the Boltzmann equation and further analytic studies are summarised by Challinor and Lasenby (1998) (Fig. 9.13). Notice that the results are presented in terms of the absolute change in intensity $\Delta I(\omega)$ which tends to zero in the high and low frequency limits. This form of distortion has been measured in a number of Abell clusters in the SuZIE experiment carried out at the CalTech Submillimetre Observatory on Mauna Kea (Benson *et al.*, 2004). Figure 9.14 shows that the expected change in sign of the Sunyaev–Zeldovich effect on either side of the frequency $\hbar\omega/kT_e = 4$ has been clearly detected.

**Fig. 9.12**

The probability distribution of photons scattered in a single Compton scattering (a) using the exact expression for Compton scattering (solid line) and (b) using the diffusion term in the Kompaneets equation (dashed line) for the case in which the hot gas has temperature $kT_e = 5.1 \text{ keV}$ or $kT_e/m_e c^2 = 0.01$. The insert shows these distributions on a linear scale. It can be seen that the distributions are broad with half-widths $\sigma \sim (kT_e/m_e c^2)^{1/2}$, that is, $\Delta\omega/\omega \sim 0.1$. The average increase in energy of the photon is $\Delta\hbar\omega/\omega = 4(kT_e/m_e c^2) = 0.04$ (Sunyaev, 1980).

9.6 Synchrotron–self-Compton radiation

The physics of inverse Compton scattering was discussed in Sect. 9.3 and is an important source of high energy radiation whenever large fluxes of photons and relativistic electrons occupy the same volume. The case of special interest in this section is that in which the relativistic electrons which are the source of low energy photons are also responsible for scattering these photons to X- and γ -ray energies, the process known as *synchrotron–self-Compton radiation*. A case of special importance is that in which the energy density of low energy photons is so great that most of the energy of the electrons is lost by synchrotron–self-Compton rather than by synchrotron radiation. This is likely to be the source of the ultra-high energy γ -rays observed in some of the most extreme active galactic nuclei.

We can derive some of the essential features of synchrotron–self-Compton radiation from the formulae we have already derived. The ratio η of the rates of loss of energy of an ultra-relativistic electron by synchrotron and inverse Compton radiation in the presence of

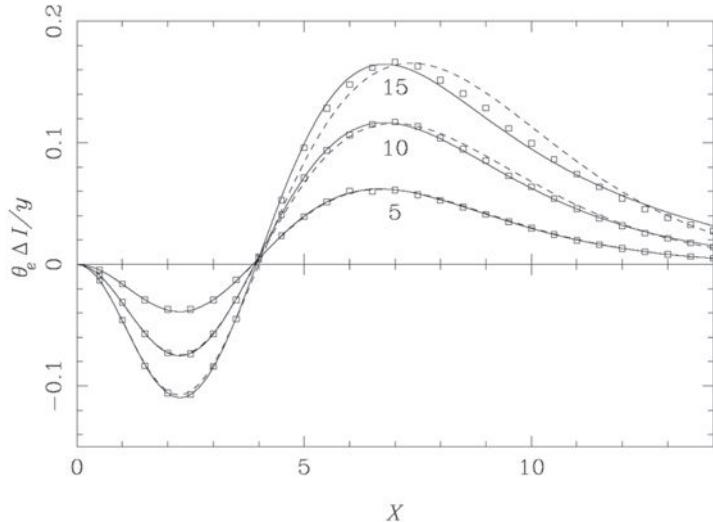


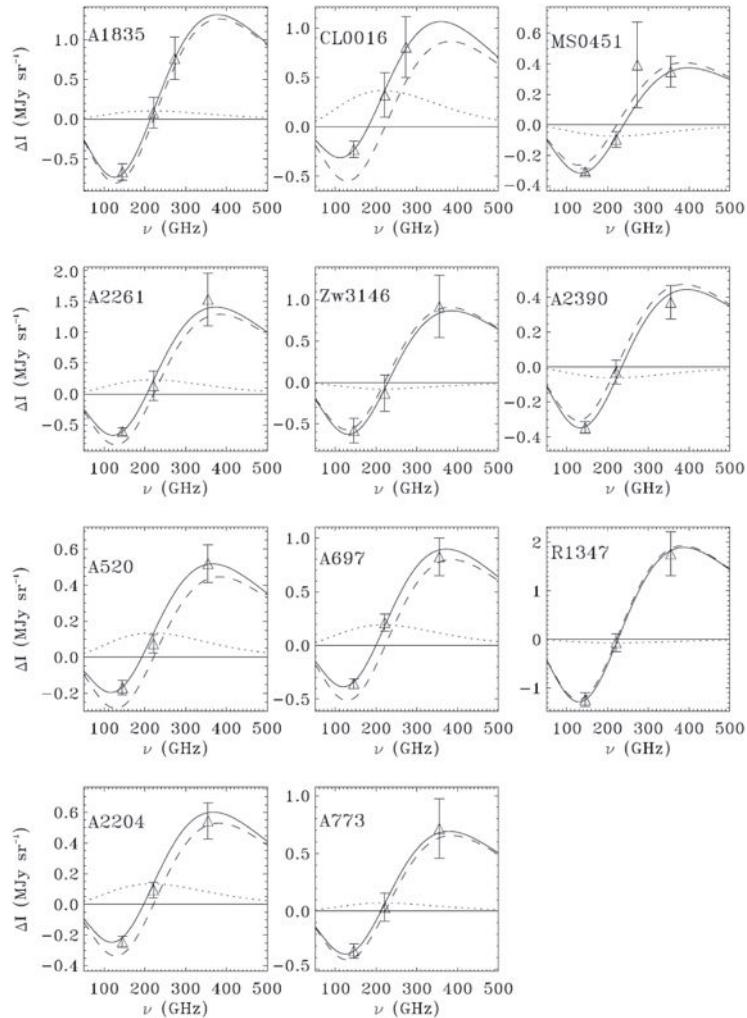
Fig. 9.13 Intensity change in units of $2(kT_0)^3/(hc)^2$, plotted against $X = \hbar\omega/kT_0$ for three values of kT_e (in keV), where $\theta_e = kT_e/m_ec^2$. The solid curves are calculated using the second-order correction to the Kompaneets equation, while the dashed lines are calculated from the first-order correction. The points are the result of a Monte Carlo evaluation of the Boltzmann collision integral by Garrett and Gull (Challinor and Lasenby, 1998).

a photon energy density u_{rad} and a magnetic field of magnetic flux density B are given by the formulae (9.42) and (9.41):

$$\eta = \frac{(dE/dt)_{\text{IC}}}{(dE/dt)_{\text{sync}}} = \frac{u_{\text{photon}}}{B^2/2\mu_0}. \quad (9.112)$$

Thus, if the synchrotron radio flux density and the X- and γ -radiation from the same source region are observed, estimates of the magnetic flux density within the source region can be made, the only problems being the upper and lower limits to the electron energy spectrum and in ensuring that electrons of roughly the same energies are responsible for the radio and X-ray emission. A good example of this procedure for estimating the magnetic flux density in the hot-spot regions in the powerful double radio source Cygnus A is discussed in Sect. 22.2.

The *synchro-Compton catastrophe* occurs if the ratio η is greater than 1. In this case, low energy radio photons produced by synchrotron radiation are scattered to X-ray energies by the same relativistic electrons. Since η is greater than 1, the energy density of the X-rays is greater than that of the radio photons and so the electrons suffer even greater energy losses by scattering these X-rays to γ -ray energies. In turn, these γ -rays have a greater energy density than the X-rays . . . , and so on. It can be seen that as soon as η becomes greater than one, the energy of the electrons is lost at the very highest energies and so the radio source should be a very powerful source of X- and γ -rays. Before considering the higher order scatterings, let us study the first stage of the process for a compact source of synchrotron radiation, so compact that the radiation is self-absorbed.

**Fig. 9.14**

The observed Sunyaev–Zeldovich spectrum associated with hot gas in clusters of galaxies. In each plot the solid line is the best-fit model for the spectral distortions, the dashed line is the thermal component of the Sunyaev–Zeldovich effect and the dotted line is the kinematic component (Benson *et al.*, 2004). The kinematic component is associated with first-order Compton scattering due to the peculiar motions of the clusters.

First of all, the energy density of radiation within a synchrotron self-absorbed radio source is estimated. As shown in Sect. 8.7, the flux density of such a source is

$$S_v = \frac{2kT_b}{\lambda^2} \Omega \quad \text{where} \quad \Omega \approx \theta^2 = \frac{r^2}{D^2} \quad \text{and} \quad \gamma m_e c^2 = 3kT_e = 3kT_b , \quad (9.113)$$

where Ω is the solid angle subtended by the source, r its physical size and D its distance. T_e is the thermal temperature equivalent to the energy of a relativistic electron with total energy $\gamma m_e c^2$. As explained in Sect. 8.7, for a self-absorbed source, the electron temperature of

the relativistic electrons is equal to the brightness temperature T_b of the source, $T_e = T_b$. The radio luminosity of the source in W Hz^{-1} is therefore

$$L_\nu = 4\pi D^2 S_\nu \approx \frac{8\pi k T_b}{\lambda^2} r^2. \quad (9.114)$$

L_ν is the luminosity per unit bandwidth and so, to order of magnitude, the bolometric luminosity is roughly νL_ν . Therefore, the energy density of radiation u_{photon} is

$$u_{\text{photon}} \sim \frac{L_\nu \nu}{4\pi r^2 c} = \frac{2k T_b \nu}{\lambda^2 c}, \quad (9.115)$$

and the ratio η is

$$\eta = \frac{u_{\text{photon}}}{B^2/2\mu_0} = \frac{\left(\frac{2k T_b \nu}{\lambda^2 c}\right)}{\left(\frac{B^2}{2\mu_0}\right)} = \frac{4k T_b \nu \mu_0}{\lambda^2 c B^2}. \quad (9.116)$$

We now use the theory of synchrotron self-absorbed sources to express the magnetic flux density B in terms of observables. Repeating the calculations carried out in Sect. 8.7,

$$\nu_g \approx \nu/\gamma^2 \quad \text{and} \quad 3kT_b = 3kT_e = \gamma m_e c^2 \quad \text{where} \quad \nu_g = eB/2\pi m_e. \quad (9.117)$$

Reorganising these relations, we find

$$B = \frac{2\pi m_e}{e} \left(\frac{m_e c^2}{3k T_b} \right)^2 \nu. \quad (9.118)$$

Therefore, the ratio of the loss rates, η , is

$$\eta = \frac{(dE/dt)_{\text{IC}}}{(dE/dt)_{\text{sync}}} = \left(\frac{81e^2 \mu_0 k^5}{\pi^2 m_e^6 c^{11}} \right) \nu T_b^5. \quad (9.119)$$

This is the key result. The ratio of the loss rates depends very strongly upon the brightness temperature of the radio source. Substituting the values of the constants, the *critical brightness temperature* for which $\eta = 1$ is

$$T_b = T_e = 10^{12} \nu_9^{-1/5} \text{ K}, \quad (9.120)$$

where ν_9 is the frequency at which the brightness temperature is measured in units of 10^9 Hz, that is, in GHz. According to this calculation, no compact radio source should have brightness temperature greater than $T_b \approx 10^{12}$ K without suffering catastrophic inverse Compton scattering losses, if the emission is incoherent synchrotron radiation.

The most compact sources, studied by very long baseline interferometry (VLBI) at centimetre wavelengths, have brightness temperatures less than the synchrotron–self-Compton limit, typically, the values found being $T_b \approx 10^{11}$ K. These observations in themselves provide direct evidence that the radiation is the emission of relativistic electrons since the temperature of the emitting electrons must be at least 10^{11} K. This is not, however, the whole story. If the time-scales of variability τ of the compact radio sources are used to estimate their physical sizes, $l \sim c\tau$, the source regions must be considerably smaller than those inferred from the VLBI observations, and brightness temperatures exceeding 10^{12} K

are found. It is likely that relativistic beaming is the cause of this discrepancy, a topic taken up in Chap. 23.

Models of synchrotron–self-Compton sources are best worked out numerically and are strongly dependent upon the input assumptions. A good impression of the forms of spectra expected is provided by the computations of Band and Grindlay (1985), who take account of the transfer of radiation within the self-absorbed source and consider both homogeneous and inhomogeneous cases. A number of important refinements are included in their computations. Of particular importance is the use of the Klein–Nishina cross-section at relativistic energies, $\hbar\omega \geq 0.5$ MeV, rather than the Thomson cross-section for photon–electron scattering. In the ultra-relativistic limit, the cross-section tends to

$$\sigma_{\text{KN}} = \frac{\pi^2 r_e^2}{h\nu} \left(\ln 2h\nu + \frac{1}{2} \right), \quad (9.121)$$

and so decreases as $(\hbar\omega)^{-1}$ at high energies. Consequently, higher order scatterings result in significantly reduced luminosities as compared with the non-relativistic calculation. Many features of such computations can be understood from Fig. 9.15a and b. The homogeneous source has the standard form of spectrum at radio frequencies, namely, a power-law distribution in the optically thin spectral region $L_\nu \propto \nu^{-\alpha}$, while, in the optically thick region, the spectrum has the self-absorbed form $L_\nu \propto \nu^{5/2}$. The relativistic boosting of the spectrum of the radio emission from the compact radio source is clearly seen. Both the low and high frequency spectral features of the radio source spectrum follow the relativistic ‘boosting’ relations

$$\nu_g \rightarrow \gamma^2 \nu_g \rightarrow \gamma^4 \nu_g \dots \quad (9.122)$$

These features are most apparent in the case of the homogeneous source. The higher order scatterings for photon energies $h\nu \gg m_e c^2$ are significantly reduced because of the use of the Klein–Nishina cross-section at high energies.

In the case of the inhomogeneous source, the magnetic field strength and number density of relativistic electrons decrease outwards as power laws, resulting in a much broader ‘synchrotron-peak’. As a result, only one Compton scattering is apparent because of the wide range of photon energies produced by the radio source.

These computations assume that the source of radiation is stationary. As will be discussed in Chap. 23, the extreme ultra-high energy γ -ray sources, which are variable over short time-scales, display many of the features expected of synchrotron–self-Compton radiation, but they must also involve relativistic bulk motion of the source regions to account for their extreme properties. As a consequence, the predictions of the models are somewhat model-dependent.

9.7 Cherenkov radiation

When a fast particle moves through a medium at a constant velocity v greater than the speed of light in that medium, it emits *Cherenkov radiation*. The process finds application in the

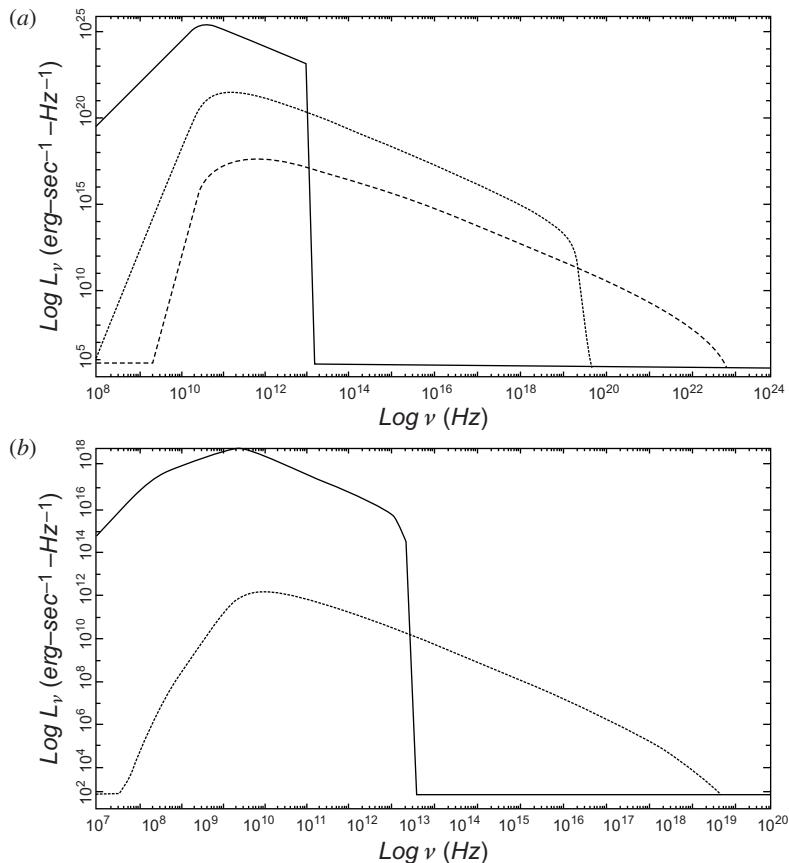


Fig. 9.15 Examples of the spectra of the synchrotron–self-Compton radiation of compact radio sources. (a) The ‘standard’ synchrotron–self-Compton spectrum of a homogeneous source with magnetic flux density 5×10^{-4} T and electron number density $N_e(\gamma) d\gamma = 4 \gamma^{-3} d\gamma m^{-3}$ in a spherical source of radius 2×10^{11} m. The solid line is the synchrotron radio spectrum, the small-dashed line the first scatterings and the large-dashed line the second scatterings. (b) The spectrum of the inhomogeneous model with inner and outer radii $r_1 = 10^9$ m and $r_2 = 10^{10}$ m, within which the magnetic flux density varies as $B = 10^{-4} (r/r_1)^{-2}$ T and the electron number density as $N_e(\gamma) d\gamma = \gamma^{-3} (r/r_1)^{-2} d\gamma m^{-3}$ for $1 \leq \gamma \leq 10^4$ (Band and Grindlay, 1985).

construction of *threshold detectors* in which Cherenkov radiation is only emitted if the particle has velocity greater than c/n . If the particles pass through, for example, lucite or plexiglass, for which $n \approx 1.5$, only those with $v > 0.67c$ emit Cherenkov radiation which can be detected as an optical signal. Particles with extreme relativistic energies can be detected in gas Cherenkov detectors in which the refractive index n of the gas is just greater than 1. A second application is in the detection of ultra-high energy γ -rays when they enter the top of the atmosphere. The high energy γ -ray initiates an electron–photon cascade (see Sect. 9.9) and, if the electron–positron pairs acquire velocities greater than the speed of light in air, optical Cherenkov radiation is emitted which can be detected by light detectors at sea-level.

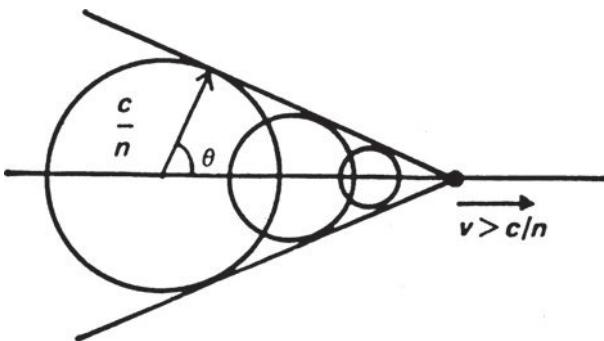


Fig. 9.16 Illustrating Huygens' construction for the wavefront of coherent radiation of a charged particle moving at constant velocity $v > c/n$ through a medium with refractive index n .

The origin of the emission is best appreciated from the expressions (6.19), the *Liénard–Wiechert potentials* $\mathbf{A}(\mathbf{r}, t)$ and $\phi(\mathbf{r}, t)$ which are repeated here:

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \left[\frac{q\mathbf{v}}{1 - (\mathbf{v} \cdot \mathbf{i}_{\text{obs}})/c} \right]_{\text{ret}} ; \quad \phi(\mathbf{r}, t) = \frac{1}{4\pi \epsilon_0 r} \left[\frac{q}{1 - (\mathbf{v} \cdot \mathbf{i}_{\text{obs}})/c} \right]_{\text{ret}}, \quad (9.123)$$

where \mathbf{i}_{obs} is the unit vector in the direction of observation from the moving charge. In the case of a vacuum, one of the standard results of electromagnetic theory is that a charged particle moving at constant velocity \mathbf{v} does not radiate electromagnetic radiation. As shown in Sect. 6.2, in a vacuum, radiation is emitted if the particle is accelerated. In the case of a medium with a finite permittivity ϵ , or refractive index n , however, the denominators of (9.123) become

$$[1 - (n\mathbf{v} \cdot \mathbf{i}_{\text{obs}})/c]_{\text{ret}}, \quad (9.124)$$

where n is the refractive index of the medium. It follows that the potentials become singular along the cone for which $1 - (n\mathbf{v} \cdot \mathbf{i}_{\text{obs}})/c = 0$, that is, for $\cos \theta = c/nv$. As a result, the usual rule that only accelerated charges radiate no longer applies.

The geometric representation of this process is that, because the particle moves superluminally through the medium, a ‘shock wave’ is created behind the particle. The wavefront of the radiation propagates at a fixed angle with respect to the velocity vector of the particle because the wavefronts only add up coherently in this direction according to Huygens’ construction (Fig. 9.16). The geometry of Fig. 9.16 shows that the angle of the wavevector with respect to the direction of motion of the particle is $\cos \theta = c/nv$.

Let us derive the main features of Cherenkov radiation in a little more detail. Consider an electron moving along the positive x -axis at a constant velocity v . This motion corresponds to a current density \mathbf{J} where¹

$$\mathbf{J} = ev \delta(x - vt) \delta(y) \delta(z) \mathbf{i}_x. \quad (9.125)$$

¹ Strictly speaking, we should multiply by N_e to create a current density, but N_e would cancel out when we revert to a single particle.

Now take the Fourier transform of this current density to find the frequency components $\mathbf{J}(\omega)$ corresponding to this motion.

$$\begin{aligned}\mathbf{J}(\omega) &= \frac{1}{(2\pi)^{1/2}} \int \mathbf{J} \exp(i\omega t) dt, \\ &= \frac{e}{(2\pi)^{1/2}} \delta(y) \delta(z) \exp(i\omega x/v) \mathbf{i}_x.\end{aligned}\quad (9.126)$$

This Fourier decomposition corresponds to representing the motion of the moving electron by a line distribution of coherently oscillating currents. Our task is to work out the coherent emission, if any, from this distribution of oscillating currents. The full treatments given in standard texts such as Jackson (1999) and Clemmow and Dougherty (1969) are quite complex. We adopt here an approach developed by John Peacock.

First, let us review some of the standard results concerning the propagation of electromagnetic waves in a medium of permittivity ϵ , or refractive index $n = \epsilon^{1/2}$. It is a standard result of classical electrodynamics that the flow of electromagnetic energy through a surface $d\mathbf{S}$ is given by the Poynting vector flux, $\mathbf{N} \cdot d\mathbf{S} = (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S}$. The electric and magnetic field strengths \mathbf{E} and \mathbf{H} are related to the electric flux density \mathbf{D} and the magnetic flux density \mathbf{B} by the constitutive relations

$$\mathbf{D} = \epsilon \epsilon_0 \mathbf{E}; \quad \mathbf{B} = \mu \mu_0 \mathbf{H}. \quad (9.127)$$

The energy density of the electromagnetic field in the medium is given by the standard formula

$$u = \int \mathbf{E} \cdot d\mathbf{D} + \int \mathbf{H} \cdot d\mathbf{B}. \quad (9.128)$$

If the medium has a constant real permittivity ϵ and permeability $\mu = 1$, the energy density in the medium is

$$u = \frac{1}{2} \epsilon \epsilon_0 E^2 + \frac{1}{2} \mu_0 H^2. \quad (9.129)$$

The speed of propagation of the waves is found from the dispersion relation $k^2 = \epsilon \epsilon_0 \mu_0 \omega^2$, that is, $c(\epsilon) = \omega/k = (\epsilon \epsilon_0 \mu_0)^{-1/2} = c/\epsilon^{1/2}$. This demonstrates the well-known result that, in a linear medium, the refractive index n is $\epsilon^{1/2}$. Another useful result is the relation between the \mathbf{E} and \mathbf{B} fields in the electromagnetic wave – the ratio E/B is $c/\epsilon^{1/2} = c/n$. Substituting this result into the expression for the electric and magnetic field energies (9.129), it is found that these are equal. Thus, the total energy density in the wave is $u = \epsilon \epsilon_0 E^2$. Furthermore, the Poynting vector flux $\mathbf{E} \times \mathbf{H}$ is $\epsilon^{1/2} \epsilon_0 E^2 c = n \epsilon_0 E^2 c$. This energy flow corresponds to the energy density of radiation in the wave $\epsilon \epsilon_0 E^2$ propagating at the velocity of light in the medium c/n . As is expected, $N = n \epsilon_0 E^2 c$. This is the result we have been seeking. It is similar to the formula used in Sects 6.2.2 and 6.2.3 but now the refractive index n is included in the right place.

We now write down the expressions for the *retarded* values of the current which contributes to the vector potential at the point \mathbf{r} (Fig. 9.17). From (6.17a), the expression for the vector potential \mathbf{A} due to the current density \mathbf{J} at distance \mathbf{r} is

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} d^3 r' = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{J}]}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \quad (9.130)$$

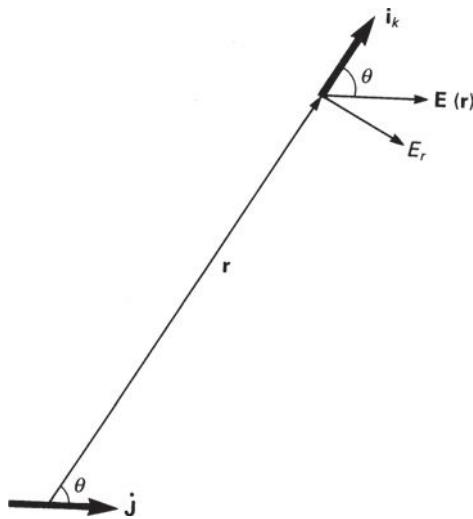


Fig. 9.17 Illustrating the geometry used in the derivation of the expressions for Cherenkov radiation.

where the square brackets refer to retarded potentials. Taking the time derivative,

$$\mathbf{E}(\mathbf{r}) = -\frac{\partial \mathbf{A}}{\partial t} = -\frac{\mu_0}{4\pi} \int \frac{[\mathbf{j}]}{|\mathbf{r} - \mathbf{r}'|} d^3 \mathbf{r}'. \quad (9.131)$$

In the far field limit, the electric field component E_r of the radiation field is perpendicular to the radial vector \mathbf{r} and so, as indicated in Fig. 9.17, $E_r = \mathbf{E}(\mathbf{r}) \times \mathbf{i}_k$, that is,

$$|E_r| = \frac{\mu_0 \sin \theta}{4\pi} \left| \int \frac{[\mathbf{j}]}{|\mathbf{r} - \mathbf{r}'|} d^3 \mathbf{r}' \right|. \quad (9.132)$$

This formula reduces to the expression (6.5) for the radiation of a point charge by the substitution $\int [\mathbf{j}] d^3 \mathbf{r}' = e\ddot{\mathbf{r}}$.

We now go through the same procedure described in Sect. 6.2.5 to evaluate the frequency spectrum of the radiation. First of all, we work out the total radiation rate by integrating the Poynting vector flux over a sphere at a large distance \mathbf{r} ,

$$\begin{aligned} \left(\frac{dE}{dt} \right)_{\text{rad}} &= \int_S n c \epsilon_0 E_r^2 dS, \\ &= \int_{\Omega} \frac{n c \epsilon_0 \mu_0^2 \sin^2 \theta}{16\pi^2} \left| \int \frac{[\mathbf{j}]}{|\mathbf{r} - \mathbf{r}'|} d^3 \mathbf{r}' \right|^2 r^2 d\Omega. \end{aligned} \quad (9.133)$$

We now assume that the size of the emitting region is much smaller than the distance to the point of observation, $L \ll r$. Therefore, we can write $|\mathbf{r} - \mathbf{r}'| = r$ and then,

$$\left(\frac{dE}{dt} \right)_{\text{rad}} = \int \frac{n \sin^2 \theta}{16\pi^2 \epsilon_0 c^3} \left| \int [\mathbf{j}] d^3 \mathbf{r}' \right|^2 d\Omega. \quad (9.134)$$

Now, we take the time integral of the radiation rate to find the total radiated energy,

$$E_{\text{rad}} = \int_{-\infty}^{\infty} \left(\frac{dE}{dt} \right)_{\text{rad}} dt = \int_{-\infty}^{\infty} \int_{\Omega} \frac{n \sin^2 \theta}{16\pi^2 \epsilon_0 c^3} \left| \int [\mathbf{j}] d^3 r' \right|^2 d\Omega dt . \quad (9.135)$$

We use Parseval's theorem to transform from an integral over time to one over frequency. Noting, as in Sect. 6.2.5, that we are only interested in positive frequencies, we find

$$E_{\text{rad}} = \int_0^{\infty} \int_{\Omega} \frac{n \sin^2 \theta}{8\pi^2 \epsilon_0 c^3} \left| \int [\mathbf{j}(\omega)] d^3 r' \right|^2 d\Omega d\omega . \quad (9.136)$$

Let us now evaluate the volume integral $\int [\mathbf{j}(\omega)] d^3 r'$. We take \mathbf{R} to be the vector from the origin of the coordinate system to the observer and \mathbf{x} to be the position vector of the current element $\mathbf{j}(\omega) d^3 r'$ from the origin. Thus, $\mathbf{r}' = \mathbf{R} - \mathbf{x}$. Now the waves from the current element at \mathbf{x} propagate outwards from the emitting region at velocity c/n with phase factor $\exp[i(\omega t - \mathbf{k} \cdot \mathbf{r}')]$ and therefore, relative to the origin at O , the phase factor of the waves, which we need to find the retarded value of $\mathbf{j}(\omega)$, is

$$\exp[i(\omega t - \mathbf{k} \cdot (\mathbf{R} - \mathbf{x}))] = \exp(-i\mathbf{k} \cdot \mathbf{R}) \exp[i(\omega t + \mathbf{k} \cdot \mathbf{x})] . \quad (9.137)$$

Therefore, evaluating $[\mathbf{j}(\omega)]$, we find

$$\left| \int [\mathbf{j}(\omega)] d^3 r' \right| = \left| i\omega \int [\mathbf{J}(\omega)] d^3 r' \right| .$$

Now we include the retarded component of $\mathbf{J}(\omega)$ explicitly by including the phase factor,

$$\left| \int [\mathbf{J}(\omega)] d^3 r' \right| = \left| \int \omega \exp[i(\omega t + \mathbf{k} \cdot \mathbf{x})] \mathbf{J}(\omega) d^3 r' \right| .$$

Using (9.126), we find

$$\begin{aligned} \left| \int [\mathbf{J}(\omega)] d^3 r' \right| &= \left| \frac{\omega e}{(2\pi)^{1/2}} \exp(i\omega t) \int \exp \left[i \left(\mathbf{k} \cdot \mathbf{x} + \frac{\omega x}{v} \right) \right] dx \right| , \\ &= \left| \frac{\omega e}{(2\pi)^{1/2}} \int \exp \left[i \left(\mathbf{k} \cdot \mathbf{x} + \frac{\omega x}{v} \right) \right] dx \right| . \end{aligned} \quad (9.138)$$

This is the key integral in deciding whether or not the particle radiates. If the electron propagates in a vacuum, $\omega/k = c$ and we can write the exponent

$$kx(\cos \theta + \omega/v) = kx(\cos \theta + c/v) . \quad (9.139)$$

Since, in a vacuum, $c/v > 1$, this exponent is always greater than zero and hence the exponential integral over all x is always zero. This means that a particle moving at constant velocity in a vacuum does not radiate.

If, however, the medium has refractive index n , $\omega/k = c/n$ and then the exponent is zero if $\cos \theta = -c/nv$. This is the origin of the Cherenkov radiation phenomenon. The radiation is only coherent along the angle θ corresponding to the Cherenkov cone derived from Huygens' construction. We can therefore write down formally the energy spectrum by using (9.67) recalling that the radiation is only emitted at an angle $\cos \theta = c/nv$. We

therefore find from (9.136)

$$\begin{aligned}\frac{dE_{\text{rad}}}{d\omega} &= \int_{\Omega} \frac{n\omega^2 e^2 \sin^2 \theta}{16\pi^3 \epsilon_0 c^3} \left| \int \exp \left[ikx \left(\cos \theta + \frac{\omega}{kv} \right) \right] dx \right|^2 d\Omega, \\ &= \frac{n\omega^2 e^2}{16\pi^3 \epsilon_0 c^3} \left(1 - \frac{c^2}{n^2 v^2} \right) \int_{\Omega} \left| \int \exp \left[ikx \left(\cos \theta + \frac{\omega}{kv} \right) \right] dx \right|^2 d\Omega.\end{aligned}$$

We now evaluate the integral. Let us write $k(\cos \theta + \omega/kv) = \alpha$. The integral therefore becomes

$$\int_{\theta} \left| \int \exp(i\alpha x) dx \right|^2 2\pi \sin \theta d\theta. \quad (9.140)$$

Let us take the line integral along a finite path length from $-L$ to L . It should be noted that there is a problem in evaluating the integral of a function which only has finite value at a specific value of θ from $-\infty$ to $+\infty$. This is why the normal derivation involves the use of contour integration to get rid of the infinites. The integral should be taken over a small finite range of angles about $\theta = \cos^{-1}(c/nv)$ for which $(\cos \theta + \omega/kv)$ is close to zero. Therefore, we can integrate over all values of θ (or α) knowing that most of the integral is contributed by values of θ very close to $\cos^{-1}(c/nv)$. Therefore, the integral becomes

$$8\pi \int \frac{\sin^2 \alpha L}{\alpha^2} \frac{d\alpha}{k}. \quad (9.141)$$

Taking the integral over all values of α from $-\infty$ to $+\infty$, we find that the integral becomes $(8\pi c/n\omega)\pi^2 L$. Therefore the energy per unit bandwidth is

$$\frac{du}{d\omega} = \frac{\omega e^2}{2\pi \epsilon_0 c^3} \left(1 - \frac{c^2}{n^2 v^2} \right) L. \quad (9.142)$$

We now ought to take the limit $L \rightarrow \infty$. However, there is no need to do this since we obtain directly the energy loss rate per unit path length by dividing by $2L$. Therefore, the loss rate per unit path length is

$$\frac{du(\omega)}{dx} = \frac{\omega e^2}{4\pi \epsilon_0 c^3} \left(1 - \frac{c^2}{n^2 v^2} \right). \quad (9.143)$$

Since the particle is moving at velocity v , the energy loss rate per unit bandwidth is

$$I(\omega) = \frac{du(\omega)}{dt} = \frac{\omega e^2 v}{4\pi \epsilon_0 c^3} \left(1 - \frac{c^2}{n^2 v^2} \right). \quad (9.144)$$

Notice that the intensity of radiation depends upon the variation of the refractive index with frequency $n(\omega)$.

9.8 Electron–positron pair production

If the photon has energy greater than $2m_e c^2$, pair production can take place in the field of the nucleus. Pair production cannot take place in free space because momentum and energy

cannot be conserved simultaneously. To demonstrate this, consider a photon of energy $\hbar\omega$ decaying into an electron–positron pair, each of which has kinetic energy $(\gamma - 1)m_e c^2$. The best one can do to conserve both energy and momentum is if the electron–positron pair moves parallel to the original direction of the photon, then

$$\begin{aligned}\text{Conservation of energy: energy of photon} &= \hbar\omega = 2\gamma m_e c^2, \\ \text{momentum of pair} &= 2\gamma m_e v = (\hbar\omega/c)(v/c).\end{aligned}$$

But,

$$\text{initial momentum of photon} = \hbar\omega/c.$$

Since v cannot be equal to c , we cannot conserve both energy and momentum in free space and this is why we need a third body, such as a nucleus, which can absorb some of the energy or momentum.

Let us quote some useful results for electron–positron pair production (Chupp, 1976; Ramana Murthy and Wolfendale, 1993):

Intermediate photon energies In the case of no screening, the cross-section for photons with energies in the range $1 \ll \hbar\omega/m_e c^2 \ll 1/\alpha Z^{1/3}$ can be written

$$\sigma_{\text{pair}} = \alpha r_e^2 Z^2 \left[\frac{28}{9} \ln \left(\frac{2\hbar\omega}{m_e c^2} \right) - \frac{218}{27} \right] \text{ m}^2 \text{ atom}^{-1}. \quad (9.145)$$

r_e is the classical electron radius and α the fine structure constant.

Ultra-relativistic limit In the case of complete screening and for photon energies $\hbar\omega/m_e c^2 \gg 1/\alpha Z^{1/3}$, the cross-section becomes

$$\sigma_{\text{pair}} = \alpha r_e^2 Z^2 \left[\frac{28}{9} \ln \left(\frac{183}{Z^{1/3}} \right) - \frac{2}{27} \right] \text{ m}^2 \text{ atom}^{-1}. \quad (9.146)$$

In both cases, the cross-section for pair production is $\sim \alpha \sigma_T Z^2$. Notice also that the cross-section for the creation of pairs through interactions with electrons is very much smaller than the above values and can be neglected.

Exactly as in Sect. 6.6, we define a radiation length ξ_{pair} for pair production

$$\xi_{\text{pair}} = \rho / N_i \sigma_{\text{pair}} = M_A / N_0 \sigma_{\text{pair}}, \quad (9.147)$$

where M_A is the atomic mass, N_i is the number density of nuclei and N_0 is Avogadro's number. If we compare the radiation lengths for pair production and bremsstrahlung by ultra-relativistic electrons, we find that $\xi_{\text{pair}} \approx \xi_{\text{brems}}$. This reflects the similarity of the Feynman diagrams for the bremsstrahlung and pair production mechanisms according to quantum electrodynamics (Leighton, 1959).

We can now put together the three main loss processes for high energy photons – ionisation losses, Compton scattering and electron–positron pair production – to obtain the total mass absorption coefficient for X-rays and γ -rays. Figure 9.18 shows how each of these processes contributes to the total absorption coefficient in lead. Notice that the energy range $500 \text{ keV} \lesssim \hbar\omega \lesssim 5 \text{ MeV}$ is a complex energy range for the experimental study of photons from cosmic sources because all three processes make a significant contribution

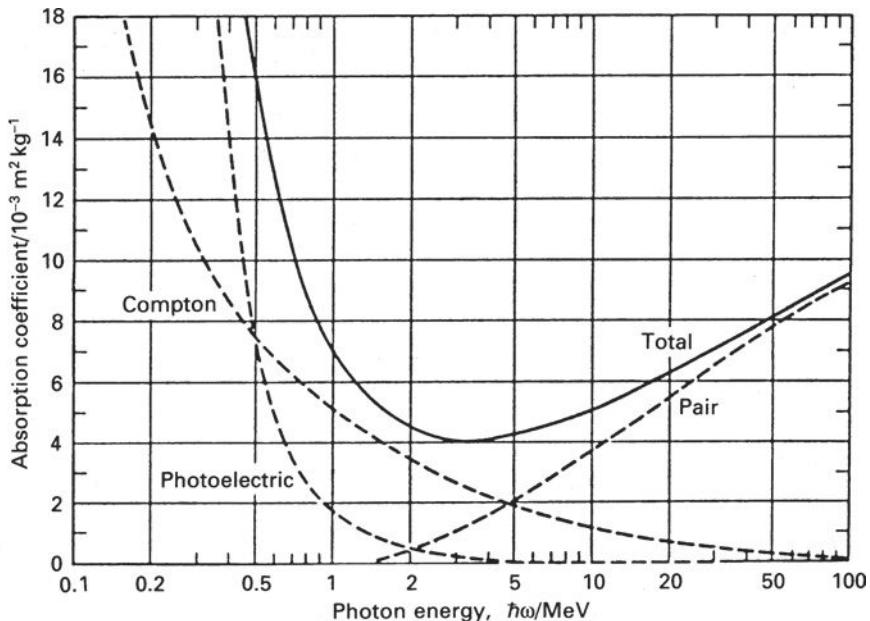


Fig. 9.18 The total mass absorption coefficient for high energy photons in lead, indicating the contributions associated with the photoelectric absorption, Compton scattering and electron–positron pair production (Enge, 1966).

to the absorption coefficient for γ -rays. Consequently, this is a particularly difficult energy range for the design and construction of γ -ray telescopes. To make matters worse, the fluxes of photons from astrophysical sources are generally low in this energy range.

9.9 Electron–photon cascades, electromagnetic showers and the detection of ultra-high energy γ -rays

We can now understand how cascades, or showers, initiated by high energy electrons or γ -rays can come about. When, for example, a high energy photon enters the upper atmosphere, it generates an electron–positron pair, each of which in turn generates high energy photons by bremsstrahlung, each of which generates an electron–positron pair, each of which . . . , and so on.

Let us build a simple model of an electron–photon cascade in the following way. In the ultra-relativistic limit, the radiation lengths for pair production and bremsstrahlung are the same, as discussed in Sect. 9.8. Therefore the probability of these processes taking place is one-half at path length ξ given by

$$\exp(-\xi/\xi_0) = \frac{1}{2} \quad \text{or} \quad \xi = R = \xi_0 \ln 2. \quad (9.148)$$

Therefore, if the cascade is initiated by a γ -ray of energy E_0 , after a distance of, on average R , an electron–positron pair is produced. For simplicity, it is assumed that the pair share

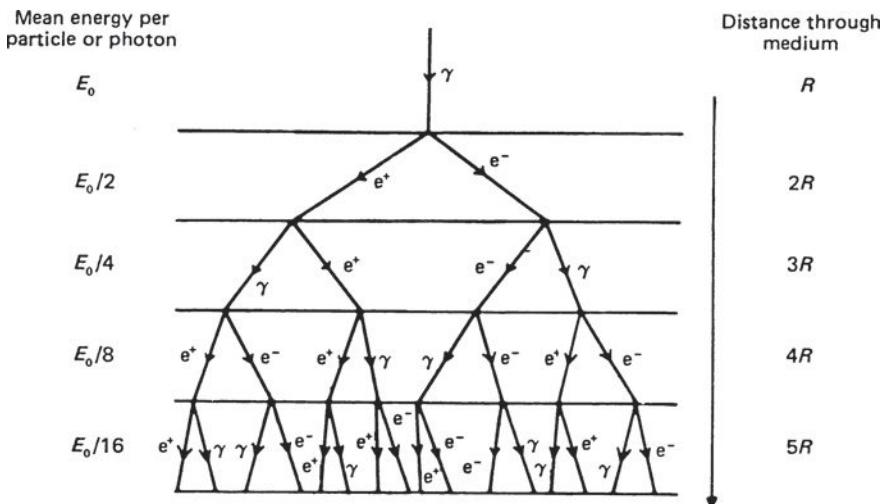


Fig. 9.19 A simple model for an electromagnetic shower.

the energy of the γ -ray, that is, $E_0/2$ each. In the next length R , the electron and positron lose, on average, half their energy and they each radiate a photon of energy $E_0/4$. Thus, we end up with two particles and two photons, all having energy $E_0/4$ after distance $2R$. This process is repeated as illustrated in Fig. 9.19 as the energy of the photons and particles is degraded through the atmosphere.

After distance nR , the number of (photons + electrons + positrons) is 2^n and their average energy is $E_0/2^n$. On average, the shower consists of $\frac{2}{3}$ positrons and electrons and $\frac{1}{3}$ photons. The cascade eventually terminates when the average energy per particle drops to the *critical energy* E_c , below which the dominant loss process for the electrons is ionisation losses rather than bremsstrahlung. This process produces copious quantities of electron-ion pairs but they are all of very low energy. In addition, with decreasing energy, the production cross-section for pairs decreases until it becomes of the same order as that for Compton scattering and photoelectric absorption, as illustrated in Fig. 9.18. Thus, the shower reaches its maximum development when the average energy of the cascade particles is about E_c . The number of high energy photons and particles is roughly E_0/E_c and the number of radiation lengths n_c over which this occurs is

$$n_c = \frac{\ln(E_0/E_c)}{\ln 2}. \quad (9.149)$$

At larger depths, the number of particles falls off dramatically because of ionisation losses which become catastrophic once the electrons become non-relativistic. These simple arguments give some impression of what needs to be included in a proper calculation. Appropriate cross-sections for different energy ranges have to be used and integrations carried out over all possible products with the relevant probability distributions. Among the first calculations to illustrate these features were the pioneering efforts of Rossi and Greisen shown in Fig. 9.20a. These calculations confirm the predictions of the simple model, namely, that the initial growth is exponential, that the maximum number of particles

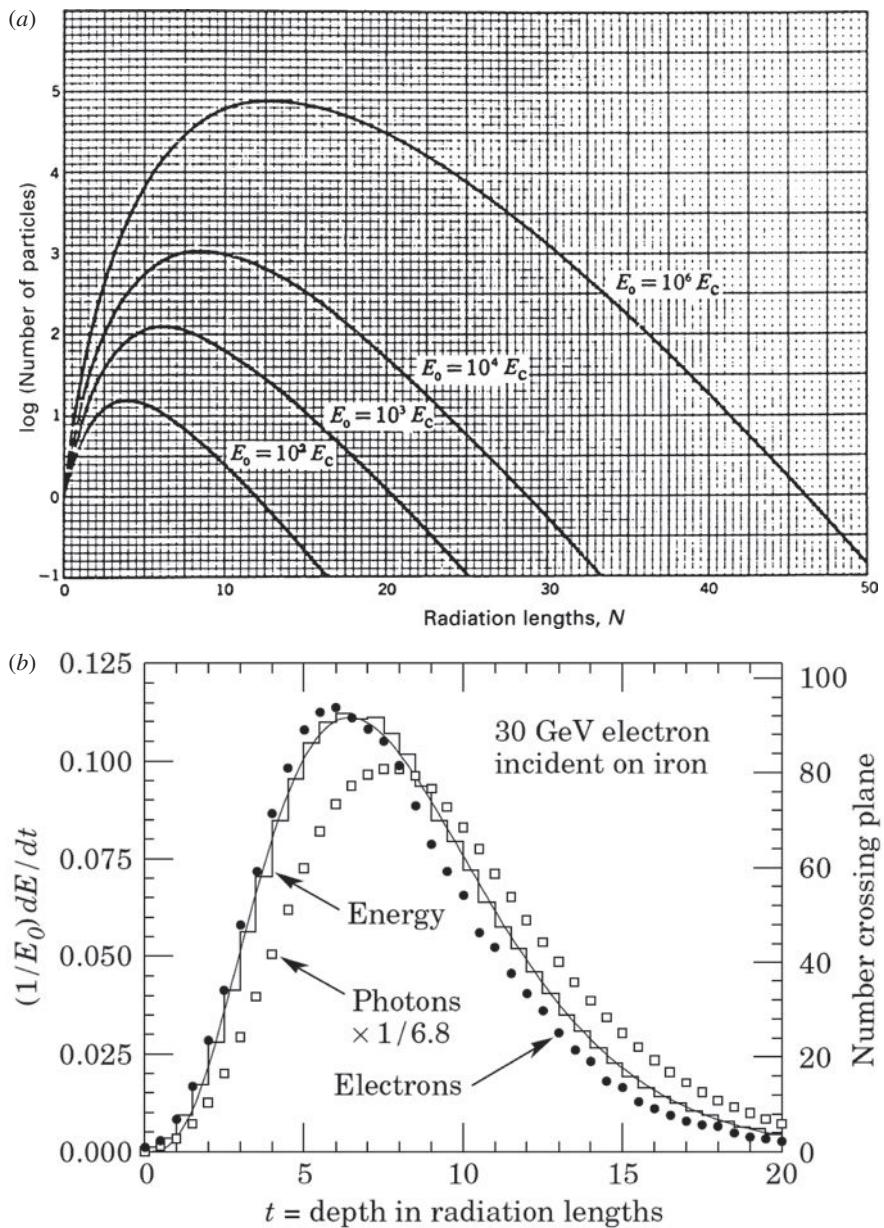


Fig. 9.20 (a) The total number of particles in a shower initiated by an electron of energy E_0 as a function of depth through the medium measured in radiation lengths N ; E_0 is the critical energy (Rossi and Greisen, 1941). (b) A more recent computation of the development of an electromagnetic shower in iron (Amsler *et al.*, 2008).

is proportional to E_0 and that after maximum development, there is a rapid attenuation of the electron flux. These computations have been considerably enhanced since these early results because of the need for precise understanding of the development of such showers which are central to the Auger ultra-high energy cosmic ray observatory and the new generation of ground-based ultra-high energy γ -ray telescopes. A more recent example of the development of an electromagnetic shower in iron is shown in Fig. 9.20b (Amsler *et al.*, 2008). Amsler and his colleagues also provide further details about the properties of electromagnetic showers.

An important feature of these results is that the showers consist only of electrons, positrons and γ -rays – there are no muons, pions and other debris produced. This helps distinguish the arrival of high energy γ -rays from other types of particle. These electron–photon cascades, or electromagnetic showers, were among the first high interactions to be detected inside cloud chambers. These showers also accompany the nuclear cascades which are considered in Chap. 10.

9.10 Electron–positron annihilation and positron production mechanisms

Perhaps the most extreme form of energy loss mechanism for electrons is annihilation with their antiparticles, the positrons. Particle–antiparticle annihilation results in the production of high energy photons and, conversely, high energy photons can collide with ambient photons to produce particle–antiparticle pairs.

There are several sources of positrons in astronomical environments. Perhaps the simplest is the decay of positively charged pions π^+ described in Sect. 10.4. The pions are created in collisions between cosmic ray protons and nuclei and the interstellar gas, roughly equal numbers of positive, negative and neutral pions being created. Since the π^0 's decay into γ -rays, the flux of interstellar positrons created by this process can be estimated from the γ -ray luminosity of the interstellar gas. A second process is the decay of long-lived radioactive isotopes created by explosive nucleosynthesis in supernova explosions. For example, the β^+ decay of ^{26}Al has a mean lifetime of 1.1×10^6 years. This element is formed in supernova explosions and so is ejected into the interstellar gas where the decay results in a flux of interstellar positrons. A third process is the creation of electron–positron pairs through the collision of high energy photons with the field of a nucleus (see Sect. 9.9).

Electron–positron pair production can also take place in *photon–photon collisions*. The threshold for this process can be worked out using similar procedures to those used in our discussion of Compton scattering. If \mathbf{P}_1 and \mathbf{P}_2 are the momentum four-vectors of the photons before the collision,

$$\mathbf{P}_1 = [\varepsilon_1/c, (\varepsilon_1/c) \mathbf{i}_1]; \quad \mathbf{P}_2 = [\varepsilon_2/c, (\varepsilon_2/c) \mathbf{i}_2], \quad (9.150)$$

Table 9.2 Threshold energies of ultra-high energy photons (ε_2) which give rise to electron–positron pairs in collision with photons of different energies (ε_1).

	ε_1 (eV)	ε_2 (eV)
Microwave Background Radiation	6×10^{-4}	4×10^{14}
Starlight	2	10^{11}
X-ray	10^3	3×10^8

conservation of four-momentum requires

$$\mathbf{P}_1 + \mathbf{P}_2 = \mathbf{P}_3 + \mathbf{P}_4 , \quad (9.151)$$

where \mathbf{P}_3 and \mathbf{P}_4 are the four-vectors of the created particles. To find the threshold for pair production, we require that the particles be created at rest and therefore

$$\mathbf{P}_3 = [m_e c, 0]; \quad \mathbf{P}_4 = [m_e c, 0] . \quad (9.152)$$

Squaring both sides of (9.151) and noting that $\mathbf{P}_1 \cdot \mathbf{P}_1 = \mathbf{P}_2 \cdot \mathbf{P}_2 = 0$ and that

$$\mathbf{P}_3 \cdot \mathbf{P}_3 = \mathbf{P}_4 \cdot \mathbf{P}_4 = \mathbf{P}_3 \cdot \mathbf{P}_4 = m_e^2 c^2 ,$$

then

$$\begin{aligned} \mathbf{P}_1 \cdot \mathbf{P}_1 + 2\mathbf{P}_1 \cdot \mathbf{P}_2 + \mathbf{P}_2 \cdot \mathbf{P}_2 &= \mathbf{P}_3 \cdot \mathbf{P}_3 + 2\mathbf{P}_3 \cdot \mathbf{P}_4 + \mathbf{P}_4 \cdot \mathbf{P}_4 , \\ 2\left(\frac{\varepsilon_1 \varepsilon_2}{c^2} - \frac{\varepsilon_1 \varepsilon_2}{c^2} \cos \theta\right) &= 4m_e^2 c^2 , \\ \varepsilon_2 &= \frac{2m_e^2 c^4}{\varepsilon_1(1 - \cos \theta)} , \end{aligned} \quad (9.153)$$

where θ is the angle between the incident directions of the photons. Thus, if electron–positron pairs are created, the threshold for the process occurs for head-on collisions, $\theta = \pi$, and hence,

$$\varepsilon_2 \geq \frac{m_e^2 c^4}{\varepsilon_1} = \frac{0.26 \times 10^{12}}{\varepsilon_1} \text{ eV} , \quad (9.154)$$

where ε_1 is measured in electron volts. This process thus provides not only a means for creating electron–positron pairs, for example in the vicinity of active galactic nuclei and hard X-ray sources, but also results an important source of opacity for high-energy γ -rays. Table 9.2 illustrates some of the examples we will encounter as our story unfolds. Photons with energies greater than those in the last column are expected to suffer some degree of absorption when they traverse regions containing large numbers of photons with the energies listed in the first column.

The cross-section for this process for head-on collisions in the ultra-relativistic limit is

$$\sigma = \pi r_e^2 \frac{m_e^2 c^4}{\varepsilon_1 \varepsilon_2} \left[2 \ln \left(\frac{2\bar{\varepsilon}}{m_e c^2} \right) - 1 \right] , \quad (9.155)$$

where $\bar{\varepsilon} = (\varepsilon_1 \varepsilon_2)^{1/2}$ and r_e is the classical electron radius. In the regime $\bar{\varepsilon} \approx m_e c^2$, the cross-section is

$$\sigma = \pi r_e^2 \left(1 - \frac{m_e^2 c^4}{\bar{\varepsilon}^2}\right)^{\frac{1}{2}} \quad (9.156)$$

(Ramana Murthy and Wolfendale, 1993). These cross-sections enable the opacity of the interstellar and intergalactic medium to be evaluated.

Electron–positron annihilation can proceed in two ways. In the first case, the electrons and positrons annihilate at rest or in flight through the interaction



When emitted at rest, the photons both have energy 0.511 MeV. When the particles annihilate ‘in flight’, meaning that they suffer a fast collision, there is a dispersion in the photon energies. It is a useful exercise in relativity to show that, if the positron is moving with velocity v and Lorentz factor γ , the centre of momentum frame of the collision has velocity $V = \gamma v(1 + \gamma)$ and that the energies of the pair of photons ejected in the direction of the line of flight of the positron and in the backward direction are

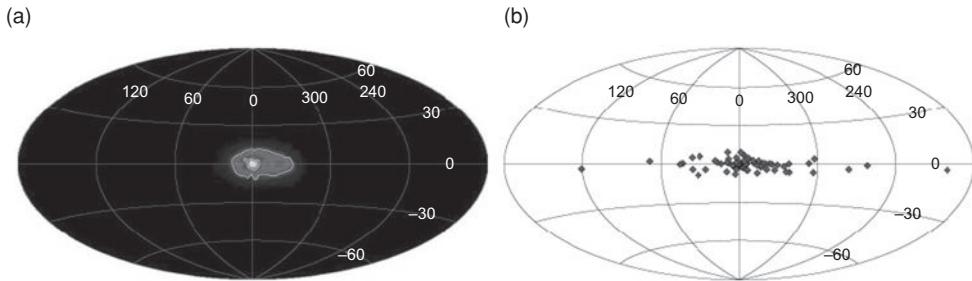
$$E = \frac{m_e c^2 (1 + \gamma)}{2} \left(1 \pm \frac{V}{c}\right). \quad (9.158)$$

From this result, it can be seen that the photon which moves off in the direction of the incoming positron carries away most of the energy of the positron and that there is a lower limit to the energy of the photon ejected in the opposite direction of $m_e c^2 / 2$.

If the velocity of the positron is small, *positronium atoms*, that is, bound states consisting of an electron and a positron, can form by radiative recombination; 25% of the positronium atoms form in the singlet 1S_0 state and 75% of them in the triplet 3S_1 state. The modes of decay from these states are different. The singlet 1S_0 state has a lifetime of 1.25×10^{-10} s and the atom decays into two γ -rays, each with energy 0.511 MeV. The majority triplet 3S_1 states have a mean lifetime of 1.5×10^{-7} s and three γ -rays are emitted, the maximum energy being 0.511 MeV in the centre of momentum frame. In this case, the decay of positronium results in a continuum spectrum to the low energy side of the 0.511 MeV line. If the positronium is formed from positrons and electrons with significant velocity dispersion, the line at 0.511 MeV is broadened, both because of the velocities of the particles and because of the low energy wing due to the continuum three-photon emission. This is a useful diagnostic tool in understanding the origin of the 0.511 MeV line. If the annihilations take place in a neutral medium with particle density less than 10^{21} m^{-3} , positronium atoms are formed. On the other hand, if the positrons collide in a gas at temperature greater than about 10^6 K, the annihilation takes place directly without the formation of positronium.

The cross-section for electron–positron annihilation in the extreme relativistic limit is

$$\sigma = \frac{\pi r_e^2}{\gamma} [\ln 2\gamma - 1]. \quad (9.159)$$

**Fig. 9.21**

(a) Observations of the whole sky in the 0.511 MeV electron–positron annihilation line made by the INTEGRAL γ -ray space observatory. (b) The right-hand panel shows the distribution of hard low mass X-ray binary stars. This stellar population has a distribution that matches the extent of the 511 keV map. (Courtesy of ESA, the Integral Science Team and G. Weidenspointner and his colleagues at the Max-Planck Institute for Extraterrestrial Physics, 2008.)

For thermal electrons and positrons, the cross-section becomes

$$\sigma \approx \frac{\pi r_e^2}{(v/c)} . \quad (9.160)$$

The 0.511 MeV electron–positron annihilation line has been detected from the direction of the Galactic Centre and observations by the ESA INTEGRAL γ -ray observatory have shown that the emission is extended along the Galactic plane with a spatial distribution similar to that of hard low mass X-ray binary stars (Fig. 9.21). We will have more to say about these observation and source of positrons as the story unfolds.

10.1 Nuclear interactions and high energy astrophysics

Nuclear physics is central to many branches of astrophysics, in particular to the understanding of the processes of energy generation in stars. In these cases, the nuclear processes occur deep in the centres of stars where the products of nucleosynthesis are generally only indirectly observable. The important exceptions to this statement are the observations of neutrinos from the Sun and the supernova SN 1987A (see Sects 2.6 and 13.1). We restrict attention here to nuclear processes in which the products of the nuclear interactions are directly observable. We need cross-sections to study the spallation reactions of high energy particles in the interstellar medium as well as production cross-sections and half-lives of radionuclides created in the spallation process and in sources of freshly synthesised material such as supernova remnants. We deal first with nuclear interactions associated with inelastic collisions of high energy protons and nuclei.

Nuclear interactions are only important when the incident high energy particle makes a more or less direct hit on the nucleus because the strong forces which hold the nucleus together are short range. Thus, the cross-section for nuclear interactions, in the sense that some form of interaction with the nucleons takes place, is just the geometric cross-section of the nucleus. A suitable expression for the radius of the nucleus is

$$R = 1.2 \times 10^{-15} A^{1/3} \text{ m}, \quad (10.1)$$

where A is the mass number. In many cases, the high energy particles have energies greater than 1 GeV. This introduces a further simplification since, at these energies, the de Broglie wavelength of the incident particle is small compared with the distance between nucleons in a nucleus. For example, the effective ‘size’ of an incident proton of energy 10 GeV can be estimated from Heisenberg’s uncertainty principle:

$$\Delta x \approx \hbar/p = \hbar/\gamma m_p v = 0.02 \times 10^{-15} \text{ m}. \quad (10.2)$$

We can therefore think of the incident proton as being a discrete, very small particle which interacts with the individual nucleons within the nucleus. The number of particles with which it interacts is just the number of nucleons along the line of sight through the nucleus. For example, a proton passing through an oxygen or nitrogen nucleus interacts, on average, with about $15^{1/3}$, that is, 2.5 of the nucleons. In fact, a reasonable model for the nuclear interactions is to consider that the incident proton undergoes multiple scattering within the nucleus.

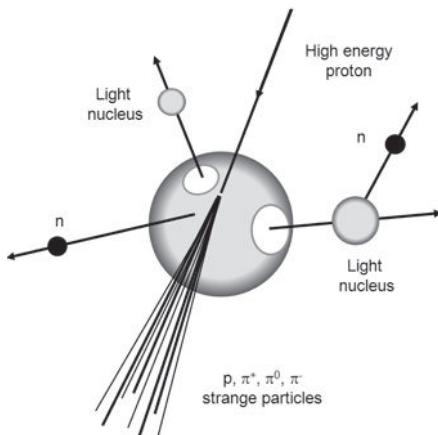


Fig. 10.1 A schematic diagram showing the principal products of the collision of a high energy proton with a nucleus.

The general picture of the interaction of a high energy proton with a nucleus can be described by the following rules.

- (i) The proton interacts strongly with an individual nucleon in a nucleus and, in the collision, pions of all charges, π^+ , π^- and π^0 , are the principal products. Strange particles may also be produced and occasionally antinucleons as well.
- (ii) In the centre of momentum frame of reference of the proton–nucleon encounter, the pions emerge mostly in the forward and backward directions but they may have lateral components of momentum of the order of $p_{\perp} \approx 100\text{--}200 \text{ MeV } c^{-1}$.
- (iii) The nucleons and pions involved in the strong interactions all possess very high forward momentum through the laboratory frame of reference and hence the products of the interaction are high energy particles.
- (iv) Each of the secondary particles is capable of initiating another collision inside the same nucleus, provided the initial collision occurred sufficiently close to the ‘front edge’ of the nucleus. Thus, a mini-nucleonic cascade is initiated inside the nucleus.
- (v) Only one or two nucleons participate in the nuclear interactions with the high energy particle and these are generally removed from the nucleus leaving it in a highly excited state. There is no guarantee that the resulting nucleus is a stable species. As a result, a variety of different outcomes may come about. Often several nuclear fragments are evaporated from the nucleus. These are called *spallation fragments* and we will have a great deal to say about them in the context of the origin of the light elements in the cosmic rays. These fragments are emitted in the frame of reference of the residual nucleus which is not given much forward momentum in the nuclear collision, virtually all of it going into tearing out the nucleons which interact with the high energy particle. Therefore, these spallation fragments are emitted more or less isotropically in the laboratory frame of reference. Neutrons are also evaporated from the ravished nucleus and other neutrons may be released from the spallation fragments. We recall that, for light nuclei, any imbalance between the numbers of neutrons and protons is fatal. These processes are summarised diagrammatically in Fig. 10.1. In high energy

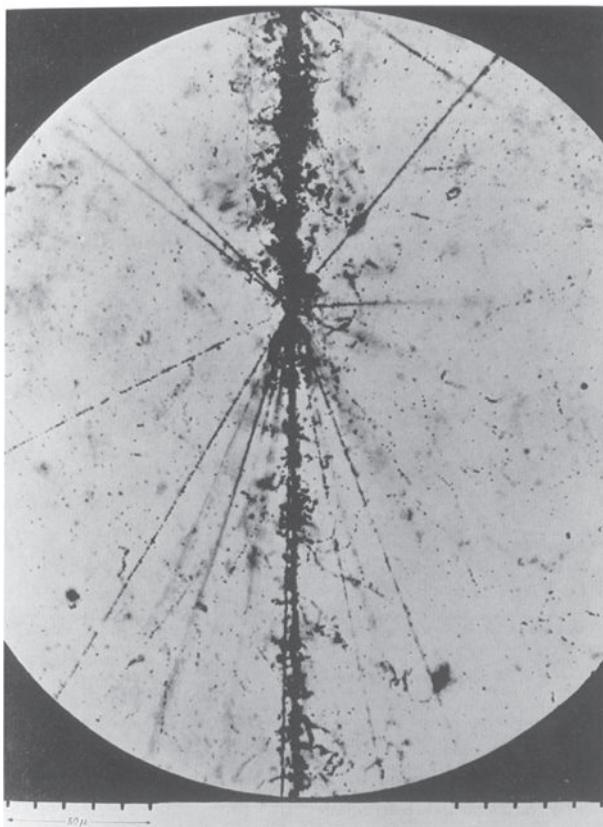


Fig. 10.2

The collision of a cosmic ray iron nucleus with a nucleus of a nuclear emulsion (Powell *et al.*, 1959).

collisions, the pions are concentrated in a rather narrow cone, the width of which is some measure of the energy of the incoming high energy particle.

From the radius for the nucleus (10.1), it is straightforward to work out the cross-section for the interaction of high energy particles with nuclei and show that the mean free path of a high energy proton in the atmosphere is about 800 kg m^{-2} , that is, very much less than the depth of the atmosphere which is about $10\,000 \text{ kg m}^{-2}$. In fact, because the proton often survives the interaction with some loss of energy, the flux of protons of a given energy falls off rather more slowly with path length. For particles of a given energy, the number density of protons falls off as $\exp -(x/L)$ where $L = 1200 \text{ kg m}^{-2}$.

For incident protons with energies greater than 1 GeV, a useful empirical rule is that, in collisions with air nuclei, roughly $2E^{1/4}$ new, high energy, charged particles are generated in the collision, where E is measured in GeV, although not necessarily all of them are pions. Pions of all charges are produced in almost equal numbers except at small energies at which charge conservation favours positively charged pions π^+ .

The most spectacular events occur when high energy nuclei undergo collisions with other heavy nuclei, for example, with the oxygen and nitrogen nuclei of our atmosphere or with the atoms of a nuclear emulsion. Figure 10.2 shows a rather impressive collision between a

cosmic ray iron nucleus and the nucleus of a nuclear emulsion. In such collisions, several pairs of nucleons undergo pion-producing collisions and not much is left of the target nucleus. This is quite a rare occurrence. Much more common are grazing encounters in which only a few nucleons interact to produce a shower of pions. The residual nuclei are left in an excited state and both eject spallation fragments as well as protons and neutrons. The important difference is that the incident high energy nucleus leaves with a stream of *relativistic* spallation fragments, protons and neutrons. This is important from several points of view. First of all, the high energy fragments can develop into separate showers and, at the very highest cosmic ray energies, $E > 10^{17}$ eV, some of those which penetrate to the surface of the Earth are found to be multi-cored; these might be due to the break up of a very high energy nucleus. Second, this mechanism produces spallation products with very high energies. This will prove to be a central topic in the study of the propagation of cosmic ray nuclei in the interstellar medium. The determination of the cross-sections for the production of the various spallation products is therefore of the greatest interest.

10.2 Spallation cross-sections

Spallation cross-sections are best determined from collider experiments in which beams of high energy particles interact with target nuclei. From these data, partial cross-sections for the production of different elements and isotopes as a function of energy can be determined. For astrophysical applications a huge range of species and particle energies are of interest.

There are therefore three approaches to the determination of spallation cross-sections. The first is to determine the cross-sections by experiment. Protons are fired at the target material and then the energy of the proton is the same as the energy per nucleon which the target nucleus possesses in the rest frame of the proton. Since hydrogen is by far the most common element in the interstellar gas, this is the dominant process involved in the splitting up of high energy nuclei, although spallation on helium nuclei also makes a significant contribution.

The results of these experiments can then be used to determine *semi-empirical relations* from which the cross-sections for rare and unstable elements and isotopes can be estimated. This procedure is similar to that used in nuclear physics in which the semi-empirical mass formula is based upon the liquid drop model of the nucleus.

A third procedure is to model the spallation process by simulating the details of particle-particle collisions inside the nucleus using Monte Carlo techniques. The trajectory of the incoming particle inside the nucleus is followed, the initial conditions being selected at random. The proton interacts randomly with the nucleons inside the nucleus and, depending upon the particles which are knocked out of the nucleus in the interaction and the energy of the excited nucleus, the parent nucleus fragments into a number of different end products, the probability of these end products being produced being described by their *partial cross-sections*. In typical Monte Carlo simulations, vast numbers of collisions are studied by high speed computer so that good statistics can be built up even for rare interaction chains.

Strenuous efforts have been made to determine spallation cross-sections for as many elements and their isotopes as is practicable. Not only have the partial cross-sections for the creation of product nuclei been determined but also the variation of these partial cross-sections with energy. The results of a major programme to achieve these goals has been published by Webber and his colleagues (Webber *et al.*, 1990a,b,c,d). Table 10.1*a* and *b* is a compilation of partial cross-sections kindly provided by Drs R. Silberberg and C. H. Tsao, who derived these from semi-empirical formulae which take into account a very wide range of nuclear data (Silberberg *et al.*, 1988). At the bottom of each column, the *total inelastic cross-section* for the break-up of the target nucleus is given. Not surprisingly, the total cross-section turns out to be similar to the geometric cross-section of the nucleus.

There is reasonable agreement between the measured cross-sections and those derived from the semi-empirical formulae. Normally, the agreement is within about 25% but there are cases in which larger discrepancies are found. The precision of the measured partial cross-sections is about 2% for the best determinations. Whilst there are some discrepancies in the absolute values of the cross-sections, the relative cross-sections for the formation of the isotopes of a particular element from a single parent are in good agreement.

Several interesting features of Table 10.1 are worth noting. There is always a large cross-section for chipping off a single nucleon or α -particle from a nucleus. This is not particularly unexpected because there are always more grazing than head-on collisions. In the spallation of ^{12}C , there is a significant cross-section for the break up of the nucleus into three α -particles. When the product nuclei are unstable, the formation of pairs of nuclei with similar masses is not favoured. This is similar to what is found in nuclear fission experiments. Another interesting point is that even nuclei are slightly favoured over odd nuclei, as can be seen from the run of the partial cross-sections for the spallation of iron with mass number. This parallels the observed abundances of the elements as a whole which favour nuclei with even numbers of nucleons and reflects the greater binding energies of nuclei with even numbers of nucleons. Finally, not all of the total cross-section is accounted for by the partial cross-sections listed in the table. This is largely because only the most important nuclei have been included.

We also need to know the energy dependence of these cross-sections, some examples described by Webber and his colleagues being shown in Fig. 10.3 (Webber *et al.*, 1990a,b,c,d). In Figs 10.3*a* and *b*, the points show the experimentally determined cross-sections and the lines are the predictions of various semi-empirical formulae. It can be seen that over the energy ranges shown in Figs 10.3*a* and *b*, the variations of the partial cross-sections with energy are quite small. On the other hand, in the spallation of iron nuclei, there are strong variations at low energies in the partial spallation cross-sections (Fig. 10.3*c*). These variations are principally associated with the difference in mass number of the parent and product nuclei. At relativistic energies, it is expected that the cross-sections should remain roughly constant and the semi-empirical formulae provided an accurate description of the partial cross-sections. These data will be used in the study of the spallation products produced when high energy protons and nuclei interact with the interstellar gas.

Table 10.1 (a) Partial cross-sections for inelastic collisions of selected heavy nuclei with hydrogen with $E = 2.3 \text{ GeV}$ per nucleon.

Product nucleus	Z	A	Parent nucleus							
			^{11}B	^{12}C	^{14}N	^{16}O	^{20}Ne	^{24}Mg	^{28}Si	^{56}Fe
Lithium	3	6	12.9	12.6	12.6	12.6	12.6	12.6	12.6	17.4
		7	17.6	11.4	11.4	11.4	11.4	11.4	11.4	17.8
Beryllium	4	7	6.4	9.7	9.7	9.7	9.7	9.7	9.7	8.4
		9	7.1	4.3	4.3	4.3	4.3	4.3	4.3	5.8
		10	15.8	2.9	1.9	1.9	1.9	1.9	1.9	4.1
Boron	5	10	26.6	17.3	16.0	8.3	7.1	6.2	5.3	5.3
		11	—	31.5	15.0	13.9	12.0	10.4	9.0	8.1
Carbon	6	10	—	3.9	3.3	2.9	2.1	1.6	1.2	0.5
		11	0.6	26.9	12.4	10.6	7.9	5.9	4.5	1.3
		12	—	—	38.1	32.7	13.5	10.1	7.6	4.7
		13	—	—	10.5	14.4	10.7	8.0	6.0	3.7
		14	—	—	—	2.3	3.9	3.0	2.2	2.1
Nitrogen	7	13	—	—	10.7	3.6	2.7	2.0	1.5	0.5
		14	—	—	—	26.3	10.9	8.1	6.1	2.9
		15	—	—	—	31.5	10.0	7.5	5.7	4.3
		16	—	—	—	—	3.4	2.6	1.9	1.6
Oxygen	8	14	—	—	—	3.4	2.5	1.9	1.4	0.3
		15	—	—	—	27.8	11.8	8.9	6.7	1.0
		16	—	—	—	—	27.0	13.5	10.2	3.9
		17	—	—	—	—	15.5	11.6	8.7	4.1
		18	—	—	—	—	4.5	4.7	3.5	2.6
Fluorine	9	16	—	—	—	—	—	1.4	1.1	—
		17	—	—	—	—	8.5	6.4	4.8	—
		18	—	—	—	—	14.4	10.8	8.1	2.4
		19	—	—	—	—	21.0	10.9	8.2	4.8
		20	—	—	—	—	—	4.2	3.1	2.3
Neon	10	18	—	—	—	—	2.8	2.1	1.6	—
		19	—	—	—	—	17.3	5.3	4.0	—
		20	—	—	—	—	—	17.8	13.4	3.6
		21	—	—	—	—	—	14.0	10.6	5.4
		22	—	—	—	—	—	8.2	5.8	4.3
		23	—	—	—	—	—	—	1.3	—
Sodium	11	20	—	—	—	—	—	1.5	1.1	—
		21	—	—	—	—	—	7.7	5.6	—
		22	—	—	—	—	—	16.8	12.7	2.3
		23	—	—	—	—	—	21.0	12.0	6.4
		24	—	—	—	—	—	—	5.2	3.7
Magnesium	12	23	—	—	—	—	—	29.8	1.6	0.6
		24	—	—	—	—	—	—	17.1	3.2
		25	—	—	—	—	—	—	18.5	6.0

Table 10.1 (cont.)

Product nucleus	Z	A	Parent nucleus							
			¹¹ B	¹² C	¹⁴ N	¹⁶ O	²⁰ Ne	²⁴ Mg	²⁸ Si	⁵⁶ Fe
Aluminium	13	25	—	—	—	—	—	—	6.3	—
		26	—	—	—	—	—	—	14.4	6.8
		27	—	—	—	—	—	—	7.6	1.7
		28	—	—	—	—	—	—	13.3	2.0
		29	—	—	—	—	—	—	21.0	6.7
		30	—	—	—	—	—	—	—	5.7
Silicon	14	27	—	—	—	—	—	—	30.7	0.4
		28	—	—	—	—	—	—	—	2.7
		29	—	—	—	—	—	—	—	6.0
		30	—	—	—	—	—	—	—	10.4
		31	—	—	—	—	—	—	—	3.1
		32	—	—	—	—	—	—	—	1.2
Total inelastic cross-section			237.8	252.4	280.9	308.8	363.3	415.7	466.0	763.4

Cross-sections measured in units of millibarns = 10^{-31} m^2 .

Data kindly provided by Drs R. Silberberg and C. H. Tsao.

Table 10.1 (b) Partial cross-sections for inelastic collisions of iron (Fe) with hydrogen with $E = 2.3 \text{ GeV}$ per nucleon.

Product nucleus	Z	σ
Silicon	14	24.1
Phosphorus	15	23.9
Sulphur	16	35.2
Chlorine	17	30.0
Argon	18	43.4
Potassium	19	41.6
Calcium	20	54.9
Scandium	21	55.5
Titanium	22	72.3
Vanadium	23	51.6
Chromium	24	79.6
Manganese	25	120.8
Iron	26	66.7

Cross-sections measured in units of millibarns = 10^{-31} m^2 .

Data kindly provided by Drs R. Silberberg and C. H. Tsao.

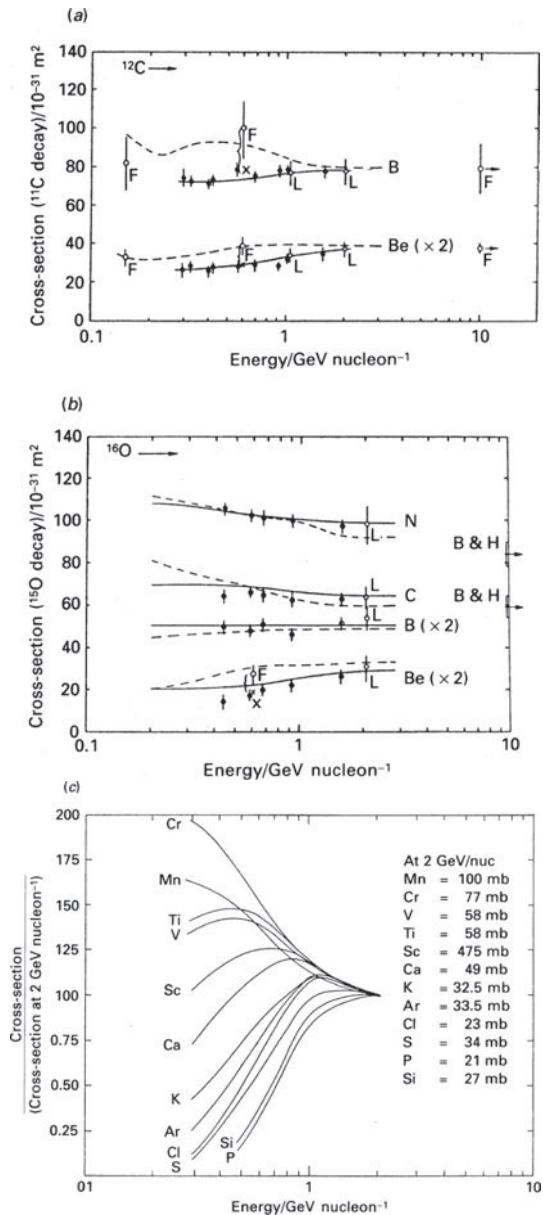


Fig. 10.3 Illustrating the energy dependence of the partial cross-sections for the formation of (a) boron and beryllium from carbon and (b) nitrogen, carbon, beryllium and boron from oxygen, both in spallation interactions with protons. The solid lines show the expectations of the semi-empirical formulae proposed by Webber and his colleagues. The dashed lines show the expectations of much earlier semi-empirical formulae of Tsao and Silberberg. (c) Relative partial cross-sections for the spallation of ^{56}Fe by protons into lighter elements as a function of energy. These cross-sections are strongly energy dependent at low energies ($1 \text{ mb} = 1 \text{ millibarn} = 10^{-31} \text{ m}^2$) (Webber *et al.*, 1990a,b,c,d; Tsao and Silberberg, 1979).

Table 10.2 Important radioactive decay chains for γ -ray line astronomy.

Decay chain	Mean life (years)	$Q/Q(^{56}\text{Ni})$	γ -ray energy (MeV)	Photons/positrons per disintegration
$^{56}\text{Ni} \rightarrow ^{56}\text{Co} \rightarrow ^{56}\text{Fe}$	0.31	1	0.847	0.2(e^+)
			1.238	1
			0.122	0.7
$^{57}\text{Co} \rightarrow ^{57}\text{Fe}$	1.1	2×10^{-2}	0.014	0.88
			1.275	0.88
$^{22}\text{Na} \rightarrow ^{22}\text{Ne}$	3.8	5×10^{-3}	1.275	0.9 (e^+)
			0.94 (e^+)	1
$^{44}\text{Ti} \rightarrow ^{44}\text{Sc} \rightarrow ^{44}\text{Ca}$	68	2×10^{-3}	1.156	1
			0.078	1
			0.068	1
$^{60}\text{Fe} \rightarrow ^{60}\text{Co} \rightarrow ^{60}\text{Ni}$	4.3×10^5	1.5×10^{-4}	1.332	1
			1.173	1
			0.059	1
$^{26}\text{Al} \rightarrow ^{26}\text{Mg}$	1.1×10^6	1.5×10^{-4}	0.85 (e^+)	0.85 (e^+)
			1.809	1

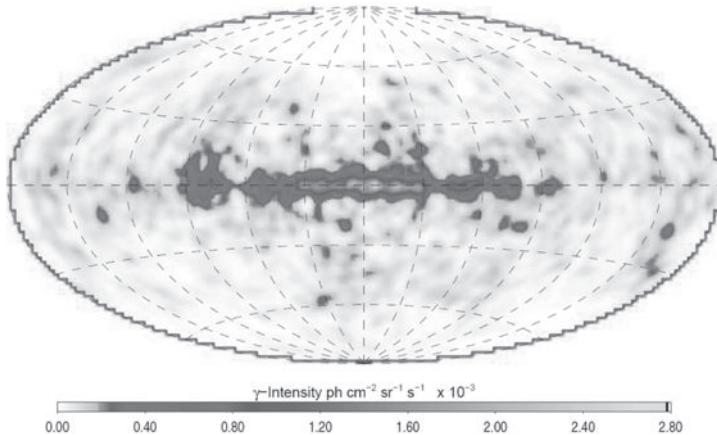
$Q/Q(^{56}\text{Ni})$ is the predicted isotopic yield of each species relative to ^{56}Ni based upon Solar System abundances of the elements and the assumption that all the Solar System abundances of ^{56}Fe , ^{57}Fe and ^{44}Ca and 1%, 0.5% and 0.1% of the ^{60}Ni , ^{22}Ne and ^{26}Mg , respectively, are produced explosively through the above chains (Ramaty and Lingenfelter 1979).

10.3 Nuclear emission lines

There are two types of nuclear process which are important in producing γ -ray lines in the spectra of astronomical sources: the decay of radioactive species created in the processes of nucleosynthesis and the collisional excitation of the nuclei by cosmic ray protons and nuclei. Highlights of the astrophysical results from γ -ray spectroscopy are summarised by Diehl and his colleagues (Diehl *et al.*, 2006b). As they emphasise, these are challenging observations since the fluxes of γ -rays are low and the background of γ -rays within the detectors is high.

10.3.1 Decay of radioactive isotopes

Stellar nucleosynthesis results in unstable as well as stable nuclei and the radioactive decay of the former are sources of γ -ray line emission. In order to be observable, there must be large enough yields of the radioactive species and their half-lives must be sufficiently short to result in detectable emission. Table 10.2 displays a list of some of the more important γ -ray lines which are expected to be observable with their half-lives.

**Fig. 10.4**

An all-sky image of the ^{26}Al γ -ray emission at 1.809 MeV as observed by the COMPTEL instrument of the Compton Gamma-ray Observatory. The image is the result of nine years of observation (Plüschke *et al.*, 2001).

In order to be observable, the radioactive nuclides must be ejected from their sources. The most likely source of most of the radioactivities in Table 10.2 is explosive nucleosynthesis so that the γ -rays emitted in the decay of the radionuclides are not absorbed in the stellar interior. This has to be the case for the radionuclides with half-lives less than one year. For the longer lived species, the radionuclides can be brought to the stellar surface if convection within the stellar interior is sufficiently vigorous and they can then be expelled in strong stellar winds, as occurs in Wolf–Rayet and asymptotic giant branch stars. As a result, sources associated with short-lived isotopes are expected to be point-like and associated with supernovae, while the longer-lived species, such as ^{26}Al and ^{60}Fe , can be expelled into the interstellar medium resulting in diffuse Galactic γ -ray line emission. Because the intensity of the longer-lived radioactive species is averaged over time-scales of order 10^6 years and because the interstellar medium is transparent at these γ -ray energies, these observations provide estimates of the average supernova rate for the Galaxy as a whole.

Thanks to observations by the Compton and INTEGRAL Gamma-ray Observatories, evidence for most of the radioactivities listed in Table 10.2 have now been observed.

- Figure 10.4 shows the Compton Gamma-Ray Observatory map of the sky in the line of ^{26}Al (Plüschke *et al.*, 2001). The spectral observations by the INTEGRAL Gamma-Ray Observatory had sufficient energy resolution to show that the material responsible for the ^{26}Al line emission partakes in the general rotation of the interstellar gas about the Galactic Centre. These observations are compelling evidence that nucleosynthesis is an ongoing process in the central regions of our Galaxy (Diehl *et al.*, 2006a). More recently, γ -ray lines of ^{60}Fe at 1.173 and 1.333 MeV have also been detected from the same general direction as the ^{26}Al line emission (Wang *et al.*, 2007).
- γ -ray lines associated with the decay of ^{56}Co were detected soon after the explosion of the supernova SN 1987A in the Large Magellanic Cloud, providing direct evidence for the radioactive origin of the decay of the light curve of supernovae and for the creation

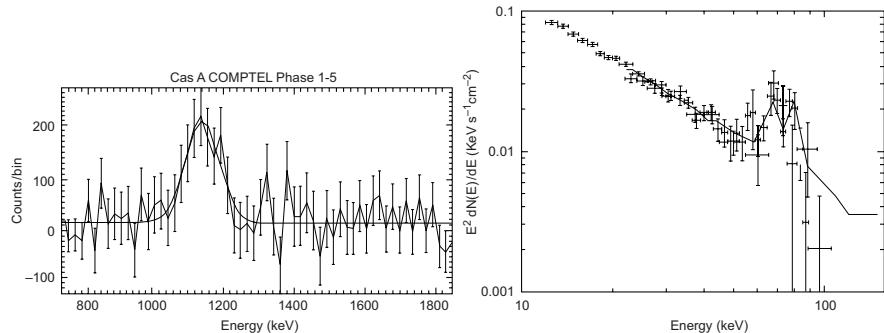


Fig. 10.5 The γ -ray lines of ^{44}Ti from the young supernova remnant Cassiopaeia A. The left-hand panel shows the 1.156 MeV line detected by the COMPTEL instrument of the Compton Gamma-Ray Observatory and the right-hand panel the 68 and 78 keV lines observed by the BeppoSAX and INTEGRAL/IBIS instruments (Iyudin *et al.*, 1994; Diehl *et al.*, 2006b).

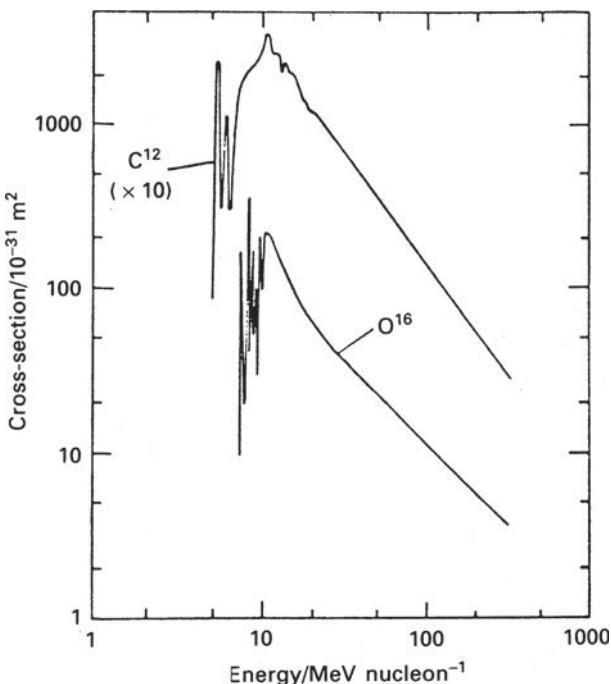
of isotopes belonging to the iron group of elements in the core-collapse of massive stars. This topic is dealt with in more detail in Sect. 13.1.

- γ -ray lines from the decay of ^{44}Ti have been detected from the young supernova remnant Cassiopaeia A (Cas A), which exploded about 350 years ago. The 1.156 MeV line was detected by the COMPTEL instrument of the Compton Gamma-ray Observatory and the 68 and 78 keV lines by the BeppoSAX and INTEGRAL/IBIS instruments (Fig. 10.5) (Iyudin *et al.*, 1994; Diehl *et al.*, 2006b). In addition, the late light curve of SN 1987A indicates that the energy source changed from ^{56}Co decays, which have a half-life of 0.31 years, to those of ^{44}Ti with a half-life of 68 years, although the γ -ray lines themselves have not been detected (see Sect. 13.1).

10.3.2 Collisional excitation of nuclei

Nuclei are excited to energy levels above the ground state by collisions with cosmic ray protons and nuclei. γ -rays are then emitted in the subsequent de-excitation of the nuclei to their ground states. These interactions may either take place in the diffuse interstellar gas, in which case the target nuclei acquire significant velocities in the collisions, or else within interstellar grains in which case the target nuclei emit the γ -rays essentially at rest. The physical process is similar to that of the collisional excitation of the electronic levels of atoms and ions. In the same way, the cross-section for excitation of the nucleus attains a maximum value for particle energies of the same order as the energy of the excited states. Examples of the cross-sections for the collisional excitation of carbon and oxygen nuclei as a function of the energy per nucleon of the incident particle are shown in Fig. 10.6. The cross-sections for collisional excitation of these nuclei are $\approx (1-2) \times 10^{-29} \text{ m}^2$ for protons with energies $\approx 8-30 \text{ MeV}$.

Evidence for these processes occurring in astrophysical environments is provided by γ -ray spectroscopic observations of solar flares. Figure 10.7 shows the γ -ray spectrum of a large flare which occurred on 23 July 2002, as observed by the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) (Lin *et al.*, 2003). γ -ray lines associated with

**Fig. 10.6**

The interaction cross-sections leading to the emission of γ -ray line emission through the excitation of ^{12}C and ^{16}O by collisions with protons as a function of the kinetic energy per nucleon of the incident particle (Ramana Murthy and Wolfendale, 1993).

a number of the abundant elements are observed, as well as lines associated with electron–positron annihilation at $\epsilon = 0.511$ MeV and the line at 2.223 MeV associated with neutron capture by hydrogen nuclei, the neutrons originating in spallation interactions induced by particles accelerated in the flare. The contributions of different line and continuum processes to the overall spectrum are indicated on the diagram. Most of the continuum radiation is non-thermal bremsstrahlung, which was discussed in Sect. 6.6. These observations are of the greatest interest from the point of view of the acceleration of charged particles in solar flares since the particles responsible for exciting the emission lines must be accelerated to MeV energies in the solar flare itself.

Ramaty and Lingenfelter carried out computations of the expected γ -ray spectrum of the interstellar medium due to the interaction of the interstellar flux of high energy particles with the interstellar gas (Ramaty and Lingenfelter, 1979). Figure 10.8 shows the predicted γ -ray emission spectrum in the general direction of the Galactic Centre due to these processes. There are considerable uncertainties in these calculations because the interstellar flux of high energy particles is poorly known in the energy range 1–100 MeV. In addition, it is not known precisely what fraction of the interstellar gas is condensed into dust grains. Nevertheless, these calculations indicate those elements which are likely to be significant γ -ray line emitters, the broad lines resulting from collisions taking place in the gas phase

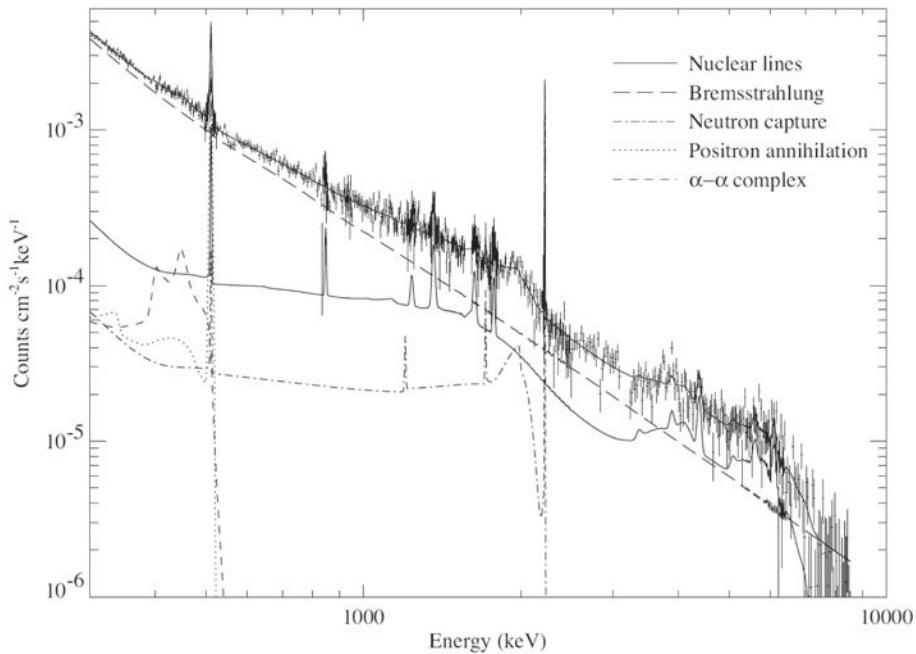


Fig. 10.7 The γ -ray spectrum of the intense γ -ray line solar flare of 23 July 2002 as measured by the RHESSI (Reuven Ramaty High Energy Solar Spectroscopic Imager) (Lin *et al.*, 2003). Modelling of the various contributions to the total spectrum are shown. The continuum is mostly non-thermal bremsstrahlung. The nuclear de-excitation lines are due to Fe, Mg, Si, Ne, C, and O, the principal lines being listed in Table 10.3. Positron annihilation and neutron capture on hydrogen result in the narrow lines at 511 keV and 2.223 MeV, respectively.

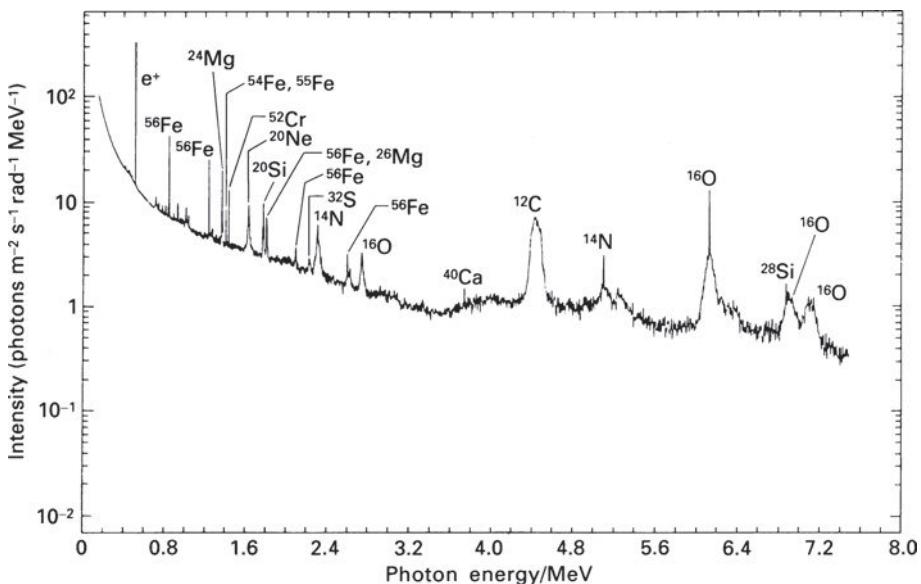


Fig. 10.8 The predicted γ -ray spectrum resulting from low energy cosmic ray interactions with the interstellar gas in the general direction of the Galactic Centre (Ramaty and Lingenfelter, 1979).

Table 10.3 Some important nuclear γ -ray lines.					
Nucleus	Energy (MeV)	Nucleus	Energy (MeV)	Nucleus	Energy (MeV)
^{12}C	4.438	^{20}Ne	1.634	^{56}Fe	0.847
^{14}N	2.313		2.613	^{56}Fe	1.238
	5.105		3.34		1.811
^{16}O	2.741	^{24}Mg	1.369		
	6.129		2.754		
	6.917		1.779		
	7.117		6.878		

and the narrow lines being produced in dust grains. A list of some of the more important lines is given in Table 10.3. Searches have been made for these γ -ray lines by the Compton Gamma-Ray and INTEGRAL Observatories, but the predicted intensities are less than the sensitivities achievable by these telescopes. To date, there are no convincing identifications of γ -ray lines due to nuclear de-excitation from the interstellar gas (Diehl *et al.*, 2006b).

10.4 Cosmic rays in the atmosphere

10.4.1 Nucleonic cascades

When high energy cosmic ray protons and nuclei enter the atmosphere, or the sensitive volume of a detector array, they initiate *nucleonic cascades*, similar to the electromagnetic cascades described in Sect. 9.9, but now including a vast array of nucleonic interactions. The interaction of a primary particle with a target nucleus was described in Sect. 10.1, Figs 10.1 and 10.2 illustrating the break-up of the target nucleus and the cosmic ray nucleus in such events. The incoming cosmic ray particles, referred to as the primary particles, give rise to secondary and subsequent generations of product nuclei. The salient features of such nucleonic cascades are as follows:

- (i) The secondary nucleons and charged pions which have sufficient energy continue to multiply through successive generations of nuclear interactions until the energy per nucleon drops below that required for pion production, that is, about 1 GeV. In the nucleonic cascade, the initial energy of the high energy particle is shared among the pions, strange particles and antinucleons, a process sometimes referred to as *pionisation*.
- (ii) The protons lose energy by ionisation losses and most of those with energies less than 1 GeV are brought to rest.
- (iii) The neutral pions π^0 have short lifetimes, 1.78×10^{-16} s, before decaying into two γ -rays, $\pi^0 \rightarrow 2\gamma$, each of which initiates an electromagnetic cascade as described

in Sect. 9.9. Many of the charged pions decay in flight into muons releasing muon neutrinos and antineutrinos:

$$\left. \begin{array}{l} \pi^+ \rightarrow \mu^+ + \nu_\mu \\ \pi^- \rightarrow \mu^- + \bar{\nu}_\mu \end{array} \right\} \text{mean lifetime} = 2.551 \times 10^{-8} \text{ s}. \quad (10.3)$$

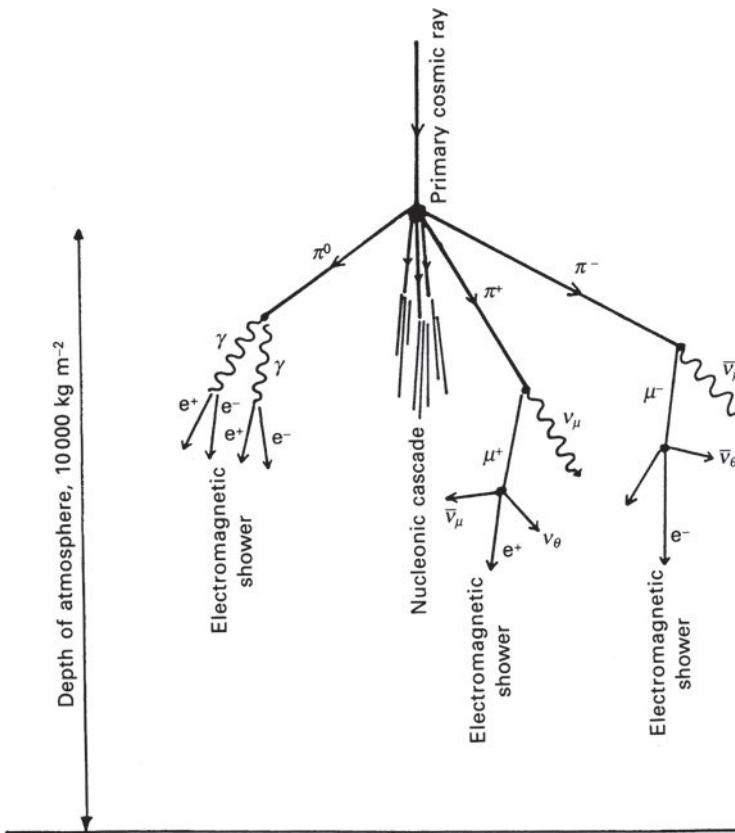
In turn, the low energy muons decay into positrons, electrons and muon neutrinos with somewhat longer mean lifetimes:

$$\left. \begin{array}{l} \mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \\ \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \end{array} \right\} \text{mean lifetime} = 2.2001 \times 10^{-6} \text{ s}. \quad (10.4)$$

For high energy cosmic rays entering the atmosphere, the muons are produced with very high energy and are highly penetrating. Because they have virtually no nuclear interaction and their ionisation losses are small, high energy muons can be observed at the surface of the Earth. In their rest frames of reference, they decay with a mean lifetime of 2.2×10^{-6} s corresponding to a distance of 660 m. To the external observer, however, they are observed to decay with a mean lifetime of $2.2 \times 10^{-6}\gamma$ s because of relativistic time dilation, where γ is the Lorentz factor, $\gamma = (1 - v^2/c^2)^{-1/2}$. As noted in all relativity textbooks, since the muons are created at an altitude of about 10 km, muons with Lorentz factors $\gamma \geq 20$ suffer little decay by the time they are observed at the surface of the Earth. These observations provide direct evidence for relativistic time dilation and length contraction. The high energy muons can penetrate quite far underground and so provide an effective means of monitoring the average intensity and isotropy of the flux of cosmic rays arriving at the top of the atmosphere.

The interactions involved in the development of nucleonic cascades are summarised in Fig. 10.9. The same processes take place within the sensitive volume of cosmic ray particle detectors. Most of the decay products are readily detectable by their ionisation losses and so, if there is sufficient depth to stop all the particles produced in the cascade, the total ionisation provides a measure of the total energy of the primary particle. We will return to this topic when we study extensive air-showers and the highest energy cosmic rays (Sect. 15.5).

As a result of these nucleonic and electromagnetic cascades, there is a distribution of the various products of nucleonic cascades through the full depth of the Earth's atmosphere. Figure 10.10 shows the vertical fluxes of particles with high and low energies, what are referred to as the hard and soft components, as observed at different heights in the atmosphere (Amsler *et al.*, 2008). We can understand qualitatively the features of this diagram in terms of the models of electromagnetic and nuclear cascades. The bulk of the observed flux is caused by primary protons having energies $E \geq 1$ GeV. The path length for interaction of these high energy protons with the atoms and molecules of the atmosphere is about 800 kg m^{-3} , compared with a total depth of about $10\,000 \text{ kg m}^{-3}$, which accounts for the rapid rise in all the products of the nucleonic cascade at the top of the atmosphere. The number of protons then falls off exponentially with path length as expected and correspondingly the numbers of pions and neutrons. The number of electrons grows exponentially to begin with, a characteristic of electron–photon cascades, and then drops off rapidly. Thus, even at the very top of the atmosphere, there are large fluxes of secondary, relativistic electrons

**Fig. 10.9**

A schematic diagram showing the development of a nucleonic cascade in the atmosphere. Such cascades initiated by high energy particles develop in exactly the same fashion inside cosmic ray telescopes.

which complicates the determination of the primary spectrum of the cosmic ray electrons in high altitude balloon flights. The high energy muons fall off rather slowly but the low energy, or soft, muons decay before reaching the surface of the Earth.

10.4.2 Radioactive nuclei produced by cosmic rays in the atmosphere

An important aspect of cosmic ray interactions in the atmosphere is the production of short-lived radioactive isotopes. Neutrons are liberated in the spallation interactions of cosmic rays with the nuclei of atoms, ions and molecules in the atmosphere, most of them eventually being absorbed by ^{14}N nuclei through the reaction



About 5% of the neutrons having energies greater than 4 MeV take part in the endothermic reaction



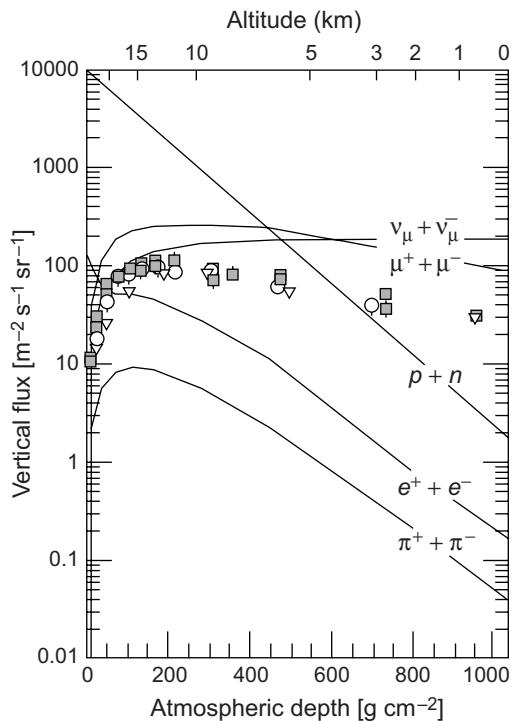


Fig. 10.10 The vertical fluxes of different components of the cosmic radiation with energies $E \geq 1$ GeV in the atmosphere (Amsler *et al.*, 2008). Most of the components are secondary or higher products of the primary cosmic rays. The points show measurements of negative muons with $E_\nu \geq 1$ GeV.

The total rate of formation of carbon-14, ^{14}C , in the atmosphere is about $2.23 \times 10^4 \text{ m}^{-2} \text{ s}^{-1}$ and that of tritium, ^3H , about $2 \times 10^3 \text{ m}^{-2} \text{ s}^{-1}$, the latter figure including tritium formed as spallation products. These radioactive products are created high up in the atmosphere where they are rapidly oxidised to form molecules such as $^{14}\text{CO}_2$ and ^3HOH . These molecules are then precipitated with CO_2 and H_2O in the normal way. The half-lives of ^{14}C and ^3H are 5568 years and 12.46 years, respectively, while their residence times in the atmosphere are about 25 years before they are absorbed in organic material or precipitated as rain and water onto the land and sea.

The abundances of ^{14}C and ^3H can therefore be used to date samples of material which contain residual organic matter, provided the rate of production of radioactive species has been constant. ^3H is used as a tracer in meteorological studies as well as being used to date agricultural products. ^{14}C is used extensively in archaeological studies and is the basis of radiocarbon dating. The success of the method depends upon calibrating the ^{14}C ages against independently estimated ages of organic samples since the production rate of ^{14}C depends upon the cosmic ray flux at the top of the atmosphere.

The calibration of radiocarbon ages against independent age estimates is a key topic, regularly reviewed in the journal *Radiocarbon*. Tree-ring dating (dendrochronology) provides a calibration of the ^{14}C scale back to times up to about 12 000 years before the

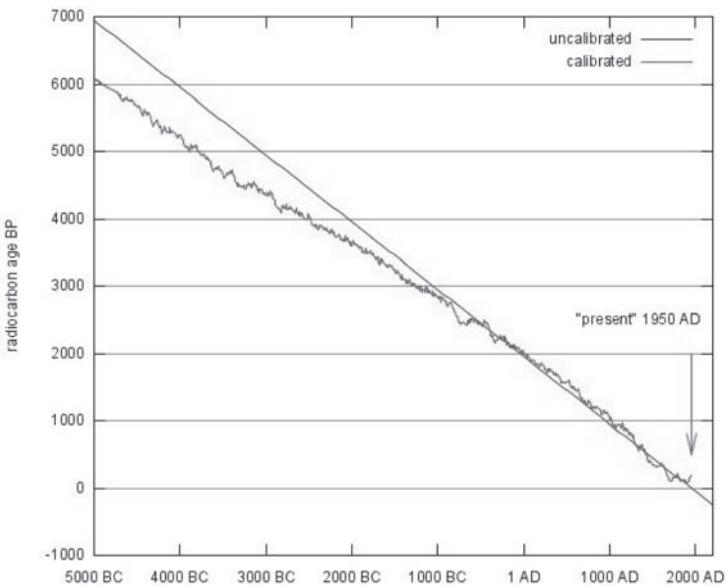


Fig. 10.11

Radiocarbon dates compared with tree-ring dates (Stuiver *et al.*, 1998). If the formation rate of ^{14}C were constant, the radiocarbon ages would agree with the tree-ring ages shown on the abscissa (straight line). The radiocarbon ages are less than the tree-ring ages at early times.

present day. The procedure is to measure the $^{14}\text{C}/^{12}\text{C}$ ratio in the rings of very ancient trees for which reliable tree-ring ages can be established. If the cosmic ray flux at the top of the atmosphere were constant there would be an exact match between the ages of organic specimens determined by radiocarbon dating and the tree-ring ages. Figure 10.11 shows that there is in fact a discrepancy between these ages which increases with increasing age before the present (Stuiver *et al.*, 1998). There is a convincing explanation for this discrepancy. Paleogeomagnetic studies have shown that the strength of the Earth's magnetic dipole has increased significantly over the last 7000 years (Damon *et al.*, 1978). The Earth's magnetic field strength affects the flux of cosmic rays incident at the top of the atmosphere because the interstellar flux of high energy particles has to diffuse through the magnetic field in the interplanetary medium and the Earth's magnetic field to reach the atmosphere. If the Earth's magnetic field strength were weaker in the past, greater fluxes of high energy particles would arrive at the top of the atmosphere resulting in a greater production rate of ^{14}C and in an underestimate of the age of the ^{14}C samples as compared the age expected if the cosmic ray flux were constant. When account is taken of these variation in the Earth's dipole moment, the interstellar cosmic ray flux appears to have been remarkably constant over the last 10 000 years. In addition, there are smaller variations associated, for example, with the 11-year solar cycle.

The dendrochronology technique has been extended to about 12 000 years before the present day and can be further extended back to about 50 000 years before the present epoch using samples of corals. Many more details of these remarkable techniques are discussed by Reimer and her colleagues (Reimer *et al.*, 2004). During the period of atmospheric nuclear

testing, the flux of ^{14}C increased by a factor of 2 within two years in the northern hemisphere because of the neutrons liberated in nuclear explosions. This distorted significantly the recent calibration curves for radioactive dating. An interesting calculation is to estimate whether or not a nearby supernova would be detectable in the ancient tree-ring data as an abrupt enhancement of the ^{14}C flux. The γ -rays emitted in the explosion arrive at the Earth at the same time as the optical signal and then release neutrons through the resonant (γ, n) interaction with the nuclei of atoms and molecules in the atmosphere. According to the calculation of Damon and his colleagues, a supernova at a distance of about 1 kpc would just be detectable as an enhanced ^{14}C signal in the tree-ring data (Damon *et al.*, 1995). Although they claimed to detect a weak signal associated with SN 1006, Menjo and his colleagues could find no evidence for such an enhancement which they argued would be masked by small changes in the ^{14}C signal because of variations in the cosmic ray flux associated with the 11-year solar cycle (Menjo *et al.*, 2005).

Plasma physics and magnetohydrodynamics are enormous subjects which play a central role in many aspects of high energy astrophysics. In this chapter, a simple introduction is provided to a number of recurring topics in the physics of diffuse plasmas. Many more details can be found in the classic text *The Physics of Fully Ionised Gases* by Spitzer (1962) and the recent authoritative survey by Kulsrud, *Plasma Physics for Astrophysics* (Kulsrud, 2005). The book *The physics of plasmas* by Fitzpatrick, available on-line, provides a clear introduction to all the topics discussed in this chapter (Fitzpatrick, 2008).

11.1 Elementary concepts in plasma physics

11.1.1 The plasma frequency and Debye length

We consider the simplest case of a fully ionised plasma consisting of protons and electrons which have equal number densities $n_p = n_e$. The electrostatic forces between the electrons and protons are very strong and ensure charge neutrality except on small scales, specifically, on scales less than the *Debye length* λ_D . Following Fitzpatrick, suppose a layer of the electrons of thickness x is displaced a distance δx relative to the ions. The net effect is to set up two oppositely charged sheets with surface charge density $\sigma = e n_e \delta x$ and the system forms a parallel plate capacitor with opposite surface charges σ on the plates. The electric field across the layer which tends to restore charge neutrality is then $E = \sigma/\epsilon_0 = e n_e \delta x/\epsilon_0$ and the equation of motion per unit surface area for the electrons in the layer is

$$(m_e n_e x) \frac{d(\delta x)}{dt} = -(e n_e x) \frac{e n_e \delta x}{\epsilon_0}, \quad \frac{d(\delta x)}{dt} = -\frac{e^2 n_e \delta x}{\epsilon_0 m_e}. \quad (11.1)$$

This is the equation of simple harmonic motion with angular frequency $\omega_p^2 = e^2 n_e / \epsilon_0 m_e$ which is known as the *angular plasma frequency*,

$$\omega_p = \left(\frac{e^2 n_e}{\epsilon_0 m_e} \right)^{1/2} = 56 n_e^{1/2} \text{ rad s}^{-1}, \quad \nu_p = \left(\frac{e^2 n_e}{4\pi^2 \epsilon_0 m_e} \right)^{1/2} = 8.97 n_e^{1/2} \text{ Hz}, \quad (11.2)$$

where n_e is in particles m^{-3} . In (11.2), the plasma frequency ν_p is also given. Notice that the same equation of motion applies for a single electron as for the electrons in bulk. The plasma frequency is a fundamental quantity in plasma physics and will appear many times in the course of this exposition. The same calculation can be carried out for the protons in which case the electron mass m_e would be replaced by the mass of the proton m_p , and so the

ion plasma frequency is $\sqrt{m_p/m_e} = \sqrt{1836} = 46$ times smaller than the electron plasma frequency.

The hydrogen plasma is assumed to be fully ionised at temperature T . The mean square speed of the electrons in the plasma in the x -direction is therefore

$$\frac{1}{2}m_e v_x^2 = \frac{1}{2}kT \quad \text{and so} \quad v_x = (kT/m_e)^{1/2}. \quad (11.3)$$

The distance a typical particle of the plasma can travel during one radian of the plasma oscillation is therefore

$$\lambda_D = \frac{v_x}{\omega_p} = \left(\frac{kT\epsilon_0}{n_e e^2} \right)^{1/2} = 69 \left(\frac{T}{n_e} \right)^{1/2} \text{ m}, \quad (11.4)$$

where the temperature of the plasma is in kelvins and the number density of electrons in particles m^{-3} . λ_D is defined to be the *Debye length*. The mass of the particle has cancelled out in deriving the expression for the Debye length and so it is the same for electrons and protons. This makes sense since this is the typical distance over which charge imbalance can take place and so should be the same for electrons and protons.

The Debye length is also the distance over which the influence of any charge imbalance is shielded by the charges in the plasma. This can be demonstrated by the simple argument given by Fitzpatrick. In thermal equilibrium, the number density of charges is given by the Boltzmann distribution

$$n = n_0 \exp(-e\Phi/kT), \quad (11.5)$$

where Φ is the electrostatic potential. Now suppose the potential distribution is perturbed by an amount $\delta\Phi$ as a result of a localised perturbation in the charge distribution $\delta\rho_{\text{ext}}$. Then, the number density of electrons and protons is modified because of the change in potential. For the protons, for example,

$$n + \delta n = n_0 \exp \left[-\frac{e(\Phi + \delta\Phi)}{kT} \right] = n - n \frac{e \delta\Phi}{kT}, \quad (11.6)$$

for small potential perturbations $\delta\Phi$. Hence,

$$\delta n = -n \frac{e \delta\Phi}{kT}. \quad (11.7)$$

A similar perturbation is present in the distribution of electrons of exactly the same magnitude but of opposite sign. Combining these results, the change in electric charge density is

$$\delta\rho = \delta\rho_{\text{ext}} - 2n \frac{e^2 \delta\Phi}{kT}. \quad (11.8)$$

We now insert this perturbation into Poisson's equation to find the potential distribution in the presence of the charge perturbation,

$$\nabla^2(\delta\Phi) = -\frac{\delta\rho}{\epsilon_0} = -\frac{\delta\rho_{\text{ext}}}{\epsilon_0} + \frac{2ne^2(\delta\Phi)}{\epsilon_0 k T}, \quad (11.9)$$

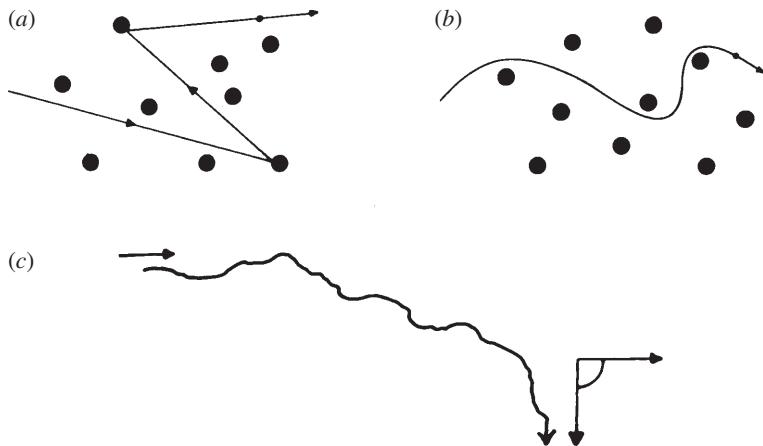


Fig. 11.1 A schematic diagram illustrating particle–particle collisions according to (a) the Drude model and (b) collisions mediated by long-range electrostatic forces. In (c) the particle is eventually deflected through 90° by the stochastic effect of a large number of distant encounters.

and hence

$$\left(\nabla^2 - \frac{2}{\lambda_D^2} \right) \delta\Phi = -\frac{\delta\rho_{\text{ext}}}{\epsilon_0}. \quad (11.10)$$

If the source of the perturbation is a charge q at the origin, we write $\delta\rho_{\text{ext}} = q \delta(\mathbf{r})$ and then (11.10) takes the familiar form:

$$\left(\nabla^2 - \frac{2}{\lambda_D^2} \right) \delta\Phi = -\frac{e \delta(\mathbf{r})}{\epsilon_0}, \quad (11.11)$$

where $\delta(\mathbf{r})$ is the Dirac δ -function. The solution of this equation is well-known:

$$\delta\Phi = \frac{q}{4\pi\epsilon_0 r} \exp\left(-\frac{\sqrt{2}r}{\lambda_D}\right). \quad (11.12)$$

This calculation illustrates the role of the Debye length in acting as a shielding distance for the influence of the charge q upon the plasma. For distances less than λ_D , the potential is the usual inverse function of distance from the charge. At distances greater than λ_D , the influence of the charge decreases exponentially because of the shielding effect of the negative charge induced by the presence of the positive charge q .

The importance of these results is that, on scales greater than the Debye length and time-scales greater than the inverse of the plasma frequency, the behaviour of individual particles is not important, but rather the bulk properties of the plasma dominate the physics. These are the scales on which the many different types of waves and instabilities occur in plasmas.

11.1.2 The diffusion of charged particles

The diffusion of charged particles and the exchange of energy between them are needed to work out the time it takes particles of different masses to come to thermal equilibrium at the same temperature and also to evaluate the electrical conductivity of fully ionised plasmas. Figure 11.1a depicts an elementary model for the diffusion of particles mediated by collisions between solid spheres. In this model, the forces involved in the collisions are very short range and associated with the repulsive effect of short range atomic forces. The case of a fully ionised plasma is different in that the interactions between electrons and ions are mediated by long range electrostatic forces and, because of the increasing numbers of electrons with increasing distance, these contribute to the forces acting on the particles, just as in our considerations of ionisation losses and bremsstrahlung. The dynamics of a particle in the plasma are illustrated schematically in Fig. 11.1b. The mean free path of the particle is defined to be the distance over which it loses all memory of its initial direction, that is, it is deflected stochastically through an angle $\theta \sim 90^\circ$ (Fig. 11.1c).

In a plasma, a charged particle is subjected to a large number of small impulses and the average of these random impulses is zero, $\langle \Delta v_\perp \rangle = 0$. Statistically, however, since the impulses are random, the root mean square velocity $\langle \Delta v_\perp^2 \rangle_{1/2}$ is non-zero and so the particle acquires net perpendicular momentum by random scattering. If the root mean square perpendicular velocity acquired per second is $\langle \Delta v_\perp^2 \rangle^{1/2}$, this velocity becomes roughly v after a time t_c , where

$$\langle \Delta v_\perp^2 \rangle t_c = v^2. \quad (11.13)$$

The time t_c is defined to be the *collision time* of the particle in the plasma in the sense that, after this number of collisions, the particle has lost all memory of its initial direction and t_c can be related to the diffusion coefficient of the particles in the plasma.

Let us carry out some simple illustrative calculations which illuminate much more complete analyses. Consider first a particle of charge Ze and velocity v interacting with identical particles in a plasma and, for simplicity, we assume all the other particles are stationary. In a single collision, as shown in Sect. 5.2, the particle receives a momentum impulse perpendicular to its direction of motion of magnitude

$$p_\perp = \frac{Z^2 e^2}{2\pi\epsilon_0 b v} \quad \text{and hence} \quad \Delta v_\perp = \frac{Z^2 e^2}{2\pi\epsilon_0 b v m}. \quad (11.14)$$

Using the same procedure as in Sect. 5.2, we find the mean square perpendicular velocity by integrating over all particles within the cylindrical volume $2\pi b db dx$ (Fig. 5.2). Hence, the mean square component of velocity perpendicular to the direction of motion acquired *in one second* is

$$\langle \Delta v_\perp^2 \rangle = \int_{b_{\min}}^{b_{\max}} \left(\frac{Z^2 e^2}{2\pi\epsilon_0 b v m} \right)^2 2\pi b N v db. \quad (11.15)$$

Therefore,

$$\langle \Delta v_\perp^2 \rangle = \frac{Z^4 e^4}{4\pi^2 \epsilon_0^2 m^2 v} 2\pi N \ln \left(\frac{b_{\max}}{b_{\min}} \right) = \frac{Z^4 e^4 N}{2\pi \epsilon_0^2 m^2 v} \ln \Lambda, \quad (11.16)$$

Table 11.1 Gaunt factors, $\ln \Lambda$, for the diffusion coefficients and electrical conductivity of a plasma as a function of electron number density and temperature (Spitzer, 1962).

T/K	Electron number density (n_e/m^{-3})							
	10^6	10^9	10^{12}	10^{15}	10^{18}	10^{21}	10^{24}	10^{27}
10^2	16.3	12.8	9.43	5.97				
10^3	19.7	16.3	12.8	9.43	5.97			
10^4	23.2	19.7	16.3	12.8	9.43	5.97		
10^5	26.7	23.2	19.7	16.3	12.8	9.43	5.97	
10^6	29.7	26.3	22.8	19.3	15.9	12.4	8.96	5.54
10^7	32.0	28.5	25.1	21.6	18.1	14.7	11.2	7.85
10^8	34.3	30.9	27.4	24.0	20.5	17.0	13.6	10.1
								4.39
								6.69

where $\Lambda = b_{\max}/b_{\min}$. Once again, we have encountered our old friend $\ln \Lambda$, a Gaunt factor. In the present instance, the maximum collision parameter b_{\max} is the Debye length for the plasma, $b_{\max} = \lambda_D = (\epsilon_0 k T / n Z^2 e^2)^{1/2}$, the typical shielding distance of a particle in the plasma. As discussed above, if $Z = 1$, the Debye length is the same for protons and electrons. The minimum collision parameter is the closest distance of approach in the classical limit, $b_{\min} = Z^2 e^2 / 8\pi \epsilon_0 m_e v^2$ (see Sect. 5.2). Therefore,

$$\frac{t_c = v^2}{\langle \Delta v_{\perp}^2 \rangle} = \frac{2\pi \epsilon_0^2 m^2 v^3}{Z^4 e^4 N \ln \Lambda} = \frac{2\pi \epsilon_0^2 m^{1/2} (3kT)^{3/2}}{Z^4 e^4 N \ln \Lambda} \quad (11.17)$$

where, in the last equality, the velocity v is taken to be the typical thermal velocity of a particle in a plasma at temperature T , $\frac{1}{2}mv^2 = \frac{3}{2}kT$.

We have plainly made some sweeping approximations in the above calculation, but the key point is that the functional dependences we have obtained are correct when all the particles are in motion with a Maxwellian distribution of velocities. Spitzer gives details of these results in his monograph *The Physics of Fully Ionised Gases* (Spitzer, 1962). In fact, the full calculation carried out by Chandrasekhar shows that our result (11.17) is within 50% of the exact answer. Spitzer's expression for what he refers to as the *self-collision time* is

$$t_c = \frac{11.4 \times 10^6 A^{1/2} T^{3/2}}{n Z^4 \ln \Lambda} \text{ seconds ,} \quad (11.18)$$

where A is defined by $m = Am_p$ and the particle number density is measured in particles m^{-3} . Appropriate values of Gaunt factors for a wide range of temperatures and densities is given in Table 11.1.

The self-collision time is closely related to the thermalisation time-scale in the sense that the particle has changed its velocity vector and hence exchanged energy with all the other particles in the plasma such that $\Delta v/v \sim \Delta E/E \sim 1$. Thus, the time t_c is also roughly the time it takes to establish a Maxwellian distribution of velocities among the particles. In some circumstances, the electrons and protons in a plasma may be far from

thermodynamic equilibrium to begin with and then (11.18) describes how long it takes the particle distributions to relax to their equilibrium values. Because of the $A^{1/2}$ dependence, the electrons come into thermal equilibrium with each other $\sqrt{1836} = 46$ times more rapidly than the protons. To complete the picture, the expression for the exchange of energy between the electrons and protons was derived in Sect. 5.2 where it was shown that, because of the large difference in masses, the energy exchange was of the order m_e/m_p , less than that between electrons. Therefore, in the notation of this section, if the thermalisation time for the electrons is τ_e , the corresponding time for protons τ_p is 46 times longer and the time τ_{pe} for the protons and electrons to come into thermal equilibrium with each other is 1836 times greater than τ_e . Thus, in certain astronomical circumstances, there may not be time for the electrons and protons to attain thermal equilibrium at the same temperature.

Let us work out the mean free path of a proton in the interplanetary medium for which the values $T = 10^6 \text{ K}$, $A = 1$, $N = 5 \times 10^6 \text{ m}^{-3}$, $Z = 1$ and $\ln \Lambda = 28$ can be adopted. Then, we find $\lambda = 3 \times 10^{13} \text{ m}$, much greater than the distance from the Earth to the Sun, $1.5 \times 10^{11} \text{ m}$. Therefore for protons, which carry all the momentum of the Solar Wind, the mean free path for electrostatic collisions is very much greater than the Sun–Earth distance. This calculation shows that the Solar Wind can be considered a collisionless plasma. It neglects, however, the central role of the interplanetary magnetic field which dominates the dynamics of the particles and the associated scattering by magnetic irregularities discussed in Sects. 7.3 and 7.4.

11.1.3 The electrical conductivity of a fully ionised plasma

We can use the results of Sect. 11.1.2 to estimate the conductivity of the plasma in what is referred to as the Lorentz approximation in which the protons are assumed to remain stationary while the current is carried by the drift of the electrons under the influence of the electric field. The Drude model for the conductivity can be used in which the mean free time between collisions τ_c due to long range interactions can be found by the same techniques exploited in Sect. 11.1.2.

Let us review first the Drude model for the mean drift velocity of particles under the influence of an electric field E_x . The electrons have a Maxwellian distribution of speeds as well as a mean drift velocity, which is assumed to be small compared with the random velocities of the particles. Then, the statistical equation of motion for the mean drift velocity $\langle v \rangle$ in the direction of the field is

$$\frac{d\langle v \rangle}{dt} = \frac{e}{m_e} E_x - \frac{\langle v \rangle}{\tau_c}, \quad (11.19)$$

where τ_c is the *mean free time* between collisions or *relaxation time*. If the electric field E_x is zero, the mean velocity in the x -direction decays to zero with characteristic time τ_c . In the steady state in the presence of the electric field E_x , the left-hand side of (11.19) is zero and so

$$\langle v \rangle = \frac{e\tau_c}{m_e} E_x. \quad (11.20)$$

If there are n_e electrons per unit volume, the current density J_x is

$$J_x = en_e \langle v \rangle = \frac{e^2 n_e \tau_c}{m_e} E_x = \sigma E_x \quad \text{where} \quad \sigma = \frac{e^2 n_e \tau_c}{m_e}. \quad (11.21)$$

This is the standard Drude expression for electrical conductivity σ of the medium.

Next, we work out the appropriate value for τ_c , the electrons diffusing through a medium consisting of stationary protons. We use the results of Sect. 11.2.2, writing m_e for m and setting $Z = 1$ in (11.17). Then,

$$\tau_c = \frac{2\pi \epsilon_0^2 m_e^{1/2} (3kT)^{3/2}}{e^4 n_i \ln \Lambda}. \quad (11.22)$$

Substituting into (11.21), we find

$$\sigma = \frac{2\pi \epsilon_0^2 (3kT)^{3/2}}{m_e^{1/2} e^2 \ln \Lambda}. \quad (11.23)$$

This approximate calculation has the same functional dependence upon the parameters of the plasma as that quoted by Spitzer (1962). A detailed discussion of the electrical conductivity of a plasma is given by Spitzer who gives the following result:

$$\sigma = \frac{32\pi^{1/2} \epsilon_0^2 (2kT)^{3/2}}{Ze^2 m_e^{1/2} \ln \Lambda} = 2.63 \times 10^{-2} \frac{T^{3/2}}{Z \ln \Lambda} (\text{ohm m})^{-1} \text{ or siemens m}^{-1}. \quad (11.24)$$

Spitzer and Härm also included the effect of electron–electron collisions and showed that, for a hydrogen plasma, the electrical conductivity is decreased by a factor of 0.582 (Spitzer and Härm, 1953). Inspection of Table 11.1 shows that the use of a Gaunt factor $\ln \Lambda = 10$ is adequate for our purposes and hence $\sigma \approx 10^{-3} T^{3/2}$ siemens m⁻¹.

Taking again the example of the interplanetary medium with $T = 10^6$ K, we find $\sigma = 10^6$ siemens m⁻¹. This value is of the same order of magnitude as the electrical conductivity of metals, which lie in the range $(1–6) \times 10^7$ siemens m⁻¹. Thus, typical cosmic plasmas have very high electrical conductivities and this has important implications for the coupling between magnetic fields and the plasma; in particular, it results in the phenomenon of *magnetic flux freezing*.

11.2 Magnetic flux freezing

Many of the plasmas encountered in high energy astrophysics, and astronomy in general, have very high electrical conductivities. In the limit of infinite conductivity, the magnetic field behaves as if it were frozen into the plasma, the phenomenon known as *magnetic flux freezing*. We present two versions of the physics of this process. In one approach, we write down the equations of magnetohydrodynamics, take the limit of infinite electrical conductivity and then find the dynamics of the fields and the plasma. The second is a more physical approach in which we study the behaviour of the flux linkage of closed circuits in a fully ionised plasma when the circuits are moved or distorted.

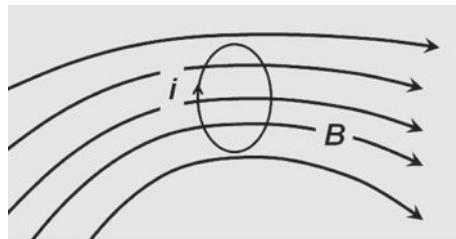


Fig. 11.2 Illustrating a current loop in a fully ionised plasma threaded by a magnetic field of magnetic flux density \mathbf{B} .

The theorem we wish to prove is the following: if we represent the magnetic field by magnetic lines of force, so that the number per unit area perpendicular to the lines is equal in magnitude to the magnetic field strength, then, when there are movements in the plasma, the magnetic field lines move and change their shape as though they were frozen into the plasma.

11.2.1 The physical approach

In the physical approach, we follow Ratcliffe's pleasant analysis in his monograph *An Introduction to the Ionosphere and Magnetosphere* (Ratcliffe, 1972). He carries out the calculations in two parts. In the first, the changes in the magnetic flux linkage in a stationary current loop are studied when the magnetic field strength changes, while in the second the effect of distorting the shape of the current loop is analysed.

It is assumed that the electrical conductivity of the plasma is infinite. Suppose there is a current loop to which no batteries are attached in the plasma (Fig. 11.2). The electromotive force \mathcal{E} induced in the circuit can only be due to the rate of change of magnetic flux ϕ linking the circuit,

$$\mathcal{E} = -\frac{d\phi}{dt} . \quad (11.25)$$

The magnetic flux ϕ consists of two parts, one part due to the current in the loop itself ϕ_i , and the other due to all external currents ϕ_{ex} . If the inductance of the loop is L , then, by definition

$$\phi_i = Li ; \quad \phi = \phi_i + \phi_{ex} . \quad (11.26)$$

If the external currents change so that ϕ_{ex} changes, then an electromotive force is induced in the circuit and the resulting current is given by

$$\frac{L di}{dt} + Ri = -\frac{d\phi_{ex}}{dt} . \quad (11.27)$$

To model the case of a collisionless plasma, the resistance of the loop is set to zero and so

$$L \frac{di}{dt} = -\frac{d\phi_{ex}}{dt} = \frac{d\phi_i}{dt} , \quad (11.28)$$

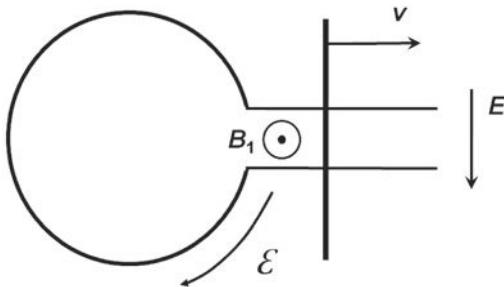


Fig. 11.3 Illustrating the conservation of magnetic flux as the shape of a loop changes.

that is,

$$\phi_i + \phi_{\text{ex}} = \text{constant}. \quad (11.29)$$

Thus, although ϕ_{ex} may change, by virtue of changing, it induces a current which exactly cancels out the effect which might have been expected. This is a consequence of the fact that the current loop has zero resistance. It is not true if R is finite but is very closely so if R is very, very small. Note that there is nothing inconsistent in assuming that there is a current i flowing without any electromotive force being present initially. Because the conductor has zero resistance, there is no means of dissipating the current. A corollary of this proof is that, if the circuit is moved, the flux will also remain unchanged because, in the frame of reference of the moving loop, only the external field changes.

What happens if the loop changes shape? Consider the specific example of the circuit shown in Fig. 11.3 which consists of a loop with parallel wires crossed by a conductor. The entire circuit is made of superconducting material and the field in the region of the parallel wires is B_1 . Now let the conductor move down the wire a distance dx at a velocity v . The strength of the induced electric field is $|E| = |v \times B| = vB_1$ in the sense shown in Fig. 11.3. The induced electromotive force due to the motion of the wire is

$$\mathcal{E} = El = vB_1l, \quad (11.30)$$

where l is the distance between the parallel wires. But $\mathcal{E} = -d\phi/dt$ and therefore the magnetic flux induced in the circuit is

$$d\phi = (vB_1l)dt \quad \text{in the sense opposite to } B_1. \quad (11.31)$$

But, because the area is bigger, more magnetic flux is enclosed by the loop. In fact, because all the changes are small,

$$d\phi = B_1l dx = (vB_1l)dt \quad \text{in the same direction as } B_1. \quad (11.32)$$

Thus, the two effects cancel exactly and there is no net change in the magnetic flux through the circuit after its shape has changed. Since the magnetic flux through the circuit is constant, $L_1i_1 = L_2i_2$, where the subscripts 1 and 2 refer to the values of L and i before and after the deformation of the circuit. The electromotive force produced when the loop is distorted

induces an increment in the current in the loop which just ‘stays around’ because there is no means of dissipating it.

To express this result mathematically, if we choose any loop C in the plasma and follow it as the shape changes due to motions in the plasma,

$$\int_S \mathbf{B} \cdot d\mathbf{S} = \text{constant} , \quad (11.33)$$

where $d\mathbf{S}$ is the increment of surface area and S refers to the total surface area bounded by the loop C . If a small circular loop is located in the magnetised plasma with $d\mathbf{S}$ parallel to \mathbf{B} and the plasma is allowed to expand uniformly, the above result leads to $B dS = B\pi r^2 = \text{constant}$, that is, $B \propto r^{-2}$. Thus, in a uniform expansion, the energy density of the magnetic field decreases as $B^2/2\mu_0 \propto r^{-4}$. This is the same result as that found in the adiabatic expansion of a gas for which the ratio of specific heats is $\gamma = 4/3$. We will return to this point at the end of the next subsection.

11.2.2 The magnetohydrodynamic approach

First, we write down the equations of magnetohydrodynamics.

- *The equation of continuity*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 , \quad (11.34)$$

where ρ is the mass density and \mathbf{v} is the velocity at a point in the fluid.

- *Force equation*

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p + \mathbf{J} \times \mathbf{B} + \mathbf{F}_v + \rho \mathbf{g} , \quad (11.35)$$

where p is the pressure, \mathbf{J} is the current density, \mathbf{B} is the magnetic flux density, \mathbf{F}_v represents viscous forces and \mathbf{g} is the gravitational acceleration. We note that $d\mathbf{v}/dt$ is a convective derivative, that is, the forces act upon a particular element of the fluid in a frame of reference which moves with that element of the plasma. This derivative is related to the partial derivatives which describe changes in the properties of a fluid at a fixed point in space:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla . \quad (11.36)$$

- *Maxwell's equations*

The equations are written in the form:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} , \quad (11.37)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} , \quad (11.38)$$

$$\nabla \cdot \mathbf{B} = 0 , \quad (11.39)$$

$$\nabla \cdot \mathbf{E} = \frac{\rho_e}{\epsilon_0} . \quad (11.40)$$

There is no displacement term $\partial \mathbf{D}/\partial t$ in (11.38) because we deal with slowly varying phenomena. Therefore, no space charge effects are present – the particles of the plasma always have time to neutralise any charge imbalance on the scale of motion of the plasma.

- *Ohm's law*

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (11.41)$$

where σ is the electrical conductivity of the plasma. Now substituting for \mathbf{E} in (11.37) using (11.41),

$$\nabla \times (\mathbf{J}/\sigma - \mathbf{v} \times \mathbf{B}) = -\frac{\partial \mathbf{B}}{\partial t}. \quad (11.42)$$

Now, eliminating \mathbf{J} between (11.38) and (11.42), we find

$$\nabla \times \left(\frac{\nabla \times \mathbf{B}}{\sigma \mu_0} - \mathbf{v} \times \mathbf{B} \right) = -\frac{\partial \mathbf{B}}{\partial t}. \quad (11.43)$$

Therefore,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \frac{\nabla \times (\nabla \times \mathbf{B})}{\sigma \mu_0}. \quad (11.44)$$

We now use the identity $\nabla \times (\nabla \times \mathbf{B}) = \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B}$. Since $\nabla \cdot \mathbf{B}$ is always zero, we find

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{1}{\sigma \mu_0} \nabla^2 \mathbf{B}. \quad (11.45)$$

- *Entropy equation*

Following Kulsrud, the entropy of a perfect gas per unit mass can be written

$$S = C_V \ln \left(\frac{p}{\rho^\gamma} \right), \quad (11.46)$$

where C_V is the specific heat capacity per unit mass, $C_V = \frac{3}{2}k/\mu m_p$ and μ is the mean molecular weight per particle (Kulsrud, 2005). In the case of a hydrogen plasma in thermal equilibrium at temperature T , the electrons and protons contribute equally to the heat capacity of the plasma and so $\mu = \frac{1}{2}$. If there is no heat flow and frictional heating and radiative heating and cooling can be neglected, the entropy of any fluid element is conserved. In diffuse plasmas, this is generally the case, except in the presence of shocks and current sheets. In the presence of magnetic fields, there is little heat transfer across field lines, although the mean free path along them is very large. Consequently, the temperature is usually nearly constant along field lines. As Kulsrud points out, this phenomenon is dramatically illustrated by the spectacular loops observed in scattering light above active sunspots which illuminate the distribution of magnetic field lines (Fig. 11.4). Neighbouring field lines can be much cooler and are not observed in scattered light.

The system of equations (11.34), (11.35), (11.45) and (11.46) form the basic equations of magnetohydrodynamics. Let us consider first the case of infinite conductivity $\sigma = \infty$,

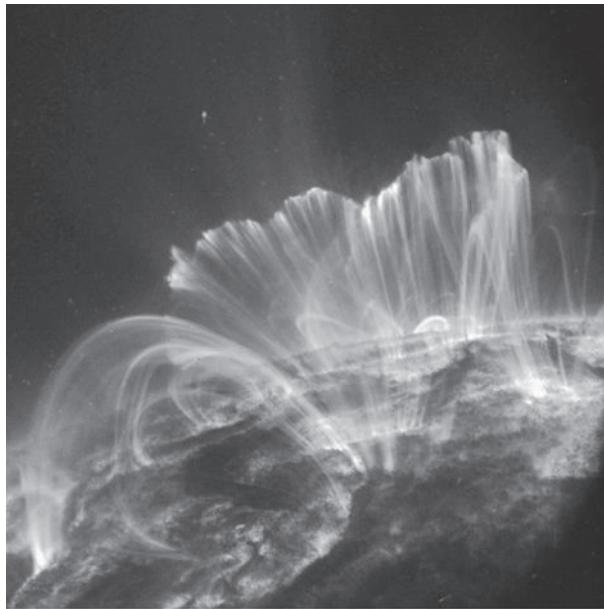


Fig. 11.4 Examples of coronal loops observed above active sunspots from observations of the surface of the Sun by NASA's Transition Region and Coronal Explorer (TRACE) spacecraft. (Courtesy of NASA and the TRACE Science Team.)

in which case (11.45) becomes

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}). \quad (11.47)$$

As in Sect. 11.2.1, consider a current loop S in the plasma and the two contributions to changes in the magnetic flux density ϕ through it with time. First, there may be changes in the magnetic flux density due to external causes, and second, there is an induced component of the flux density due to motion of the loop. The first contribution is

$$\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}. \quad (11.48)$$

The second contribution results from the fact that, because of the motion of the loop, there is an induced electric field $\mathbf{E} = \mathbf{v} \times \mathbf{B}$. Because $\nabla \times \mathbf{E} = -\partial \phi / \partial t$, there is an additional contribution to the total magnetic flux through the loop,

$$\int_S \frac{d\mathbf{B}}{dt} \cdot d\mathbf{S} = - \int_S \nabla \times (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{S}. \quad (11.49)$$

Adding together both contributions, we obtain

$$\begin{aligned} \frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} &= \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} - \int_S \nabla \times (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{S} \\ &= \int_S \left(\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) \right) \cdot d\mathbf{S} = 0, \end{aligned}$$

because of (11.47). Thus, the magnetic flux through the loop is constant, in other words *magnetic flux freezing*, the same result derived in Sect. 11.2.1.

Kulsrud has emphasised that care has to be taken in interpreting the way in which magnetic fields change as the density of the plasma changes (Kulsrud, 2005). The three symmetrical examples he gives are as follows:

- (i) Consider first the squashing of a cylinder of magnetised plasma in the radial direction, the uniform magnetic field being parallel to the axis of the cylinder. Both the mass and magnetic flux in the cylinder are conserved and so $\rho \propto r^{-2}$ and $B \propto r^{-2}$ and so

$$B/\rho = \text{constant} . \quad (11.50)$$

- (ii) Next, suppose the area of the cylinder is unchanged, but the length l is extended. Then, the magnetic flux is unchanged, but the plasma density decreases as l^{-1} . Therefore, in this case,

$$B/\rho l = \text{constant} . \quad (11.51)$$

- (iii) Finally, consider the isotropic expansion or contraction of the plasma towards the origin. Both the mass and magnetic flux within a sphere of radius r are conserved and so ρr^3 and Br^2 are both constants. Therefore,

$$B/\rho^{2/3} = \text{constant} . \quad (11.52)$$

Thus, the value of n in the relation $B/\rho^n = \text{constant}$ depends upon the nature of the geometric distortion of the magnetic field and plasma configuration, even in these symmetric cases.

Another important result is the time it would take a magnetic field to diffuse out of a particular region as a result of the finite electrical conductivity of the medium. If the plasma is at rest, $v = 0$, (11.45) becomes

$$\frac{\partial \mathbf{B}}{\partial t} - \frac{1}{\sigma \mu_0} \nabla^2 \mathbf{B} = 0 . \quad (11.53)$$

This diffusion equation can be used to estimate, to order of magnitude, the time it takes the magnetic field to diffuse out of a region by the usual procedure of writing $\partial B/\partial t \sim B/\tau$, where τ is a characteristic diffusion time and $\nabla^2 B \approx B/L^2$, where L is the scale of the system. Therefore,

$$\frac{B}{\tau} \approx \frac{1}{\sigma \mu_0} \frac{B}{L^2}; \quad \tau \approx \sigma \mu_0 L^2 . \quad (11.54)$$

Let us apply this result to a number of important astrophysical cases.

- First we consider the collapse of a main sequence star to a *white dwarf*. If the star collapsed isotropically by a factor of 100 in radius to form a white dwarf, the magnetic flux density would increase by a factor of 10^4 and so, if the initial magnetic flux density were 10^{-2} T, the white dwarf would have $B \approx 10^2$ T, similar to the values observed. To check that the flux freezing assumption is appropriate, the diffusion time-scale for the magnetic field from the white dwarf can be found using (11.54) with the electrical

conductivity $\sigma \approx 10^{-3} T^{3/2}$ siemens m⁻¹. Assuming $T = 10^6$ K and that the radius of the white dwarf is 10⁷ m, the diffusion time is about 3×10^6 years, very much longer than the collapse time of the star. Thus, the flux freezing assumption holds good and provides a wholly plausible origin for the magnetic fields of white dwarfs.

- The same calculation can be repeated for the collapse of the core of a main sequence star to a *neutron star*. In this case, the collapse is by a factor of about 10⁵ in radius, so that the field of the neutron star would be 10⁸ T, again consistent with the observed magnetic flux densities of neutron stars. Assuming the temperature of the newly formed neutron star is $T = 10^8$ K, the diffusion time for the magnetic field would be 3000 years, very much greater than the collapse time of the core of a massive star which is a matter of seconds. Again, it is wholly plausible that flux freezing accounts for the origin of the magnetic field in neutron stars.
- For the case of protostars collapsing from densities of order 10⁶ m⁻³ found in the cores of giant molecular clouds to 10³⁰ m⁻³ in main sequence stars, the isotropic collapse would be by a factor of 10⁸ in radius and so, according to the flux freezing argument, even if the initial field had magnetic flux density 3×10^{-10} T, that inside the main sequence star would be 3 × 10⁶ T, far greater than the observed values and greater than the thermal pressure within a main sequence star. The problem is compounded by the fact that the diffusion time for a star with central temperature $T \sim 10^6$ K is of the order of 10¹⁰ years. In fact, such magnetic fields would be strong enough to halt the collapse of the star. In this case, there must be other processes which lead to the diffusion of magnetic fields out of the protostar, for example, ambipolar diffusion associated with the mixed neutral and ionised gas.

Let us write the condition for magnetic flux freezing in a slightly different way by returning to (11.45),

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \frac{1}{\sigma \mu_0} \nabla^2 \mathbf{B} .$$

The condition for magnetic flux freezing is that the first term on the right-hand side of this equation far exceeds the second. Suppose we are interested in phenomena on the scale L . Then, to order of magnitude, the ratio of the first to the second terms on the right-hand side is

$$R_m = \sigma \mu_0 \frac{\nabla \times (\mathbf{v} \times \mathbf{B})}{\nabla^2 \mathbf{B}} \sim \sigma \mu_0 \frac{(v B/L)}{(B/L)^2} = \sigma \mu_0 v L , \quad (11.55)$$

where v is the velocity of the plasma. The quantity R_m is known as the *magnetic Reynolds number* and is a measure of the importance of magnetic flux freezing on the scale L . Thus, in the collapse of a main sequence star to a white dwarf, the velocity of collapse is of order 50 km s⁻¹ and so $R_m \sim 10^{14}$. Therefore, it is a very secure assumption that, on the scale of collapse of the star to a white dwarf, the magnetic field is frozen into the plasma.

These examples are sufficient to demonstrate that the diffusion times for magnetic fields in typical cosmic plasmas are long and generally much greater than dynamical time-scales. In these circumstances, magnetic flux freezing is a good approximation. We will find numerous applications of this concept in the course of the exposition.

Table 11.2 Typical parameters of the Solar Wind.

Particle velocity	$\sim 350 \text{ km s}^{-1}$
Particle flux	$\sim 1.5 \times 10^{12} \text{ m}^{-2} \text{ s}^{-1}$
Particle concentration	$\sim 10^7 \text{ m}^{-3}$
Energy of proton	$\sim 500 \text{ eV}$
Energy density of protons	$\sim 4 \times 10^{-10} \text{ J m}^{-3}$
Temperature	$\sim 10^6 \text{ K}$
Magnetic flux density	$\sim 5 \times 10^{-9} \text{ T}$
Energy density in magnetic field ($B^2/2\mu_0$)	$\sim 10^{-11} \text{ J m}^{-3}$

These figures refer to the normal Sun. In high speed streams, velocities up to $700\text{--}800 \text{ km s}^{-1}$ are found and the particle concentrations are $\sim 5 \times 10^6 \text{ m}^{-3}$ so that the particle fluxes are more or less the same.

11.2.3 The Solar Wind

The *Solar Wind* is the outflow of hot, ionised material from the corona of the Sun. The temperature of the corona exceeds 10^6 K , resulting in a steady outflow of hot coronal gas. There is plentiful evidence for the presence of strong magnetic fields in the surface layers and corona of the Sun, as indicated by the remarkable coronal loops of hot plasma which stream along the field lines (Fig. 11.4). The plasma and the magnetic field are strongly tied together by magnetic flux freezing and therefore the dynamics depend upon which component has the greater energy (or mass) density. From the properties of the Solar Wind listed in Table 11.2, the kinetic energy of the protons is much greater than that of the magnetic field and therefore the magnetic field is dragged outwards by the inertia of the Solar Wind.

The Sun rotates once every 26 days on its axis and the Solar Wind is released radially outwards with more or less constant radial velocity of the order of 350 km s^{-1} . The particles are tied to magnetic field lines rooted in the Sun and therefore the magnetic field in the Solar Wind takes up a spiral pattern. This is illustrated schematically in Fig. 11.5a which shows the dynamics of particles ejected at constant radial velocity from the Sun as it rotates. The dynamics are the same as those of a rotating garden sprinkler.

Both slow and fast motions are observed in the Solar Wind, in general high speed flows originating along open fields lines towards the polar regions of the Sun. The speeds are smaller closer to the equatorial plane where the field lines are closed. These phenomena are illustrated by the Solar Wind-velocity diagram shown in Fig. 11.5b which was obtained by the Ulysses space mission of the European Space Agency and NASA, which had the great advantage of making observations from an orbit which passed over the north and south poles of the Sun.

In addition to defining the basic structure of the magnetic field in the Solar Wind, the Voyager and Pioneer spacecraft confirmed the tight wrapping of the spiral field beyond about 20–25 AU. Superimposed upon this basic pattern, there is a myriad of other phenomena. For example, the Solar Wind is not uniform over all latitudes and, in particular, at periods

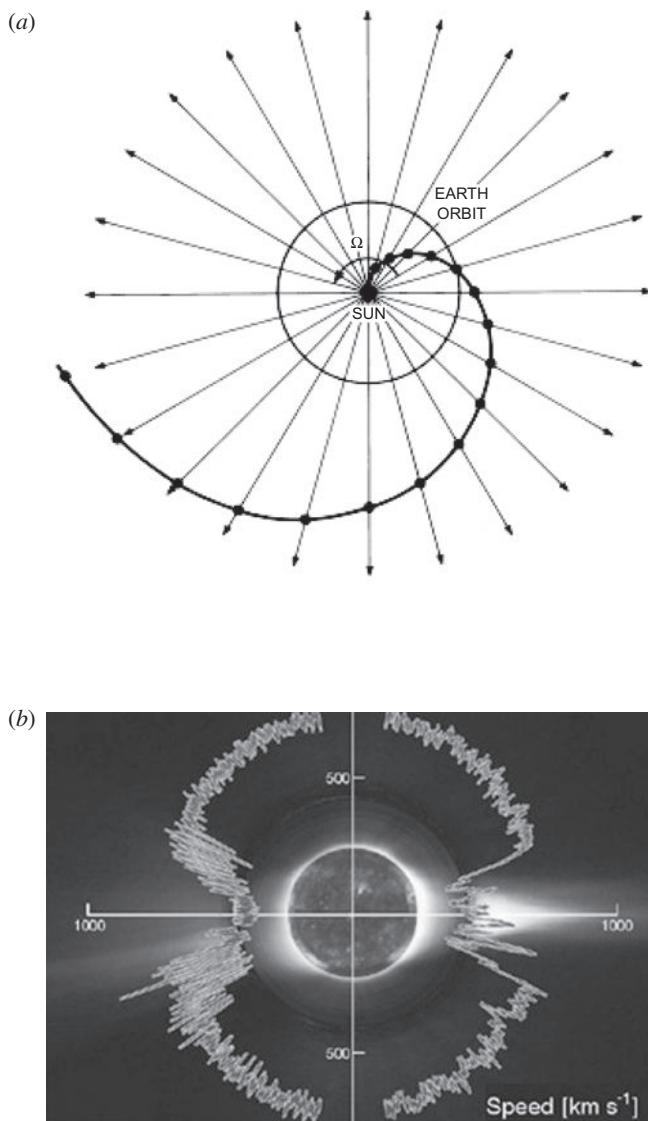


Fig. 11.5 (a) A schematic diagram showing how the magnetic field of the Solar Wind takes up a spiral configuration. The plasma leaving the solar corona moves out more or less radially and the magnetic field is dragged with it. The diagram shows the dynamics of plasma associated with one field line while the Sun rotates through half a rotation. At large distances, the spiral is Archimedean. (b) The Ulysses mission of the European Space Agency and NASA measured the speed of the Solar Wind as it leaves the Sun in 2007. The Ulysses spacecraft flew over the Sun's poles, enabling the velocity of the Solar Wind to be measured as a function of solar latitude. The observations revealed a high speed wind blowing from high latitudes and a slower wind flowing from the equatorial regions. (Courtesy of ESA, NASA and the Ulysses Science Team.)

when there is a high level of solar activity, there are fast streams in the Solar Wind, among the most energetic of these being associated with coronal mass ejection events. These result in shock waves propagating outwards through the interplanetary medium which bring with them a wide variety of new phenomena in the plasma physics and magnetohydrodynamics of the Solar Wind.

The outflow of material from the Sun also modifies the structure of the magnetic field of the Earth and the shape of the distorted magnetic dipole has been determined in detail from satellite studies. The Solar Wind is highly supersonic when it encounters the Earth's magnetic field and hence a *shock front* forms resulting in the characteristic 'stand-off' behaviour seen in front of blunt objects when they move supersonically.

11.3 Shock waves

Shock waves are found ubiquitously in high energy astrophysics. It is useful to derive some of their basic properties which find application in as diverse fields as star formation in the spiral arms of galaxies, the high velocity outflows from young stars, extragalactic radio sources and active galactic nuclei. The basic physics is set out in two classic texts, *Fluid Mechanics* by Landau and Lifshitz (1987), in particular Chap. 9, and *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* by Zeldovich and Raizer (2002).

Perturbations in a gas are propagated away from their source at the speed of sound in the medium. Therefore, if a disturbance is propagated at a velocity greater than the speed of sound, it cannot behave like a sound wave. There is a discontinuity between the regions behind and ahead of the disturbance, the latter region having no prior knowledge of its imminent arrival. These discontinuities are called *shock waves*. They commonly arise in explosions and where gases flow past obstacles at supersonic velocities or, equivalently, objects more supersonically through a gas. The basic phenomenon is the flow of gas at a supersonic velocity relative to the local velocity of sound.

11.3.1 The basic properties of plane shock waves

We assume that there is an abrupt discontinuity between the two regions of fluid flow. In the undisturbed region ahead of the shock wave, the gas is at rest with pressure p_1 , density ρ_1 and temperature T_1 – the speed of sound is c_1 . Behind the shock wave, the gas moves supersonically at speed $U > c_1$ and its pressure, density and temperature are p_2 , ρ_2 and T_2 , respectively (Fig. 11.6a). It is convenient to transform to a reference frame moving at velocity U in which the shock wave is stationary (Fig. 11.6b). In this reference frame, the undisturbed gas flows towards the discontinuity at velocity $v_1 = |U|$ and, when it passes through it, its velocity becomes v_2 away from the discontinuity.

The behaviour of the gas on passing through the shock wave is described by a set of conservation relations. First, mass is conserved on passing through the discontinuity and

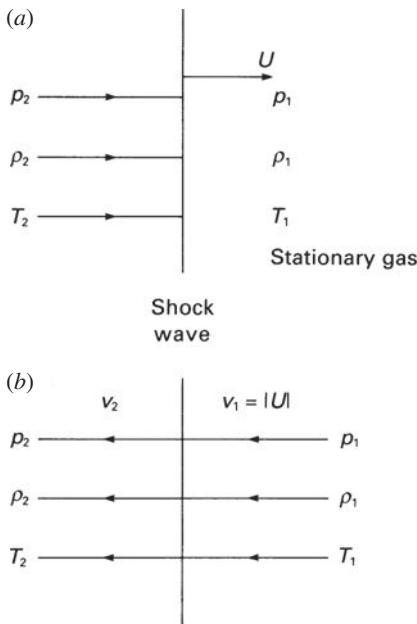


Fig. 11.6 (a) A shock wave propagating through a stationary gas at a supersonic velocity U . The velocity U is supersonic with respect to the sound velocity in the stationary medium c_1 . (b) The flow of gas through the shock front in the frame of reference in which the shock front is stationary.

hence

$$\rho_1 v_1 = \rho_2 v_2 . \quad (11.56)$$

Second, the energy flux, that is, the energy passing per unit time through unit area parallel to v_1 is continuous. One of the standard results of fluid dynamics is that the energy flux through a surface normal to the vector \mathbf{v} is $\rho \mathbf{v} (\frac{1}{2} v^2 + w)$ where w is the enthalpy per unit mass, $w = \varepsilon_m + pV$, ε_m is the internal energy per unit mass and V is the specific volume $V = \rho^{-1}$, that is, the volume per unit mass. We consider only plane shock waves which are perpendicular to v_1 and v_2 and so the conservation of energy flux implies

$$\rho_1 v_1 \left(\frac{1}{2} v_1^2 + w_1 \right) = \rho_2 v_2 \left(\frac{1}{2} v_2^2 + w_2 \right) . \quad (11.57)$$

Notice that it is the enthalpy per unit mass and not the energy per unit mass ε which appears in this relation. The reason is that, in addition to internal energy, work is done on any element of the fluid by the pressure forces in the fluid and this energy is available for doing work. Another way of looking at this relation is in terms of Bernoulli's equation of fluid mechanics in which the quantity $\frac{1}{2} v^2 + w = \frac{1}{2} v^2 + \varepsilon_m + p/\rho$ is conserved along streamlines which is the case for flow at normal incidence through the shock wave.

Finally, the momentum flux through the shock wave should be continuous. For the perpendicular shocks considered here, the momentum flux is $p + \rho v^2$ and hence

$$p_1 + \rho_1 v_1^2 = p_2 + \rho_2 v_2^2 . \quad (11.58)$$

Notice that the pressure p , being a force per unit area, contributes to the momentum flux of the gas. The three conservation relations (11.56), (11.57) and (11.58) are often referred to as the *shock conditions*.

For simplicity, we study shock waves in a perfect gas for which the enthalpy is $w = \gamma pV/(\gamma - 1)$, where γ is the ratio of specific heat capacities and V the specific volume. Landau and Lifshitz show how many elegant results can be obtained for such perfect gases. First, we define the mass flux per unit area $j = \rho_1 v_1 = \rho_2 v_2$. Then, from (11.58), the equation of momentum conversation, we find

$$j^2 = (p_2 - p_1)/(V_1 - V_2). \quad (11.59)$$

In addition, we obtain an expression for the velocity difference

$$v_1 - v_2 = j(V_1 - V_2) = [(p_2 - p_1)(V_1 - V_2)]^{1/2}. \quad (11.60)$$

The next step is to find the ratio V_2/V_1 as a function of p_1 and p_2 for a perfect gas. We begin with the equation of conservation of energy flux (11.57) and substitute as follows:

$$w_1 + \frac{1}{2}v_1^2 = w_2 + \frac{1}{2}v_2^2; \quad w_1 + \frac{1}{2}j^2V_1^2 = w_2 + \frac{1}{2}j^2V_2^2. \quad (11.61)$$

Using (11.59), this expression reduces to

$$(w_1 - w_2) + \frac{1}{2}(V_1 + V_2)(p_2 - p_1) = 0. \quad (11.62)$$

We can now substitute the perfect gas expression, $w = \gamma pV/(\gamma - 1)$ into the relation (11.62) with the result,

$$\frac{V_2}{V_1} = \frac{p_1(\gamma + 1) + p_2(\gamma - 1)}{p_1(\gamma - 1) + p_2(\gamma + 1)}, \quad (11.63)$$

the relation between the pressures and specific volumes on either side of the shock. We can now find the relation between T_2 and T_1 from the perfect gas law, $p_1 V_1/T_1 = p_2 V_2/T_2$,

$$\frac{T_2}{T_1} = \frac{p_2}{p_1} \frac{V_2}{V_1} = \frac{p_2}{p_1} \frac{p_1(\gamma + 1) + p_2(\gamma - 1)}{p_1(\gamma - 1) + p_2(\gamma + 1)}. \quad (11.64)$$

Also, using expression (11.63), we can eliminate V_2 from (11.59) for the flux density j ,

$$j^2 = \frac{(\gamma - 1)p_1 + (\gamma + 1)p_2}{2V_1}. \quad (11.65)$$

From (11.65), we find the velocities of the gas in front of and behind the shock

$$v_1^2 = j^2 V_1^2 = \frac{V_1}{2}[(\gamma - 1)p_1 + (\gamma + 1)p_2], \quad (11.66)$$

$$v_2^2 = j^2 V_2^2 = \frac{V_2}{2} \frac{[p_1(\gamma + 1) + p_2(\gamma - 1)]^2}{p_1(\gamma - 1) + p_2(\gamma + 1)}. \quad (11.67)$$

It is convenient to write these results in terms of the *Mach number* M_1 of the shock wave which is defined to be $M_1 = U/c_1 = v_1/c_1$ where c_1 is the velocity of sound of the undisturbed gas, $c_1 = (\gamma p_1/\rho_1)^{1/2}$. Thus,

$$M_1^2 = v_1^2/(\gamma p_1/\rho_1) = v_1^2/\gamma p_1 V_1. \quad (11.68)$$

Substituting (11.68) into (11.66), the *pressure ratio* is then

$$\frac{p_2}{p_1} = \frac{2\gamma M_1^2 - (\gamma - 1)}{(\gamma + 1)}. \quad (11.69)$$

From the mass conservation equation (11.56) combined with (11.66) and (11.67), the density ratio is

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} = \frac{(\gamma - 1)p_1 + (\gamma + 1)p_2}{(\gamma + 1)p_1 + (\gamma - 1)p_2} = \frac{(\gamma + 1)}{(\gamma - 1) + 2/M_1^2}. \quad (11.70)$$

Finally, from expressions (11.64), (11.69) and (11.70), we find the temperature ratio

$$\frac{T_2}{T_1} = \frac{[2\gamma M_1^2 - (\gamma - 1)][2 + (\gamma - 1)M_1^2]}{(\gamma + 1)^2 M_1^2}. \quad (11.71)$$

In the limit of very strong shocks, $M_1 \gg 1$, we find the following results

$$\frac{p_2}{p_1} = \frac{2\gamma M_1^2}{(\gamma + 1)}, \quad (11.72)$$

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma + 1)}{(\gamma - 1)}, \quad (11.73)$$

$$\frac{T_2}{T_1} = \frac{2\gamma(\gamma - 1)M_1^2}{(\gamma + 1)^2}. \quad (11.74)$$

Thus, in the strong shock limit, the temperature and pressure can become arbitrarily large, but the density ratio attains a maximum value of $(\gamma + 1)/(\gamma - 1)$. For example, a monatomic gas has $\gamma = \frac{5}{3}$ and hence $\rho_2/\rho_1 = 4$ in the limit of very strong shocks. These results demonstrate how efficiently strong shock waves can heat gas to very high temperatures as is found in supernova explosions and supernova remnants.

What is happening in the shock front? The undisturbed gas is both heated and accelerated as it passes through the shock front and, in the case of ordinary gases, this is mediated by their atomic or molecular viscosities. It can be shown that the acceleration and heating of the gas takes place over a physical scale of the order of a few mean free paths of the atoms, molecules or ions of the gas. This makes physical sense because it is over this scale that energy and momentum can be transferred between gas molecules. Thus, the shock front is expected to be very narrow and the heating takes place over this short distance.

11.3.2 The supersonic piston

A common situation in high energy astrophysics involves an object being driven supersonically into a gas, or equivalently, supersonic gas flowing past a stationary object. A illustrative example, set as a problem by Landau and Lifshitz, is that of a piston driven supersonically into a cylinder containing stationary gas (Fig. 11.7) (Landau and Lifshitz, 1987). A shock wave forms ahead of the piston and the gas behind the shock moves at the velocity of the piston U . In the frame of reference of the shock front, which moves at some as yet unknown velocity v_s , the velocity of inflow of the stationary gas is $v_1 = |v_s|$ and the

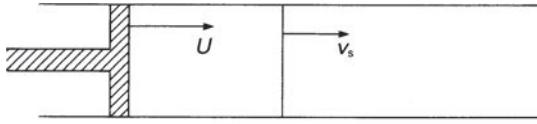


Fig. 11.7 Illustrating the flow of gas in the case of a piston which moves at a supersonic velocity U with respect to the velocity of sound in the stationary medium.

gas behind the shock moves at velocity v_2 . As yet we do not know v_1 and v_2 , but we know that their difference is $v_1 - v_2 = U$.

First of all, from (11.60),

$$v_1 - v_2 = U = [(p_1 - p_2)(V_1 - V_2)]^{1/2}. \quad (11.75)$$

Substituting for V_2 using equation (11.63) and squaring expression (11.75), the expression can be written in terms of the pressure ratio p_2/p_1 ,

$$\left(\frac{p_2}{p_1}\right)^2 - \left(\frac{p_2}{p_1}\right) \left[2 + (\gamma + 1) \frac{U^2}{2p_1 V_1}\right] + \left[1 - \frac{(\gamma - 1)U^2}{2p_1 V_1}\right] = 0. \quad (11.76)$$

We can now write $\gamma p_1 V_1 = c_1^2$, where c_1 is the speed of sound in the undisturbed medium, and solve for p_1/p_2 .

$$\frac{p_2}{p_1} = 1 + \frac{\gamma(\gamma + 1)U^2}{4c_1^2} + \frac{\gamma U}{c_1} \left[1 + \frac{(\gamma + 1)^2 U^2}{16c_1^2}\right]^{1/2}. \quad (11.77)$$

The velocity $v_1 = |v_s|$ follows from expression (11.66),

$$v_1^2 = \frac{V_1}{2} [(\gamma - 1)p_1 + (\gamma + 1)p_2] = \frac{c_1^2}{2\gamma} \left[(\gamma - 1) + (\gamma + 1)\frac{p_2}{p_1}\right]. \quad (11.78)$$

Some simple algebra shows that, substituting for p_2/p_1 using (11.77),

$$v_s = \frac{(\gamma + 1)}{4} U + \left[c_1^2 + \frac{(\gamma + 1)^2 U^2}{16}\right]^{1/2}. \quad (11.79)$$

This is the elegant result we have been seeking since it determines the length of the column of shocked gas ahead of the piston for any supersonic velocity U . In the case of a very strong shock wave $U \gg c_1$, (11.79) reduces to

$$v_s = (\gamma + 1)U/2. \quad (11.80)$$

Thus, the ratio of the position of the shock front to the position of the piston is $v_s/U = (\gamma + 1)/2$. For a monatomic perfect gas $\gamma = \frac{5}{3}$ and hence $v_s/U = \frac{4}{3}$. Thus, all the gas which was originally in the tube between $x = 0$ and the position of the shock wave is squeezed into a smaller distance $(v_s - U)t$. It follows that the density increase over the undisturbed gas is $\rho_2/\rho_1 = v_s/(v_s - U) = (\gamma + 1)/(\gamma - 1)$, the same result we found in (11.73).

This simple calculation gives some impression of what is expected when supersonically moving gas encounters an obstacle or is ejected into a stationary gas. Ahead of the obstacle there is a shocked region which runs ahead of the advancing piston. This is expected to

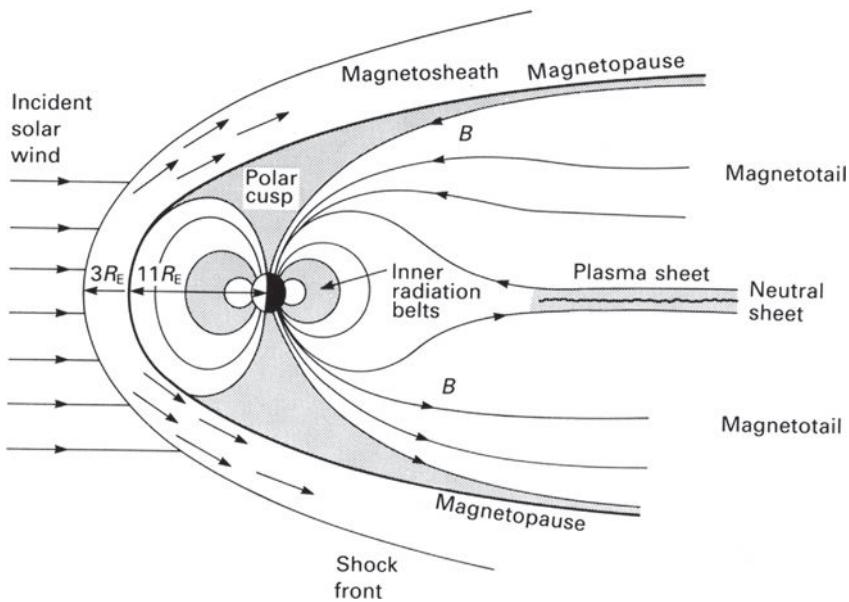


Fig. 11.8 A schematic diagram showing the structure of the Earth's magnetosphere. The names of the various regions are shown on the diagram.

occur when a supernova ejects a sphere of hot gas into the interstellar medium. It also shows that there is a *stand-off distance* between the shock front and the supersonic ejecta and this is observed in the flow of the Solar Wind past the Earth's magnetic dipole.

11.4 The Earth's magnetosphere

The Solar Wind is highly supersonic when it encounters the Earth. To a rough approximation, the Earth and its associated magnetic field act as a spherical obstacle in the outflowing Solar Wind and, consequently, if this were a problem in gas dynamics, a stand-off shock would be expected to form in front of it. The example of the shocked zone in front of a supersonic piston developed in Sect. 11.3.2 provides a simple picture of what might be expected. The important difference is that the gas can flow round the sides of the obstacle and so, while the shock wave is perpendicular at the equator, it becomes oblique with increasing geomagnetic latitude as shown in Fig. 11.8. In the case of oblique shocks, the component of flow velocity parallel to the shock wave is continuous whilst the normal component of the flow satisfies the shock conditions derived in Sect. 11.3.1. As a result, the streamlines are refracted on passing through the oblique shock. Note that the velocity of the flow behind the shock can become supersonic if the shock wave is sufficiently oblique.

Despite the differences between the case of a solid obstacle placed in a supersonic gas flow and the Solar Wind flowing past the Earth, the structures observed in the vicinity of the Earth can be described rather well by classical gas dynamics. The magnetic field and

particle distributions in the vicinity of the Earth have been well determined by space probe experiments, resulting in the structure shown schematically in Fig. 11.8. There is a bow shock, similar to that in front of a solid object, at a stand-off distance of about $14R_E$ from the centre of the Earth in the direction of incidence of the Solar Wind, where R_E is the radius of the Earth. Closer to the Earth, there is a boundary known as the *magnetopause* at a distance of about $11R_E$ which acts as the surface of the region within which the Earth's magnetic field is dynamically dominant. For the purpose of visualisation, the magnetopause may be thought of as the surface of a solid obstacle. The Solar Wind plasma flows past the Earth between the shock wave and the magnetopause. The whole region within the magnetopause is known as the *magnetosphere*, meaning the region in which the magnetic field of the Earth is the dominant dynamical influence. The typical density enhancement across the bow shock is observed to be about a factor of 2–4, typical of the values expected for strong shocks in a monatomic gas.

The Earth's dipole magnetic field is strongly perturbed by the flow of the Solar Wind and so, although it can be well represented by a magnetic dipole close to the surface of the Earth, further away it is distorted as shown in Fig. 11.8. Perhaps the most significant distortion is the fact that the magnetic field lines on the downstream side of the Earth are stretched out by the drag exerted by the Solar Wind. The magnetospheric cavity is stretched out into a long cylindrical region which has radius about $25R_E$ at the distance of the Moon's orbit, that is, at a distance of about $60R_E$. This region is known as the *magnetotail*. The magnetic field lines are oppositely directed on either side of the equatorial plane, those in the northern region heading towards the Earth while those in the southern region point away from the Earth. Between the two regions is a thick layer of hot plasma which is known as the *plasma sheet*. The magnetic field lines run in opposite directions on either side of the plasma sheet and so there must be a surface of zero magnetic field separating the two regions, which is known as a *neutral sheet*. The magnetic field changes sign through the neutral sheet and so an induced electric current flows in the plasma sheet – particles can be accelerated in its vicinity. If the plasma moves in such a way as to bring together regions of oppositely directed magnetic field, the magnetic field lines can 'annihilate', converting the magnetic field energy into particle energy by virtue of the electric fields created as magnetic flux is convected into the neutral sheet. The Solar Wind particles flowing past the magnetotail are coupled into the magnetotail by instabilities acting at the magnetopause. The Kelvin–Helmholtz instability, which results when a fluid streams past a stationary fluid, enables Solar Wind particles to be entrained within the magnetosphere.

This picture of the Earth's magnetosphere provides an explanation for the phenomena of the *aurorae* observed at high geomagnetic latitudes. From Fig. 11.8, it can be seen that particles accelerated in the region of the magnetotail can drift along the magnetic field lines to high geomagnetic latitudes and be deposited in what is known as the *auroral zone*. Electrons with energies 0.5–20 keV entering the upper layers of the atmosphere at about 90–130 km excite oxygen atoms producing the green 558 nm and red 630 nm lines of oxygen characteristic of the aurorae.

There are a number of points of special interest about the structure of the magnetosphere. First of all, standard gas dynamics can be used to understand the overall structure of the magnetosphere, despite the fact that the plasma is collisionless on the scale of an

astronomical unit. The reason for this is the presence of the magnetic field which is frozen into the collisionless plasma. Despite the fact that the particles have very long mean free paths, the presence of even a very weak magnetic field ties the particles together. The fact that this works so well in the Earth's magnetosphere shows that this simplification can also be used in other astrophysical environments.

Related to this point is the fact that there is clear evidence for a *shock wave discontinuity* at the boundary of the magnetosheath. As described in Sect. 11.2.1, the thickness of the shock wave should be of the same order as the mean free path of the particles, despite the fact that the plasma is collisionless. The magnetic field is frozen into the plasma and the particles of the plasma gyrate about the magnetic field direction at the gyrofrequency. The effective friction and viscosity needed to transfer momentum and energy through the shock wave are provided by the magnetic stresses which couple the particles of the plasma. The distance over which energy and momentum are transferred is, to order of magnitude, the gyroradius of a proton in the interplanetary magnetic field. The mechanism by which energy is transferred is likely to be through various forms of plasma wave interaction involving the magnetic field. This is a somewhat complex subject but is of the greatest importance for astrophysical plasmas. The shock wave which bounds the magnetopause is one of the best examples known of a *collisionless shock wave*.

We have stated that the Solar Wind flows supersonically and, in the case of an ordinary gas, the flow is supersonic with respect to the local sound speed. Within the magnetosphere, however, the dynamics are dominated by the energy density and pressure of the magnetic field. In this case, the appropriate sound speed is the *Alfvén speed* $v_A = B/(\mu_0\rho)^{1/2}$. All sound speeds are roughly the square root of the ratio of the energy density of the medium to its inertial mass density $v \approx (\varepsilon/\rho)^{1/2}$ where ε is the energy density in the medium. Since the magnetosphere is magnetically dominated, $\varepsilon = B^2/2\mu_0$ and hence $v \approx B/(\mu_0\rho)^{1/2}$. The exact answer is the Alfvén speed quoted above which is the speed at which hydromagnetic waves can be propagated in a magnetically dominated plasma. Inserting appropriate values for the magnetosphere, $B = 5 \text{ nT}$, $n = 10^7 \text{ m}^{-3}$, we find $v_A = 35 \text{ km s}^{-1}$. Thus, the flow of the Solar Wind is certainly highly supersonic with respect to the Alfvén velocity within the magnetosphere. If any region of space is magnetically dominated, the appropriate sound speed is the Alfvén speed rather than the standard sound speed in the gas. Often, the flow of the Solar Wind is described as super-Alfvénic rather than supersonic.

11.5 Magnetic buoyancy

One of the remarkable features of magnetic flux freezing is that it gives substance to Faraday's concept of magnetic lines of force. The plasma and magnetic field are tied together and movements in the plasma are mirrored in the motions of the field lines which adjust themselves so that

$$\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} = 0 . \quad (11.81)$$

The magnetic field can therefore be stretched and distorted by motions in the plasma, be they ordered or turbulent, and so energy can be transferred from the kinetic energy of the plasma to the magnetic field. The concept of *tubes of force* therefore plays a central role in the magnetohydrodynamics of cosmic plasmas and their evolving topology can be visualised in terms of the response of tubes of force to motions in the plasma. These motions are most vividly displayed in the phenomena observed in the solar atmosphere and corona where the evolution of sunspots, solar flares and their associated magnetic fields can be observed evolving in real time (Fig. 11.4). The texts *Solar Magnetohydrodynamics* by Priest and *The Physics of Solar Flares* by Tandberg-Hanssen and Emslie provide full discussions of these and other magnetohydrodynamic phenomena (Priest, 1982; Tandberg-Hanssen and Emslie, 1988).

An important aspect of the physics of flux tubes is the concept of *magnetic buoyancy*. Following the exposition of Tandberg-Hanssen and Emslie, suppose an isolated magnetic flux tube is located in a plane-parallel stratified atmosphere. The number density of protons in the atmosphere is n_0 and that inside the flux tube is n_i . The atmosphere and the magnetic flux tube are assumed to be in pressure balance in a gravitational potential gradient and hence $p_0 = p_i$. The buoyancy arises from the fact that, since the inertial mass density in the magnetic field is much less than the mass deficit outside and inside the tube, the mass density inside the flux tube is less than that in the flux tube surrounding it and consequently, in the presence of a gravitational field, the lighter volume ‘floats up’ the potential gradient. Assuming that the material inside and outside the flux tube are at the same temperature and that the plasma is fully ionised, the electrons and ions each contribute a pressure nkT and so the equation of pressure balance is

$$2n_0kT = \frac{B^2}{2\mu_0} + 2n_i kT . \quad (11.82)$$

Therefore,

$$n_i = n_0 - \frac{B^2}{4\mu_0 kT} . \quad (11.83)$$

The buoyancy force acting upon the flux tube in the potential gradient is therefore

$$F = (n_0 - n_i)m_p g V = \frac{B^2 m_p g V}{4\mu_0 kT} , \quad (11.84)$$

where m_p is the mass of the proton and V is the volume of the flux tube. For an atmosphere in hydrostatic equilibrium, $d\rho/dx = -\rho g$ and, since $p = 2\rho_0 kT/m_p$, the scale height of the atmosphere H , defined by $d\rho_0/\rho_0 = dx/H$, is

$$H = \frac{2kT}{m_p g} . \quad (11.85)$$

Therefore,

$$F = \frac{B^2 V}{2\mu_0 H} . \quad (11.86)$$

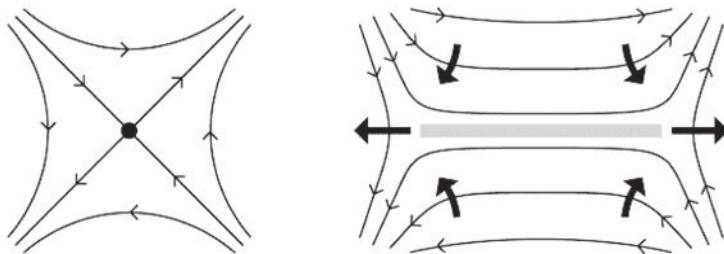


Fig. 11.9 Illustrating the process of reconnection of magnetic field lines about an X-point in the magnetic field distribution.
(Courtesy of Prof. Eric Priest.)

After the tube has risen a height H , it has acquired a kinetic energy

$$\frac{1}{2}Mu^2 \approx \frac{1}{2}\rho_0 Vu^2 = FH = B^2V/2\mu_0 , \quad (11.87)$$

because of the work done by the force F in accelerating the flux tube. The resulting velocity of the tube u is therefore

$$u = (B^2/\mu_0\rho_0)^{1/2} . \quad (11.88)$$

This is the local Alfvén speed $v_A = (B^2/\mu_0\rho_0)^{1/2}$. Thus, the flux tube rises up through the atmosphere at roughly the local Alfvén speed. In the solar atmosphere, the flux tubes are tied to the material of the outer layers of the Sun at their footpoints and so it is natural that the flux tubes develop into loop-like structures driven by the buoyancy of the magnetic field. This property of the buoyancy of magnetic flux tubes is very general and occurs wherever the matter density inside the tube is less than that outside and the system is located in a gravitational potential gradient. Similar processes are expected to take place in the magnetic fields confined to the plane of the Galaxy and in accretion discs. More details of these concepts and their more general applicability are given by Parker (1979).

11.6 Reconnection of magnetic lines of force

The magnetic fields in the surface layers of the Sun contain large amounts of energy which is available for powering energetic phenomena such as solar flares. Energy is released because of the finite electrical conductivity of the plasma which not only enables the field lines to diffuse relative to the plasma, but also leads to the dissipation of the energy of the magnetic field with consequent heating of the plasma. This process is particularly effective if the magnetic field lines run in opposite directions, as is the case in current sheets. The magnetic field lines can reconnect with the resistive dissipation of energy. Magnetic reconnection, illustrated in Fig. 11.9, takes place in solar flares in which the changing topology of the magnetic field lines has been observed. Similar processes are also inferred to take place in the magnetotail of the Earth's magnetosphere. Magnetic reconnection is also observed in large plasma machines such as tokamaks.

Despite the empirical evidence for magnetic reconnection, the detailed microphysics is not fully understood, largely because the electrical conductivity of cosmic plasmas is so high that the dissipation time-scales are generally predicted to be very much longer than those observed. The simplest estimate of the time-scale involved in the release of magnetic energy is the diffusion time-scale (11.54) derived in Sect. 11.2.2, $\tau_c \approx \sigma \mu_0 L^2$ where $\sigma = 10^{-3} T^{3/2}$ siemens m⁻¹. For typical solar flares, representative values for pre-flare conditions are $T = 2 \times 10^6$ K, $L = 10^7$ m, $B = 0.03$ T and $n = 10^{16}$ m⁻³ (Tandberg-Hanssen and Emslie, 1988). The resulting dissipation time-scale is of the order of 10⁷ years, far in excess of the time-scale associated with solar flares, which are of the order of hours or less. This process is clearly inadequate to account for the rate at which energy is extracted from the magnetic field.

The underlying problems are the very large values of the electrical conductivity and the large length-scales over which dissipation takes place. The issues involved have been clearly expounded by Kulsrud, Priest, Forbes, Tandberg-Hansen and Emslie, among others (Kulsrud, 2005; Priest, 1982; Priest and Forbes, 2000; Tandberg-Hanssen and Emslie, 1988). An important advance was made in the pioneering papers by Sweet and Parker (Sweet, 1958; Parker, 1957) who realised that in neutral sheets, the physical scales could be very much reduced in the direction perpendicular to the sheet. The Sweet–Parker mechanism represented a dramatic improvement over the simple dissipation model described above.

The model is illustrated in Fig. 11.10. The magnetic field reverses direction along the x -axis and oppositely directed field lines are convected towards the x -axis at velocity v in the y -direction, the sheet being taken to be infinite in the z -direction. To conserve mass in the steady state, the inflow of plasma and magnetic field are balanced by outflow along the $\pm x$ -directions. The object of the calculation is to work out the rate at which magnetic field energy is dissipated by ohmic losses and the time-scale over which it is released. A closed loop path is constructed about the dissipation region and then Ampère's theorem in integral form is used to find the current flowing through the loop. Since $\mathbf{J} = \text{curl} \mathbf{B}/\mu_0$, this relation can be written in integral form using Stokes' theorem,

$$\int_S \mathbf{J} \cdot d\mathbf{S} = \frac{1}{\mu_0} \int_C \mathbf{B} \cdot d\mathbf{l}, \quad (11.89)$$

where J is the current density passing through the loop and the integral on the right-hand side is taken round the closed loop. For the geometry shown in Fig. 11.10, we find, to order of magnitude,

$$lLJ \approx 2BL/\mu_0, \quad J \approx 2B/l\mu_0, \quad (11.90)$$

where l is the width of the loop and L its length, as indicated in the diagram. Thus, as the value of l decreases, the current density J in the reconnection region increases so that, even if the conductivity of the region is very high, it would appear that there can be efficient ohmic losses in the neutral sheet if the width of the dissipation region l is narrow enough. A lower limit to the width of this region is set by the gyroradii of the particles in the field. If the resistivity of the plasma is $\eta = \sigma^{-1}$, the dissipation rate is $\eta J^2 = 4\eta B^2/\mu_0^2 l^2$ per unit volume.

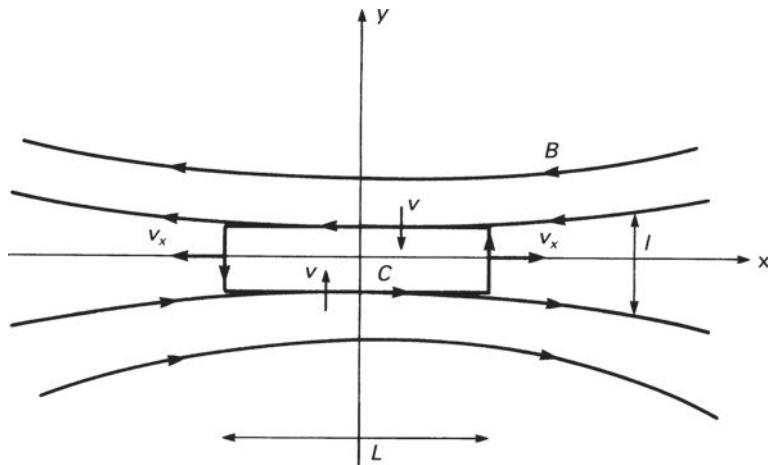


Fig. 11.10

Illustrating the process of magnetic field line reconnection according to the Sweet–Parker picture. The B_x magnetic field component reverses direction at the x -axis leading to a large current density in the z -direction. Magnetic field lines are convected into the neutral sheet in the y -direction and this is balanced by the outflow of material along the positive and negative x -axes. The dimensions of the reconnection region are shown on the diagram. It is assumed that the geometry extends indefinitely in the $\pm z$ -direction (Tandberg-Hanssen and Emslie, 1988).

What has been omitted from this argument is the influence of the gas pressure in the neutral sheet. The plasma and the magnetic field are convected into the reconnection region and cannot be compressed indefinitely. In the steady state, the dissipation of the energy in the magnetic field heats up the plasma and contributes to the pressure in the current layer. Furthermore, in the steady state, the pressure balance must be preserved along the y -axis. Since the magnetic field is zero on the axis of the current sheet, pressure balance requires the thermal pressure in the current sheet to be equal to the magnetic pressure just outside the current layer. Therefore, on axis, the pressure of the gas must be of order $p_0 \approx B^2/2\mu_0$.

Now, in the current sheet, we can neglect the magnetic field and so the equation of motion of the plasma along the x -axis is

$$\rho \frac{dv_x}{dt} = -\frac{\partial p}{\partial x}. \quad (11.91)$$

In the steady state $\partial v_x / \partial t = 0$ and, since $d/dt = \partial/\partial t + (\mathbf{v} \cdot \nabla)$, (11.91) can be written in Eulerian coordinates,

$$\rho v_x \frac{\partial v_x}{\partial x} = -\frac{\partial p}{\partial x}. \quad (11.92)$$

Integrating from $x = 0$ to $x = \pm\infty$ and setting $p_\infty = 0$, $p_0 = \frac{1}{2}\rho v_x^2$. But, we have shown that $p_0 = B_x^2/2\mu_0$ and so the velocity of escape of the material along the x -axis is of the order of the Alfvén speed $v_x \approx B/(\mu_0\rho)^{1/2} = v_A$, as might have been expected. This outflow is balanced by inflow along the y -axis and hence, by mass conservation, the speed at which the material is convected into the dissipation region is $v = (l/L)v_A$.

The dissipation rate by ohmic losses is equal to the rate at which magnetic energy is convected into the reconnection region, that is,

$$\int_V \eta J^2 dV = \int_S \frac{B^2}{2\mu_0} v dS . \quad (11.93)$$

Therefore, per unit length in the z -direction,

$$\eta J^2 (Ll) = \frac{B^2}{2\mu_0} 2vL , \quad \eta J^2 l = \frac{B^2 v}{\mu_0} . \quad (11.94)$$

But, from (11.90), $J = 2B/\mu_0 l$ and hence

$$v = 4\eta/l\mu_0 . \quad (11.95)$$

Combining this expression with the relation $v = (l/L) v_A$,

$$v^2 = \frac{4\eta}{\mu_0 L} v_A , \quad l^2 = \frac{4\eta L}{\mu_0 v_A} . \quad (11.96)$$

Notice that the thickness of the reconnection region l has disappeared from the expression for v .

It is now convenient to introduce a ‘longitudinal’ magnetic Reynolds number R_m , the *Lundquist number S*, in which the length-scale L is the length of the neutral sheet and v the Alfvén speed v_A ,

$$S = \sigma \mu_0 v L = \frac{\mu_0 v_A L}{\eta} . \quad (11.97)$$

Note that the Lundquist number S is the magnetic Reynolds number R_m with $v = v_A$. Therefore, the reconnection velocity v_r into the neutral sheet is

$$v_r = \left(\frac{4\eta}{\mu_0 L} v_A \right)^{1/2} = 2v_A/S^{1/2} , \quad (11.98)$$

and the thickness of the neutral sheet is

$$l = \left(\frac{4\eta L}{\mu_0 v_A} \right)^{1/2} = 2L/S^{1/2} . \quad (11.99)$$

Adopting the values for a typical solar flare given above, we find $v_A = 7 \times 10^6 \text{ m s}^{-1}$ and $S = 2 \times 10^{14}$. Therefore, the velocity at which magnetic field lines are convected into the neutral sheet is only 10^{-7} of the Alfvén speed. This is, however, a significant improvement over the time-scale for the diffusive dissipation of energy over a length-scale L which is $\tau_D \sim \sigma \mu_0 L^2$. The diffusive velocity can be written as $v_D \sim L/\tau_D \sim 1/\sigma \mu_0 L \sim v_A/S$, which is longer than the reconnection velocity v_r by a factor of roughly $S^{1/2}$. Thus, the reconnection time is 10^7 times less than the diffusive time-scale and so of the order of a year. This figure is still very much longer than the time-scales associated with solar flares, but a very significant advance over the diffusive time-scale. We can also estimate the amount of energy released in this reconnection model. The total amount of magnetic energy in the neutral sheet is $(B^2/2\mu_0)V$ where $V \sim L^2 l \sim L^3/S^{1/2}$. Inserting the above values into these relations, we find $E \sim 3 \times 10^{23} \text{ J}$, the energy of a somewhat modest solar flare but, as noted above, this energy is released over a time-scale of a year rather than hours or less.

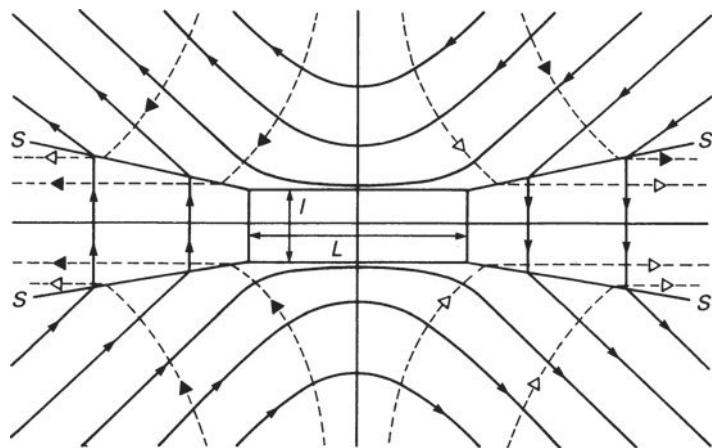


Fig. 11.11 The geometry of reconnection according to Petschek (1964). The solid lines represent the magnetic field lines and the dashed lines the streamlines of the plasma flow. The standing shock waves are labelled S. It can be seen that the magnetic field lines do indeed reconnect in this picture (Tandberg-Hanssen and Emslie, 1988).

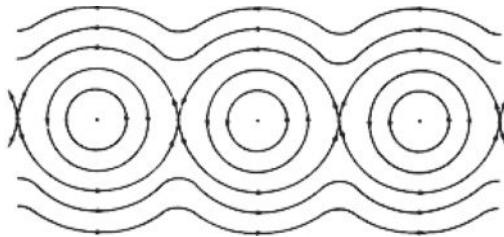


Fig. 11.12 Illustrating the formation of magnetic islands and O and X-type neutral points as a result of the development of the tearing mode instability in a neutral current sheet (Tandberg-Hanssen and Emslie, 1988).

In 1964, it was pointed out by Petschek that the dissipation rate can be increased if standing shock waves form on either side of the neutral sheet, creating the geometry shown in Fig. 11.11 (Petschek, 1964). The magnetic field lines reconnect as shown in the sketch. According to Petschek's analysis, the reconnection velocity can be as large as $v_A / \ln S$. The structure of these neutral sheets and their associated shock waves requires careful attention to the detailed microphysics and goes far beyond what can be covered here. Priest and Forbes generalised the models for the reconnection of magnetic field lines in neutral sheets and showed that the reconnection velocity can almost be as large as the Alfvén velocity v_A , but the reconnection speed is critically dependent upon the boundary conditions (Priest and Forbes, 1986).

There are a number of ways in which the energy release can be modified within the neutral current sheet. The current sheet has been found to be susceptible to *tearing mode instabilities* in which the sheet breaks up into a number of X and O-neutral points as illustrated in Fig. 11.12. As a result, the current sheet is converted into a layer of current

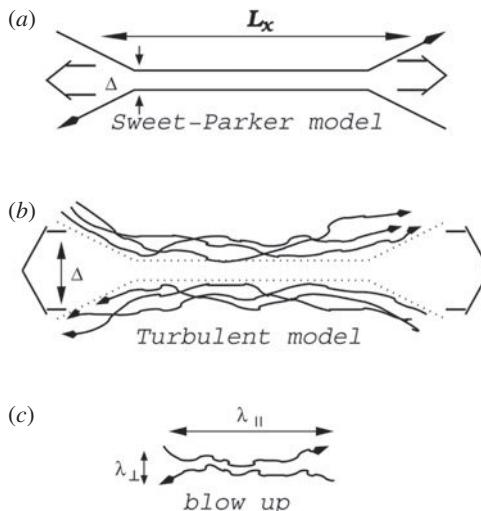


Fig. 11.13

(a) The Sweet–Parker model of reconnection. (b) Reconnection in a weakly stochastic magnetic field according to Lazarian and Vishniac (1999). The outflow is limited by the diffusion of magnetic flux lines which depends on the stochasticity of the field lines. (c) An individual small reconnection region. Reconnection over small patches of the magnetic field distribution determines the local reconnection rate. The global reconnection rate is substantially larger than in the Sweet–Parker case as many independent patches come together (Lazarian *et al.*, 2004).

filaments. The flow pattern is different from that in the simple neutral current sheet with magnetic islands collapsing and dissipating energy with a much smaller length-scale than that of the current sheet itself. The effect of the instability is not necessarily to enhance the reconnection rate but rather it makes the process impulsive and bursty.

In addition to these instabilities, the resistivity of the plasma may be enhanced because of the phenomenon of *anomalous resistivity*. The resistivity of the plasma may be significantly increased because of the presence of waves or turbulence in the plasma. The effect of these waves is to move the particles of the plasma coherently so that an individual electron interacts with the collective influence of a large number of particles rather than with a single particle. An example of the type of plasma instability which could have this effect in the neutral sheet is the *ion-acoustic instability* in which the drift velocity of the plasma exceeds the ion sound speed $c_i = (kT/m_p)^{1/2}$. This condition is likely to be satisfied in the neutral current sheets in solar flares.

The picture of reconnection developed above is essentially a two-dimensional representation of what is in fact a three-dimensional problem. In three dimensions, topologically tubes of magnetic flux can cross each other and this leads to an enhanced reconnection rate at many different points within the reconnection volume. If the medium is even mildly turbulent, the reconnection rate can be significantly enhanced by the process which Lazarian and Vishniac describe as field wandering induced by turbulence (Lazarian and Vishniac, 1999). They found that, once mild turbulence is included into three-dimensional simulations of the distribution of the magnetic flux tubes, the reconnection speed is much faster than the Sweet–Parker rate and is independent of the resistivity of the plasma. The

difference between the models is illustrated in Fig. 11.13, which shows the results of computer simulations of the process of reconnection according to the Sweet–Parker model and in the presence of a turbulent plasma (Kowal *et al.*, 2009). Lazarian makes the point that it is now feasible to include turbulence properly into computations of the physics of astrophysical plasmas because of the exponential growth in computer power over recent years.

This simplified discussion disguises a host of issues in the magnetohydrodynamics and plasma physics of the physics of reconnection of magnetic field lines. One of the main concerns is whether or not the models are fully self-consistent when the many plasma effects and instabilities are taken into account. The books *Magnetic Reconnection* by Priest and Forbes and *Plasma Physics for Astrophysics* by Kulsrud provide more details of many of these issues (Priest and Forbes, 2000; Kulsrud, 2005).

There is no doubt that the reconnection of magnetic field lines is a key process in many astrophysical plasmas, including those involved in star formation, in extragalactic radio sources and in the accretion discs about compact objects.



PART III

HIGH ENERGY ASTROPHYSICS IN OUR GALAXY

12.1 The interstellar medium in the life cycle of stars

The understanding of the nature and physical properties of the interstellar medium is of the first importance astrophysically since new stars are formed in dense regions of the interstellar gas and the medium is continually replenished by mass loss from stars and by metal-rich material processed in supernova explosions. Thus, the interstellar medium plays a key role in the birth-to-death cycle of stars. The same diagnostic tools are applicable to the study of diffuse gas and magnetic fields anywhere in the Universe, be they galaxies, the intergalactic gas or the environs of active galactic nuclei. Furthermore, interstellar gas will prove to be an essential ingredient in the fuelling of active galactic nuclei.

The mass of the interstellar gas amounts to about 5% of the visible mass of our Galaxy. In the Galactic plane close to the Sun, the overall gas density is to about 10^6 particles m⁻³, but there are very wide variations in density and temperature from place to place throughout the interstellar medium.

12.2 Diagnostic tools – neutral interstellar gas

12.2.1 Neutral hydrogen: 21-cm line emission and absorption

Neutral hydrogen emits line radiation at a frequency $\nu_0 = 1420.4058$ MHz ($\lambda_0 = 21.1$ cm) through an almost totally forbidden hyperfine transition in which the spins of the electron and proton change from being parallel to antiparallel. The spontaneous transition probability is $A_{21} = 2.85 \times 10^{-15}$ s⁻¹ for the ground state of hydrogen, that is, about once every 10^7 years. Although this is a very rare transition, there is so much neutral hydrogen in the Galaxy that the line is readily detectable. Because there are two possible orientations of the spins of both the electron and the proton, there are four stationary states, three degenerate in the upper state and one in the lower state. Because of the very small transition probability, collisions and other processes have time to establish an equilibrium distribution of hydrogen atoms in the upper and lower states, labelled 2 and 1, respectively, and so the ratio of the number of atoms in these states is given by the Boltzmann distribution $N_2/N_1 = (g_2/g_1) \exp(-h\nu_0/kT)$. T is the excitation temperature and g_2 and g_1 are the statistical weights of the upper and lower levels, $g_2/g_1 = 3$. The excitation temperature T

is called the *spin temperature* T_s . Under all cosmic conditions $h\nu_0/k = 7 \times 10^{-2}$ K $\ll T_s$ and therefore $N_2/N_1 = 3$.

If the emitting region is optically thin, only spontaneous emission need be considered and so the emissivity κ_{21} of the gas is

$$\kappa_{21} = \frac{g_2}{g_2 + g_1} N_H A_{21} h \nu_0 = \frac{3}{4} N_H A_{21} h \nu_0 , \quad (12.1)$$

where N_H is the number density of neutral hydrogen atoms.

If the neutral hydrogen is distributed along the line of sight from the observer, the flux density received within solid angle Ω , say, the solid angle subtended by the beam of the radio telescope, is

$$S = \int \frac{\kappa_{21}(r)}{4\pi r^2} \Omega r^2 dr ; \quad I = \frac{S}{\Omega} = \frac{3}{16\pi} A_{21} h \nu_0 \int N_H dr , \quad (12.2)$$

where r is distance along the line of sight. $I = S/\Omega$ is the intensity of radiation in that direction and is a measure of the total *column density* of neutral hydrogen along the line of sight $\int N_H dr$. In this calculation I is measured in W m⁻² and is equal to the integral of the intensity of radiation per unit bandwidth I_ν over the line profile $I = \int I_\nu d\nu$.

Because of its very small transition probability, the natural linewidth of the 21-cm line is very narrow. If the neutral hydrogen is in motion relative to the observer, Doppler shifts of the 21-cm line emission can be readily measured by making observations with a multi-channel 21-cm line receiver. This provides a very powerful tool for investigating the dynamics of neutral hydrogen in our own and in other galaxies.

Non-thermal radio sources such as supernova remnants and extragalactic radio sources have smooth synchrotron spectra at radio wavelengths and therefore, if neutral hydrogen clouds lie along the line of sight to the radio source, absorption features in the radio source spectrum are expected. The absorption coefficient for 21-cm line absorption can be worked out using the same technique discussed in the case of thermal bremsstrahlung absorption at radio wavelengths in Sect. 6.5.2. The relation (6.51) can be used in the low frequency limit $h\nu \ll kT$ in which case the black-body intensity is $I_\nu = 2kT/\lambda^2$ and so

$$\chi_\nu I_\nu = \chi_\nu \frac{2kT}{\lambda^2} = \frac{\kappa_{21}}{4\pi} . \quad (12.3)$$

If $\Delta\nu$ is the linewidth of the neutral hydrogen profile, the emissivity per unit frequency interval is

$$\kappa_{21} = \frac{3}{4} N_H A_{21} h \frac{\nu_0}{\Delta\nu} . \quad (12.4)$$

Therefore, the absorption coefficient χ_ν is

$$\chi_\nu = \frac{3}{32\pi} \frac{A_{21} h c^2}{\nu_0^2 k T_s} \frac{\nu}{\Delta\nu} N_H . \quad (12.5)$$

If the radio source has brightness temperature $T_b \gg T_s$, its observed spectrum is

$$I_\nu = I_0(\nu) \exp(-\tau_\nu) ; \quad \tau_\nu = \chi_\nu l , \quad (12.6)$$

where l is the path length through the cloud. Evidently, the interpretation of the absorption spectrum requires knowledge of the spin temperature T_s of the intervening cloud. The absorption profile cannot normally be fitted by a single Gaussian function but consists of a number of components with different velocities and linewidths resulting from a combination of systematic and random velocities of the clouds along the line of sight to the radio source. The neutral hydrogen absorption measurements give information about the small scale structure and velocity dispersion of the neutral hydrogen along the line of sight on the scale of the angular size of the background source, whereas the emission profiles provide information on the scale of the beamwidth of the radio telescope.

12.2.2 Molecular radio lines

Long before the advent of radio astronomy, it was known that there exist significant abundances of molecules in interstellar space. The molecules CH, CH^+ and CN possess electronic transitions in the optical waveband and absorption features associated with these were well known features of the spectra of bright stars. The advantage of observing molecules at centimetre and millimetre wavelengths is that, unlike the optical waveband, there is no extinction because of interstellar dust. The first interstellar molecule to be detected at radio wavelengths was the hydroxyl radical OH which was observed in absorption against the bright radio source Cassiopaeia A in 1963. Soon afterwards, the hydroxyl lines were observed in emission, the surprise being that the sources were very compact and variable in intensity. The corresponding brightness temperatures were very great indeed, $T_b \geq 10^9$ K, implying that some form of maser action must be involved. A key discovery was the great intensity of the carbon monoxide molecule CO, first observed in 1970. Since that date, the number of detected molecular species has multiplied rapidly (Table 12.1). In dusty regions of interstellar space, where the molecules are protected from dissociating optical and ultraviolet radiation, complex organic molecules with up to 13 constituent atoms have been discovered. The molecules observed are composed of the most abundant elements: hydrogen (and deuterium), nitrogen, carbon, sulphur, silicon and oxygen and their isotopes. In some sources, the molecular line spectra are so rich that the noise in the spectra is the result of the superposition of a myriad of weak molecular lines.

Molecules can emit line radiation associated with transitions between electronic, vibrational and rotational levels. The highest energy transitions are those associated with *electronic transitions* and normally these lie in the optical region of the spectrum. *Vibrational transitions* are associated with the molecular binding between atoms of the molecule which can be represented by a simple harmonic oscillator; transitions between these vibrational levels typically lie in the infrared spectral region $h\nu \sim 0.2$ eV.

The lowest energy transitions are those between rotational energy levels. The frequencies of these *rotational transitions* can be found from the rules of quantisation of angular momentum. According to quantum mechanics, the angular momentum \mathbf{J} is quantised such that it can only take discrete values given by the relation $J^2 = j(j+1)\hbar^2$ where the angular momentum quantum number j takes integral values, $j = 0, 1, 2, \dots$. The energy of each of these stationary states is given by exactly the same formula which relates energy and angular momentum in classical mechanics, $E = \mathbf{J}^2/2I$, where I is the moment of inertia

Table 12.1 This list of interstellar molecules is arranged in columns showing the numbers of atoms which make up each molecule. The data are taken from the web site http://www.astrochymist.org/astrochymist_ism.html maintained by D.E. Woon. In each column, the order is by date of publication of the discovery according to Woon's table. Isotopic species have generally not been listed. Tentative detections are indicated by a question mark. This table was compiled in January 2009.

2	3	4	5	6	7	8	9
CH	H ₂ O	NH ₃	HC ₃ N	CH ₃ OH	CH ₃ CHO	CHOOC ₃	CH ₃ OCH ₃
CN	HCO ⁺	H ₂ CO	HCOOH	CH ₃ CN	CH ₃ CCH	CH ₃ C ₃ N	CH ₃ CH ₂ OH
CH ⁺	HCN	HNCO	CH ₂ NH	NH ₂ CHO	CH ₃ NH ₂	C ₇ H	CH ₃ CH ₂ CN
OH	OCS	H ₂ CS	NH ₂ CN	CH ₃ SH	CH ₂ CHCN	CH ₃ COOH	HC ₇ N
CO	H ₂ S	C ₃ N	H ₂ CCO	C ₂ H ₄	HC ₅ N	CH ₂ OHCHO	CH ₃ C ₄ H
H ₂	HNC	HNCS	C ₄ H	C ₅ H	C ₆ H	C ₆ H ₂	C ₈ H
SiO	N ₂ H ⁺	HOCO ⁺	SiH ₄	CH ₃ NC(?)	c-C ₂ H ₄ O	CH ₂ CHCHO	CH ₃ CONH ₂
CS	C ₂ N	C ₃ H	c-C ₃ H ₂	HC ₂ CHO	CH ₂ CHOH	CH ₂ CCHCN	C ₈ H ⁻
SO	SO ₂	C ₃ O	CH ₂ CN	H ₂ CCCC	C ₆ H ⁻	NH ₂ CH ₂ CN	CH ₂ CHCH ₃
SiS	HDO	HCNH ⁺	C ₅	HC ₃ NH ⁺			
NS	HCO	H ₃ O ⁺	SiC ₄	C ₅ N			
C ₂	HNO	C ₃ S	H ₂ CCC	C ₄ H ₂			
NO	OCN ⁻	c-C ₃ H	CH ₄	HC ₄ N			
HCl	HCS ⁺	C ₂ H ₂	HCCNC	c-H ₂ C ₃ O			
NaCl	HOC ⁺	HC ₂ N	HNCCC	CH ₂ CNH			
AlCl	c-SiC ₂	H ₂ CN	H ₂ COH ⁺	C ₅ N ⁻			
KCl	MgNC	SiC ₃	C ₄ H ⁻				
AlF	C ₂ S	CH ₃	CNCHO				
PN	C ₃	C ₃ N ⁻					
SiC	CO ₂	PH ₃ (?)					
CP	CH ₂	HCNO					
NH	C ₂ O						
SiN	NH ₂						
SO ⁺	N ₂ O						
CO ⁺	MgCN						
HF	H ₃ ⁺						
LiH(?)	SiCN						
SH	AlNC						
FeO(?)	SiNC						
N ₂	HCP						
CF ⁺	CCP						
O ₂							
PO							

In addition, there are molecules with 10 atoms, (CH₃)₂CO, HOCH₂CH₂OH, CH₃CH₂CHO and CH₃(C≡C)₂CN, 11 atoms, H(C≡C)CN and CH₃C₆N, 12 atoms C₆H₆ and 13 atoms, H(C≡C)CN.

of the molecule about its rotation axis. When a photon is emitted or absorbed, one unit of angular momentum has to be created or absorbed and hence j changes by one unit. The selection rule for these electric dipole transitions is therefore $\Delta j = \pm 1$. The energy of the photon emitted in the rotational transition from the stationary state j to that corresponding to $j - 1$ is therefore

$$h\nu = E(j) - E(j-1) = [j(j+1) - (j-1)j]\hbar^2/2I = j\hbar^2/I. \quad (12.7)$$

For a diatomic molecule composed of atoms of masses M_1 and M_2 , the moment of inertia is $I = \mu r_0^2$ where μ is the reduced mass of the molecule $\mu = M_1 M_2 / (M_1 + M_2)$ and r_0 is the equilibrium spacing of the atomic nuclei. Therefore, $\nu = j\hbar/4\pi^2\mu r_0^2$. This calculation illustrates an important feature of the rotational spectrum of molecules – the rotational lines are equally spaced in frequency, often referred to as the *rotational ladder* of the molecule's spectrum. For CO, for example, $\mu = 6.859$ atomic mass units = 1.11×10^{-26} kg and $r_0 = 1.128 \times 10^{-10}$ m. Therefore, the lowest frequency rotational transition, $j = 1 \rightarrow 0$, is 115 GHz or $\lambda = 2.6$ mm. The next transitions in the rotational ladder have frequencies 230 GHz ($j = 2 \rightarrow 1$), 345 GHz ($j = 3 \rightarrow 2$), and so on. Corresponding results are found for more complex molecules involving more than two atoms. The transition probabilities depend upon the net electric dipole moment of the molecule and so symmetrical molecules such as hydrogen H₂ do not emit electric dipole radiation, but asymmetrical molecules such as CO and HC₁₁N are sources of millimetre line emission.

Other molecules, such as the hydroxyl radical OH and formaldehyde H₂CO, have permitted transitions in the radio waveband through molecular doubling processes. In the case of a diatomic molecule such as OH, the doubling results from the interaction between the electronic motions in the molecule and the rotation of the molecule as a whole.

Generally, molecular line emission provides information about denser regions of the interstellar gas than the 21-cm line emission because the molecules are fragile and can be dissociated by optical and ultraviolet photons. They are therefore predominantly found in dense molecular clouds with densities $N_{\text{H}} \approx 10^{9-10} \text{ m}^{-3}$ within which the molecules are shielded from the interstellar flux of high energy photons by dust and also by *self-shielding* by the molecular hydrogen at the peripheries of the clouds. The higher frequency transitions of a particular rotational ladder have larger transition probabilities and so can be used to determine much higher molecular densities within the clouds.

The most common molecule is expected to be molecular hydrogen, H₂, but, because it has no electric dipole moment, no rotational transitions are observed. Molecular hydrogen was, however, detected by the Copernicus satellite in absorption in the ultraviolet region of the spectrum through its electronic transitions. These observations confirmed that H₂ is present in large quantities in the interstellar gas. The next most abundant molecule is carbon monoxide, CO, which, as shown above, emits strong permitted line radiation at 2.6 mm and its harmonics. Strong CO radiation has been detected throughout the Galaxy and provides complementary information to that provided by surveys of the 21-cm line of neutral hydrogen. The importance of the CO observations is that, wherever there exist CO molecules, there must also exist H₂. The excitation mechanism for the CO molecules is collisions with hydrogen molecules and so the CO observations provide a measure of the number density of H₂ molecules.

Table 12.1 contains a wide variety of different types of molecule – organic molecules, inorganic molecules, free radicals and molecular ions. There is also a great range in the size of the molecules. Many consist of two atoms but very much larger examples are observed, the record holder being the acetylenic chain molecule HC_{11}N with thirteen atoms. Several important patterns are discernible in Table 12.1. For example, there is the remarkable sequence of acetylenic chain molecules HCN , HC_3N , HC_5N , HC_7N , HC_9N , HC_{11}N – there must be some simple mechanism for lengthening a pre-existing chain. Searches have been made for the simplest amino acid molecules, such as glycine, but to date no confirmed detection has been reported.

The Universe contains an overwhelming majority of hydrogen atoms and so the existence of many unsaturated species, that is, species containing double and triple bonds, is remarkable. If a giant molecular cloud were in thermodynamic equilibrium at a temperature of, say, 50 K, the only species expected would be saturated molecules such as CH_4 , NH_3 , H_2O , and so on. There would be no CO nor any of the unsaturated multiply-bonded species such as HC_{11}N . The inference is that the interstellar medium must be very far from thermodynamic equilibrium. The principal reactions which determine the abundances of the different molecular species are gas-phase reactions and chemical reactions taking place on grain surfaces. Besides their obvious interest for *interstellar chemistry*, the existence of these molecules provides an important tool for probing the physical conditions and velocity fields deep inside star-forming regions. Some of the largest redshift galaxies discovered in the submillimetre waveband and large redshift radio-quiet quasars have been detected by their millimetre line emission, providing evidence for the early build up of the heavy elements in these galaxies.

12.2.3 Optical and ultraviolet absorption lines

Atoms observed in absorption in the optical waveband must possess excited states within about 4 eV of the ground state. It turns out that relatively few of the more abundant species satisfy this criterion, the most important being the transitions of Na I , Ca II , Ca I , K I , Ti II and Fe I . These absorption lines have been observed in stellar spectra, the strongest being those of Ca II and Na I which are both doublets, the pairs of lines being known as the H and K lines of calcium at $\lambda 396.85$ and $\lambda 393.37$ nm, respectively, and the D lines of sodium, D1 $\lambda 589.59$ and D2 $\lambda 589.00$ nm. The ultraviolet region of the spectrum, 100–300 nm, corresponds to higher energy transitions and a very much wider range of interstellar atoms and molecules can be studied, in particular, atomic and molecular hydrogen and essentially all the common heavy elements. The Orbital Astronomical Observatories, OAO-II and Copernicus, and the International Ultraviolet Observatory (IUE) revolutionised studies of the interstellar medium, and absorption lines associated with all the common elements in various stages of ionisation have been detected.

The interpretation of interstellar absorption spectra requires knowledge of atomic absorption cross-sections as a function of frequency $\sigma(\nu)$. For an atom at rest, the absorption cross-section may be calculated quantum mechanically in the case of simple atoms or, in most cases, derived from laboratory experiments. The frequency dependence of the absorption cross-section depends upon the mechanism of line broadening. For interstellar

absorption lines the most important are *Doppler broadening*, which may result either from the random motions of the absorbing atoms in the gas or from bulk motions within the clouds, and *radiation or natural broadening* which results from the fact that the atom remains only a finite time Δt in an excited state. A rough estimate of the natural linewidth can be found from Heisenberg's uncertainty principle $\Delta E \approx h/\Delta t$ and so $\Delta v \approx \Delta t^{-1}$.

In the simplest optically thin case, the optical depth of the line τ_v is a measure of the total column density of the atomic species, $\tau_v = \int \sigma_i N_i dl$. The story becomes more complicated when τ_v is very large because natural broadening of the lines becomes important. Astronomers work in terms of the equivalent width W of the absorption lines which is the amount of energy extracted from the continuum expressed as a linewidth,

$$W = \int \left(1 - \frac{I_v}{I_{vc}}\right) dv, \quad (12.8)$$

where I_{vc} is the continuum spectrum expected in the absence of the absorption line. The relation between W and the column density of the species is known as the *curve of growth*.

Ultraviolet observations of this type have resulted in a number of important discoveries about the nature of the interstellar gas. For example,

- (i) Molecular hydrogen H₂ has been discovered in large quantities in the interstellar gas but there are wide variations in its abundance relative to atomic hydrogen. H₂ molecules can only survive if they are shielded from optical and ultraviolet photons in regions with density $N_H \geq 10^9 \text{ m}^{-3}$.
- (ii) The interstellar abundances of the heavy elements are less than their cosmic values by factors up to 10^3 – 10^4 . A considerable fraction of these ‘missing’ elements is locked up in interstellar dust grains.
- (iii) Atomic deuterium has been detected with abundance relative to neutral hydrogen of about 1.5×10^{-5} . This value is remarkably constant wherever deuterium has been detected in the interstellar gas and is a very high abundance for such a fragile element. A convincing case can be made that deuterium was synthesised in the non-equilibrium conditions during the first few minutes of the Hot Big Bang (Longair, 2008).
- (iv) Highly ionised oxygen O VI has been detected as a broad absorption feature in the spectra of the majority of hot stars. This is evidence for a hot component of the interstellar gas having $2 \times 10^5 \leq T \leq 10^6 \text{ K}$. Similar broad features have been observed in the lines of C IV in the spectra of halo stars and of B stars in the Magellanic Clouds. These are attributed to absorption in a highly ionised, hot gaseous halo about our Galaxy.

12.2.4 X-ray absorption

The process of photoelectric absorption was described in Sect. 9.1. If the standard cosmic abundances of the elements are assumed, the dependence of the absorption coefficient upon photon energy shown in Fig. 9.2 is obtained, displaying the characteristic K-absorption edges of the common elements. A useful smooth approximation to that absorption curve is

$$\tau_x = 2 \times 10^{-26} \left(\frac{h\nu}{1 \text{ keV}}\right)^{-8/3} \int N_H dl, \quad (12.9)$$

where $\int N_{\text{H}} \, dl$ is the *column depth* of atomic hydrogen, expressed in atoms m^{-2} , and $h\nu$ in keV. The absorption may take place within the source itself or in the intervening medium, for example, in our own Galaxy.

12.3 Ionised interstellar gas

12.3.1 Thermal bremsstrahlung

Thermal bremsstrahlung emission and absorption were discussed in some detail in Sect. 6.5. The characteristic signature of bremsstrahlung is that the emissivity spectrum in $\text{W m}^{-3} \text{Hz}^{-1}$ is flat up to frequencies $h\nu \approx kT$, beyond which there is an exponential cut-off. The intensity of radiation per unit bandwidth depends upon the combination of parameters $N_e^2 T^{-1/2}$ and so the bremsstrahlung intensity observed along the line of sight is

$$I_\nu = A \int N_e^2 T^{-1/2} \, dr , \quad (12.10)$$

where the constant A is given in (6.47). At radio wavelengths, diffuse regions of ionised hydrogen at $T \approx 10^4 \text{ K}$ are strong sources of bremsstrahlung. If the region is compact, the region becomes optically thick and the absorption coefficient can be derived using Kirchhoff's law (Sect. 6.5.2). The radio spectra of the most compact regions of ionised hydrogen found in the vicinity of regions of star formation have the form $I_\nu \propto \nu^2$ at centimetre wavelengths, the signature of bremsstrahlung absorption (Fig. 6.4). Provided the source is homogeneous, both T and N_e can be found from such spectra. At the very lowest radio frequencies, $\nu \leq 10 \text{ MHz}$, thermal bremsstrahlung absorption by the diffuse ionised interstellar gas becomes important and the Galactic plane is observed in absorption against the background of Galactic non-thermal radio emission (Ellis, 1982).

At X-ray wavelengths, bremsstrahlung has been observed from the diffuse intergalactic gas in rich clusters of galaxies (Fig. 4.5) and from the shells of supernova remnants. Emission lines of very highly ionised species such as Fe xxv have also been observed in these sources, confirming the presence of a very hot gas with $T \approx 10^7\text{--}10^8 \text{ K}$. The soft X-ray emission from the plane of the Galaxy is interpreted as the diffuse thermal bremsstrahlung of the hot component of the interstellar gas which is also responsible for the ultraviolet O vi absorption lines. The temperature of gas responsible for O vi lines lies in the range $(1\text{--}3) \times 10^6 \text{ K}$.

12.3.2 Permitted and forbidden transitions in gaseous nebulae

Strong emission lines are observed from high-density regions of the interstellar gas which are excited by the ultraviolet emission of hot stars. These may be either regions in which massive young stars have formed or the vicinity of hot dying stars such as the central stars in planetary nebulae. The mechanism of heating and ionising the gas is photoexcitation and photoionisation, that is, exactly the same process described in Sect. 9.1 but at much lower energies, specifically at energies $h\nu \geq 13.6 \text{ eV} = E_1$, the ionisation potential of hydrogen.

In the process of photoionisation, photons in the high energy tail of the Planck distribution with energy $h\nu \geq E_1$ are responsible for the ionisation of the gas. The reason for this is the large cross-section of hydrogen atoms for photoionisation by photons with energies $h\nu \geq E_1$. The resulting temperature of the ionised gas about the hot star is very much less than E_1/k partly because, in a simple approximation, it can be shown that $T_{\text{gas}} \approx T_*$ where T_* is the effective temperature of the stellar atmosphere and partly because of cooling by line emission. Thus, typical temperatures in the gas are about 5000–20 000 K, compared with $T_{\text{gas}} = 10^5$ K which would be required for collisional ionisation of neutral hydrogen, that is, $kT_{\text{gas}} \approx E_1$. The book *The Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* by Osterbrock and Ferland can be strongly recommended, both for its clear exposition of the basic atomic physics involved and of how emission lines can be used as diagnostic tools to measure physical conditions in gaseous nebulae such as regions of ionised hydrogen, planetary nebulae, the shells of supernova remnants and the environments of active galactic nuclei (Osterbrock and Ferland, 2005).

Hydrogen recombination lines are amongst the strongest lines observed in the spectra of gaseous nebulae and are responsible for a large part of their cooling. The ratio of intensities of the Balmer lines is known as the *Balmer decrement* and is relatively insensitive to physical conditions, unless the particle densities are very high, $N_e \geq 10^{14} \text{ m}^{-3}$ when the effects of self-absorption and collisional excitation of the Balmer series become important. The intensities of the hydrogen recombination lines do not provide direct information about the particle densities in the line-emitting regions. For example, the $\lambda 486.1 \text{ nm H}\beta$ line of the Balmer series in which the principal quantum number n changes from 4 to 2 is one of the strongest lines in the spectra of regions of ionised hydrogen. The line intensity is

$$\begin{aligned} L(H\beta) &= N_e N_p \alpha h v_{H\beta} V \epsilon \\ &= 2.28 \times 10^{-26} N_e^2 T_e^{-3/2} b_4 \epsilon V \exp(9800/T_e) \text{ W} \end{aligned} \quad (12.11)$$

where α is the recombination coefficient appropriate to the $H\beta$ transition, V is the volume of the source, b_4 is a factor representing the departure of the population of the upper level of the $H\beta$ transition from thermal equilibrium, T_e is the electron temperature of the gas and ϵ is the *filling factor* which is the fraction of the volume of the source which is filled with gas; if the volume is uniformly filled with gas, $\epsilon = 1$. The intensity of the $H\beta$ line thus measures $\int N_e^2 T_e^{-3/2} dl$ through the source region. Values for b_4 are given in tables by Pengelly (1964). For temperatures $T \approx (1-2) \times 10^4 \text{ K}$, b_4 lies in the range 0.1 – 0.4 depending upon the physical conditions. There is no direct way of disentangling N_e from this study without further physical considerations.

Hydrogen recombination lines have been observed from the diffuse warm component of the interstellar gas. According to Reynolds, diffuse $H\alpha$ emission is present over the entire sky and, at Galactic latitudes $|b| > 10^\circ$, follows the cosec $|b|$ law expected of the emission of a thin disc (Reynolds, 1990). The intensity of this emission provides a measure of $\int N_e^2 dl$ whereas the dispersion measures of pulsars determine $\int N_e dl$ (Sect. 12.3.3) so that the clumpiness of the ionised gas can be found. Further information on the temperature and density of the diffuse ionised gas is obtained from observations of the forbidden lines

of [N II], [S II] and [O III]. The properties of the diffuse warm gas responsible for these lines and the diffuse H α emission are similar to those labelled ‘Intercloud medium’ in Table 12.3 below.

Another application of hydrogen recombination lines is in the study of very high order transitions $n \rightarrow n - 1$ with $n \geq 100$, which result in photons with energies in the radio waveband. These have been detected from many diffuse regions of ionised hydrogen and provide a further probe of physical conditions. Because the radio emission is not attenuated by interstellar dust, it provides a valuable tool for studying distant regions of ionised hydrogen, the presence of which are only known from their radio bremsstrahlung. Since the linewidths are narrow, the radio recombination line velocities can be used as spiral arm tracers in the more distant parts of the Galaxy. Remarkably, similar recombination lines have been observed at low radio frequencies, $v \sim 15\text{--}30$ MHz, associated with the recombination of carbon atoms, but with very large principal quantum numbers, for example, $n = 631$ at 26.12 MHz and $n = 768$ at 14.7 MHz.

The other strong emission lines observed in the optical spectra of gaseous nebulae are the *forbidden lines*. Because the gas in gaseous nebulae is relatively cool, $T_e \approx 5000\text{--}20\,000$ K, collisions can only excite those energy levels within a few eV of the ground state. For the common elements such as C, N, O, Ne, S, the only accessible levels are metastable levels which have excitation potentials less than about 5 eV. In these elements the low-lying levels are associated with two, three or four electrons in incomplete p shells. An example of such a term diagram, that of doubly ionised oxygen O⁺⁺ or O III, is shown in Fig. 12.1 in which there are two 2p² states within 5 eV of the ground state (Moore and Merrill, 1968). The only way in which electrons in these levels can return to the ground state by a radiative transition is through the transitions shown on the Grotrian diagram which violate the rules for electric dipole transitions, that is, they are *forbidden* transitions. The levels above the ground state can become highly populated by electron collisions in a low density plasma because there are no selection rules for the collisional excitation of an atom or ion. This large population of ions in these metastable states is more than enough to compensate for the small spontaneous transition probability for magnetic dipole or electric quadrupole transitions between these levels and accounts for the high intensities of the forbidden emission lines.

Another type of transition which violates the selection rules for electric dipole transitions is the class of *semi-forbidden transitions* which are less highly forbidden than the above examples. These transitions result in *intercombination lines* in which only a single selection rule is violated. A well-known example is the semi-forbidden transition associated with doubly ionised carbon, which is denoted C III] $\lambda 190.9$ nm.

Forbidden lines provide diagnostic tools for determining densities and temperatures in emission line regions. The strengths of the lines are determined by the competing processes by which de-excitation takes place following excitation by electron collisions. If the density is low, *radiative de-excitation* results in the emission of a photon and the intensity of the line is proportional to the rate of collisional excitation. If, however, the density is high, de-excitation by electron collisions is more important and leads to the suppression of the intensity of the emission line. There is thus a *critical density* above which forbidden line

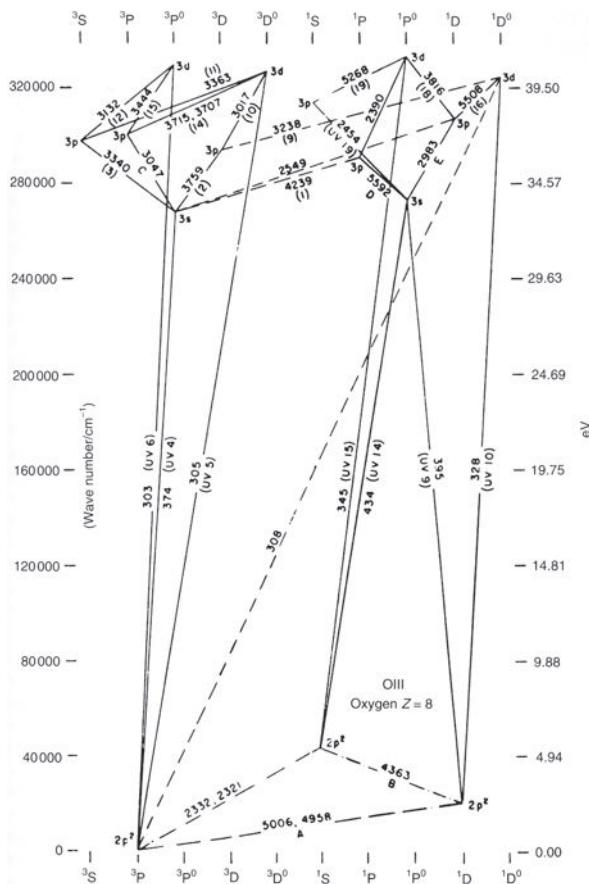


Fig. 12.1 The term diagram for doubly ionised oxygen O^{III}. The forbidden transitions observed in the optical waveband originate from low-lying levels associated with the ¹S and ¹D configurations of the 2p² shell electrons (Moore and Merrill, 1968).

emission is rapidly quenched – critical densities for a number of the common ions are listed in Table 12.2 (Osterbrock and Ferland, 2005).

Critical densities can also be evaluated for the semi-forbidden lines and, because of their greater spontaneous transition probabilities, much greater electron densities can be studied. For example, for C III], the critical density is $N_e \approx 10^{16} \text{ m}^{-3}$. In order to make estimates of parameters such as the electron density and electron temperature, it is essential to measure the *ratios* of different forbidden lines originating from the same region. More detailed studies involve using line ratios among the low level forbidden lines of a particular ion which are sensitive to both density and temperature. Osterbrock and Ferland provide an elegant description of the techniques by which this can be achieved (Osterbrock and Ferland, 2005). Notice that, in contrast to other techniques, this method enables particle densities to be determined directly in the regions under study.

Table 12.2 The critical densities for collisional de-excitation of some common ions. All values are calculated for $T = 10\,000$ K. (Osterbrock and Ferland, 2005).

Ion	Level	Critical density (N_e/m^{-3})	Ion	Level	Critical density (N_e/m^{-3})
C II	$^2P_{3/2}$	8.5×10^7	O III	1D_2	7.0×10^{11}
			O III	3P_2	3.8×10^9
C III	3P_2	5.4×10^{11}	O III	3P_1	1.7×10^9
N II	1D_2	8.6×10^{10}	Ne II	$^2P_{1/2}$	6.6×10^{11}
N II	3P_2	3.1×10^8	Ne III	1D_2	7.9×10^{12}
N II	3P_1	1.8×10^8	Ne III	3P_0	2.0×10^{10}
N III	$^2P_{3/2}$	3.2×10^9	Ne III	3P_1	1.8×10^{11}
N IV	3P_2	1.4×10^{12}	Ne V	1D_2	1.6×10^{13}
O II	$^2D_{3/2}$	1.6×10^{10}	Ne V	3P_2	3.8×10^{11}
O II	$^2D_{5/2}$	3.1×10^9	Ne V	3P_1	1.8×10^{11}

12.3.3 The dispersion measure of pulsars

Estimates of the column density of free electrons in the Galaxy, $\int N_e dl$, may be obtained from the delay times in the arrival of radio signals as a function of frequency. In a plasma, a wavepacket propagates at the group velocity v_{gr} which is a function of frequency. At frequencies well above the gyrofrequency of the electrons in the plasma, $\nu \gg v_g$, the group velocity depends only upon the plasma frequency ν_p and is given by $v_{gr} = c [1 - (\nu_p/\nu)^2]^{1/2}$, where ν_p is the plasma frequency,

$$\nu_p = \left(\frac{e^2 N_e}{4\pi^2 \epsilon_0 m_e} \right)^{1/2} = 8.98 N_e^{1/2} \text{ Hz}, \quad (12.12)$$

where N_e is measured in electrons m^{-3} . At radio wavelengths, $\nu \approx 10^2 - 10^3$ MHz, $\nu_p/\nu \ll 1$ and hence

$$v_{gr} = c \left[1 - \frac{1}{2} \left(\frac{\nu_p}{\nu} \right)^2 \right]. \quad (12.13)$$

If a pulse of radio waves is emitted at time $t = 0$, the arrival time of the signals T_a is therefore a function of frequency, that is,

$$T_a = \int_0^l \frac{dl}{v_{gr}} = \int_0^l \frac{dl}{c} \left[1 + \frac{1}{2} \left(\frac{\nu_p}{\nu} \right)^2 \right] = \frac{l}{c} + \frac{e^2}{8\pi^2 \epsilon_0 m_e c} \frac{1}{\nu^2} \int_0^l N_e dl. \quad (12.14)$$

Thus, by measuring the arrival time of the pulse T_a as a function of frequency ν , the electron column density along the line of sight to the source $\int N_e dl$ can be found. Inserting the

numerical values of the constants into (12.14), we find

$$T_a = 4.15 \times 10^9 \frac{1}{\nu^2} \int N_e dl \text{ seconds ,} \quad (12.15)$$

where the electron density is measured in electrons m⁻³, the distance *l* in parsecs and *v* in Hz.

For the procedure to be practicable, sources are required which emit sharp pulses of radiation over a wide range of frequencies. Pulsars, which are discussed in Sect. 13.3, are ideal for this purpose and estimates of $\int N_e dl$, which is known as the *dispersion measure*, are readily made for all of them. These data provide estimates of $\int N_e dl$ in roughly 2000 directions through the interstellar gas. If it is assumed that the electron density is uniform in the plane of the Galaxy, the dispersion measure provides an estimate of the distance of the pulsar. Improved distances can be found by adopting a more detailed picture for the distribution of the ionised gas in the Galaxy (Taylor and Cordes, 1993; Cordes and Lazio, 2002).

12.3.4 Faraday rotation of linearly polarised radio signals

The partially ionised interstellar gas is permeated by the Galactic magnetic field and hence constitutes a magnetised plasma, or *magnetoactive medium*. Under typical interstellar conditions, both the plasma frequency $\nu_p = 8.98 N_e^{1/2}$ Hz and the gyrofrequency $\nu_g = 2.8 \times 10^{10} B$ Hz, where *B* is measured in tesla, are much less than typical radio frequencies, $10^7 \geq \nu \geq 10^{11}$ Hz. Under these conditions, the position angle of the electric vector of linearly polarised radio emission is rotated on propagating along the magnetic field direction. This phenomenon is known as *Faraday rotation*.

Faraday rotation results from the fact that the modes of propagation of radio waves in a magnetised plasma are elliptically polarised in opposite senses, that is, they can be right- or left-handed elliptically polarised waves. These are the natural modes of propagation of the waves because, under the influence of the perturbing electric field of the waves, the electrons are constrained to move in spiral paths about the magnetic field direction (Sect. 7.1). Therefore, when a linearly polarised signal is incident upon a magnetoactive medium, it can be resolved into equal components of oppositely handed elliptically polarised radiation. In the limit $\nu_g/\nu \ll 1$, the refractive indices *n* of the two modes are different:

$$n^2 = 1 - \frac{(\nu_p/\nu)^2}{1 \pm (\nu_g/\nu) \cos \theta} , \quad (12.16)$$

where θ is the angle between the direction of wave propagation and the magnetic field direction. The phase velocities of the two modes are different and so one sense of elliptical polarisation runs ahead of the other. When the elliptically polarised components are added together at depth *l* through the region, the result is a linearly polarised wave rotated with respect to the initial direction of polarisation. From the dispersion relation (12.16), the difference in refractive indices under the conditions $\nu_p/\nu \ll 1$, $\nu_g/\nu \ll 1$, is

$$\Delta n = \frac{\nu_p^2 \nu_g}{\nu^3} \cos \theta . \quad (12.17)$$

On propagating a distance dl through the region, the phase difference between the two modes is

$$\Delta\phi = \frac{2\pi v \Delta n}{c} dl . \quad (12.18)$$

On summing the two elliptically polarised waves, the direction of the linearly polarised electric vector is rotated through an angle $\Delta\theta = \Delta\phi/2$, that is,

$$\Delta\theta = \frac{\pi v_p^2 v_g}{cv^2} \cos\theta dl . \quad (12.19)$$

For $v_g \cos\theta$ we write $2.8 \times 10^{10} B_{\parallel}$ Hz where B_{\parallel} is the component of B parallel to the line of sight in tesla. Therefore,

$$\theta = \frac{\pi}{cv^2} \int_0^l v_p^2 v_g \cos\theta dl , \quad (12.20)$$

or, rewriting the formula in more convenient units,

$$\theta = 8.12 \times 10^3 \lambda^2 \int_0^l N_e B_{\parallel} dl , \quad (12.21)$$

where θ is measured in radians, λ in metres, N_e in particles m⁻³, B_{\parallel} in tesla and l in parsecs. The quantity θ/λ^2 is known as the *rotation measure* and is measured in radians m⁻² – it provides information about the integral of $N_e B_{\parallel}$ along the line of sight. In addition, the sign of the rotation gives information about the weighted mean direction of the magnetic field along the line of sight. If θ/λ^2 is negative, the magnetic field is directed away from the observer; if θ/λ^2 is positive, the field is directed towards the observer.

Many Galactic and extragalactic radio sources emit linearly polarised radio emission and therefore, by measuring the variation of the position angle of the electric vector with frequency, estimates of $\int N_e B_{\parallel} dl$ may be obtained for many different lines of sight through the Galaxy. An estimate of the strength of the Galactic magnetic field can be found by combining observations of the Faraday rotation of the linearly polarised emission of pulsars with their dispersion measures. The former gives an estimate of $\int N_e B_{\parallel} dl$ and the latter $\int N_e dl$. We therefore obtain a weighted estimate of the strength of the magnetic field along the line of sight,

$$\langle B_{\parallel} \rangle \propto \frac{\text{rotation measure}}{\text{dispersion measure}} \propto \frac{\int N_e B_{\parallel} dl}{\int N_e dl} . \quad (12.22)$$

In addition to rotation of the plane of polarisation, the radio emission is *depolarised* with increasing wavelength. If the radio emission originates from a region of size l in which the magnetic flux density B and the plasma density N_e are uniform, the radiation is fully polarised at high enough frequencies because internal Faraday rotation within the region is proportional to λ^2 and so tends to zero as the wavelength tends to zero. At long wavelengths, however, because there is substantial rotation of the plane of polarisation through the source region, the polarisation vectors originating from different depths within the region add up at different angles as the radiation leaves the source. When the plane of polarisation of the radiation is rotated by $\theta = \pi$ radians through the source region,

the net degree of polarisation decreases. In this model, the frequency at which significant depolarisation is observed provides information about the integral of $N_e B_{\parallel} l$ through the source region. Whereas the rotation of the plane of polarisation provides information about $\int N_e B_{\parallel} dl$ from the source to the Earth, the depolarisation provides information about the source regions themselves. This process is often referred to as *Faraday depolarisation*.

The above analysis only applies to the simplest magnetic field configurations. If there are irregularities or fine structure in the magnetic field and plasma distribution, the contributions of each region to the total polarisation have to be summed. In addition, if the magnetic field distribution is stretched in some direction, we may obtain polarised emission, but the depolarisation would depend upon how the electric field vectors are rotated on passing through different regions within the source.

12.4 Interstellar dust

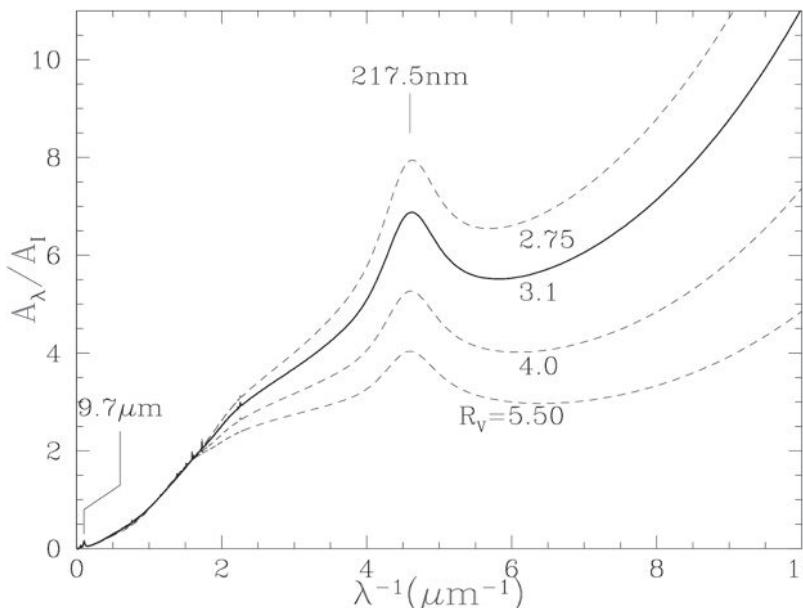
A vital component of the interstellar medium is dust which causes the patchy obscuration seen in the optical image of the Galaxy (Fig. 1.2). Interstellar dust is inferred to contain a large fraction of the heavy elements present in the interstellar medium because the gaseous phase is significantly under-abundant in these elements. Dust is present in most environments in the Universe, unless it is heated to temperatures above the material's sublimation temperature, which is about 10^3 K. Dust shells are observed to form about dying stars and supernovae when the temperature of the ejected material falls below roughly this temperature.

Throughout the optical and infrared wavebands, the effect of dust extinction can be described by an extinction law $S \propto e^{-\tau}$, where the optical depth τ of the medium depends upon wavelength λ roughly as $\tau \propto \lambda^{-x}$ – in the optical waveband, $x \approx 1$ and in the infrared waveband $1.6 \lesssim x \lesssim 1.8$. This attenuation is often written in terms of apparent magnitudes as $m(\text{obs}) = m + A_{\lambda}$, where m is the apparent magnitude in the absence of extinction and A_{λ} is referred to as the *total extinction* at wavelength λ or in one of the standard wavebands. The term *extinction* is used to include the attenuation of the radiation due to both absorption and scattering. The extinction amounts typically to about $0.7\text{--}1.0$ mag kpc $^{-1}$ for the local interstellar medium in the V waveband.

Examples of the extinction curves as a function of inverse wavelength along different lines of sight through the interstellar medium in our Galaxy are shown in Fig. 12.2. The slope of the extinction curve in the optical waveband can be characterised by the quantity

$$R_V = \frac{A_V}{A_B - A_V} = \frac{A_V}{E(B - V)}, \quad (12.23)$$

where $E(B - V)$ is known as the *reddening* or *selective absorption* and R_V is the ratio of total to selective absorption in the V waveband. For many sight-lines through the Galaxy, $R_V = 3.1$, corresponding to $x \approx 1$, but there are variations about this value as indicated by the plots in Fig. 12.2. For precise work, the extinction coefficient has to be determined along each line of sight. The strong dependence of the extinction coefficient upon wavelength

**Fig. 12.2**

The extinction A_λ relative to the extinction at $\lambda = 900 \text{ nm}$ ($\lambda^{-1} = 1.1 \mu\text{m}^{-1}$), as a function of inverse wavelength λ^{-1} , for Milky Way regions characterized by the quantity of $R_V = A_V/E(B-V)$, where A_V is the extinction at $B = 440.0 \text{ nm}$, A_V is that at $V = 550 \text{ nm}$, and the ‘reddening’ $E(B-V) = A_B - A_V$. The extinction increases rapidly in the vacuum ultraviolet ($\lambda^{-1} > 5 \mu\text{m}$) for regions with $R_V \lesssim 4$. The normalization is approximately $A_V/N_{\text{H}} \approx 2.6 \times 10^{22} \text{ cm}^2$ per hydrogen nucleon. The silicate absorption feature at $9.7 \mu\text{m}$ and the diffuse interstellar bands are just visible (Draine, 2004).

explains why obscuration can affect optical and ultraviolet observations very severely and yet have a modest effect in the infrared waveband. For example, in the direction of the Galactic Centre, the attenuation in the V waveband, $\lambda = 0.55 \mu\text{m}$, amounts to about 30 magnitudes, a factor of 10^6 in flux density. At $2 \mu\text{m}$, the attenuation would be only 8 magnitudes, a factor of 1600, and at $5 \mu\text{m}$ only 3 magnitudes or a factor of 15.

Dust grains absorb and scatter electromagnetic waves efficiently at wavelengths less than or equal to their physical sizes but are transparent at longer wavelengths. This can be demonstrated by writing the cross-section for scattering and absorption in terms of the physical cross-section of the grain πa^2 times an *extinction efficiency factor* Q so that the cross-section is $\sigma = Q\pi a^2$, where a is the radius of the spherical grain. Exact results for all wavelengths can be calculated for spherical particles with isotropic dielectric constants using the *Mie theory of scattering and absorption*. This approach involves finding exact solutions of Maxwell’s equations for plane-parallel incident light. Figure 12.3 shows the result of computations of Mie scattering and absorption for spherical silicate grains with a complex dielectric constant $\epsilon = 3 + 0.1i$, where the imaginary term represents absorption by the grain material. The results are shown as a function of the size parameter $x = 2\pi a/\lambda$. At large values of x , corresponding to short wavelengths, the total cross-section for scattering and absorption tends to $Q = 2$, that is, $\sigma = 2\pi a^2$. At values of

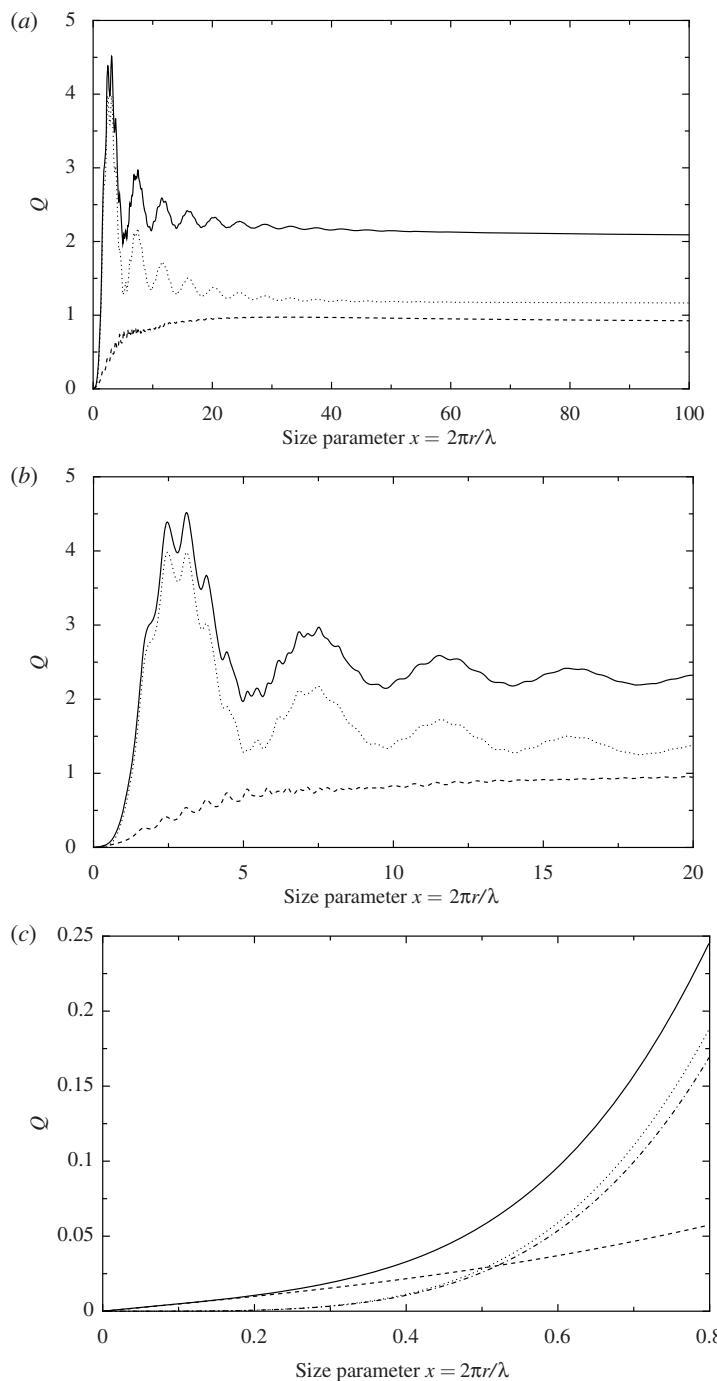


Fig. 12.3

The extinction efficiency Q as a function of size parameter $x = 2\pi r/\lambda$ for silicate spheres with isotropic dielectric constant $\epsilon = 3 + 0.1i$. (a) The numerical solution for a wide range of values of x including both absorption (dashed line) and scattering (dotted line); $Q \rightarrow 2$ as $x \rightarrow \infty$. (b) The values of Q for $x \leq 20$ showing the detailed structure of the extinction efficiency for both scattering and absorption. (c) Details of the function Q for $x \leq 0.8$ in the same notation as (b). The scattering component of the extinction follows closely the Rayleigh scattering law $Q \propto \lambda^{-4}$ which is shown by the dot-dash line but reduced by 10% since it lies almost exactly along the dotted curve. (Courtesy of Bojan Nikolic.)

$x \ll 1$, corresponding to wavelengths much greater than the size of the grain, the cross-section varies as λ^{-4} , as expected for Rayleigh scattering. The physical reason for this is that the grain as a whole feels the same electric field of the wave and so experiences coherent polarisation, resulting in Rayleigh scattering.

Similar calculations can be carried out for other materials such as graphite, but the extinction efficiency factors are sensitive to anisotropies associated with the large sheets of hexagonal benzene rings within the graphite grains. Despite these complications, Fig. 12.3 illustrates the general result that the extinction cross-section is about $2\pi a^2$ for wavelengths $\lambda \ll a$ and changes as λ^{-4} at wavelengths $\lambda \gg a$. There must therefore be a wide range of grain sizes present in the interstellar medium to account for the fact that the extinction coefficient of the interstellar gas extends rather smoothly from ultraviolet through optical to infrared wavelengths (Fig. 12.2). Superimposed upon this continuum absorption curve, there are several prominent features. The strongest is the broad absorption feature observed at about 217.5 nm which is present in the Galactic extinction curve. This feature corresponds rather closely with the excitation energy of the $\pi \rightarrow \pi^*$ transition associated with π -orbitals of the hexagonal lattice and it is commonly assumed that this is evidence for graphite in interstellar grains. A natural extension of this model is that the feature might be associated with similar excitations associated with sheets of large PAH molecules described below. There are also diffuse interstellar bands in the optical waveband but these have remained unidentified despite an enormous amount of work by many authors.

Dust absorption features have also been discovered in the infrared waveband, for example, the 3.1 μm water ice feature and the prominent silicate absorption and emission features at 9.7 and 18 μm . The nature of the grains is therefore likely to be somewhat complex. In a popular picture, the grains contain graphite or silicon cores surrounded by water ice mantles. A key role of dust grains is in the formation of molecules. Atoms and molecules are adsorbed onto grain surfaces where they can migrate, combine with other species and then return to the interstellar medium. Thus, the grains act as a ‘catalyst’ for the formation of organic molecules. This is almost certainly the origin of many of the species listed in Table 12.1.

Much of the study of interstellar dust grains focussed upon the properties of particles roughly 0.1–1 μm in size but there is also evidence for a population of very much smaller grains from studies of the infrared continuum spectra of reflection nebulae (Sellgren, 1984). The emission is associated with transient heating of very small dust grains. For grains with dimension 1 μm , the energy of the absorbed photons is thermalised and reradiated at the temperature to which the grains are heated. For grains only about 1 nm in size, this is no longer the case. An incident ultraviolet photon can raise the temperature of the grain to about 1000 K and then the grain cools rapidly, resulting in a quite different non-equilibrium continuum spectrum. The necessary number of very small dust grains can be explained as an extrapolation of the grain size distribution from larger sizes. These tiny grains can be thought of as large molecules.

This concept was taken further by Leger and Puget who sought to account for the strong unidentified emission features observed in the infrared region of the spectrum (Leger and Puget, 1984). Prominent lines are observed at wavelengths λ 3.28, 6.2, 7.7, 8.6 and 11.3 μm in the spectra of a wide variety of Galactic and extragalactic sources (Fig. 12.4). These lines

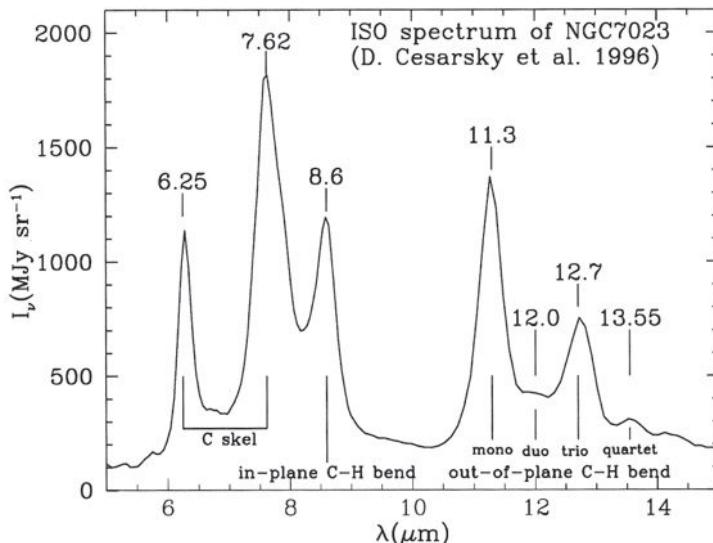


Fig. 12.4 The PAH emission features in the 5–15 μm spectrum of the reflection nebula NGC 7023 obtained with the ISO Observatory by Cesarsky and his colleagues (Draine, 2003).

are associated with various bending and stretching modes of small aromatic molecules known as *polycyclic aromatic hydrocarbons*, or *PAHs*. The molecules typically consist of about 50 carbon atoms in the form of planes of hexagonal benzene rings. For the PAH coronene, for example, Leger and Puget computed that, at a temperature of 600 K, spectral features should be observed at λ 3.3, 6.2, 7.6, 8.8 and 11.9 μm . These features were identified as follows: the feature at 3.3 μm with the C–H stretching mode, those at 6.2 and 7.7 μm with the C–C stretching modes, that at 8.6 μm with the in-plane bending mode and that at 11.3 μm with the C–H out-of-plane bending mode. In the last case, other features are expected depending upon the number of nearby hydrogen atoms. The excitation of these modes is associated with the absorption of a single UV photon which transiently raises the temperature of the molecule to about 1000 K.

The net result of these studies is that interstellar dust must be composed of a number of different components. An excellent discussion of the necessary range of different types of dust particles necessary to account for the observations is given by Draine (2003).

Interstellar dust grains perform a number of different functions. First of all, dust absorbs ultraviolet and optical radiation and therefore, within dust clouds, molecules are protected from the interstellar flux of dissociating radiation. The second process is the *reradiation* of the radiation absorbed by the dust grains. This is an efficient energy loss mechanism for stars which are in the process of formation or have just formed. Stars form in the densest regions of giant molecular clouds and the ultraviolet radiation emitted by them is absorbed by the dust grains. The grains are heated to a temperature which is determined by the balance between the energy absorbed from the radiation field and their rate of radiation. They radiate more or less like little black-bodies, the Planck distribution being modified by the emissivity function $\kappa(\nu)$ of the material of the grains. Thus, the emissivity of the

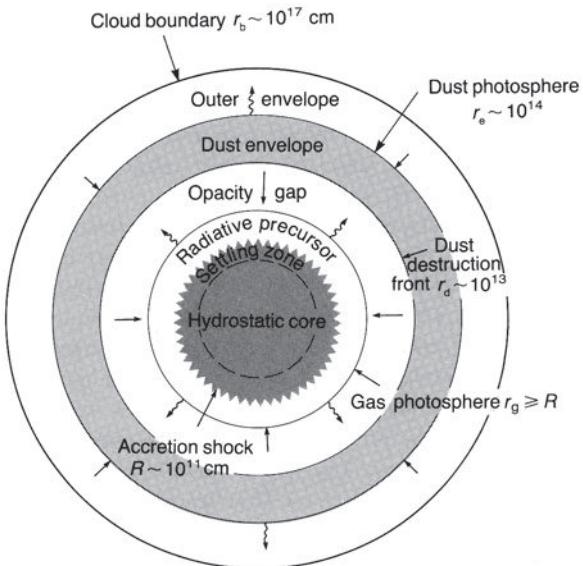


Fig. 12.5 A diagram illustrating the structure of an accreting protostar according to the analysis of Shu and his colleagues. The various regions are described in the text (Stahler *et al.*, 1980).

grain can be written $\epsilon(\nu) = \kappa(\nu)B(\nu)$ where $B(\nu)$ is the Planck distribution. Hildebrand has shown that, to a good approximation, $\kappa(\nu) \propto \nu$ at wavelengths $\lambda < 100 \mu\text{m}$ and $\kappa(\nu) \propto \nu^2$ at much longer wavelengths, $\lambda > 1 \text{ mm}$ (Hildebrand, 1983). The grains radiate away the absorbed energy very rapidly at roughly the temperature to which they are heated, which is typically about 30–100 K for the far-infrared sources found in dense molecular clouds. At wavelengths $\lambda \sim 30\text{--}100 \mu\text{m}$ the dust is transparent and so the energy of the star can be radiated away very efficiently. This picture explains why intense far-infrared emission is the signature of sites of star formation. In addition, many galaxies, particularly those in which there is active star formation such as late-type spiral and irregular galaxies as well as the colliding galaxies, show extreme far-infrared luminosities with the characteristic emission spectra of heated dust.

An important application of these ideas is in understanding the early evolution of protostars and stars which have just evolved onto the main sequence. Figure 12.5 shows the expected structure of a protostar. There is a central *hydrostatic core* and the outer regions are associated with an accretion flow as the star builds up its mass. In the outer envelope, the matter and dust are optically thin and can radiate away their thermal energy very efficiently. The infall in this region is therefore close to isothermal. Eventually the matter and dust densities increase to values such that the dust becomes optically thick, the radius at which the optical depth is unity being referred to as the *dust photosphere*. At smaller radii, there is a dust envelope within which the temperature increases with decreasing radius until it becomes hot enough for the dust to evaporate, at $T \approx 2300 \text{ K}$ for graphite grains. Within this radius, the radiative transfer is determined by the properties of the gas rather than the dust. The gas is accreted onto the hydrostatic core and, since the latter acts as a ‘solid body’,

an accretion shock is formed which has the effect of dissipating the kinetic energy of the infalling gas and radiating away its binding energy.

This picture indicates why protostars are expected to be intense infrared sources. The binding energy of the matter accreted onto the protostar is transported by radiation from the accretion shock and this energy is trapped and degraded in the dust envelope. The energy is eventually radiated away at the temperature of the dust photosphere. Models such as those of Adams and Shu show a broad maximum corresponding to the superposition of the emission from grains at different temperatures in the dust photosphere, the typical temperatures being about 100 K (Adams and Shu, 1985). The predicted spectra are similar to those observed in a number of sources inferred to be protostellar objects (see Fig. 12.6).

A ‘standard’ scenario of star formation has been described by Shu and his colleagues which synthesises these ideas into a general picture illustrated schematically in Fig. 12.7 (Shu *et al.*, 1987). The process begins with the collapse of cool density enhancements within giant molecular clouds and, in the early stages, the energy source in protostars and pre-main-sequence stars is the accretion of matter onto the core of the protostar rather than nuclear energy generation. Because the infalling matter is bound to have some angular momentum, a rotating disc forms perpendicular to the rotation axis. The removal of the gravitational binding energy of the accreted matter is effected by the reradiation of heated dust at far-infrared wavelengths at which the protostellar cloud is transparent. At some stage, a stellar wind breaks out along the rotation axis of the system, creating a bipolar outflow. Finally, when the accretion phase is completed, all that is left is the newly formed star with a circumstellar disc. One of the more striking discoveries of the IRAS mission was that objects with the spectral characteristics corresponding to each of these stages have been observed (Fig. 12.6). Objects at the earliest stages in their evolution are purely far-infrared sources. At later stages, the emission from the star and a protoplanetary, or accretion, disc can be observed.

12.5 An overall picture of the interstellar gas

12.5.1 Large scale dynamics

Most of the gas in the Galaxy is confined to the Galactic plane and moves in circular orbits about the Galactic Centre, the inward force of gravitational attraction being balanced by centrifugal forces. The gravitational potential in which the gas moves is defined by the mass distribution of the stars and of the Galactic dark matter. The kinematics of the interstellar neutral hydrogen and molecules therefore act as probes of the gravitational potential field and so provide information about the distribution of mass in the Galaxy. The disc of the Galaxy is in a state of *differential rotation*, the mean rotational velocity of the material as a function of distance from the Galactic Centre, its *rotation curve*, being shown in Fig. 12.8 (Fich and Tremaine, 1991). The distance from the Galactic Centre to the local standard of

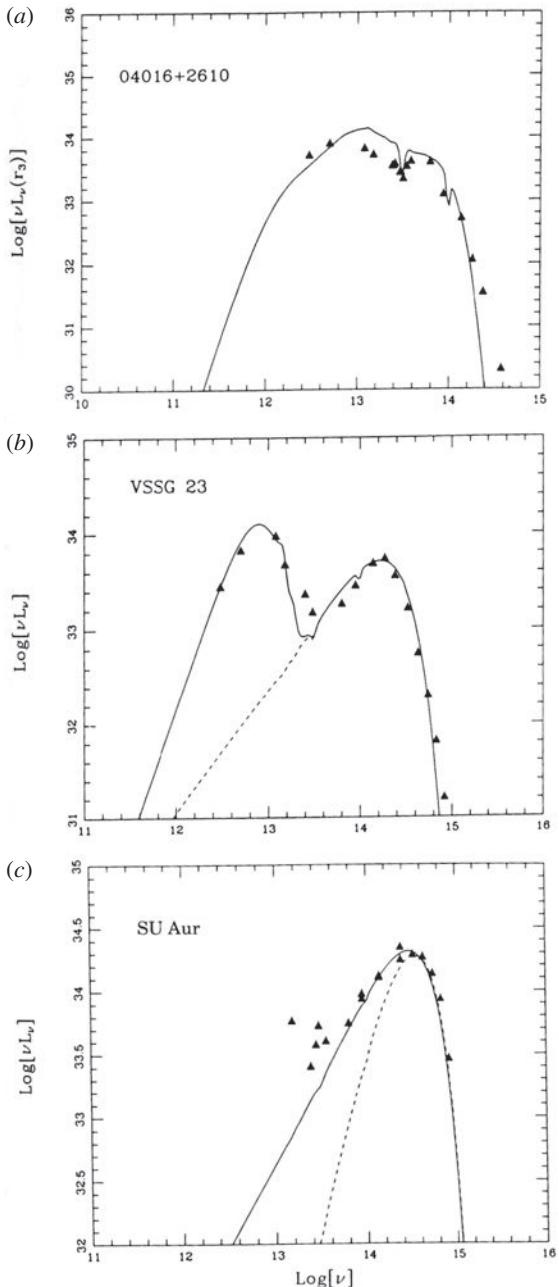
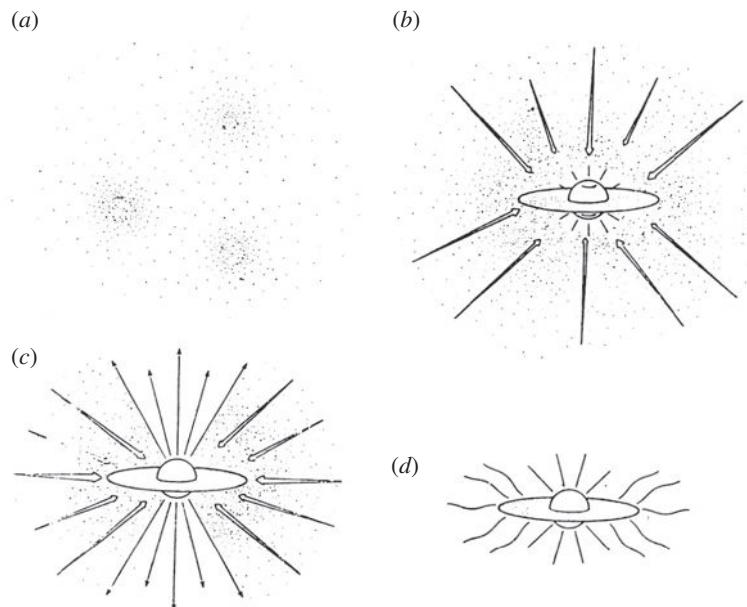
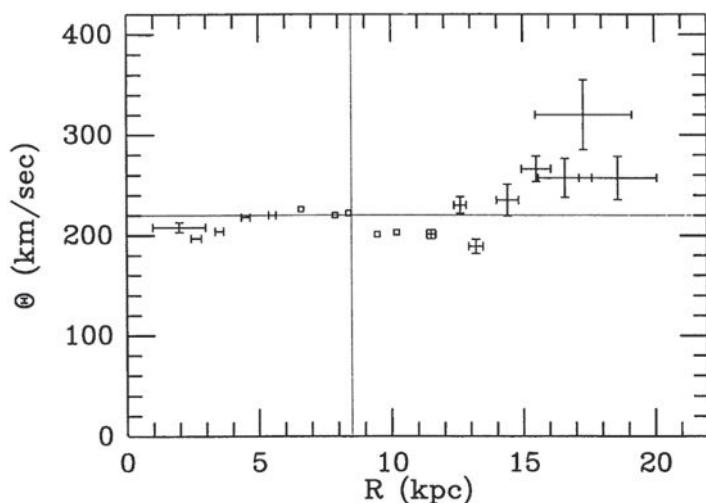


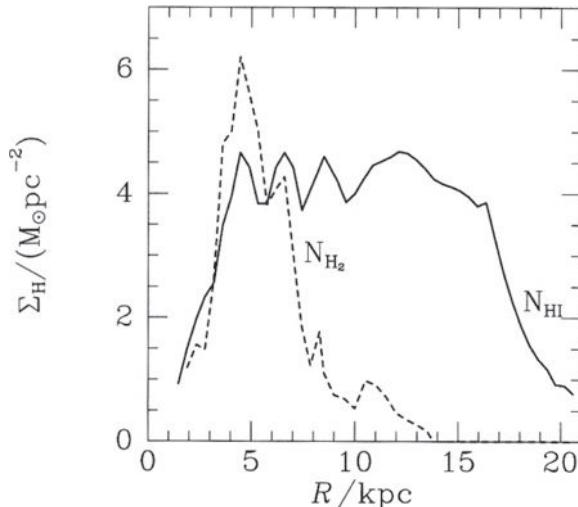
Fig. 12.6 Comparison of the theoretical and observed spectra of sources in the Taurus and Ophiuchus molecular clouds. The ordinate is ν/L_ν representing the energy emitted at each frequency. All the sources have mass of order $1 M_\odot$. (a) The source 14016+2610 is inferred to be a protostar during its main infall phase, that is, the star and disc are embedded in an infalling dust envelope. (b) In VSSG 23 it is inferred that an intense wind has broken out along the rotation axis revealing the newly born star surrounded by a nebular disc. (c) The source SU Aur is a T Tauri star with a small infrared excess. The disc has disappeared leaving an isolated pre-main sequence star (Adams *et al.*, 1987).

**Fig. 12.7**

A schematic representation of a plausible scenario for the formation of stars. (a) Density inhomogeneities collapse under their own self-gravity. (b) The main accretion phase in which an accreting core has formed and infall of matter onto that core takes place. The binding energy of the accreted matter is removed by radiation which is absorbed by dust and reradiated in the far-infrared waveband. (c) Jets of material burst out of the accreting star along its rotation axis producing the characteristic 'bipolar outflows' observed in most young stars. (d) The accretion of material ceases and the system is left with a young, hydrogen-burning star and a rotating dust disc (Shu *et al.*, 1987).

**Fig. 12.8**

An average rotation curve for our Galaxy adopting the 1985 IAU recommended values for the Sun-Centre distance of 8.5 kpc and a mean local rotation velocity about the Galactic Centre of 220 km s^{-1} (Fich and Tremaine, 1991).

**Fig. 12.9**

The radial distribution of atomic and molecular hydrogen as deduced from radio surveys of the Galaxy in the 21-cm line of atomic hydrogen and from millimetre surveys of the molecular emission lines of carbon monoxide, CO (Binney and Merrifield, 1998).

rest at the Sun is taken to be 8.5 kpc and the mean rotation velocity about the Galactic Centre in the solar vicinity is 220 km s⁻¹. The rotation velocities are remarkably constant over radial distances from 3 to 15 kpc, with some evidence for an increase in the rotation velocity beyond 15 kpc. These results are inconsistent with solid body rotation, for which $v_{\text{rot}} \propto r$, or Keplerian orbits for which $v_{\text{rot}} \propto r^{-1/2}$. As is discussed in Sect. 3.5.2, these data provide evidence for dark matter in the outer regions of our Galaxy. Similar rotation curves are found in other giant spiral galaxies (Fig. 3.11).

The distribution of neutral hydrogen in the Galaxy was determined as long ago as the 1950s and, more recently, carbon monoxide surveys have defined the distribution of the molecular gas. The neutral and molecular hydrogen are closely confined to the plane of the Galaxy, the typical half-widths being about 120 and 60 pc, respectively. They have, however, very different distributions with distance from the Galactic Centre. The neutral hydrogen extends from about 3 kpc to beyond 15 kpc from the Centre, whereas the molecular component appears to form a thick ring between radii $3 \lesssim r \lesssim 8$ kpc (Fig. 12.9). The evidence of spiral arm tracers such as O and B stars and H II regions suggests that our Galaxy possesses a rather tightly wound spiral structure. Features possibly related to spiral arms have been observed in the local distribution of neutral hydrogen, the giant molecular clouds also having a tendency to be found in spiral arm regions.

Whilst the overall distribution of the gas is determined by the gravitational potential defined by the stars and the dark matter, some mechanism is needed to enhance the average gas density from about 10⁶ m⁻³ to values at least 100–1000 times greater in giant molecular clouds and to result in conditions favourable for the formation of stars in the vicinity of spiral arms. One mechanism for achieving this is through the formation of a density wave in the distribution of stars and dark matter in the Galactic disc. The *density wave theory*

of spiral structure is based upon considerations of the stability of a differentially rotating disc of stars to axial perturbations. It is found that a spiral density wave in the stellar distribution tends to propagate either inwards or outwards from the centre of the disc thus destroying the spiral perturbation. There must therefore be some forcing mechanism which maintains the spiral pattern in the stellar distribution. This might be associated either with gravitational interactions with companion galaxies or possibly with perturbations associated with the ellipsoidal distribution of stars in the central bulge of the Galaxy (see Fig. 1.4 and Sect. 1.4.2).

Assuming that the density wave in the stellar disc is maintained, the behaviour of the cold interstellar gas under its influence can be studied. The sound speed in the neutral and cold gas is very low and so the gas tends to collect at the potential minima of the density wave. It turns out that the velocity the gas acquires in falling into the potential minima is supersonic. Shock waves form along the trailing edge of the stellar density wave and a large increase in gas density behind the shock is expected since the compressed gas can cool effectively. This picture can explain the formation of clouds of neutral and molecular gas in the vicinity of spiral arms and is consistent with the observed location of young objects relative to the underlying spiral density wave defined by the old stellar populations.

Spiral density waves are not the only means of forming giant molecular clouds. Supernova explosions, for example, lead to strong shock waves propagating through the interstellar gas and, in the late stages of expansion, cooling of the compressed gas can lead to the formation of cool dense clouds. It is significant that the largest star-formation rates are found in the most irregular galaxies and not in those with the most beautifully developed spiral structures. Once the first stars are formed in a molecular cloud complex, the most massive explode over a time-scale of 10^{6-7} years and the supersonic motion of the shells of the resulting supernova remnants can trigger the next generation of star formation. This picture can be modelled as a *percolation process* occurring throughout the disc of a galaxy and has had success in explaining the observation of spiral features in galaxies.

12.5.2 Heating mechanisms

Left on its own, the interstellar gas would cool to a low temperature but this is in conflict with the observation of gaseous phases at a wide range of different temperatures. The hottest gas is produced by *supernova explosions*. A shock wave propagates ahead of the supersonically expanding shell of cooling gas and heats the interstellar gas to high temperatures. Cox and Smith first showed that heating by supernova explosions could lead to about 10% of the volume of the interstellar gas being heated to a high temperature (Cox and Smith, 1974). The collisions of old shells of supernova remnants can lead to reheating of the swept up gas as the kinetic energy of expansion is converted into heat. Cox and Smith predicted that the hot component would form tunnels through the interstellar gas as a result of the overlapping of old supernova remnants. At least some part of the soft X-ray emission from the plane of our Galaxy is likely to be associated with this hot gas. Observations by the far-ultraviolet Wide Field Camera of the ROSAT satellite showed that the Solar System is probably located within a large bubble of hot gas of diameter about 500 pc, consistent with this picture. It is also probable that the hot gas inferred to be present in the halo of our

Galaxy from observations of C IV and O VI lines of highly ionised carbon and oxygen has attained a dynamical equilibrium in the gravitational field of the disc and halo. It is natural that the hot gas should expand to form such a hot halo since its scale height is expected to be much greater than that of the stars of the disc.

A second important heating mechanism is the *ultraviolet radiation* of young stars. The youngest of these remain embedded in the gas clouds out of which they formed. The heated gas can be recognised by the strong emission lines of hydrogen and oxygen. The gas temperature is determined by the balance between photoionisation of the neutral gas by ultraviolet radiation and recombination of the ionised component, resulting in a temperature of typically 10^4 K (Osterbrock and Ferland, 2005). Older blue stars, no longer embedded in regions of ionised hydrogen, can ionise and heat the surrounding regions. This form of local heating is observed in the ultraviolet spectra of certain O and B stars. A region of ionised gas has also been observed about a binary X-ray source in which very high excitation species are observed, these being attributed to ionisation and heating by the source.

As discussed in Chap. 17, the flux of cosmic rays observed in the vicinity of the Solar System is probably typical of the flux of high energy particles present throughout the interstellar medium. The ionisation losses of these particles are important sources of heating and ionisation of both the diffuse neutral gas and the gas in giant molecular clouds. The heating rate is poorly known because the greatest heating rates are associated with cosmic rays of relatively low energy for which the energy spectra are poorly known because of the effects of solar modulation. Adopting the spectrum of high energy protons observed at the top of the atmosphere without taking account of the effects of solar modulation, the ionisation rate of the interstellar gas by ionisation losses is found to amount to about $10^{-17} N_{\text{H}}$ electrons s⁻¹, the average energy of each electron being about 35 eV – N_{H} is the number density of neutral hydrogen atoms. This estimate takes account of the production of secondary electrons by the primary electrons released in the process of ionisation. Not all this energy is available for heating the gas since much of it goes into exciting the atoms of the gas. The heating rate could be significantly greater than this figure once the effects of solar modulation are taken into account. On the other hand, it is unlikely to be very much greater than this figure because a local energy density of cosmic rays of about 1 MeV m⁻³ can be accounted for in terms of the observed energies of supernova remnants and their rate of occurrence in the Galaxy. Ionisation losses are almost certainly the origin of the small but significant abundance of free electrons present in molecular clouds which are crucial for interstellar chemistry.

There are other potential sources of heating. For example, the intergalactic flux of ultraviolet ionising radiation, mass loss from all types of star, including stellar winds and bipolar outflows from young stars, infall of matter from intergalactic space, and so on. There are thus good reasons why the interstellar medium should be far from equilibrium.

12.5.3 Cooling mechanisms

Radiation is the principal means by which the thermal energy of the interstellar gas is lost and therefore by observing line and continuum emission at frequencies close to the peak of the black-body spectrum appropriate to that phase of the gas, the cooling processes

can be observed directly. For very hot ionised gas, at temperatures in excess of 10^7 K, the principal cooling mechanism is the *bremsstrahlung* or *free-free emission* of the free electrons in the plasma (Sect. 3.5.2). At lower temperatures, 10^4 – 10^7 K, the emission is due to *bound-bound* and *bound-free* transitions of hydrogen, helium and heavy elements. This temperature regime is more difficult to study observationally because most of the radiation is emitted in the unobservable ultraviolet region of the spectrum. At least part of the soft X-ray radiation detected in the plane of the Galaxy is associated with the radiation of gas at a temperature of about 10^6 K and the spectrum can be attributed to the bound-free emission of different elements which, when summed, results in a smooth steep spectrum which extends to soft X-ray and far-ultraviolet wavelengths.

Much of the gas observed in bright regions of ionised hydrogen has a temperature of about 10^4 K. The gas is excited by radiation from hot blue stars which have strong fluxes of radiation in the ultraviolet continuum. At 10^4 K, the main cooling mechanism for the gas is line radiation, the resonance lines of hydrogen or the forbidden transitions of singly and doubly ionised oxygen, [O II] and [O III], respectively (Sect. 12.3.2) These lines give ionised hydrogen clouds their characteristic red glow on colour photographs.

At temperatures less than 10^4 K, the ionised gas recombines and very few free electrons are present. Between 10^3 and 10^4 K, the principal radiation loss mechanism is the line emission of neutral or singly ionised carbon, nitrogen and oxygen associated with forbidden transitions of low lying energy levels. Observations from high flying aircraft such as the Kuiper Airborne Observatory have shown that the lines of [O I] (63 and 145 μm), [C I] (609 and 370 μm) and [C II] (157.7 μm) are particularly strong and are likely to be among the most important coolants of the interstellar gas in this temperature range.

At temperatures below about 10^3 K, *interstellar dust* can survive and plays a key role in determining the state of the gas at low temperatures. As described above, dust absorbs ultraviolet and optical radiation and therefore, within dust clouds, molecules are protected from the interstellar flux of dissociating radiation. Within the dust clouds, there are two important cooling processes. The first is molecular line emission associated either with rotational transitions of asymmetric molecules such as carbon monoxide, CO, and water vapour, H₂O, or, in some cases, with the infrared forbidden rotational and rotational-vibrational transitions of molecular hydrogen, H₂. In some regions these lines are so strong that they must be the dominant cooling mechanism. The second is the reradiation of optical and ultraviolet radiation absorbed by dust grains in the far-infrared waveband, the process described in Sect. 12.4.

12.5.4 The overall state of the interstellar gas

The picture which emerges is one in which many different processes contribute to the heating and cooling of the interstellar gas under different circumstances. The term *the violent interstellar medium* is often used, reflecting the fact that the medium is far from stationary, being constantly buffeted by supernova explosions and the winds from young stars and bipolar outflows as well as by large scale dynamical phenomena. In spite of the complexity of the interstellar medium, it is useful to have some reference figures to describe its various phases (Table 12.3). The diffuse phases have roughly the same pressure,

Table 12.3 The principal phases of the interstellar gas. (Courtesy of Dr. John Richer.)

Names	Main constituent	Detected by	Volume of interstellar medium	Fraction by mass	N (m^{-3})	Temperature (K)
‘Molecular clouds’	H ₂ , CO CS, etc	Molecular lines. Dust emission	~0.5%	40%	$\geq 10^9$	10–30
‘Diffuse clouds’	H, C, O with	21-cm		5%	40%	$10^6\text{--}10^8$
‘H I clouds’	some ions,	emission &				80
‘Cold neutral medium’	C ⁺ , Ca ⁺	absorption				
‘Intercloud medium’	H, H ⁺ , e ⁻ Ionisation fraction 10–20%	21-cm emission & absorption H α emission	40%	20%	$10^5\text{--}10^6$	8000
‘Coronal gas’	H ⁺ , e ⁻ Highly ionised species, O ⁵⁺ , C ⁺³ , etc	O VI Soft X-rays 0.1–2 keV	~50%	0.1%	$\sim 10^3$	$\sim 10^6$

$p = NkT$, and so they must be more or less in pressure equilibrium throughout much of the interstellar medium. Within the giant molecular clouds densities greater than 10^9 m^{-3} are found.

Why are some phases conspicuously present while others are not? The probable causes are *thermal instabilities* in the diffuse gas. The condition for a phase of the gas to be thermally unstable was first derived by Field in terms of a generalised heat-loss function \mathcal{L} , which is defined as the energy loss rate minus the rate of energy gain per unit mass of material per second (Field, 1965). In the stability analysis, it is assumed that the energy losses are by radiation and that the gas is optically thin. In the classic analysis of Field, Goldsmith and Habing, the heating was assumed to be due to the ionisation losses of low energy cosmic rays (Sect. 5.4) (Field *et al.*, 1969). Thus, the generalised loss rate can be written

$$\mathcal{L}(N, T) = \Lambda(N, T) - \Gamma , \quad (12.24)$$

where $\Lambda(N, T)$ is the cooling rate of the gas and Γ is the total heating rate. In the equilibrium state, there is balance between the heating and cooling rates so that $\mathcal{L} = 0$ and the gas is in pressure equilibrium. Field showed that the equilibrium state is unstable if $(\partial \mathcal{L} / \partial T)_p < 0$ (Field, 1965). The origin of this instability is clearly described by Shu (1992). Suppose in some region the density increases so that the rate of energy loss also increases. The region contracts and the decrease in thermal energy is partly or wholly offset by the work done by the surrounding medium on the perturbed cloud. The system is stable if the resulting pressure is more than sufficient to maintain pressure equilibrium but, if it is not,

the perturbation continues to collapse until a new equilibrium state is attained at a higher density and lower temperature.

In the analysis of Field, Goldsmith and Habing, it was shown that there are two stable phases of the interstellar medium at temperatures less than 10^4 K, one at about 8000 K and the other at the lower temperature of about 80 K, corresponding to two of the entries in Table 12.3. Between these temperatures, cooling due to the atomic and ionic lines described in Section 12.5.2 causes the gas to be unstable. This analysis gave rise to the concept of the *two-phase model* of the interstellar medium. Extending the analysis to higher temperatures, the existence of the hot coronal gas can be explained as well. Thus, although there is every reason to expect the interstellar medium to be in a state of continual flux, it is rather natural that the principal components listed in Table 12.3 should be in approximate pressure equilibrium.

12.6 Star formation

Star formation is important for high energy astrophysics because the star-formation rate is related to the rate of formation of the heavy elements and to the frequency of supernovae. The explosions of supernovae in the vicinity of molecular clouds may also stimulate the star-formation process. The subject of star formation is enormous and is comprehensively discussed in the book *The Formation of Stars* by Stahler and Palla (2005). Only those aspects needed for our future purposes are briefly reviewed here.

12.6.1 The initial mass function and the Schmidt–Kennicutt law

The *initial mass function* $\xi(M)$ describes the birth rate of stars of different masses. It is not trivial to determine this function observationally because stars are observed at widely differing stages of their evolution. The *luminosity function* of stars describes the numbers with different luminosities and can be converted into a mass function from the mass–luminosity relation. This function, however, underestimates the birth rate of stars more massive than $1 M_\odot$ since their lifetimes are shorter than the age of the Galaxy and so the statistics have to be corrected for the lifetimes of stars of different mass. A determination of the initial mass function for stars in the solar neighbourhood is shown in Fig. 12.10 from which it is apparent that it is a monotonically decreasing function of increasing mass. It is often convenient to adopt the *Salpeter initial mass function* $\xi(M) dM \propto M^{-2.35} dM$, shown as a dashed line in Fig. 12.10, as a reasonable approximation for stars with masses roughly that of the Sun (Salpeter, 1955). More recent determinations have suggested that the function can be described by the log-normal distribution function proposed by Miller and Scalo (1979)

$$\xi(\log M) dM \propto \exp[-C_1(\log M - C_2)^2] dM, \quad (12.25)$$

where C_1 and C_2 are constants (Fig. 12.10). Note that this function is a global average derived from local samples of stars.

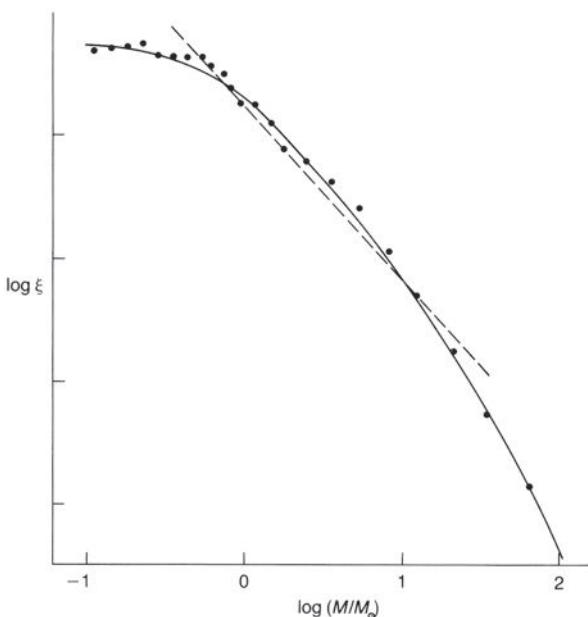


Fig. 12.10 An estimate of the initial mass function of stars derived by Miller and Scalo and their best-fitting log-normal distribution. Also shown as a dashed line is the initial mass function of power-law form proposed by Salpeter (Salpeter, 1955; Miller and Scalo, 1979).

Another important relation is dependence of the star-formation rate upon the density of the interstellar gas. This was first determined by Schmidt who studied the variation of the star-formation rate at different heights perpendicular to the Galactic plane as a function of gas density (Schmidt, 1959). His favoured solution was that the star-formation rate varies as the square of the gas density. Kennicutt compared the global star-formation rates in spiral and star-forming galaxies with their mean gas densities (Kennicutt, 1998). The mean star-formation rate was estimated from the H α intensity distribution and the total gas density from neutral hydrogen and CO observations in 61 normal spiral galaxies as well as far-infrared and CO observations of 36 infrared-selected starburst galaxies. This enabled the strong correlation between star-formation rate and gas density to be determined over a very wide range of gas densities and star-formation rates (Fig. 12.11). The disk-averaged star-formation rates and gas densities can be well represented by a Schmidt law $\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}^{1.40 \pm 0.15}$, where the Σ s refer to mean surface densities. This relation, often referred to as the *Schmidt–Kennicutt law*, is commonly used in constructing models of galaxy evolution. Note that the law refers to global averages rather than to any particular star-formation region and is an empirical result.

12.6.2 Regions of star formation

Stars form within giant molecular clouds, the typical properties of which are listed in Table 12.3. The giant molecular clouds have sizes vastly greater than the prominent regions

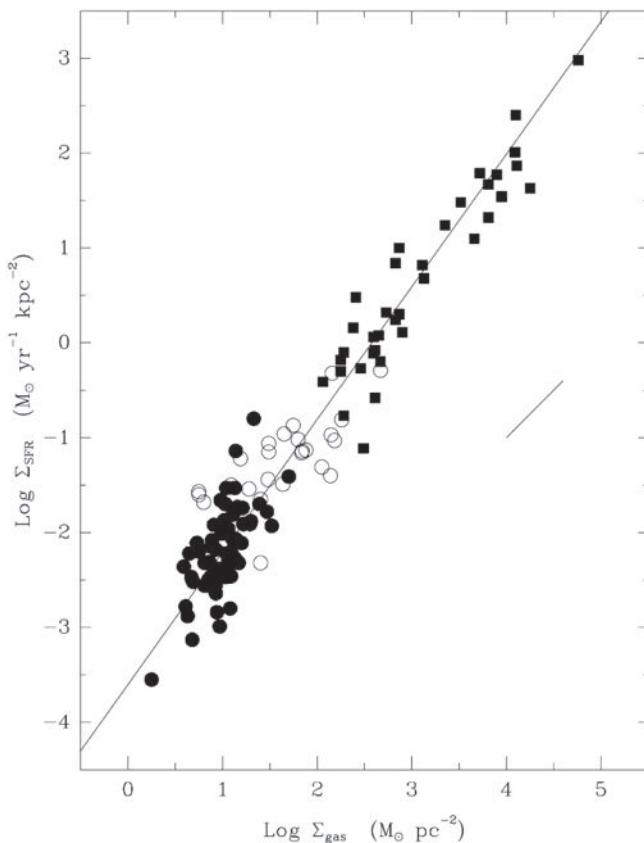


Fig. 12.11 The correlation between star-formation rate and surface gas density for a sample of 61 spiral galaxies and 36 star-forming galaxies (Kennicutt, 1998). The filled circles are normal disc galaxies and the squares starburst galaxies. The open circles show the star-formation rates and gas densities for the central regions of normal disc galaxies. The straight line corresponds to $\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}^{1.40}$.

of ionised hydrogen such as the Orion Nebula. Fig. 12.12a is an optical photograph of the constellation of Orion, created by patching together a number of 6° Schmidt Telescope plates. The Orion Nebula is the most prominent region of ionised hydrogen towards the top right of the box labelled Orion A Molecular Cloud. The Orion Nebula is dwarfed by the Orion Molecular Cloud which extends over about 16° on the sky, roughly the same size as the constellation of Orion. The southern region of the Orion giant molecular clouds is shown in higher resolution in Fig. 12.12b, which shows that there is a great deal of fine structure within the molecular clouds, each density enhancement being a potential site of star formation. There are large quantities of dust associated with the clouds which protect the interstellar molecules from being photodissociated by the interstellar flux of ionising radiation. Consequently, we tend to see optically only those regions of ionised hydrogen which lie close to the front surface of the clouds. The Orion Nebula, for example, is probably a ‘blister’ on the front surface of the Orion giant molecular clouds.

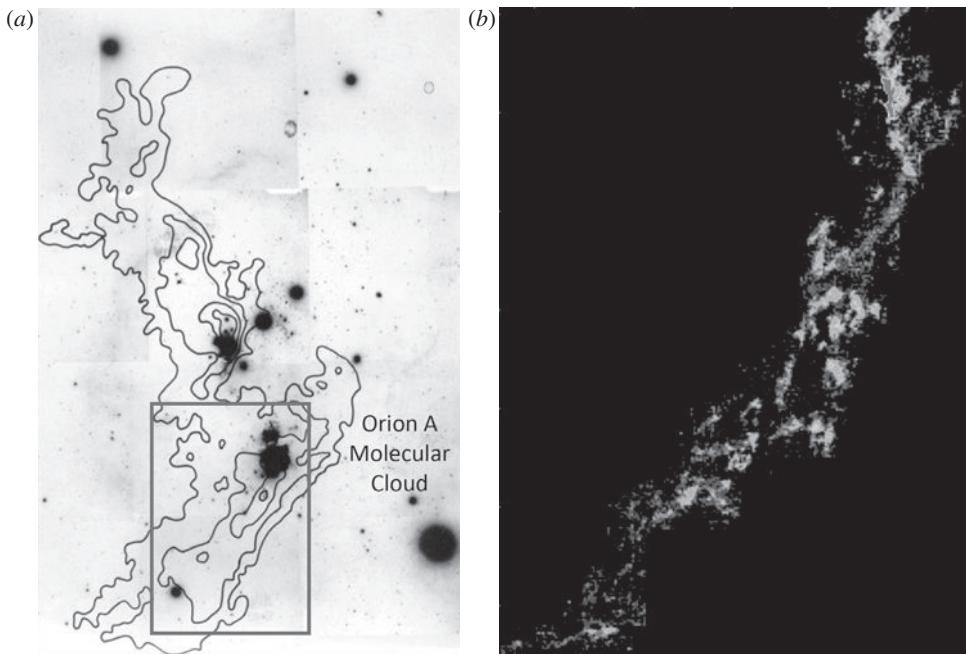


Fig. 12.12 (a) An optical image of the constellation of Orion superimposed upon which are contours of the CO emission showing the extent of the Orion giant molecular clouds. The familiar bright stars of the constellation of Orion can be seen. (Courtesy of the Royal Observatory Edinburgh). (b) A high resolution map of the Orion A Molecular Cloud, indicated by the box in (a), observed in the 49 GHz line of CS by the 45 m radio telescope of the Nobeyama Radio Observatory, Japan. (Courtesy of K. Tatematsu.) The most intense molecular line emission is associated with the Orion Nebula. The many compact knots are sites of the next generation of new stars.

The youngest objects observed in the clouds are the hot far-infrared sources. Optically, the Trapezium stars, seen close to the centre of Fig. 12.13, are the brightest stars in the region of Orion, but at far-infrared wavelengths most of the luminosity is associated with the region to the north-west of it where the Becklin–Neugabauer (B–N) object and the Kleinmann–Low Nebula are located. The B–N object is a compact far-infrared source with far-infrared luminosity about 10^5 times that of the Sun. Its spectrum is sharply peaked in the far-infrared region of the spectrum, typical of the emission spectrum of reradiated dust (Fig. 12.6a below). There is no region of ionised hydrogen surrounding the B–N object, suggesting that the stars must be very newly formed or even in the process of formation.

An important feature of the far-infrared sources found in star-forming regions is that virtually all of them are associated with *bipolar outflows*. A number of these are associated with the optical emission line nebulae known as *Herbig–Haro, or HH, objects* which are found in the vicinity of stars in the process of formation. Observations at millimetre and infrared wavelengths have shown that molecular outflows from the protostar are powered by highly collimated molecular beams ejected in opposite directions from the infrared sources.

**Fig. 12.13**

A composite infrared image of the Orion Nebula as observed in the J, H and K infrared wavebands made using the near-infrared camera ISAAC on the ESO 8.2-m VLT Antu telescope. A total of 81 individual ISAAC images were merged to form this mosaic. The four bright Trapezium stars are in the centre of the image. To the north-west of these are the obscured luminous far-infrared sources, the Becklin–Neugebauer object and the Kleinmann–Low nebula. (Courtesy of Mark McCaughrean and ESO.)

Figure 12.14a shows the remarkable bipolar outflow source HH211 (Gueth and Guilloteau, 1999). The central region is obscured optically and so the structures shown in Fig. 12.14a were observed at infrared or millimetre wavelengths. The underlying image shows the distribution of molecular hydrogen as observed in the $2.12\text{ }\mu\text{m}$ infrared S0 vibrational line of H_2 – this emission is associated with shock-excitation of molecular gas. The contour map shows the structure of the jets in the CO $j = 1 \rightarrow 0$ rotational transition observed by the IRAM millimetre interferometer on the Plateau de Bure. In the very centre of the image is a compact submillimetre source, the source of the outflow, which contains a protostar, or very young star. The velocities of the jets powering the bipolar outflows, as measured from the Doppler shifts of the molecular lines, are found to be highly supersonic, jet velocities as large as $50\text{--}100\text{ km s}^{-1}$ being observed. Similar structures have been observed in other Herbig–Haro objects. Figure 12.14b shows an image of HH34 taken with the FORS2 instrument on the 8-metre Kueyen Telescope of the VLT. The structure is similar to that of HH211. Figure 12.14c shows a Hubble Space Telescope image of the central core of the Herbig–Haro object HH30. Images taken at different epochs have shown that the proper motions of the jet correspond to velocities of about 200 km s^{-1} .

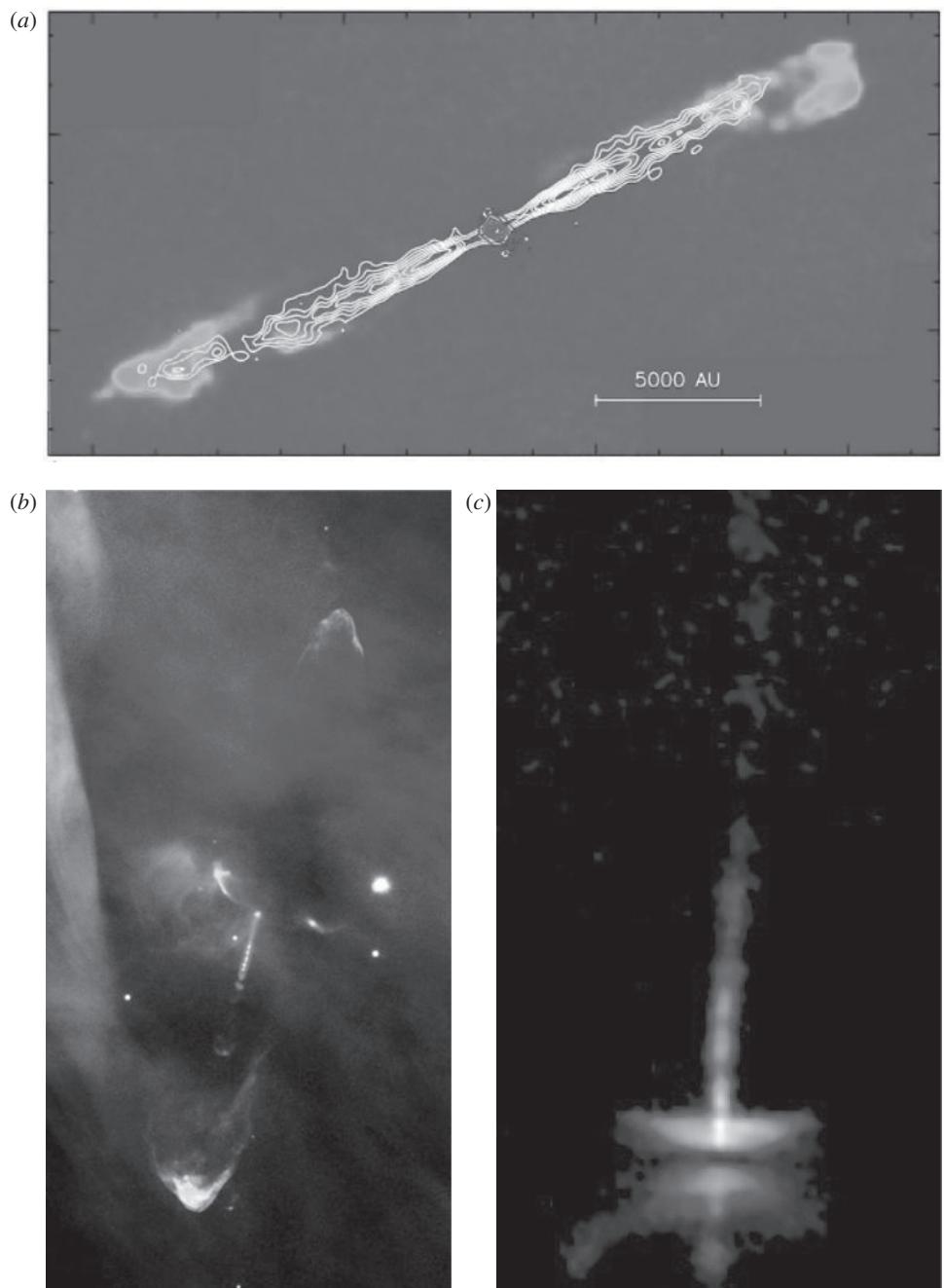


Fig. 12.14 (a) The bipolar outflow source HH211 observed in the H_2 line at $2.12 \mu\text{m}$ superimposed upon which are contours of the $\text{CO} j = 1 \rightarrow 0$ rotational transition. A submillimetre source is observed at the location of the protostar (Gueth and Guilloteau, 1999). (b) The bipolar outflow in the Herbig–Haro object HH34 taken with the FORS2 instrument on the Kueyen Telescope of the VLT. (Courtesy of ESO.) (c) The Herbig–Haro object HH30 observed by the Hubble Space Telescope. The image shows the jet originating close to the protostar which is obscured by a disc of material seen edge-on. The proper motion of the features in the jet correspond to velocities of order 200 km s^{-1} . (Courtesy of Alan Watson, NASA, ESA and the Space Telescope Science Institute.)

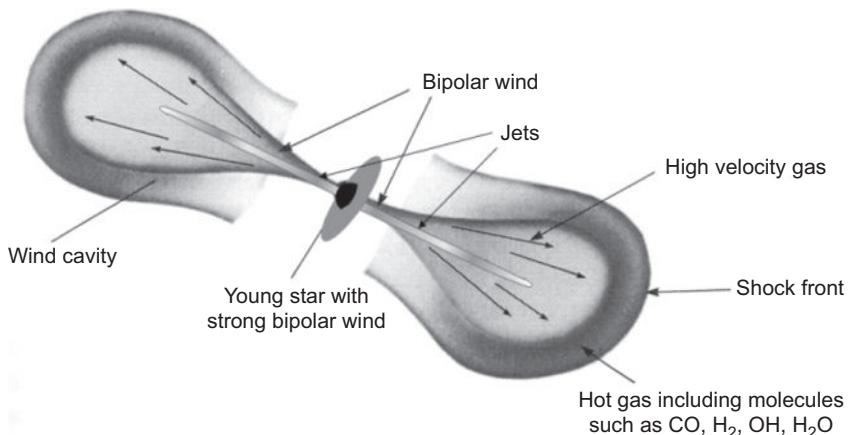


Fig. 12.15 A schematic diagram illustrating the characteristics of a typical bipolar outflow source. The outflow is supersonic and compresses the surrounding molecular gas. Some of the gas is ejected in narrow jets which are aligned with the polar axis of the protostar and its protoplanetary disc. The heating of the molecular gas by the outflow and the cooling by molecular line emission results in a temperature of about 2000 K at which it can be observed through its infrared molecular line emission.

Polarisation observations of the infrared molecular hydrogen emission in Orion show, in addition to a molecular hydrogen reflection nebula, polarisation vectors parallel to the molecular outflow. This is interpreted as evidence for a magnetic field in the outflow. A schematic representation of the structure of a bipolar outflow is shown in Fig. 12.15. It is striking that the structures seen in protostellar objects are very similar in appearance to those observed in extragalactic radio sources, the big differences being that in the case of protostars, the jets consist of molecular material ejected from the vicinity of a star in the process of formation, whereas in the case of the extragalactic radio sources, the jets consist of relativistic particles and magnetic fields and are on a scale about a million times greater than those of the protostars.

12.6.3 Issues in the theory of star formation

An outline of a plausible scenario for the formation of stars was given towards the end of Sect. 12.4. That summary disguises the fact that there are three problems which have to be solved to understand how regions with densities about 10^9 m^{-3} , typical of giant molecular clouds, can collapse to form stars with about 10^{30} times greater densities. First of all, there is an *energy* problem. To form a stable star, the protostar must get rid of its gravitational binding energy – this is solved by the radiative loss of energy by the reradiation of dust grains in the collapsing protostar. Second, any cloud possesses some angular momentum and, because of conservation of angular momentum, the rotational energy increases during collapse. Unless there is some way of getting rid of angular momentum, the growth of rotational energy will halt the collapse in the equatorial plane – this is the *angular momentum* problem. Third, if there is a *magnetic field* present in the collapsing cloud, its field strength is amplified during

Table 12.4 The Jeans criterion and the contents of giant molecular clouds.

	GMC	Clump	Dense core
Size	50 pc	10 pc	0.1 pc
Mass	$10^5 M_\odot$	$30-10^3 M_\odot$	$3-10 M_\odot$
Number density	10^8 m^{-3}	$5 \times 10^8 \text{ m}^{-3}$	$5 \times 10^{10} \text{ m}^{-3}$
Temperature	15 K	10 K	10 K
Jeans length	4 pc	1.5 pc	0.15 pc
Jeans mass	$600 M_\odot$	$100 M_\odot$	$30 M_\odot$

collapse and this could become sufficiently strong to halt collapse in the equatorial plane (Sect. 11.2.2).

Gravity ensures that, on a large enough scale, a gas cloud of any density and temperature is unstable because of the *Jeans instability*. If a uniform medium is perturbed, the self-gravitation of the perturbation causes the region to collapse, but this is resisted by internal pressure gradients. The criterion for collapse is therefore that the gravitational force should exceed the internal pressure forces. The force of gravity acting on 1 m^3 of matter at the edge of the uniform cloud of mass M , radius R and density ρ is $\sim GM\rho/R^2$ while the force associated with the pressure gradient which prevents collapse is $dp/dr \sim p/R$. When the former exceeds the latter, collapse occurs. Since the speed of sound c_s is approximately $(p/\rho)^{1/2}$ and $M \sim \rho R^3$, the condition $GM\rho/R^2 > p/R$ reduces to $R \geq R_J = c_s/(G\rho)^{1/2}$, where the characteristic length-scale R_J is known as the *Jeans length*. It is the largest scale a cloud can have before collapse under self-gravity is inevitable. We can also define a *Jeans mass* as the mass contained within the region which has scale R_J . To order of magnitude,

$$M_J \sim \rho R_J^3 \approx 10^5 \frac{T^{3/2}}{\mu_H^2 N^{1/2}} M_\odot, \quad (12.26)$$

where μ_H is the mean molecular weight of the particles contributing to the pressure relative to the mass of the hydrogen atom. The time-scale for collapse of the unstable region is roughly $\tau \approx R_J/c_s \sim (G\rho)^{-1/2}$. I have given elsewhere a more formal derivation of these results starting from the equations of gas dynamics coupled with Poisson's equation for the gravitational potential (Longair, 2008).

The values of the Jeans length and Jeans mass for the typical structures observed in giant molecular clouds are listed in Table 12.4. It is clear from these figures that giant molecular clouds are unstable against fragmentation and collapse. As the collapse proceeds, the density increases, but the cloud continues to remain cool because of radiation by molecular lines and dust emission. Therefore, the Jeans length becomes smaller and fragmentation continues. It is therefore natural that giant molecular clouds are the seats of active star formation. The fragmentation ceases when the cloud becomes optically thick to radiation which is expected to occur for masses $M \sim 0.01 M_\odot$. In fact, all the stars we observe have masses greater than $0.1 M_\odot$ and this is attributed to the fact that stars with mass less than about $0.08 M_\odot$ are not hot enough in their centres for nuclear burning to take place.

Another problem concerns the origin of the bipolar outflows. The rotation axis of the accretion disc provides a natural axis for the ejection of matter. According to the picture of Shu, Adams and Lizano, the collimation may be associated with the escape of a stellar wind as it escapes from the accreting envelope along the path of least resistance, which is along the rotation axis (Shu *et al.*, 1987). The hot wind may be associated with the dissipation of energy in the boundary layer between the accretion disc and the stellar surface or with the hot innermost layers of the accretion disc. The hot gas may be channelled by the magnetic field in the ‘magnetosphere’ of the accreting star along the polar directions. We return to these problems when we tackle the origin of jets and beams in extragalactic radio sources. Undoubtedly, magnetic fields are involved in the collimation of the jets observed in the bipolar outflows.

12.7 The Galactic magnetic field

12.7.1 Faraday rotation in the interstellar medium

We have already shown in Sect. 12.3.4 how measurements of the Faraday rotation of the plane of polarisation of polarised radio waves provides information about the rotation measure, which is proportional to the quantity $\int N_e B_{\parallel} dl$ along the line of sight to a radio source. A plot of the magnitude of the rotation measure as a function of Galactic latitude b shows that the rotation measures of extragalactic radio sources increase towards low Galactic latitudes (Fig. 12.16a). If the Galactic magnetic field were uniform and ran parallel to the plane of the Galaxy and if the electron density were uniform, the path length through the Galactic disc would be proportional to $\text{cosec } b$ and the component of the magnetic field along the line of sight would be proportional to $\cos b$. Therefore it would be expected that the rotation measure would vary as $\cot |b|$. This relation provides a reasonable upper envelope to the distribution of points in Fig. 12.16a and so most of the Faraday rotation of extragalactic radio sources originates within our own Galaxy rather than in the sources themselves. There is, however, a large scatter in the values of the rotation measures at any given Galactic latitude, in particular, even at low Galactic latitudes there are some sources with very small rotation measures. There must, therefore, be considerable irregularities in the distribution of the product $N_e B_{\parallel}$ along the line of sight.

If the magnitudes and signs of the rotation measures are plotted in Galactic coordinates (Fig. 12.16b), there is general clustering of rotation measures of the same sign in different directions, which is evidence that there is some overall order in the Galactic magnetic field. The signs of the rotation measures change about Galactic longitude 180° , particularly in the southern Galactic hemisphere, suggesting that the parallel component of magnetic field changes direction at this longitude. This evidence is consistent with a model in which the magnetic field runs predominantly parallel to the plane of the Galaxy in the direction of the local spiral arm. The sense of the field is such that it points away from the Earth in the direction of galactic longitude roughly 90° . The magnetic field directions in some spiral galaxies are parallel to the spiral structure as is beautifully

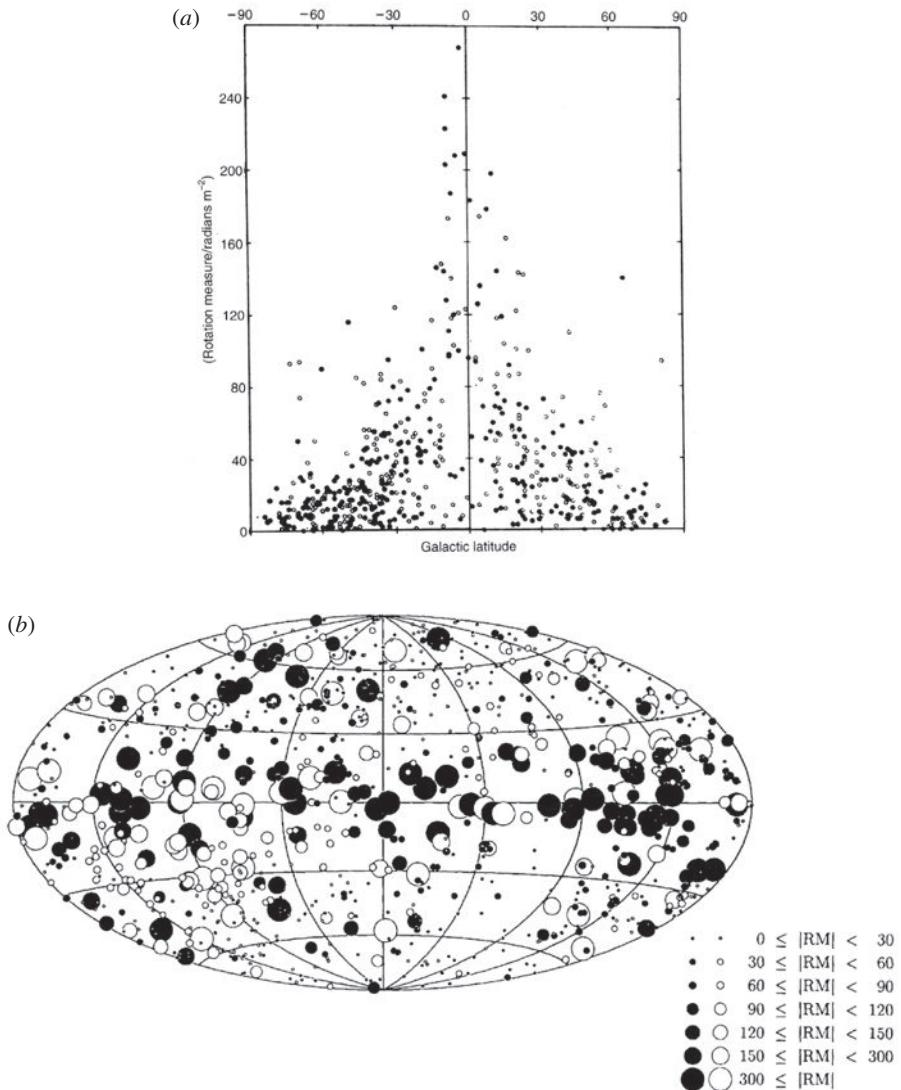


Fig. 12.16 (a) The variation of the rotation measures of extragalactic radio sources with galactic latitude. The largest rotation measures are found close to the Galactic plane (Whiteoak, 1974). (b) The magnitudes and signs of the rotation measures of 976 extragalactic radio sources plotted in galactic coordinates (Wielebinski, 1993).

illustrated by high sensitivity radio observations of the galaxy M51 by Neininger (1992) (Fig. 12.17).

Another use of this technique is to combine the rotation measures of pulsars with their dispersion measures, which provide measures of $\int N_e dl$. If attention is restricted to pulsars at distances less than 2 kpc in the Galactic plane, it is found that they are consistent with a uniform magnetic field of strength 2.5×10^{-10} T running parallel to the Galactic plane

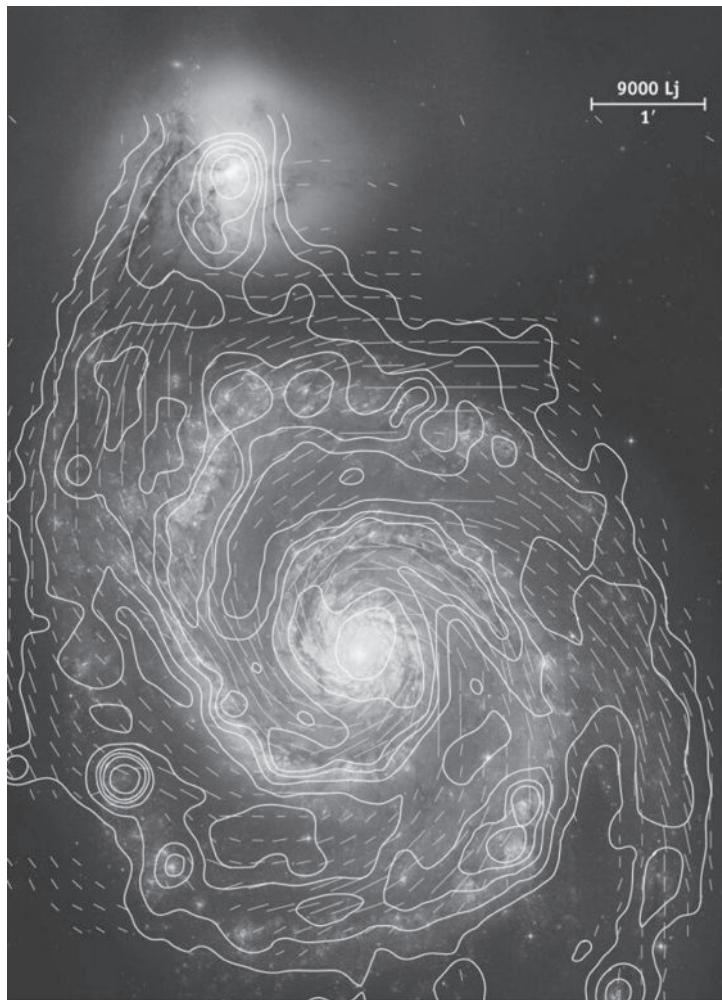


Fig. 12.17 The magnetic field distribution in the spiral galaxy M51 as observed by the Effelsberg 100 m telescope and the Very Large Array superimposed upon an HST image of the galaxy. (Courtesy of the Max-Planck-Institut für Radioastronomie, Bonn and the NRAO, Charlottesville, USA.)

in the direction of longitude $l = 90^\circ$ (Heiles, 1976). It is apparent, however, that there are also large scale irregularities in the field on large and small angular scales.

12.7.2 Optical polarisation of starlight

At optical wavelengths, the light of reddened stars is often found to be polarised. The degree of polarisation depends upon wavelength and can be written empirically in terms of the maximum degree of polarisation p_{\max} at the wavelength λ_{\max} as

$$p(\lambda) = p_{\max} \exp [-K (\ln(\lambda/\lambda_{\max}))] , \quad (12.27)$$

where $\lambda_{\max} \approx 550$ nm and $K \approx 1$ (Serkowski, 1973). The degree of polarisation is strongly correlated with extinction

$$p_{\max} \leq 0.03 A(\lambda_{\max}) \text{ mag}^{-1}, \quad (12.28)$$

where $A(\lambda_{\max})$ is the extinction at wavelength λ_{\max} (Serkowski *et al.*, 1975). The polarisation is naturally attributed to differential extinction by aligned dust grains, the alignment being due to the presence of a large scale magnetic field along the line of sight to the star. The dust grains must be significantly non-spherical and sufficiently aligned so that the extinction is about 6% greater in one polarisation than in the other. Percentage polarisations less than the empirical relation (12.28) are attributed to the fact that along some lines of sight the magnetic field is disordered or that the grains are not so well aligned. The extinction occurs preferentially in that polarisation of the incident light waves which has the electric field vector parallel to the long axes of the grains. Therefore the transmitted radiation is polarised parallel to the minor axes of the grains.

Despite the fact that the polarisation of starlight was discovered as long ago as 1949 by Hall and Hiltner, the understanding of the physical processes involved in the alignment of the dust grains have proved elusive because of the complexities of understanding the physical properties of the grains and some quite subtle pieces of physics involved in the magnetic properties of asymmetric dust grains. An excellent survey of these physical processes and problems of grain alignment is given by Draine (2004).

Two separate phenomena contribute to the alignment mechanism (Spitzer, 1968). First, if elongated dust grains are described by prolate spheroids with principal axes $a_1 > a_2 = a_3$ and the principal moments of inertia about the grain axes are I_1 , I_2 and I_3 , the moment of inertia about the major axis I_1 is smaller than those about the minor axes I_2 and I_3 . Let $I_2 = I_3 = \gamma I_1$, where $\gamma > 1$. In statistical equilibrium, the rotational energy about each principal axis is the same, $\frac{1}{2}I_1\omega_1^2 = \frac{1}{2}I_2\omega_2^2 = \frac{1}{2}I_3\omega_3^2$ and therefore $I_2\omega_2 = I_3\omega_3 = \gamma^{1/2}I_1\omega_1$. Therefore, the angular momentum vectors of the rotating grains in equilibrium lie preferentially perpendicular to the major axis of the grain. Consequently, there is greater extinction for the polarisation parallel to the major axis of the grain and so the light is polarised parallel to the rotation axis of the grain.

The second part of the story is more complicated than the first and concerns the alignment of the rotation axis of the grains with the magnetic field direction. First of all, because of the equipartition of energy, the grains must be rotating quite rapidly. Equating the total angular rotational energy $\frac{1}{2}I\omega^2$ to $\frac{3}{2}kT$ and setting $I = \frac{2}{5}mr^2$ for spherical dust grains of mass m and radius a , the root mean square angular velocity of the grains is

$$\langle \omega^2 \rangle^{1/2} = \left(\frac{15kT}{2ma^2} \right)^{1/2} = 4.6 \times 10^4 \left(\frac{T}{100 \text{ K}} \right)^{1/2} \left(\frac{3000 \text{ kg m}^{-3}}{\rho} \right)^{1/2} \left(\frac{10^{-7} \text{ m}}{a} \right)^{5/2} \text{ Hz}. \quad (12.29)$$

Thus, for the typical properties of grains responsible for extinction in the optical wavebands, the grains have angular rotation speeds of order 10^5 rad s⁻¹ or 10^4 Hz.

In addition, the grains may become charged. The most important processes for charging the grains are collisions with electrons and ions and photoelectric emission. The first process

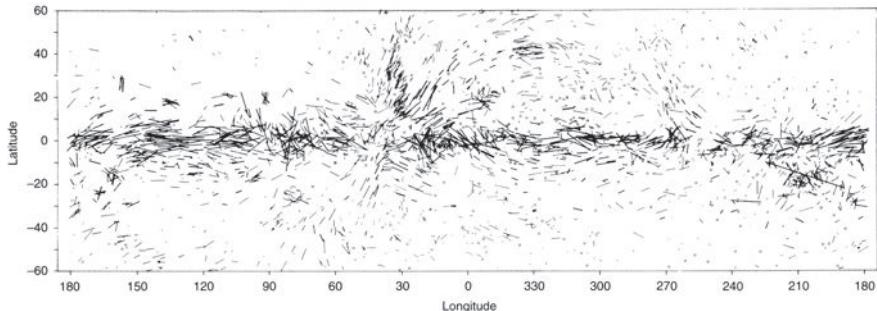
tends to make the grains negatively charged since the electrons have much greater speeds than the ions, while the second process, by ejecting electrons from the grains, tends to make the grains positively charged. Draine describes clearly the condensed matter physics involved in these processes and shows that, under interstellar collisions, the balance may tip either way, depending on the composition of the grains, their sizes, the electron density n_e and temperature T and the spectrum and intensity of the ultraviolet background due to starlight (Draine, 2004). The computations of Draine indicate how the charge on the grain depends upon their sizes, chemical compositions and the medium in which they are located. The typical small grain, $a \leq 70$ nm, picks up one or two positive or negative charges. Larger grains, $a \sim 0.5$ μm , can typically maintain about 10 electronic charges. The electric charges on the grains play important roles in a number of key processes in the physics of the interstellar medium, for example, coupling the grains and neutral particles to the magnetic field, increasing the drag on the grains due to Coulomb interactions with ions in the gas and the injection of energetic photoelectrons into the gas, which is a heating mechanism for interstellar gas.

As Draine has emphasised, there are many well established physical processes by which grains can be aligned by a magnetic field. He lists the following processes:

- The Rowland effect: a charged, spinning dust grain will develop a magnetic moment due to its circulating charge.
- The Barnett effect: a spinning dust grain with unpaired electron spins will spontaneously magnetise.
- Suprothermal rotation due to dust–gas temperature differences.
- Suprothermal rotation due to photoelectric emission.
- Suprothermal rotation due to H_2 formation.
- Viscoelastic dissipation of rotational kinetic energy due to time-varying stresses in a grain which is not rotating around a principal axis.
- Barnett dissipation of rotational kinetic energy due to the electron spin system.
- Dissipation of rotational kinetic energy due to the nuclear spin system.
- Suprothermal rotation due to starlight torques.
- Fluctuation phenomena associated with Barnett dissipation and coupling to the nuclear spins.

Some of these processes are referred to as suprothermal in the sense that, although the grains have temperatures ~ 30 K, they are not in thermal equilibrium with the incident UV starlight, nor with the hot gas, nor with the energetic particles associated with the ejection of photoelectrons and H_2 molecules. Hence, the grains can be spun up to angular velocities exceeding their thermal values, $\frac{1}{2}I\omega^2 = \frac{3}{2}kT$.

In the original picture of Greenstein and Davis, rotation about an axis parallel to a magnetic field is favoured because the component of magnetisation of a paramagnetic material about that axis does not change, whereas rotation about the other axes results in the direction of magnetisation changing continuously and internal couples result in the damping of rotation about these axes by paramagnetic dissipation (Davis and Greenstein, 1951). The problem with this mechanism is that random collisions tend to destroy the alignment. The full complexities of grain alignment mechanisms are described in Draine's

**Fig. 12.18**

The polarisation of stars as a function of Galactic coordinates. The magnitudes of the vectors indicate the strength of the polarisation and the directions of the vectors indicate the planes of polarisation of the light (Matthewson and Ford, 1970).

survey which is strongly recommended. His conclusion is that if the grains are spun up to suprathermal rotation, disalignment by random collisions is no longer important and that the alignment results from the combined effects of the Davis–Greenstein alignment and starlight torques.

The net result is that elongated grains rotate with their minor axes parallel to the magnetic field direction and so the electric vector of the transmitted radiation is parallel to the magnetic field direction. From a study of the polarisation properties of about 6000 stars, Mathewson and Ford derived the map of their polarisation vectors shown in Fig. 12.18, the lengths of the lines indicating the percentage polarisation (Mathewson and Ford, 1970). All stars plotted lie within 3 kpc of the Sun. The magnetic field runs predominantly parallel to the Galactic plane in agreement with the observations of the intrinsic polarisation of the Galactic radio emission. These observations have suggested that the uniform magnetic field component runs in the general direction of the local spiral arm, $l \approx 50^\circ\text{--}80^\circ$. There are also large scale irregularities in the magnetic field distribution, some of which are associated with Galactic loops such as the North Polar Spur, the prominent feature which extends towards the north Galactic pole from $l \approx 30^\circ$ (see also Fig. 1.8a). Thus, the polarisation vectors provide information about the overall field direction, but the detailed physics is not secure enough to enable estimates of the magnitude of the magnetic flux density to be made.

12.7.3 Radio emission of spinning dust grains

A consequence of the finite electric dipole moment of dust grains is that, since they must be spinning, they radiate dipole radiation according to the Larmor formula (6.8),

$$-\frac{dE}{dt} = \frac{|\vec{p}|^2}{6\pi\epsilon_0 c^3}, \quad (12.30)$$

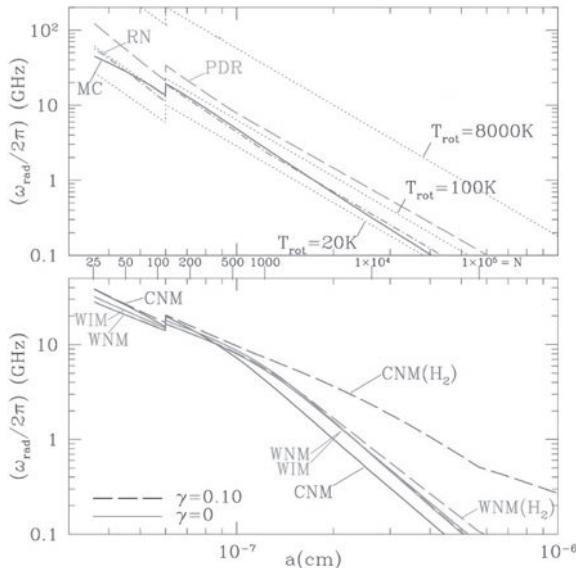


Fig. 12.19 Effective rotation rate ω as a function of grain radius a for various environmental conditions. The acronyms have the following meanings: CNM – cold neutral medium; WNM – warm neutral medium; WIM – warm ionised medium; MC – molecular cloud; DC – dark cloud; RN – reflection nebula; PDR – photodissociation region; CNM(H_2) and WNM(H_2) include torques due to H_2 formation on grains. The thermal rotation rates at $T = 20, 100$, and 8000 K are also shown. The number of atoms N in a grain is indicated between the two panels (Draine and Lazarian, 1998).

where p is the electric dipole moment. Writing $p = p_0 e^{i\omega t}$, where ω is the angular frequency of rotation, and averaging over a cycle of the emission, the radiation rate is

$$-\left(\frac{dE}{dt}\right)_{av} = \frac{p_0^2 \omega^4}{12\pi\epsilon_0 c^3}. \quad (12.31)$$

The electric dipole moment p_0 of the grain is the sum of its intrinsic dipole moment p_i and the dipole moment associated with the charge it acquires by the processes described in the last section, $p_e = Zea_e$, where a_e is the displacement of the charge from the centre of momentum of the grain. Thus, $p_0 = p_i + Zea_e$. Draine and Lazarian adopt a typical value of a_e of about 0.01 times the radius of the grain (Draine and Lazarian, 1998).

In order to work out the dipole emission of the dust grains, three sets of data are needed – the numbers of small grains, their dipole moments p_0 and their angular velocities. Draine and Lazarian provide detailed calculations of what is involved in determining these input data (Draine and Lazarian, 1998). Of particular significance is the fact that, according to (12.29), small charged dust grains with dimensions of the order of 10^{-9} m , the size of typical PAH molecules, radiate at about 10 GHz. As they emphasise in their analysis, this is a rather crude estimate since the processes which excite and damp the rotation of the grains are far from thermodynamic equilibrium. The results of their detailed calculations for the rotation frequency–grain radius relation are shown in Fig. 12.19 for different phases of the interstellar medium. The discontinuity at $N = 120$ atoms is an artifact of the assumption

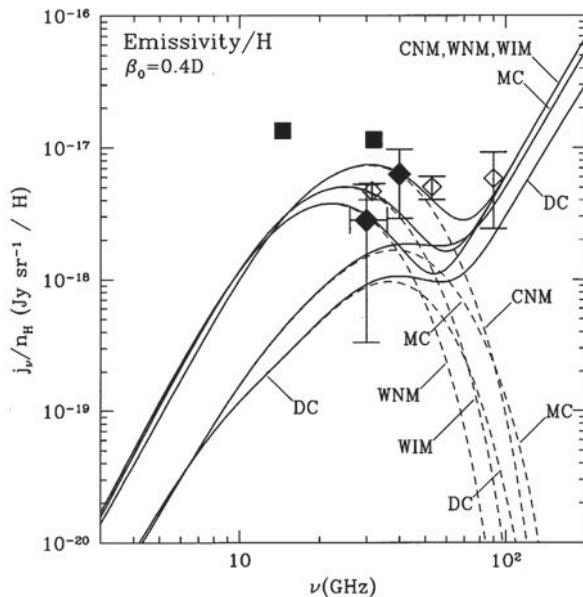


Fig. 12.20

The emissivity per hydrogen atom due to rotating dust grains for the phases of the interstellar medium listed in the caption to Fig. 12.19. Solid line is total emissivity; dashed line is rotational emission from ultra-small spinning grains (Draine and Lazarian, 1998).

that larger grains are spherical and small grains are planar. It is apparent that, for grains which radiate at $\nu \sim 10$ GHz, the thermal equilibrium rotation rates are a reasonable approximation for rotational temperatures $T_{\text{rot}} \sim 20-100$ K. The important point is that grains with radii $a \sim 10^{-9}$ m are expected to radiate at frequencies 10–100 GHz. This process may therefore contribute to the Galactic background radiation in the wavebands at which observations of the minute fluctuations in the Cosmic Microwave Background Radiation are carried out. Draine and Larazian adopt the typical charges found from their detailed theoretical calculations and use the log-normal size distribution of small grains needed to account for the emission observed in the 12 and 25 μm wavebands which are attributed to PAH emission.

The predicted spectra for the different phases of the interstellar medium are compared with the COBE observations in Fig. 12.20. The form of the emission spectra of the small grains can be understood as follows. The one-to-one relation between angular frequency and grain radius and the size distribution of the grains determine the number of emitters which radiate at frequency ν . The emission is then weighted as ν^4 according to the Larmor radiation formula (12.31). The computations shown in Fig. 12.20 suggest that it is entirely plausible that rotating charged dust grains contribute to the Galactic background radio emission in the 10–100 GHz waveband. Evidence for the detection of the radio emission from rotating dust grains from a number of Galactic sources is discussed by Davies (2006).

12.7.4 Zeeman splitting of 21-cm line radiation

The Galactic magnetic field strength may also be estimated from the Zeeman splitting of the 21-cm neutral hydrogen line. The observational problem is formidable since the splitting amounts to only 28 GHz T^{-1} and the expected magnetic field strengths are only about $10^{-9} - 10^{-10} \text{ T}$. Thus, the radio spectrometers must be sensitive enough to detect splittings of about 10 Hz in 1420 MHz. If, however, the magnetic field runs parallel to the line of sight, Zeeman splitting results in two circularly polarised components with opposite senses of circular polarisation on opposite sides of the line centre. The splitting is always much less than the width of the absorption line and therefore the technique adopted is to observe an intense radio or millimetre absorption line and to search for an excess of oppositely circularly polarised radiation on either side of the line centre.

Magnetic field strengths have been measured by this technique using the 21-cm line of neutral hydrogen in the direction of a number of intense radio sources and are found to be greater than 10^{-9} T . It is probable that these strong magnetic fields are associated with the high density gas clouds responsible for the formation of the absorption line rather than with the general interstellar medium. Similar observations have been made of OH absorption lines and even stronger magnetic field strengths, about 10^{-8} T , have been found. These high magnetic field strengths are likely to be associated with the dense clouds in which the OH absorption takes place.

12.7.5 The radio emission from the Galaxy

The diffuse Galactic radio emission and its polarisation are attributed to the synchrotron radiation of ultrarelativistic electrons spiralling in the Galactic magnetic field. As discussed in Sect. 8.9, there are problems in deriving a unique value for the magnetic field strength from these observations but the values are more or less in agreement with the other independent pieces of evidence.

12.7.6 Summary of the information on the Galactic magnetic field

The various techniques described above provide complementary information about different aspects of the Galactic magnetic field. The distribution of the rotation measures of pulsars and extragalactic radio sources and of the optical polarisation vectors are convincing evidence that there exists some large scale order. In the vicinity of the Sun, the uniform component of the field runs roughly in the direction $l = 90^\circ$ along the local spiral arm. A mean value of the magnetic flux density of $(2-3) \times 10^{-10} \text{ T}$ is consistent with much of the evidence but there must be significant fluctuations about this value with $\Delta B/B \sim 1$ on a wide range of scales. In clouds, the Zeeman splitting experiments indicate that somewhat stronger magnetic fields are present.

The stars described in Chap. 3 are held up by the thermal pressure of hot gas, the source of energy being nuclear energy generation in their central regions. As evolution proceeds from the main sequence, up the giant branch and towards the final phases when the outer layers of the giant star are ejected, nuclear processing continues until the available nuclear energy resources of the star are exhausted. The more massive the star, the more rapidly it evolves and the further it can proceed along the path to the synthesis of iron, the most stable of the chemical elements. In the most massive stars, $M \geq 8 M_{\odot}$, it is likely that the nuclear burning can proceed all the way through to iron whereas in less massive stars, the oxygen flash, which occurs when core burning of oxygen begins, may be sufficient to disrupt the star. In any case, at the end of these phases of stellar evolution, the core of the star runs out of nuclear fuel and collapses until some other form of pressure support enables a new equilibrium configuration to be attained.

Possible equilibrium configurations which can exist when the nuclear fuel runs out are as *white dwarfs*, *neutron stars* or *black holes*. In white dwarfs and neutron stars, the star is supported by degeneracy pressure associated with the fact that electrons, protons and neutrons are fermions and so only one particle can occupy any single quantum mechanical state. White dwarfs are held up by *electron degeneracy pressure* and can have masses up to about $1.4 M_{\odot}$. In neutron stars, *neutron degeneracy pressure* is responsible for the pressure support and they can have masses up to about $1.4 M_{\odot}$, possibly slightly higher if the neutron star is rapidly rotating. More massive dead stars must be black holes. This knowledge does not help us decide which types of star become white dwarfs, neutron stars or black holes. For example, low mass stars with $M < 2 M_{\odot}$, can in principle end up in any of the three forms. Even stars with masses very much greater than $2 M_{\odot}$ can form white dwarfs or neutron stars if they lose mass sufficiently rapidly. Computations of mass loss during the late stages of stellar evolution have shown that even $10 M_{\odot}$ stars can lose mass very effectively towards the ends of their lifetimes and form non-black hole remnants.

13.1 Supernovae

13.1.1 The historical supernovae and supernova typology

The formation of neutron stars and black holes must be associated with the rapid liberation of huge amounts of energy, the gravitational binding energy of a $1 M_{\odot}$ neutron star being

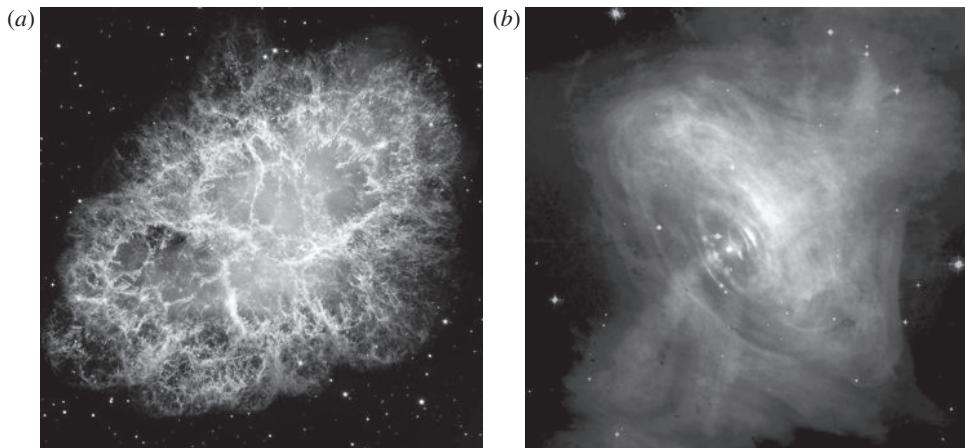


Fig. 13.1 (a) The Crab Nebula, also known as M1 and NGC 1952, as observed by the Hubble Space Telescope, (b) A composite X-ray-optical image of the Crab Nebula made by the Chandra X-ray Observatory. The bright central X-ray source is the Crab Nebula pulsar which has pulse period 33.2 ms and which is the energy source for the nebula. Jets of material originating at the pulsar are observed perpendicular to the disc of material which is illuminated by the pulsar emission. (Courtesy of the ESA, NASA, the Chandra Science Team and the Space Telescope Science Institute.)

about 10^{46} J and the time-scale for collapse of the central iron core of a massive star is only a matter of seconds. These events can be naturally associated with the violent events known as *supernovae* in which the star as a whole explodes and its envelope is ejected at high velocity. Ultimately, the ejection of the outer layers of the pre-supernova star gives rise to the formation of *supernova remnants*. Supernovae are thus extremely violent and luminous stellar explosions in which the optical luminosity of the star at maximum light can be as great as that of a small galaxy.

Five supernovae have been observed in our own Galaxy during the last millennium – SN 1006, SN 1054 which gave rise to the Crab Nebula (Fig. 13.1), SN 1181, associated with the supernova remnant 3C 58, Tycho’s supernova of 1572 and Kepler’s supernova of 1604 (Stephenson and Green, 2002). In each of these cases, when the star exploded, it became the brightest in the sky. The supernova 1006 probably reached apparent magnitude –7, about a thousand times brighter than the brightest stars. These five supernovae are all relatively nearby – more distant supernovae would have been obscured because of interstellar dust in the plane of the Galaxy. For example, the supernova which gave rise to the supernova remnant Cassiopeiae A must have exploded about 350 years ago but it was not recorded by astronomers. Presumably it was too faint to be observed with the naked eye, although its distance is only about 3.4 kpc. The most recent Galactic supernova G1.9+0.3 exploded close to the Galactic centre about 150 years ago and was identified as an expanding radio and X-ray source (Green *et al.*, 2008). The most recent bright supernova was SN 1987A which exploded in the Large Magellanic Cloud in 1987 (see Sect. 13.1.5). It reached apparent magnitude 3 and is of outstanding importance for understanding supernovae and the late stages of stellar evolution.

Table 13.1 Supernovae Types I and II.

Type	Characteristics
Type I – absence of hydrogen lines in optical spectrum	
Type Ia	Absence of hydrogen lines in spectrum; singly ionised silicon Si II at 615.0 nm observed near peak light.
Type Ib	Neutral helium (He I) line at 587.6 nm observed but no strong silicon absorption feature at 615.0 nm.
Type Ic	Helium lines are weak or absent; no strong silicon absorption feature 615.0 nm.
Type II – hydrogen lines present in optical spectrum	
Type IIP	Reaches a ‘plateau’ in its light curve.
Type IIL	Displays a linear decrease in its light curve
Type IIn	These supernovae contain relatively narrow features compared with the usual broad emission lines of Type II supernovae.
Type IIb	These supernovae have spectra similar to Type II at early times but to Type Ib/c at later times.

Supernovae are classified into two basic types, Type I and Type II, the key distinction being the presence or absence of the Balmer series of hydrogen in their optical spectra at maximum light. Within each type, various subtypes have been defined on the basis of other spectral features and differences in their light-curves – more details of the classification criteria are given in Table 13.1. The differences between the two types can be naturally explained if the Type II explosions occur in progenitor stars which have hydrogen envelopes, whereas the Type I supernovae occur in objects which have lost these envelopes, either because of strong mass-loss from their surface layers, or because they involve the explosion of white dwarfs which lost their hydrogen envelopes when they were formed. Type Ia supernovae are found in all types of galaxy with no preference for star-forming regions, indicating that they are associated with old or intermediate-age stellar populations. In contrast, all the other types are found in the vicinity of star-forming regions. A compelling case can be made that the Type Ia supernovae are associated with thermonuclear explosions of accreting white dwarf stars, whereas all the others are associated with the core collapse of massive stars which have lost their outer layers.

13.1.2 Type Ia supernovae

The Type Ia supernova form a particularly important subgroup since have remarkably standard properties. Excellent summaries of the vast literature on the observation and theory of these objects is provided by Leibundgut (2000) and by Hillebrandt and Niemeyer (2000). Their light curves, meaning the variation of their luminosities with time, are all remarkably similar (Fig. 13.2a). This similarity becomes even more impressive if the correlation between maximum luminosity and the width of the light curve about maximum luminosity is taken into account (Phillips, 1993). Even before this correlation is taken into account,

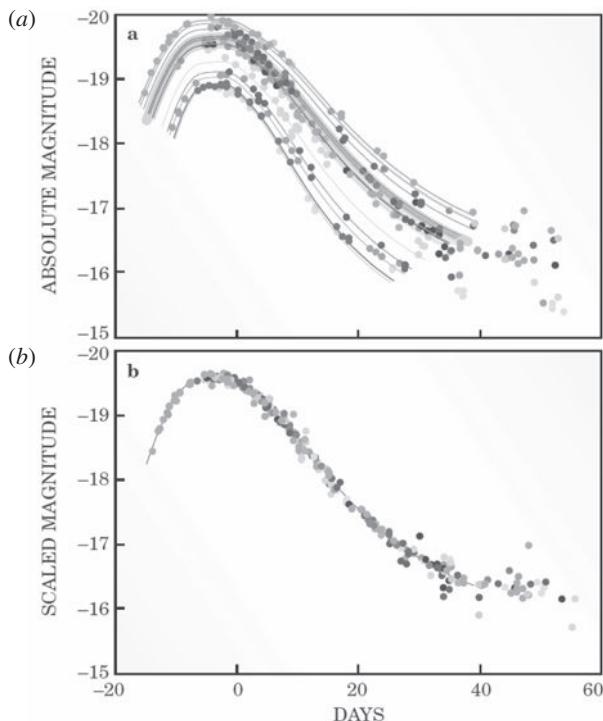


Fig. 13.2 (a) The light curves of a number of Type Ia supernovae, illustrating the luminosity–width relation. (b) Once account is taken of the luminosity–width relation, the light curves of Type Ia supernovae are remarkably similar and so can be used as distance indicators which can be observed to redshifts greater than one (2003).

the dispersion in absolute magnitudes at maximum is less than 0.5 magnitudes. Once this empirical correlation is included, the light curves lie on top of each other (Fig. 13.2b). The Type Ia supernovae are the most luminous supernovae known, their typical absolute B magnitudes being $M_B = -19.5 \pm 0.1$. Thus, the light curves and the width-luminosity relation enable the absolute magnitudes of very distant supernovae to be determined rather precisely and this has proved to be one of the most important means of determining the redshift-distance relation out to redshifts of one and greater – the resulting values of the cosmological parameters Ω_0 and Ω_Λ are in excellent agreement with many independent estimates of these parameters. These observations provide compelling evidence that our Universe is accelerating and is dominated dynamically by *dark energy* with a negative pressure equation of state $p = w\varrho c^2$, where $w \approx -1$.¹

Type Ia supernovae are quite rare events. The usual way of expressing their frequency of occurrence is in terms of *supernova units* (SNu), the number of events per century for a galaxy of luminosity $10^{10} L_\odot(B)$. In these units, the frequency is about 0.2 per century, or one every 500–600 years for a galaxy of luminosity $10^{10} L_\odot(B)$. This is significantly less

¹ For many more details of these topics, see my book *Galaxy Formation* (Longair, 2008).

than the rate for supernovae in general which is about 0.6 per century in the above units. In surveys of extragalactic supernovae, however, the Type Ia and other classes of supernovae are observed with roughly the same frequency because the Type Ia events are typically about two magnitudes more luminous than the others and so can be observed within a larger volume of space. Intriguingly, it has been possible to identify Tycho's supernova remnant of 1572 as originating from a Type Ia supernova by the light-echo technique. Rest and his colleagues found evidence for light echoes from dust clouds in the general direction of Tycho's remnant, the motion of the light echoes corresponding to vectors which converged at the supernova remnant (Rest *et al.*, 2008). Optical spectroscopy by Krause and his colleagues with the 8.2 metre Subaru Telescope showed that the spectrum of the light echo was identical to the spectrum of a Type Ia supernova at maximum light (Krause, 2008b). Thus, at least one of the supernovae observed in the last millennium was of Type Ia.

The consensus of opinion is that Type Ia supernovae are associated with the nuclear explosion of carbon–oxygen white dwarfs with masses close to the Chandrasekhar mass of $1.4 M_{\odot}$, the critical mass above which they are gravitationally unstable (Sect. 13.2.2). If white dwarfs were driven over the critical mass, say, by the accretion of mass from a binary companion, collapse to a neutron star must take place. Computations of the evolution of accreting white dwarfs indicate, however, that there are circumstances under which, before collapse takes place, the stars can be disrupted by the thermonuclear energy release associated with the fusion reactions of carbon and oxygen. Support for this picture is provided by the spectroscopic observation of intermediate mass elements such as silicon, calcium, magnesium, sulphur and oxygen in the spectra of Type Ia supernovae at maximum light. The evolution of high mass stars and the formation of carbon–oxygen cores were described in Sect. 2.7.2. In addition, the nuclear reactions involved in carbon and oxygen burning were outlined, indicating how elements up to the iron peak are synthesised. The end point of the thermonuclear reactions is the formation of ^{56}Ni which undergoes successive electron-capture (ec) and β^+ decays to form ^{56}Co and then to ^{56}Fe :



The first reaction has a half-life of only 6.1 days while the second has a half-life of 77.1 days. 1.72 MeV of energy is liberated in the decay of each ^{56}Ni nucleus in the form of γ -rays, while the average γ -ray energy released in each decay of the ^{56}Co nucleus is 3.5 MeV. It is therefore possible to work out the amount of ^{56}Ni produced in the supernova explosion directly from the bolometric luminosity of the supernova. The ratio of abundances of ^{56}Ni and ^{56}Co to iron should decrease as the parent nuclei decay.

This picture can naturally account for the form of the Type Ia supernova light curves. The rise to maximum light is rapid, about half a magnitude per day; the maximum of the light curve can be approximated by a Gaussian function, as can be appreciated from Fig. 13.2a. The colours also evolve very rapidly about maximum, from blue, $(B - V) \approx -0.1$, at 10 days before maximum, to red, $(B - V) \approx 1.1$, 30 days after maximum. After about 50 days, the luminosity decreases exponentially, the bolometric luminosity decreasing on average about 0.025 magnitudes per day (Leibundgut, 2000). The luminosity at maximum

is naturally explained as the energy deposited into the expanding envelope from the decay of the ^{56}Ni nuclei. The later exponential decay can be associated with continuing energy release associated with the decay of ^{56}Co nuclei. From those Type Ia supernovae which have well determined bolometric luminosities, synthesised masses of ^{56}Ni nuclei of between 0.1 and $1 M_{\odot}$ have been determined. These events are therefore among the most important sources of iron nuclei through the decay chain (13.1). Because Type Ia supernovae are associated with the accretion onto white dwarfs, this source of enrichment of the interstellar media of galaxies proceeds over cosmological time-scales.

The circumstantial evidence for the accreting white dwarf picture is compelling and was summarised long ago by Woosley and Weaver (1986), but it has proved very much more difficult to understand the physics of the explosion mechanism. In addition, the problems of understanding the subsequent radiative transfer though the expanding envelope are highly non-trivial. These issues are reviewed in some detail by Hillebrandt and Niemeyer (2000). The problem of the explosion mechanism is to discover processes which do not result in the formation of neutron stars and which can synthesise the heavy elements in their observed abundances. The limit of stability for white dwarfs as described by the Chandrasekhar mass, $M \approx 1.4 M_{\odot}$, provides an attractive explanation for the uniformity of Type Ia supernovae. A binary companion provides a source of mass to be accreted onto the white dwarf, but the process of accretion has to be fine-tuned or else different types of source would be created. For example, if the binary companion were a main sequence star, mass transfer onto the white dwarf would lead to the steady burning of hydrogen or helium in the surface layers – such systems are identified with *cataclysmic variables* and *novae*. If the companion were a giant star, the result would be a *symbiotic star*. If the companion were a white dwarf, the system would inspiral because of energy loss by gravitational radiation and the coalescence of the two white dwarfs could give rise to a Type Ia supernova. The key point is that the progenitor of the Type Ia supernova should increase in mass towards the Chandrasekhar limit, if the favoured explosion mechanism is to be effective.

The physics of the explosion itself is a major challenge, many of the difficult issues being carefully described in the review by Hillebrandt and Niemeyer (2000). As they express it, the carbon and oxygen nuclear burning rates are very sensitive to temperature, $\dot{S} \propto T^{12}$, and so nuclear burning takes place in thin layers which propagate either conductively as subsonic *deflagrations*, or flames, or by shock compression as supersonic *detonations*. Support for the former picture is provided by the finding that, unlike the detonation picture which converts most of the carbon and oxygen into iron, the deflagration model can reproduce the observed spectra of Type Ia supernovae at maximum light. There are, however, complex issues associated with the stability of the nuclear burning layers. These need to be studied by two-dimensional and three-dimensional numerical simulations. In particular, in the deflagration model in which the motions are subsonic, turbulence may develop as a result of Rayleigh–Taylor instabilities and associated secondary instabilities. Because the nuclear reactions take place in thin layers, turbulence has the effect of increasing the surface area over which burning can take place and so of enhancing the overall rate of energy generation. Hillebrandt and Niemeyer discuss the merits of many variants of these explosion mechanisms – prompt detonation, pure turbulent deflagration, delayed detonation pulsational delayed detonation, and so on. These are important areas of current research. In the deflagration process, the

Table 13.2 Evolution of a $15M_{\odot}$ star. Most of the table is from the paper by Woosley and Janka (2005), but the specific nuclear reactions are from the review by Arnett (2004).

Stage	Time Scale	Reaction	Ash or product	Temperature (10^9 K)	Density (gm cm^{-3})	Luminosity (solar units)	Neutrino losses (solar units)
Hydrogen	11 My	pp	He	0.035	5.8	28,000	1800
		CNO	He, N, Na				
Helium	2.0 My	$3\alpha \rightarrow ^{12}\text{C}$	C	0.18	1390	44,000	1900
		$^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$	O				
Carbon	2000 y	$^{12}\text{C} + ^{12}\text{C}$	Ne, Na	0.81	2.8×10^5	72,000	3.7×10^5
			Mg, Al				
Neon	0.7 y	$^{20}\text{Ne}(\gamma, \alpha)^{16}\text{O}$	O, Mg, Al	1.6	1.2×10^7	75,000	1.4×10^8
Oxygen	2.6 y	$^{16}\text{O} + ^{16}\text{O}$	Si, S, Ar, Ca	1.9	8.8×10^6	75,000	9.1×10^8
Silicon	18 d	$^{28}\text{Si}(\gamma, \alpha)$	Fe, Ni, Cr, Ti...	3.3	4.8×10^7	75,000	1.3×10^{11}
Iron core collapse	1 s	Neutronisation	Neutron star	>7.1	> 7.3×10^9	75,000	> 3.6×10^{15}

whole star is disrupted before the star reaches the Chandrasekhar mass and so no neutron star is formed.

13.1.3 Core-collapse supernovae and the formation of neutron stars and black holes

All other types of supernovae are believed to be formed as a result of the core collapse of massive stars. Woosley and Janka (2005) have reviewed the physics of core collapse and shown how the process may well be involved in the formation of a wide range of high energy astrophysical events including γ -ray bursts. An outline of the evolution of massive stars was presented in Sect. 2.7.2, Fig. 2.21 illustrating the successive burning of shells of heavier and heavier elements until an iron core is formed. As Woosley and Janka express it,

Indeed, the inner parts of a massive star can be thought of as just one long contraction, beginning with the star's birth, burning hydrogen on the main sequence, and ending with the formation of a black hole or neutron star. Along the way, the contraction 'pauses', sometimes for millions of years, as nuclear fusion provides the energy necessary to replenish what the star is losing to radiation and neutrinos.

This statement is reinforced by Table 13.2, taken from their paper, which quantifies the physical conditions found at various stages in the evolution of the central core of a $15 M_{\odot}$ star. These data complement those illustrated in Fig. 2.20 for the evolution of a $5 M_{\odot}$ star. As the temperature in the core increases, the time-scale for nuclear burning decreases. In particular, after helium burning, the time-scales are drastically reduced because of the

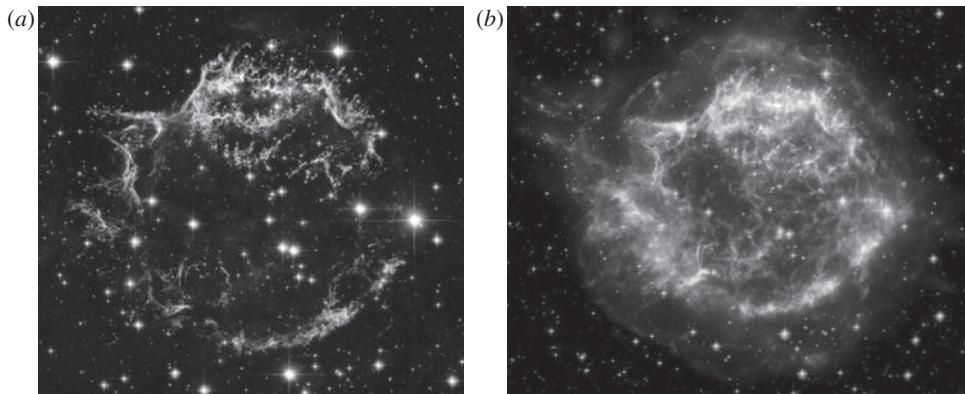


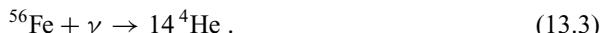
Fig. 13.3 (a) Cassiopaeia A (Cas A), as observed by the Hubble Space Telescope; (b) a composite X-ray-infrared-optical image of Cas A by the Chandra X-ray Observatory, the Spitzer Infrared Space Observatory and the Hubble Space Telescope. (Courtesy of the ESA, NASA, the Chandra Science Team and the Space Telescope Science Institute.)

enormous neutrino luminosity which greatly exceeds the optical luminosity of the star. The reason for this is that, as the central temperature approaches 10^9 K, thermal populations of electrons and positrons are created. Neutrino–antineutrino pairs are created by electron–positron annihilation and, because of their very small cross-section for interaction with matter, these escape unimpeded from the star. As the nuclear reactions proceed through the sequence of carbon, neon, oxygen and silicon burning, the neutrino losses increase dramatically, as can be seen in Table 13.2. Because nuclear burning is needed to replenish the huge neutrino energy loss, the time-scales for the later stages of the nuclear burning chain become very short indeed, silicon burning lasting only about 18 days.

Eventually, an iron core of about $1.5 M_\odot$ is formed at temperatures exceeding 7.3×10^9 K. Then further energy loss processes come into play. Energetic electrons interact with protons to form neutrons through the inverse β decay process

$$p + e^- \rightarrow n + \nu_e . \quad (13.2)$$

In addition, thermal high energy γ -rays lead to the photodisintegration of iron nuclei, which can be written schematically as



These processes lead to an enormous neutrino luminosity, more than $10^{15} L_\odot$ and an energy release of 3×10^{46} J, corresponding to about 10% of the rest-mass energy of the $1.5 M_\odot$ iron core. The removal of the pressure support from the iron core results in collapse to a proto-neutron star on a time-scale of about 1 second.

This energy release is more than enough to account for the kinetic energy of the material ejected in core-collapse supernovae, which typically is about $(1 - 2) \times 10^{44}$ J. An example of such a kinetic energy release is the supernova remnant Cassiopaeia A (Fig. 13.3) which has been identified as a Type IIb supernova by the same light-echo technique as described above for the case of Tycho’s supernova (Krause *et al.*, 2008a). Willingale and his colleagues

made observations of Cas A with the XMM-Newton X-ray Observatory and found that the total energy of the expanding X-ray nebula corresponded to 10^{44} J (Willingale *et al.*, 2003).

Originally it was believed that the collapse of the iron core to form a neutron star would result in a sudden halt to the collapse and a ‘bounce’ in which a strong shock wave would expel the outer layers of the star. It is now believed that the energy loss due to photodisintegration and neutrino emission is sufficiently great to cause the shock to stall while the proto-neutron star continues to accrete mass at a very high rate. According to Woosley and Janka, if this accretion continued even for one second, collapse to a black hole would be the outcome. This is a rerun of the old problem of the mechanism by which a core-collapse supernova explosion can be initiated. The problem is to use the 3×10^{46} J of neutrino energy to expel the outer layers of the star at high velocity.

Woosley and Janka (2005) describe two- and three-dimensional hydrodynamical computations which provide clues to the mechanisms which enable the vast neutrino luminosity of the collapse phase to be tapped. The efficiency of the absorption of neutrinos depends sensitively upon the details of the density and temperature structure surrounding the proto-neutron star and the determination of this structure is a very demanding problem in computational fluid dynamics. They discuss a promising model in which a significant fraction of the neutrinos is absorbed by the huge flux of electron–positron pairs, resulting in a ‘bubble’ of radiation, electrons and positrons, at the expanding edge of which a shock wave is formed. Their simulations show the turbulent nature of the region just outside the proto-neutron star which results in the inhomogeneous expulsion of the outer metal-rich layers of the pre-supernova star. This process can account for the fact that the different chemical species observed in the supernova remnant Cas A have different spatial distributions. These are challenging and complex computations and many key issues are currently being studied – does this process actually lead to the explosion of the star, what is the dependence of the outcome of the explosion upon the mass of the star, what happens when rotation and magnetic fields are included, and so on? With increasing computer power, many more insights into the physics of core-collapse supernova explosions are expected.

13.1.4 Steady-state hydrostatic and explosive nucleosynthesis

Supernova explosions are the origin of most of the heavy elements found in nature. There are two principal ways in which nucleosynthesis can take place in stars. The first is *steady-state hydrostatic nucleosynthesis* in which the elements are built up successively in a sequence of core and shell burning as illustrated in Figs 2.20 and 2.21 and by the entries in Table 13.2. Many of the common elements up to the iron peak are synthesised in this way and then expelled into the interstellar medium in supernova explosions. In addition, further nuclear processing can take place by the process of *explosive nucleosynthesis* which takes place during the explosion itself. Unlike steady hydrostatic nucleosynthesis, explosive nucleosynthesis results in a ‘non-equilibrium’ distribution of element abundances. Pioneering computations of explosive nucleosynthesis were carried out by Arnett, Clayton and their colleagues in the 1960s (see, for example, Arnett and Clayton, 1970). The nuclear reactions which took place during the rapid expansion of shells of carbon, oxygen and silicon from very high initial temperatures were followed and the abundances of the product nuclei

Table 13.3 Important processes in the synthesis of various isotopes^{a,b} (Woosley, 1986).

¹² C	He	³² S	O, EO	⁴⁹ Ti	ESi ^c , EHe ^c
¹³ C	H, EH	³³ S	EO	⁵⁰ Ti	nnse
¹⁴ N	H	³⁴ S	O, EO	⁵⁰ V	ENe, nnse
¹⁵ N	EH ^c	³⁶ S	EC, Ne, ENe	⁵¹ V	ESi ^c
¹⁶ O	He	³⁵ Cl	EO, EHe, ENe	⁵⁰ Cr	EO, ESi
¹⁷ O	EH, H	³⁷ Cl	EO, C, He	⁵² Cr	ESi ^c
¹⁸ O	H, EH, He	³⁶ Ar	EO, ESi	⁵³ Cr	ESi ^c
¹⁹ F	EH, He(?)	³⁸ Ar	O, EO	⁵⁴ Cr	nnse
²⁰ Ne	C	⁴⁰ Ar	?, Ne, C	⁵⁵ Mn	ESi ^c , nse ^c
²¹ Ne	C, ENe	³⁹ K	EO, EHe	⁵⁴ Fe	ESi, EO
²² Ne	He	⁴⁰ K	He, EHe, Ne, ENe	⁵⁶ Fe	ESi ^c , nse, α nse ^c
²² Na	EH, ENe	⁴¹ K	EO ^c	⁵⁷ Fe	nse ^c , ESi ^c , α nse ^c
²³ Na	C, Ne, ENe	⁴⁰ Ca	EO, ESi	⁵⁸ Fe	He, nnse, C, ENe
²⁴ Mg	Ne, ENe	⁴² Ca	EO, O	⁵⁹ Co	α nse ^c , C
²⁵ Mg	Ne, ENe, C	⁴³ Ca	EHe, C	⁵⁸ Ni	α nse, ESi
²⁶ Mg	Ne, ENe, C	⁴⁴ Ca	EHe	⁶⁰ Ni	α nse ^c
²⁶ Al	ENe, EH	⁴⁶ Ca	EC, C, Ne, ENe	⁶¹ Ni	α nse ^c , ENe, C, EHe ^c
²⁷ Al	Ne, ENe	⁴⁸ Ca	nnse	⁶² Ni	α nse ^c , ENe, O
²⁸ Si	O, EO	⁴⁵ Sc	EHe, Ne, ENe	⁶⁴ Ni	ENe
²⁹ Si	Ne, ENe, EC	⁴⁶ Ti	EO	⁶³ Cu	ENe, C
³⁰ Si	Ne, ENe, EO	⁴⁷ Ti	EHe ^c	⁶⁵ Cu	ENe
³¹ P	Ne, ENe	⁴⁸ Ti	ESi ^c	⁶⁴ Zn	EHe ^c , α nse ^c

^a The most important process is listed first and additional (secondary) contributions follow.

^b The coding of the different nuclear reactions is as follows:

H = hydrogen burning

EH = explosive hydrogen burning, novae

He = hydrostatic helium burning

EHe = explosive helium burning

C = hydrostatic carbon burning

(esp. Type I supernovae)

Ne = hydrostatic neon burning

EC = explosive carbon burning

O = hydrostatic oxygen burning

ENe = explosive neon burning

Si = hydrostatic silicon burning

EO = explosive oxygen burning

nse = nuclear statistical equilibrium (NSE)

ESi = explosive silicon burning

α nse = α -rich freeze out of NSE

nnse = neutron-rich NSE

^c Radioactive progenitor.

compared with the observed cosmic abundances of the elements up to the iron peak. The results of these endeavours were summarised by Woosley (1986) who carried out computations for a variety of different astrophysical circumstances, for example, in low and high mass stars, in novae and in Types I and II supernovae. He provided a helpful summary of the likely processes of formation of many of the isotopes in the periodic table (Table 13.3). The code at the bottom of the table lists the various processes of nucleosynthesis, the key distinction being between those with the prefix ‘E’ meaning ‘explosive nucleosynthesis’ and those without an ‘E’ indicating ‘hydrostatic nucleosynthesis’. Many of the most

abundant elements are synthesised by steady-state hydrostatic processes, for example, the CNO cycle synthesising ^{12}C and ^{16}O , carbon burning producing ^{20}Ne and oxygen burning producing ^{28}Si and ^{32}S . On the other hand, the processes responsible for creating many of the other isotopes involve explosive nucleosynthesis, for example, most of the heavy elements between sulphur (S) and iron (Fe). Important radioactive species such as ^{26}Al are attributed to explosive nuclear burning.

A second important aspect of explosive nucleosynthesis is the r-process which was discussed briefly in Sect. 2.7.2. The process involves creating conditions in which elements of the iron group are irradiated by neutrons which are successively added to these nuclei before they undergo β decays. As a result the neutron excess found in heavy elements beyond the iron group can be synthesised, but it requires an environment in which a large flux of neutrons is created. These conditions are believed to occur within a few seconds of the collapse of the iron core to a proto-neutron star. The favoured picture is clearly described by Woosley and Janka (2005). A huge flux of neutrinos and antineutrinos lasting only about 10 seconds is created during the formation of the neutron star. These interact with the electron–positron pairs and the unbound neutrons and protons in the atmosphere of the neutron star. The outer layers of the neutron star are neutron-rich and so the antineutrinos in the cooling wind are more abundant than the neutrinos. In the resulting interactions with neutrons and protons in the atmosphere of the neutron star, a neutron excess is created. As the outflow cools, the protons and neutrons combine to create α -particles, which in turn create nuclei up to the iron group. In the neutron-rich environment, these ‘seed’ iron-group nuclei are then converted into heavy elements beyond the iron peak by the r-process. According to Woosley and Janka, about $10^{-5} M_{\odot}$ of material of the wind is ejected, about 10–20% consisting of r-process elements which would be enough to account for their observed abundances.

One of the attractive features of this version of the r-process is that it can account rather naturally for the observation that despite the fact that some stars are very metal-poor, the relative abundances of the heavy elements beyond the iron peak to lower mass elements are remarkably constant. According to the picture described above, the formation of the heavy elements does not depend upon the pre-existing abundance of iron-peak elements, but rather may be thought of as a ‘primary’ process of heavy element production. The heavy elements are synthesised directly in the extreme conditions of the atmosphere of a proto-neutron star and all trace of the material from which the neutron star formed has been eliminated. Therefore, the heavy element abundances do not depend upon the past nucleosynthetic history of the stellar material.

13.1.5 The supernova SN 1987A

One of the most exciting and important astronomical events of recent years was the explosion of a supernova in one of the dwarf companion galaxies of our own Galaxy, the Large Magellanic Cloud. This supernova, known as SN 1987A, was first observed on 24 February 1987 and reached about third visual magnitude by mid-May 1987 (Fig. 13.4). It is classified as a peculiar Type IIP supernova in that the light curve showed a much more gradual increase to maximum light than is typical of Type II supernovae. After 80 days it reached

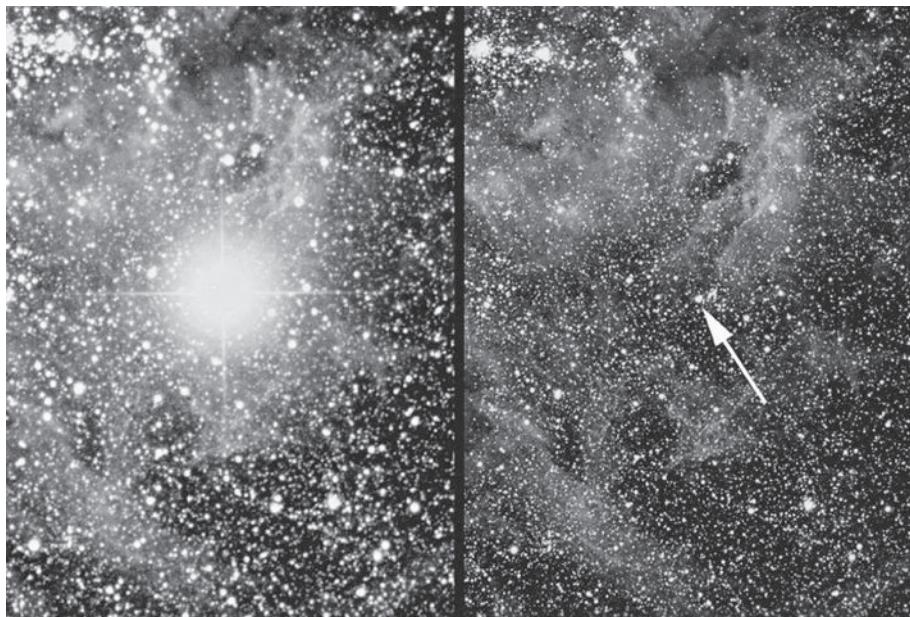


Fig. 13.4 The field of the supernova 1987A before (right) and after (left) the supernova explosion which was first observed on 24 February 1987. (Courtesy of David Malin and the Anglo-Australian Observatory.)

maximum light and its bolometric luminosity then remained roughly constant at magnitude 4 for about 2 months during which time its surface temperature declined rapidly. It was subluminous as compared with a typical Type II supernova.

The supernova coincided precisely with the position of the massive early-type B3 supergiant star Sanduleak –69 202, which disappeared following the supernova explosion. The fact that the progenitor was a highly luminous blue star was a surprise because it was expected that the supernova would have marked the end point of evolution of a red supergiant star. A clue to the evolution of the progenitor was provided by the observation of dense gas shells about the supernova. The progenitor probably did evolve to become a red giant but strong mass-loss blew off the outer layers resulting in a blue rather than a red supergiant star. The early phases of development of the light curve suggested a smaller envelope than is usual for the B-star and a lower abundance of heavy elements than the standard cosmic abundances, roughly one third of the solar value. This last result is consistent with the general trend of the heavy element abundances of stars in the Large Magellanic Cloud. The progenitor star must have been massive, $M \approx 20 M_{\odot}$, consistent with the mass of the B-star Sanduleak –69 202. Stellar evolution models have been developed in which the progenitor first became a red giant and then, because of strong mass loss, moved to the blue region of the H-R diagram for 10^4 years before exploding as a supernova.

One of the pieces of great good fortune was that, at the time of the explosion, neutrino detectors were in operation at the Kamiokande experiment in Japan and at the Irvine–Michigan–Brookhaven (IMB) experiment located in an Ohio salt-mine in the USA. Both experiments were designed for an entirely different purpose, namely the search for evidence

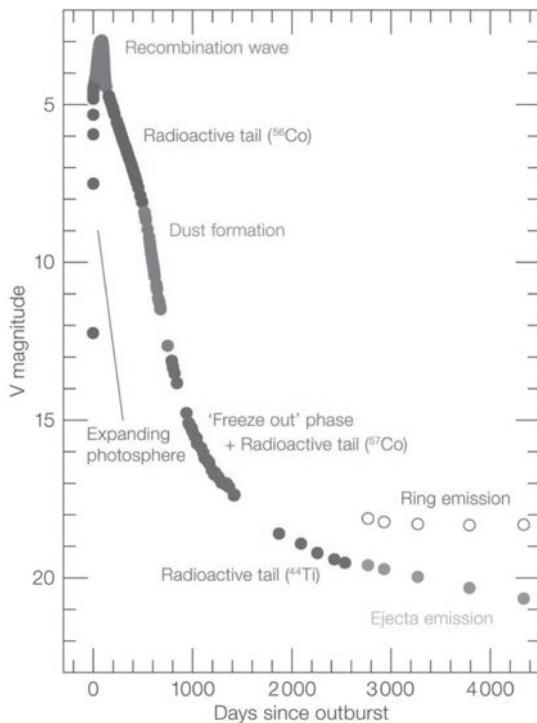


Fig. 13.5 The light curve of the supernova 1987A over a 20 year time-period. Characteristic phases in the evolution of the supernova are indicated on the diagram. (Courtesy of the European Southern Observatory.)

of proton decay, but the signature of the arrival of a burst of neutrinos was convincingly demonstrated in both experiments. Only 20 neutrinos with energies in the range 6–39 MeV were detected (12 at Kamiokande and 8 at IMB) but they arrived almost simultaneously at the two detectors, the duration of the pulse being about 12 seconds (Bahcall 1989). The neutrino energy liberated by the supernova was of the same order as that expected from the formation of a neutron star, $E \approx 10^{46}$ J. In so far as the neutrino energy spectrum could be determined from the small number of neutrinos detected, it was consistent with the extremely high temperature, $T \sim 7 \times 10^9$ K, expected when a neutron star forms. The time-scale of 12 seconds is consistent with what would be expected during the core-collapse phase of a Type II supernova. This observation, coupled with the measured energies of the neutrinos, enable limits to be set to the rest mass of the neutrino of $m_{\nu_e} \leq 20$ eV.

What makes the identification of the neutrino pulse with the supernova wholly convincing is the fact that the supernova was only observed optically some hours after the neutrino pulse. The neutrinos escape more or less directly from the centre of the collapse of the progenitor whereas the optical light has to diffuse out through the expanding supernova envelope. These observations provide strong observational support for the essential correctness of our understanding of the late stages of stellar evolution.

The light curve of the supernova has now been followed for over 20 years from the initial explosion (Figure 13.5). After the initial outburst, the luminosity decayed exponentially

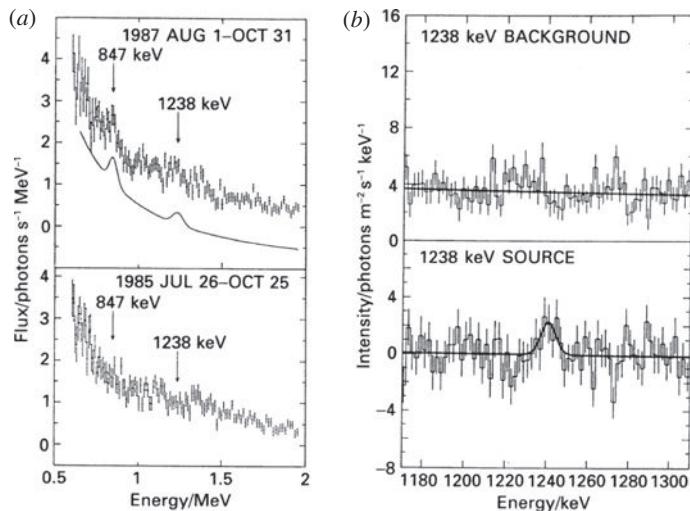
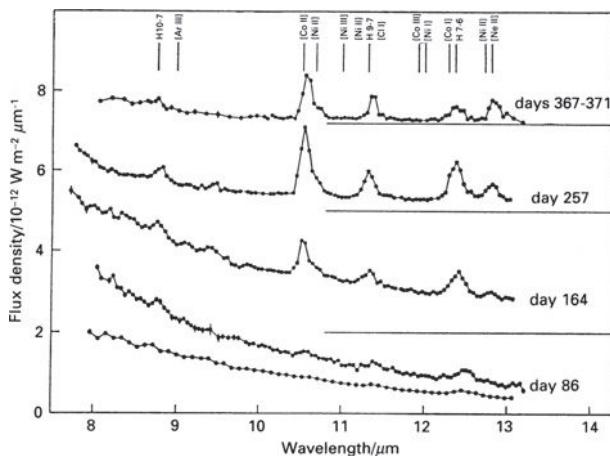


Fig. 13.6 Observations of the γ -ray lines of ^{56}Co from the supernova SN 1987A. (a) The background-subtracted spectrum obtained by the Gamma-Ray Spectrometer on the Solar Maximum Mission. The expected profiles of the two ^{56}Co lines plus a power-law continuum are shown as a solid line. The equivalent spectrum obtained in 1985 before the explosion of the supernova is also shown. The presence of an excess at the expected positions of both lines is apparent (Matz *et al.*, 1988). (b) Balloon observations of the 1238 keV line of ^{56}Co made by the Jet Propulsion Laboratory group (Mahoney *et al.*, 1988).

with a half-life of about 77 days until roughly 800 days after the explosion. This decay is convincingly associated with the energy release associated with the decay of ^{56}Co nuclei formed by the decay of ^{56}Ni through the decay chain (13.1) which has a half-life of 77.1 days. The mass of ^{56}Ni synthesised in the supernova explosion was inferred to be about $0.07 M_{\odot}$.

As soon as the supernova exploded, strenuous efforts were made to detect γ -ray lines of ^{56}Co from space missions and from dedicated balloon flights once the envelope of the supernova became transparent to γ -rays, about 6 months to a year after the explosion. Many observations were made of the 1238 and 847 keV lines of ^{56}Co and, although the signal-to-noise ratio is not large, the evidence for their existence is convincing. Figure 13.5 shows observations made by the Solar Maximum Mission and balloon observations carried out by the Jet Propulsion Laboratory group in 1987. Additional evidence for the presence of substantial quantities of cobalt and nickel in the supernova is provided by infrared spectroscopic observations in the 7–13 μm waveband in which the forbidden lines of [Co I] and [Ni II] have been observed (Fig. 13.6). Analyses of these spectra indicate that the abundance of cobalt decreases with time when the envelope of the supernova becomes optically thin as expected.

The light curve of SN 1987A in the V waveband decreased more rapidly after about 500–600 days but, at that same time, the far-infrared flux increased so that the total luminosity continued to decrease exponentially. In addition, observations of the near-infrared lines of iron indicated that less than $0.075 M_{\odot}$ of iron was present. At about the same time, the

**Fig. 13.7**

The development of the 8–13 μm spectrum of supernova SN 1987A during its first year. The positions of fine structure and hydrogenic lines are shown. The presence of strong lines of cobalt and nickel can be seen (Aitken *et al.*, 1988).

emission lines showed absorption of the redshifted gas. These observations are consistent with the formation of dust within the supernova ejecta after about 500 days. Eventually the dust became optically thin and then the light curve decreased less rapidly. The natural interpretation of this phenomenon is that a longer lived radioactive nuclide had taken over from ^{56}Co , the expected candidate being ^{57}Co which has a half-life of 1.1 years. Eventually, this energy source was replaced by the even longer-lived radionuclide ^{44}Ti which has a mean lifetime of 68 years. These successive radioactive energy sources are indicated in Fig. 13.5. The totality of these observations provides unambiguous confirmation of the radioactive origin of the supernova light curve and for the formation of iron-peak elements in supernova explosions.

SN 1987A provided information about the pre-supernova phase and the surrounding interstellar medium. The supernova outburst illuminated the material ejected in previous mass-loss events, particularly from the period of strong mass-loss during the red giant phase about 10^4 years before the supernova explosion. One of the most unexpected discoveries was the observation by the Hubble Space Telescope of rings of emission in the forbidden line of doubly ionised oxygen [O III] about the supernova (Fig. 13.8a). This ring was excited by the initial outburst of ultraviolet radiation from the supernova. The ultraviolet spectrum of the supernova was regularly monitored by the International Ultraviolet Explorer (IUE) and, when the burst of ultraviolet radiation encountered the ring, forbidden ultraviolet emission lines were observed which increased to a maximum intensity after a certain time. From these data, Panagia and his colleagues found the diameter of the ring to be $(1.27 \pm 0.07) \times 10^{16}$ m. Combining this dimension with the observed angular diameter of the ring, 1.66 ± 0.03 arcsec, a distance of 51 ± 3 kpc is obtained (Panagia *et al.*, 1991). This is a remarkably accurate distance for the Large Magellanic Cloud and in excellent agreement with independent estimates. In particular, the distance has also been estimated using the Baade–Wesselink technique applied to the expanding photosphere of the supernova (see

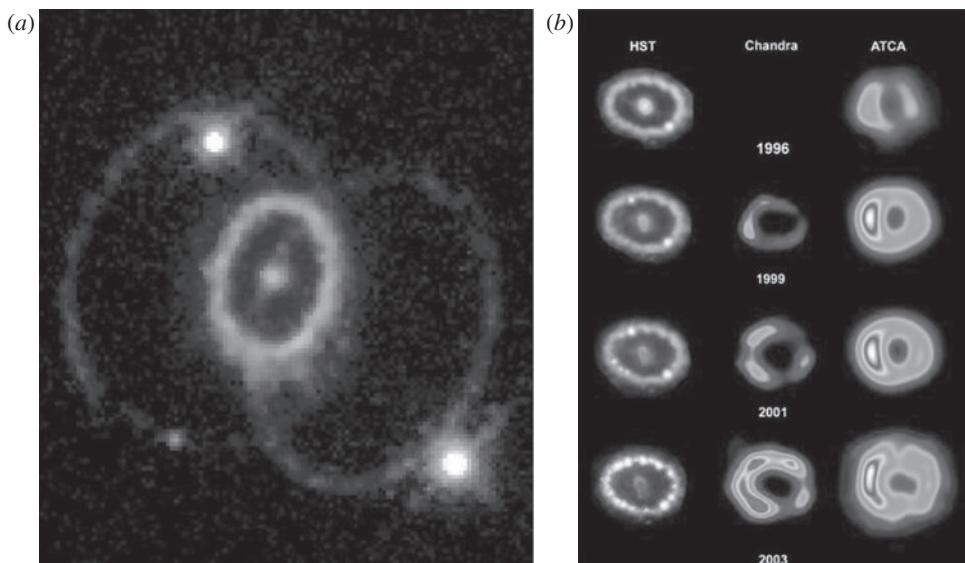
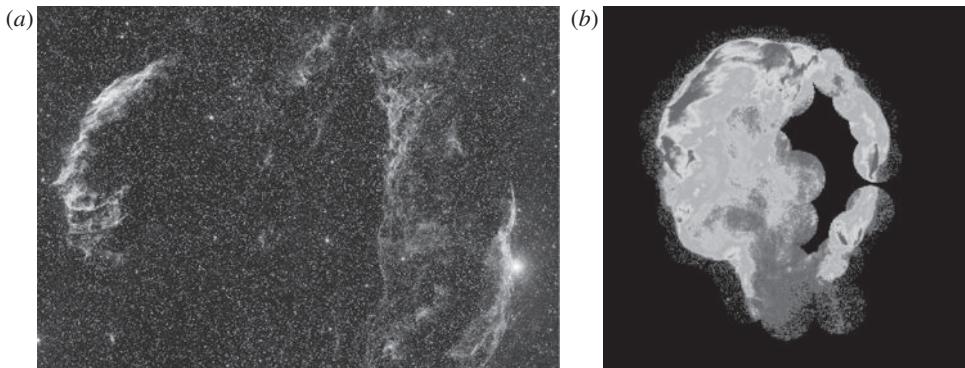


Fig. 13.8 (a) A Hubble Space Telescope image of the ring of ionised gas about the supernova SN 1987A excited by the ultraviolet radiation emitted in the initial outburst. This image was taken in the forbidden line of doubly ionised oxygen [O III] (Panagia *et al.*, 1991). (b) Composite image of the evolution of the rings about SN 1987A as observed by the Hubble Space Telescope in the optical waveband, the Chandra X-ray Observatory in the X-ray waveband and the Australia Telescope Compact Array in the radio waveband. (Courtesy of R. McCray, D. Burrows, S. Park and R. Manchester.)

Appendix A). Schmidt, Kirshner and Eastman found a distance of 49 ± 3 kpc using this technique (Schmidt *et al.*, 1992).

The ring of gas must have been created during the mass-loss phase of the progenitor star. If the outflow during the red to blue supergiant transition was in the form of a bipolar outflow, the circular ring may well have formed in the equatorial plane of the outflow, similar to what is believed to occur in the bipolar outflows about protostars and young stars. Alternatively, the ring may have formed from the debris resulting from the merger of the progenitor with a companion star. Whatever the origin of the ring, McCray and his colleagues predicted in 1994 that within the succeeding 10 years, the expanding envelope of the supernova would crash into the ring, resulting in a major increase in its luminosity (Luo *et al.*, 1994). Figure 13.8b shows that this event did indeed occur about 2002. The images show the time evolution of the structure of the ring at optical, X-ray and radio wavelengths. The Hubble Space Telescope optical image shows gas at a temperature of about 10^4 K in hot spots where the supernova blast wave has collided with the ring. The X-ray images show an expanding shell of gas at temperature about 10^8 K which is initially inside the ring. When the shell encounters the ring, it increases dramatically in X-ray luminosity. The radio emission observed by the Australia Telescope Compact Array is identified as the synchrotron radiation of electrons accelerated in the shock wave and gyrating in the magnetic field of the expanding nebula. Eventually, the blast wave will propagate beyond the ring and may well illuminate earlier events in the mass loss history of the progenitor star.

**Fig. 13.9**

(a) The Cygnus Loop (NGC 6960-92) observed in red light by the Palomar 48-inch Schmidt Telescope (photograph from the Hale Observatories). It is an old supernova remnant, probably about 50 000 years old. (b) The Cygnus Loop observed by the ROSAT X-ray Observatory. (Courtesy of the Max Planck Institut für Extraterrestrische Physik, Munich.)

13.1.6 Final things

Two other aspects of supernovae are of special importance in the context of high energy astrophysics. The first is that the kinetic energy of the matter ejected in the explosion is a powerful source of heating for the ambient interstellar gas. The shells of supernova remnants are observable until they are about 100 000 years old (Fig. 13.9a). At most stages they are observable as intense X-ray sources, in the early stages through the radiation of hot gas originating in the explosion itself and in the later stages through the heating of the ambient gas to a high temperature as the shock wave advances ahead of the shell of expelled gas (Fig. 13.9b). In both cases the emission mechanism is the bremsstrahlung of hot ionised gas. Thus, the kinetic energy of the expanding supernova remnant is a powerful heating source for the interstellar gas, regions up to about 50 pc about the site of the explosion being heated to temperatures of 10^6 K or greater (see Sect. 12.5.3).

The second important aspect is that supernovae are sources of very high energy particles. Direct evidence for this comes from the synchrotron radio emission of supernova remnants. This topic is central to the study of high energy processes in astrophysics and the physical processes involved and their many ramifications are discussed in Chap. 18.

13.2 White dwarfs, neutron stars and the Chandrasekhar limit

13.2.1 The internal structure of degenerate stars

In both white dwarfs and neutron stars, there is no internal heat source – the stars are held up by degeneracy pressure. In the centres of stars at an advanced stage in their evolution, the densities become high and the use of the pressure formulae for a classical gas is no longer appropriate. The combination of Heisenberg's uncertainty principle, $\Delta p \Delta x \approx \hbar$,

and Fermi's exclusion principle for fermions ensure that at very high densities, when the interparticle spacing becomes small, the particles of the gas must possess large momenta and cannot occupy the same quantum state. These large quantum mechanical momenta provide the pressure of the degenerate gas.

First of all, we work out the physical conditions under which degeneracy pressure is important. If the electron–proton plasma is in thermal equilibrium at temperature T , the root mean square velocity of the particles is given by $\frac{1}{2}m\langle v^2 \rangle = \frac{3}{2}kT$ and hence the typical momentum of the particles is $p = mv \approx (3mkT)^{1/2}$. According to Heisenberg's uncertainty principle, the interparticle spacing at which quantum mechanical effects become important is $\Delta x \approx \hbar/\Delta p$ and hence, setting $\Delta p = p$, the density of the plasma, which is mostly contributed by the protons, is

$$\frac{\rho \approx m_p}{(\Delta x)^3} \approx m_p \left(\frac{3mkT}{\hbar} \right)^{3/2}, \quad (13.4)$$

where m is the mass of the particle. Because the electrons are much lighter than the protons and neutrons, they become degenerate at much larger interparticle spacings and hence at lower densities than the protons and neutrons. Thus, the density at which degeneracy occurs in the non-relativistic limit is proportional to $T^{3/2}$.

We can use order-of-magnitude methods to work out the equation of state of degenerate matter in the non-relativistic regime. In general, the relation between pressure and energy density can be written $p = (\gamma - 1)\varepsilon$ where p is the pressure, ε is the energy density of the matter or radiation which provides the pressure and γ is the ratio of specific heat capacities. In the non-relativistic regime, the energy of an electron in the degenerate limit is $E = \frac{1}{2}m_e v^2 = p^2/2m_e \approx \hbar^2/2m_e a^2$, where $a \approx \Delta x$ is the interelectron spacing. Therefore, to order of magnitude, the energy density of the material is $\varepsilon \approx E/a^3 = \hbar^2/2m_e a^5$. Since the density of matter is $\rho \sim m_p/a^3$, it follows that $p \propto \rho^{5/3}$ and hence the ratio of specific heat capacities is $\gamma = 5/3$. The pressure of the gas is therefore roughly

$$p \approx \frac{\hbar^2}{3m_e a^5} \approx \frac{\hbar^2}{3m_e} \left(\frac{\rho}{m_p} \right)^{5/3}. \quad (13.5)$$

Kippenhahn and Weigert (1990) give the proper expression for the pressure of a non-relativistic degenerate gas applicable for any chemical composition of the stellar material. The material can be in any state of ionisation and so, following their conventions, the density of material ρ can be written in terms of the atomic mass unit m_u in three ways:

$$\rho = (n + n_e)\mu m_u = n\mu_0 m_u = n_e\mu_e m_u, \quad (13.6)$$

where n_e is the number density of electrons and n the number density of nuclei in the plasma; μm_u , $\mu_0 m_u$ and $\mu_e m_u$ are the average particle masses per free particle (μ), per nucleus (μ_0) and per electron (μ_e) respectively. Thus, for a fully ionised hydrogen plasma, $\mu = 0.5$, $\mu_0 = 1$ and $\mu_e = 1$; for fully ionised helium, $\mu = 1.33$, $\mu_0 = 4$ and $\mu_e = 2$; for fully ionised iron, $\mu = 56/29 \approx 2$, $\mu_0 = 28$ and $\mu_e = 2$. For mixtures and partially ionised gases, the values of the μ s differ from these cases. The equation of state for a non-relativistic

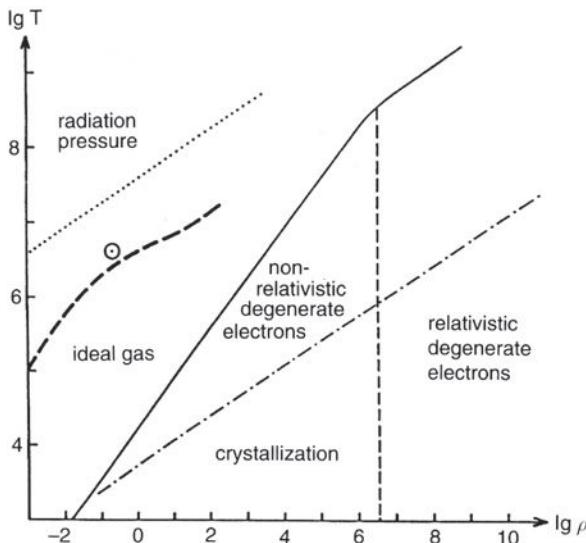


Fig. 13.10 A sketch of the density–temperature plane showing the regions in which different types of equation of state are applicable. In addition to the regions discussed in the text, the diagram also shows the regions in which radiation pressure exceeds the gas pressure and also the region in which the degenerate gas is expected to become a solid; that is, it represents the melting temperature of the stellar material. The heavy dashed line shows the location of the Sun from its core to envelope (Kippenhahn and Weigert, 1990).

degenerate gas is then

$$p = \frac{(3\pi^2)^{2/3} \hbar^2}{5} \left(\frac{\rho}{\mu_e m_u} \right)^{5/3}. \quad (13.7)$$

Equating the pressure of a degenerate electron gas in the non-relativistic limit (13.7) to the pressure of a classical gas $p = \rho k T / \mu m_u$, the critical density is

$$\frac{T}{\rho_{\text{cr}}^{2/3}} = \frac{(3\pi^2)^{2/3} \hbar^2}{5m_e m_u^{2/3} k} \frac{\mu}{\mu_e^{5/3}} \quad \text{or} \quad \rho_{\text{cr}} = 2.38 \times 10^{-5} \left(\frac{T}{\mu} \right)^{3/2} \mu_e^{5/2} \text{ kg m}^{-3}, \quad (13.8)$$

where T is the temperature in kelvins. Figure 13.10 is a plot of density against temperature showing the regions in which different forms of the equation of state apply. Also plotted is a line showing the conditions of temperature and density from the centre to the surface of the Sun. It can be seen that, in stars like the Sun, the equation of state can always be taken to be that of a classical gas. When the star moves off the main sequence, however, the central regions contract and, although there is a modest increase in temperature, the matter in the core can become degenerate and this plays a crucial role in the evolution of stars on the giant branch. Ultimately, in the white dwarfs, the densities are typically about 10^9 kg m^{-3} and so they are degenerate stars.

The next consideration is whether or not the electrons are relativistic. To order of magnitude, we can find the condition for the electrons to become relativistic by setting $\Delta p \approx m_e c$ in Heisenberg's uncertainty relation and then, by the same arguments as above, the density

is

$$\rho \sim \frac{m_p}{(\Delta x)^3} \sim m_p \left(\frac{m_e c}{\hbar} \right)^3 \sim 3 \times 10^{10} \text{ kg m}^{-3}. \quad (13.9)$$

A better calculation, with exactly the same physics but expressed in a slightly different way is to require the Fermi momentum of a degenerate Fermi gas in the zero temperature limit to be $m_e c$ (Kippenhahn and Weigert, 1990). In this case, the density at which the electrons become relativistic is

$$\rho = \frac{m_u}{3\pi^2} \left(\frac{m_e c}{\hbar} \right)^3 \mu_e = 9.74 \times 10^8 \mu_e \text{ kg m}^{-3}. \quad (13.10)$$

This limit is indicated in Fig. 13.10. In the centres of the most massive white dwarfs, the densities attain these values and so the equation of state for a relativistic degenerate electron gas has to be used. This feature determines the upper mass limit for white dwarfs and neutron stars.

We can repeat the order-of-magnitude calculation to find the pressure of a relativistic degenerate electron gas. In this case, $E \approx pc \approx \hbar c/a$ and hence $\varepsilon \approx E/a^3 \approx \hbar c/a^4$. Since $\rho \sim m_p/a^3$, $p \propto \rho^{4/3}$ and $\gamma = 4/3$. The pressure of the gas is roughly

$$p \approx \frac{\hbar c}{3a^4} \approx \frac{\hbar c}{3} \left(\frac{\rho}{m_p} \right)^{4/3}. \quad (13.11)$$

The exact result derived from the Fermi–Dirac distribution in the ground state is as follows:

$$p = \frac{(3\pi^2)^{1/3} \hbar c}{4} \left(\frac{\rho}{\mu_e m_u} \right)^{4/3}. \quad (13.12)$$

Corresponding results for degenerate neutrons are obtained if we substitute neutrons for the electrons in the above expressions and set $\mu_e = 1$. Then, the expressions for the pressure of a degenerate neutron gas in the non-relativistic and relativistic limits are

$$\text{Non-relativistic} \quad p = \frac{(3\pi^2)^{2/3} \hbar^2}{5} \left(\frac{\rho}{m_n} \right)^{5/3} \quad (13.13)$$

$$\text{Relativistic} \quad p = \frac{(3\pi^2)^{1/3} \hbar c}{4} \left(\frac{\rho}{m_n} \right)^{4/3}. \quad (13.14)$$

In both cases, the pressure is independent of the temperature and so it is remarkably straightforward to find solutions for the internal pressure and density structures inside these stars.

13.2.2 The Chandrasekhar limit for white dwarfs and neutron stars

Because the pressure is independent of the temperature for degenerate stars, we only need the first two equations of stellar structure (2.6) to carry out the analysis,

$$\frac{dp}{dr} = -\frac{GM\varrho}{r^2}; \quad \frac{dM}{dr} = 4\pi r^2 \varrho. \quad (13.15)$$

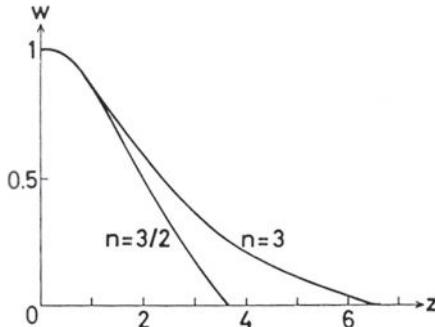


Fig. 13.11 Solutions of the Lane–Emden equation for values of the polytropic index $n = 3/2$ and 3, corresponding to ratios of specific heat capacities $\gamma = 5/3$ and $4/3$, respectively. In both cases, the density falls to zero at a finite value of z .

Eliminating M between these equations, we find a second-order differential equation relating p and ρ ,

$$\frac{d}{dr} \left(\frac{r^2}{\rho} \frac{dp}{dr} \right) + 4\pi G \rho r^2 = 0. \quad (13.16)$$

As shown in Sect. 13.2.1, the pressure p depends upon the density ρ as $p = \kappa \rho^\gamma$ with $\gamma = 5/3$ and $4/3$ in the non-relativistic and relativistic cases. Solutions of this type are known as *polytropes* and are written in terms of the *polytropic index* n such that $\gamma = 1 + (1/n)$. Thus, if $\gamma = 5/3$, $n = 3/2$ and if $\gamma = 4/3$, $n = 3$. The next step is to change variables so that (13.16) is reduced to a more manageable form. Firstly, we write the density at any point in the star in terms of the central density ρ_c as $\rho(r) = \rho_c w^n$. Then, we write the distance r from the centre in terms of the dimensionless distance z ,

$$r = az \quad \text{where} \quad a = \left[\frac{(n+1)\kappa\rho_c^{(1/n)-1}}{4\pi G} \right]^{1/2}. \quad (13.17)$$

With a little bit of algebra, (13.16) becomes

$$\frac{1}{z^2} \left[\frac{d}{dz} \left(z^2 \frac{dw}{dz} \right) \right] + w^n = 0. \quad (13.18)$$

This equation is known as the *Lane–Emden equation*.

Kippenhahn and Weigert (1990) give a very accessible account of the solutions of this equation and how these can be used to obtain insights into many different phases of stellar evolution. Analytic solutions exist only for $n = 0, 1$ and 5 . For all values of n less than 5 , the density goes to zero at some finite radius z_n which corresponds to the surface of the star at radius $R = az_n$. The solutions of the Lane–Emden equation for $\gamma = 5/3$ ($n = 3/2$) and $\gamma = 4/3$ ($n = 3$) are displayed in Fig. 13.11. The values of z at which w goes to zero are $z_{3/2} = 3.654$ and $z_3 = 6.897$ for $n = 3/2$ and 3 , respectively. From the definition of a , we find the relation between the central density of the star ρ_c and its radius R since the latter lies at a fixed value of z for a given value of n . From (13.17), it follows that

$$\rho_c \propto R^{2n/(1-n)}. \quad (13.19)$$

Thus, for $n = 3/2$, then $\rho_c \propto R^{-6}$ so that the central density increases as the radius decreases but, notice, much faster than R^{-3} .

Next, we can find the mass–radius relation by integrating the density distributions shown in Fig. 13.11 from $r = 0$ to R :

$$\begin{aligned} M &= \int_0^R 4\pi\rho r^2 dr = 4\pi\rho_c \int_0^R w^n r^2 dr , \\ &= 4\pi\rho_c a^3 \int_0^{z_n} w^n z^2 dz = 4\pi\rho_c \left(\frac{r}{z}\right)^3 \int_0^{z_n} w^n z^2 dz . \end{aligned} \quad (13.20)$$

But from (13.18), we observe that

$$\int_0^{z_n} z^2 w^n dz = - \left(z^2 \frac{dw}{dz} \right)_R . \quad (13.21)$$

Therefore, we find

$$M = 4\pi\rho_c \left(\frac{R}{z_c}\right)^3 \left[-z^2 \left(\frac{dw}{dz} \right) \right]_R . \quad (13.22)$$

For any polytrope, the expression in square brackets in (13.22) is a constant for a fixed value of n . The figures quoted by Kippenhahn and Weigert for $[-z^2(dw/dt)]_R$ are 2.71406 if $n = 3/2$ and 2.01824 if $n = 3$. Therefore, from (13.19),

$$M \propto \rho_c R^3 \propto R^{(3-n)/(1-n)} . \quad (13.23)$$

Thus, if $n = 3/2$, then $M \propto R^{-3}$, then that is, the greater the mass of the star, the smaller its radius and the greater the central density. Consequently, the central density increases rapidly with increasing mass until a critical density is reached at which the relativistic equation of state with $n = 3$ has to be used instead of $n = 3/2$. From (13.23), it follows immediately that the mass of a relativistic degenerate star is independent of its radius.

The mass of the star in the extreme relativistic case $n = 3$ is found from (13.22),

$$M = 2.018244 \times \pi \left(\frac{\kappa}{\pi G} \right)^{3/2} = \frac{(3\pi)^{3/2}}{2} \left(\frac{\hbar c}{G} \right)^{3/2} \times \frac{2.01824}{(\mu_e \mu_u)^2} = \frac{5.836}{\mu_e^2} M_\odot . \quad (13.24)$$

In white dwarf stars, the chemical abundances have evolved through to helium, carbon or oxygen and therefore we expect the limiting mass for the white dwarfs to correspond to $\mu_e = 2$. Therefore,

$$M_{\text{Ch}} = 1.46 M_\odot . \quad (13.25)$$

This is the famous *Chandrasekhar mass*.

The same analysis can be carried out for neutron stars for which $m_u = m_n$ and $\mu_e = 1$. The formal result found from (13.24) is that the upper limit is $M_{\text{ns}} \leq 5.73 M_\odot$. As discussed by Shapiro and Teukolsky (1983), this is a significant overestimate because a general relativistic treatment is needed, as well as a more realistic equation of state. The relativity parameter $2GM/Rc^2$ for neutron stars of mass $1M_\odot$ and radius $R = 10$ km is 0.15 and so the effects of general relativity cannot be neglected in the stability analysis. The effect of

general relativity is to make the effective force of gravity stronger since the gravitational potential energy contributes to the total mass. The various considerations which Shapiro and Teukolsky give in their treatment of this problem suggest that the upper limit for neutron stars must be less than about $3M_{\odot}$.

The expressions (13.24) and (13.25) are such important results that it is worthwhile giving a more physically intuitive analysis of the problem. Using the approximate methods described in Sect. 13.2.1, the total internal energy of the star in the ultra-relativistic limit is

$$U = V\varepsilon = 3Vp \approx V\hbar c \left(\frac{\rho}{m_p} \right)^{4/3}. \quad (13.26)$$

According to the *virial theorem* (Sect. 3.2.3), the total internal energy U is one-half of the total gravitational potential energy Ω_g , that is,

$$2U = |\Omega_g|; \quad 2V\hbar c \left(\frac{\rho}{m_p} \right)^{4/3} = \frac{1}{2} \frac{GM^2}{R}. \quad (13.27)$$

Now, $V \approx R^3$ and $\rho V = M$. Therefore, the left-hand side of (13.27) becomes

$$2U = \frac{2\hbar c}{R} \left(\frac{M}{m_p} \right)^{4/3}. \quad (13.28)$$

The key point is that, because we have used a relativistic equation of state, the left-hand side of equation (13.27) depends upon the radius as R^{-1} , exactly the same dependence as the gravitational potential energy. Just as in the analysis proceeding for the Lane–Emden equation, the mass of the star does not depend upon its radius. From (13.27), we find

$$M \approx \frac{1}{m_p^2} \left(\frac{\hbar c}{G} \right)^{3/2} \approx 2M_{\odot}, \quad (13.29)$$

dropping constants of order unity. Furthermore, this is an upper limit to the mass of the star because, inspection of (13.27) and (13.28) shows that $|\Omega_g| \propto M^2$ while $U \propto M^{4/3}$. Therefore, with increasing mass, the gravitational energy always exceeds twice the internal energy of the star since both energies depend upon the radius R in the same way – consequently, there is no equilibrium state. For lower mass stars, the question of whether or not the star is stable depends upon how close n is to 3 since stable degenerate stars are found for $n < 3$.

The Chandrasekhar mass depends only upon fundamental constants. One of the more intriguing ways of rewriting (13.29) is in terms of a ‘gravitational fine structure constant’, α_G . The fine structure constant in electrodynamics is $\alpha = e^2/4\pi\epsilon_0\hbar c$. The equivalent formula for gravitational forces can be found by replacing $e^2/4\pi\epsilon_0$ in the inverse square law of electrostatics $F = e^2/4\pi\epsilon_0 r^2$, by GM^2 in Newton’s law of gravity, $F = Gm_p^2/r^2$ where m_p is the mass of the proton. Thus, $\alpha_G = Gm_p^2/\hbar c$. Putting in numerical values, $\alpha^{-1} = 137.04$ and $\alpha_G = 5.6 \times 10^{-39}$, the ratio of these constants is $\alpha_G/\alpha = 2.32 \times 10^{40}$, reflecting the differing strengths of the electrostatic and gravitational forces. Therefore, the Chandrasekhar mass is roughly

$$M \approx m_p \alpha_G^{-3/2}. \quad (13.30)$$

In other words, stars are objects which typically consist of about 10^{60} protons. The calculation applies equally to white dwarfs and neutrons stars, the only difference being that the neutrons stars are very much denser than the white dwarfs.

13.3 White dwarfs

The determination of the internal structures of white dwarfs and neutron stars depends upon detailed knowledge of the equation of state of the degenerate electron and neutron gases (Shapiro and Teukolsky, 1983; Camenzind, 2007). The case of white dwarfs is the more straightforward. At the typical densities found in white dwarfs, $\rho \sim 10^9 \text{ kg m}^{-3}$, the equation of state is well understood, the main uncertainty being the chemical composition of the star. Spectroscopic observations of their surface properties show that most white dwarfs have lost their hydrogen envelopes. For stars with masses roughly that of the Sun, nuclear burning results in the formation of a degenerate helium core surrounded by a hydrogen-burning shell. Eventually, helium burning in the core is initiated in a ‘helium flash’ in which the degeneracy is relieved and helium burning proceeds to form a carbon–oxygen core. The temperature never becomes high enough to initiate carbon burning. More massive stars also form carbon–oxygen cores (Fig. 2.20) while the most massive stars can form iron cores. The fate of these stars therefore depends upon whether there is sufficient mass loss for them to end up as white dwarfs or whether they undergo catastrophic collapse to neutron stars or black holes.

The thermal energy of the star is derived from the internal energy with which the star was endowed when it was formed. The cooling times for white dwarfs are about 10^9 – 10^{10} years, very much longer than the thermal cooling time-scale for a star like the Sun because their surface areas are very much smaller than those of main sequence stars. For the white dwarf stars in star clusters, the ages of the clusters are of the same order as the cooling lifetimes of the white dwarfs. Because of their high surface temperatures and small diameters, the white dwarfs lie below the main sequence on the H-R diagram (Fig. 13.12). The solid lines represent the cooling curves for black-bodies with the masses and radii of white dwarfs, for a given mass the luminosity L being proportional to T^4 .

13.4 Neutron stars

The interiors of neutron stars consist of zones of increasing density until the material attains nuclear densities in bulk. Let us follow the physics of ultra-dense material as the density increases. With increasing density, the degenerate electron gas becomes relativistic and, when the total energy of the electron exceeds the mass difference between the neutron and the proton, $E = \gamma m_e c^2 \geq (m_n - m_p)c^2 = 1.29 \text{ MeV}$, the inverse β decay process, $p + e^- \rightarrow n + \nu_e$, can convert protons into neutrons. In a non-degenerate electron gas, the

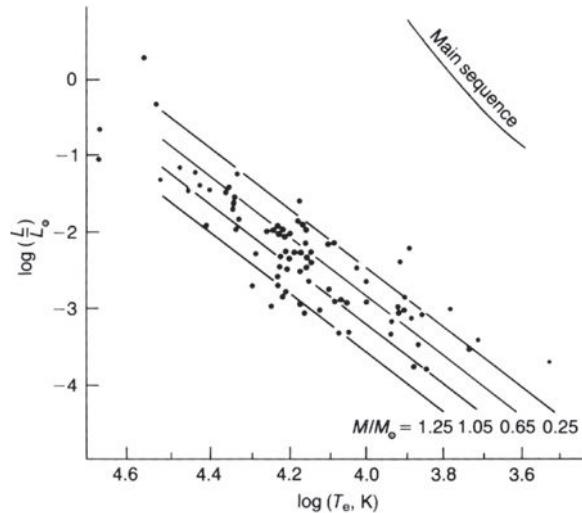


Fig. 13.12 Comparison of the theoretical Hertzsprung–Russell diagram for white dwarfs with their observed properties. The location of the cooling curve on the H–R diagram depends upon the mass of the white dwarf (Shapiro and Teukolsky, 1983).

neutrons would decay into protons and electrons with a mean lifetime of 14.8 minutes, corresponding to a half-life of 10.2 minutes, but this is not possible if the electron gas is degenerate as there are no available states for the ejected electron to occupy. This stabilisation takes place when the Fermi energy of the degenerate electron gas is greater than the kinetic energy of the emitted electrons. For a hydrogen plasma, the critical density at which stabilisation takes place can be found as follows. The total energy of the electron must be $E \geq E_{\text{tot}} = (m_n - m_p)c^2 = 1.29 \text{ MeV}$. The critical Fermi momentum p_F follows from the standard relation between total energy and momentum,

$$p_F = \gamma m_e v = \left(\frac{E_{\text{tot}}^2}{c^2} - m_e^2 c^2 \right)^{1/2}. \quad (13.31)$$

The number density of a degenerate electron gas is given by the usual formula $n_e = (8\pi/3h^3)p_F^3$, from which we find the total density $\rho = n_e m_u \mu_e$. Taking $\mu_e = 1$ for a hydrogen plasma, $\rho = 1.2 \times 10^{10} \text{ kg m}^{-3}$. This process is often referred to as *neutronisation*.

For heavier nuclei, which are expected to form the bulk of the matter in white dwarfs and proto-neutron stars, the situation is more complicated. At densities $\sim 10^{10} \text{ kg m}^{-3}$, the nuclei form a non-degenerate Coulomb lattice and the nuclei are the conventional stable elements such as carbon, oxygen and iron. As the density increases, the inverse β decay reaction favours the formation of neutron-rich nuclei. However, the energies needed to achieve this transition are greater than in the case of protons because the neutrons are degenerate within the nuclei and therefore the electron must be sufficiently energetic to exceed the Fermi energy within the nucleus. If the nuclei become too neutron-rich, however, they begin to break up and an equilibrium state is set up consisting of neutron-rich nuclei, a free neutron gas and a degenerate relativistic electron gas. This process of releasing

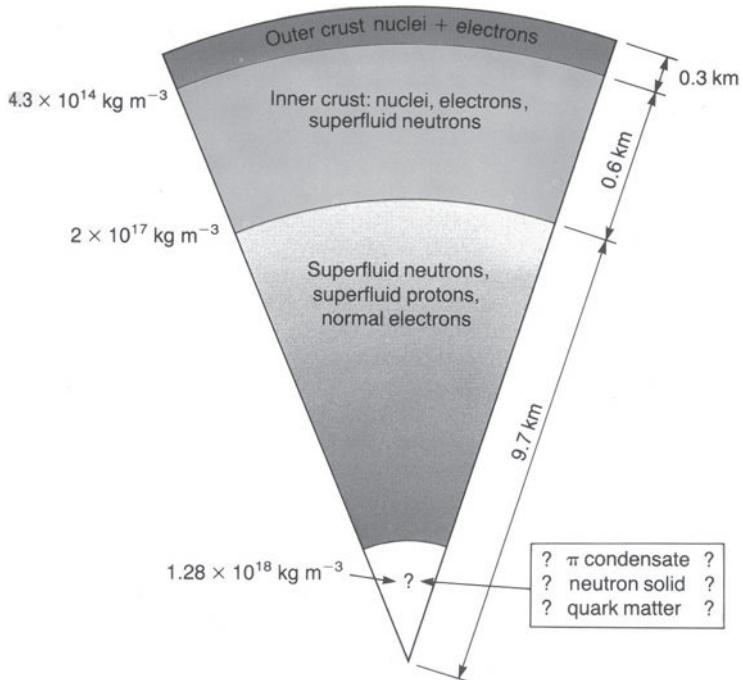


Fig. 13.13 A representative model showing the internal structure of a $1.4 M_{\odot}$ neutron star.

neutrons from the neutron-rich nuclei is referred to as *neutron drip* and sets in at a density of about $4 \times 10^{14} \text{ kg m}^{-3}$.

These processes result in profound changes in the equation of state such that stable stars cannot form until much higher central densities are attained, $\sim 10^{17} \text{ kg m}^{-3}$, at which the neutron-drip process has converted almost all of the matter into neutrons. The degeneracy pressure of the neutron gas prevents collapse under gravity and results in the formation of a *neutron star*. The underlying physics is the same as for white dwarfs, the difference being that the neutrons are about 2000 times more massive than the electrons and consequently, according to (13.4), degeneracy sets in at a correspondingly higher density. In addition, a general relativistic treatment is needed to determine the structures of the most massive neutron stars.

The internal structures of *neutron stars* are less well determined because of uncertainties in the equation of state of degenerate nuclear matter. The problems involved in determining the equation of state are elegantly presented by Shapiro and Teukolsky (1983) – a much more recent survey of the detailed physics of all classes of compact object is provided by Camenzind (2007). Figure 13.13 shows a representative example of the internal structure of a neutron star. The various zones in the model are as follows:

- (i) The *surface layers* are taken to be the regions with densities less than about 10^9 kg m^{-3} . The matter consists of atomic polymers of ^{56}Fe in the form of a close packed solid. In the presence of strong surface magnetic fields, the atoms become cylindrical.

The matter behaves like a one-dimensional solid with high conductivity parallel to the magnetic field and with essentially zero conductivity across it.

- (ii) The *outer crust* is the region with density in the range $10^9 \leq \rho \leq 4.3 \times 10^{14} \text{ kg m}^{-3}$ and consists of a solid region composed of matter similar to that found in white dwarfs, that is, heavy nuclei forming a Coulomb lattice embedded in a relativistic degenerate gas of electrons. When the energies of the electrons become large enough, inverse β decay increases the numbers of neutron-rich nuclei which would be unstable on Earth. For example, ^{62}Ni forms at a density of $3 \times 10^{11} \text{ kg m}^{-3}$, ^{80}Zn at $5 \times 10^{13} \text{ kg m}^{-3}$, ^{118}Kr at $4 \times 10^{14} \text{ kg m}^{-3}$, and so on.
- (iii) The *inner crust* has density between about 4.3×10^{14} and about $2 \times 10^{17} \text{ kg m}^{-3}$. It consists of a lattice of neutron-rich nuclei together with free degenerate neutrons and a degenerate relativistic electron gas. As the density increases, more and more of the nuclei begin to dissolve and the neutron fluid provides most of the pressure.
- (iv) The *neutron liquid* phase occurs at densities greater than about $2 \times 10^{17} \text{ kg m}^{-3}$ and consists mainly of neutrons with a small concentration of protons and electrons.
- (v) In the very centre of the neutron star, a *core region* of very high density, $\rho \geq 3 \times 10^{18} \text{ kg m}^{-3}$, may or may not exist. The existence of this phase depends upon the behaviour of matter in bulk at very high energies and densities. It is not clear if there is a phase transition to a neutron solid or to quark matter or to some other phase of matter quite distinct from the neutron liquid. Many of the models of stable neutron stars do not possess this core region but it is certainly not excluded that quite exotic forms of matter could exist in the centres of massive neutron stars. These issues and their implications for the structure and stability of neutron stars are clearly described by Camenzind (2007).

A consequence of the fact that a neutron star may be thought of as one huge nucleus containing about 10^{60} nucleons is that the inner regions are likely to be superfluid and the protons superconducting. It is interesting to contrast the physical processes in neutron stars with those in laboratory superfluids and superconductors. ^3He , for example, becomes superfluid at a low enough temperature. The ^3He atoms are fermions and so, in order to create a Bose condensation, ^3He atoms pair up with opposite spins so that they obey Bose-Einstein statistics. The physical causes of pairing are long-range attractive forces between ^3He atoms which result in the fluid being in a lower energy state if pairs of helium atoms remain correlated. If the energy difference Δ between the ‘paired-state’ and the ‘unpaired-state’ is greater than kT , where T is the temperature of the fluid, the system remains in the lower energy state, the particles forming long range pairs.

Pairing processes are also responsible for the phenomenon of superconductivity in metals. At low temperatures, almost all the electronic states up to the Fermi level of the metal are filled and the electrical conductivity is associated with the very small fraction which are close to the Fermi level. At low enough temperatures, these conduction electrons can form pairs with opposite spins due to long range attractive forces associated with interactions between the electrons and the lattice vibrations. If the energy gap Δ associated with the energy difference between the paired and unpaired states is greater than kT , the lower energy state with the electrons forming *Cooper pairs* is preferred and, since the

pairs of electrons are bosons, they prefer to occupy the same state. As Weisskopf (1981) expresses it, the Cooper pairs form a superconducting ‘frozen crust’ on top of the Fermi distribution.

There is no attractive force between free neutrons, but there is a net attractive force between the neutrons within an atomic nucleus which is mediated by bulk nuclear forces. In the central regions of a neutron star, these result in long range attractive forces between pairs of neutrons, the interaction energy being about 3 MeV. This energy is much greater than that corresponding to the typical internal temperatures of neutron stars which are probably of the order of $kT \sim 1\text{--}10$ keV. Therefore, it is likely that the neutrons in the central regions of neutron stars form pairs and are superfluid. The free neutrons can form a superfluid in the inner crust among the neutron rich nuclei (region 3). Likewise, in region 4, the liquid neutron phase, in which the nuclei have dissolved into neutrons and protons, the neutron fluid is expected to be superfluid. The protons in the quantum liquid phase (region 4) are expected to be superconducting. In all these phases, the electrons remain ‘normal’ in the sense that the interactions between them are not sufficient to produce superconductivity at these temperatures. These phenomena do not have an important influence upon the overall internal structure of the neutron star, but they have a profound impact upon its internal rotation and upon the behaviour of its internal magnetic field.

To anticipate the discussion of Sect. 13.5, the observation of polarised radio emission from radio pulsars and, in particular, the observed rotation of the plane of polarisation within the pulses, provide powerful evidence for the presence of a magnetic field in pulsars. Field strengths in the range $10^6\text{--}10^9$ T are inferred from the observed rate of deceleration of pulsars (Sect. 13.5). Further evidence for such intense magnetic fields is provided by the observation of a cyclotron radiation feature in the X-ray spectra of the X-ray pulsars such as Her X-1 (Sect. 8.2). There is no problem in accounting for the strength of such fields because the magnetic field is very strongly coupled to the ionised plasma by *magnetic flux freezing* (Sect. 11.2). When a star collapses spherically, the magnetic field strength increases as $B \propto r^{-2}$ because of conservation of magnetic flux and so, if a star like the Sun possessed a magnetic field of strength 10^{-2} T, there is no problem in accounting for a field strength of 10^8 T if the star collapsed to only 10^{-5} of its initial radius. It might be thought that the magnetic field would be expelled from the central regions of the neutron star because of the superconducting proton fluid. The presence of the normal relativistic degenerate electron gas, however, ensures that the magnetic field can exist within the central regions.

The rotation of neutron stars is responsible for the observation of their pulsed emission at radio and X-ray wavelengths, the pulses being attributed to the passage of a beam of radiation from the poles of the neutron star across the line of sight to the observer. The observed rotation periods of the neutron stars can be compared with the maximum which they could possess. A rough estimate of this may be made by assuming that the neutron star would break up centrifugally if its rotational kinetic energy were greater than half its gravitational potential energy, that is, the star would no longer satisfy the virial theorem. For a $1 M_\odot$ neutron star, the break-up rotational period is about half a millisecond. This is shorter than the observed rotation periods of all pulsars, although pulsars with periods in the range 1–10 ms, the *millisecond pulsars*, are well known objects, the shortest

period being only 1.5 ms which is within a factor of about 3 of the break-up rotational period.

Let us turn to the observational evidence for the existence of neutron stars.

13.5 The discovery of neutron stars

Neutron stars play a central role in many different contexts in high energy astrophysics. The story of neutron stars in various guises can be conveniently told in a historical sequence, emphasising the different astronomical technologies which contributed to their discovery.

13.5.1 ‘Normal’ radio pulsars

Radio pulsars came as a more or less complete surprise when they were discovered by Hewish and Bell in 1967 (Hewish *et al.*, 1968). Hewish had established that the fluctuating radio signals of compact radio sources at low radio frequencies were due to electron density fluctuations in the interplanetary medium. This provided a new method for finding compact radio sources, many of which were quasars, and also of studying the properties of the Solar Wind. The key technological development was the need to build a large enough array at low radio frequencies so that fluctuations in the flux densities of the sources could be detected on the time-scale of 0.1 second. The first sky surveys began in July 1967 and Jocelyn Bell, Hewish’s graduate student, discovered a strange source which seemed to consist entirely of scintillating radio signals (Fig. 13.14a). In November 1967, the source was observed using a receiver with a shorter time-constant and the signal was found to consist entirely of a series of pulses with a pulse period of about 1.33 s (Fig. 13.14b). The source PSR 1919+21 was the first *pulsar* to be identified and over the next few months three further examples were discovered with pulse periods in the range 0.25 to almost 3 s. This remarkable story has been described by Hewish (1986) and Bell-Burnell (1983). Authoritative surveys of the properties and physics of pulsars are provided by the books by Lyne and Graham Smith (2006) and by Lorimer and Kramer (2005).

The pulsars were soon identified with isolated, rotating, magnetised neutron stars following the proposals by Gold (1968) and Pacini (1967; 1968). The key observations were the very stable, short periods of the pulses and the observation of polarised radio emission. To account for the observation of radio pulses, the magnetic axis of the star and its rotation axis must be misaligned. The pulses are assumed to originate from beams of radio emission emitted along the magnetic axis as illustrated in Fig. 13.15. The discovery of pulsars in the Crab Nebula and the Vela supernova remnant were of special importance because they are both young pulsars with ages more or less consistent with the ages of the remnants. These observations proved conclusively that neutron stars are formed in supernova explosions. The very short period of the Crab pulsar, 33 ms, enabled other possible candidates as the parent bodies of the radio pulsars, except neutron stars, to be excluded. We will use the term ‘normal’ radio pulsars to mean radio pulsars which are isolated, rotating magnetised neutron stars with periods $P \gtrsim 30$ ms.

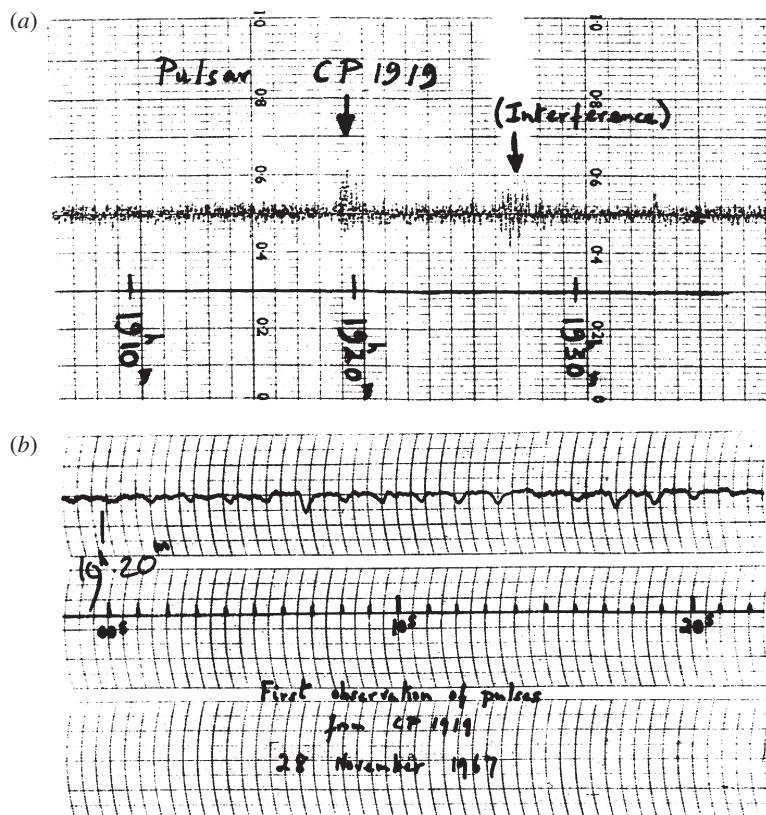
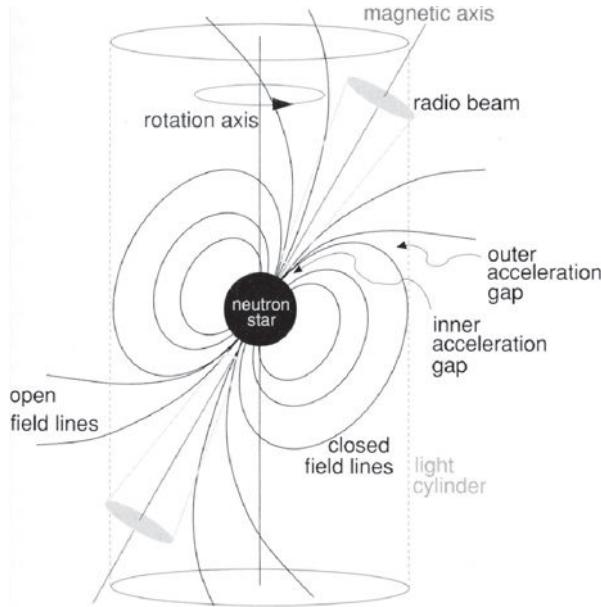


Fig. 13.14 The discovery records of the first pulsar to be discovered, PSR 1919+21. (a) The first record of the strange scintillating source labelled CP 1919. Note the subtle differences between the signal from the source and the neighbouring signal due to terrestrial interference. (b) The signals from PSR 1919+21 observed with a shorter time-constant than the discovery record, showing that the signal consists entirely of regularly spaced pulses with period 1.33 s (Hewish *et al.*, 1968; Hewish, 1986).

The pulse periods of pulsars P can be measured with very high accuracy indeed and one of the most important parameters is the rate at which the pulse period changes with time, \dot{P} . Normal radio pulsars are slowing down and the rate of loss of rotational energy can be described by a *braking index* n which is defined by $\dot{\Omega} = -\kappa \Omega^n$, where Ω is the angular frequency of rotation. The braking index provides information about the energy loss mechanism responsible for slowing down the rotation of the neutron star. Among the most important of these is magnetic braking. In order to produce pulsed radiation from the magnetic poles of the neutron star, the magnetic dipole must be oriented at an angle with respect to the rotation axis and then the magnetic dipole displays a varying dipole moment as observed at a large distance (Fig. 13.15). As a result, the pulsar loses energy by electromagnetic radiation which is extracted from the rotational energy of the neutron star. By exact analogy with the radiation of an electric dipole (Sect. 6.2.2), a magnetic dipole of

**Fig. 13.15**

A schematic model of a pulsar as a magnetised rotating neutron star in which the magnetic and rotation axes are misaligned. The radio pulses are assumed to be due to beams of radio emission from the poles of the magnetic field distribution and are associated with the passage of the beam across the line of sight to the observer (Lorimer and Kramer, 2005). Typical neutron star parameters are $M \approx 1.4 M_{\odot}$, radius ≈ 10 km, magnetic flux density $10^5 - 10^9$ T.

magnetic dipole moment p_m radiates electromagnetic radiation at a rate

$$-\frac{dE}{dt} = \frac{\mu_0 |\vec{p}_m|^2}{6\pi c^3}. \quad (13.32)$$

This expression can be simply derived by replacing the electrostatic term $|\vec{p}|^2/4\pi\epsilon_0$ in the expression (6.8) by the corresponding magnetostatic term $\mu_0 |\vec{p}_m|^2/4\pi$, where p_m is the magnetic dipole moment of the neutron star. In the case of a rotating magnetic dipole, $p_m = p_{m0} \sin \Omega t$, where Ω is the angular velocity of the neutron star and p_{m0} is the component of the magnetic dipole perpendicular to the rotation axis. Consequently,

$$-\left(\frac{dE}{dt} \right) = \frac{\mu_0 \Omega^4 p_{m0}^2}{6\pi c^3}. \quad (13.33)$$

This magnetic dipole radiation extracts rotational energy from the neutron star. If I is the moment of inertia of the neutron star,

$$-\frac{d(\frac{1}{2}I\Omega^2)}{dt} = -I\Omega \frac{d\Omega}{dt} = \frac{\mu_0 \Omega^4 p_{m0}^2}{6\pi c^3}. \quad (13.34)$$

Consequently, $d\Omega/dt \propto \Omega^3$ and so the braking index for magnetic dipole radiation is $n = 3$. The braking index n can be estimated if the second derivative of the pulsar angular frequency $\ddot{\Omega}$ can be measured. If $\dot{\Omega} = -\kappa \Omega^n$, $\ddot{\Omega} = -n\kappa \dot{\Omega}^{(n-1)}$. Dividing the latter by the

former, $n = \Omega\ddot{\Omega}/\dot{\Omega}^2$, and so

$$n = \frac{\Omega\ddot{\Omega}}{\dot{\Omega}^2} = \frac{\nu\ddot{\nu}}{\dot{\nu}^2} = 2 - \frac{P\ddot{P}}{\dot{P}^2}. \quad (13.35)$$

The age of the pulsar can be estimated if it is assumed that its deceleration can be described by a constant braking index n throughout its lifetime. Integrating $\dot{\Omega} = -\kappa\Omega^n$,

$$\frac{1}{(n-1)} \left[\Omega^{-(n-1)} - \Omega_0^{-(n-1)} \right] = \kappa\tau, \quad (13.36)$$

where τ is the age of the pulsar and Ω_0 is its initial angular velocity. If $n > 1$ and $\Omega_0 \gg \Omega$, the age of the pulsar can be estimated,

$$\tau = \frac{\Omega^{-(n-1)}}{\kappa(n-1)} = -\frac{\Omega}{(n-1)\dot{\Omega}} = \frac{P}{(n-1)\dot{P}}. \quad (13.37)$$

It is conventional to set $n = 3$ to derive the age of pulsars and so $\tau = P/(2\dot{P})$.

Braking indices n have been measured for a number of pulsars. For example, for the Crab pulsar, $n = 2.515 \pm 0.005$; for PSR B1509–58, $n = 2.837 \pm 0.001$; for PSR B0540–69, $n = 1.81 \pm 0.07$; and for PSR J1119–6127, $n = 3.0 \pm 0.1$ (Lyne and Graham-Smith, 2006). In the case of the Crab pulsar, it has been possible to measure the third derivative of the angular frequency with respect to time, $d^3\Omega/dt^3$, and it is also consistent with the value $n = 2.515$. The problem of extending this type of measurement to other radio pulsars is that glitches (Sect. 13.6) and timing noise prevent good estimates of $\ddot{\Omega}$ to be found from short data runs. Thus, although magnetic braking may be the cause of the deceleration in some cases, it cannot be the whole story. The quantity κ in the definition of the braking index, $\dot{\Omega} = -\kappa\Omega^n$, may vary if, for example, the moment of inertia of the neutron star I , the magnetic flux density B or the angle of inclination of the magnetic axis to the rotation axis α change with time. Then, it is straightforward to show that

$$n_{\text{obs}} = \frac{\nu\ddot{\nu}}{\dot{\nu}^2} = n + \frac{\dot{\kappa}}{\kappa} \frac{\nu}{\dot{\nu}}. \quad (13.38)$$

Using the relation $\tau = P/(2\dot{P})$, the typical lifetime for normal pulsars is about 10^5 – 10^8 years. The Crab Nebula pulsar has a large spin-down rate and, using the formula, $\tau = P/(2\dot{P})$, a characteristic age of $\tau = 1400$ years is found, roughly the same as the age of the Crab Nebula which was observed to explode in 1054.

It can be seen from equation (13.34) that the rate of loss of rotational energy from the neutron star can be determined directly from the slow-down rate of the pulsar. This relation can be rewritten as follows:

$$-\frac{dE_{\text{rot}}}{dt} = -I\Omega \frac{d\Omega}{dt} = -4\pi I\dot{P}P^{-3}. \quad (13.39)$$

A particularly interesting result for the Crab pulsar is that the rate at which it loses rotational energy, $dE/dt \sim 6.4 \times 10^{31}$ W, is similar to the energy requirements of the surrounding supernova remnant in non-thermal radiation and bulk kinetic energy of expansion, $dE/dt \sim 5 \times 10^{31}$ W. The origin of the continuous supply of high energy particles to the Nebula had been a major mystery prior to the discovery of the Crab pulsar because the radiation

lifetimes of the particles emitting X-ray and optical synchrotron radiation in the nebula are much less than the age of the supernova remnant. The continuous injection of energy into the nebula from the pulsar solves this problem.

If the magnetic braking mechanism is responsible for the slow-down of the neutron star, estimates can be made of the magnetic flux density at the surface of the neutron star. Approximating the magnetic field at the surface of the neutron star by a dipole field, the magnetic flux density at its surface is

$$\mathbf{B} = \frac{\mu_0 p_{m0}}{4\pi r^3} [2 \cos \theta \mathbf{i}_r + \sin \theta \mathbf{i}_\theta]. \quad (13.40)$$

Thus, at $r = R$, the surface magnetic field strength is $B_s \approx \mu_0 p_{m0}/4\pi R^3$. Substituting into (13.34), we find

$$-\frac{d\Omega}{dt} = \frac{\mu_0 \Omega^3 p_{m0}^2}{6\pi c^3 I} = \frac{\mu_0 \Omega^3}{6\pi c^3 I} \left(\frac{4\pi R^3 B_s}{\mu_0} \right)^2 = \frac{8\pi \Omega^3 R^6 B_s^2}{3\mu_0 c^3 I}. \quad (13.41)$$

For a uniform sphere rotating about its axis, $I = 2MR^2/5$, and so we find

$$B_s = -\left(\frac{3\mu_0 c^3 M \dot{\Omega}}{20\pi \Omega^3 R^4}\right)^{1/2} = \left(\frac{3\mu_0 c^3 M}{80\pi^3 R^4}\right)^{1/2} (P \dot{P})^{1/2} \approx 3 \times 10^{15} (P \dot{P})^{1/2} \text{ T}. \quad (13.42)$$

These relations can be conveniently summarised in a plot of P against \dot{P} , a $P - \dot{P}$ diagram, which can be thought of as the pulsar equivalent of the Hertzsprung-Russell diagram. Figure 13.16 was derived from a very large sample of pulsars studied by Manchester and his colleagues with the Parkes Radio Telescope (Manchester *et al.*, 2005). The large clump of pulsars with values in the range $-12 \gtrsim \log \dot{P} \gtrsim -17$ and $-1 \lesssim P \lesssim -0.5$ are the normal radio pulsars. The lines showing the ages and magnetic flux densities are derived from (13.37) and (13.42) respectively. It can be seen that the ages range from young pulsars with ages of the order of $10^3 - 10^4$ years, many of which are associated with the remnants of the supernovae in which they were formed, to old pulsars with ages up to about 10^8 years. The magnetic flux densities lie in the range $10^7 - 10^9$ T. It should be emphasised that these ages and magnetic fields are indicative values and should be considered order of magnitude estimates. Other classes of pulsar will be introduced in the course of this section. The location of pulsars on this diagram may be interpreted as an evolutionary sequence in the sense that, as normal pulsars grow old, they are spun-down by magnetic braking and consequently become less luminous according to (13.39). The absence of normal pulsars to the bottom right of the diagram can be attributed to their longer periods and to decay of their magnetic flux densities. This region of the diagram is often referred to as the ‘graveyard’ for dead pulsars and the bounding locus to the bottom right of the diagram as the ‘death line’, meaning that they are no longer observable as normal radio pulsars.

13.5.2 Neutron stars in binary systems – X-ray binaries

The next event, which was to have a profound influence upon thinking in high energy astrophysics, was the discovery of neutron stars in binary X-ray sources by the UHURU satellite in 1971. The UHURU X-ray observatory was the first satellite dedicated exclusively

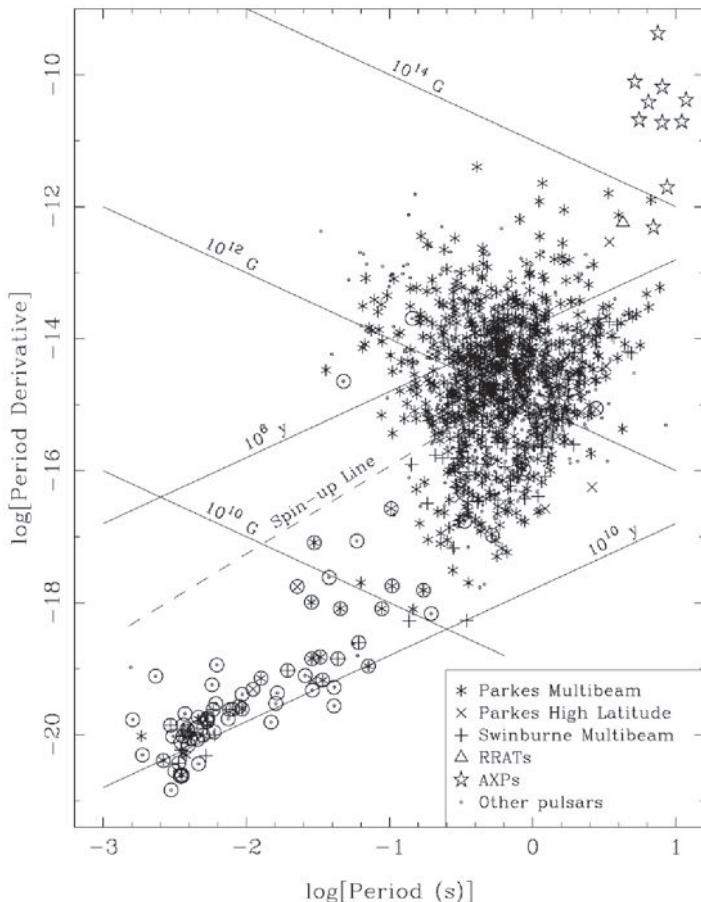
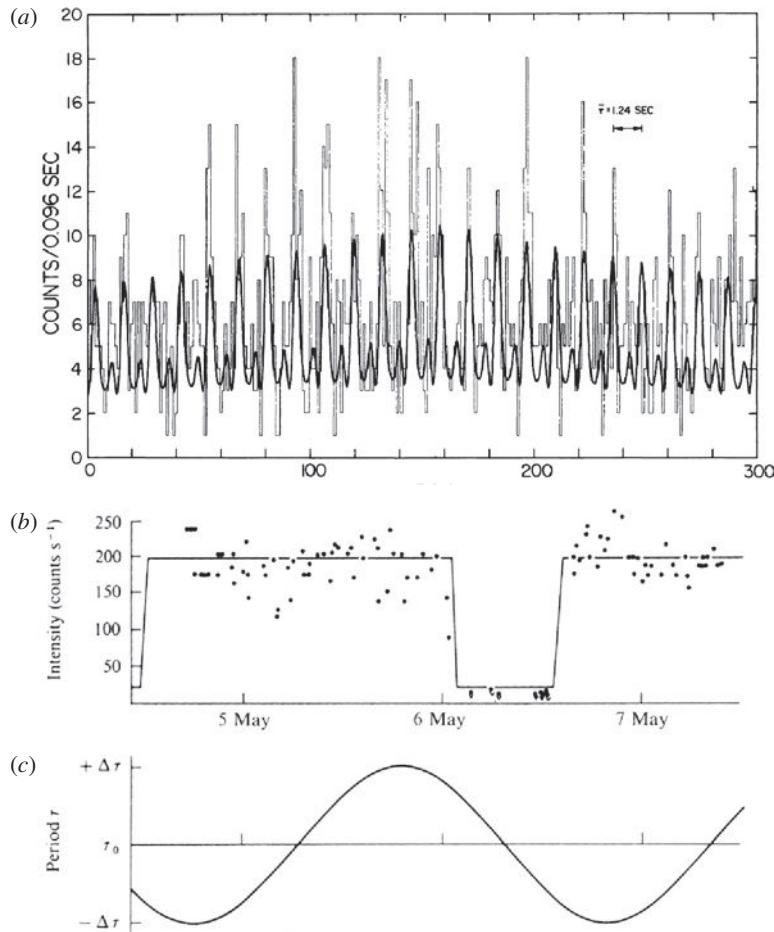


Fig. 13.16 A plot of \dot{P} versus P for pulsars, known as the $P-\dot{P}$ diagram (Manchester, 2005, from data described in Manchester *et al.*, 2005). The different symbols refer to different large pulsar surveys. The symbols enclosed in circles represent pulsars which are members of binary systems. Lines of constant age derived from the formula $\tau = P/2\dot{P}$ are shown. The magnetic flux densities are derived from (13.43), assuming the deceleration of the pulsar is due to magnetic braking. The upper limit to the spin-up periods for dead pulsars according to the models of van den Heuvel is also shown (van den Heuvel, 1987).

to X-ray astronomy and carried out the first systematic survey of the whole sky. Observations of the source Centaurus X-3 (Cen X-3) were first made in January 1971 and showed a clear periodicity with a pulse period of about 5 s, longer than that of any known radio pulsar. The pulsation period was not stable but seemed to vary with time (Giacconi *et al.*, 1971). The source was reobserved in May 1971 and it was found that the period of the X-ray pulsations varied sinusoidally with a period of 2.1 days. This suggested that the X-ray source was a member of a binary system, the change in period of the pulses being due to the Doppler shift of the X-ray pulses in the binary orbit. Then, on 6 May, the source disappeared, only to reappear half a day later. This pattern repeated roughly every two days – the X-ray source was being occulted by the primary star in the binary system (Schreier *et al.*, 1972). With

**Fig. 13.17**

(a) The discovery record of the pulsating X-ray source Her X-1. The histogram shows the number of counts observed in successive 0.096 s bins. The continuous line shows the best-fitting harmonic curve to the observations, taking account of the varying sensitivity of the telescope as it swept over the source (Tananbaum *et al.*, 1972). (b) The rate of arrival of X-ray photons from Her X-1, showing the eclipse of the source by the primary star. The source is observed for about 34 hours and then is eclipsed for 6 hours. (c) Variations in the arrival time of pulses from Her X-1. The sinusoidal variation of the pulse arrival time is naturally attributed to the orbital motion of the X-ray source in a binary system.

these clues, the primary star was identified with a massive blue star with the same binary period of 2.1 days as the X-ray source (Krzeminski, 1974). Soon after this discovery, another similar source was discovered, the source Hercules X-1 (Her X-1) which had a pulse period of 1.24 s and an orbital period of 1.7 days (Fig. 13.17) (Tananbaum *et al.*, 1972).

The short period of the X-ray source Her X-1 was compelling evidence that the parent body must be a neutron star, similar to those of the radio pulsars. The energy source was, however, quite different, the *accretion* of matter from the primary star onto the neutron

star. The subject of accretion will be dealt with in detail in Chap. 14 where it is shown that, according to a simple Newtonian calculation, the accretion luminosity onto an object of mass M and radius r is roughly $0.5 \dot{m}c^2(r_g/r)$, where $r_g = 2GM/c^2 = 3(M/M_\odot)$ km is the Schwarzschild radius of an object of mass M and \dot{m} is the mass accretion rate (Sect. 14.2.1). According to this estimate, the accretion of matter onto a $1 M_\odot$ neutron star with radius 10 km can liberate about 10% of the rest-mass energy of the infalling matter. When the effects of general relativity are taken into account, the upper limit to the energy release is 5.72% of the rest-mass energy for accretion onto a non-rotating black hole, roughly an order of magnitude greater than can be liberated by nuclear fusion reactions.

A second calculation is to work out the typical temperature needed to account for the observed X-ray luminosities of binary X-ray sources. Taking the luminosity of a typical luminous X-ray binary to be 10^{30} W and assuming that it is black-body radiation from the surface of a neutron star, the lower limit to the temperature of the emitting region is about 10^7 K. Thus, it is entirely natural that the radiation should be emitted in the X-ray waveband.

A third argument concerns the steady-state X-ray luminosity of accreting compact objects. If the luminosity of the source were too great, the radiation pressure acting on the infalling gas would prevent matter falling onto the surface of the compact object (Sect. 14.2.2). Assuming the radiation pressure acting on the matter is due to Thomson scattering, the critical luminosity, known as the *Eddington luminosity*, depends only upon the mass of the gravitating body,

$$L_{\text{Edd}} = 1.3 \times 10^{31} (M/M_\odot) \text{ W}. \quad (13.43)$$

If other sources of opacity are also important, these increase the radiation pressure and result in a lower value for the critical luminosity above which accretion is suppressed. The luminosities of the binary X-ray sources in the Galaxy and the Magellanic Clouds are more or less consistent with this upper limit. Their luminosity function extends up to about 10^{31} W, above which it cuts off rather sharply.

These arguments show how naturally accretion can account for the properties of binary X-ray sources and also illustrate the importance of accretion as a source of energy in astrophysics. The ramifications of these ideas are profound and they have been extended to the cases of accretion onto black holes, both the stellar mass variety present in a number of X-ray binaries and the supermassive examples which are present in active galactic nuclei. Many more details of the physics of accretion are taken up in Chap. 14.

A wide variety of different types of accreting X-ray sources has been identified. In the case of *high mass X-ray binaries*, the primary star is a massive *late O or early B type star*, these being among the most luminous and massive stars known with short main sequence lifetimes of order 10^7 years. The massive star is responsible for most of the optical light, while the compact object, either a neutron star or black hole, is the dominant source of X-rays. Examples of high mass X-ray binaries include Cygnus X-1, Vela X-1 and 4U 1700–37. Much more common are the *low-mass X-ray binaries* in which the primary star is a low mass main sequence star, with mass, luminosity and temperature similar to those of the Sun. A number of low mass X-ray binaries have been identified as members of globular

clusters. There are numerous variants on this theme, including X-ray bursters, symbiotic X-ray binaries, X-ray pulsars and soft X-ray transients.

Since the X-ray sources are members of binary systems, the masses of the neutron stars can be estimated using the classical techniques of dynamical astronomy. In the best cases, the velocity curves of both the primary and secondary stars are measured. In the case of high mass binaries, the O and B stars are sufficiently bright for accurate measurements of the variation of radial velocity with orbital phase to be made. The velocity curve of the X-ray pulsar can be found from the Doppler shifts of its X-ray pulse period. The X-ray pulsars have periods which range from a fraction of a second to about 15 minutes, the lower end of this range being similar to the periods found in the normal radio pulsars. From the amplitude of the velocity excursions about the mean value for the members of the binary, the ratio of masses of the two stars, M_1/M_2 , can be measured. Absolute values of the masses cannot be determined, however, because only the quantity $(M_1 + M_2) \sin^3 i$ can be estimated, where i is the angle of inclination of the orbit to the plane of the sky. It is therefore necessary to estimate the angle i to make progress. The X-ray source can be considered a point object and so the X-ray source may be occulted by the primary if the plane of the orbit lies close to the line of sight from the Earth. In a number of cases, such X-ray occultations are observed with periods equal to those of the binary orbits. In addition, the X-ray source itself may influence the surface properties of the primary star, either by distorting the figure of the surface into an ovoid shape because of the gravitational influence of the neutron star, or possibly by heating up the face of the primary star closest to the X-ray source, thus causing that face to be more luminous optically when pointing towards the Earth. In the first case, the optical luminosity of the primary is expected to vary at half the period of the binary, whereas, in the second, the optical luminosity varies with the same period as the binary period. There is evidence for both of these phenomena among the binary X-ray sources.

For the low mass systems, it is much more difficult to measure the velocity curve for the faint primary star, the light of which can be overwhelmed by the light from the accretion disc. If the velocity curve can be measured for the X-ray pulsar, this is equivalent to the study of classical *single-line spectroscopic binaries*, in which high-resolution optical spectroscopy can provide the radial velocity of only one star as a function of orbital phase. In these cases, observations of the velocity curve of the X-ray source determine the *mass function* of the binary system,

$$f(M_X, M_0, i) = \frac{M_0^3 \sin^3 i}{(M_X + M_0)^2}, \quad (13.44)$$

where M_X is the mass of the X-ray pulsar and M_0 is the mass of the primary star. Thus, further assumptions are needed to derive the masses of these stars. The best black hole candidates are such single-line spectroscopic binaries.

These procedures have been used to estimate the masses of the neutron stars in X-ray binaries and some examples of these are shown in Fig. 13.18. Also included in this diagram are the masses of the components of the binary radio pulsar systems for which masses can

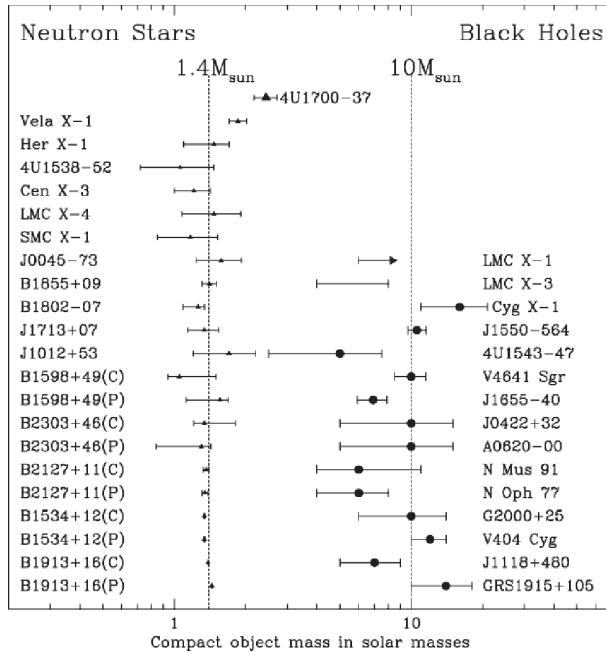
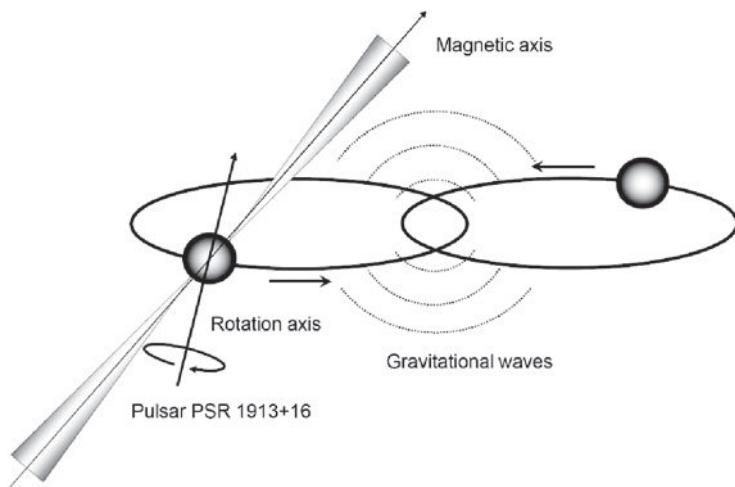


Fig. 13.18 Examples of mass estimates for the neutron stars and black holes in X-ray binary systems and binary radio pulsars for which good mass determinations are available from their velocity curves and other information (Clark *et al.*, 2002).

be found from very accurate pulsar timing (see Sect. 13.5.3). The derived masses of the neutron stars are consistent with the theoretical expectation that their masses should be close to $1.4 M_{\odot}$.

13.5.3 Binary pulsars

The next important advance was the discovery of the binary pulsar PSR B1913+16 by Hulse and Taylor (1975). Up till that time, all pulsars were inferred to be solitary objects since their pulse periods exhibited no periodic Doppler shifts which could be associated with their motion in a binary system. The pulsar PSR B1913+16 was the first to exhibit binary motion with the remarkably short binary period of only 7.75 hours and large orbital eccentricity, $e = 0.617$. The corresponding dimensions of the major and minor axes are 6.4 and 5 light-seconds respectively – for reference, the diameter of the Sun is 4.6 light-seconds. In the case of PSR B1913+16, only one of the pair of neutron stars is a pulsar (Fig. 13.19). Both neutron stars are so inert and compact that the binary system is very ‘clean’ and so can be used in some of the most sensitive tests of general relativity yet devised. Essentially, the pulsar is a perfect clock in a rotating frame of reference. I have described the use of binary pulsars in tests of general relativity and in estimating the masses of the neutron stars in

**Fig. 13.19**

A schematic diagram showing the binary pulsar PSR B1913+16. As a result of the ability to measure precisely many parameters of the binary orbit by ultra-precise pulsar timing, the masses of the two neutron stars have been measured with very high precision.

my book *Galaxy Formation* (Longair, 2008). Suffice to say that the observed acceleration of the binary orbit and the precession of its elliptical orbit are entirely consistent with the expectations of general relativity. Particularly spectacular is the observation that the period of the binary orbit changes as $-d\Omega/dt \propto \Omega^5$, exactly as expected for energy loss due to the quadrupole emission of gravitational waves.

As the computing power available to undertake searches for binary systems in pulsar timing data increased, many more pulsars in binary systems were discovered. Lyne and Graham-Smith (2006) provide an excellent survey of these systems and the means of discovering them. These studies culminated in the discovery in 2003 of the double pulsar PSR J0737–3039, in which both neutron stars are observed as pulsars (Lyne *et al.*, 2004). This system has an orbital period of only 2.4 hours and so the orbital velocities and accelerations are correspondingly greater than those of PSR B1913+16. Because the kinematics of both pulsars could be determined, remarkably precise values for the masses of both components of the binary could be obtained in a matter of years (Kramer *et al.*, 2006). Some properties of the binary system J0737–3039 are given in Table 13.4. Already, the measurements of the Shapiro time delay have provided a strong-field test of relativistic gravity, showing that the observations agree with the predictions of general relativity to 0.05% accuracy. The decay of the orbit due to the emission of gravitational radiation has also been confirmed, with the result that the two neutron stars will coalesce in about 85 My.

The masses obtained from the relativistic binary systems are the most accurate in astronomy. A compilation of masses of neutron stars in binary systems is shown in Fig. 13.20. They all have masses about $1.4 M_{\odot}$, consistent with the expectations of detailed theoretical studies of their stability.

Table 13.4 Various observed and derived parameters for the binary pulsar system PSR J0737–3039 (Lyne *et al.*, 2004; Kramer *et al.*, 2006).

Pulsar	PSR J0737–3039A	PSR J0737–3039B
Spin frequency (Hz)	44.054069392744(2)	0.36056035506(1)
Spin frequency derivative (s^{-2})	$-3.4156(1) \times 10^{-15}$	$-0.116(1) \times 10^{-15}$
Eccentricity	0.0877775(9)	0.0877775(9)
Distance (pc)	~ 500	~ 500
Characteristic age (My)	210	50
Surface magnetic flux density (T)	6.3×10^5	1.6×10^8
Spin-down luminosity (W)	5.8×10^{26}	1.6×10^{23}
Mass M_\odot	1.3381(7)	1.2489(7)

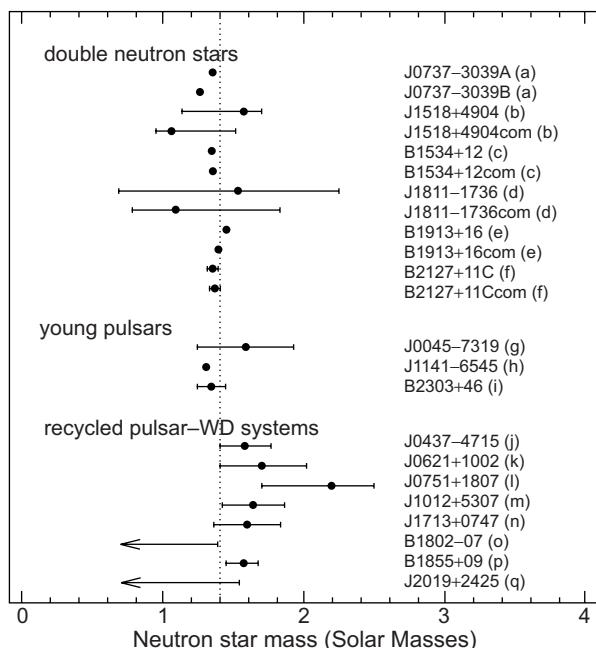


Fig. 13.20 The masses of neutron stars which are members of binary systems. The vertical dotted line indicates a mass of $1.4 M_\odot$ (Stairs, 2004; Lorimer and Kramer, 2005).

13.5.4 Millisecond pulsars

Until the early 1980s, the Crab Nebula pulsar had the shortest known rotation period, the natural assumption being that, because of its youth, it was still rotating rapidly and would in due course spin down to become a normal isolated radio pulsar. In 1982, the millisecond pulsar B1937+21 was discovered by Backer and his colleagues (Backer *et al.*, 1982). The radio source 4C 21.53 was known to be a highly polarised radio source with a steep radio

spectrum at low radio frequencies, similar in character to the radio properties of the Crab Nebula pulsar. In a remarkable analysis of the time-series data from the very bright radio source 4C 21.53, Backer and his colleagues discovered that it was indeed a pulsar with pulse period 1.56 ms, the first of the millisecond pulsars. The demands of computation made the search for similar objects prohibitive until the exponential growth in computer power enabled surveys for similar sources to be carried out effectively in the 1990s. Over 100 millisecond pulsars are now known, the majority of them being members of binary systems.

The millisecond pulsars have very stable pulse periods, from which it is inferred that they must have relatively weak magnetic fields (see (13.42)). Furthermore, because they have much smaller values of \dot{P} than those with periods greater than about 0.1 s, they must have much greater ages, in the most extreme cases of the order of the age of the Universe, $\sim 10^{10}$ years. The millisecond pulsars form a distinctive group of objects to the bottom left of the $P - \dot{P}$ diagram (Fig. 13.16), those which are members of binary systems being enclosed in circles. The fact that the majority of the millisecond pulsars are members of binary systems provides a natural explanation for their short periods. Mass transfer from the primary star to the neutron star transports angular momentum, resulting in spin-up of the neutron star. A weak pulsar magnetic field is a considerable advantage because the magnetic pressure determines the accretion radius about the star and, if this is weak, angular momentum transfer can occur close to the surface of the neutron star resulting in a large spin-up. There is a maximum spin-up rate which is limited by the value of the surface magnetic field strength of the pulsar (van den Heuvel, 1987). This limit can be written $P = 1.9B_g^{6/7}$ ms where B_g is the surface magnetic field strength measured in units of 10^5 T (see Sect. 14.4.2). This relation is plotted on Fig. 13.16 in which virtually all the millisecond pulsars lie below the limiting spin-up line. If the companion star explodes, disruption of the system may occur resulting in the creation of isolated millisecond pulsars. In this picture, a dead pulsar can become alive again if it is a member of a binary system because its period is spun up and the pulsar then recrosses the ‘death-line’. Although the magnetic fields are weak, this is more than compensated for by the fast rotation speeds of the millisecond pulsars.

The association of millisecond pulsars with close binary systems suggested that they should also be present in globular clusters because low mass X-ray binaries are often found in such clusters. This has indeed proved to be the case. Lyne and Graham-Smith provide a catalogue of pulsars in globular clusters, particularly important being the 21 pulsars detected in the nearby globular cluster 47 Tucanae, all of them with pulse periods in the range $2 \lesssim P \lesssim 6$ ms.

13.5.5 Magnetars, soft γ -ray repeaters and anomalous X-ray pulsars

Neutron stars are also the parent bodies of the classes of X- and γ -ray pulsar known as *soft γ -ray repeaters* and *anomalous X-ray pulsars*. These objects were discovered in sky surveys at X- and γ -ray wavelengths by space observatories such as the Rossi X-ray Timing Explorer (RXTE) and SWIFT satellites. These pulsars have pulse periods $2 \lesssim P \lesssim 15$ s, longer than normal radio pulsars, but with very large spin-down rates. These classes of pulsar are shown as stars to the top right of the $P - \dot{P}$ diagram (Fig. 13.16), from which it

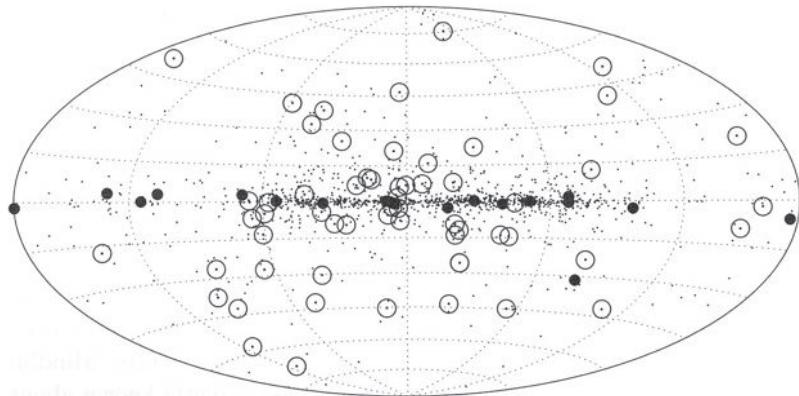


Fig. 13.21 The Galactic distribution of pulsars in a Hammer–Aitoff projection. The dots are normal radio pulsars. Pulsar-supernova associations are shown as larger filled circles. Millisecond pulsars are indicated by dots with open circles (Lorimer and Kramer, 2005).

can be seen that the inferred magnetic flux densities are very large and the lifetimes very short. Many of the anomalous X-ray pulsars exhibit intense outbursts. Collectively, these objects belong to a class of extreme pulsars known as *magnetars*. The source of energy cannot be the loss of rotational kinetic energy since, according to the relation (13.39), these neutron stars are rotating too slowly. Rather, it is inferred that the source of energy must involve the internal magnetic field of the neutron star which is amplified to magnetic flux densities far exceeding those present in the population of normal radio pulsars (Thompson and Duncan, 1995, 1996). These classes of object may well be extreme examples of normal radio pulsars with very strong magnetic fields.²

13.6 The galactic population of neutron stars

The Galactic distributions of the different classes of pulsar discussed in the last section are shown in Fig. 13.21 (Lorimer and Kramer, 2005). The vast majority of the objects, indicated by small dots, are the normal radio pulsars and they are strongly concentrated towards the Galactic equator. Distances can be estimated from the dispersion measures of the pulsars combined with a model for the Galactic distribution of interstellar ionised hydrogen. Analyses of these data show that the pulsars are associated with the spiral arm populations of the Galaxy (Cordes and Lazio, 2002, 2003). Those pulsars associated with supernova remnants, indicated by large filled circles in Fig. 13.21, are found close to the Galactic plane and are certainly members of the youngest stellar populations in our Galaxy.

² A catalogue of magnetars can be found at <http://www.physics.mcgill.ca/~pulsar/magnetar/main.html>, maintained by the McGill pulsar group.

In contrast, the millisecond pulsars, indicated by circled points in Fig. 13.21, are much more isotropically distributed about our location in the Galaxy. Since the majority of the millisecond pulsars are associated with binary systems with much greater ages than normal radio pulsars, they belong to much older stellar populations which have a somewhat broader distribution in Galactic latitude.

Despite their concentration towards the Galactic plane, the normal radio pulsars have a significantly greater scale height, $h \sim 600$ pc, perpendicular to the Galactic plane (Lyne and Graham-Smith, 2006). A natural explanation for this difference is that the normal radio pulsars have very much larger space velocities than typical spiral arm populations. Timing measurements and radio interferometric observations of pulsars have shown that they have very large transverse velocities on the sky, the mean space velocity at birth in the plane of the Galaxy being estimated to be about 450 km s^{-1} . Cordes and his colleagues find that the velocity distribution can be better described by a two-component birth velocity distribution with mean velocities of 90 and 500 km s^{-1} (Arzoumanian *et al.*, 2002). About 15% of the pulsars have birth velocities exceeding 1000 km s^{-1} . An extreme example is the pulsar associated with the Guitar Nebula which has projected velocity 1600 km s^{-1} .

Most of the velocity vectors have large components perpendicular to the Galactic plane which accounts for their very much broader distribution in Galactic latitude as compared with other young Galactic objects. Arzoumanian and his colleagues estimate that roughly half the normal radio pulsars have velocities exceeding the escape velocity of 500 km s^{-1} from the plane of the Galaxy. The ages of the pulsars derived from their kinematic behaviour can be compared with those derived from their indicative ages from their slow-down rates, $\tau = P/2\dot{P}$, and there is reasonable agreement between these estimates, certainly for the younger systems. Interestingly, the millisecond pulsars have much smaller space velocities, $\sim 100 \text{ km s}^{-1}$, as compared with the normal radio pulsars.

Gunn and Ostriker (1970) suggested that the large birth velocities of normal radio pulsars could be attributed to the disruption of a close binary system when one of the stars explodes as a supernova. The smaller velocities could be attributed to this mechanism, but it cannot account for large birth velocities, $v \gtrsim 500 \text{ km s}^{-1}$. The likely explanation for these high velocities is asymmetric collapse and explosion of the core of the pre-supernova star as the neutron star is formed. The simulations described by Woosley and Janka (2005) offer some promise of understanding how these could come about.

Large samples of normal radio pulsars are now available and the sky surveys are sufficiently complete for the luminosity functions, space densities and rates of formation of pulsars to be estimated. These are highly non-trivial calculations since many selection effects influence the probability of detecting a neutron star as a radio pulsar, including the fact that the pulsar radiation has to be beamed and so we observe only a fraction of the total pulsar population. These complications are discussed by Lyne and Graham-Smith (2006) who conclude that the birth rate of pulsars is about once every 60–300 years. Comparing this figure with the rate of supernovae in our Galaxy, it is entirely plausible that most, if not all, radio pulsars were born in supernova explosions.

13.7 Thermal emission of neutron stars

Following the considerations of Sect. 13.1.3, neutron stars are expected to be very hot when they form. They cool by thermal radiation from their surfaces and by neutrino emission from their interiors. Below temperatures of about 10^9 K, the neutron star is transparent to neutrinos and so they provide a very efficient means of getting rid of the thermal energy of the star. After about 300 years, the predicted *surface* temperatures of the neutron stars are about 2×10^6 K and remain in the range about 0.5 to 1.5×10^6 K for at least 10^4 years. The search for thermal X-rays from the surfaces of young neutron stars was one of the initial targets of the early X-ray astronomy missions.

The theory and observation of thermal X-ray emission from the surfaces of neutron stars are reviewed in some detail by Zavlin (2009). In binary X-ray sources, the thermal emission from their surfaces is overwhelmed by the accretion luminosity and so the pulsars needed for this study should be isolated objects. Even then, in many of the young pulsars associated with supernova remnants, the thermal emission from the surface may be masked by the non-thermal emission associated with high energy particles in the strong magnetic fields in the pulsar magnetosphere. The best opportunities are provided by pulsars with ages of order 10^4 – 10^6 years. Pulsars much older than about 10^6 years are expected to have too low temperatures for their X-ray emission to be detectable. There is also the issue of whether the emission originates from the entire surface of the neutron star, or from the polar regions which can be heated by relativistic particles accelerated in these regions. The latter process might make older pulsars detectable as X-ray sources.

The simplest assumption is that the surfaces of neutron stars radiate as black bodies but, as in the case of the stars, this is an oversimplification. The problem is that the gravitational acceleration at the surface of the neutron star is enormous and it is threaded by an extremely strong magnetic field. The scale height H of the atmosphere of the neutron star is very small, $H \sim kT_s/m_pg \sim 0.1$ – 10 cm, where g is the gravitational acceleration at its surface. Furthermore, the enormous magnetic fields change the structures of the atoms, making them linear rather than spherical, with the result that the emission through the surface layers is anisotropic.

Zavlin (2009) has discussed in some detail the comparison of theory and observations for a selected number of pulsar candidates in which there is good evidence that the soft X-ray thermal radiation from the neutron star has been detected. These studies have been aided enormously by the imaging and spectroscopic capabilities of the *Chandra* and *XMM-Newton* X-ray observatories. The pulsar in the Vela supernova remnant and the young pulsars PSR J0538+2817 and PSR B2334+61 have thermal components with $T \sim 10^6$ K. A group of three ‘middle-aged’ pulsars, PSR B0656+14, PSR B1055–52 and the source known as ‘Geminga’ have ages of $(1$ – $5) \times 10^5$ years and have spectra which can be fitted by a three-component model. At high energies, the radiation spectrum is non-thermal and associated with the pulsar magnetosphere. The other two can be modelled as hot and cool black-bodies, the hot component originating from a hot spot on the pulsar, presumably the polar cap regions, and the cool component from the bulk of the stellar surface.

Perhaps the most remarkable sources are seven ‘truly isolated’ radio-quiet neutron stars discovered by the ROSAT X-ray observatory as X-ray pulsars with periods in the range 3.5–11 s. Their spin-down rates indicate ages of order 2×10^6 years and surface magnetic flux densities $B \sim 3 \times 10^9$ T. Their spectra can be fitted by pure black-bodies with no need for a non-thermal component. The surface black-body temperatures lie in the range $T \approx (0.7\text{--}3) \times 10^6$ K. These are probably the best candidates for genuine cooling neutron stars, uncontaminated by non-thermal emission (Haberl, 2007).

13.8 Pulsar glitches

Pulsar periods are remarkably stable once account is taken of their steady decelerations. There are, however, discontinuous changes in the slow-down rates and two types of behaviour have been identified. One type is known as *timing noise*, what Lyne and Graham Smith (2006) describe as ‘a generally noisy and fairly continuous erratic behaviour’. The second type is much more dramatic and consists of large discontinuous changes in the pulsar’s rotation speed, what are referred to as *glitches*. These phenomena occur about once every few years in the Crab Nebula and Vela pulsars. Glitches are rare phenomena and have been observed in fewer than 40 pulsars. They are observed most frequently in the younger pulsars, one third of all known examples occurring in the Crab Nebula and Vela pulsars. In the cases of the Crab Nebula and Vela pulsars, the frequency changes correspond to $\Delta\nu/\nu \sim 10^{-7}$ and 10^{-6} , respectively. Glitches are of special interest because they enable unique insights to be gained into the internal structures of neutrons stars and the behaviour of the superfluid components in their interiors.

The nature of the discontinuity in rotation frequency is illustrated in Fig. 13.22 in which the pulsar eventually settles down to a steady slow-down following the glitch. This phenomenon can be attributed to changes in the moment of inertia of the neutron star as it slows down. An attractive model to explain the general features of Fig. 13.22 is provided by a two-component model for the interior of the neutron star in which the superfluid neutron component is very weakly coupled to the other components, namely the normal component, the crust and the charged particles. Let us call the moments of inertia of these components I_s and I_n , respectively. After a glitch has taken place, it is assumed that the angular frequency of the normal component decreases discontinuously. Following Shapiro and Teukolsky (1983), the rate at which the superfluid component is spun up is determined by the coupling between the superfluid and normal components, the coupling being described by a time constant τ_c . This quantity is the relaxation time for frictional dissipation which is also the time-scale for exchange of angular momentum between the two components. The change of angular frequency with time $\dot{\Omega}$ is then governed by two linear differential equations:

$$I_n \dot{\Omega} = -\alpha - \frac{I_n(\Omega - \Omega_s)}{\tau_c}; \quad I_s \dot{\Omega}_s = \frac{I_n(\Omega - \Omega_s)}{\tau_c}, \quad (13.45)$$

where α describes the loss of rotational energy due to external torques, for example, by magnetic dipole radiation. These equations can be solved to find the rate of change of Ω

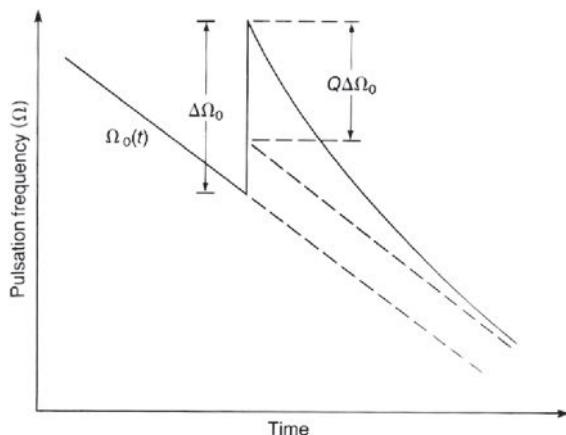


Fig. 13.22 Illustrating the phenomenon of *glitches*. The pulse period increases smoothly as the rotation rate of the neutron star decreases but there are sudden discontinuities in the pulse period, following which the steady increase in period continues. The variation of the pulse period with time during the glitches provides information about the internal structure of the neutron star (Shapiro and Teukolsky, 1983).

with time,

$$\Omega(t) = \Omega_0(t) + \Delta\Omega_0 [Qe^{-t/\tau_c} + 1 - Q], \quad (13.46)$$

where Q is a *healing parameter* which describes the degree to which the angular frequency returns to its extrapolated value $\Omega_0(t) = \Omega_0 - \alpha t/I$, the pulsar angular frequency in the absence of the glitch where Ω_0 is a constant. The significance of these quantities is illustrated in Fig. 13.22. The expression (13.46) is called the *glitch function* and can give a good description of the behaviour of the angular frequency of the neutron star following a glitch. The values of τ_c are related to the physical processes of coupling between the superfluid and normal components of the neutron star. For the Vela pulsar, τ_c is of the order of months while for the Crab pulsar, it is of the order of weeks. These are very long time-scales and indicate that a considerable fraction of the neutron fluid must be in the superfluid state. The two-component model can provide a good explanation of the glitches observed in the Crab and Vela pulsars. Of particular interest is the fact that the values of τ_c and Q for different glitches in the same pulsar seem to be more or less the same, as required by the model.

One mechanism by which the moment of inertia of the neutron star can change is as a result of a *starquake*, by analogy with the deformations of the Earth's crust which take place during an earthquake. The crust takes up an equilibrium configuration in which the gravitational, centrifugal and the solid state forces in the crust are in balance. As the pulsar slows down, the centrifugal forces weaken and the crust then attempts to establish a new equilibrium figure with a lower moment of inertia. In a starquake, the crust establishes its new shape by cracking the surface. Since the moment of inertia decreases, this results in a speed-up of the normal component, that is, the crust, the normal component, the charged particles and the magnetic field. As these components are weakly coupled to the neutron

superfluid, the latter is spun-up through frictional forces over a time-scale τ_c , resulting in a slow down of the crust.

The shape of the crust of a neutron star turns out to be similar to that of a rotating liquid mass, the ellipticity being given by the ratio of its rotational energy to its gravitational binding energy, $\epsilon \approx E_{\text{rot}}/E_{\text{grav}}$. Estimates of ϵ for the Crab Nebula and Vela pulsars are $\epsilon \approx 10^{-3}$ and 10^{-4} , respectively. The changes in ellipticity during glitches can also be evaluated for these pulsars, $\Delta\epsilon = \Delta I/I = -\Delta\Omega/\Omega$. The values for the Crab Nebula and Vela pulsars are $\Delta\epsilon \sim 10^{-8}$ and 10^{-6} , respectively. Note that these changes can be attributed to the shrinkage of the neutron star by only a fraction of a millimetre. Thus, it is quite feasible that the Crab Nebula pulsar has undergone glitches at the observed rate of about once every four years over the last 1000 year. In the case of the Vela pulsar, however, the glitches occur roughly every 2.5 years which is too frequent given that the age of the supernova remnant is about 10^4 years.

More detailed physical models for glitches involve the properties of the rotating neutron superfluid. Superfluid liquids display many remarkable properties, in particular, on a macroscopic scale, the fluid must rotate irrotationally, that is, within the superfluid $\nabla \times \mathbf{v} = 0$. In a superfluid, angular velocity is quantised so that in the lowest energy state $\oint \mathbf{v} \cdot d\mathbf{l} = h/2m_n$, where \mathbf{v} is the velocity of the fluid and $2m_n$ is the mass of a neutron pair. These requirements mean that the rotation of the neutron fluid is the sum of a discrete array of vortices rotating parallel to the rotation axis. The finite vorticity of the fluid is confined to the very core of each vortex tube which consists of normal fluid. Feynman (1972) provides an explanation of how this comes about. In the case of the Crab Nebula pulsar, the number of vortex lines per unit area is about $2 \times 10^9 \text{ m}^{-2}$. The relevance of vortices to the origin of pulsar glitches concerns how they interact with the crustal material. In some models the vortices are *pinned* to nuclei in the crust and in others, they thread the spaces between them. As the star slows down, angular momentum is transferred outwards by the migration of the vortices. If the vortices are pinned, this process is jerky and may lead to small glitches. In the case of the giant glitches, there may be a catastrophic unpinning of the vortices, leading to a large change in rotation speed. This line of reasoning leads to a somewhat complex discussion of how the different superfluid and normal components interact within the various regions of the neutron star. Many of these issues are described by Lyne and Graham-Smith (2006).

13.9 The pulsar magnetosphere

The immediate environment of a pulsar is referred to as its *magnetosphere* by analogy with the magnetically dominated regions around the Earth (see Sect. 11.4). A pulsar may be taken to be a non-aligned rotating magnet with a quite enormous magnetic dipole moment. The electrodynamics of magnetised rotating neutron stars turns out to be a problem of daunting complexity, as described by Mestel in his authoritative book *Stellar Magnetism* (Mestel, 1999). Even the simpler case of an aligned rotating magnet does not have a complete solution. In the original model of Pacini (1967; 1968), the electrodynamics were taken to be that of a magnetised, rotating, perfectly conducting sphere in a vacuum. Then, the

vacuum radiation loss formula (13.32) could be used to estimate the slow-down rate by magnetic dipole radiation.

In the simplest approximation, a pulsar can be taken to be a perfectly conducting sphere with magnetic dipole moment \mathbf{p}_0 aligned with the axis of rotation. A uniform magnetic field is frozen into the sphere which rotates at angular frequency Ω . An induced electric field $\mathbf{E}_i = (\mathbf{v} \times \mathbf{B})$ would therefore be expected to be present, but it is cancelled out by the electric charges which reorganise themselves so that

$$\mathbf{E} + (\mathbf{v} \times \mathbf{B}) = \mathbf{E} + [(\boldsymbol{\Omega} \times \mathbf{r}) \times \mathbf{B}] = 0 , \quad (13.47)$$

because of the infinite conductivity of the medium. As a result, there is a charge distribution within the star which can be found from the relation $\operatorname{div} \mathbf{E} = \rho_e/\epsilon_0$. At the surface of the star, this charge distribution has to be matched to the external vacuum solution of Laplace's equation $\nabla^2 \mathbf{E} = 0$. It was shown by Larmor (1884) that the external electrostatic potential is of quadrupolar form,

$$\phi = -\frac{B_0 \Omega R^5}{6r^3} (3 \cos^2 \theta - 1) , \quad (13.48)$$

where B_0 is the polar magnetic flux density and R is the radius of the neutron star. As a result, there is a surface charge distribution on the sphere.

Goldreich and Julian (1969) realised that the vacuum approximation would not be applicable for pulsar magnetospheres because of the enormous strength of the induced electric fields at the surface of the neutron star. Differentiating (13.48) in the radial direction, there are enormous radial electric fields at the surface of the neutron star, to order of magnitude,

$$E \approx \Omega R B_0 \approx 6 \times 10^{12} P^{-1} \text{ V m}^{-1} , \quad (13.49)$$

where $B_0 = 10^8$ T and the period of the pulsar P is in seconds. The ratio of the Lorentz force to the gravitational force acting on an electron is of order $e(\mathbf{v} \times \mathbf{B})/(GMm_e/R^2) \sim e\Omega R^3 B/GMm_e \approx 10^{12}$ for the case of the Crab Nebula pulsar. Thus, not only is the structure of the pulsar magnetosphere completely dominated by electromagnetic forces, but also the induced electric fields at the surface of the star are so strong that the forces on charges in its surface layers exceed the work function of the surface material and consequently they are dragged off the surface, resulting a plasma surrounding the neutron star. It is therefore inevitable that there is a fully conducting plasma surrounding the neutron star and electric currents can flow in the magnetosphere.

The result is a complex distribution of magnetic and electric fields in the magnetosphere of the neutron star. The induced electric field $\mathbf{E}_i = (\mathbf{v} \times \mathbf{B})$ is neutralised by the flow of charges in the plasma so that the net field is reduced to zero, $\mathbf{E} + (\mathbf{v} \times \mathbf{B}) = 0$, and the space charge distribution can be found from Maxwell's equation $\operatorname{div} \mathbf{E} = \rho_e/\epsilon_0$ where $\rho_e = e(n_+ - n_-)$ is the electric charge density. Performing this calculation, the charge distribution within the undistorted corotating dipolar magnetic field is found to be

$$\rho_e = \epsilon_0 \nabla \cdot \mathbf{E} = \epsilon_0 B_0 \left(\frac{R^3}{r^3} \right) (1 - 3 \cos^2 \theta) . \quad (13.50)$$

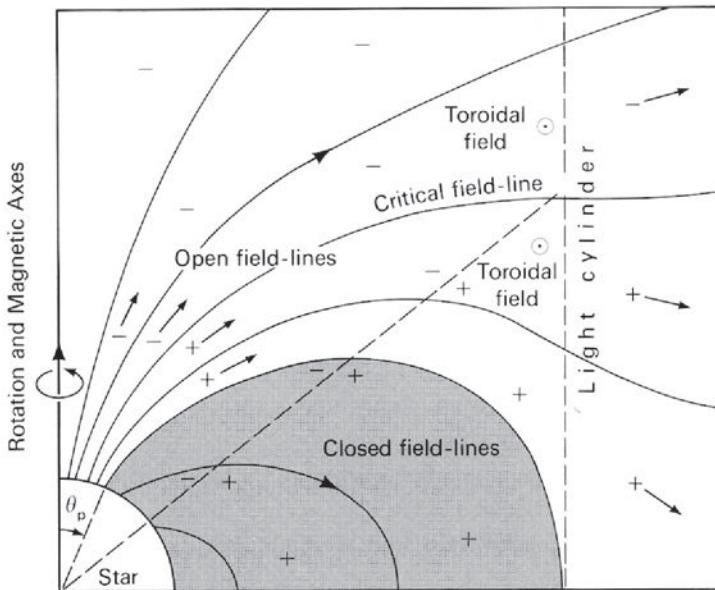


Fig. 13.23

A diagram illustrating the magnetic field and charge distribution about a rotating magnetised neutron star according to the analysis of Goldreich and Julian (1969). The magnetic axis is taken to be parallel to the rotation axis of the neutron star. The charge distribution within the magnetosphere is shown. The *light cylinder* is defined as that radius at which the rotational speed of the corotating particles is equal to the speed of light. Particles attached to closed magnetic field lines corotate with the star and form a corotating magnetosphere. The magnetic field lines which pass through the light cylinder are open and are swept back to form a toroidal field component. Charged particles stream out along these open field lines. The critical field line is at the same potential as the exterior interstellar medium and divides regions of positive and negative current flows from the star. The plus and minus signs indicate the sign of the electric charges in different regions about the neutron star as given by (13.50). The dashed line shows the *zero charge cone* (Manchester and Taylor, 1977).

This distribution has the important property of separating positive and negative charges along *zero charge cones* at an angle $\theta = \arccos(1/3)^{1/2} = 54^\circ 44'$ to the magnetic axis of the neutron star (Holloway and Pryce, 1981).

The resulting magnetic field and charge distribution about an aligned rotating magnetised neutron star is illustrated in Fig. 13.23. A key role is played by the *light cylinder*, or *corotation radius*, at $r_c = c/\Omega$, at which the speed of rotation of material corotating with the neutron star is equal to the speed of light. Within the light cylinder, the closed field lines take up more or less the usual dipole configuration and there is a closed field line which is tangential to the light cylinder. Particles attached to closed magnetic field lines corotate with the star and form the corotating magnetosphere. Those field lines which extend beyond the light cylinder are open and particles dragged off the poles of the neutron star can escape to infinity. Beyond the light cylinder, the charged particles are tied to the magnetic field lines and, just as in the case of the solar wind, the magnetic field takes up a spiral configuration when viewed from above. The magnetic stresses associated with the sweeping back of the magnetic field lines in the vicinity of the light cylinder result in the deceleration of the

pulsar and the energy loss rate turns out to be the same as that given by the magnetic dipole radiation formula (13.32) (Mestel, 1999).

An important feature of the model is the *polar cap* region defined by the field lines which are tangential to the light cylinder. Charged particles within the polar cap are tied to open field lines and so they can escape to infinity. As we will discuss in the next section, these regions are of importance in understanding the intense radio emission of pulsars. The angle subtended by the polar cap can be worked out in the dipole approximation using the fact that, along any dipolar magnetic field line the quantity $\sin^2 \theta / r$ is conserved. Therefore, for small values of θ_{pc} , the angular radius of the polar cap regions is $\theta_{\text{pc}} = (R/r_c)^{1/2} = (\Omega R/c)^{1/2}$ and the radius of the polar cap region is $R_{\text{pc}} = \theta_{\text{pc}} R \approx R^{3/2}/r_c^{1/2}$. Thus, for the case of an aligned pulsar with period 0.1 s, $r_c/R \approx 500$ and the polar angle is $\theta_{\text{pc}} \approx 2.6^\circ$. This structure can be naturally associated with the beaming of the radio emission necessary to account for the radio pulsar phenomenon.

From (13.48), the potential difference $\Delta\phi$ between the pole and the radius of the polar cap can be found by setting $r = R$ and taking the angle θ_{pc} to be small,

$$\Delta\phi \approx \frac{\Omega B_0 R^2 \theta_{\text{pc}}^2}{2} \approx \frac{\Omega^2 R^3 B_0}{2c}. \quad (13.51)$$

Taking $B_0 = 10^8$ T and expressing the period P in seconds, the potential difference is $\Delta\phi \approx 6 \times 10^{12}/P^2$ V. Thus, enormous potential differences are experienced by charged particles within the polar cap regions.

The case of the non-aligned rotator is illustrated in Fig. 13.15 (Lorimer and Kramer, 2005). According to Mestel (1999), many of the features of the aligned rotator reappear in the non-aligned rotator. Again, the distributions of charges and currents are complex and there are no complete solutions for the distributions of charges and fields.

13.10 The radio and high energy emission of pulsars

The physical mechanism by which the *radio pulses* are generated remains a challenging problem. A requirement of all models of the radio emission mechanism is that the radiation cannot be incoherent radiation. The *brightness temperature* of the emission $T_b = (\lambda^2/2k)(S_\nu/\Omega)$ can be estimated from the known distances of the pulsars, the duration of the pulses and their observed flux densities S_ν . Typically, brightness temperatures in the range 10^{23} – 10^{26} K are found. This far exceeds the conceivable temperature of material within the pulsar magnetosphere. A solution of this problem is to associate the radiation with some form of coherent radiation in which the particles radiate in bunches rather than singly. In order to radiate coherently, bunches of, say, N charged particles must have dimension less than the wavelength of the emitted radiation and then, because the radiated power depends upon the square of the oscillating charge, the intensity of

the radiation can be N^2 times that of an individual charge. Alternatively, the emission might be some form of maser emission associated with plasma phenomena in the pulsar magnetosphere.

The infrared, optical, X- and γ -ray pulses observed in the Crab Nebula pulsar have similar pulse profiles to that observed at radio wavelengths. There is however an important distinction between the radio pulses and those observed at higher energies in that the brightness temperatures of the radiation at optical and shorter wavelengths would be consistent with the incoherent emission of high energy electrons. The brightness temperature of the X-ray emission is about 10^{11} K which corresponds to electron energies of about 10 MeV. Thus, the emission at the higher energies can be the incoherent radiation of high energy electrons accelerated in the pulsar's magnetosphere.

In both cases, the acceleration of charged particles to high energies in the pulsar magnetosphere is a requirement of the models and this may take place in what are referred to as the *gap regions* in the pulsar magnetosphere. In Fig. 13.15, these regions are referred to as the *inner* and *outer acceleration gaps* and are potential sites for the acceleration of electrons. The inner acceleration gap region lies just above the polar cap. Ruderman and Sutherland (1975) pointed out that, although the electrons may be dragged from the surface of the neutron star, the protons are very tightly bound indeed to the crust of the neutron star. As a result, rather than a neutral stream, a flux of electrons is dragged from the surface of the polar cap along open field lines and are accelerated by the intense induced electric fields just above the polar caps. The accelerated particles very rapidly lose any velocity component perpendicular to the magnetic field lines by synchrotron radiation, but, as they stream out along the curved magnetic field lines, they radiate *curvature radiation* (Sect. 8.3). In the strong magnetic fields close to the pulsar, the high energy photons produced by the curvature radiation interact with the transverse component of the magnetic field to produce electron–positron pairs. The threshold for this process is given by the relation $h\nu B_{\perp} \geq 4 \times 10^{14}$ eV T, where the energy of the photons $h\nu$ is expressed in eV and the magnetic flux density in tesla (Erber, 1966). In turn, these electrons and positrons radiate high energy photons which generate electron–positron pairs, and so on. The results is an electron/positron–photon cascade, similar to those described in Sect. 9.9. The electron and positrons are accelerated in opposite directions, resulting in a ‘spark’ which ‘shorts out’ the continuous process of particle acceleration, resulting in a finite sized ‘gap region’ above the polar cap. Similar cascades could be initiated by $\gamma\gamma$ interactions if ambient low energy photons were present in the gap regions.

Ruderman and Sutherland proposed that the electron–photon cascades would result in bunches of particles which would radiate coherently in sheets, resulting in the extreme values of the radio brightness temperatures observed in radio pulsars. It is natural to associate these processes with the open field lines emanating from the poles of the magnetic field structure. The sparks produced in the inner acceleration gap region may account for the fine structure observed in the individual pulses. An attractive feature of this model is that it may explain why most pulsars have periods about 1 s and magnetic field strength about 10^8 T – only if the magnetic field is strong enough and if the rotation period is sufficiently short, will electron–photon cascades take place. According to this theory, there is a ‘death-line’ in

the magnetic field–period relation for pulsars such that $B T_p^2 \geq 10^7 \text{ T s}^{-1}$, where T_p is the period of the pulsar in seconds (see Fig. 13.16).

The *outer acceleration gaps* have a slightly different origin and are associated with the *zero charge cones* shown in Fig. 13.23. Again, if the protons were removed from the vicinity of the zero charge cones, the resulting perturbations to the electric field result in a narrow charge-free zone about the cones. Details of the formation of the outer gap regions are described by Holloway and Pryce (1981). Within these zones, electrons can be accelerated under the influence of the induced electric field and processes similar to those described above in the case of the inner acceleration gap can take place giving rise to beams of radiation along the zero charge cones. It may well be that the high energy infrared, optical and X-ray emission could be associated with the *incoherent* radiation of electrons accelerated in these gap regions. Lyne and Graham-Smith (2006) provide a helpful summary of the problems associated with interpreting the pulsed radio and high energy emission from pulsars in terms of processes occurring in the inner and outer acceleration gaps.

13.11 Black holes

Neutron stars are the last known forms of stable star. In a prescient paper presented to the Royal Society of London in 1783, John Michell noted that, if a star were sufficiently massive, the escape velocity from its surface would exceed the speed of light and so light would not escape from it (Michell, 1784). In classical terms, the escape velocity from the surface of a star of mass M and radius r is $v = (2GM/r)^{1/2}$. Setting v equal to the speed of light c , the radius of such a star would be $r_g = 2GM/c^2$. In addition, for objects more compact than neutron stars, the attractive force of gravity becomes so strong that no physical force can prevent collapse to a physical singularity. These two properties explain the origin of the term *black hole* – no radiation can emerge from within the radius r_g and so the object is ‘black’, and at the same time the singularity represents a ‘hole’ in space-time into which matter can collapse but from which it cannot emerge (Wheeler, 1968). It turns out that the form of the expression for r_g is exactly the same as the *Schwarzschild radius* of a black hole of mass M in general relativity, although the coordinate r has a different meaning from the Newtonian definition of distance. This spherical surface plays the role of the black surface of the black hole.

To do full justice to the physics and mathematics of black holes is an enormous undertaking and the enthusiast may wish to tackle the books by Chandrasekhar *The Mathematical Theory of Black Holes* (1983) and by Frolov and Novikov *Black Hole Physics: Basic Concepts and New Developments* (1998). These studies go far beyond the scope of this book. I have already given a number of heuristic arguments in my book *Theoretical Concepts in Physics* (TCP2) which make plausible the physical content of the *General Theory of Relativity* (Longair, 2003). These are not repeated here but it is helpful to summarise a number of key results.

13.11.1 General relativity and the Schwarzschild metric

Within months of the Einstein's definitive formulation of general relativity in 1915, Schwarzschild (1916) discovered the metric of space-time about a point mass M ,

$$ds^2 = \alpha dt^2 - \frac{1}{c^2} [\alpha^{-1} dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)] , \quad \text{where } \alpha = \left(1 - \frac{2GM}{rc^2}\right) . \quad (13.52)$$

This metric is known as the *Schwarzschild metric* and has the same meaning as the interval ds^2 in special relativity, the *Minkowski metric*,

$$ds^2 = dt^2 - \frac{1}{c^2}[dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)] . \quad (13.53)$$

In both cases, the metrics are written in spherical polar coordinates. The coordinates r and t have, however, different meanings in the metrics (13.50) and (13.51). In particular, the metric (13.50) describes the interval between events in the curved space-time about the point mass.

In the limit of large distances from the point mass, $r \rightarrow \infty$, the two metrics are the same. They are different, however, at small values of r , reflecting the influence of the mass M upon the geometry of space-time. These differences are explored in more detail in Chap. 17 of TCP2. Consider the case of two events in space-time separated by coordinates dr, dt with $d\theta = d\phi = 0$, that is, the spatial increment is in the radial direction. In special relativity, dr and dt are the distance and time increments in some inertial frame of reference S and the *proper time* interval $d\tau$ between events is

$$d\tau = ds = \left(dt^2 - \frac{dr^2}{c^2}\right)^{1/2} . \quad (13.54)$$

In general relativity, the distance coordinate r appearing in the metric (13.50) gives the correct result for proper distances measured perpendicular to r , as can be seen from inspection of the metric – the proper distance dl is $dl = r d\theta$ and hence r is often referred to as an *angular diameter distance*. The increment of proper distance in the radial direction in the reference frame in which the point mass is at rest is *not* dr , but is given by the radial component of the metric, $dx = \alpha^{-1/2} dr$; dx is an increment of *geodesic distance* in the radial direction. In general relativity, geodesics replace straight lines as the shortest distance between two points. Thus, the r 's appearing in metrics (13.50) and (13.51) have somewhat different meanings. The reason for this is that the geometry of space-time is not flat and the curvature of space-time varies from point to point.

In general relativity, there are three different time intervals. Unlike special relativity, the rate at which time passes depends upon the gravitational potential, a result which follows directly from the principle of equivalence (see TCP2, Sect. 17.2.2). In Newtonian terms, the gravitational potential at radius r is $-GM/r$. In the case of the Schwarzschild metric, the time interval between events according to an observer who is stationary at the point r in S is $dt' = \alpha^{1/2} dt$. Thus, the time interval dt' depends upon the gravitational potential in which the observer is located and it only reduces to dt in the limit of very large distances

from the origin, $r \rightarrow \infty$, at which the gravitational potential goes to zero. The time interval dt is a convenient *coordinate time* since every observer can relate the time dt' to dt because it only depends upon the gravitational potential in which the observer is located.

In general relativity, the increment of *proper time* is

$$ds = d\tau = dt \left[\alpha - \frac{1}{c^2 \alpha} \left(\frac{dr}{dt} \right)^2 \right]^{1/2} = dt' \left(1 - \frac{1}{c^2} \frac{dx'^2}{dt'^2} \right)^{1/2}, \quad (13.55)$$

again for the case $d\theta = d\phi = 0$. Since $dx' = 0$ in the proper frame of reference S' , the observer in S' must be moving at a velocity $v = dx/dt'$ through S . This is the *free-fall velocity* of the observer towards the origin. In other words, the proper time interval is

$$d\tau = dt' \left(1 - \frac{v^2}{c^2} \right)^{1/2} = \frac{dt'}{\gamma}, \quad (13.56)$$

where γ is the Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$. The expression (13.54) is the standard time dilation formula between the frame S and the proper frame S' of an observer in free-fall towards the point mass in the radial direction.

The metric (13.50) can be used to derive the formula for the redshift of electromagnetic waves emitted from the point r from the origin as observed at infinity. If the time interval $\Delta t'$ corresponds to the period of the waves emitted at the point r , the observed period of the wave at infinity Δt is given by

$$\Delta t' = \left(1 - \frac{2GM}{rc^2} \right)^{1/2} \Delta t. \quad (13.57)$$

Therefore, the emitted and observed frequencies, ν_{em} and ν_∞ , respectively, are related by

$$\nu_\infty = \nu_{\text{em}} \left(1 - \frac{2GM}{rc^2} \right)^{1/2}. \quad (13.58)$$

This expression is the general relativistic result corresponding Michell's insight of 1783. If the radiation is emitted from radial coordinate $r = 2GM/c^2$, the frequency of any wave is redshifted to zero frequency. This means that no information can reach infinity from radii $r \leq r_g = 2GM/c^2$. The radius r_g is known as the *Schwarzschild radius*. According to general relativity, no radiation can escape from within this radius and so the surface $r = r_g$ is 'black'. Notice that, in this treatment, the r coordinate is an angular diameter distance.

The other key aspect of the Schwarzschild metric concerns the dynamics of test masses in the field of the point mass. It is simplest to begin with the Newtonian expressions for the conservation of energy and angular momentum of a test particle in a gravitational field. The velocity of the particle v can be written in spherical polar coordinates $v^2 = \dot{r}^2 + (r\dot{\theta})^2$ and so, writing the gravitational potential energy of the particle as $-GmM/r$, conservation of energy results in the relations

$$\frac{1}{2}mv^2 - \frac{GMm}{r} = \frac{1}{2}mv_\infty^2, \quad \dot{r}^2 + (r\dot{\theta})^2 - \frac{2GM}{r} = v_\infty^2, \quad (13.59)$$

where v_∞ is the velocity of the particle at infinity. The conservation of angular momentum, $m\dot{\theta}r^2 = \text{constant}$, can be written in terms of the *specific angular momentum* of the particle

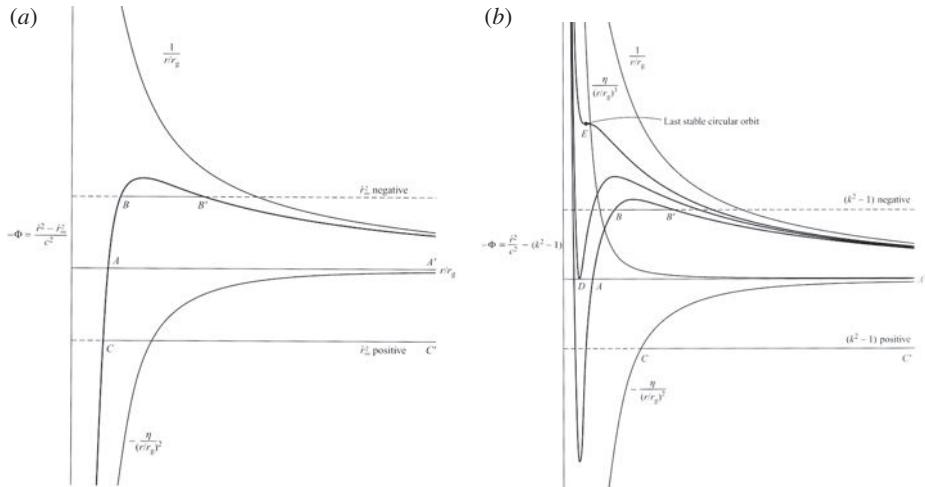


Fig. 13.24

(a) Illustrating the variation of the potential $-\Phi = (r^2 - r_\infty^2)/c^2$ according to Newtonian gravity. There are two contributions to the potential, the gravitational potential varying as r^{-1} and the negative centrifugal potential varying as r^{-2} . (b) Illustrating the variation of the potential $-\Phi = (r^2/c^2) - (k^2 - 1)$ according to general relativity. There is in addition an attractive gravitational potential varying as r^{-3} . The various loci illustrate the origin of black holes and the last stable orbit according to general relativity (Longair, 2003).

$h = \dot{\theta}r^2$, that is, its angular momentum per unit mass. Then (13.57) can be written

$$\dot{r}^2 + \frac{h^2}{r^2} - \frac{2GM}{r} = \dot{r}_\infty^2 , \quad (13.60)$$

where \dot{r}_∞ is the radial velocity of the particle at infinity. In this formulation, r and t have their usual Newtonian meanings in Euclidean geometry. Notice that, according to (13.58), so long as h is non-zero, the particle cannot reach $r = 0$ because the energy term associated with the centrifugal potential, h^2/r^2 , becomes greater than the gravitational potential energy $2GM/r$ for small enough values of r .

The corresponding equation obtained from a general relativistic treatment is as follows:

$$\dot{r}^2 + \frac{h^2}{r^2} - \frac{2GM}{r} - \frac{2GMh^2}{r^3c^2} = (k^2 - 1)c^2 , \quad (13.61)$$

where h and k are constants of the motion (see TCP2, Sect. 17.6). This equation looks formally similar to (13.58) but there are important differences, most importantly, the extra term $(-2GMh^2/r^3c^2)$ on the left-hand side. Note also that r is an angular diameter distance and \dot{r} means dr/ds where s is proper time. The constants h and $(k^2 - 1)$, which come out of the formal analysis of the dynamics of the particle according to the Schwarzschild metric, are equivalent to the Newtonian constants h and \dot{r}_∞^2 . The additional term $(-2GMh^2/r^3c^2)$ has the crucial effect of enhancing the attractive force of gravity, even if the particle has a finite specific angular momentum h . The greater the value of h , the greater the enhancement. One way of thinking about this result is to recall that the kinetic energy associated with rotational motion about the point mass contributes to the inertial mass of a test particle and thus enhances the gravitational force upon it.

A number of elegant results can be found from the analysis of (13.59). First of all, it can be seen by inspection that, for sufficiently small values of r , the general relativistic term $-2GMh^2/r^3c^2$ becomes greater than the centrifugal potential term. It is a useful exercise to show that, if the specific angular momentum of a particle is $h \leq 2r_g c$, it inevitably falls in to $r = 0$, where r_g is the Schwarzschild radius, $r_g = 2GM/c^2$. Another important related result is that there is a *last stable circular orbit* about the point mass with radius $r = 3r_g$. There do not exist circular orbits with radii less than this value, the particles spiralling rapidly in to $r = 0$. This is why the black hole is called a ‘hole’ – matter inevitably collapses in to $r = 0$ if it comes too close to the point mass. These results are derived and illustrated in Sect. 17.6.2 of TCP2.

The Schwarzschild radius r_g plays a special role in the theory of black holes. Inserting the values of the constants,

$$r_g = \frac{2GM}{c^2} = 3 \left(\frac{M}{M_\odot} \right) \text{ km}. \quad (13.62)$$

Thus, for the Sun, $r_g = 3$ km, which is negligible compared with its radius of 695 980 km. A neutron star with $M = M_\odot$ and radius $r = 10$ km has $2GM/rc^2 = 0.3$ and so general relativity is no longer a small correction factor in determining the stability of these stars. Indeed, the last stable orbit has radius 9 km and so these neutron stars have just avoided becoming black holes.

In the metric (13.50), there appears to be a singularity at the Schwarzschild radius $r_g = 2GM/c^2$. It can be shown, however, that this is not a physical singularity, perhaps the simplest demonstration of this being the fact that a particle can fall through r_g and into $r = 0$ in a finite proper time. The apparent singularity in the metric results from the particular choice of coordinates in which the metric (15.50) has been written – if Kruskal or Finkelstein coordinates are used, the singularity disappears; see, for example, Shapiro and Teukolsky (1983), Sect. 12.6, or Rindler (2001), Sect. 8.6.

At $r = 0$, there is a real *physical singularity* and, according to classical general relativity, the infalling matter collapses to a singular point. This is an unappealing situation, but it may be of some comfort that this Schwarzschild singularity is unobservable because no information can arrive to the external observer from within the Schwarzschild radius r_g . There has been a great deal of study of the nature of these singularities, in particular, the general conditions under which black holes inevitably form. The singularity theorems developed by Penrose and Hawking give a precise answer to this question (see, for example, Hawking and Ellis 1973). The conditions for the avoidance of a singularity are quite extreme. If, for example, matter had a negative pressure equation of state $p = -\varrho c^2$, it might be possible to avoid the singularity. The singularity theorems are based upon classical theories of gravity, such as general relativity, or theories in which the force of gravity is attractive. As yet, there is no quantum theory of gravity but intuitively one feels that taking proper account of quantum phenomena in collapse to a black hole must be necessary to obtain a deeper understanding of the nature of these singularities.

13.11.2 The general case of black holes in general relativity

The discussion of Sect. 13.11.1 gives some impression of the physics of the simplest forms of black hole in general relativity. In 1916, the next spherically symmetric solution of the Einstein field equations was discovered, the case of a point mass M with electric charge Q . The external solutions for the space-time geometry are given by the Reissner–Nordström metric which we will not reproduce here; see Misner, Thorne, and Wheeler (1973) for more details. Then, in 1962, Kerr discovered the general solution for a black hole with angular momentum J (Kerr, 1963). In 1965, a further generalisation of the Kerr metric was discovered by Newman and his colleagues for the case of a rotating system with finite electric charge by solving the combined Einstein and Maxwell field equations (Newman *et al.*, 1965). Only later was it realised that the metric describes a rotating black hole with finite electric charge. Whereas the Schwarzschild solution is determined entirely by the mass of the black hole, the Kerr metric depends upon both its mass and angular momentum. It took some years and a great deal of analysis before it was realised just how powerful these solutions are. In 1971, Carter showed that the only possible solutions for uncharged axisymmetric black holes were the Kerr solutions (Carter, 1971) and in 1972 Hawking showed that all stationary black holes must be either static or axisymmetric so that the Kerr solutions indeed included all possible forms of black hole (Hawking, 1972). Thus, isolated black holes can only possess these three properties, mass M , charge Q and angular momentum J – all other quantum numbers and properties are radiated away when the black hole is formed. Black holes are the simplest macroscopic objects in nature.

Rotating black holes, often referred to as *Kerr black holes*, possess a number of properties which are of the greatest relevance for many aspects of high energy astrophysics. The mathematical analysis of their properties is of quite daunting complexity – see, for example, Chandrasekhar (1983) and his footnote on page 530. We can only scrape the surface of the subject, indicating some of the properties of rotating black holes which have application in understanding astrophysical phenomena. To give some flavour for the subject, the *Kerr metric* can be written in Boyer–Lindquist coordinates,

$$\begin{aligned} ds^2 = & \left(1 - \frac{2GMr}{\rho c^2}\right) dt^2 - \frac{1}{c^2} \left[\frac{4GMra \sin^2 \theta}{\rho c} dt d\phi \right. \\ & \left. + \frac{\rho}{\Delta} dr^2 + \rho d\theta^2 + \left(r^2 + a^2 + \frac{2GMra^2 \sin^2 \theta}{\rho c^2}\right) \sin^2 \theta d\phi^2 \right]. \end{aligned} \quad (13.63)$$

I have preserved the dimensional constants G and c and adopted the same sign conventions as in the metrics (13.50) and (13.51). In this formalism, the black hole rotates in the positive ϕ direction and the symbols have the following meanings: $a = (J/Mc)$ is the angular momentum of the black hole per unit mass and has the dimensions of distance; $\Delta = r^2 - (2GMr/c^2) + a^2$ and $\rho = r^2 + a^2 \cos^2 \theta$. If the black hole is non-rotating, $J = a = 0$ and the Kerr metric reduces to the Schwarzschild metric (13.50). The properties of this metric are described by Misner, Thorne, and Wheeler (1973) and by Shapiro and Teukolsky (1983). Let us summarise some of the more important properties of Kerr black holes.

- (i) The metric coefficients are independent of t and ϕ and so the metric is stationary and axisymmetric about the rotation axis.
- (ii) Just as in the case of the Schwarzschild metric, the metric coefficient of dr^2 becomes singular at a certain radial distance, in the case of the Kerr metric when $\Delta = 0$. This radius corresponds to the surface of infinite redshift or to the *horizon* of the rotating black hole. The radius at which this occurs is the solution of

$$\Delta = r^2 - \left(\frac{2GMr}{c^2} \right) + a^2 = 0. \quad (13.64)$$

Taking the larger of the two roots of this quadratic equation, the horizon occurs at radius r_+ given by

$$r_+ = \frac{GM}{c^2} + \left[\left(\frac{GM}{c^2} \right)^2 - \left(\frac{J}{Mc} \right)^2 \right]^{1/2}. \quad (13.65)$$

This spherical surface has exactly the same properties as the Schwarzschild radius in the case of non-rotating black holes. Particles and photons can fall in through this radius but they cannot emerge outwards through it according to general relativity.

- (iii) If the system has too much angular momentum J , no black hole is formed. Inspection of (13.63) shows that this maximum angular momentum is $J = GM^2/c$. This result is of special importance because matter falling into a black hole adds angular momentum as well as mass to the black hole. For a maximally rotating Kerr black hole, the horizon radius is $r_+ = GM/c^2$, which is half the radius of a Schwarzschild black hole of the same mass, $r_g = 2GM/c^2$.
- (iv) The effect of the rotation of the black hole is that, as one approaches closer to the black hole, the phenomenon of the *dragging of inertial frames* becomes more and more important. As expressed by Thorne (1986),

The hole's rotation drags *all* physical objects near it into orbital motion in the same direction as the hole rotates; nothing can resist. The nearer one is to the horizon, the stronger the dragging effect. Therefore, it is inevitable that our fiducial observers have some radius-dependent finite angular velocity as seen by a distant static (undragged) observer.

This also means that any gyroscope will be observed to precess relative to the distant stars. Before the horizon equivalent to the Schwarzschild radius is reached, the dragging of the inertial frames becomes so strong that there is a limiting radius, known as the *static limit*, within which no observer can remain at rest relative to the background stars. All observers within this limit must rotate in the same direction as the black hole. No matter how hard an observer tries to remain static, for example by firing rocket engines in the direction opposite to the rotation, rotation relative to the background stars cannot be prevented. Within the static radius, light cones point in the direction of rotation. The static radius about the black hole r_{stat} is

$$r_{\text{stat}} = \frac{GM}{c^2} + \left[\left(\frac{GM}{c^2} \right)^2 - \left(\frac{J}{Mc} \right)^2 \cos^2 \theta \right]^{1/2}, \quad (13.66)$$

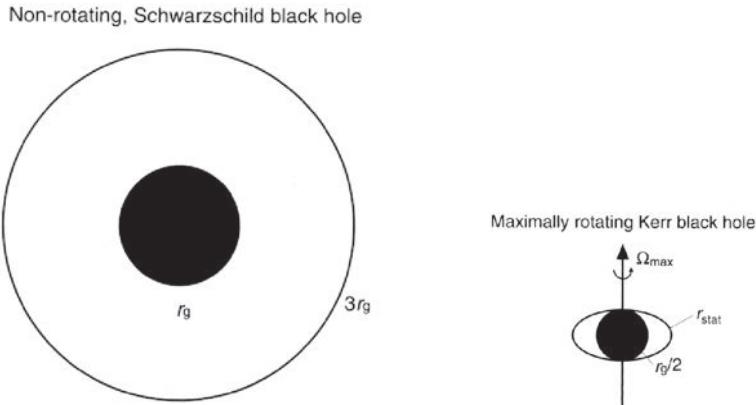


Fig. 13.25 A schematic diagram illustrating the various regions about a maximally rotating black hole compared with the corresponding dimensions for a non-rotating black hole of the same mass.

where θ is the polar angle measured relative to the rotation axis of the black hole. In the polar directions $\theta = 0$ and π , r_+ and r_{stat} coincide but they are distinct in all other directions. Only in the case of no rotation $J = 0$ do the two radii coincide for all values of θ . A diagram showing the relative dimensions of rotating and non-rotating black holes of the same mass is shown in Fig. 13.25.

- (v) The shrinking of the horizon for rotating black holes described in (iii) is an important result because it suggests that, in the case of a maximally rotating black hole, matter can move in stable circular orbits about the black hole with much smaller radii as compared with the Schwarzschild case. Consequently, it is expected that more of the rest-mass energy of the infalling matter can be extracted as compared with the non-rotating case.

In the case of a Schwarzschild black hole, the maximum energy which can be released corresponds to the binding energy of the matter on the last stable circular orbit at $r = 3r_g = 6GM/c^2$. The fraction $[1 - (8/9)^{1/2}] = 0.0572$ of the rest mass energy can be released in this process.³ For Kerr black holes, the results depend upon whether the matter moves in the direction of rotation of the black hole (corotation) or in the opposite direction (counter-rotation). The most interesting cases are the maximally rotating black holes for which $J = GM^2/c$. The last stable circular orbits lie at $r = r_+ = GM/c^2$ for corotating test particles and at $r_+ = 9r_+ = 9GM/c^2$ for counter-rotating particles. The latter orbit lies outside the static limit, which is not a surprise. For the corotating case, the maximum binding energy is a fraction $(1 - 1/\sqrt{3}) = 0.423$ of the rest-mass energy of the orbiting material and $(1 - \sqrt{(25/27)}) = 0.0377$ for counter-rotating orbits (Shapiro and Teukolsky, 1983). The corotating result is of the greatest interest because up to 42.3% of the rest-mass energy of the material can be released as it spirals into the black hole through a sequence of almost circular orbits. This is the process by which energy is liberated by friction in accretion discs

³ I have given a proof of this result in TCP2, Sect. 17.10 (Longair, 2003).

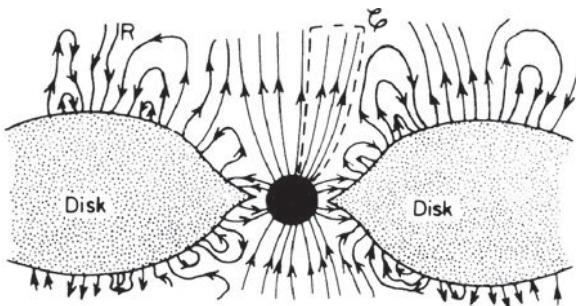


Fig. 13.26 A possible configuration for the magnetic field distribution about a black hole which is surrounded by a thick accretion disc. The magnetic field lines are linked to both the accretion disc and the surrounding medium (Thorne *et al.*, 1986).

about black holes and is the probable source of energy in some of the most extreme astrophysical objects (see Chap. 14). Note that the energy available is very much greater than that attainable from nuclear fusion processes which at most can release about 1% of the rest-mass energy of the matter.

- (vi) The region of space-time between the radii r_+ and r_{stat} is referred to as the *ergosphere* and it has a number of important properties which may be of importance astrophysically. Although matter has to rotate in the same direction as the black hole within the ergosphere, particles can escape from it by the *Penrose process* which depends upon the existence of negative energy orbits within the region defined by the static limit. Penrose (1969) showed that, if a particle enters one of these negative energy orbits and splits into two, one of them can fall down the hole while the other escapes to infinity with greater energy than the original particle had when it fell in. The source of the extra energy is the rotational energy of the black hole. The reason that the volume between the horizon at r_+ and the static limit at r_{stat} is called the *ergosphere* is that this is the region from which energy can be extracted by the Penrose process.
- (vii) Thus, the rotational energy of the black hole can be tapped to provide energy to the exterior Universe. Penrose (1969) showed that the fraction of the rest-mass energy of a rotating black hole which can be made available to the exterior Universe is

$$1 - 2^{-1/2} [1 + [1 - (J/J_{\max})^2]^{1/2}]^{1/4}. \quad (13.67)$$

For maximally rotating black holes, this percentage amounts to 29%, again demonstrating their importance as energy sources for high energy astrophysical phenomena.

- (viii) As already stated, an isolated black hole cannot possess any properties other than mass, electric charge and angular momentum. There is no reason, however, why a magnetic field cannot be associated with a black hole provided it is ‘tied to’ the surrounding medium. This is illustrated schematically in Fig. 13.26 which shows the electrodynamics of such configurations may well be important in understanding the origin of jets of high energy particles observed in active galactic nuclei. In their remarkable book *Black Holes – the Membrane Paradigm*, Thorne, Price and

Macdonald (1986) show that the electrodynamics of all types of black holes can be understood if the horizon is replaced by a two-dimensional membrane with the resistivity of free space.

- (ix) In addition to a finite resistivity, the surface of a black hole has a temperature defined by

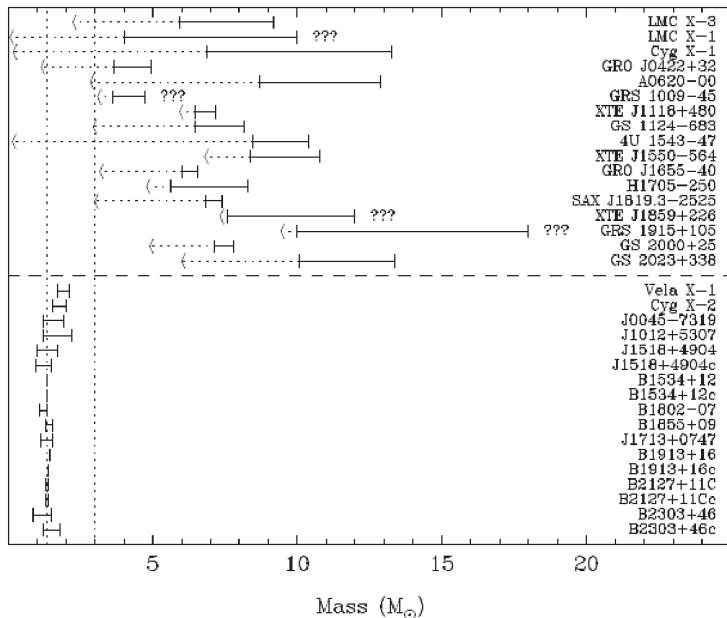
$$T = \frac{\hbar c^3}{8\pi GMk} \approx \frac{10^{-7}}{(M/M_\odot)} \text{ K}. \quad (13.68)$$

According to general relativity, nothing can escape from a black hole and therefore it cannot emit radiation. In a remarkable analysis, Hawking showed, however, that, when quantum electrodynamical processes are considered at the horizon of a black hole, there is a finite probability of particles escaping to infinity (Hawking, 1975). A simple picture of this process is that virtual photons are created from the vacuum at the horizon and, before they can annihilate, one of them falls into the hole while the other escapes to infinity. Hawking predicted that the emitted particles should have a thermal spectrum at the temperature given by (13.66). Evidently, the rate of evaporation of solar mass black holes by this process is very low indeed. Since the temperature is inversely proportional to mass, however, if very low mass black holes were formed in the very early Universe, all those with mass less than about 10^{12} kg would have evaporated by now. Primordial black holes with roughly this mass should be ending their existence at the present epoch and would be expected to result in a strong pulse of electromagnetic radiation. Inserting $M = 10^{12}$ kg into (13.66), the temperature of the thermal emission should be about 10^{11} K so that a pulse of γ -rays would be expected.

We will find that many of these ideas find application in different contexts in high energy astrophysics.

13.11.3 Observational evidence for black holes

Evidence for the presence of black holes in binary X-ray sources and in active galactic nuclei is now wholly convincing. We will deal with the latter in Chap. 19. Galactic X-ray sources which are members of binary systems have proved to be a rich hunting ground for systems consisting of a stellar mass black hole and a normal companion. Those binary X-ray sources which exhibit X-ray pulsations are identified as X-ray pulsars, but there are others in which the X-ray source is luminous and highly variable in intensity on millisecond time-scales, but does not exhibit X-ray pulsations. These are prime candidates for the presence of black holes in Galactic X-ray binary systems. The first compelling example of such system was the bright X-ray source Cyg X-1, which was identified with a ninth magnitude blue supergiant star which turned out to be the primary star of a binary system with a period of 5.6 days (Bolton, 1972; Webster and Murdin, 1972). The mass function of the system was determined from the velocity curve and, assuming the primary star had mass about $10 M_\odot$, the invisible X-ray source was found to have mass about $20 M_\odot$.

**Fig. 13.27**

A compilation of mass estimates for the 17 black holes, the arrows indicating the value of the mass function, which is a firm lower limit for the mass of the black hole (Orosz, 2007). The solid lines indicate the best estimates of the masses of black holes. In the lower half of the figure, the masses of neutron stars in binary neutron star systems are also shown.

Since that time the number of black hole candidates has increased dramatically. The techniques of mass measurement are exactly the same as those described in Sects 13.5.2 and 13.5.3 for estimating the masses of neutron stars in binary systems. In the present cases, however, velocity curves are only available for the normal stellar companion and so masses are estimated using the techniques developed for *single line spectroscopic binaries*. The upper half of Fig. 13.27 contains similar information to Fig. 13.18, and shows a compilation of mass estimates for 17 black hole candidates, the solid bars indicating the best estimates of the masses of the unseen companion. The importance of Fig. 13.27 is that it includes arrows which indicate the value of the mass function of the binary system – these provide firm lower limits to the masses of the optically invisible black hole candidates (Orosz, 2007). Most of these examples provide convincing evidence for the presence of optically invisible companions with masses which must exceed $3 M_{\odot}$, a generous upper limit for the stability limit for neutron stars. The lower half of the diagram shows the corresponding diagram for neutron star binaries in which radio or X-ray pulsars are detected.

A particularly important discovery has been the identification of the X-ray binary system M33 X-7/M33 X-7 (X-ray binary) in the nearby spiral galaxy M33 (Orosz *et al.*, 2007). The system has an orbital period of 3.45 days and the primary star is a very luminous O7III-8III star with mass $70 \pm 6.9 M_{\odot}$. The importance of this system is that it is the only known example of an eclipsing X-ray binary which contains a massive black hole. The X-ray eclipses are shown as a function of orbital phase in Fig. 13.28a, the duration of the eclipse enabling the inclination of the orbit to be determined. In addition, the shape of

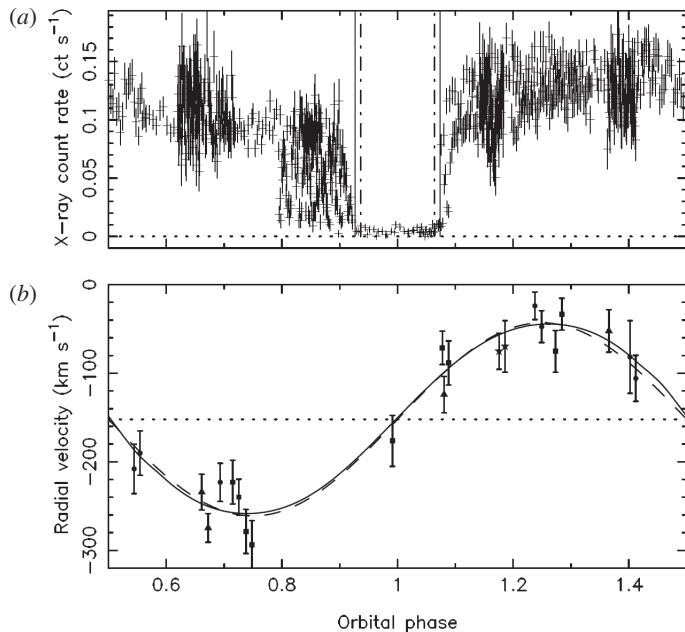


Fig. 13.28 (a) The X-ray light curve of the X-ray binary system M33 X-7 in the nearby spiral galaxy , showing the eclipse of the X-ray source by the luminous primary O star. (b) The velocity curve of the primary star was determined from observations with the Gemini North telescope (Orosz *et al.*, 2007).

the primary star is distorted by the presence of the black hole companion, providing further information about the geometry of the system. With knowledge of the inclination of the orbit of the binary system, the mass of the black hole was found to be $15.65 \pm 1.45 M_{\odot}$, the most massive stellar mass black hole yet identified.

Another black hole binary system of special interest is GRS 1915+105. This source is a *microquasar* in which radio jets are observed to move away from the γ -ray source at superluminal speeds. The source lies in the Galactic plane and so the properties of the companion star have been inferred from observations in the infrared waveband. Infrared spectral observations have shown that the ‘donor’ star has mass about $1 M_{\odot}$ and is a class III giant star. The distance of the source is about 11 kpc through the plane of the Galaxy and so suffers 25–30 magnitudes optical extinction. The period of the binary system was found to be 33.5 days and, from its velocity curve (Fig. 13.29), the mass function was found to be $9.5 \pm 3.0 M_{\odot}$ (Greiner *et al.*, 2001). Assuming the axis of superluminal ejection is perpendicular to plane of the orbit of the binary system, the angle of inclination could be estimated and hence the mass of the black hole to be about $14 \pm 4 M_{\odot}$. In this system, the distance between the donor star and the accreting black hole is about half the distance between the Earth and the Sun. The importance of these observations is that they fully confirm the inference of Mirabel and Rodrigues (1998) that microquasars are small scale versions of the extreme physical phenomena observed in distant quasars – the radio jets are unquestionably powered by accreting black holes.

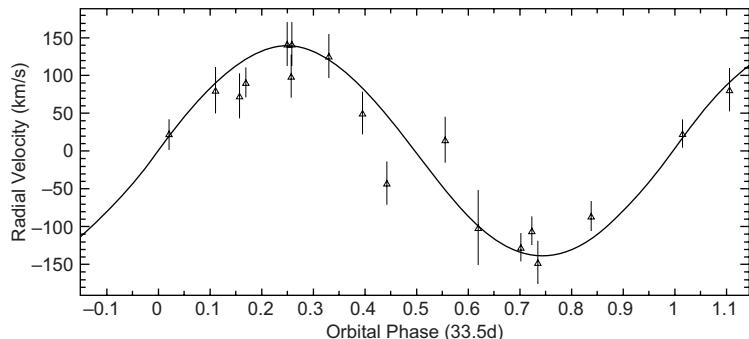


Fig. 13.29 The velocity curve of the microquasar GRS1915+105 as determined from observations with the ISAAC infrared spectrograph on the Antu telescope of the ESO VLT (Greiner *et al.*, 2001).

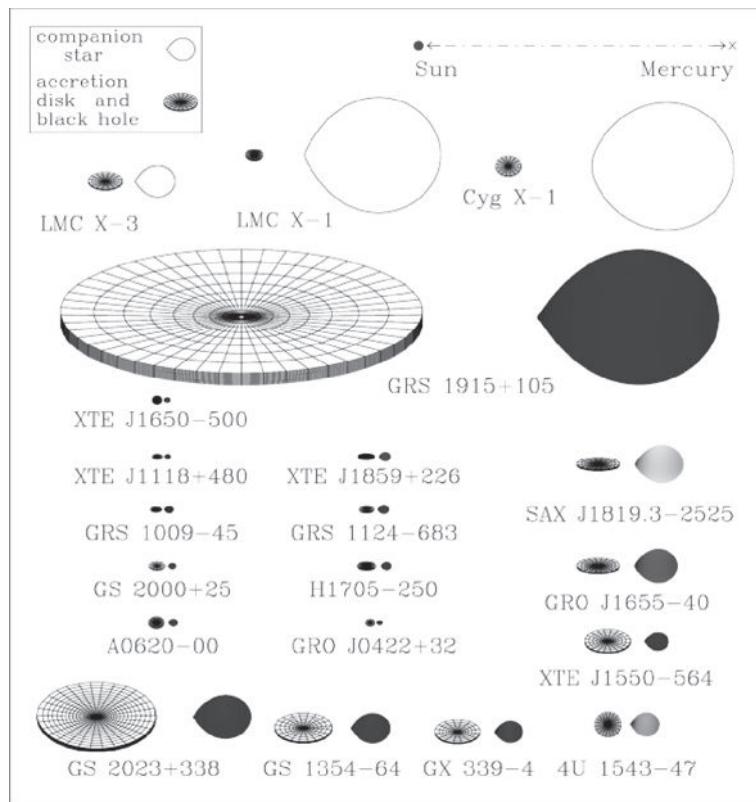


Fig. 13.30 Schematic sketches, to scale, of plausible models for the 20 stellar mass black hole candidates and their associated accretion discs discussed by Orosz (2007).

Orosz (2007) has provided a helpful diagram (Fig. 13.30) showing sketches of 20 dynamically confirmed black hole binaries high mass companions. The 17 objects with low mass companions have the typical temperatures of stars with masses in the range $1 \lesssim M \lesssim 3 M_{\odot}$ while the massive companions of Cyg X-1, LMC X-1 and LMC X-3 all have considerably higher temperatures. Our next task is to understand the physics of the accretion discs inferred to be present in these systems.

14.1 Introduction

Accretion means the accumulation of diffuse gas or matter onto some object under the influence of gravity. Accretion from the interstellar medium onto stationary and moving stars was the subject of a number of pioneering papers by Bondi, Lyttelton and Hoyle and, in the light of subsequent studies, these have proved to provide quite accurate predictions for the rate of accretion (Hoyle and Lyttleton, 1939; Bondi and Hoyle, 1944; Bondi, 1952). The subject was reinvigorated in the 1960s by the realisation that accretion of matter onto supermassive black holes is a remarkably effective means of accounting for the extreme properties of active galactic nuclei and, even more, by the discovery of intense X-ray sources associated with binary systems in our Galaxy. The discovery of these objects and the ensuing flourishing of theory ushered in a new epoch in high energy astrophysics. Accretion was also applied to binary systems involving white dwarf stars and these processes could account for the properties of cataclysmic variables.

Let us begin our analysis by deriving some of the simple relations which show how naturally accretion can account, in principle, for many of the key features of galactic X-ray sources and active galactic nuclei.

14.2 Accretion—general considerations

14.2.1 The efficiency of the accretion process

Consider the accretion of matter onto a star of mass M and radius R . If the matter falls onto the star in free-fall from infinity, it acquires kinetic energy as its gravitational potential energy becomes more negative. Considering a proton falling in from infinity, we can write

$$\frac{1}{2}m_p v_{\text{ff}}^2 = \frac{GMm_p}{r}. \quad (14.1)$$

When the matter reaches the surface of the star at $r = R$, it is rapidly decelerated and, assuming all the matter accumulates on the surface of the star, the kinetic energy of free-fall is radiated away as heat. If the rate at which mass is accreted onto the star is \dot{m} , the rate at which kinetic energy is dissipated at the surface of the star is $\frac{1}{2}\dot{m}v_{\text{ff}}^2$ and hence the

luminosity of the source is

$$L = \frac{1}{2} \dot{m} v_{\text{ff}}^2 = \frac{GM\dot{m}}{R}. \quad (14.2)$$

It is convenient to introduce the Schwarzschild radius, $r_g = 2GM/c^2$, of a star of mass M . Inserting this radius into (14.2),

$$L = \frac{1}{2} \dot{m} c^2 \left(\frac{r_g}{R} \right). \quad (14.3)$$

This expression for the luminosity can be written $L = \xi \dot{m} c^2$, where ξ is the *efficiency of conversion* of the rest-mass energy of the accreted matter into heat. According to the above calculation, the efficiency is roughly $\xi = (r_g/2R)$ and so depends upon how compact the star is. For a white dwarf star with $M = M_\odot$ and $R \approx 5 \times 10^6$ m, $\xi \approx 3 \times 10^{-4}$. For a neutron star with mass $M = M_\odot$ and $R = 10$ km, $\xi \sim 0.15$. In the case of nuclear energy generation, the greatest release of nuclear binding energy occurs in the conversion of hydrogen into helium for which $\xi \approx 7 \times 10^{-3}$. Thus, accretion onto neutron stars is an order of magnitude more efficient as an energy source than nuclear energy generation.

A simple interpretation of (14.3) suggests that we would do even better if the matter were accreted onto a black hole. There is, however, no solid surface onto which the matter can be accreted as there is in the case of a neutron star. If matter fell radially into the black hole, no heat would be released. In practice, what happens is that the infalling matter must acquire some angular momentum and then conservation of angular momentum prevents matter from falling directly into the black hole. The energy release by this process is considered in more detail in Sect. 14.3.3, but first let us give some justification for these statements.

Consider a small mass element m falling from a large distance towards the black hole. Because of small fluctuations in the gravitational potential, the element is bound to acquire some small amount of angular momentum \mathcal{L} with respect to the black hole, $\mathcal{L} = I\Omega = mv_\perp r$. Because of conservation of angular momentum, $mv_\perp r = \text{constant}$, and hence the rotational energy of the element increases as it approaches the hole, $E_{\text{rot}} = \frac{1}{2} I\Omega^2 = \frac{1}{2} \mathcal{L}^2/I \propto r^{-2}$ since $I = mr^2$. Thus, as the mass element approaches the hole, its rotational energy increases more rapidly than its gravitational potential energy which increases only as r^{-1} . This is a familiar result in the analysis of orbits according to Newtonian dynamics (Fig. 13.24a) – no matter how small the angular velocity of a mass element at a very large distance, this is sufficient to prevent collapse to $r = 0$. As discussed in Sect. 13.11.1, this result is modified in the general relativistic treatment because of the presence of an ‘attractive’ $-(2GMh^2/r^3c^2)$ term in a general relativistic treatment (Fig. 13.24a). The net result is that particles with specific angular momenta $h \leq 2r_g c$ can fall directly in to $r = 0$. This is, however, a tiny specific angular momentum relative to the typical values found in, say, mass transfer in close binary systems.

Thus, matter is prevented from falling into the black hole in directions perpendicular to the rotation axis by centrifugal forces. The matter can, however, collapse along the rotation axis of the infalling material so that a disc of matter forms about the black hole. The matter in this *accretion disc* can only fall into the black hole if it loses its angular momentum and this is achieved by viscous forces acting in the disc. The viscosity has two effects. First, it transfers angular momentum outwards as illustrated by the results of an exact model

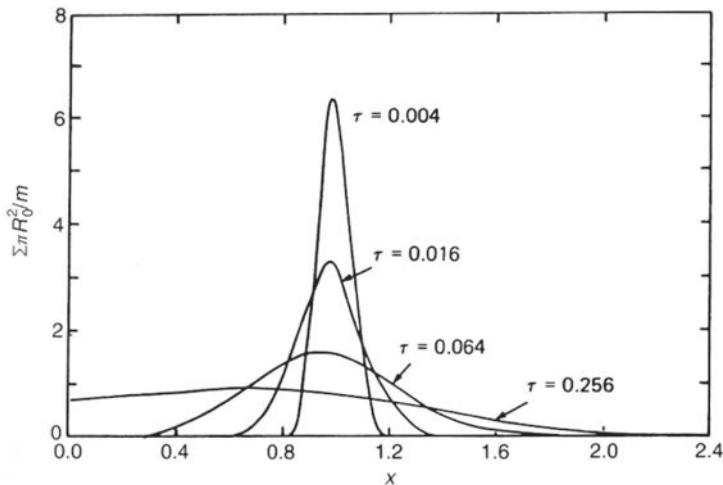


Fig. 14.1 The evolution of a ring of matter of mass m which is placed initially at radius $r = R_0$. The ring spreads out under the action of viscous forces in the ring. The surface density Σ is shown as a function of radius $x = r/R_0$ and the dimensionless time variable $\tau = 12\nu t R_0^{-2}$. ν is the kinematic viscosity of the matter of the disc (Pringle 1981).

calculation of the *spreading* of a thin viscous disc (Fig. 14.1). It can be seen that viscosity redistributes the angular momentum so that some of the matter spreads outwards taking angular momentum with it and thus allows the rest of the matter to spiral inwards (Pringle and King, 2007). At the same time, the viscosity acts as a frictional force which results in the dissipation of heat. The matter in the accretion disc drifts gradually inwards until it reaches the *last stable orbit* about the black hole, at which point it spirals irretrievably into the black hole (Sect. 13.11).

Thus, the maximum energy which can be released by accretion onto black holes is given by the energy which has to be dissipated in order to reach the last stable orbit about the black hole. These binding energies were given in Sects 13.11.1 and 13.11.2. For Schwarzschild black holes, $\xi = 0.06$, whilst for maximally rotating Kerr black holes $\xi = 0.426$. Thus, black holes, and, in particular, maximally rotating black holes, are the most powerful energy sources we know of in the Universe and accretion is the process by which the energy can be released.

14.2.2 The Eddington limiting luminosity

It might seem as though we could generate arbitrarily large luminosities by allowing matter to fall at a sufficiently great rate onto a black hole. There is, however, a limit to the luminosity because, if the luminosity were too great, radiation pressure would blow away the infalling matter. This limiting luminosity, known as the *Eddington luminosity*, is found by balancing the inward force of gravity against the outward pressure of the radiation.

It is assumed that the infalling matter is fully ionised and that the radiation pressure force is provided by Thomson scattering of the radiation by the electrons in the plasma. By assuming that the pressure is due to Thomson scattering, the smallest cross-section for

the processes which impede the loss of radiation from the system is adopted (see Sect. 2.4 and Fig. 2.7). Consider the forces acting on an electron–proton pair at distance r from the source. The inward gravitational force is

$$f_{\text{grav}} = \frac{GM}{r^2}(m_p + m_e) \approx \frac{GMm_p}{r^2}. \quad (14.4)$$

The radiation pressure acts upon the electron but, because the plasma must remain neutral, this pressure is communicated to the protons by the electrostatic forces between the protons and the electrons. Expressing this result in terms of plasma physics, the electrons and protons cannot be separated by more than typically a Debye length $\lambda_D = v_e/\omega_p = 70(T_e/N_e)^{1/2}$ m, where v_e is the thermal speed of the electrons, $\omega_p = (N_e e^2/m_e \epsilon_0)^{1/2}$ is the angular plasma frequency and N_e is the number density of the electrons (Sect. 11.1).

Each photon gives up a momentum $p = h\nu/c$ to the electron in each collision. Therefore, the force acting on the electron is the momentum communicated to it per second by the incident flux density of photons N_{ph} . Thus, $f = \sigma_T N_{\text{ph}} p$, where $\sigma_T = 6.653 \times 10^{-29} \text{ m}^{-2}$ is the Thomson cross-section. The flux density of photons at distance r from the source is $N_{\text{ph}} = L/4\pi r^2 h\nu$, where L is the luminosity of the source, and so the outward force on the electron is $f = \sigma_T L/4\pi r^2 c$. Equating this force to the inward gravitational force, we find

$$\frac{\sigma_T L}{4\pi r^2 c} = \frac{GMm_p}{r^2}; \quad L_E = \frac{4\pi GMm_p c}{\sigma_T}. \quad (14.5)$$

L_E is known as the *Eddington luminosity* and is the maximum luminosity a spherically symmetric source of mass M can emit in a steady state. The limiting luminosity is independent of the radius r and depends only upon the mass M of the emitting region. This argument is closely related to the vibrational instability which leads to the upper mass limit for main sequence stars discussed in Sect. 2.9.1. Stars with mass $M \sim 60M_\odot$ are essentially radiating at the Eddington luminosity, as may be confirmed by comparing the mass–luminosity relation for main sequence stars (Sect. 2.5) with the Eddington limit.

Introducing the Schwarzschild radius $r_g = 2GM/c^2$, the expression for the Eddington luminosity can be rewritten

$$L_E = \frac{2\pi r_g m_p c^3}{\sigma_T} = 1.3 \times 10^{31} \left(\frac{M}{M_\odot} \right) \text{ W}. \quad (14.6)$$

This is an important result. It is interesting to compare the X-ray luminosities of binary X-ray sources in our Galaxy and in the Magellanic Clouds with this limit for typical stellar masses (Fig. 14.2). The X-ray luminosities extend up to more or less the maximum allowed values for masses $M \sim 1-10 M_\odot$. It is possible to exceed the Eddington limit by adopting different geometries for the source region but not by a large factor. Also, in non-steady-state situations such as in supernova explosions, the Eddington limit can be exceeded by a large margin. For many purposes, we will find that the spherically symmetric result provides useful constraints for models of high energy astrophysical phenomena.

The observation of pulsed X-ray emission with pulse periods similar to those of radio pulsars is convincing evidence that the pulsating binary X-ray sources contain rotating, magnetised neutron stars. In this case, we can work out a lower limit to the temperature of the source if it radiates at the Eddington limit. Any source loses energy most efficiently if it

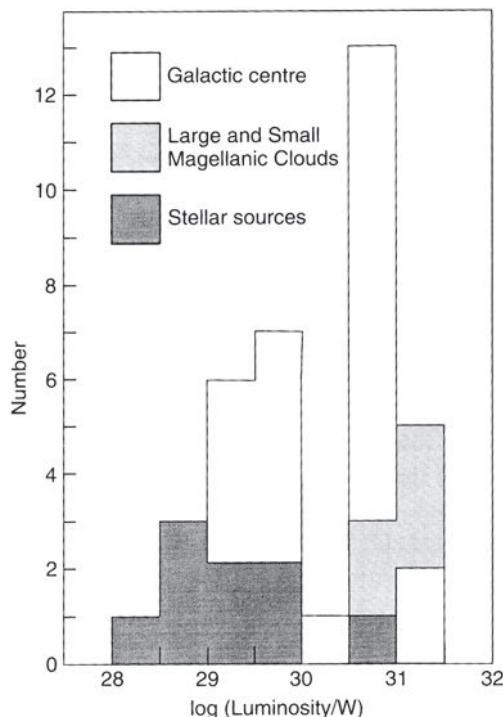


Fig. 14.2 A histogram of the X-ray luminosities of bright X-ray sources in the 1–10 keV energy band in the Galaxy and the Magellanic Clouds. The key to the shading is shown on the diagram. The Eddington limiting luminosity for a $1 M_{\odot}$ object is 1.3×10^{31} W (Margon and Ostriker, 1973).

radiates as a black-body at temperature T . Therefore, equating the black-body luminosity of a solar mass neutron star to the Eddington limit, we find

$$4\pi r_{\text{NS}}^2 a T^4 = L_{\text{Edd}} = 1.3 \times 10^{31} \left(\frac{M}{M_{\odot}} \right) \text{ W}. \quad (14.7)$$

For solar mass neutron stars with radii $r_{\text{NS}} = 10$ km, $T \geq 10^7$ K, and so it is entirely natural that neutron stars accreting at close to the Eddington limit should emit most of their energy at X-ray wavelengths. The same calculation can be performed for accretion onto white dwarfs which have radii about 1000 times greater than those of neutron stars. Then the typical black-body temperatures are about 3×10^5 K and so they are expected to be strong emitters of ultraviolet radiation. This expectation is in good agreement with the observed properties of *cataclysmic variables* which are binary systems in which the compact companion is a white dwarf. These stars provide a paradigm for the processes which may take place in the more extreme cases of accreting neutron stars and black holes.

14.2.3 Black holes in X-ray binaries and active galactic nuclei

The above results can be used to investigate the case of accreting black holes in X-ray binaries and active galactic nuclei. In the case of X-ray pulsars, the evidence is overwhelming that the energy source is accretion – the X-ray pulse periods less than about 1 second are convincing evidence for rotating, magnetised neutron stars. In the case of black holes, there is not expected to be such a clear signature of a compact object but variability might be expected on a time-scale of roughly the light-travel time across the last stable orbit about the black hole. The light-travel time across the Schwarzschild radius of a black hole is

$$T_{\min} \approx \frac{r_g}{c} = 10^{-5} \left(\frac{M}{M_\odot} \right) \text{ s}. \quad (14.8)$$

This is the minimum time-scale which can be associated with an object of mass M , unless we appeal to phenomena such as relativistic beaming which allows the causality limit, $\tau > l/c$, to be violated (see Sect. 23.2). In practice, we would expect variations on time-scales somewhat longer than this value. For example, the last stable orbit about a spherically symmetric black hole lies at a radius $3r_g$ from the centre of the hole.

For solar mass black holes, the expected temperatures of the radiating matter are similar to those given by (14.7) because the last stable orbit occurs at a radius $r = 3r_g = 9(M/M_\odot)$ km and so, if $M \sim M_\odot$, temperatures $T \sim 10^7$ K are expected at the inner edge of the accretion disc. In fact, in the black hole candidates discussed in Sect. 13.11.3, ‘flickering’ of the X-ray intensity is observed on a time-scale of tens of milliseconds (Fig. 14.3a). The light curve of Cygnus X-1 (Cyg X-1) does not show regular pulsations but rather a characteristic flickering pattern of variation. Similar random flickerings are observed in the variable X-ray emission observed in the active galaxies MCG-6-30-15 and NGC 5506 (Fig. 14.3b). These resemble the flickering of Cyg X-1 but on very much longer time-scales as expected according to the scaling relation (14.8). These flickering light curves are interpreted as the signatures of black holes on both the stellar and galactic scale.

The same type of causality argument may be used to investigate the masses of active galactic nuclei in a slightly different way. As is apparent from Fig. 14.3, active galactic nuclei not only exhibit enormous luminosities, but also rapid variability at X-ray and optical wavelengths. Wandel and Mushotzky (1986) estimated the masses of the active galactic nuclei by two independent methods. First, they used a version of (14.8) to find masses from the shortest variability time-scales observed. Second, they made dynamical estimates of the masses of active galactic nuclei from the properties of their broad emission line regions. The clouds responsible for the emission lines are in motion about the nucleus and, by studying the processes of ionisation and excitation of the emission line regions, estimates of the distances of the clouds from the nucleus can be made. Since both the distances of the clouds from the nucleus and their velocities are known, mass estimates can be found by balancing the centrifugal forces acting on the clouds by their gravitational attraction for the nucleus (see Sect. 20.3 for more details).

$$\frac{Gm_{\text{cl}}M}{r^2} = \frac{m_{\text{cl}}v^2}{r}; \quad M = \frac{v^2r}{G}. \quad (14.9)$$

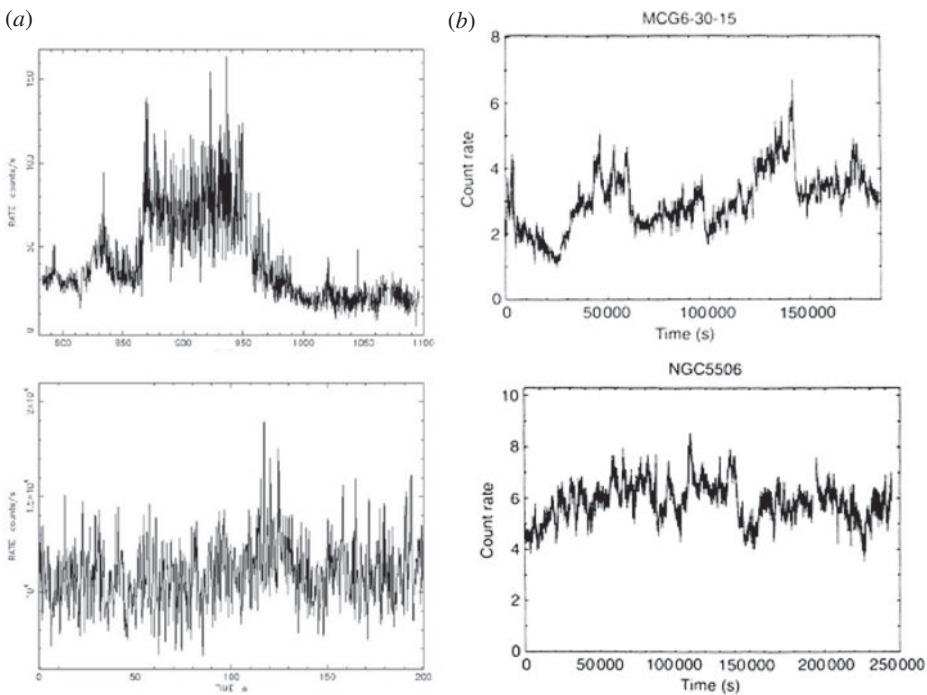


Fig. 14.3 Examples of the ‘flickering’ of the X-ray emission of compact Galactic and extragalactic X-ray sources. (a) The X-ray binary source Cyg X-1 as observed by the Rossi X-ray Timing Explorer satellite showing the X-ray variability from 780 to 1100 days (top) and from 0 to 200 seconds (bottom). Variations on time-scales very much less than one second are observed. (Courtesy of James Lochner, NASA Goddard Space Flight Center.) (b) The active galaxies MCG-6-30-15 (top) and NGC 5506 (bottom) as observed in long exposures by the EXOSAT X-ray telescope showing variability on time scales ranging from minutes to hours (Seward and Charles, 2010).

Wandel and Mushotzky found that, for their sample of active galactic nuclei, these independent estimates of the masses of the central objects were in good agreement (Fig. 14.4a). The masses of the black holes inferred to be present in these active galactic nuclei range from about 10^6 to $10^9 M_\odot$. Furthermore, when they compared the luminosities of the sources with the corresponding Eddington luminosities, none of the nuclei was found to exceed that limit (Fig. 14.4b). Thus, even the most luminous active galactic nuclei could be accounted for by this model. These arguments suggest that we can use these objects to elucidate the behaviour of matter in strong gravitational fields.

Another interesting calculation is to work out the typical temperature of the emitting gas if the radiation originates from roughly the last stable orbit about the black hole by equating the Eddington luminosity to the thermal emission of the disc. For simplicity, it is assumed that the radiation originates from a spherical surface of radius $r = 3r_g = 9(M/M_\odot)$ km about the black hole. Substituting this value of r for r_{NS} in (14.7),

$$T \approx 2 \times 10^7 \left(\frac{M}{M_\odot} \right)^{-1/4} \text{ K}. \quad (14.10)$$

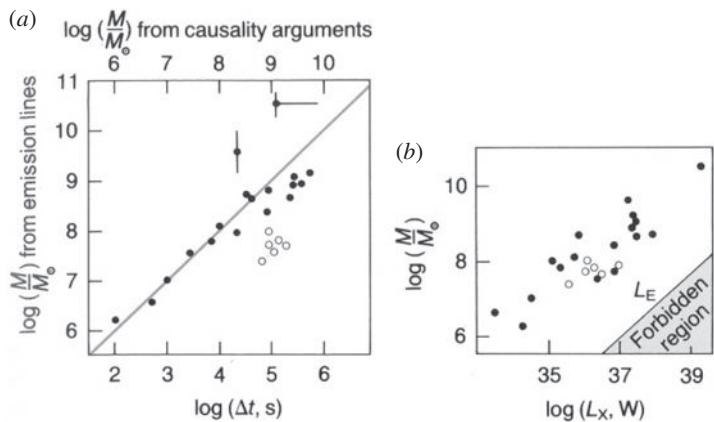


Fig. 14.4 (a) Comparison of mass estimates of active galactic nuclei from the variability of their X-ray emission and from dynamical estimates. The filled circles are quasars and Type I Seyfert galaxies and the open circles are Type II Seyfert galaxies. (b) Comparison of the inferred masses and luminosities with the Eddington limiting luminosity, $L_E = 1.3 \times 10^{31} (M/M_\odot)$. All the active galactic nuclei lie well below the Eddington limit (Wandel and Mushotzky, 1986).

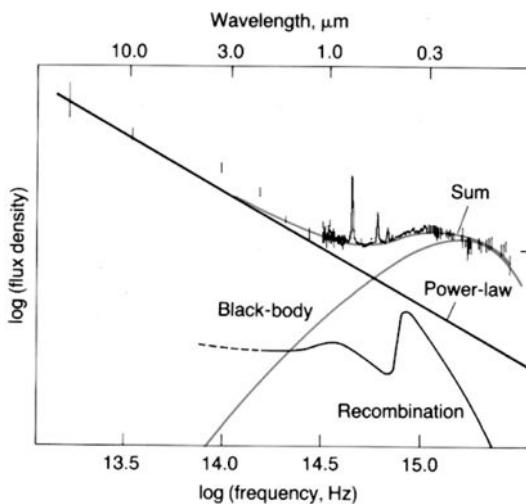


Fig. 14.5 The optical-ultraviolet spectrum of the quasar 3C 273. The continuum has been decomposed into a ‘power-law’ component, a component associated with recombination radiation and a ‘blue-bump’ component which has been represented by a black-body curve. The prominent Balmer series in the optical waveband which led to the discovery of the large redshift of 3C 273 can be seen (Malkan and Sargent, 1982).

This calculation suggests that, for supermassive black holes with $M = 10^8 M_\odot$, the thermal emission would have temperature $T \sim 2 \times 10^5 \text{ K}$. It is significant that many active galactic nuclei have strong ultraviolet continua, 3C 273 being a good example of a galaxy with a ‘blue bump’ (Fig. 14.5). This rough argument certainly cannot be the whole story since these nuclei are just as powerful emitters in the X-ray as in the ultraviolet waveband.

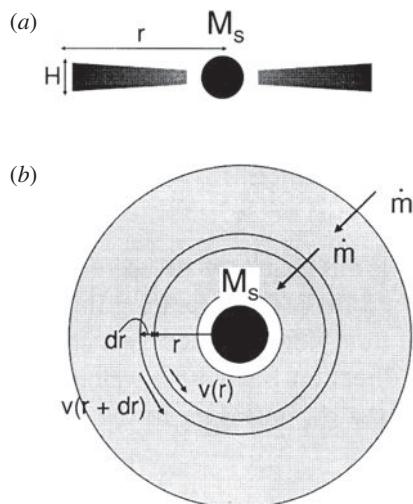


Fig. 14.6 A schematic diagram illustrating the geometry of a thin accretion disc. (a) A side view and (b) a view from above the disc.

14.3 Thin accretion discs

The simplest cases of disc accretion are the thin discs. Many of their essential features are described in *Accretion Power in Astrophysics* by Frank, King and Raine (2002) and in *Astrophysical Flows* by Pringle and King (2007). In the simplest picture, the discs are taken to be in a steady state with a constant rate of mass accretion \dot{m} into the disc. The matter in the disc would take up Keplerian orbits if there were no viscous forces present. These forces are, however, essential in order to transfer the angular momentum of the accreted material outwards and so allow the gas to move inwards to more tightly bound orbits. First of all, let us work out the conditions under which the thin disc approximation is appropriate.

14.3.1 Conditions for thin accretion discs

The geometry of a thin accretion disc is shown in Fig. 14.6. It is assumed that the mass of material in the disc is much less than the mass of the central star, $M_{\text{disc}} \ll M_s$. The condition for hydrostatic support in the direction perpendicular to the plane of the disc, that is, in the z -direction, is

$$\frac{\partial p}{\partial z} = -\frac{GM_s\rho \sin\theta}{r^2}. \quad (14.11)$$

For a thin disc, $\sin\theta \approx z/r$ and, to order of magnitude, $\partial p/\partial z \approx p/H$, where H is the thickness of the disc. Therefore, the condition for hydrostatic support in the z -direction is

$$\frac{p}{H} \approx \frac{GM_s\rho H}{r^3}. \quad (14.12)$$

Although the gas drifts slowly in towards the centre, it moves in roughly Keplerian orbits about the star and so is in centrifugal equilibrium, that is, at any radius,

$$\frac{v_\phi^2}{r} = \frac{GM}{r^2}; \quad \frac{GM}{r} = v_\phi^2. \quad (14.13)$$

Substituting for GM/r from (14.13) into (14.12),

$$\frac{p}{\rho} \approx v_\phi^2 \frac{H^2}{r^2}. \quad (14.14)$$

But, the speed of sound of the material of the disc is $c_s^2 \approx p/\rho$ and therefore

$$\frac{H}{r} \approx \frac{c_s}{v_\phi} = \frac{1}{\mathcal{M}}, \quad (14.15)$$

where \mathcal{M} is the *Mach number* of the rotation velocity of material of the disc relative to the local sound speed in the disc. Thus, the condition that the thin disc approximation can be adopted is that the rotation velocity of the disc is very much greater than the sound speed. To express this condition in other words, internal pressure gradients should not inflate the disc. This is exactly the same condition which is found for the confinement of cool gas to the plane of the disc of a spiral galaxy such as our own.

14.3.2 The role of viscosity—the α parameter

Viscosity plays a central role in determining the structure of accretion discs. Consider the case of unidirectional flow of the fluid in the positive x -direction but with a velocity gradient in, say, the y -direction. The force acting on unit area in the $x-z$ plane, that is the shear stress, is given by the expression

$$f_x(y) = \eta \frac{\partial v_x(y)}{\partial y}, \quad (14.16)$$

where η is defined to be the *dynamic* or *shear viscosity*. In the case of two-dimensional flow, the expression for the shear stress acting at the point (x, y) is

$$f_{xy} = \eta \left[\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right]. \quad (14.17)$$

Because of the symmetry of the stress tensor, $f_{xy} = f_{yx}$ (see, for example, Sect. 15 of Landau and Lifshitz 1987). For an accretion disc in which the gas moves in circular orbits, this expression can be converted to polar coordinates in which $x = r \cos \phi$ and $y = r \sin \phi$. Writing the components of the velocity in cylindrical coordinates,

$$v_x = v_r \cos \phi - v_\phi \sin \phi; \quad v_y = v_r \sin \phi + v_\phi \cos \phi. \quad (14.18)$$

The differentials can be converted into polar coordinates through the relations

$$\frac{\partial}{\partial x} = \cos \phi \frac{\partial}{\partial r} - \frac{\sin \phi}{r} \frac{\partial}{\partial \phi}; \quad \frac{\partial}{\partial y} = \sin \phi \frac{\partial}{\partial r} + \frac{\cos \phi}{r} \frac{\partial}{\partial \phi}. \quad (14.19)$$

Because of the cylindrical symmetry of the problem, it is convenient to evaluate the differentials at $\phi = 0$ in which case

$$f_{xy} = \eta \left[\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right] = \frac{1}{r} \left(\frac{\partial v_x}{\partial \phi} \right)_{\phi=0} + \left(\frac{\partial v_y}{\partial r} \right)_{\phi=0}, \quad (14.20)$$

$$\frac{1}{r} \left(\frac{\partial v_x}{\partial \phi} \right)_{\phi=0} = -\frac{v_\phi}{r}; \quad \left(\frac{\partial v_y}{\partial r} \right)_{\phi=0} = \frac{\partial v_\phi}{\partial r}. \quad (14.21)$$

Hence,

$$f = \eta \left(-\frac{v_\phi}{r} + \frac{\partial v_\phi}{\partial r} \right) = \eta r \frac{\partial}{\partial r} \left(\frac{v_\phi}{r} \right) = \eta r \frac{\partial \Omega}{\partial r}. \quad (14.22)$$

Thus, there is no shear stress if the fluid rotates as a solid body, $\Omega(r) = \text{constant}$.

Next, we work out the torque acting on an annulus of the disc of thickness dr at radius r from the centre (Fig. 14.6b). If the disc has thickness H , the torque acting on the inner edge of the annulus at r is

$$G = \eta r^2 (2\pi r H) \frac{\partial \Omega}{\partial r}. \quad (14.23)$$

The torque on the outer edge of the annulus is

$$G(r + dr) = G(r) + \frac{\partial G}{\partial r} dr, \quad (14.24)$$

and so the net torque acting on the annulus is the difference of these torques $(\partial G / \partial r) dr$. The equation of motion of the annulus is therefore

$$\frac{\partial \mathcal{L}}{\partial t} = \frac{\partial G}{\partial r} dr = \frac{\partial}{\partial r} \left[\eta r^2 (2\pi r H) \frac{\partial \Omega}{\partial r} \right] dr, \quad (14.25)$$

where $\mathcal{L} = 2\pi r^2 H \rho v_\phi dr$ is the angular momentum of the annulus. Hence,

$$r^2 \frac{\partial v_\phi}{\partial t} = \frac{\eta}{\rho} \frac{\partial}{\partial r} \left[r^3 \frac{\partial}{\partial r} \left(\frac{v_\phi}{r} \right) \right]. \quad (14.26)$$

It is conventional to write $\nu = \eta/\rho$ where ν is the *kinematic viscosity* and so the equation of motion for the circumferential velocity v_ϕ is

$$\frac{\partial v_\phi}{\partial t} = \frac{\nu}{r^2} \frac{\partial}{\partial r} \left[r^3 \frac{\partial}{\partial r} \left(\frac{v_\phi}{r} \right) \right]. \quad (14.27)$$

This is the relation we have been seeking. There is, however, a problem because the molecular viscosity is very small and so the *Reynolds number* $\mathcal{R} \approx L^2/vT \approx VL/\nu$ is very large; L , T and V are typical dimensions of length, time and velocity for the flow. If viscous forces are to play an important role, the Reynolds number should be less than unity as, for example, in the flow of a viscous fluid about a cylinder. At large Reynolds numbers, $\mathcal{R} \geq 10^3$, the flow becomes turbulent (Landau and Lifshitz, 1987; Feynman *et al.*, 1965; Batchelor, 1970).

Let us evaluate the Reynolds number for the material of a thin disc. V is the speed of rotation of the material of the disc and L the characteristic scale over which the

velocity changes which is the typical radial distance from the centre of the disc. According to kinetic theory, the dynamic viscosity is $\eta = \rho v = \frac{1}{3} \rho v_s \lambda$ where ρ is the mass density of the gas, v_s is the internal sound speed, which is roughly the same as the typical speed of the particles in the gas, and λ is the mean free path of the particles. To order of magnitude, $\mathcal{R} \sim VL/v_s \lambda$.

Consider the case of an accretion disc about a neutron star of mass $1 M_\odot$ radiating at the Eddington limit. The luminosity of the source is 1.3×10^{31} W and this can be equated to the accretion luminosity $L = \frac{1}{2} \dot{m} c^2 (r_g/r)$, where r is the radius of the region from which most of the luminosity is generated. Adopting $r = 10r_g$, corresponding to three times the radius of the neutron star, the accretion rate is $\dot{m} \sim 10^{15}$ kg s⁻¹. For a thin disc, the accretion rate is also $\dot{m} = 2\pi r \times H \times \rho v_r = 2\pi r H \rho v_r$, where H is the width of the disc at radius r and v_r is the radial inward drift velocity. From (14.15), $H = r/\mathcal{M}$ where \mathcal{M} is the Mach number of the rotation velocity of the disc relative to the sound speed in the disc. The Mach number \mathcal{M} can be estimated from the ratio of the Keplerian velocity of the disc at $R = 10r_g$ and the sound speed from the thermal velocity of hot gas at 10⁷ K. From these, we find $\mathcal{M} \sim 200$. The only remaining quantity to be estimated is the radial drift velocity v_r . This velocity is certainly subsonic with respect to the speed of sound in the disc and so we write $v_r = \beta v_s$ where $\beta \ll 1$. Inserting these values into $\dot{m} = 2\pi r H \rho v_r$, the mass density in the disc is $\rho \sim 0.1 \beta^{-1}$ kg m⁻³. Thus, if $\beta = 0.01$, a typical value found in more detailed models of the accretion discs responsible for the X-ray emission of neutron stars in close binary systems, the mass densities are large, ~ 10 kg m⁻³.

The reason for carrying out this calculation is that the mass density ρ is needed to work out the mean free path λ of the particles in the disc. According to (11.17), the mean free path of a proton is

$$\lambda \approx v_s t_c = 11.4 \times 10^6 \frac{T^{3/2} A^{1/2} v_s}{N Z^4 \ln \Lambda} \text{ m.} \quad (14.28)$$

Inserting these values into the expression for the Reynolds number, $\mathcal{R} \sim VL/v_s \lambda \sim 10^{12}$.

This is a key result for the astrophysics of accretion discs about neutron stars and white dwarfs. The result has two implications. Firstly, molecular viscosity associated with the deflection of charged particles in the plasma cannot play a role in determining the structure of the disc. According to the above analysis, the flow would be strongly turbulent. The result also provides a possible solution to the problem. The generation of turbulence within the disc results in a *turbulent viscosity* which can perform all the functions of molecular viscosity but now the transport of momentum is associated with the motion of turbulent eddies in the plasma. There is also likely to be a magnetic field in the disc and this provides a further means of transporting momentum on a large scale. The turbulent viscosity ν_{turb} of the material of the disc can be written $\nu_{\text{turb}} \sim \lambda_{\text{turb}} v_{\text{turb}}$, where λ_{turb} is the scale of the eddies and v_{turb} is their rotational or ‘turn-over’ velocity.

To overcome this problem, Shakura and Sunyaev (1973) introduced the following prescription for the turbulent viscosity, $\nu = \alpha v_s H$, where v_s is the speed of sound in the disc and H is its scale height in the z -direction. The turbulent eddies must have dimension less than the scale height of the disc and the turn-over velocities must be less than the speed of sound. The advantage of this formalism is that analytic solutions can be found for the

structure of thin accretion discs in terms of the single parameter α – these solutions are often referred to as α -discs. Empirical values for the parameter α can be found once self-consistent solutions of the equations are found and the models compared with the observed properties of X-ray binary systems.

An important advance in understanding the nature of the turbulent viscosity of accretion discs was made by Balbus and Hawley (1991) who showed that if there is a magnetic field in the disc and there is even a small vertical component, the instability first described by Velikhov (1959) and Chandrasekhar (1981) results in the generation of magnetohydrodynamic turbulence in the disc. Frank and his colleagues (2002) provide a helpful physical picture of the nature of the instability. As they note, this is only one way in which turbulent viscosity can be generated.

14.3.3 The structure of thin discs

It is useful to continue with the analysis of the structure of α -discs in more detail since they give insight into the physical content of the models. Consider steady-state discs in which the inflow velocity v_r is determined by the viscosity ν . Conservation of mass implies that in the steady state the mass flow through any radius is a constant,

$$\dot{m} = 2\pi r v_r \int \rho dz = \text{constant} , \quad (14.29)$$

where the integral takes account of inflow through the thickness of the disc. It is convenient to work in terms of the *surface density* of the disc $\Sigma = \int \rho dz$ and so

$$\dot{m} = 2\pi r v_r \Sigma = \text{constant} . \quad (14.30)$$

Firstly, the torque G acting on the cylindrical surface at radius r from the centre of the disc is given by (14.23),

$$G = 2\pi r^3 \nu \Sigma \frac{d\Omega}{dr} . \quad (14.31)$$

From (14.24), the net torque acting on the annulus between radii r and $r + \Delta r$ is

$$G(r) - G(R + \Delta r) = -\frac{\partial G}{\partial r} \Delta r , \quad (14.32)$$

where G is given by (14.31). We now work out the rate of transport of angular momentum through the surface at radius r . The mass transfer per second through the surface is given by (14.30) and therefore the angular momentum transfer through radius r is

$$\dot{m} v_\phi r = 2\pi r^3 \Sigma v_r \Omega . \quad (14.33)$$

In the same way, the angular momentum transfer through radius $r + \Delta r$ is

$$(\dot{m} v_\phi r)_{r+\Delta r} = (2\pi r^3 \Sigma v_r \Omega)_{r+\Delta r} . \quad (14.34)$$

Making a Taylor expansion of (14.34) and subtracting (14.33), the net rate of transfer of angular momentum into the annulus is

$$\Delta \mathcal{L} = 2\pi \frac{d}{dr} (r^3 \Sigma v_r \Omega) \Delta r , \quad (14.35)$$

per unit time. Equating (14.32) and (14.35),

$$\frac{dG}{dr} = 2\pi \frac{d}{dr} (r^3 \Sigma v_r \Omega) . \quad (14.36)$$

Integrating (14.36),

$$G = 2\pi r^3 \Sigma v_r \Omega + C , \quad (14.37)$$

or, using (14.31) for G ,

$$v \Sigma \frac{d\Omega}{dr} = \Sigma v_r \Omega + \frac{C}{2\pi r^3} , \quad (14.38)$$

where C is a constant to be found by the matching of the rotational velocity at the surface of the star to the velocity of rotation at the inner edge of the accretion disc. The matching of these velocities takes place through a boundary layer of thickness b and this results in slight complications which are clearly described by Frank, King and Raine (2002). Within the boundary layer, matter is dragged about the star in solid body rotation and so out to radius $R_* + b$, $d\Omega/dr = 0$. Hence, at the radius $r_* + b$,

$$C = -2\pi(r^3 \Sigma v_r \Omega)_{r_*+b} . \quad (14.39)$$

At this radius, the speed of rotation is the velocity of material in a Keplerian orbit about the star, $\Omega^2 = GM_*/r^3$. Hence,

$$C = -2\pi[r^{3/2} \Sigma v_r (GM_*)^{1/2}]_{r_*+b} . \quad (14.40)$$

But $\dot{m} = 2\pi r v_r \Sigma = \text{constant}$ at any radius and so

$$C = -\dot{m}[GM_*(r_* + b)]^{1/2} . \quad (14.41)$$

Thus, if $b \ll r_*$,

$$C = -\dot{m}(GM_*r_*)^{1/2} . \quad (14.42)$$

The constant C represents the rate of transfer of angular momentum into the boundary layer. In the disc itself, the velocities are close to Keplerian and so, substituting (14.42) and (14.30) into (14.38), we find the pleasant result

$$v \Sigma = \frac{\dot{m}}{3\pi} \left[1 - \left(\frac{r_*}{r} \right)^{1/2} \right] . \quad (14.43)$$

The next result we need is the rate of dissipation of energy by the viscous forces acting in the disc. The expression for the heat generated per unit volume is found in standard text-books.¹ In Cartesian coordinates,

$$-\left(\frac{dE}{dt} \right) = \frac{1}{2} \eta \sum_{i,j} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right)^2 , \quad (14.44)$$

¹ See, for example, Landau and Lifshitz (1987), *Fluid Mechanics*, Sect. 16.

where $\mathbf{v} = (v_x, v_y, v_z)$ is the fluid velocity. This expression can be converted into cylindrical polar coordinates using the relations $x = r \cos \phi$; $y = r \sin \phi$, $z = z$ and $v_x = v_r \cos \phi - v_\phi \sin \phi$, $v_y = v_r \sin \phi + v_\phi \cos \phi$, $v_z = 0$. Therefore, in the case of axial symmetry, the dissipation rate is

$$-\left(\frac{dE}{dt}\right) = \eta r^2 \left(\frac{d\Omega}{dr}\right)^2. \quad (14.45)$$

Integrating over the z -coordinate, the dissipation rate is

$$-\left(\frac{dE}{dt}\right) = \int \eta r^2 \left(\frac{d\Omega}{dr}\right)^2 dz = \nu \Sigma r^2 \left(\frac{d\Omega}{dr}\right)^2. \quad (14.46)$$

Substituting for $\nu \Sigma$ from (14.43) and assuming the orbits in the disc are Keplerian,

$$-\left(\frac{dE}{dt}\right) = \frac{3G\dot{m}M_*}{4\pi r^3} \left[1 - \left(\frac{r_*}{r}\right)^{1/2}\right]. \quad (14.47)$$

This is the result we have been seeking. The energy dissipation rate does not depend explicitly upon the viscosity ν which has been absorbed into the requirement of steady-state accretion at a constant rate \dot{m} . On the other hand, other properties of the disc such as its surface density depend upon ν . In the steady state, this heat energy is dissipated as radiation and so the luminosity of the disc is found by integrating the heat dissipation rate (14.47) from r_* to infinity:

$$L = \int_{r_*}^{\infty} \left(-\frac{dE}{dt}\right) 2\pi r dr = \frac{G\dot{m}M_*}{2r_*}. \quad (14.48)$$

This is a sensible result. The matter falling in from infinity passes through a series of bound Keplerian orbits for which the kinetic energy is equal to half the gravitational potential energy. The matter dissipates half its potential energy in falling from infinity to radius r and this is the source of the luminosity of the disc. When the matter reaches the boundary layer, it has only liberated half its gravitational potential energy. If the matter is then brought to rest on the surface of the star, the rest of the gravitational potential energy can be dissipated. Thus, the boundary layer can be just as important a source of luminosity as the disc itself.

This result can be written in terms of the rate of dissipation of energy, or luminosity $L(r) \Delta r$, in an annulus between radii r and $r + \Delta r$. Multiplying (14.47) by $2\pi r \Delta r$, the luminosity of the annulus is

$$L(r) \Delta r = -\left(\frac{dE}{dt}\right) = \frac{3G\dot{m}M_*}{2r^2} \left[1 - \left(\frac{r_*}{r}\right)^{1/2}\right] \Delta r. \quad (14.49)$$

This expression shows that more than just the release of gravitational binding energy is involved when matter moves from a Keplerian orbit at $r + \Delta r$ to one at r which would be $(G\dot{m}M_*/2r^2) \Delta r$. The difference between this expression and (14.49) represents the net flow of energy into the annulus Δr associated with the transport of angular momentum outwards. Thus, although the total energy released in reaching the surface of the star is half the gravitational potential energy, at any radius, the energy dissipation rate consists of both the energy loss due to angular momentum transport as well as the release of gravitational

binding energy. From (14.49), the energy dissipation rate at distances $r \gg r_*$ is

$$L(r) \Delta r = - \left(\frac{dE}{dt} \right) = \frac{3G\dot{m}M_*}{2r^2} \Delta r , \quad (14.50)$$

which is three times the rate of release of gravitational binding energy.

14.3.4 Accretion discs about black holes

Next, the analysis can be extended to the case of thin accretion discs about black holes. The physics is exactly the same as above, the only issue being the boundary condition which replaces (14.42). In the case of black holes, matter drifts inwards through the accretion disc until it reaches the last stable orbit, at which radius the material spirals into the black hole. As discussed in Sect. 13.11.1, the angular momentum now helps rather than hinders the collapse of matter into the black hole since, crudely speaking, the rotational energy contributes to the inertial mass of the infalling matter. Suppose the last stable orbit has radius r_1 which for a non-rotating, spherically symmetric black hole would be $r_1 = 3r_g = 6GM/c^2$. According to a classical calculation, the condition that the matter reaches the last stable orbit is that the rotational energy of the matter should be less than or equal to half the gravitational potential energy, that is, $\frac{1}{2}\mathcal{L}^2/I \leq \frac{1}{2}GMm/r_1$ where \mathcal{L} is the angular momentum of a mass element m and $I = mr_1^2$ is its moment of inertia about the hole on the last stable orbit. It is convenient to work in terms of the *specific angular momentum* $J = \mathcal{L}/m$. Therefore, the condition that the matter falls into the hole is

$$J \leq (GMr_1)^{1/2} . \quad (14.51)$$

The inner boundary condition is therefore that the angular momentum with which the material arrives at the last stable orbit should be less than this critical value. The angular momentum, as well as the matter, is consumed by the black hole. It is conventional to write the angular momentum of the material as it arrives at r_1 as $\mathcal{L} = \beta\dot{m}(GMr_1)^{1/2}$, where $\beta \leq 1$. This boundary condition replaces (14.42). Therefore, for thin accretion discs about black holes, the luminosity of the disc between radii r and $r + \Delta r$ is

$$L(r) \Delta r = - \left(\frac{dE}{dt} \right) = \frac{3G\dot{m}M_*}{2r^2} \left[1 - \beta \left(\frac{r_1}{r} \right)^{1/2} \right] \Delta r . \quad (14.52)$$

The total luminosity is found by integrating (14.52) from r_1 to infinity,

$$L = \left(\frac{3}{2} - \beta \right) \frac{G\dot{m}M_*}{r_1} . \quad (14.53)$$

14.3.5 The temperature distribution and emission spectra of thin discs

Let us now make a simple estimate of the temperature distribution and emission spectrum of a thin accretion disc. It is assumed that the disc is optically thick to radiation and that there are sufficient scatterings to ensure that the emission can be approximated as black-body radiation at each point in the disc. The disc radiates from its top and bottom surfaces and so the heat dissipated between r and $r + \Delta r$ (14.53) can be equated to the black-body

emission from these surfaces, that is, to $2\sigma T^4 \times 2\pi r \Delta r$, where σ is the Stefan–Boltzmann constant:

$$\sigma T^4 = \frac{3G\dot{m}M_*}{8\pi r^3}; \quad T = \left(\frac{3G\dot{m}M_*}{8\pi r^3\sigma} \right)^{1/4}. \quad (14.54)$$

At radii $r \gg r_*$, the temperature of the disc increases towards the centre as $T \propto r^{-3/4}$. Hence, the sound speed in the disc increases towards the centre and more detailed calculations are needed to ensure that the thin disc approximation remains valid.

Assuming that each annulus of the disc radiates like a black-body at the temperature given by (14.54), the form of the integrated spectrum of the disc can be derived. The total intensity of the disc is proportional to the surface area at temperature T times the black-body intensity at that temperature. Therefore,

$$I(\nu) \propto \int_{r_1}^{r_{\max}} 2\pi r B(T, \nu) dr, \quad (14.55)$$

where $B(T, \nu) \propto \nu^3 [\exp(h\nu/kT) - 1]^{-1}$ is the Planck function. But, $T \propto r^{-3/4}$ and so $dr \propto (1/T)^{1/3} d(1/T)$. Therefore, the integral over r can be converted into an integral over $(1/T)$. Preserving the dependence upon frequency, we find

$$I(\nu) \propto \int_{T^{-1}(r_1)}^{T^{-1}(r_{\max})} \left(\frac{1}{T} \right)^{4/3} \nu^3 \left[\exp \left(\frac{h\nu}{kT} \right) - 1 \right]^{-1} \left(\frac{1}{T} \right)^{1/3} d \left(\frac{1}{T} \right). \quad (14.56)$$

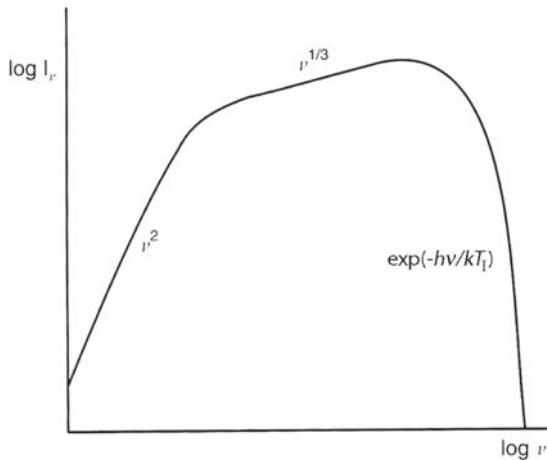
Changing the variable of integration from $(1/T)$ to $x = (h\nu/kT)$,

$$I(\nu) \propto \nu^{1/3} \int_{x(r_1)}^{x(r_{\max})} x^{4/3} (\exp x - 1)^{-1} x^{1/3} dx. \quad (14.57)$$

The definite integral is a constant, all the dependences upon frequency are outside the integral and so $I(\nu) \propto \nu^{1/3}$. Thus, in the black-body approximation, the predicted spectrum of a thin, optically thick accretion disc should have the form $I(\nu) \propto \nu^{1/3}$ between the frequencies corresponding to r_1 and r_{\max} as illustrated in Fig. 14.7. At frequencies less than that corresponding to the temperature of the disc at r_{\max} , the spectrum tends towards a Rayleigh–Jeans spectrum $I_\nu \propto \nu^2$, while at high frequencies the spectrum decreases exponentially as $\exp(-h\nu/kT_1)$.

14.3.6 Detailed models of thin discs

The results obtained in Sects 14.3.1–14.3.5 are almost as far as we can go without solving the equations for realistic models of thin accretion discs. Shakura and Sunyaev (1973) showed that, adopting the α -disc prescription for the turbulent viscosity, eight equations can be derived which can be solved in closed form in terms of eight unknown parameters. This analysis is straightforward, if lengthy. The results are quoted and discussed in some detail by Frank, King and Raine (2002). Many of the properties of thin discs turn out to depend only weakly upon the viscosity parameter α , which is encouraging from the point of view of the confrontation with observation, but disappointing from the perspective of understanding more about the nature of the viscosity in the disc.

**Fig. 14.7**

A schematic representation of the emission spectrum of an optically thick accretion disc. The exponential cut-off at high energies occurs at frequency $\nu = kT_l/h$, where T_l is the temperature of the innermost layers of the thin accretion disc. At frequencies less than that corresponding to the temperature of the disc at r_{\max} , the spectrum tends towards a Rayleigh–Jeans spectrum $I_\nu \propto \nu^2$.

An important aspect of these calculations is the opacity of the material of the disc. The same considerations concerning the appropriate opacities for different ranges of density and temperature discussed in Sect. 2.4 in the context of stellar structure come into play. For temperatures in the range $10^4.5$ – 10^7 K, the Rosseland mean opacity is dominated by bound–free and free–free absorption, what is referred to as *Kramers opacity*, for which $\kappa \propto \varrho T^{-7/2}$. At higher temperatures, the opacity is dominated by electron scattering. Frank, King and Raine (2002) provide a helpful diagram which shows the conditions under which the different opacities are important for thin accretion discs about white dwarfs, neutron stars and black holes for different assumed values of the accretion rate (Fig. 14.8). The diagonal lines indicate the transitions from one regime to another. From right to left, the line $T_c = 10^4$ K corresponds to the transition between temperatures at which Kramers opacity is dominant and those at which molecular and atomic processes are more important. The line $\kappa(\text{electron scattering}) = \kappa_R(\text{Kramers})$ is the locus at which electron scattering and Kramers opacity are equally important. The line $P_r = P_g$ shows the locus to the left of which the radiation pressure exceeds the gas pressure.

It can be seen that Kramers opacity is always the dominant process in accretion onto white dwarfs. In the case of accretion onto neutron stars and black holes, the disc can be divided into three regions, the outer region in which the gas pressure is much greater than the radiation pressure and the opacity is dominated by Kramers opacity, an intermediate region in which the gas pressure is still dominant but the source of opacity is electron scattering and an inner region in which the radiation pressure is dominant and electron scattering is the most important source of opacity (Shakura and Sunyaev, 1973; Novikov and Thorne, 1973). As a result, in the innermost regions, the emitted spectrum of the disc cannot be approximated by a black-body spectrum. In these regions, the disc may be optically thin to

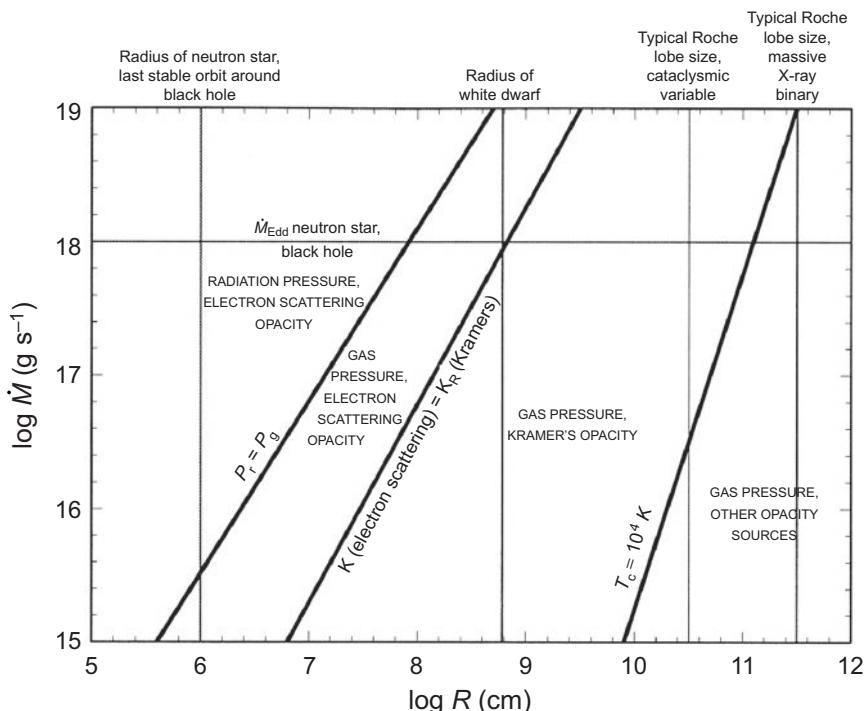


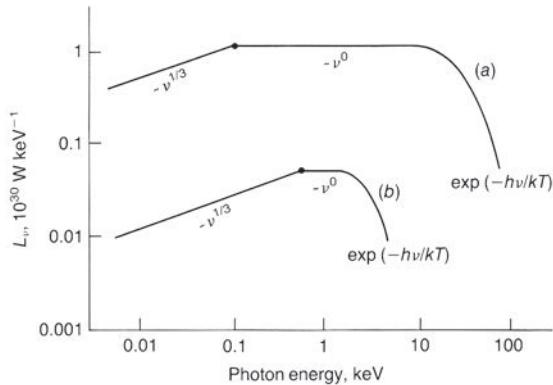
Fig. 14.8

The physical regimes for steady α -discs about $1 M_\odot$ objects as a function of the mass accretion rate (Frank *et al.*, 2002). The radii of solar mass white dwarfs, neutron stars and black holes are indicated, as well as the dimensions of the Roche lobes of the binary star systems involved in massive X-ray binaries and symbiotic stars. The regions of the discs in which gas or radiation pressure dominate are shown, as well as those regions in which Kramers opacity and electron scattering are dominant. The horizontal line shows the accretion rate for a solar mass neutron star or black hole radiating at the Eddington limit.

free-free absorption, even when multiple scattering of the radiation by the electrons is taken into account and then the predicted emission continuum spectrum is flatter, characteristic of optically thin bremsstrahlung, $I(\nu) \propto \nu^0$ (Sect. 6.3). Examples of the predicted spectra for two assumptions about the accretion rate and the viscosity parameter are shown in Fig. 14.9.

14.4 Thick discs and advective flows

Figure 14.8 shows that, in the cases of neutron stars and black holes, if the accretion rates are high, the radiation pressure towards the inner edge of the accretion disc exceeds the thermal pressure. As discussed by Frank, King and Raine (2002), the balance between radiation pressure and gravity then resembles the case of the Eddington luminosity and the thickness of the disc is determined by exactly the same condition. In addition, close to the inner edge of the accretion disc, $r \approx H$, and so the thin disc approximation breaks down.

**Fig. 14.9**

The spectra of the radiation emitted by an accretion disc about a spherical, non-rotating black hole (Shakura and Sunyaev, 1973). In model (a), the black hole mass is $1 M_\odot$, the accretion rate $10^{-8} M_\odot \text{ year}^{-1}$ and the accretion takes place at the Eddington limiting luminosity. In model (b), the mass accretion rate is lower, $10^{-10} M_\odot \text{ year}^{-1}$ and the luminosity is 10^{29} W . The radiation generated in the outer cool regions of the disc result in a power-law spectrum $L(\nu) \propto \nu^{1/3}$. In the inner regions, electron scattering is the dominant source of opacity and the spectrum is approximately independent of frequency. The temperature of the exponential tail corresponds to the surface temperature of the inner regions of the disc modified by the effects of electron scattering (Shapiro and Teukolsky, 1983).

To express this in another way, the rotational velocity becomes comparable to the velocity of sound and so the disc inflates, creating a thick disc. Consequently, many of the other approximations break down in these regions.

Thick discs represent a very much more complicated problem than the thin discs and no fully self-consistent solution has yet been found. Figure 14.10 shows the structure of a Newtonian rotationally supported vorticity-free torus, meaning that the angular velocity depends upon radius as $\Omega(r) \propto r^{-2}$. The lines show the equipotential surfaces of such structures. Stability analyses of such tori have shown, however, that they are globally unstable (Papaloizou and Pringle, 1984; Pringle and King, 2007), meaning that they are subject to instabilities similar to those found in the density wave theory of spiral structure. An enormous amount of analysis has been carried out on these structures, but a discussion of these would go far beyond the scope of this text.² Nonetheless, these structures have a number of attractive features for galactic X-ray sources with jets and for active galactic nuclei. Thick discs have ‘funnels’ along their rotation axes and these may be related to the collimation of beams of particles and relativistic gases observed to be ejected at very high velocities in Galactic and extragalactic sources (Fig. 14.10). A discussion of the problems involved in the construction of thick disc models is given by Frank, King and Raine (2002).

Another major topic in the astrophysics of accretion flows concerns the role of *advection* of mass and energy. The above analysis has been entirely concerned with the diffusive transport of energy and angular momentum mediated by molecular or turbulent viscosity. In addition, energy can be transported in bulk by *convective* and *advection*

² See, for example, the reviews by Narayan and Goodman (1991) and Narayan (1991).

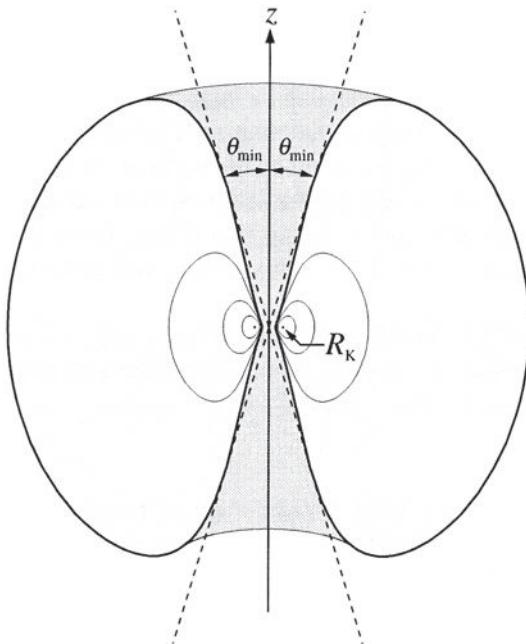


Fig. 14.10 A perspective view of a Newtonian vorticity-free torus for which $\Omega(r) \propto r^{-2}$. The toroidal lines are equipotential surfaces. The diagram shows the axial ‘funnels’ which may be associated with the collimation of jets from compact sources. R_K is the radius at which the angular velocity attains the Keplerian velocity about the compact object (Frank *et al.*, 2002).

processes. The term *advective* refers to the transport of energy and mass in bulk by large scale motions in the fluid. A simple example, already discussed in Sect. 11.2, is *magnetic flux freezing* in a perfectly conducting plasma in which the magnetic field lines are advected by motions in the plasma – the magnetic field follows the bulk motions in the plasma. The term *convection* is used to describe flows in which both diffusive and advective motions are important.

Advective flows are of special interest in the case of accretion onto black holes. In the case of advective flows onto objects with a ‘solid surface’, kinetic energy is released as heat when the flow encounters the surface and so, although the details of the flow can become complex, the standard rate of energy release by accretion comes about through the dissipation of the kinetic energy of the advected material. In the case of black holes, however, there is no solid surface and the infalling matter can be advected through the black hole horizon without releasing the kinetic energy of the flow. Thus, under conditions in which advective motions dominate the flow close to the last stable orbit, the luminosity of the source can be significantly less than that of a body with a solid surface. These considerations are of particular interest in the context of accretion onto supermassive black holes in active galactic nuclei and we take up this topic in more detail in the discussion of Sect. 20.7, which is based on the presentation of Frank, King and Raine (2002).

14.5 Accretion in binary systems

14.5.1 Binary star systems

The discussion of stellar structure and evolution in Chap. 2 referred almost entirely to single stars. We know, however, that a large fraction of the stars in our Galaxy belong to binary systems and this can strongly affect their evolution, particularly if they are members of close binary systems. Double star systems can be detected either as *visual binaries* or, if spectroscopic observations show periodic variations of the radial velocities of one or both stars, as *spectroscopic binaries*. The periods of binary stars can range from a few hours in the case of close binary systems to thousands of years, the upper limit being set by the limited period over which precise observations have been carried out (Griffin, 1985). More than half the stars in our Galaxy are members of binary systems and the determination of their orbits provides a means of estimating stellar masses.

From the perspective of the present study, the close binary systems are of the greatest interest since their close proximity affects their evolution. The separations of close binaries can range from a few times the radii of the stars to systems in which the stars share a common envelope – the latter are known as *contact binaries*. Close binary systems are known containing a wide range of stellar types, from massive binaries containing O and B stars, through intermediate mass binaries, to systems containing compact stars, either white dwarfs, neutron stars or black holes, and to systems such as the binary pulsars consisting of pairs of neutron stars.

The most instructive way of understanding how binary stars evolve is through consideration of the equipotential surfaces of the rotating system. In a frame of reference rotating with the binary system, a centrifugal potential associated with their binary motion is added to the gravitational potential of each star. At radial distance r from the centre of momentum of the binary system, the equipotential surfaces are defined by

$$\phi = \frac{GM_1}{r_1} + \frac{GM_2}{r_2} - \Omega^2 r^2 = \text{constant}, \quad (14.58)$$

where r_1 and r_2 are the distances from the centres of the stars of masses M_1 and M_2 to the point at r . This results in the forms of equipotential surface shown in Fig. 14.11. There is a critical equipotential surface encompassing both stars which is referred to as the *Roche lobe* of the binary system. The equipotential surfaces within the Roche lobe show that the shapes of the stars are significantly distorted from spheres if they fill a significant fraction of the Roche lobe.

In the extreme case of contact binary systems, the common envelope of the binary lies outside the Roche lobe. These binaries have very short periods, less than about half a day, and have distinctive light curves which resemble sine waves rather than typical eclipse light curves. Examples of this type of binary are the *W UMa-type binaries*. The common envelope leads to a quite different internal structure, for example, the stars having a common convective envelope. This results in a number of important differences as compared with the

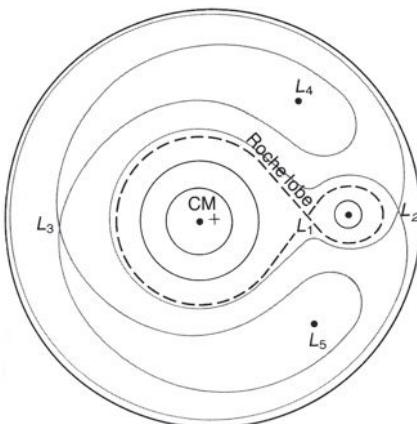


Fig. 14.11 Illustrating the equipotential surfaces of a binary star system in the rotating frame of reference. The equipotential surfaces shown correspond to the sum of the equipotential surfaces for each star plus a centrifugal potential term to take account of the rotation of the stars about their common centre of momentum (CM). In this example, the mass ratio of the stars is 10:1. These surfaces define the shapes of the stars in the binary system. The equipotential surface which connects both stars is known as the *Roche lobe* of the system, shown by the dashed line, and the point at which the two lobes touch is called the *inner Lagrangian point* L_1 . Other turning points in the value of the potential are labelled L_2 , L_3 , etc. In close binary stars, the common surface of the stars may lie outside the Roche lobe (Shapiro and Teukolsky, 1983).

properties of isolated stars. For example, the mass–luminosity relation becomes $L \propto M$, rather than $L \propto M^4$ as found for isolated main sequence stars.

For non-contact close binaries, the stars do not fill their Roche lobes and the stars evolve more or less as normal stars. Interesting phenomena occur as the stars evolve off the main sequence. The more massive of the pair evolves off the main sequence at an earlier time than the less massive star and, as it becomes a red giant, expands to fill its Roche lobe. Matter seeks the lowest gravitational potential and this is achieved if matter passes through the Lagrangian point L_1 onto the secondary companion. In this way, the mass of what was initially the less massive star increases at the expense of the more massive star. In the case of massive binaries, this can lead to the secondary component becoming more massive than the primary. Sequences of events such as this can lead to apparently anomalous situations in which a low mass white dwarf is found as a companion to an intermediate or high mass star.

There are many variations upon this theme of mass transfer in close binary systems (Fig. 14.12). What was initially the secondary star may now evolve into a red giant and the reverse process of mass transfer back onto the original primary can occur. The end point could be the formation of a binary white dwarf or a white dwarf–neutron star pair. Another intriguing variation occurs if one of the stars undergoes a supernova explosion. These explosions may be associated with either low or high mass stars as described in Sect. 13.1. In the process, a considerable amount of mass is ejected from the system and the binary may either remain bound or unbound. If the less massive star explodes due, for example, to mass

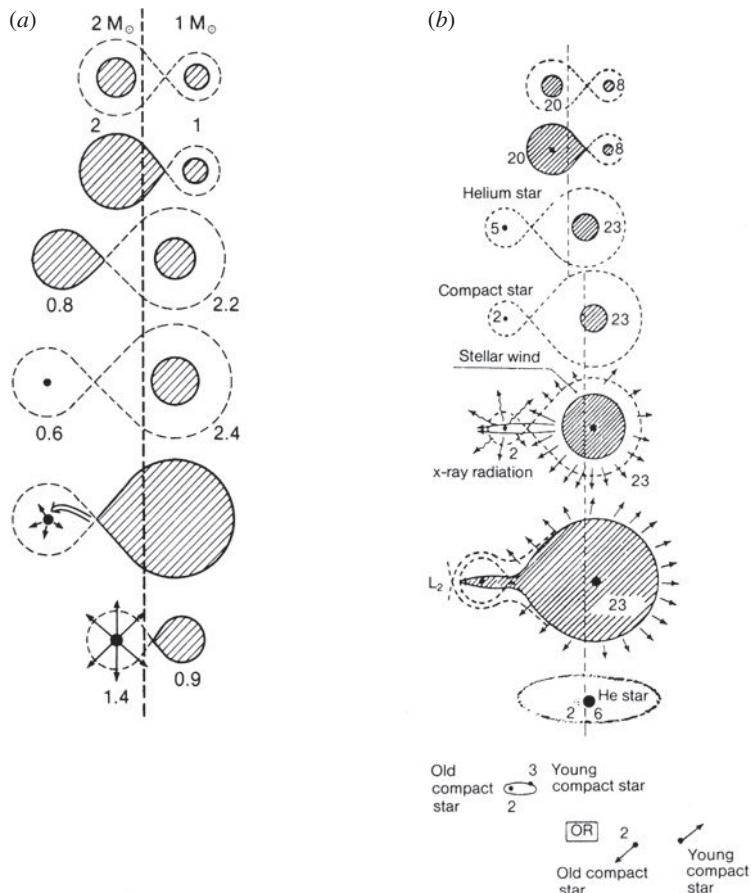


Fig. 14.12 Examples of the evolution of stars in binary systems:

(a) A low mass binary. The stages of evolution are as follows. (i) Both stars on the main sequence; (ii) mass transfer from the more massive of the stars to the lighter once the former becomes a red giant; (iii) formation of a light subgiant and a more massive main sequence companion; (iv) formation of a white dwarf and a main sequence stars; (v) mass transfer onto the white dwarf when the second star becomes a giant; (vi) the white dwarf exceeds the Chandrasekhar mass and explodes as a Type Ia supernova.

(b) A massive binary. (i) Main sequence phase; (ii) the more massive star fills the Roche lobe as it becomes a red giant and mass transfer begins; (iii) end of the first mass transfer phase; the first Wolf-Rayet phase begins; (iv) the helium star has exploded leaving $2 M_{\odot}$ remnant which may be a neutron star or a black hole; (v) the massive star becomes a giant star and mass is transferred onto the compact star which becomes a strong X-ray source; (vi) large scale mass-loss onto the secondary star begins; (vii) the second Wolf-Rayet phase in which the more massive star has lost its envelope leaving a roughly $6 M_{\odot}$ helium star ; (viii) the $6 M_{\odot}$ star explodes as a supernova. The binary may or may not be disrupted in the explosion depending upon the mass of the remnant (Karttunen *et al.*, 2007)

transfer from the primary, the system can remain bound and this provides a mechanism for creating a binary system containing a neutron star. If the more massive star explodes as a Type II supernova, the system may be disrupted and the components disperse at high velocities, roughly those of the stars in the binary orbit. This is a plausible explanation for the fact that radio pulsars are observed to have large proper motions, corresponding to transverse velocities of typically 300 km s^{-1} . Asymmetric stellar collapse may also be important in accounting for some of the extreme velocities observed (Woosley and Janka, 2005).

Various classes of binary system have been associated with most of these permutations. For example, the *symbiotic stars* are associated with binaries in which mass transfer occurs between a red giant star and a compact companion which may be a dwarf main sequence star or a white dwarf. The *cataclysmic variables* probably consist of semi-detached binaries in which the ‘cataclysmic’ variability is associated with an accretion disc about a white dwarf. The *X-ray binaries* consist of a main sequence star and a neutron star or black hole.

As mass transfer takes place, the inner regions of the star are exposed and material can be deposited onto the surface of the secondary component. These processes provide an explanation for some of the abundance anomalies found in the surfaces of some binary stars. It is also a mechanism by which the *Wolf-Rayet stars* can lose their hydrogen envelopes, exposing the products of hydrogen and helium burning.

14.5.2 Feeding the accretion disc

In binary systems, the process of accretion may occur in two ways which are illustrated in Fig. 14.13. All normal stars emit stellar winds, ‘quiescent’ mass loss in the case of stars like the Sun and much more violent winds in the cases of luminous O and B stars in which mass loss rates as high as $10^{-5} M_{\odot} \text{ y}^{-1}$ are observed (Sect. 2.9). In one picture, the compact companion is embedded in an outflowing stellar wind (Fig. 14.13a). As in the case of the Earth’s magnetosphere, there is a bow shock in the upstream direction from the compact star (Sect. 11.4) and accretion onto the star takes place within this cavity. The accretion process may approximate more closely to spherical rather than to thin disc accretion.

The second model illustrated in Fig. 14.13b involves *Roche lobe overflow*. The equipotential surfaces of a close binary star system are distorted in the rotating frame of reference when the stars fill a substantial fraction of the Roche lobe (Fig. 14.11). If the normal star fills its Roche lobe in the course of evolution, matter is transferred through the inner Lagrangian point L_1 as the overflowing matter seeks a lower gravitational potential. As illustrated in Fig. 14.13b, the result is a stream of plasma flowing from the inner Lagrangian point into the accretion disc. It is therefore expected that there will be a ‘hot spot’ at the point where the accreting matter meets the accretion disc and there is evidence for such a spot in observations of certain cataclysmic variable stars.

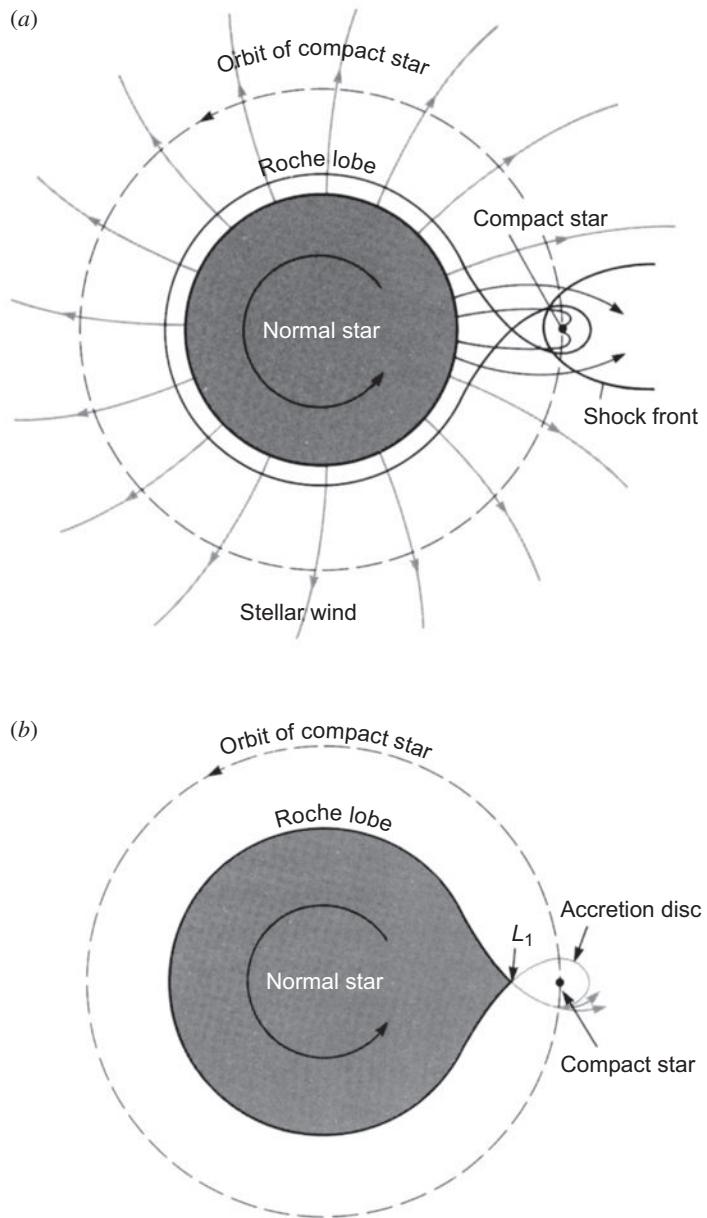


Fig. 14.13 Illustrating two ways in which accretion onto stars in binary systems may take place. In (a), the massive star has a strong stellar wind and the compact star is embedded in the outflow from it. In (b), the normal star expands to fill its Roche lobe and matter passes through the Lagrangian point L_1 onto the compact star. An accretion disc is formed about the compact star (Shapiro and Teukolsky, 1983).

14.5.3 The role of magnetic fields

So far, we have neglected the role of magnetic fields except to remark that they are almost certainly implicated in the turbulent viscosity within the accretion disc. The compact stars themselves possess magnetic fields and these can strongly influence the accretion of matter onto the compact star. Direct evidence for the presence of strong magnetic fields in X-ray binaries is provided by the observation of electron cyclotron radiation features in the X-ray spectra of X-ray binaries (Sect. 8.2). If the magnetic field is of dipolar form, the magnetic flux density at distance r from the centre of the star is given by (13.40). Approximately this can be written in the form $B \approx (R_*/r)^3 B_s$, where B_s is the typical magnetic flux density at the surface of the compact star at radius R_* .

Consider the case of spherical accretion onto the compact star. The magnetic pressure at radial distance r is $p_{\text{mag}} \approx B^2/2\mu_0 \approx (B_s^2/2\mu_0)(R_*/r)^6$. The infalling matter can be considered to have fallen from rest at infinity and so its infall velocity at radius r is $v = (2GM_*/r)^{1/2}$. The pressure which this infalling gas exerts upon the magnetic field is known as the *ram pressure* p_{ram} and is the rate at which momentum is transported inwards at radius r per unit area, that is, $p_{\text{ram}} = \rho v^2$. The same type of pressure acts at the interface between the Solar Wind and the Earth's magnetosphere and is also responsible for the containment of the high energy particles in extended extragalactic radio sources. This ram pressure is balanced by the magnetic pressure of the compact star at a radius r_M , known as the *Alfvén radius*, such that

$$\rho v^2 = \frac{B_s^2}{2\mu_0} \left(\frac{R_*}{r_M} \right)^6. \quad (14.59)$$

In the case of spherically symmetric accretion, the mass accretion rate is $\dot{m} = 4\pi r^2 \rho v$ and so (14.59) can be reorganised to find r_M in terms of the accretion rate:

$$r_M = \left(\frac{2\pi^2}{G\mu_0^2} \right)^{1/7} \left(\frac{B_s^4 R_*^{12}}{M_* \dot{m}^2} \right)^{1/7}. \quad (14.60)$$

For a solar mass neutron star accreting at the Eddington luminosity, $L = \dot{m}\eta c^2 = 1.3 \times 10^{31}$ W. Adopting $\eta = 0.1$, $B_s = 10^8$ T and $R_* = 10$ km, we have $r_M = 10^3$ km, that is, about 100 times the radius of the neutron star. Thus, in the case of luminous X-ray sources, the immediate vicinity of the neutron star is magnetically dominated. Matter can, however, be accreted onto the surface of the neutron star and hence releasing the maximum amount of binding energy of the infalling matter, if the matter flows along the magnetic field lines onto the poles of the rotating neutron star, as illustrated in Fig. 14.14. The opening angle of the polar cone depends upon the angle of inclination between the axis of the magnetic field and the axis of the disc, the surface area of the polar cap relative to the surface area of the star being $\sim R_*/r_M$, a result similar to that found in Sect. 13.10 for radio pulsars. This reasoning leads to the concept that there is an *accretion column* associated with the infall of matter onto strongly magnetic neutron stars. Only if the magnetic field is very weak can accretion take place directly onto the surface of the neutron star, that is, if $r_M \leq r_*$.

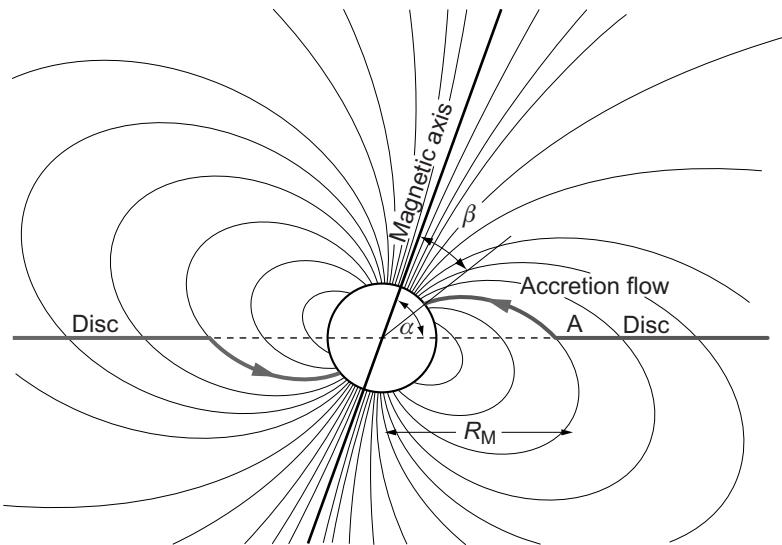


Fig. 14.14

Illustrating the accretion of matter from an accretion disc onto the polar caps of a magnetised neutron star or white dwarf (Frank *et al.*, 2002). Within the magnetically dominated region, the matter is channelled down the magnetic field lines onto the magnetic poles of the neutron star where its binding energy is released, resulting in strong heating of the plasma. A hot accretion column is formed above the magnetic poles from which intense X-radiation is emitted.

Repeating this calculation for white dwarfs, the radius of the star is taken to be 5000 km and the surface magnetic flux density $B_s = 10^3$ T (Sect. 8.2). The Alfvén radius for a solar mass white dwarf radiating at the Eddington limit is then 6×10^4 km, about 10 times the radius of the white dwarf and so again, the region in the vicinity of the star is magnetically dominated. Since the Alfvén radius depends upon the magnetic flux density and mass accretion rate as $B_s^{4/7} \dot{m}^{-2/7}$, it follows that those white dwarfs with strong magnetic fields and low accretion rates are magnetically dominated, leading to accretion onto the poles of the magnetic field structure.

The physics of an accretion disc in the presence of a magnetically dominated region about the star is a complex problem. The interaction between the magnetic field structure and the plasma of the accretion disc has to be understood and this depends upon the angle between the magnetic axis of the compact star and the axis of the accretion disc. The radius within which the magnetic field of the star dominates the dynamics of the accretion disc can be found by equating the torque exerted by the accretion disc on the magnetic field structure to the magnetic torque associated with the distorted magnetic field distribution of the star at that radius. This is a non-trivial calculation, particularly when the effect of instabilities at the interface between the disc and the magnetic field structure are taken into account. The result is similar to that derived above for the case of spherical accretion. According to Frank, King and Raine (2002), the critical radius is slightly smaller than the Alfvén radius r_M found above, typically being about half that value. The reason why this radius is not so different from the spherically symmetric case is the very strong radial dependence of the pressure of the magnetic field, $p_{\text{mag}} \propto r^{-6}$.

The accretion disc exerts a torque upon the magnetosphere of the star which in turns transmits the torque to the star itself, leading to the speeding up of the rotation of the star. Spin-up is observed in the periods of X-ray binary stars and it is also likely that such accretion torques lead to the formation of the millisecond pulsars described in Sect. 13.5.4. To a very good approximation, the angular velocity of the material in a thin accretion disc at radius r can be taken to be the Keplerian value, $\Omega_K^2(r) = (GM_*/r^3)$. If angular momentum is to be transferred inwards, the angular velocity of the disc at radius r_M must be greater than Ω , the angular velocity of the neutron star and its associated magnetosphere, which corotate within radius r_M , that is $\Omega_K(r_M) > \Omega$. Consider the case in which $\Omega_K(r_M) \gg \Omega$, that is, the compact star is a slow rotator. The rate of inward transfer of angular momentum at radius r_M is given by (14.42) which can be rewritten $G = \dot{m}r_M^2\Omega_K(r_M)$. This angular momentum is absorbed by the star and so

$$I\dot{\Omega} = \dot{m}r_M^2\Omega_K(r_M), \quad (14.61)$$

where I is the moment of inertia of the star.

The rate of speed-up of the compact star can now be related to observable properties of the binary system. If the luminosity of the source is due to accretion, the mass accretion rate is directly related to the luminosity L through the relation $L = G\dot{m}M/r_*$, since, as indicated in Fig. 14.14, the energy release by accretion continues down to the surface of the star. Therefore, the rate of spin-up of the compact star can be expressed in terms of its luminosity. Inserting this expression and (14.60) into (14.61),

$$\frac{\dot{P}}{P} = -\frac{1}{2\pi I} \left(\frac{2\pi^2}{\mu_0^2 G^6} \right)^{1/14} (B_s^{2/7} R_*^{12/7} M_*^{-3/7}) (L^{6/7} P), \quad (14.62)$$

where it is assumed that the value of $\Omega_K(r_M)$ is given by the Keplerian value at radius r_M and P is the period of rotation of the compact star, $P = 2\pi/\Omega$. The speed-up rate depends principally upon the product of the rotation period of the compact star and the luminosity to the power 6/7. Inserting typical values expected for a neutron star, $M_* = M_\odot$, $R_* = 10^4$ km and $B_s = 10^8$ T, we find

$$\log_{10} \left(\frac{\dot{P}}{P} \right) = -4.4 + \log_{10} \left(PL_{30}^{6/7} \right), \quad (14.63)$$

where the luminosity L_{30} is measured in units of 10^{30} W.

This dependence of the spin-up rate upon luminosity and period can be tested using observations of binary X-ray sources observed in the hard energy waveband by the BATSE experiment of the Compton Gamma-ray Observatory. Figure 14.15 shows two examples of the relation between luminosity and rotational frequency for the sources A0535+262 and GRO J1744–28 which were observed during giant outbursts. The square symbols refer to the rise of the luminosity to maximum and the diamond symbols to the outburst decline. The dotted curves are power laws with slope 6/7 which are in good agreement with

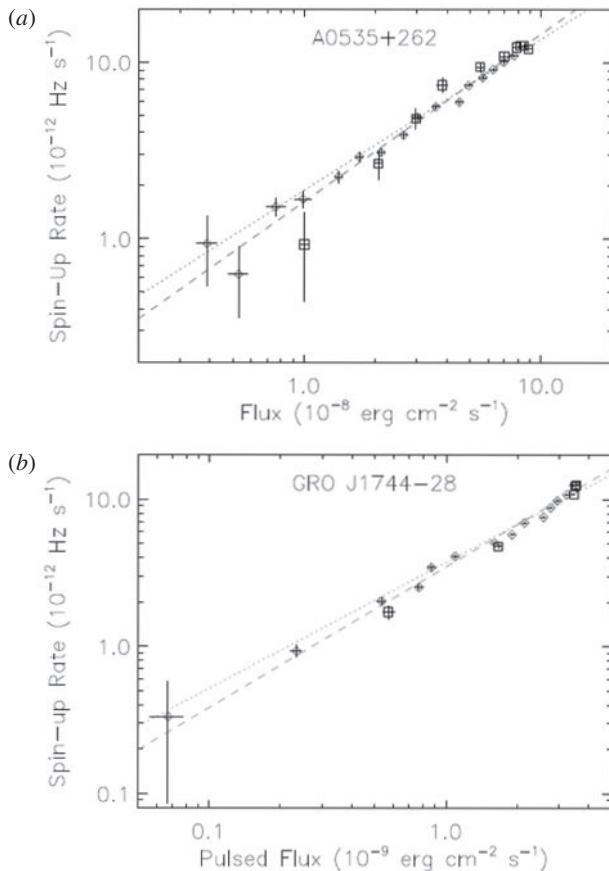


Fig. 14.15

The relation between luminosity and rotational frequency for two X-ray binary sources observed by the BATSE instrument of the Compton Gamma-ray Observatory (Bildsten *et al.*, 1997). (a) The spin rate of A0535+262 during the 1994 giant outburst versus the flux density in the 20–100 keV energy band. (b) The spin-rate of GRO J1744–28 during the December 1995–March 1996 outburst versus the flux density in the 20–50 keV waveband. Square symbols refer to the outburst rise, and the diamond symbols to the outburst decline. The dotted curves are power laws with slope 6/7, while the dashed curves are best-fit powers laws. The best fit index for A0535+262 is 0.951(26) and for GRO J1744–28 is 0.957(26).

the expectation of the accretion theory of spin-up (Bildsten *et al.*, 1997). These are two particularly clear cases in which the simple theory can account of the observations. In other cases, more complex spin-up patterns are observed, presumably associated with variations in the accretion rate onto the neutron star.

Another pleasant result can be derived from the simple picture of spin-up associated with the accretion of matter in binary systems. Spin-up can only occur if the Keplerian angular velocity of the matter in the disc is greater than the angular velocity of the star. Spin-up ceases when these angular velocities are equal. An upper limit to the angular velocity of a neutron star can therefore be found by equating Ω to $\Omega_K(r_M)$. Expressing this condition in

terms of the period of the neutron star and using the result (14.60) for r_M ,

$$P = 2\pi \left(\frac{r_M^3}{GM_*} \right)^{1/2} = \frac{2\pi}{(GM_*)^{1/2}} \left(\frac{2\pi^2}{G\mu_0^2} \right)^{3/14} \left(\frac{R_*^{12}}{M_*\dot{m}^2} \right)^{3/14} B_s^{6/7}. \quad (14.64)$$

Thus, the greater the mass accretion rate \dot{m} , the shorter the minimum period of the neutron star, but the stronger the magnetic field, the less effective this process is in spinning-up neutron stars. The minimum period to which the neutron star could be spun-up can be found by adopting the maximum accretion rate associated with the Eddington limiting luminosity $L = G\dot{m}M_*/R_* = 1.3 \times 10^{31}$ W for a neutron star with $M = M_\odot$ and $R_* = 10^4$ m,

$$P_{\min} \approx 2B_s^{6/7} \text{ ms}, \quad (14.65)$$

where the period P is measured in milliseconds and the magnetic field strength B_s in units of 10^5 T. This is exactly the result found by van den Heuvel (1987) which was quoted in Sect. 13.5.4. The fact that most of the millisecond radio pulsars are members of binary systems strongly supports the picture in which they are spun-up by accretion. Notice that millisecond pulsars have relatively weak magnetic fields, typically $\leq 10^5$ T, which is an advantage according to (14.64). To derive the line labelled ‘spin-up’ in Fig. 13.16, (14.64) is combined with the relation between B_s and the spin-down rate of the radio pulsar (13.42), $B_s = 3 \times 10^{15}(P\dot{P})^{1/2}$ T. The result is the line $\dot{P} \leq 2 \times 10^{-15}P^{4/3}$ plotted on Fig. 13.16 which shows that even the shortest period millisecond pulsars can be accounted for by the process of spin-up by accretion torques.

14.6 Accreting binary systems

14.6.1 Cataclysmic variables

The cataclysmic variables are close binary systems of particular interest because they display a wide range of different types of accretion phenomena and, in some of them, direct evidence has been found for the presence of accretion discs. All aspects of the study of cataclysmic variables are discussed in the comprehensive text *Cataclysmic Variable Stars* by Warner (1995). The donor star is usually a late-type star near to or on the main sequence in which Roche-lobe overflow takes place onto the compact star which is normally a white dwarf. The periods of the binaries range from about 80 minutes to several hours.

The best known examples of the class are the classical novae and the dwarf novae. In *classical novae*, the optical brightness increases by between 6 and 19 magnitudes and only one such major outburst is ever observed. The more luminous the outburst, the shorter its duration. The enormous release of energy is attributed to the accumulation of sufficient high temperature material at the boundary of the degenerate core of the white dwarf that thermonuclear reactions involving the conversion of hydrogen into helium takes place. Because the equation of state of the degenerate material is independent of temperature, the temperature of the material increases at constant pressure and this leads to an exponentially increasing rate of the release of energy in a *thermonuclear runaway*. In contrast, in *dwarf*

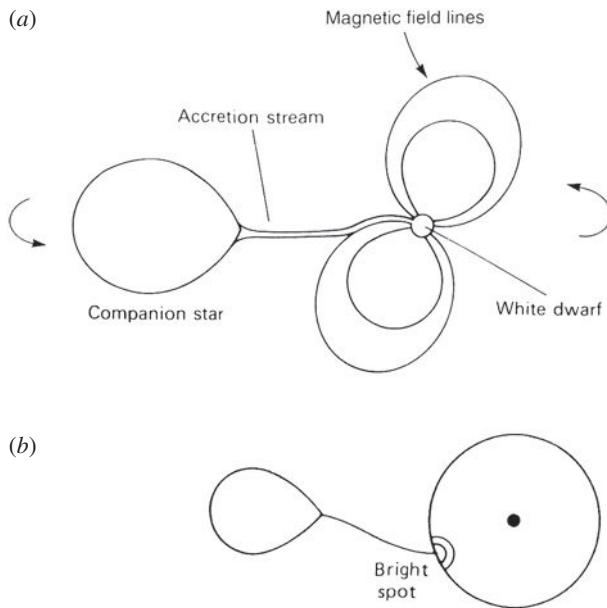


Fig. 14.16

(a) Illustrating the accretion of matter in a ‘polar’ or AM Herculis system. The magnetic field is very strong in these systems ($B \geq 10^3$ T) and the rotation of the white dwarf is locked to the rotation of the binary system as a whole. No accretion disc forms and the matter streams more or less directly onto the surface of the white dwarf. (b) Illustrating the formation of a ‘hot-spot’ at the outer boundary of an accretion disc in a cataclysmic binary system (Frank *et al.*, 2002)

novae, the star regularly brightens by 2–5 magnitudes and this is assumed to be associated with the process of accretion. Intermediate between these types are the objects known as *recurrent novae* in which outbursts take place on the time-scale of decades. In some cases the companion stars are giant stars rather than main sequence stars. Close relatives of these are the *symbiotic stars* in which the accreting star is a main sequence star rather than a dwarf star. There are also systems known as *novae-like stars* which are similar to novae but which are in permanent outburst.

In some cataclysmic variables, the white dwarfs have very strong magnetic fields, $B \geq 10^3$ T, and the rotation of the star is phase-locked to the orbit of the binary. The objects called *polars* or *AM Herculis stars* are examples of this type in which the accreting matter flows along the magnetic field lines from the donor star onto the poles of the white dwarf (Figure 14.16a). In other cases, the rotation period of the white dwarf is shorter than the period of the binary orbit and these objects are referred to as *intermediate polars*. Both of these classes of magnetic cataclysmic variable exhibit high and low states of activity, presumably associated with different states of mass accretion.

The dwarf novae are of particular interest since the optical and ultraviolet emission from the disc can be observed directly. In other close binary systems, the light is dominated by the emission of the normal star, for example, in X-ray binaries containing O and B stars or in low mass X-ray binaries which emit reprocessed X-ray light. In some dwarf novae,

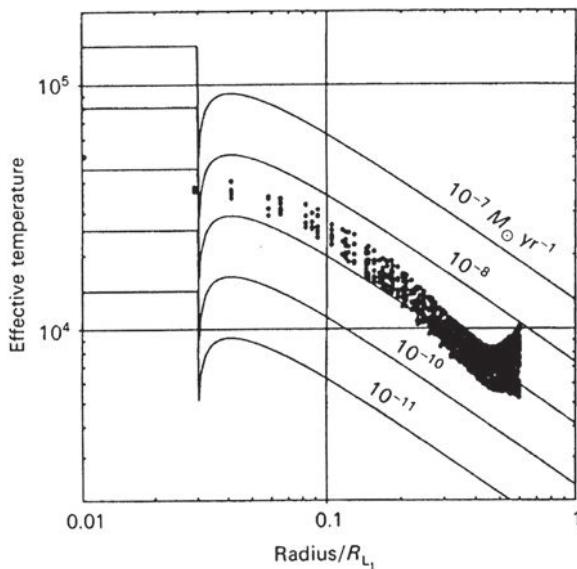
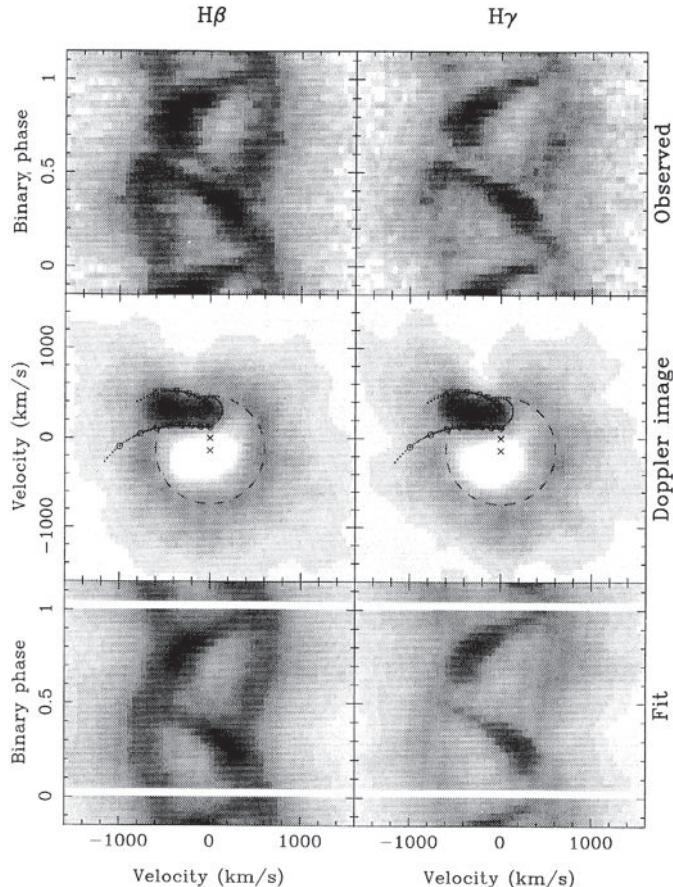


Fig. 14.17 The results of eclipse mapping to determine the temperature distribution in the accretion disc in the dwarf nova Z Cha. The observations were made during an outburst so that the emission from the disc dominates the optical light. The lines show the expected temperature distributions for optically thick steady-state discs with different accretion rates (Horne and Marsh, 1986; Frank *et al.*, 2002).

the normal star eclipses the compact star and its accretion disc and so the two-dimensional temperature distribution in the disc can be reconstructed by very precise rapid photometry in a number of colours, a procedure known as *eclipse mapping*. Figure 14.17 shows the results of reconstructing the temperature distribution in the accretion disc of the dwarf nova Z Cha during an outburst (Horne and Marsh, 1986). The temperature distribution follows closely the relation $T \propto r^{-3/4}$ as expected from (14.54), implying that, during the outburst, the disc was optically thick and that the steady-state accretion rate was about $10^{-9} M_{\odot} \text{ y}^{-1}$. In their quiescent states, however, the inferred temperature distributions of the accretion discs are not consistent with the expectations of the steady-state models, probably because the accretion rates are low and the discs are no longer optically thick.

In the standard disc accretion picture, the stream of matter flowing through the inner Lagrangian point L_1 impacts the disc at some point and causes a local rise in temperature of the disc in a ‘hot spot’ (Fig. 14.16b). The geometry of the hot spot and disc can be mapped using the procedure known as *Doppler tomography*. As the binary system rotates, different projections of the disc and the hot spot are observed and, by taking many high resolution spectra of the system at different binary phases, the intensity distribution in the disc can be reconstructed. Figure 14.18 shows the results of carrying out this procedure for the binary system U Gem by Marsh and his colleagues (1990). The emission lines of H β and H γ are bright in this system and the central panels show the reconstructed brightness distribution of the disc. Most of the line emission comes from the hot spot, although emission is also observed from the disc. This set of observations was taken whilst the system was in its

**Fig. 14.18**

Doppler tomography of the emission lines of $H\beta$ and $H\gamma$ in the dwarf nova system U Gem during its quiescent phase. The top panels show the observed spectra as a function of orbital phase while the central panels show the reconstructed Doppler images fitted to these observations. The lower panels show the computed velocity-phase data expected from the model-fits. The gaps in the lowest panels correspond to the eclipses which were not included in the fits. The reference circles have radii corresponding to 600 km s^{-1} (Marsh *et al.*, 1990).

quiescent state. During an outburst, the emission from the hot spot is swamped by the emission from the disc.

A further test of the standard thin disc model is its continuum spectrum. According to the simplest model of an optically thick steady-state disc, the continuum spectrum is expected to be of the form $I(\nu) \propto \nu^{1/3}$ (14.57). A better approximation is to assume that the disc radiates like a stellar atmosphere from its upper and lower surfaces. According to Frank, King and Raine (2002), this somewhat more elaborate treatment does not improve markedly the fit of the theoretical spectra to the observations.

In the discussion of thin accretion discs of Sect. 14.3.3, it was shown that only half of the gravitational potential energy of the accreted matter is dissipated in reaching the boundary layer about the star at radius $r_* + b$, the rest of the energy being dissipated in the boundary

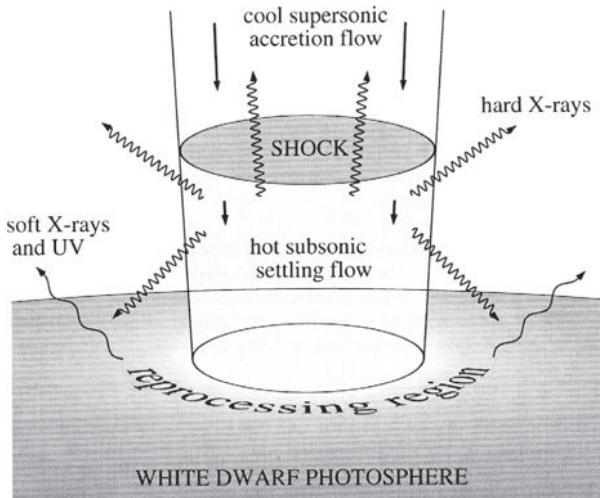


Fig. 14.19 Illustrating the geometry of the accretion column at the poles of a strongly magnetised white dwarf (Frank *et al.*, 2002).

layer itself. It is therefore expected that the highest energy radiation is emitted from the boundary layer close to the stellar surface. Since the temperatures of the innermost layers are expected to be about 10^5 K, cataclysmic variables should be strong emitters in the EUV and soft X-ray wavebands, as is found to be the case.

What was more surprising was the detection of high energy X-rays in the 1–10 keV waveband from cataclysmic variables. In most cases, the spectra can be described by thermal bremsstrahlung spectra with temperatures in the range $kT \approx 1\text{--}5$ keV. Matter is funnelled by the magnetic field onto the poles of a magnetic white dwarf where it is brought to rest, in the process dissipating its kinetic energy of collapse as thermal energy. The matter acquires a high velocity in falling onto the star and so a shock front is formed above the stellar surface as shown in Fig. 14.19. An upper limit to the temperature of the gas in the shocked region can be found by equating the kinetic energy of infall of the gas to the thermal energy of the plasma once it has been thermalised in the shocked region, that is,

$$\frac{1}{2}(m_p + m_e)v^2 \approx \frac{1}{2}m_p v_{ff}^2 = \frac{1}{2}m_p v_{ff}^2 = \frac{GM_* m_p}{R_*} = \frac{3}{2}(kT_p + kT_e). \quad (14.66)$$

If the energy of infall is shared between the electrons and protons on passing through the shock front and thermal equilibrium is attained, that is, $T_p = T_e$, then

$$T_e = \frac{GM_* m_p}{3k R_*}. \quad (14.67)$$

For the case of accretion onto the poles of a solar mass white dwarf with $R_* = 10^4$ km, $T_e = 5.4 \times 10^8$ K and $kT_e \approx 50$ keV. This is an upper limit to the thermal energy of the plasma in the polar cap since it assumes that all the energy of infall goes into heating the plasma. As indicated in Fig. 14.19, some of the energy heats the surface layers of the white dwarf where it is reradiated as far ultraviolet and soft X-ray emission. Thus, there

is no problem, in principle, in understanding the origin of very hot gas in the vicinity of the poles of accreting white dwarfs. If the region remains optically thin to radiation, these regions would be expected to be sources of high energy bremsstrahlung emission. Frank, King and Raine (2002) discuss the many problems in constructing realistic models for the accretion columns associated with white dwarfs. To a good approximation, the problem can be reduced to a one-dimensional flow problem with the hard X-ray emission escaping from the cylindrical surface of the shocked polar cap as indicated in Fig. 14.19. A further surprise was the detection of strong winds from cataclysmic variables. Characteristic P-Cygni profiles have been observed with wind velocities up to 5000 km s^{-1} . Thus, not only is mass being accreted onto the white dwarf but it is also being returned to the interstellar medium in the form of winds.

The dwarf nova phenomenon is assumed to be associated with abrupt enhancements of the accretion flow onto the white dwarf. In one model, the enhanced accretion rate is attributed to instabilities in the disc. The question of the stability of the material of an accretion disc is a huge subject which goes beyond the scope of this volume. It turns out to depend very sensitively upon the assumed value of the viscosity of the disc as described by the α -parameter. An alternative picture is that the phenomenon is caused by short-lived mass transfer events from the donor star. Whatever the origin of the enhanced accretion rate, there is evidence that the X-ray emission originates close to the surface of the white dwarf. In the case of the dwarf nova HT Cas, the X-ray emission is occulted by the white dwarf and the depth of the eclipse is found to be very deep, consistent with it being total. The eclipse transition is also found to be short, which sets a limit of 1.15 times the white dwarf radius for the size of the X-ray emission region (Mukai *et al.*, 1997). It is likely that the X-ray emission originates within the boundary layer about the white dwarf.

14.6.2 Novae and Type Ia supernovae

The principal features of the *classical novae* are summarised by Starrfield (1988) and Woosley (1986). Classical novae are ‘new stars’ which can increase in brightness by more than 10 magnitudes in a stellar explosion. The rise to maximum light occurs over a period of a few days and can remain at that luminosity for several months at roughly the Eddington luminosity for a $1 M_\odot$ white dwarf, $L = 1.3 \times 10^{31} \text{ W}$. The total energy release amounts to about 10^{38} J . There is evidence for the ejection of material from the nova with velocities between a few hundred and a few thousand km s^{-1} . The ejecta contain large abundances of elements synthesised by the CNO cycle at high temperatures. The event rate for classical novae is estimated to be about 30 y^{-1} in our own Galaxy and about 38 y^{-1} in M31. Thus, although they are much less powerful explosions than supernovae, they occur much more frequently.

The preferred model for classical novae involves the accretion of a critical mass of hydrogen-rich fuel onto the surface of a white dwarf. The accreted material is compressed and heated to a high temperature, $T > 10^7 \text{ K}$, at the base of the accreted layer which becomes degenerate and, as the critical mass is approached, nuclear burning of hydrogen into helium through the CNO cycle can take place. The critical mass of hydrogen is estimated to be about $10^{-5} - 10^{-4} M_\odot$. If the matter is sufficiently degenerate, *thermonuclear runaway*

takes place because of the very strong dependence of the reaction rate of the CNO cycle upon temperature, $\epsilon \propto T^{17}$ (Sect. 2.4). The CNO cycle (2.28) involves the addition of four protons to the carbon nucleus with two intermediate β^+ decays of ^{13}N and ^{15}O . Other side-chains involve the β^+ decays of ^{14}O and ^{17}F . As described by Starrfield (1988), during the early stages of thermonuclear runaway, the time-scales for the reactions involving the addition of the protons to the parent nuclei are longer than the half-lives of the above β decay nuclei. As the temperature increases above 10^8 K, however, the time scale for the addition of protons becomes shorter than the half-life of the β decays with the result that these unstable nuclei become abundant and the rate at which nuclear energy generation proceeds is stabilised since the proton capture processes have to wait until the β decays have taken place. As a result, the thermonuclear runaway stabilises at a temperature of about 10^8 K. A further feature of this process is that, because of the very strong temperature dependence of the energy generation rate in the CNO cycle, the layers become unstable to convection and so these unstable nuclei can be convected to the surface layers of the accreted material where their β decays provide a further source of heating. Once the peak temperature is reached and the envelope begins to expand, these unstable nuclei continue to provide a source of energy for the nova. These ideas have been studied in detail in computer simulations of nova explosions and can account for the main features of classical novae including the abundances of the rarer CNO isotopes. The mass ejected in classical novae is about $10^{-4} M_\odot$ and so most of the accreted layer is expelled in the explosion.

Type Ia supernovae were described in Sect. 13.1.2 and also involve the accretion of mass onto the surface of a white dwarf. The difference is that in this case the effect of accretion of mass is to increase the central temperature of the star to such high values that the nuclear burning of carbon and oxygen takes place in the core. The resulting nuclear deflagration at temperatures $T \geq 10^9$ K disrupts the whole star.

14.6.3 Low mass X-ray binaries

In the low mass X-ray binaries, the donor star is normally a main sequence star with mass $M \sim M_\odot$. Members of this class are often referred to as *Galactic bulge sources* because they lie in the general direction of the Galactic Centre. A number of them have been identified with binary systems in globular clusters which are members of the bulge population with a distribution similar that of the Galactic bulge stars and amongst the oldest systems in our Galaxy. All the stars with mass greater than about $1 M_\odot$ in these clusters have completed their evolution on the main sequence and the giant branch.

Unlike the case of the cataclysmic variables, it proved much more difficult to obtain definitive evidence for the binary nature of the low mass X-ray binaries. This is now understood to be a selection effect in the sense that what is observed at X-ray wavelengths is strongly dependent upon the angle of inclination i of the plane of the binary to the plane of the sky. The brightest and most luminous sources are those in which the angle of inclination is small so that the orbital plane is viewed face-on. The compact X-ray source is then observed unobscured by the disc or the binary companion but this geometry also means that it is difficult to determine the binary properties of the orbit. Lower luminosity X-ray sources display evidence of eclipses and ‘dips’ in their X-ray light curves, examples

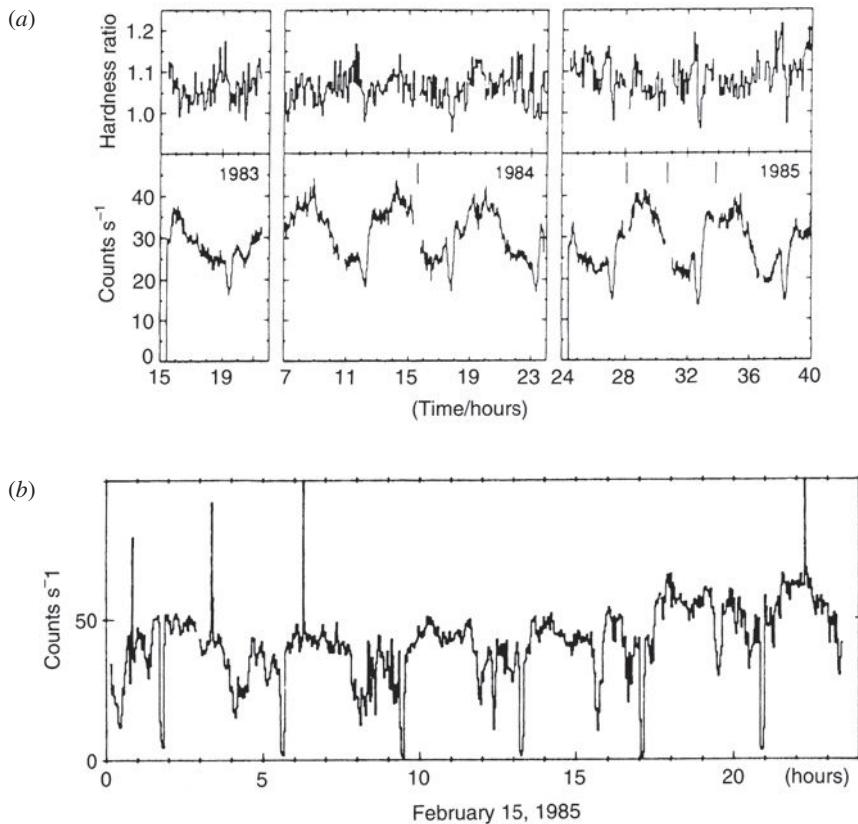
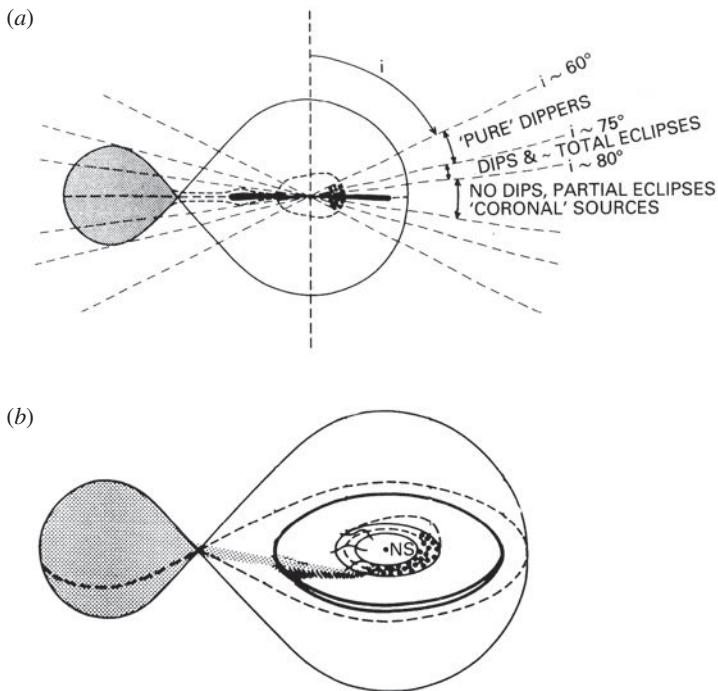


Fig. 14.20

Examples of the X-ray light curves of two low-mass X-ray binary systems illustrating a number of eclipse and absorption features. (a) The light curve of the 5.57 hour binary system X1822–371 determined by the EXOSAT X-ray telescope in 1983, 1984 and 1985. The upper panel shows the hardness ratio of the X-ray spectrum. (b) The light curve of the 3.83 hour period binary XBT0748–676. (Watson and King, 1991).

of some of these features being shown in Fig. 14.20. It is significant that the ratio of their X-ray to optical luminosities are much smaller than those systems inferred to be observed at small orbital inclinations. This strongly suggests that the X-ray emission is strongly attenuated when observed at large angles of inclination.

In modelling these sources, it is assumed that a standard accretion disc is the source of the X-ray luminosity, most of it originating from the inner region of the disc, but much more has to be added to account for the various observed features of their light curves. There appear to be two essential additions to the basic model. The first is related to the observation that, in a number of sources, the binary eclipses are gradual and do not completely occult the X-ray source. This is interpreted as evidence for an *accretion disc corona* which may fill a substantial fraction of the Roche lobe of the neutron star. The X-ray luminosity of the corona is only a fraction of the total luminosity of the source and so it might be scattered radiation from the compact X-ray source or the emission of a hot corona about the neutron star which might result from heating by the X-ray source or from instabilities in the accretion disc.

**Fig. 14.21**

(a) A schematic diagram illustrating the necessity of incorporating a thick absorbing screen in the structure of the accretion disc in low mass X-ray binaries. In this sketch, absorption features are only observed at angles of inclination greater than about 60° . It is assumed that the ‘accretion disc corona’ observed in X-rays fills a large fraction of the Roche lobe of the neutron star (Frank *et al.*, 2002). (b) A schematic perspective view of the model for a low mass X-ray binary with penetrating gas streams (Frank *et al.*, 1987).

The second essential component is some thickening of the accretion disc to account for statistics of absorption features seen in the low mass X-ray binaries. This geometry is indicated schematically in Fig. 14.21a. It can be seen that unless the disc is somewhat thicker than the standard thin disc, there is only a small likelihood of observing eclipses and absorption features in the X-ray light curves. One possibility discussed by Frank, King and Lasota (1987) is that it is associated with the fact that the stream of plasma which comes through the Lagrangian point L_1 is likely to be broader than the thickness of the thin disc at the point where the stream encounters the disc. This can lead to streams of material skimming over the surface of the disc until they reach centrifugal equilibrium at some radius. This process would lead to the formation of cool clouds which can provide thickening of the disc and the necessary X-ray absorption (Fig. 14.21b).

Another variant upon accretion in low mass X-ray binaries is provided by the *X-ray burst sources*, or *bursters*. By far the more common are the Type I bursts in which the X-ray luminosity typically increases by about a factor of 10 with rise times only about 1–10 seconds. The bursts last for less than a minute and, during their decline from peak intensity, the X-ray spectra steepen to low energies (Fig. 14.22). The spectra can be well-characterised by black-body spectra which cool as the burst evolves. The Type I burst sources occur at

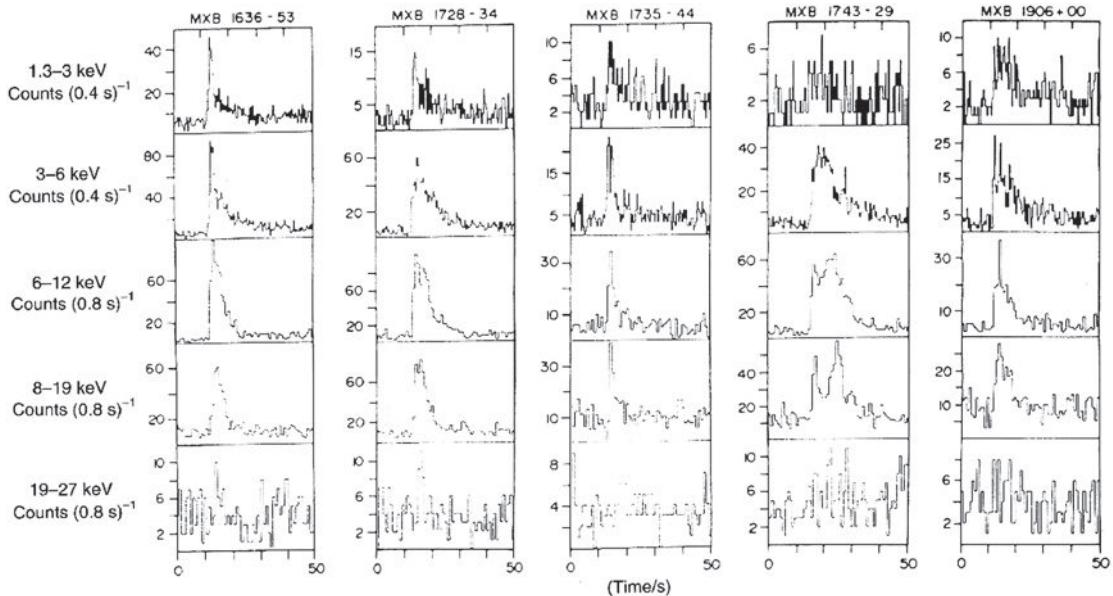


Fig. 14.22 Examples of Type I X-ray bursts from five different sources. The light curves are shown in five different energy ranges and it can be seen that the gradual decay persists longer at the low energies than it does at high energies, indicating that the source cools as the source fades away (Frank *et al.*, 2002, courtesy of Waller Lewin).

intervals of hours to days and so the total energy liberated in the bursts is small compared to the luminosity of the steady-state component. Typical statistics for the Type I bursts are that the interval between bursts is roughly 1000 times the duration of the burst but during the burst the luminosity is enhanced by a factor of 10. Thus, the *average* luminosity of the steady component is about 100 times that of the bursts. A further clue is provided by the observation that no X-ray pulsar has been observed in a Type I burst source suggesting that the magnetic fields in the neutron stars are weak.

These pieces of evidence can be accounted for by a model in which the accretion disc extends inwards to the surface of the neutron star. The steady X-ray emission is attributed to the radiation of the accretion disc while the burst emission is associated with a *thermonuclear runaway* of the accreted matter as it builds up on the surface of the neutron star. The latter is the same process which can account for the energy release in novae (Sect. 16.5.3) but now applied to accretion onto neutron stars. We can compare the average energy released by accretion, E_{acc} , with the energy it would release when it delivers its available nuclear energy at the surface of the neutron star, E_{nuc} ,

$$\frac{E_{\text{acc}}}{E_{\text{nuc}}} \leq \frac{\eta m_p c^2}{0.007mc^2}, \quad (14.68)$$

since the maximum amount of nuclear energy which can be released is the binding energy per nucleon of the helium nucleus, which is $0.007m_p c^2$. For the case of neutron stars we can take $\eta \sim 0.1$ and so the accretion disc can release at least an order of magnitude more energy than the burning of the nuclear fuel on the surface of the neutron star. When account

is taken of the fact that not all the nuclear energy is emitted as black-body radiation from the surface of the neutron star, an average luminosity ratio of the steady to burst component of the order of 100 can be naturally explained.

In the course of the thermonuclear explosion, the X-ray luminosity of the neutron star approaches the Eddington limit. If the luminosity of an X-ray burster is observed to saturate at a particular value, this can be taken to correspond roughly to the Eddington critical luminosity and then further information may be obtained about the dimensions of the source region. In 10 Type I burst sources for which temperatures have been measured, the ratio of the limiting luminosity of the burst to the inferred surface area from which the luminosity is emitted has been found to be roughly a constant. The typical radius of the source is inferred to be about 7 km for these bursters, providing further evidence for the presence of neutron stars in the Type I X-ray bursters.

There are only two examples known of Type II bursters, the sources MXB 1730–335, which is also known as the *rapid burster*, and GRO 1744–28. They are similar to Type I bursts but in addition there are much more rapid bursts with repetition periods of seconds to minutes. The energy of each burst is apparently proportional to the waiting time until the next burst, as if the burst had exhausted the supply of energy for the moment. The origin of this behaviour within the context of the standard accretion disc picture is not clear.

The low mass X-ray binaries are sources of many intriguing astrophysical phenomena which provide insights into the physics of neutron stars and the interaction between the accretion disc and the neutron star. An example is the modulation of the light curve of the low mass X-ray binary Her X-1. In addition to the X-ray pulsar which has period 1.24 s and a binary period of 1.7 days, the source undergoes a 35 day cycle during which it is strong for 9 days and then relatively faint for the remaining 26 days. This behaviour is accompanied by changes in the pulse profiles. The likely explanation of this behaviour is precession of the rotation axis of the neutron star in its binary orbit so that the X-ray beam from one of the poles is pointing towards the observer during the bright phase. As precession changes the orientation of the magnetic poles with respect to the observer, the intensity observed from the bright pole decreases but radiation from the other pole is observed. This model has important implications for the internal structure of the neutron star because the crust of the neutron star and the associated magnetic field must be decoupled, or unpinned, from the neutron superfluid since otherwise the neutron star would possess too much angular momentum and would not undergo appreciable precession. It is also necessary that the crust be sufficiently rigid to maintain a significant oblateness.

This is only the beginning of a complex story which has emerged from detailed studies of low mass X-ray binaries by space telescopes such as the EXOSAT Observatory and the Rossi Timing X-ray Explorer (RXTE). Most of the bright Galactic bulge sources do not contain X-ray pulsars but exhibit unstable periodic behaviours which are referred to as *quasi-periodic oscillations* (QPOs). Observations by the EXOSAT Observatory showed that sources such as Cyg X-2, Sco X-1 and GX5–1 exhibit rapid fluctuations in their intensities. Power spectrum analysis of the source intensities as a function of time showed, not a single sharp line spectrum indicative of a stable period, but rather a broad peak spanning a range of frequencies. An example of quasi-periodic oscillations in the source Cyg X-2 is shown in Fig. 14.23 (Hasinger and van der Klis, 1989). As indicated in the figure, the phenomenon is

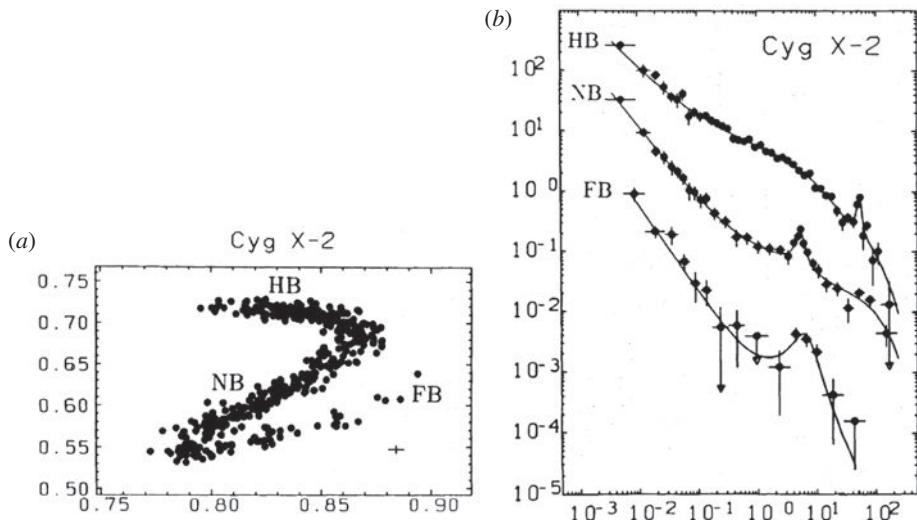


Fig. 14.23

(a) The X-ray colour–colour diagram for the low mass X-ray binary Cyg X-2, showing the various branches identified by Hasinger and van der Klis. The colours are ratios of flux densities in the soft (3–4.5 keV : 1–3 keV) and hard X-ray (6–20 keV : 4.5–6 keV) wavebands. (b) The power spectra of intensity variations for the three branches identified in (a), showing the presence of quasi-periodic oscillations (Hasinger and van der Klis, 1989). The abscissa is the frequency in Hz.

only observed when the X-ray sources have specific X-ray spectra, as indicated by their X-ray colour indices. The mean frequency of the oscillations moves to higher frequencies as the intensity of the source increases. For example, in the source GX5–1, the central frequency increases systematically from 20 to 36 Hz as the source intensity increases by almost 50%. This behaviour is observed in most sources which exhibit quasi-periodic oscillations, although there are some exceptions. Observations with the RTXE have discovered QPOs with central frequencies in the millisecond range (van der Klis, 2000).

The short periods of the QPOs indicate that these phenomena must be associated with processes occurring very close to the surface of the neutron star. In order that the accretion disc extend in towards the surface of the neutron star, the magnetic fields must be relatively weak, $B \leq 10^6$ T, consistent with the fact that the sources are not observed as X-ray pulsars. A favoured picture is one in which the oscillations are associated with ‘beats’ between the rotation frequency of the neutron star and the Keplerian frequency of rotation of matter at the inner edges of the accretion disc. In one version of this model, the accretion disk and the magnetosphere, which corotates with the neutron star, rotate at different frequencies and inhomogeneities in the accretion disk interact with anisotropies in the magnetosphere. There are, however, alternative theoretical suggestions for the QPO phenomenon. These include models in which the QPO frequency is interpreted as a mode of oscillation of the accretion disk, or a relativistic precession model in which inclined eccentric orbits exhibit nodal and relativistic periastron precession, and a photon bubble model in which these form close to the surface of the neutron star and then rise up the gravitational potential gradient under the influence of buoyancy.

14.6.4 High mass X-ray binaries

The *high mass X-ray binaries* were among the first binary X-ray sources to be identified following the success of the UHURU mission in the 1970s. The donor stars are massive OB stars and the X-ray sources are associated with their binary companions which are neutron stars or black holes. The high mass X-ray binaries can be divided into two sub-classes. The more common are known as Be X-ray binaries in which the donor star is a Be star, meaning that the B-star displays prominent hydrogen emission lines. The Be stars belong to luminosity classes III–V and the emission lines are associated with a disc of circumstellar material. In the other class the donor star is a supergiant star belonging to luminosity classes I and II (Liu *et al.*, 2006; Lewin and van der Klis, 2006). Liu and his colleagues provide a catalogue of 114 high mass X-ray binaries in our own Galaxy. There is also a large population of high mass X-ray binaries in the Magellanic Clouds.

According to the statistics of Liu and his colleagues, about 60% of the Galactic sources are known or suspected Be X-ray binaries, while 32% are associated with supergiants. In the Be systems, the compact source is a neutron star which has a wide, moderately eccentric orbit so that the neutron star only passes through the dense circumstellar disc associated with the Be star for a small fraction of its orbit. None of these systems is known to host a black hole. X-ray outbursts are observed when the neutron star passes through the dense disc, accreting matter from the low velocity, high density wind about the Be star. In the cases of those X-ray binaries associated with early-type supergiant stars, the compact X-ray source orbits the star deep inside a highly supersonic wind, the fuelling of the X-ray source being either by Roche lobe overflow or by accretion from the strong stellar wind.

The OB supergiants have large mass-loss rates, $\dot{m} \sim 10^{-6} M_{\odot} \text{ y}^{-1}$, and accretion can take place through the capture of a certain fraction of the outflow from the giant star by the compact star (Fig. 14.13a). Within a certain radius, known as the *capture radius* or *accretion radius*, matter is inevitably captured by the compact object. The compact star of mass M_X moves in its binary orbit at velocity v_X and so, assuming the wind is radial with velocity v_W , the velocity of the wind relative to the star v_t is the vector sum of these two velocities, $v_t^2 = v_X^2 + v_W^2$. The accretion rate onto the compact star can then be determined in terms of v_t .

In Sect. 5.2 it was shown that the impulse which a charged particle receives on passing a stationary charge is given by the inward force at the distance of closest approach b times the duration of the collision – this calculation gives exactly the correct answer if the collision time is taken to be the time it takes the moving particle to travel a distance $2b$. The momentum impulse in the case of gravitational encounters in the stellar wind is found by exactly the same argument (Fig. 14.24). The gravitational force of attraction per unit mass at distance b is GM_X/b^2 and the duration of this force is $2b/v_t$. The momentum impulse inwards is therefore $\Delta p = 2GM_X/bv_t$. The result is that the outflowing wind is deflected towards the axis of the flow as illustrated in Fig. 14.24. At some distance l downstream, the particles with collision parameter b collide on the axis of the flow. The perpendicular component of the velocity goes to zero and so the necessary condition that the matter be captured by the star is that the gravitational potential energy of matter at l be greater than

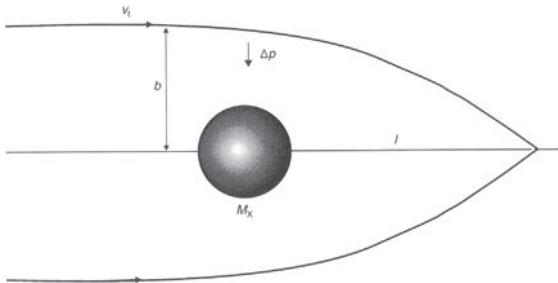


Fig. 14.24 Illustrating the process of accretion by a star of mass M_X in a stellar wind of velocity v_t .

its kinetic energy outwards, which is approximately $\frac{1}{2}v_t^2$ per unit mass. The distance l is roughly $b v_t / v_\perp$ where $v_\perp = \Delta p$ since we consider unit mass. Reorganising these relations, the condition $GM_X/l \geq \frac{1}{2}v_t^2$ reduces to $b \leq 2GM_X/v_t^2$. This critical radius is known as the *capture radius* R_c and can be written

$$R_c = b = \frac{2GM_X}{v_X^2 + v_W^2}. \quad (14.69)$$

The X-ray luminosity of the neutron star in the steady state is $L = \eta \dot{m} c^2$ with mass accretion rate $\dot{m} = (\dot{m}_P/4)(R_c/R_P)^2$; \dot{m}_P is the mass-loss rate from the donor star and R_P is the distance of the neutron star from the centre of the star. Assuming the orbit is circular, the X-ray luminosity of the neutron star is

$$L_X \approx \frac{\eta \dot{m}_P}{4} \left(\frac{2GM_X}{R_P} \right)^2 v_W^{-4}, \quad (14.70)$$

where it is assumed that the wind velocity v_W is much greater than the orbital velocity v_X of the neutron star. The X-ray luminosity is directly proportional to the mass-loss rate of the donor star and is sensitive to the wind velocity. This simple model can account for the X-ray luminosities of many of the high mass X-ray binary systems. Notice that this process is very inefficient compared with that of Roche lobe overflow since in the latter case almost all the material lost by the donor star is channelled through the Lagrangian point into the accretion disc. The sources are, however, observable in the X-ray waveband because of the very high mass-loss rates of the OB stars.

14.7 Black holes in X-ray binaries

The physics of black holes and the observational evidence for their presence in X-ray binary systems were discussed in some detail in Sect. 13.11. Twenty convincing cases of black holes with masses greater than $3 M_\odot$ were illustrated in Fig. 13.30. These sources have been the subject of intense study by X-ray telescopes with the capability of making fast timing observations, in particular, the Rossi X-ray Timing Mission (RXTE) which has transformed the study of black holes in X-ray binary systems. The X-ray properties of these 20 systems

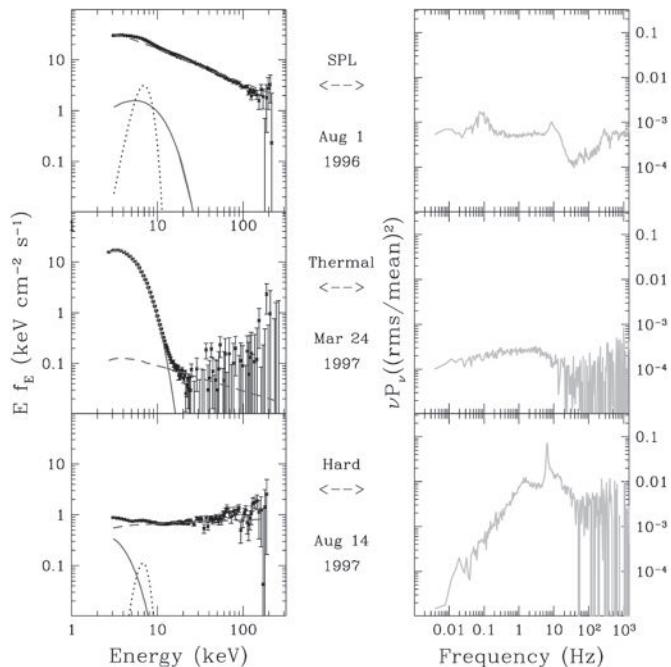


Fig. 14.25 Illustrating sample X-ray spectra and power density spectra for the X-ray nova GRO J1655–40 in three different outburst states (Remillard and McClintock, 2006). The spectra have been decomposed into a thermal component (solid line), a power-law component (dashed line) and the relativistically broadened iron K α line (dotted line). The power spectra are plotted as $\nu \log P_\nu$.

and 20 other candidate systems have been surveyed in detail by McClintock and Remillard (2006) and Remillard and McClintock (2006).

As described by these authors, the first two black hole binaries to be discovered, Cygnus X-1 and LMC X-3, are persistently bright in X-rays, while the third A 0620–00 was quite different in character in that it was discovered as an X-ray nova in 1975. At peak intensity, it became the brightest non-solar X-ray source ever observed. Over the following year, the luminosity decayed back to ‘quiescence’ whilst at the same time the optical counterpart of the outburst faded by 7.1 magnitudes, revealing the K-dwarf companion star. Of the 20 black hole binaries discussed by Remillard and McClintock, 17 are X-ray novae like A 0620–00 and only three are persistently bright.

As is abundantly clear from their survey, a bewildering variety of behaviour is observed among the examples they discuss. They classify the various active states in which the black hole binaries are generally found according to the intensity spectra and power density spectra of the variability, as illustrated in Fig. 14.25. They refer to these states as the *thermal*, *hard* and *steep power-law* states and their quantitative definitions are given in Table 14.1. The parameters in the table have the following significance: the disc fraction f is the ratio of the thermal disc flux density to the total flux density, both unabsorbed in the 2–20 keV energy band; the photon spectral index Γ is defined by $n(\varepsilon) \propto \varepsilon^{-\Gamma}$; r is the power density

Table 14.1 Outburst states of black hole binaries (after Remillard and McClintock 2006).

State name	Definitions of X-ray state
Thermal	Disc fraction $f > 75\%$ QPOs absent or very weak: QPO amplitude $a < 0.005$ Power continuum level $r < 0.075$
Hard	Disc fraction $f < 20\%$ $1.4 < \Gamma < 2.1$ Power continuum level $r < 0.1$
Steep power law (SPL)	Presence of steep power-law component with $\Gamma > 2.4$ Power continuum level $r < 0.15$ <i>Either</i> $f < 0.8$ and 0.1–30 Hz QPOs present with $a > 0.01$ <i>or</i> disc fraction $f < 50\%$ with no QPOs

spectrum of the variability from 0.1 to 10 Hz as a fraction of the average source count rate; and a is the integrated amplitude of the QPOs detected in the 0.1–10 Hz frequency range.

In their reviews, McClintock and Remillard provide many examples of how the various contributions to the total spectrum vary as the bursts develop and reference should be made to these to obtain an appreciation of the complexities involved. The separation into the various components in Fig. 14.25 is physically motivated, some being better understood than others.

14.7.1 The thermal state

The thermal component is taken to be the emission of the inner regions of the accretion disc about the black hole and the standard model of a thin disc developed in Sect. 14.3.4 can be used to model this component as a first approximation. If it is assumed that the accretion disc is optically thick, the temperature distribution close to the black hole can be found by combining the (14.52) with (14.54) so that

$$T(r) = \left[\frac{3GM\dot{m}}{8\pi\sigma r^3} (1 - (r_i/r)^{1/2}) \right]^{1/4}, \quad (14.71)$$

where it is assumed that there is an inner cut-off radius to the accretion disc at r_i . The luminosity of the optically thick disc is therefore

$$L = \int_{r_i}^{r_{\max}} 4\pi r \sigma T^4(r) dr \approx 4\pi r_i^2 \sigma T_i^4, \quad (14.72)$$

where the radiation is assumed to be emitted from both sides of the disc and T_i is the temperature at the inner edge of the disc at r_i . The flux density observed by a distant observer is

$$S = \frac{L \cos i}{2\pi r^2} = \frac{2 \cos i}{D^2} r_i^2 T_i^4, \quad (14.73)$$

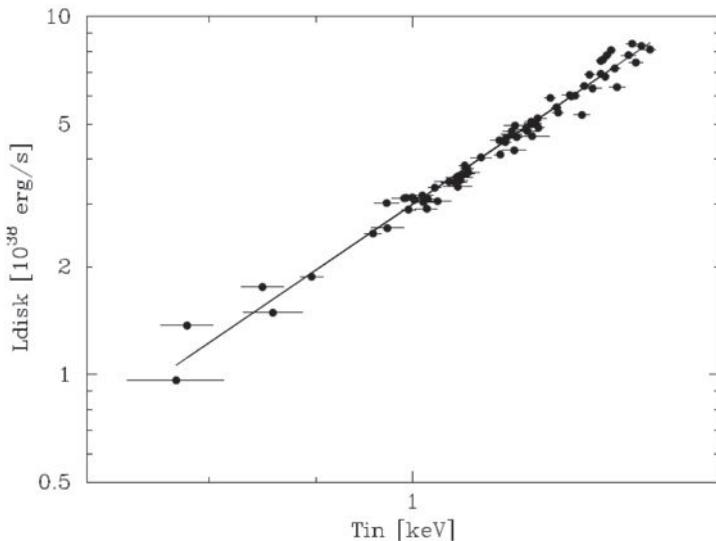


Fig. 14.26 The estimated thermal luminosity L_{disk} of LMC X-3 plotted against the observed temperature T_{in} . The distance and inclination are assumed to be $D = 50 \text{ kpc}$ and $i = 66^\circ$. The solid line is the relation $L_{\text{disk}} \propto T_{\text{in}}^4$ (Kubota and Makishima, 2005).

where the distance of the source is D and i is the angle of inclination between the plane of the disc and the plane of the sky. Both the luminosity L and the temperature T_i can be measured for the thermal component and so the quantity $r_i \cos^{1/2} i$ can be found.

This relation is found to be a good fit to the data for many sources in the thermal state. The example of LMC X-3 is given in Fig. 14.26 which shows that the luminosity–temperature relation (14.72) is an excellent fit to the 80 data points obtained from RXTE observations (Kubota and Makishima, 2005). A similar X-ray luminosity–temperature relation has been observed in the thermal state of many black hole binaries except at very high disk temperatures. The inference is that there is a fixed inner circular orbit about the black hole, which can be identified with the last stable circular orbit about the black hole. Precise determination of this radius requires fully relativistic models for the accretion disc and detailed understanding of the physical processes in the inner regions of the disc. Suffice to say that the measured values are in encouraging agreement with expectation on the basis of the mass estimates of the black holes.

14.7.2 The hard state

One of the most striking correlations in these studies is that between the hard state and the presence of radio jets. The transition is abrupt in that, when the radio jets switch on, the state changes to the hard state. The emission of the radio jets is assumed to be synchrotron radiation because of the remarkable similarity between the microquasar phenomenon and the powerful extragalactic double radio sources such as Cygnus A. It is therefore perfectly

plausible that the hard state of the black hole X-ray binaries is also associated in some way with the processes responsible for the acceleration of the electrons responsible for the radio emission. According to McClintock and Remillard, recent observations of black hole binaries in the hard state suggest that both synchrotron and inverse Compton radiation contribute to the hard spectrum, with the Compton component presumably originating at the base of the jet.

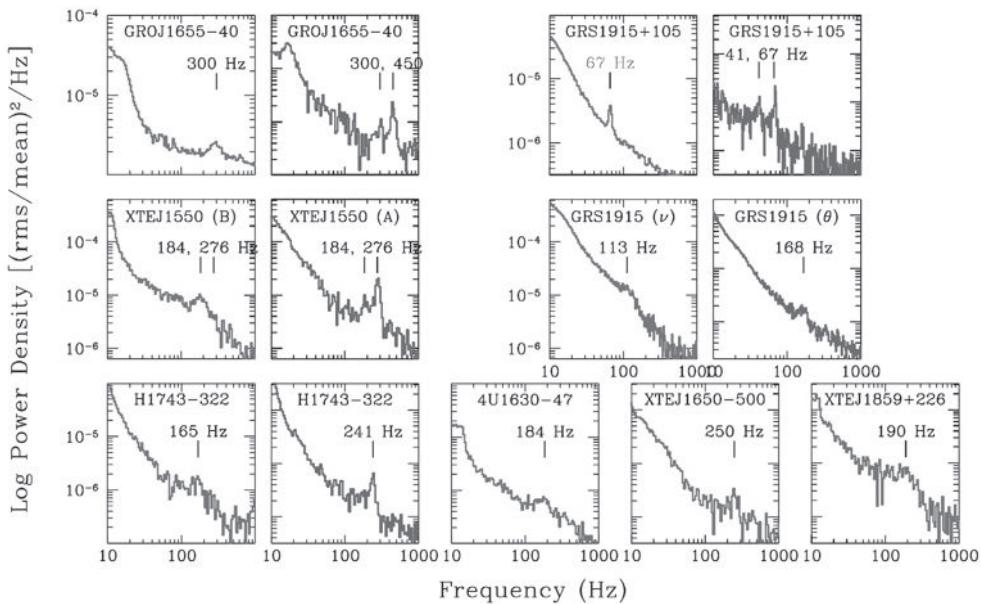
14.7.3 The steep power-law state

The origin of the steep power-law state is not well understood. In some sources the spectrum extends to 1 MeV or higher without a spectral break. This state is similar to the thermal state, both showing thermal and steep spectral components, but a very much more powerful power-law component is observed and QPOs have been detected, unlike the thermal state (see Fig. 14.25). The steep power-law state tends to dominate the binary black hole spectra as the luminosity approaches the Eddington limit and is associated with the high-frequency QPOs. Most models for this component assume that it involves Comptonisation, as illustrated by the spectra in Fig. 9.9. Consequently, there needs to be a source of low energy photons and hot corona associated with the accretion disc.

14.7.4 Quasi-periodic oscillations, QPOs

Quasi-period oscillations with low frequencies, 0.1–30 Hz, have been observed in many of the 20 black hole binaries. The amplitude of the power-spectrum of the oscillations generally peaks at photon energies greater than 10 keV and so must be related to the non-thermal components of the X-ray spectrum. The frequencies of the QPOs increase with X-ray luminosity during the outbursts. The models for these oscillations are similar to those discussed in Sect. 14.6.3 in the context of low mass X-ray binaries .

High frequency QPOs with oscillation frequencies in the range 40–450 Hz have been detected in seven sources (Fig. 14.27). These frequencies are of special importance because they are similar to the orbital frequency of matter on the last stable circular orbit about the black hole. For the case of a Schwarzschild black hole, the last stable orbit lies at $3r_g = 6GM/c^2$ and the velocity of matter on the last stable orbit is $c/2$ (see Sect. 20.5). Therefore, the orbital frequency of matter is roughly $2.7 \times 10^3 (M/M_\odot)^{-1}$ Hz. It can be seen that the frequencies of the high frequency QPOs would correspond to Schwarzschild black holes with masses of roughly $5\text{--}15 M_\odot$. A important feature of the high frequency QPOs is that they do not change in frequency with sizeable changes in luminosity, unlike the low frequency QPOs, suggesting that they are a more stable feature of the innermost regions of the accretion disc. Another feature observed in the power density spectra of four of the sources is the presence of pairs of QPOs with frequency ratios of 3:2. These different frequencies are not detected simultaneously, but they suggest some form of resonance behaviour, which is not understood.

**Fig. 14.27**

High frequency quasi-periodic oscillations observed in black hole binary and black hole candidate systems. Where two power density spectra are shown for a source, the right-hand trace is for the range 13–30 keV; the left-hand trace is for a broader energy range, which may be either 2–30 or 6–30 keV (Remillard and McClintock, 2006).

14.7.5 Iron fluorescence lines

In a number of black hole X-ray binaries, the broadened fluorescence line of neutral iron has been observed. The physics of the broadening of the fluorescence lines is discussed in more detail in the context of active galactic nuclei in Sect. 20.5, but the same considerations apply to black hole binaries. As explained in that section, the rest frequency of the iron fluorescence line is 6.4 keV but the line is asymmetrically broadened if it is emitted by an accretion disc which extends in to the last stable orbit about the black hole. The model for this process involves an X-ray corona which surrounds the accretion disc, the X-ray fluorescence line being excited by the incidence of the X-ray flux upon the cool material of the accretion disc. The asymmetric broadening of the line results from a combination of Doppler and gravitational redshifts. The asymmetric broadening of the line can be seen in Fig. 14.28, the profile on the low energy side of the line extending to about 3 keV (Miller *et al.*, 2002). Such an extreme broadening of the line can be naturally explained if the emission originates close to the last stable orbit about the black hole. In this case, the authors argue that the broadening is so extreme that a spinning black hole is necessary to account for the observed broadening (see Sect. 20.5).

These observations add compelling evidence to the view that X-ray observations of black hole binaries provide a unique and powerful tool for studying the effects of relativistic gravity in the strong field limit.

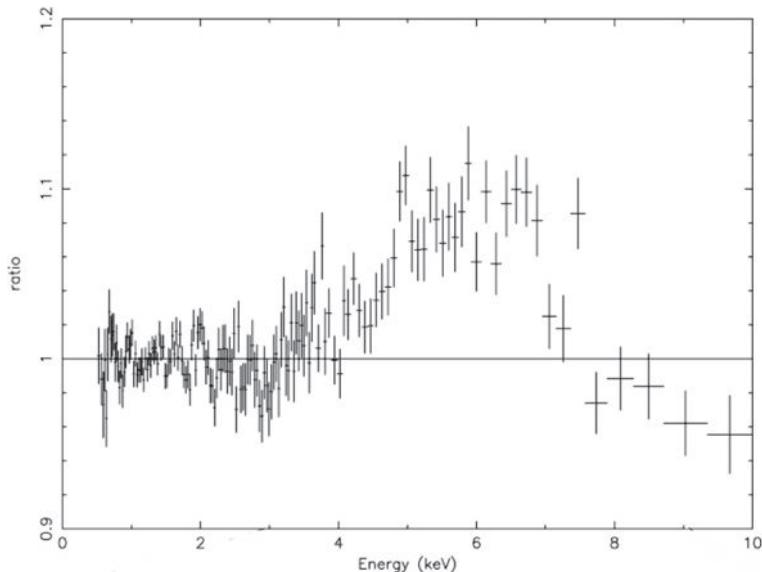


Fig. 14.28 The broad asymmetric iron fluorescence line at 6.4 keV in the black hole X-ray binary XTE J1650–500 (Miller *et al.*, 2002). A thermal disc component and power-law spectrum have been used to remove the underlying continuum.

14.8 Final thoughts

In this chapter, we have no more than scraped the surface of one of the most exciting and important areas of contemporary high energy astrophysics. I have selected topics in which the central role of accretion phenomena is well established, but these must be considered no more than an introduction to a vast area of contemporary research. There is undoubtedly the real possibility of studying the behaviour of matter in gravitational fields in the strong field limit and probing the innermost regions of accretion discs about Schwarzschild and Kerr black holes – and they are in our own backyard in the Galaxy.

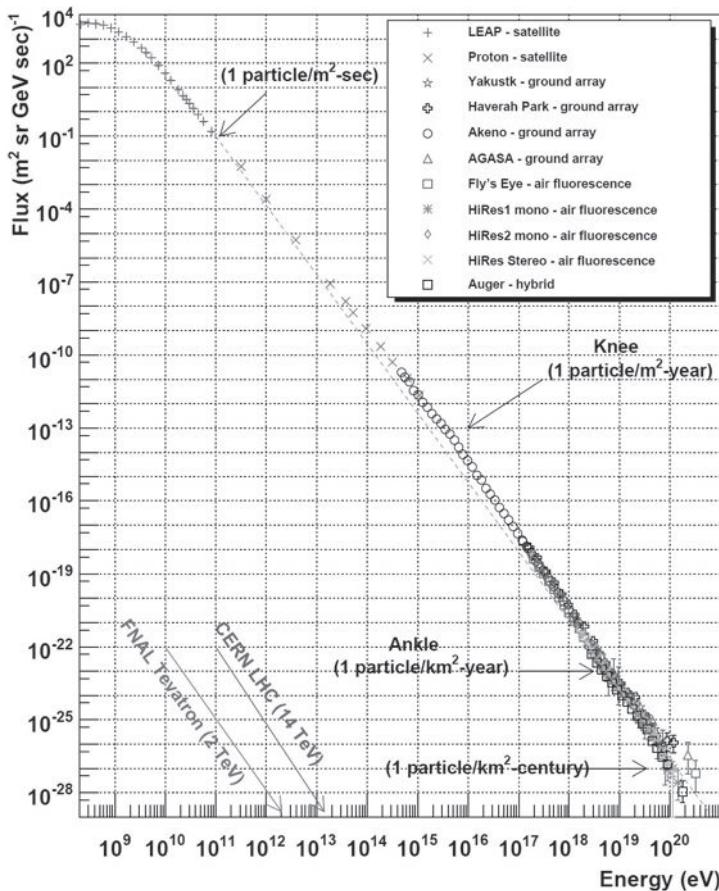
Cosmic ray protons, nuclei and electrons are the only particles which have been detected from sources outside the Solar System. As observed at the top of the atmosphere, about 98% of the particles are protons and nuclei whilst about 2% are electrons. Of the protons and nuclei, about 87% are protons, 12% are helium nuclei and the remaining 1% are heavier nuclei.

Figure 15.1 provides a quick overview of the complete cosmic ray spectrum. A very wide range of energies is observed and the spectrum can be described by power-law distributions over many decades in energy. There are, however, important features in the spectrum, including the ‘knee’ at 10^{15} eV and the ‘ankle’ at 10^{18} eV. It is convenient to consider first cosmic rays with energies in the range 10^9 – 10^{15} eV and then those with higher energies. This division corresponds to the different techniques which are used to detect the cosmic rays, particle detectors in space observatories in the lower energy range and the cosmic ray air-shower technique at the higher energies. These are distinguished by the change of symbols in Fig. 15.1 which shows the results of a large number of experiments.

15.1 The energy spectra of cosmic ray protons and nuclei

The energy spectra of cosmic rays can be well represented by power-law energy distributions as illustrated in Fig. 1.16, which shows the differential energy spectra for protons, helium, carbon and iron nuclei as a function of the kinetic energy per nucleon of the particles (Simpson, 1983). At energies less than about 1 GeV nucleon $^{-1}$, the energy spectra of all four species show a pronounced cut-off relative to the power law observed at high energies. The energy and shape of the cut-off vary with the phase of the solar cycle so that the fluxes of low energy particles decrease during periods of high solar activity and are at a maximum during phases of low solar activity. This phenomenon, known as *solar modulation*, is due to the fact that the cosmic ray particles diffuse in towards the Earth from interstellar space through the outflowing Solar Wind. The greater the solar activity, the greater the disturbances in the interplanetary magnetic field which impede the propagation of particles with energies less than about 1 GeV nucleon $^{-1}$ to the Earth. The solar modulation can be described by a diffusive-convective model in which the particles are scattered by the observed spectrum of magnetic irregularities in the Solar Wind (Jokipii, 1973, see also Sect. 7.4).

Once account is taken of the effects of solar modulation, the *differential energy spectra* of the various cosmic ray species can be well represented by power-law distributions of the

**Fig. 15.1**

The overall differential energy spectra of cosmic rays from various experiments. Prominent features in the spectrum are indicated – the ‘knee’ at 10^{15} eV and the ‘ankle’ at 10^{18} eV. The frequencies of arrival of particles of different energies are indicated, as well as the energies attainable in various accelerator experiments (from www.physics.utah.edu/~whanlon/spectrum.html).

form $N(E) dE = KE^{-x} dE$ in the energy range $1 < E < 10^4$ GeV nucleon $^{-1}$ as can be seen in Fig. 1.16. A more recent version of the energy spectra of different cosmic ray species is shown in Fig. 15.2 which extends to the highest energies at which clear discrimination of their charges can be achieved. At higher energies, the energy spectra are obtained from cosmic ray air-shower data and the identification of the species is estimated from the signatures of the masses of the particles in the air-shower data. The intensity of primary nucleons in the energy range from several GeV to somewhat beyond 100 TeV is given approximately by

$$I_N(E) \approx 1.8 \times 10^4 E^{-x} \text{ nucleons m}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ GeV}^{-1}, \quad (15.1)$$

with $x = 2.7$ and E the energy per nucleon in GeV. There are, however, significant differences between the energy spectra of different elements. The general trend is that those

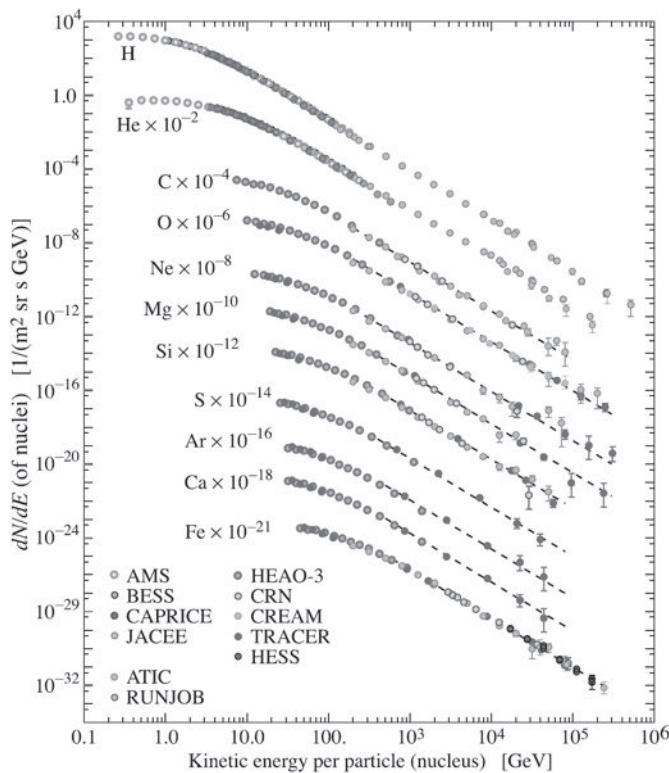
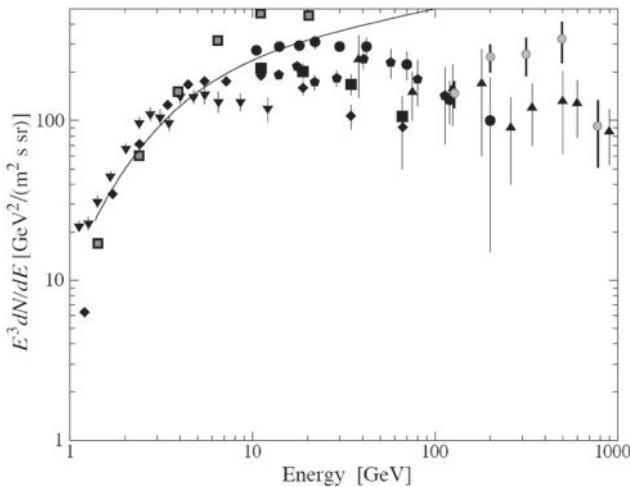


Fig. 15.2 The differential energy spectra of different cosmic ray species as a function of kinetic energy (Amsler *et al.*, 2008). Note that these data refer to the energy per particle rather than the energy per nucleon (see Table 7.1). The scaling factors used to display the spectra of the different elements are clearly shown on the diagram.

species which are identified as primary species, meaning those accelerated in the sources, have somewhat flatter spectra than those identified as secondary species, meaning that they are primarily created in nuclear collisions of the primary species with the nuclei of atoms and molecules of the interstellar gas, the process known as *spallation*. This result is interpreted as differences in the path lengths for spallation as a function of particle energy (see Sect. 15.5.2).

The energy spectra of *cosmic ray electrons* have been determined both from high flying balloon experiments and from satellites. In Fig. 15.3, the differential energy spectrum is displayed in the form $E^3 N(E)$ so that a power-law distribution of electron energies $N(E) \propto E^{-3}$ would correspond to a straight line parallel to the abscissa. The spectrum of the electrons is strongly influenced by the effects of solar modulation at energies $E \lesssim 1$ GeV, but at energies greater than about 10 GeV the observations are free of these effects. In the high energy range, $E \geq 10$ GeV, the observations can be described by a spectrum of power-law form given by Webber (1983),

$$N(E) dE = 700 E^{-3.3} dE \text{ particles m}^{-2} \text{s}^{-1} \text{sr}^{-1}, \quad (15.2)$$

**Fig. 15.3**

Differential spectrum of cosmic ray electrons plus positrons plotted as $(dN/dE) E^3$. The line shows the proton spectrum multiplied by 0.01 (Amsler *et al.*, 2008).

where the energies of the particles are measured in GeV. This spectrum is somewhat steeper than that of cosmic ray protons and nuclei which is shown by the solid line in Fig. 15.3. This spectrum has, however, been significantly influenced by the effects of synchrotron radiation losses and hence probably does not reflect the injection spectrum of the electrons. Fortunately, there is complementary information about the spectrum of ultra-relativistic electrons in the interstellar medium from observations of their synchrotron radio emission (Sect. 8.9).

The positron fraction decreases from about 0.2 below 1 GeV to about 0.1 at 2 GeV and to about 0.05 at the highest energies for which it is measured (5–20 GeV). The ratio of antiprotons to protons is $\sim 2 \times 10^{-4}$ at about 10–20 GeV. The \bar{p}/p ratio shows a strong dependence on the phase and polarity of the solar cycle in the opposite sense to that of the positron fraction, suggesting that the antiparticles are secondary products. There is no evidence for a significant primary component of either positrons or of antiprotons.

15.2 The abundances of the elements in the cosmic rays

The chemical abundances of the cosmic rays provide important clues to their origin and to the processes of propagation from their sources to the Earth. These can be compared with the Solar System and interstellar abundances of the elements.

15.2.1 The Solar System abundances of the elements

The determination of the chemical abundances of the elements is a vast subject, introductions to the subject being provided by Tayler in *The Origin of the Chemical Elements*

(1972) and the monograph by Pagel *Nucleosynthesis and Chemical Evolution of Galaxies* (1997). The problem is that there are enormous differences in the elemental abundances of a representative sample of terrestrial material as compared with those found on the surface of the Moon, the outer layers of the Sun or in a typical sample of interstellar gas. Each sample has been influenced to a greater or lesser extent by its past history. The best estimates are derived from the photospheric abundances of the Sun, from meteoric samples and from estimates of chemical abundances in the interstellar gas. The meteoritic abundances have the advantage that they have not undergone fractionation and other terrestrial geophysical phenomena and are believed to have chemical abundances similar to that of the primitive solar nebula. The chondritic meteorites have chemical abundances which are similar to those found in the solar photosphere, and so the abundances in these meteorites can be used to fill in the abundances of some of the rarer elements in the periodic table. In Table 15.1, abundances are given for the Solar System based upon studies of the solar atmosphere and meteorites. Also shown are estimates by Meyer (1979) of local interstellar abundances which are in reasonable agreement with the adopted Solar System abundances.

What guarantee is there that these abundances are typical of ordinary matter in the Universe? Generally, for stars like the Sun, there is not a great deal of variation in the chemical abundances. The Sun is about 4.6×10^9 years old compared with the age of the Galaxy, which is at least twice that age, and so it was formed from interstellar material which had already been rather thoroughly processed by previous generations of massive star formation. Studies of the evolution of the chemical abundances of the elements show that the typical cosmic abundances of the elements are attained rather rapidly and this strongly suggests that on average the abundances found in stars like the Sun may well be fairly typical of the average cosmic abundances in the Galaxy at the present day. For our present purposes, we adopt the figures given in Table 15.1 as representative of the chemical abundances of the matter in stars at the present epoch.

15.2.2 The chemical abundances in the cosmic rays

A wide range of data on the chemical abundances in the cosmic rays has been summarised by Simpson (1983). Many of the best data were obtained by the French–Danish experiment flown in the HEAO-C2 experiment and from balloon experiments in which large detector packages were flown at high altitude. Simpson's summary of the abundances of the elements in the cosmic rays in different energy ranges is compared with the Solar System abundances and estimated abundances for the local interstellar medium in Table 15.1. These data are displayed graphically in Fig. 15.4 which reveals the overall similarities and differences between the abundances in the cosmic rays and those of typical Solar System and interstellar matter. The following features are apparent:

- the abundance peaks at carbon, nitrogen and oxygen and at the iron group are present both in the cosmic ray and Solar System abundances;

Table 15.1 Galactic cosmic ray elemental abundances at 1 AU, normalised to Si = 100, compared with the Solar System and local interstellar abundances. The energy intervals are given in MeV nucleon⁻¹.

Element	Cosmic rays		Average at 1000–2000	Solar System	Local Galactic
	70–280	600–1000			
He	41700 ± 3000	27030 ± 580			(0.27 ± 0.06) × 10 ⁶
Li	100 ± 6	136 ± 3		5.0 × 10 ⁻³	
Be	45 ± 5	67 ± 2	69.4 ± 10	8.1 × 10 ⁻⁵	
B	210 ± 9	233 ± 4	212 ± 10	3.5 × 10 ⁻³	
C	851 ± 29	760 ± 16	684 ± 27	1110	1300 ± 300
N	194 ± 8	208 ± 5	188 ± 6	231	230 ± 100
O	777 ± 28	707 ± 15	607 ± 28	1840	2300 ± 500
F	18.3 ± 1.3	17.0 ± 1.1	13.5 ± 2.3	0.078	0.093 (1.6)
Ne	112 ± 6	113 ± 3	100 ± 3	240	270 (1.7)
Na	27.3 ± 3.4	25.8 ± 1.1	21.3 ± 3.2	6	5.6 ± 0.9
Mg	143 ± 6	142 ± 4	125 ± 12	106	105 ± 3
Al	25.2 ± 3.0	28.2 ± 1.2	22.2 ± 3.2	8.5	8.4 ± 0.4
Si	100	100	100	100	100 ± 3
P	4.0 ± 0.7	5.3 ± 0.5	5.3 ± 1.6	0.65	0.96 ± 0.20
S	16.4 ± 1.2	23.1 ± 1.1	19.6 ± 0.9	50	45 ± 13
Cl	3.6 ± 0.5	6.4 ± 0.5	4.7 ± 0.4	0.47	0.47 (1.6)
A	6.3 ± 0.6	10.2 ± 0.7	8.2 ± 1.2	10.6	9.0 (1.7)
K	5.1 ± 0.6	7.2 ± 0.5	6.3 ± 0.4	0.35	0.36 ± 0.12
Ca	13.5 ± 1.0	16.1 ± 0.9	13.1 ± 1.2	6.25	6.2 ± 0.8
Sc	2.9 ± 0.5	4.5 ± 0.5	3.3 ± 1.1	0.003	0.0035 ± 0.0005
Ti	10.7 ± 0.9	10.2 ± 0.7	9.1 ± 0.9	0.24	0.27 ± 0.04
V	5.7 ± 0.6	6.7 ± 0.5	4.6 ± 0.3	0.025	0.026 ± 0.005
Cr	10.9 ± 1.0	11.8 ± 0.8	9.1 ± 0.8	1.27	1.30 ± 0.12
Mn	7.2 ± 1.2	8.2 ± 0.7	6.3 ± 0.4	0.93	0.79 ± 0.17
Fe	60.2 ± 3.2	69.8 ± 2.0	60.5 ± 7.6	90.0	88 ± 6
Co	0.2 ± 0.1		0.4 ± 0.2	0.22	0.21 ± 0.03
Ni	2.9 ± 0.4	3.7 ± 0.5	2.8 ± 0.6	4.78	4.8 ± 0.6
Cu			0.038 ± 0.006		0.052 (1.6)
Zn			0.035 ± 0.005		0.135 (1.6)

The cosmic ray data are taken from the review by Simpson (1983) who gives detailed references to the sources of these data. The data on the Solar System and local interstellar abundances are taken from the same review, supplemented by data from Cameron (1973).

- the odd–even effect in the relative stabilities of the nuclei according to atomic number known to be present in the Solar System abundances of the elements is also present in the cosmic rays but to a somewhat lesser degree;
- the light elements, lithium, beryllium and boron are grossly overabundant in the cosmic rays relative to their Solar System abundances;

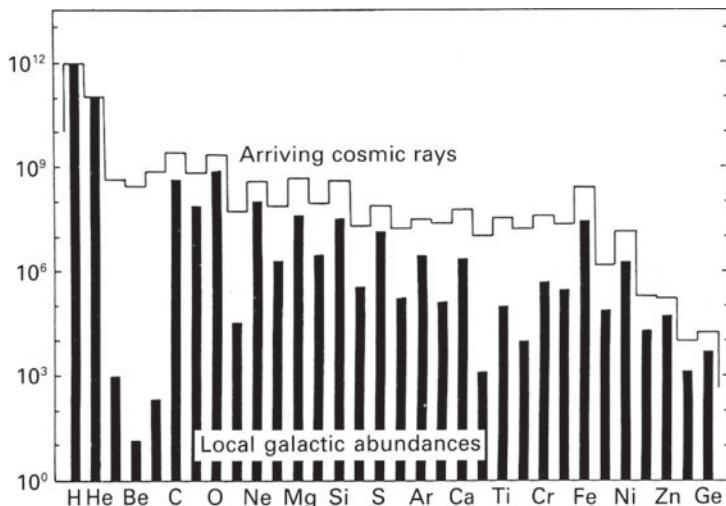


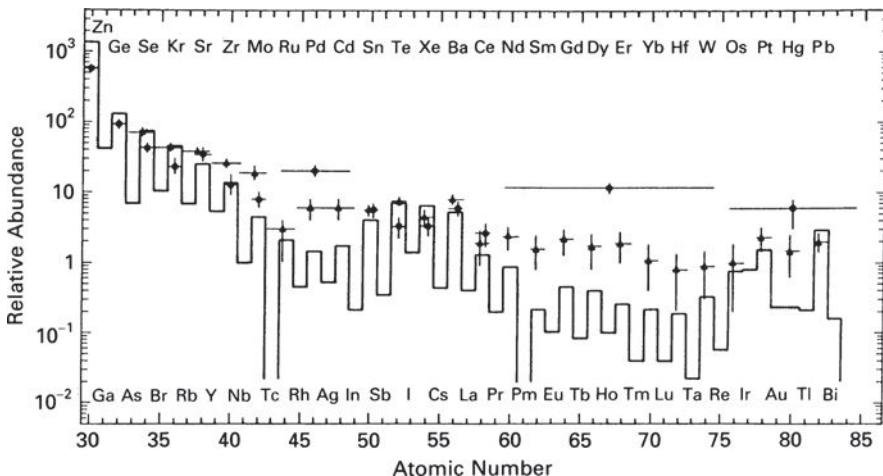
Fig. 15.4

The cosmic abundances of the elements in the cosmic rays (solid line) compared with the Solar System abundances (solid histogram). The data have been normalised to a relative abundance of hydrogen of 10^{12} (Lund, 1984).

- there is an excess abundance in the cosmic rays of elements with atomic and mass numbers just less than those of iron, that is, elements between about calcium and iron;
- there is an underabundance of hydrogen and helium in the cosmic rays relative to the heavy elements.

Overall, the distribution of element abundances in the cosmic rays is not so different from those of typical Solar System abundances. Some of the differences, specifically items (iii) and (iv) listed above, can be attributed to spallation in the interstellar medium between the sources of the cosmic rays and their arrival at the Earth. In these collisions, the cosmic ray nuclei are chipped away and fragmented, resulting in the production of nuclei with atomic and mass numbers less than those of the primary nuclei.

The abundances shown in Table 15.1 extend to just beyond the iron peak. The HEAO-C3 and BUGS experiments were designed to measure the abundances of elements beyond the iron peak, up to elements such as uranium. There is good agreement between these two experiments, a summary of the data being displayed in Fig. 15.5. The resolution of the experiments was not sufficient to resolve individual elements and so the data have been binned. In the case of the HEAO-C3 experiment, the bins shown in Fig. 15.5 are quite broad; in the case of the Ariel-VI data, the data are presented for even nuclei. The areas of good agreement are the abundances of heavy cosmic rays with atomic numbers in the range up to about 42, the peak at $50 < Z < 58$ and the high Z peak at $78 < Z < 84$. Between the peaks, the Ariel-VI observations show somewhat larger abundances of the elements in the cosmic rays as compared with the HEAO results. Between these peaks, the cosmic ray abundances are significantly greater than the Solar System abundances of these elements. Again, a natural explanation for this phenomenon is the spallation of elements in the peaks as they encounter the atoms and molecules in the interstellar gas. The abundances of the

**Fig. 15.5**

The abundances of the elements with very large atomic numbers in the cosmic rays. The abundances observed in the Ariel-VI observations are indicated by filled triangles and the HEAO-C3 observations by filled circles. The solid line indicates the cosmic abundances of the elements. In each case, the abundances are normalised to 10^6 iron nuclei (Wefel, 1988).

heavy primary elements relative to iron are similar in both the cosmic rays and in the Solar System.

15.2.3 Isotopic abundances of cosmic rays

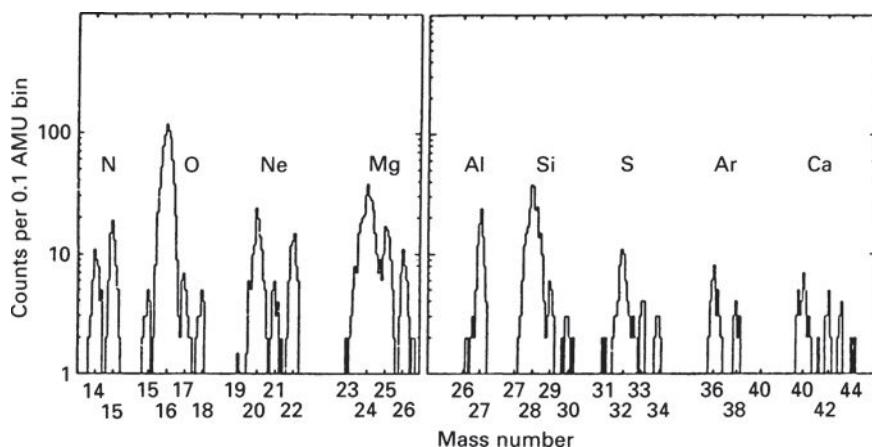
In addition to overall chemical abundances, isotopic abundances are available for a number of species. The very lightest stable elements, ${}^1\text{H}$, ${}^2\text{H}$, ${}^3\text{He}$ and ${}^4\text{He}$, form a special group of isotopes. Most of the helium was synthesised in the Hot Big Bang through the p p chain as the matter and radiation cooled down through a temperature of about 10^9 K – this process is described in all the standard textbooks on cosmology (see, for example, Longair, 2008). About 24% of helium by mass, corresponding to a helium-to-hydrogen ratio by number of about 0.08:1, is synthesised in the early stages of the Big Bang for a wide range of reasonable initial conditions. The local interstellar abundances of these elements are compared with the cosmic ray abundances in Table 15.2. ${}^2\text{H}$ and ${}^3\text{He}$ are present in very much greater abundances in the cosmic rays than they are in the interstellar medium and can be attributed to spallation reactions between the four species concerned. These elements provide an independent check of the spallation models – their abundances are so much greater than those of the other elements that the spallation products of elements such as carbon, oxygen and nitrogen contribute little to the relative isotopic abundances of hydrogen and helium.

Some of the species created in spallation reactions are radioactive and hence, if the production rates of the different isotopes of a given element are known, information can be obtained about the time that it has taken these samples to reach the Earth from their sources.

Table 15.2 Isotope ratios of hydrogen and helium.

Isotope ratio	60 MeV nucleon ⁻¹	80 MeV nucleon ⁻¹	200 MeV nucleon ⁻¹	Cosmic abundance
² H/ ¹ H		(4.4 ± 0.5) × 10 ⁻²	(5.7 ± 0.5) × 10 ⁻²	1.0 × 10 ⁻⁵
³ He/ ⁴ He		(9.5 ± 1.5) × 10 ⁻²	(11.8 ± 0.7) × 10 ⁻²	3.0 × 10 ⁻⁵
² H/ ⁴ He	0.21 ± 0.09		0.31 ± 0.03	1.0 × 10 ⁻⁴

All measurements except those at 60 MeV nucleon⁻¹ were made near solar minimum. The figures in the table are the *number ratios* of isotopes rather than their mass ratios (Simpson 1983).

**Fig. 15.6**

Examples of the determination of the isotopic abundances of the elements nitrogen, oxygen, neon, magnesium, aluminium, silicon, sulphur, argon and calcium. The charge resolutions in these observations ranges from 0.23 amu for oxygen to 0.29 amu for silicon (Webber *et al.*, 1985).

The most famous of these ‘cosmic ray clocks’ is the isotope ¹⁰Be which has a radioactive half-life of 1.5×10^6 years and so is a very useful discriminant for determining the typical lifetime of the spallation products in the vicinity of the Earth.

A third aspect of isotopic abundances concerns the sources of the cosmic rays. Quite subtle tests of the origin of the cosmic rays are possible with the ability to determine the isotopic abundances of the heavy elements. It is found that the most common isotopes of the heavy elements are the same as the most common isotopes found in the Solar System and in the local interstellar medium but, in a number of cases, significantly greater abundances of relatively rare isotopes are found among the cosmic ray particles. Examples of these are shown in Fig. 15.6 in which the neon, magnesium and silicon isotopes have been clearly resolved. The advantage of using isotopic abundances is that the differential effects of spallation should be a second-order effect because the cross-sections for spallation should not be very different for isotopes of the same element. Thus, the isotopic ratios of a particular heavy element should be relatively insensitive to the spallation history of the species. In

the example shown in Fig. 15.6, the ratio of $^{22}\text{Ne}/^{20}\text{Ne}$ is about four times greater than the value found in the Solar System. For elements which have been well studied, there is a greater abundance of neutron-rich isotopes as compared with the Solar System abundances. For example, the isotopic abundances of $^{25}\text{Mg}/^{24}\text{Mg}$, $^{26}\text{Mg}/^{24}\text{Mg}$, $^{29}\text{Mg}/^{28}\text{Mg}$ and $^{30}\text{Si}/^{28}\text{Si}$ are each about 1.6 times greater in the cosmic rays than they are in the Solar System abundances. Evidently, a source is required which favours the production of neutron-rich elements.

15.3 The isotropy and energy density of cosmic rays

Figure 1.16 shows that the fluxes of cosmic rays with energies less than about 10 GeV are significantly influenced by solar modulation and hence information about the arrival directions of these cosmic rays at the Solar System is lost. In fact, only relatively high energy protons and nuclei penetrate to the vicinity of the Earth undeflected by the magnetic field in the interplanetary medium. A measure of the deflection suffered by a particle is the ratio of its radius of gyration in the interplanetary magnetic field to the scale of the Solar System. For a relativistic proton, the gyroradius is

$$r_g = 3 \times 10^9 \gamma (B/10^{-9} \text{ T}) \text{ m} \quad (15.3)$$

where the magnetic flux density B is measured in tesla and $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor (Sect. 7.1). Adopting the local value of the magnetic flux density in the interplanetary medium $B = 10^{-9}$ T, relativistic protons with $\gamma = 10^3$, that is, energies of 10^{12} eV, have gyroradii which are $3 \times 10^{12} = 20$ AU, that is, 20 times the distance from the Sun to the Earth, corresponding roughly to the radius of the orbit of the planet Uranus about the Sun. Thus, particles with these energies and greater are likely to preserve information about their arrival directions at the Solar System when they arrive at the top of the Earth's atmosphere.

Observations of the arrival directions of high energy protons and nuclei can be undertaken using the Earth's atmosphere itself as a 'convertor'. Very high energy protons and nuclei create pions in collisions with the nuclei of atoms and molecules in the upper layers of the atmosphere. The charged pions decay into muons which have half-lives such that they decay long before they reach the surface of the Earth unless they are highly relativistic, $\gamma \geq 20$ (Sect. 10.4.1). This process is the basis of studies of the isotropy of cosmic rays using underground muon detectors which are designed to be sensitive to primary particles entering the atmosphere with energies of 10^{12} eV and greater.

These data are summarised together with data on much higher energy particles in Fig. 15.7 which is taken from the review by Hillas (1984) (see also Wdowczyk and Wolfendale, 1989). In the range 10^{13} – 10^{14} eV, the distribution of arrival directions of high energy particles is remarkably uniform, anisotropy only being detected at the level of less than 1 part in 10^3 . Quantitatively, the amplitude of the best-fitting first harmonic function to the distribution of arrival directions of these cosmic rays corresponds to 0.06%. Thus, the net streaming velocity of the flux of high energy particles relative to the local frame of reference in our

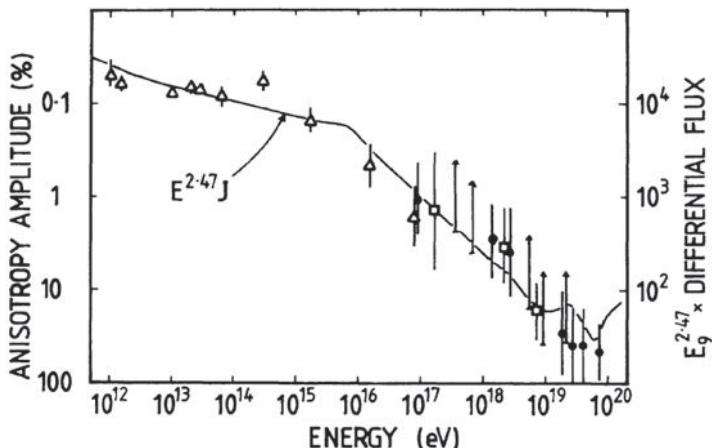


Fig. 15.7 The amplitude of the anisotropy in the distribution of arrival directions of cosmic rays as a function of energy. In each case, a best-fitting sine wave has been fitted to the data and the percentage amplitude of this harmonic measured. The solid line shows the shape of the differential spectrum of cosmic rays (Hillas, 1984).

Galaxy is small. This, in turn, has implications for the diffusion of high energy particles in the interstellar medium and also for the escape of these particles from the Galaxy.

We obtain a lower limit to the local energy density in high energy particles by assuming that the flux of particles observed at the top of the atmosphere is representative of that present in the local interstellar medium. Although the spectrum of the cosmic rays extends to very high energies, there is little total energy in these because of the steepness of the energy spectrum of the particles, $dN \propto E^{-2.7} dE$. The maximum of the observed proton spectrum corresponds to about 2 protons $m^{-2} s^{-1} sr^{-1} MeV^{-1}$ at an energy of about 1 GeV. Wdowczyk and Wolfendale (1989) find that the total energy density of cosmic rays with energies greater than 1 GeV is about $1 MeV m^{-3}$ (or $1 eV cm^{-3}$). This energy density is similar to that present in the interstellar magnetic field, $B^2/2\mu_0 \approx 0.2 MeV m^{-3}$. It is also similar to the local energy density in starlight (about $0.3 MeV m^{-3}$) and to the energy density of the Microwave Background Radiation (about $0.3 MeV m^{-3}$). Some of these coincidences have real physical significance while others have no deep astrophysical significance.

15.4 Gamma ray observations of the Galaxy

Just as the Galactic radio emission traces the distribution of high energy electrons and magnetic fields in the Galaxy, so the distribution of γ -radiation can provide information about high energy protons and the overall distribution of interstellar gas. In collisions between high energy particles and protons and nuclei of atoms and molecules of the interstellar gas, pions of all charges, π^+ , π^0 and π^- are produced. The positive and

negative pions decay into positive and negative muons which in turn decay into positrons and electrons with relativistic energies (see Fig. 10.9). The latter may make a contribution to the low energy electron spectrum and the predicted presence of positrons provides a direct test of the importance of the pion production mechanism in interstellar space. The neutral pions decay almost instantly into two γ -rays. In proton–proton collisions, the cross-section for the production of a pair of high energy γ -rays is roughly the geometric size of the proton, $\sigma_\gamma \approx 10^{-30} \text{ m}^2$. The spectrum of γ -rays produced in such collisions has a broad maximum centred about 70 MeV and is a characteristic signature of the neutral pion decay process.

The γ -ray image of the Galaxy is shown in Fig. 1.14 from observations made by the EGRET instrument of the Compton Gamma-ray Observatory. Numerous discrete γ -ray sources have been detected but there is also diffuse γ -ray emission from the plane of the Galaxy. In addition to the π^0 process, there are expected to be contributions from inverse Compton scattering of starlight and relativistic bremsstrahlung of ultra-relativistic electrons. Thus, in analysing the γ -ray maps of the sky, it is necessary to achieve consistency with the distribution of relativistic electrons as defined by the radio map of the sky (Fig. 1.9) and the distribution of neutral and molecular gas (Fig. 1.8) which are the main targets for the cosmic ray protons.

Let us first make a crude estimate of the γ -ray luminosity of the Galaxy due to the π^0 decay process. Stecker (1975) found the γ -ray luminosity of the Galaxy to be roughly 10^{32} W at γ -ray energies $\varepsilon \geq 100 \text{ MeV}$ from analysis of the SAS-2 data. The probability P_{coll} that a high energy proton undergoes an inelastic collision per second with a nucleus of the interstellar gas is

$$P_{\text{coll}} = \sigma_{\text{pp}} N c, \quad (15.4)$$

where σ_{pp} is the proton–proton inelastic cross-section which is about $2.5 \times 10^{-30} \text{ m}^2$, N the number density of the interstellar gas and it is assumed that the proton travels at the speed of light. Roughly one-third of the pions produced in each collision are neutral pions which decay into two γ -rays with mean energy 180 MeV. Therefore, if we suppose the disc of the Galaxy has volume V and is filled uniformly with interstellar gas and high energy particles, the number of collisions per second of high energy protons of energy E is $V N_{\text{CR}}(E) \sigma_{\text{pp}} N c$ where $N_{\text{CR}}(E)$ is the number density of high energy particles with energy E . Therefore the total energy liberated as γ -rays is roughly

$$L_\gamma = \frac{1}{3} \sigma_{\text{pp}} N c \sum N_{\text{CR}}(E) E = \frac{1}{3} P_{\text{coll}} \epsilon_{\text{CR}} V, \quad (15.5)$$

where ϵ_{CR} is the local energy density of high energy particles. Taking the disc to have half-thickness 200 pc and radius 8 kpc, $V \approx 2 \times 10^{60} \text{ m}^3$. Adopting $N = 10^6 \text{ m}^{-3}$ and $\epsilon_{\text{CR}} \approx 10^6 \text{ eV m}^{-3}$, $L_\gamma \approx 10^{32} \text{ W}$, in rough agreement with the total observed γ -ray luminosity of the Galaxy.

In an ambitious programme, Strong and his colleagues (2004) carried out a detailed analysis of the γ -ray maps of the sky derived from the EGRET observations and combined these with analyses of the radio synchrotron maps of the sky, estimates of the diffuse background radiation in the optical and infrared waveband and the distribution of neutral

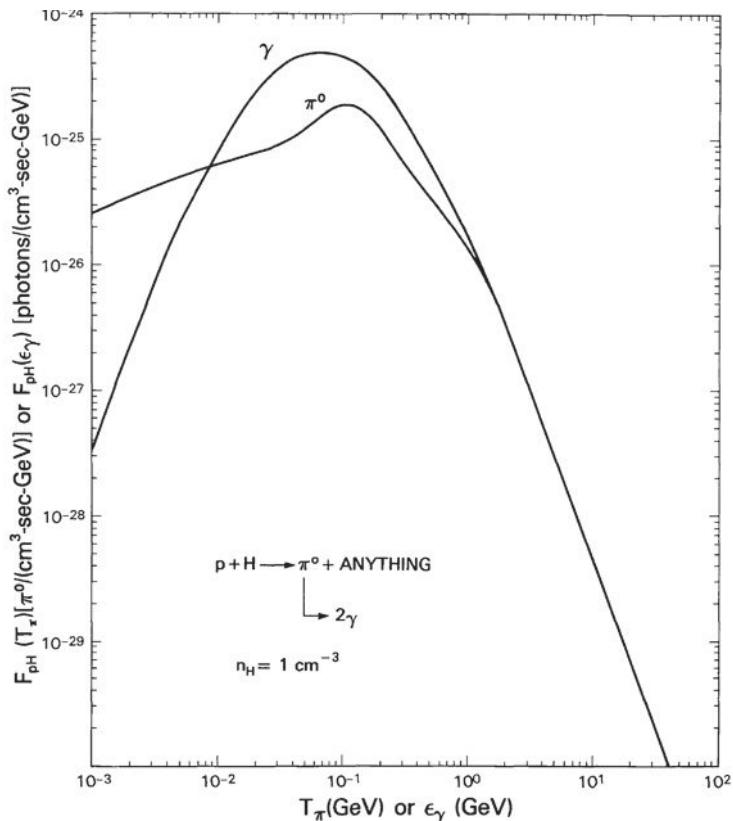


Fig. 15.8 The secondary π^0 and γ -ray emissivities resulting from the interaction of the local demodulated cosmic ray proton spectrum with unit density of atomic hydrogen (Dermer, 1986).

and molecular gas in the plane of the Galaxy. To estimate the γ -ray emission due to interstellar π^0 production, they use the estimates of the π^0 production rates and resulting γ -ray emission by Dermer (1986), improving the pioneering calculations of Stecker (1975).¹ Dermer's calculations show the predicted intensity spectrum of γ -rays per unit volume assuming the local spectrum of cosmic ray protons, corrected for the effects of solar modulation (Fig. 15.8). The models of Strong and his colleagues include the emission of secondary electrons and positrons, as well as being constrained by antiproton production. The results of these computations for different regions of the Galactic plane and a region at high Galactic latitudes is shown in Fig. 15.9. Strong and his colleagues emphasise that their model has achieved good agreement with a very wide range of observational data but make no claim for the uniqueness of the result. The computations make the important point

¹ For more details of these and other calculations, see the monograph *High Energy Radiation from Black Holes: Gamma Rays, Cosmic Rays, and Neutrinos* by Dermer and Menon (2009).

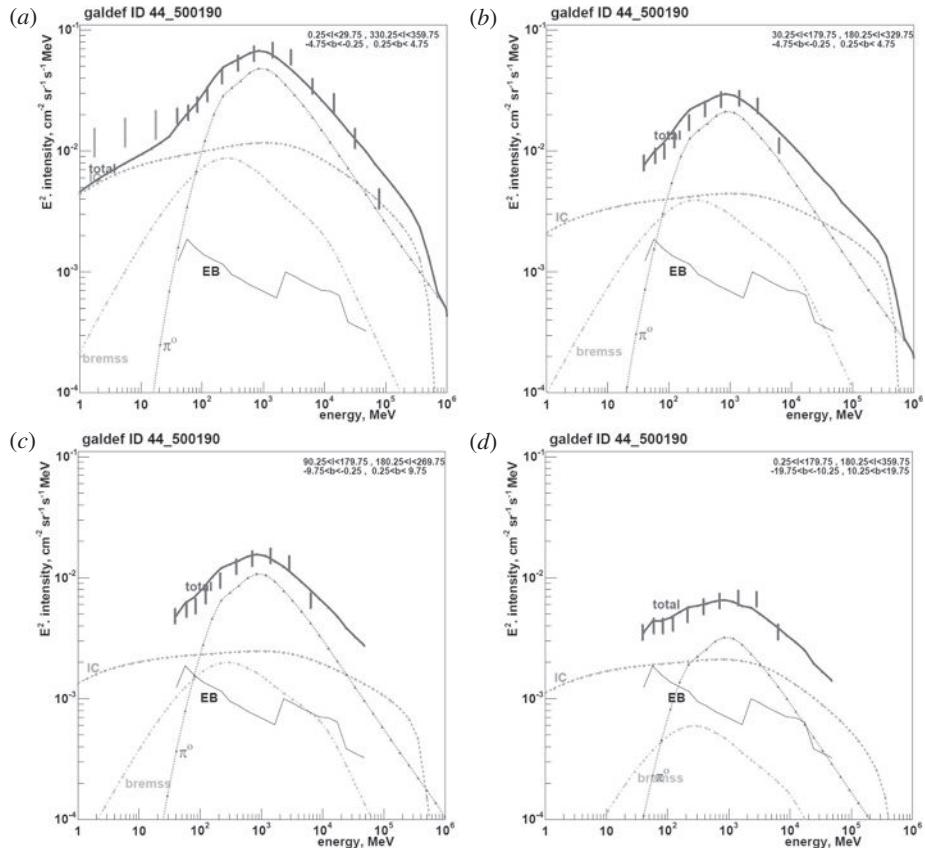


Fig. 15.9 Comparison of the observed and predicted γ -ray spectra in different regions of the Galaxy (Strong *et al.*, 2004). The regions are: (top left, *a*) the inner Galaxy $330^\circ < l < 30^\circ$, $|b| = 0\text{--}5^\circ$; (top right, *b*) Galactic plane avoiding inner Galaxy $30^\circ < l < 330^\circ$, $|b| = 0\text{--}5^\circ$; (bottom left, *c*) outer Galaxy $90^\circ < l < 270^\circ$, $|b| = 0\text{--}10^\circ$; (bottom right, *d*) intermediate latitudes $0^\circ < l < 360^\circ$, $|b| = 10\text{--}20^\circ$. The observations are shown by vertical bars. The predicted contributions of π^0 decays, inverse Compton scattering (IC) and bremsstrahlung (brems), as well as the inferred extragalactic background component, are shown by different symbols.

that satisfactory models need to accommodate a very wide range of different observational data and constraints.

From our perspective, the important points are illustrated by the γ -ray spectra in panels *a*, *b* and *c* of Fig. 15.9. The γ -ray spectra at low Galactic latitudes all show the characteristic π^0 decay peak at about 70 MeV, confirming that this is the dominant process in these regions. In contrast, at higher galactic latitudes (panel *d*), the inverse Compton scattering process becomes much more important and this trend continues to high Galactic latitudes. The important point is that the γ -ray observations provide compelling evidence that the local flux of cosmic rays is similar to that present throughout the plane of the Galaxy and so the γ -ray maps in these regions provide tracers of the distribution of high energy protons and nuclei and of interstellar matter in the Galaxy.

15.5 The origin of the light elements in the cosmic rays

The elemental abundances in the cosmic rays were compared with the Solar System abundances in Tables 15.1 and 15.2 and Fig. 15.4. Three obvious differences need to be explained, in each case the cosmic rays having much greater elemental abundances as compared with the Solar System abundances. These are:

- the light elements lithium, beryllium and boron;
- the ratio of ${}^3\text{He}$ to ${}^4\text{He}$;
- the elements just lighter than iron.

These cosmic ray nuclei can be created as fragmentation products of the interactions of high energy nuclei with cold interstellar matter, the process known as *spallation*. The table of fragmentation probabilities describing the ways in which heavy nuclei are split up in inelastic collisions with hydrogen or helium nuclei in the interstellar gas was presented in Sect. 10.2 and in Tables 10.1a and b. We begin with the transfer equation for cosmic rays, including the effects of spallation.

15.5.1 The transfer equation for light nuclei

The diffusion-loss equation (7.41) for the propagation of high energy nuclei through the interstellar gas in the presence of fragmentation gains and loss of a particular species i was discussed in Sect. 7.5,

$$\frac{\partial N_i}{\partial t} = D \nabla^2 N_i + \frac{\partial}{\partial E} [b(E)N_i] + Q_i - \frac{N_i}{\tau_i} + \sum_{j>i} \frac{P_{ji}}{\tau_j} N_j . \quad (15.6)$$

N_i is the number density of nuclei of species i and is a function of energy; the term $D \nabla^2 N_i$ is the diffusion term; $\partial/\partial E[b(E)N_i]$ takes account of the effect of energy gains and losses upon the energy spectrum of the particles; and Q_i is the rate of injection of particles of species i from sources per unit volume. The last two terms describe the effects of spallation gains and losses. τ_i and τ_j are the spallation lifetimes of particles of species i and j . The spallation products of all species with $j > i$ provide contributions to N_i , as indicated by the sum in the last term of equation (15.6). P_{ji} is the probability that, in an inelastic collision involving the destruction of the nucleus j , the species i is created.

First of all, we neglect diffusion and energy losses. The light nuclei have very low cosmic abundances and so it can be assumed that there is no injection of these particles from the sources of high energy particles, $Q_i = 0$. It is convenient to write the transfer equation in terms of the number of kg m^{-2} traversed by the particles, as described in Sect. 5.4.2 for the case of ionisation losses. The path length through the interstellar gas is $\xi = \rho x = \rho v t$, where v is the velocity of the particle and ρ the gas density. Therefore, the simplified transfer equation for light nuclei is

$$\frac{dN_i(\xi)}{d\xi} = -\frac{N_i(\xi)}{\xi_i} + \sum_{j>i} \frac{P_{ji}}{\xi_j} N_j(\xi) , \quad (15.7)$$

where ξ_i and ξ_j are the mean free paths for inelastic collisions expressed in kg m^{-2} . In adopting this transfer equation, it is assumed that *all* the particles traverse the same amount of material between 0 and ξ , that is, there is a one-to-one relation between path length ξ and species produced. This is referred to as a *slab model* and is clearly an oversimplification. To follow the evolution of the chemical abundances of all species, equations similar to (15.7) are written for the isotopes of every element in the periodic table, so that, even without taking account of the other terms in the full diffusion-loss equation, there is a fairly large matrix to be inverted.

From the table of abundances of the cosmic rays (Table 15.1), carbon, nitrogen and oxygen are far more abundant than all the other heavy elements and therefore the spallation of this group of elements, which are referred to as the *medium* or M group of elements, are bound to be the main progenitors of the light or L elements, lithium, beryllium and boron.

Initially there are no particles in the L group at $\xi = 0$. The differential equations which describe how the abundances of the L and M groups of elements change with path length are

$$\frac{dN_M(\xi)}{d\xi} = -\frac{N_M(\xi)}{\xi_M} \quad ; \quad \frac{dN_L(\xi)}{d\xi} = -\frac{N_L(\xi)}{\xi_L} + \frac{P_{ML}}{\xi_M} N_M(\xi). \quad (15.8)$$

Integrating the first equation of (15.8),

$$N_M(\xi) = N_M(0) \exp(-\xi/\xi_M). \quad (15.9)$$

Multiplying the second equation of (15.8) by an integrating factor $\exp(\xi/\xi_L)$ and integrating,

$$\frac{d}{d\xi} [\exp(\xi/\xi_L) N_L(\xi)] = \frac{P_{ML}}{\xi_M} \exp\left(\frac{\xi}{\xi_L} - \frac{\xi}{\xi_M}\right) N_M(0), \quad (15.10)$$

$$\frac{N_L(\xi)}{N_M(\xi)} = \frac{P_{ML} \xi_L}{(\xi_L - \xi_M)} \left[\exp\left(\frac{\xi}{\xi_M} - \frac{\xi}{\xi_L}\right) - 1 \right]. \quad (15.11)$$

In our simplified treatment, average values of ξ_L , ξ_M and P_{ML} are adopted for the L and M group of elements from Table 10.1a. For illustrative purposes, we adopt a weighted average total inelastic cross-section for the M elements of 280 millibarns and a weighted average fragmentation probability P_{ML} of 0.28. The average total inelastic cross-section for destruction of the L group of elements is taken to be 200 millibarns. Notice that, following our prescription of Sect. 5.1, the total inelastic cross-sections are roughly proportional to the geometric cross-sections of the nuclei. For collisions with hydrogen nuclei, the corresponding values of the mean free paths are $\xi_M = 60 \text{ kg m}^{-2}$ and $\xi_L = 84 \text{ kg m}^{-2}$.

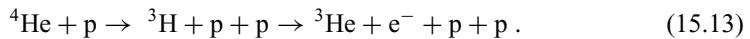
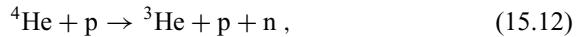
The data presented in Table 15.1 indicate that the observed ratio of the abundances of the L to the M elements in the cosmic rays is $N_L(\xi)/N_M(\xi) = 0.25$. Inserting these values into (15.11), the typical path length through which the M elements would have to pass in order to create the observed abundance ratio of the L to M elements is $\xi = 48 \text{ kg m}^{-2}$, of the same order of magnitude as the mean free path of the M elements which is hardly surprising. Furthermore, the abundance-weighted cross-sections for the production of lithium, beryllium and boron from carbon, nitrogen and oxygen can be compared with

Table 15.3 Abundance-weighted probabilities of creating lithium, beryllium and boron by the spallation of carbon, nitrogen and oxygen nuclei.

	Abundance-weighted probability of formation (millibarns)	Measured abundance relative to [Si] = 100
Lithium	24	136
Beryllium	16.4	67
Boron	35	233

the observed abundance ratios of these light elements (Table 15.3). The predicted relative probabilities of creating lithium, beryllium and boron are in roughly the same proportions as the observed abundances.

Exactly the same type of calculation can be performed for the production of ${}^3\text{He}$ by the spallation of ${}^4\text{He}$ in the interstellar gas. The principal interactions are:



The solutions of the transfer equations give a value for the path length ξ of about 50 kg m^{-2} , suggesting that the helium nuclei have traversed roughly the same amount of material as the L group of elements (Shapiro, 1991).

There is, however, a problem with the spallation of iron nuclei because their inelastic cross-section, $\sigma_{\text{Fe}} = 764$ millibarns, is more than twice those of the M nuclei and therefore the iron would be much more severely depleted by spallation. The ratio of spallation products to residual iron would be

$$\frac{[\text{products}]}{[\text{primaries}]} = \frac{1 - \exp(-\xi/\xi_{\text{Fe}})}{\exp(-\xi/\xi_{\text{Fe}})}. \quad (15.14)$$

Taking $\xi = 50 \text{ kg m}^{-2}$ and $\xi_{\text{Fe}} = 22 \text{ kg m}^{-2}$, the ratio of spallation products to residual iron primaries would be $[\text{products}]/[\text{primaries}] = 8.7$. Table 10.1 shows that many of the spallation products of iron are just lighter than the parent nucleus, about one-third of the total cross-section resulting in nuclei such as Mn, Cr, V. Therefore, if ξ were 50 kg m^{-2} , the abundances of these elements in total would be significantly greater than that of iron, contrary to the abundances shown in Fig. 15.1. Another way of demonstrating this discrepancy is to compare the ratio of abundances of the elements between chlorine and vanadium to that of iron, $[\text{Cl} \rightarrow \text{V}]/[\text{Fe}]$, since most of the products of the spallation of iron fall in this range of atomic numbers. The observed value is 1.5, significantly smaller than 8.7.

There is an obvious way of resolving this discrepancy. It has been assumed that all the high energy particles traverse the same amount of matter in reaching the Earth. In fact, there must be a distribution of path lengths. Removing the source, spallation and energy



Fig. 15.10

Models for the path length distribution for the propagation of high energy particles in the interstellar gas: (a) a 'slab' model in which all high energy particles traverse the same path length of interstellar gas; (b) a model in which high energy particles arriving at the Earth traverse different path lengths of interstellar gas so that a wider range of spallation products can be explained; (c) a Gaussian distribution of path lengths expected in diffusion models and an exponential distribution with a low energy cut-off.

loss terms from the diffusion-loss equation (15.6),

$$\frac{\partial N}{\partial t} = D \nabla^2 N - \frac{N}{\tau_e}, \quad (15.15)$$

where τ_e is the characteristic escape time from the system. If the particles diffused from their sources to the observer, $\tau_e = \infty$, and we would expect a *Gaussian distribution* of path lengths between the source and the observer. If the particles remained within the confinement volume for a characteristic escape time τ_e , $D = 0$, and *exponential distribution* of path lengths is expected since, if

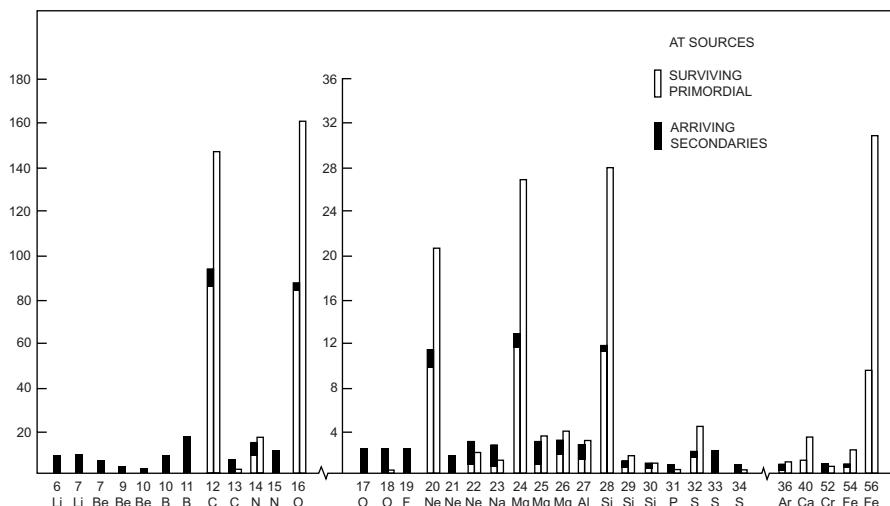
$$\frac{\partial N}{\partial t} + \frac{N}{\tau_e} = 0, \text{ then } N \propto \exp(-t/\tau_e) \text{ or } N \propto \exp(-\xi/\xi_e). \quad (15.16)$$

These models correspond to different ways in which cosmic ray particles diffuse and escape from the Galaxy.

A simple example shows how a distribution of path lengths helps solve the problem. Suppose one-third of the matter traverses a path length of 100 kg m^{-2} and two-thirds traverses a negligible amount. The one-third which traverses 100 kg m^{-2} suffers a large amount of spallation. From (15.11), $[L]/[M] = 0.6$ while essentially all the iron is converted into iron products. The remaining two-thirds survives unaltered. Adding together the products of these two paths,

$$[L]/[M] = 0.25: \quad [\text{products of Fe}]/[\text{Fe}] = 0.5, \quad (15.17)$$

in rough agreement with observation. Models for the path length distribution are shown in Fig. 15.10, including the slab model in *a*, the two-component model discussed above in *b* and a Gaussian and a truncated exponential path length distributions in *c*. The truncated exponential path length distribution has been found to give satisfactory agreement with the observed cosmic ray abundances when the full set of transfer equations is solved. This exponential path length model is often referred to as a *leaky box model* in which the high energy particles diffuse freely inside the confinement volume and are reflected at its boundaries. There is a certain probability that the particles escape from this volume on each encounter with the boundary and this results in an exponential path length distribution

**Fig. 15.11**

The relative isotopic abundances of the cosmic rays as observed near the Earth and as inferred to be present in their sources, once the effects of spallation between the sources and the Earth are taken into account. The abundances have been normalised to 100 for $^{12}\text{C} + ^{13}\text{C}$. The hatched histograms show the inferred source abundances and the neighbouring histograms the observed abundances, the open parts showing the surviving primary elements and the black parts the amount produced by spallation (Shapiro, 1991; Wefel, 1991).

$N \propto e^{-t/\tau_e}$, where τ_e is the characteristic escape time from the confinement volume as shown above.

The results of detailed computations are discussed by Simpson (1983) and Wefel (1991), Fig. 15.11 displaying typical results of these calculations (Shapiro, 1991). In the figure, the inferred source isotopic abundances are indicated by the shaded histogram and next to these are the abundances observed in the vicinity of the Earth, showing separately the amount of each isotope which is primordial (open histogram) and how much is produced by the spallation of heavier elements (black histogram). The source abundances and the cosmic abundances of the elements are compared in Table 15.4 (Wefel, 1991). In addition to all the isotopes of the light elements and the elements just lighter than iron, species such as ^{15}N , ^{17}O , ^{18}O , ^{19}F and ^{21}Ne are also secondary species. On the other hand, substantial fractions of the common elements, carbon, oxygen, neon, magnesium and silicon, have survived unaffected by spallation interactions between their sources and the Earth.

Inspection of the results in Table 15.4 shows that there are significant differences between the source abundances of the cosmic rays and the local Galactic abundances. Elements up to about neon are underabundant as compared with the local Galactic abundances. If the species are ordered by their first ionisation potentials, there is a systematic trend in the underabundances (Fig. 15.12). Elements with first ionisation potentials greater than about 10 eV are underabundant by about a factor of 5 relative to the cosmic abundances. Hydrogen and helium show even greater discrepancies.

The same pattern of underabundances is found in the solar cosmic rays. This correlation is interpreted as providing information about the ionisation state of the region in which

Table 15.4 Cosmic ray source abundances compared with the local Galactic abundances, both normalised to Si] = 100 (Wefel, 1991).

Element	Cosmic ray source abundance	Local Galactic abundance
(1990 update)		
H	$8.9 \pm 2.2 \times 10^4$	$2.7 \pm 0.3 \times 10^6$
He	2.4×10^4	$2.6 \pm 0.7 \times 10^5$
C	431 ± 34	1260 ± 330
N	19 ± 9	225 ± 90
O	511 ± 20	2250 ± 560
F	<2.5	0.09 ± 0.06
Ne	64 ± 8	325 ± 160
Na	6 ± 4	5.5 ± 1.0
Mg	106 ± 6	105 ± 3
Al	10 ± 4	8.4 ± 0.4
Si	100	100
P	<2.5	0.9 ± 0.2
S	12.6 ± 2.0	43 ± 15
Cl	<1.6	0.5 ± 0.3
Ar	1.8 ± 0.6	11 ± 5
K	<1.9	0.3 ± 0.1
Ca	5.1 ± 0.9	6.2 ± 0.9
Sc	<0.8	$3.5 \pm 0.5 \times 10^{-3}$
Ti	<2.4	0.27 ± 0.04
V	<1.1	0.026 ± 0.005
Cr	2.2 ± 0.6	1.3 ± 0.1
Mn	1.7 ± 1.7	0.8 ± 0.2
Fe	93 ± 6	88 ± 6
Co	0.32 ± 0.12	0.21 ± 0.03
Ni	5.1 ± 0.5	4.8 ± 0.6
Cu	0.06 ± 0.01	0.06 ± 0.03
Zn	0.07 ± 0.01	0.10 ± 0.02
Ga	$5.6 \pm 2.8 \times 10^{-3}$	$\sim 3.7 \times 10^{-3}$
Ge	$7.4 \pm 1.0 \times 10^{-3}$	$\sim 11.4 \times 10^{-3}$

the particles were accelerated. Those elements with ionisation potentials less than about 10 eV would be ionised and the local Galactic composition is preserved after acceleration. For elements with higher ionisation potentials, there must be some suppression mechanism which reduces the element abundances by about a factor of 5. Some other mechanism is required to account for the even lower abundances of hydrogen and helium.

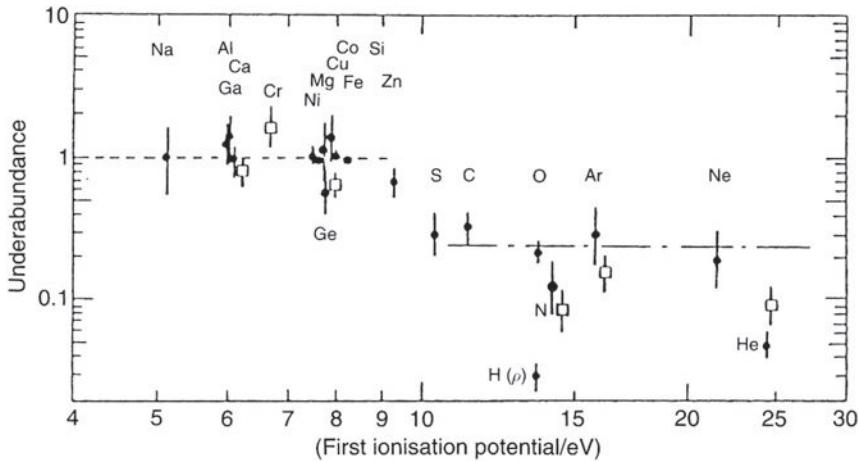


Fig. 15.12 The ratio of the cosmic ray source abundances to the local Galactic abundances as a function of the first ionisation potential of the elements. Open squares show recent improved estimates of some of the abundance ratios (Wefel, 1991).

15.5.2 Variations in the chemical composition of cosmic rays with energy

As discussed in Sect. 15.1, the relative abundances of the cosmic rays change with energy, Fig. 15.13 displaying the variation of the boron-to-carbon [B/C] and the chromium-to-iron [Cr/Fe] ratios with energy per nucleon. Some of these variations are associated with differences in primary injection spectra and others with variations of the secondary to primary production ratios as a function of energy.

The simplest interpretation of the energy dependence of the secondary to primary ratio is that the path length of the primary particles through the interstellar gas changes with energy. Suppose the energy dependence of the path length for escape from the Galaxy is of the form $\xi_e(E) = \xi_0(E/E_0)^{-\alpha}$, so that, if α is positive, the path length decreases at high energies. We therefore include a loss term of the form $-N_L/\xi_e(E)$ on the right-hand side of (15.8) to represent the escape of particles from the system. Let us study the steady state solution $dN_L/dt = 0$, meaning the abundances reach a steady value under the combination of spallation losses, escape from the system and injection from sources,

$$-\frac{N_L}{\xi_e(E)} + \frac{P_{ML}}{\xi_M} N_M(\xi) - \frac{N_L}{\xi_L} = 0 . \quad (15.18)$$

Rearranging this expression,

$$N_L = \left(\frac{P_{ML} N_M(\xi)}{\xi_M} \right) \Bigg/ \left(\frac{1}{\xi_e(E)} + \frac{1}{\xi_L} \right) . \quad (15.19)$$

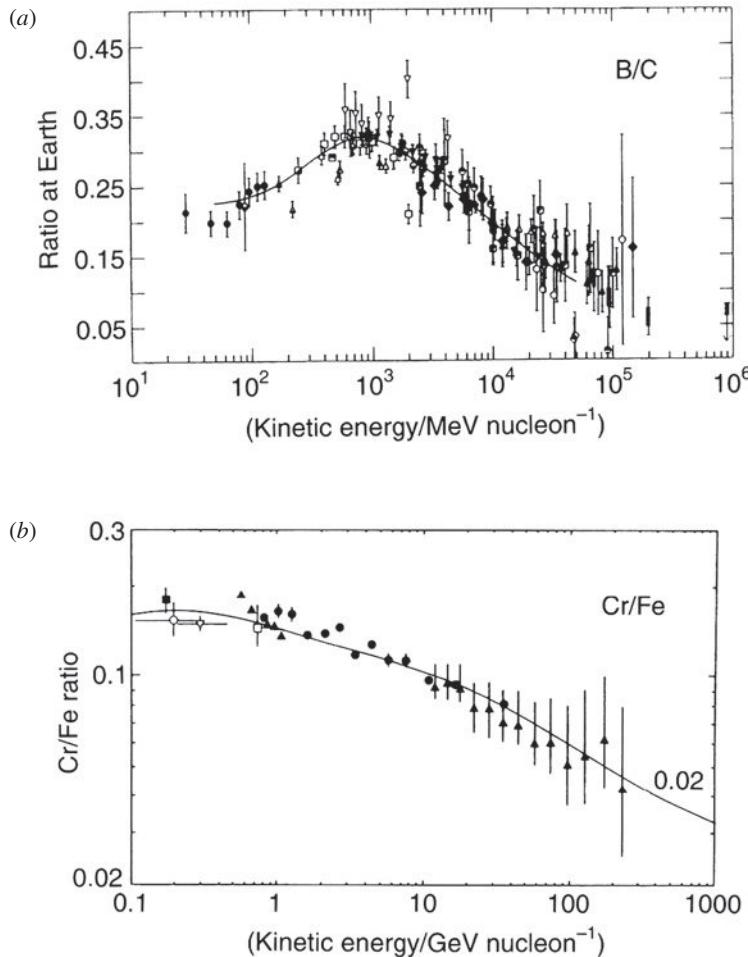


Fig. 15.13

(a) The boron-to-carbon ratio as a function of energy. The curve shows the results of propagation calculations by Garcia-Munoz and his colleagues (1987). (b) The chromium-to-iron ratios as a function of energy. The curve shows the results of full propagation calculations by Mewaldt and Webber (1990) which assumed that the source Cr/Fe ratio was 0.02 (Wefel, 1991).

In the high energy limit, the escape path length is much less than the spallation path length, $\xi_E \ll \xi_L$ and so (15.19) reduces to

$$\frac{N_L(\xi)}{N_M(\xi)} = P_{ML} \frac{\xi_e(E)}{\xi_M} . \quad (15.20)$$

Since P_{ML} and ξ_M are independent of energy, the energy dependence of the ratio of secondary to primary particles is directly related to the energy dependence of the escape path length $\xi_e(E)$. From the analysis of Garcia-Munoz and his colleagues (1987), a dependence of the form $\xi_e(E) \propto E^{-0.6}$ can provide a good fit to observations of the boron-to-carbon ratio at energies $E \geq 1$ GeV nucleon⁻¹. At lower energies, the energy dependence is in the opposite sense, as can be appreciated from Fig. 15.13a.

In the case of the chromium-to-iron ratio shown in Fig. 15.13b, Fig. 15.11 shows that only part of the chromium is created by spallation and some of it is primordial. Therefore, because the escape path length is assumed to decrease with increasing energy, there is only a small contribution from spallation at a high enough energy, and the primordial chromium abundance should be observed. This is illustrated by the model calculations of Mewaldt and Webber (1990) shown in Fig. 15.13b.

In addition to the variation of the secondary to primary ratios of the cosmic ray abundances with energy, there are differences in the energy spectra of primary nuclei which are assessed by Wefel (1991). The spectral index of high energy protons is $x = 2.73 \pm 0.09$ over the energy interval $100\text{--}10^5$ MeV while in the high energy range, $E > 10^4$ MeV, the spectral index of helium nuclei is slightly steeper, $x = 2.87 \pm 0.13$. On the other hand, species such as carbon, oxygen and iron have spectral indices closer to $x = 2.5$. Presumably, these spectral indices reflect differences in the processes of acceleration of the high energy particles.

A clue to the sites of acceleration is provided by isotopic anomalies of individual cosmic ray species, as noted in Sect. 15.2.3. The largest anomaly is associated with the ratio of the isotopes of neon, the isotopic ratio [$^{22}\text{Ne}/^{20}\text{Ne}$] being about four times its cosmic abundance. Smaller but significant isotopic excesses are also found for the abundance ratios [$^{25,26}\text{Mg}/^{24}\text{Mg}$] and [$^{29,30}\text{Si}/^{28}\text{Si}$]. These anomalies suggest that at least some of the cosmic rays have originated in neutron-rich environments. Similar isotope anomalies of neon are found in the isotopic abundances of the *Wolf-Rayet stars*. These massive stars undergo substantial amounts of mass-loss resulting in abnormal chemical abundances in their atmospheres.

15.6 The confinement time of cosmic rays in the Galaxy and cosmic ray clocks

According to the calculations of Sect. 15.5, high energy particles traverse typically about 50 kg m^{-2} of matter on their route to the Solar System. This path length can be converted into an escape time τ_e of the cosmic rays from our vicinity in the Galaxy if a value for the mean density $\bar{\rho}$ of matter through which they travelled is assumed $\xi = \bar{\rho}v\tau$. If the number density of the interstellar gas is $N = 10^5\text{--}10^6 \text{ m}^{-3}$ and assuming the particles propagate at the speed of light, the particles remain within our locality for $\tau_e \approx 3 \times 10^6 - 3 \times 10^7$ years.

Some of the species created in the spallation reactions are radioactive and hence their abundances can be used to ‘date’ the samples of cosmic rays observed near the Earth. Examples of such radioactive spallation products are given in Table 15.5. The most famous of these is the radioactive isotope of beryllium ^{10}Be which has a half-life similar to the escape time τ_e found above. If the particles are relativistic, the lifetime measured by the observer is $\gamma\tau_r$ where γ is the Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$. ^{10}Be is produced in significant quantities in the spallation of carbon and oxygen, the fraction of the total spallation cross-section for the production of ^{10}Be being about 10% of the total production

Table 15.5 Decay half-lives of some important radioactive isotopes created in spallation reactions (Yanasak *et al.*, 2001).

Isotope	Decay mode	Half-life of decay mode
¹⁰ Be	β^-	1.51×10^6 years
¹⁴ C	β^-	5700 years
²⁶ Al	β^+	8.73×10^5 years
	e capture	8.45×10^6 years
³⁶ Cl	β^-	3.07×10^5 years
	e capture	1.59×10^7 years
⁵⁴ Mn	β^-	$(6.3 \pm 1.7) \times 10^5$ years
	e capture	312 days

cross-section of beryllium (see Table 10.1a). The ¹⁰Be nuclei undergo β^- decays into ¹⁰B. Therefore, the relative abundances of the isotopes of Be and B provide a measure of whether or not all the ¹⁰Be has decayed and consequently an estimate of the mean age of the cosmic rays observed in our vicinity.

Another term has to be introduced into the diffusion-loss equation (15.6) to describe the radioactive decay of species *i*. If the e-folding time is τ_r , the additional term has the form $-N_i(\xi)/\tau_r$. If the typical time it takes the cosmic rays to reach the Earth from their sources is less than τ_r , the ratio of [¹⁰Be]/[⁷Be + ⁹Be + ¹⁰Be] should correspond to the relative production rates of these species, about 10%. If the escape time τ_e is much longer than τ_r , this ratio should be very much less than 10%. The isotopic abundance ratio should be of order τ_r/τ_e of the production ratio if $\tau_e \gg \tau_r$.

Let us give a simple derivation of the expected abundance of ¹⁰Be in a model in which there is a characteristic escape time τ_e of the cosmic rays from the Galaxy. It is assumed that the system has attained a steady state so that $\partial N_i/\partial t = 0$ and that all the beryllium isotopes are produced by spallation of the M elements. The production rate of the isotope *i* is

$$C_i = \sum_{j>i} \frac{P_{ij}}{\tau_j} N_j . \quad (15.21)$$

With these simplifications, the steady state abundance of a non-radioactive isotope is derived from the simplified transfer equation,

$$-\frac{N_i}{\tau_e(i)} + C_i - \frac{N_i}{\tau_{\text{spal}}(i)} = 0; \quad N_i = \frac{C_i}{\frac{1}{\tau_e(i)} + \frac{1}{\tau_{\text{spal}}(i)}} , \quad (15.22)$$

where $\tau_{\text{spal}}(i)$ is the time-scale over which the isotope *i* is destroyed by inelastic collisions.

If the isotope *j* is radioactive, another loss term is added to the transfer equation. If $\tau_r(j)$ is the characteristic decay time of the radioactive nucleus, then

$$-\frac{N_j}{\tau_e(j)} + C_j - \frac{N_j}{\tau_{\text{spal}}(j)} - \frac{N_j}{\tau_r(j)} = 0; \quad N_j = \frac{C_j}{\frac{1}{\tau_e(j)} + \frac{1}{\tau_r(j)} + \frac{1}{\tau_{\text{spal}}(j)}} . \quad (15.23)$$

The steady-state ratio of the ^{10}Be and ^7Be isotopes is therefore

$$\frac{N(^{10}\text{Be})}{N(^7\text{Be})} = \frac{\frac{1}{\tau_e(^7\text{Be})} + \frac{1}{\tau_{\text{spal}}(^7\text{Be})}}{\frac{1}{\tau_e(^{10}\text{Be})} + \frac{1}{\tau_r(^{10}\text{Be})} + \frac{1}{\tau_{\text{spal}}(^{10}\text{Be})}} \frac{C(^{10}\text{Be})}{C(^7\text{Be})}. \quad (15.24)$$

If the time-scale for the destruction of the beryllium isotopes by spallation is much greater than their escape times $\tau_{\text{spal}} \gg \tau_e$, we obtain a simpler expression,

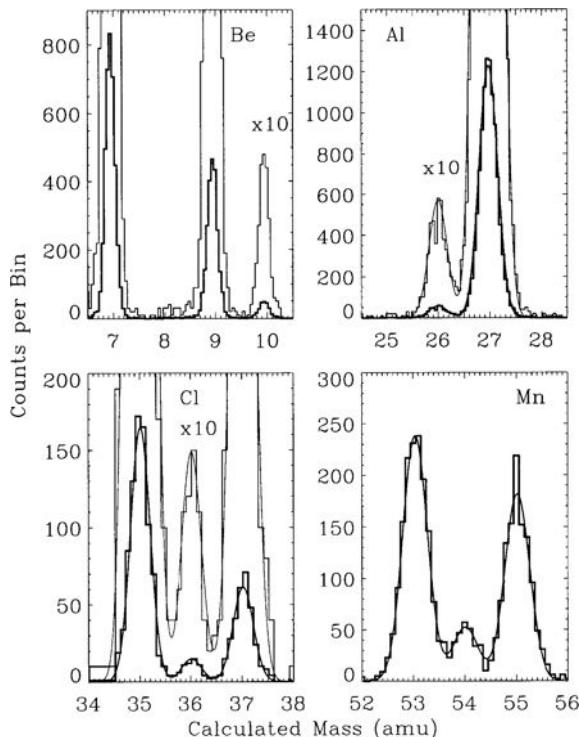
$$\frac{N(^{10}\text{Be})}{N(^7\text{Be})} = \frac{\frac{1}{\tau_e(^7\text{Be})}}{\frac{1}{\tau_e(^{10}\text{Be})} + \frac{1}{\tau_r(^{10}\text{Be})}} \frac{C(^{10}\text{Be})}{C(^7\text{Be})}. \quad (15.25)$$

The experiments carried out by the Chicago group from the IMP-7 and IMP-8 satellites resolved the isotopes of beryllium, the ratio $[^{10}\text{Be}] / [^7\text{Be} + ^9\text{Be} + ^{10}\text{Be}] = 0.028$ being found (Simpson, 1983). Inserting this value into (15.25), an escape time of 10^7 years is obtained. Combining this result with the spallation path length through the interstellar gas of 50 kg m^{-2} , the average interstellar density through which the particles have travelled is about $3 \times 10^5 \text{ m}^{-3}$.

Observations of the radioactive isotopes of beryllium, aluminium, chlorine and manganese with energies $70 \leq E \leq 400 \text{ MeV nucleon}^{-1}$ by the Cosmic-Ray Isotope Spectrometer (CRIS) of the Advanced Composition Explorer (ACE) are shown in Fig. 15.14 (Yanasak *et al.*, 2001). These observations resolved very clearly the secondary isotopes of these elements with excellent statistics. From Table 10.1, the production ratio of ^9Be to ^{10}Be should be about 2:1, whereas it is apparent that the observed value in the cosmic radiation is about 10:1, indicating that there has been time for most of the ^{10}Be to have decayed. Yanasak and his colleagues carried out very much more detailed solutions of the full transfer equations, including a leaky-box path length distribution, ionisation losses and rates of attachment and stripping of single electrons. The continuous lines in Fig. 15.14 show the expectations of their best-fit model which can account simultaneously for all the observed abundances of the radioactive isotopes. They found a unique confinement time of τ_e of $1.5 \pm 0.16 \times 10^7$ years, consistent with an average interstellar hydrogen number density of $n_H = (0.34 \pm 0.04) \times 10^6 \text{ m}^{-3}$.

15.7 The confinement volume for cosmic rays

These data show that the cosmic ray particles do not escape unimpeded from our Galaxy. If the confinement volume of the cosmic rays in the Galaxy had dimension, say, 1–10 kpc, then, if they propagated freely at a velocity close to that of light, they would escape from the Galaxy in about $3 \times 10^3 - 3 \times 10^4$ years, very much less than the confinement times deduced in Sect. 15.6. Furthermore, their distribution on the sky would be highly anisotropic, most of the flux coming from the general direction of the central regions of

**Fig. 15.14**

Mass histograms for radioactive isotopes of beryllium, aluminium, chlorine and manganese in the cosmic radiation as measured by the Cosmic-Ray Isotope Spectrometer (CRIS) of the Advanced Composition Explorer (ACE) (Yanasak *et al.*, 2001). The light histograms and fitted curves are magnified 10 times relative to the heavy histograms and lines, showing more clearly the abundances of the rarer radioactive isotopes.

the Galaxy, in conflict with the very high degree of isotropy of cosmic rays with energies $10^{11} - 10^{14}$ eV (Fig. 15.7). Therefore, the high energy particles must either diffuse from their sources, as described by a Gaussian distribution of path lengths, or be ‘confined’ within some volume, as described by the leaky box model, before escaping into intergalactic space. These path-length distributions are illustrated in Fig. 15.10c.

Let us assume that an isotropic diffusion model can be used to describe the dynamics of particles within the disc of the Galaxy. Then, the particles should diffuse a distance roughly the half-thickness of the Galactic disc, about 300 pc, in about 10^7 years. If τ_e is the typical escape time, the diffusion coefficient D can be roughly estimated by simplifying the diffusion equation $D \nabla^2 N - N/\tau_e = 0$ by writing

$$\frac{DN}{L^2} \sim \frac{N}{\tau_e} : \quad D \approx L^2/\tau_e = 3 \times 10^{23} \text{ m}^2 \text{ s}^{-1}. \quad (15.26)$$

The diffusion coefficient D is related to the mean free path of the particles by $D = \frac{1}{3} \lambda v$ and hence $\lambda \sim 0.1$ pc. This mean free path can be interpreted as the typical scale of inhomogeneities in the interstellar medium and, indeed, there is abundant evidence that

there exist irregularities on this scale, associated with supernova shells, regions of ionised hydrogen, and so on.

We can now work out how anisotropic the flux of cosmic rays would be by estimating the net streaming velocity V of the high energy particles because of diffusion:

$$D \frac{dN}{dx} = VN; \quad V \approx \frac{D}{L} = 10^{-4}c. \quad (15.27)$$

Therefore, if we were located at the edge of the Galactic disc, we would expect to observe a net streaming velocity of about $10^{-4}c$ which would be reflected in an anisotropy of the cosmic ray flux. The relation between the measured cosmic ray anisotropy $\delta = (I_{\max} - I_{\min})/(I_{\max} + I_{\min})$ and the streaming velocity of the particles is $V = [\delta/(x + 2)]c$ where x is the spectral index of the differential energy spectrum of the cosmic rays. This is known as the *Compton–Getting effect* and is similar to the anisotropies which are discussed in the context of relativistically moving sources in Sect. 23.2. Taking the observed anisotropy to be $\delta = 6 \times 10^{-4}$ and $x = 2.5$, $V \approx 10^{-4}c$, consistent with the above analysis.

The scattering of high energy particles by irregularities in the magnetic field was discussed in Sect. 7.3 in the context of the diffusion of cosmic rays from interstellar space to the vicinity of the Earth through the outflowing Solar Wind. Also, in Sect. 7.4, the scattering of high energy particles by Alfvén and hydromagnetic waves was shown to limit their bulk streaming velocity along the magnetic field lines to the Alfvén speed, $v_A = B_0/(\mu_0\rho)^{1/2}$. Taking the density of the ionised component of the interstellar gas to be $N = 10^5 \text{ m}^{-3}$ and the magnetic field strength $B_0 = 3 \times 10^{-10} \text{ T}$, $v_A = 2 \times 10^4 \text{ m s}^{-1} \sim 10^{-4}c$. As discussed in Sect. 7.4, however, the instability is only effective if the waves are not damped. The presence of neutral particles in the interstellar plasma damp the waves by neutral–ion collisions, the damping rate of the waves given by Kulsrud and Pearce (1969) for temperatures of 10^3 and 10^4 K being $\Gamma^* = \Gamma_0 n_H = (3.3 \text{ and } 8.4) \times 10^{-9} N_H \text{ s}^{-1}$ where N_H is the number density of neutral hydrogen atoms.

At low cosmic ray energies, $E \sim 3 \text{ GeV}$, the high energy particles are sufficiently numerous that the growth rate of the instability exceeds the damping rate and so the streaming velocities are restricted to the Alfvén velocity in the neutral intercloud medium. The typical time-scale for the growth of the instability is, however, a strong function of energy since it depends upon the number of high energy particles that can resonate with the wave, $N(\geq E)$. The time-scale $\tau = \Gamma^{-1}$ for the growth of the instability varies as $E^{1.7}$ and so the high energy particles are not scattered as strongly as those of low energy. The streaming instabilities automatically lead to an escape time from the Galaxy which is energy dependent, as is inferred must be the case from the energy dependence of the spallation products of common nuclei (Sect. 15.5.2). Skilling (1971) showed that, according to this model, the mean free path of the particles should be proportional to $E^{-0.4}$, in reasonable agreement with the observations. According to Cesarsky (1980), particles with energies $E > 100 \text{ GeV}$ should not be confined to the Galaxy.

The net result is that, in regions where there are high densities of neutral material, the particles can diffuse rapidly, but as soon as they encounter fully ionised plasmas, their streaming velocity drops to the Alfvén speed. This provides a physical realisation for the leaky box model for the propagation of high energy particles in the Galactic disc. Within

the Galactic disc, the plasma consists of a mixture of ionised and neutral material whilst outside the disc the gas is ionised. The particles can stream freely inside the disc but outside it they are limited to the Alfvén velocity. By continuity at the interface, the rate of loss of particles from the disc of the Galaxy must be $N_{\text{int}}c = N_{\text{ext}}v_A$. Thus, if the net outward streaming velocity from the disc is only v_A , the typical residence time inside the disc is $(L/c) \times (c/v_A) \approx 10^7$ years, similar to the figures derived above from the abundances of the radioactive isotope clock. Many of these topics concerning the confinement volume of cosmic rays are discussed in the excellent review by Cesarsky (1980).

15.8 The Galactic halo

The diffusion of particles from the Galaxy would be expected to lead to a halo of high energy particles about the Galaxy. For cosmic ray electrons, Webster (1974) showed that the distribution of radio emission at low radio frequencies can be interpreted in terms of a somewhat flattened radio emitting ellipsoid with semi-major axis 10 kpc and semi-minor axis about 3–4 kpc perpendicular to the plane. According to Beck and his colleagues, the radio distribution of most edge-on spiral galaxies have scale heights of only about 1 kpc perpendicular to their discs but there are some exceptions (Beck *et al.*, 1996). A good example is the edge-on barred spiral galaxy NGC 253 which possesses a radio halo which extends to at least 9 kpc above the galactic disc (Fig. 15.15). A halo of similar dimensions is observed in the X-ray waveband.

A second piece of evidence is provided by ultraviolet observations of blue halo stars and B stars in the Magellanic Clouds. Highly ionised lines of CIV and SiIV are observed as absorption troughs with velocity widths up to about 150 km s^{-1} . The gas responsible for these broad absorption troughs is associated with our own Galaxy, the lines of sight passing through the halo of our Galaxy. Savage and de Boer (1979) interpreted these lines as originating in a hot gaseous halo about our Galaxy with scale height about 5 kpc perpendicular to the Galactic plane. In Jenkins' survey of the use of these lines to probe the Galactic halo, the data are consistent with a somewhat flattened halo, similar in shape to that inferred by Webster for the radio halo (Jenkins, 1987). Such halo gas must be hot with temperature $T \sim 10^6$ because, otherwise, it would be confined to the plane of the Galaxy. The temperatures and scale heights of the gas estimated from these observations are consistent with it being in hydrostatic equilibrium in the Galactic gravitational potential. Such a hot halo could be maintained by *galactic fountains* associated with the cavities of very hot gas in the plane of the Galaxy created by supernova explosions (Shapiro and Field, 1976). These bubbles of very hot gas rise out of the Galactic plane under the influence of buoyancy and cool as they expand into the upper reaches of the halo. In their original model, Shapiro and Field associated the collapse of the cooled clouds back onto the disc of the galaxy with the high velocity clouds of neutral hydrogen observed to be approaching the Galactic plane.

Another reason to suppose that there should be a halo of high energy particles is associated with Parker's classical analysis of the stability of the gaseous disc of the Galaxy if it is

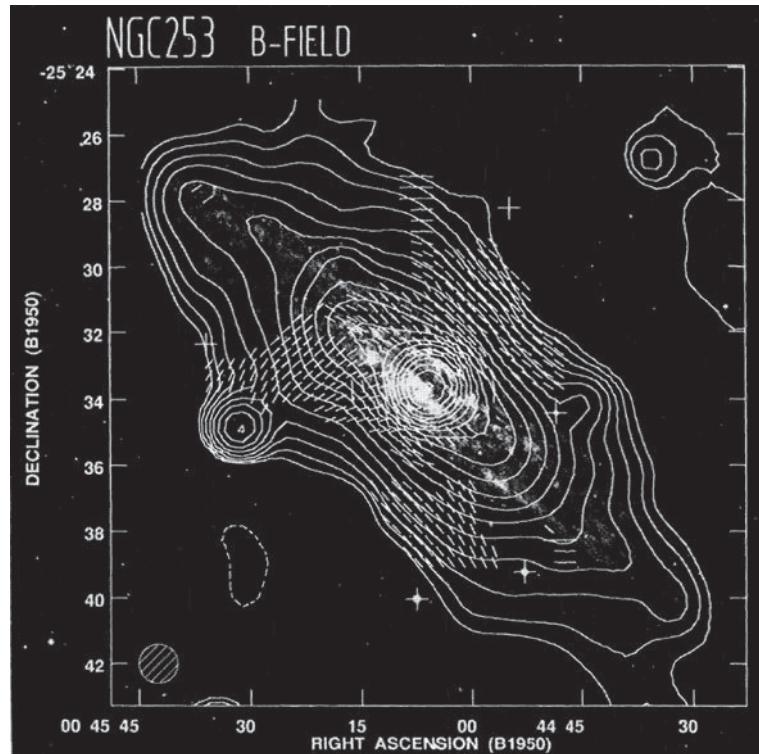


Fig. 15.15 A radio map of the galaxy NGC253 at 1.4 GHz made with the VLA. The magnetic field directions, shown by the short lines, were inferred from polarisation observations of the radio emission at 1.4, 5 and 10.4 GHz made with the VLA and 100-metre Effelsberg radio telescope (Beck *et al.*, 1994).

threaded by a magnetic field (Parker, 1965). He showed that the gaseous disc of our Galaxy is unstable to perturbations of the magnetic field lines which lie predominantly parallel to the Galactic plane. As discussed in Sect. 15.3, there is observed to be rough equality of the energy densities in high energy particles and in the Galactic magnetic field. The energy density in the Galactic magnetic field is $B^2/2\mu_0$ which, for $B = 3 \times 10^{-10}$ T, is 2×10^5 eV m $^{-3}$. The best estimate for the energy density of cosmic rays in the local interstellar medium was about 10 6 eV m $^{-3}$. It is hardly a coincidence that these values are of the same order of magnitude because of the close coupling between the relativistic gas and the interstellar magnetic field.

In Parker's analysis, the unperturbed state consists of a magnetic field running parallel to the plane of the disc embedded in partially ionised cold gas and the relativistic gas (Fig. 15.16). As shown in Sect. 7.2, the high energy particles conserve their adiabatic invariants in a varying magnetic field and so are tied to the magnetic field lines, just as in the case of magnetic flux freezing (Sect. 11.2). The neutral component is also coupled to the ionised gas by particle collisions and, thus, the magnetic field is also coupled to the cold gas which contains most of the inertia and which consequently 'holds down' the magnetic field.

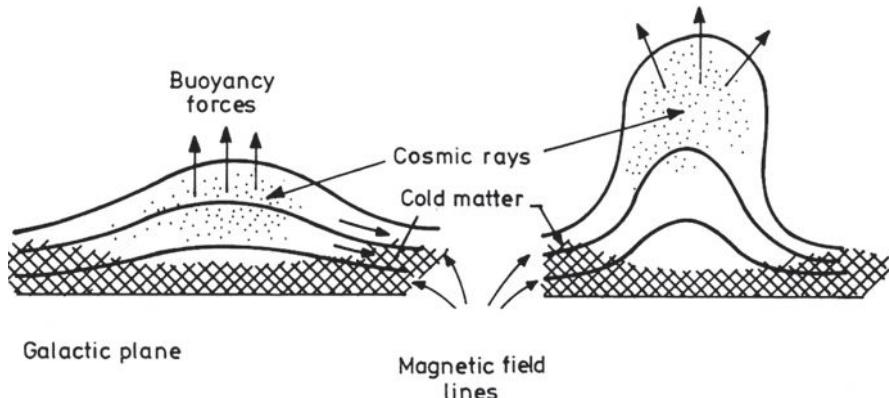


Fig. 15.16

Illustrating the formation of magnetic loops and expulsion of relativistic gas from the Galactic disc due to the instability described by Parker. These are often referred to as Parker's loops.

Suppose there is a kink in the magnetic field lines which causes them to bulge out of the plane. The cold matter, having a low temperature, falls down the potential gradient to the minimum value, whilst the relativistic gas ‘inflates’ the kink. The relativistic gas in the kink forms a ‘bubble’ of relativistic plasma which rises buoyantly up the potential gradient and out of the Galactic plane. The natural tendency of a mixture of relativistic gas and magnetic field is to expand if unrestrained. Parker suggested that this mechanism is responsible for the formation of Galactic loops and for the formation of a halo of cosmic rays about the plane of the Galaxy. The instability suggests a physical reason for the local equality of the cosmic ray and the magnetic field energy densities. If $\varepsilon_{\text{CR}} \ll \varepsilon_B$, the pressure is dominated by the magnetic field which remains securely tied to the cold matter. If, however, the opposite inequality is true, $\varepsilon_B \ll \varepsilon_{\text{CR}}$, the local pressure is dominated by the relativistic gas which forms a bubble and the instability begins to grow, expelling the local excess pressure into the halo. Therefore, we expect that, in general, the high energy particle and magnetic field energy densities should be comparable. The dominant scale length for the instability is probably of the order of the scale height of the disc, about 200–300 pc.

15.9 The highest energy cosmic rays and extensive air-showers

In the 1930s Auger and his colleagues showed that coincidences of cosmic ray events were observed between particle detectors separated by up to 300 m on the ground (Auger *et al.*, 1939). They correctly inferred that these coincidences were associated with nucleonic cascades triggered by individual very high energy cosmic rays at the top of the atmosphere. The showers were the secondary, tertiary and high order generations of particles and were named *extensive air-showers*. The fact that 10^6 or more relativistic particles were detected at ground level meant that the initiating particles had very high energies. A great deal of experimental and theoretical work was carried out during the 1960s and 1970s to understand

these showers and a number of very large cosmic ray air-shower arrays were constructed specifically to detect the most energetic of these particles.

The basic processes involved in the development of extensive air-showers are the nucleonic cascades discussed in Sect. 10.4.1 and the electron–positron, or electromagnetic, cascades described in Sect. 9.9. In both cases, the initial growth of the cascade is exponential until the typical energy per particle is degraded to about 1 GeV in the case of pion production. Thus, the greater the energy of the primary particle, the further the extensive air-shower penetrates through the atmosphere before development of the air-shower ceases. The depth through the atmosphere at which the maximum development of the shower takes place is known as X_{\max} and is a key quantity in measuring the properties of the particles.

The products of the nucleonic and electromagnetic showers pass through the atmosphere at a speed very close to that of light and so they arrive in ‘sheets’ or ‘pancakes’ at the surface of the Earth. The high energy cosmic rays enter the atmosphere at different angles to the vertical and so their arrival directions can be determined by very accurate timing of the shower as the particles pass through different elements of the air-shower array. The path length of a cosmic ray through the atmosphere at angle θ to the vertical is referred to as the *slant path length*. In large experiments such as the Akeno Giant Air Shower Array (AGASA) in Japan, there were 111 scintillator detectors and 27 muon detectors regularly spaced over an area roughly $14 \times 8 \text{ km}^2$. Hence, very large air showers could be detected and the energy of the primary particles estimated from the total measured signal. The most energetic particles with $E \sim 10^{20} \text{ eV}$ reach maximum development at roughly the vertical path length through the atmosphere, that is, about $10\,000 \text{ kg m}^{-2}$.

An important method of tracking the development of the showers makes use of the fact that the charged particles excite nitrogen molecules in the atmosphere resulting in the isotropic emission of fluorescent radiation. On average about 4 fluorescent photons m^{-1} per charged particle are emitted at wavelengths between about 300 nm to 400 nm. The intensity of the fluorescent radiation is proportional to the flux of charged particles and so, by measuring the flux of fluorescent radiation, the development of the shower through the atmosphere can be determined. Figure 15.17 shows examples of the type of data which can be accumulated by this means.

Pioneering observations using this technique were carried out using *fly’s eye telescopes* in which a large number of small optical elements observe a large solid angle of the sky so that the flux of fluorescent photons along the path of the incoming ultra-high energy cosmic ray can be measured directly. If more than one fly’s eye telescope is used separated by a several kilometres, the geometry of the shower can be determined. This requires the fly’s eye telescopes to be located in regions of clear dark sky and these are found at the Dugway Proving Grounds in Utah, USA which hosted the HiRes I and II system and the site of the Auger project on the Pampa Amarilla in western Argentina.

It is apparent from Fig. 15.1 that the spectrum of high energy cosmic rays extends to about 10^{20} eV which vastly exceeds the energies which can be attained in terrestrial particle physics laboratories. The investment in major facilities such as the HiRes and Auger projects is justified by the physical and astrophysical problems of understanding the origin of these particles and their significance for particle physics, astrophysics and cosmology. Some impression of the scale of these projects is given by the map of the

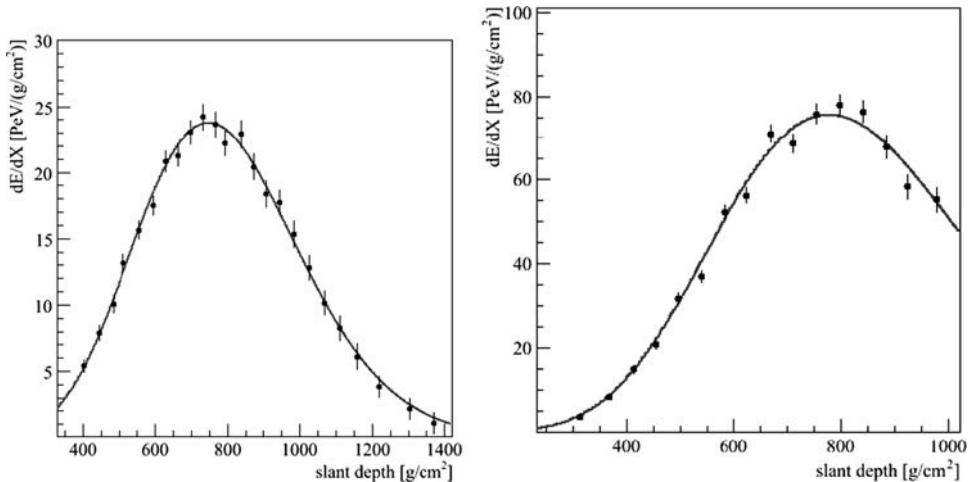


Fig. 15.17

Examples of the development of two ultra-high energy cosmic ray showers from measurements of their nitrogen fluorescent emission. The units on the ordinate are the numbers of charged particles. The inferred number density of charged particles as a function of slant path length has been fitted by a Gaisser–Hillas function (15.29) (Mattiae, 2010).

Auger array in Fig. 15.18. The surface array consists of 1600 water Cherenkov detector stations separated by 1.5 km, covering a total area of 3000 km^2 . There are four fluorescence telescope enclosures, each containing six telescopes giving 24 telescopes in total. The 24 telescopes are oriented to span the total ‘active volume’ of the atmosphere over the array, as indicated by the six lines from each of the telescope enclosures shown in Fig. 15.18.

15.10 Observations of the highest energy cosmic rays

A compilation of a large number of experiments to measure the spectrum of cosmic rays with energies $E \geq 10^{14}$ eV is shown in Fig. 15.19 (Amsler *et al.*, 2008). The results are presented in the form of a differential energy distribution $j(E)$, multiplied by the energy of the particle E raised to the power 2.7, that is, $j(E) E^{2.7}$. The low energy spectrum of cosmic rays follows the law $j(E) \propto E^{-2.7}$ quite closely and therefore, if the high energy particles were to follow the same law, the function $j(E) E^{2.7}$ should be independent of energy E . Part of the scatter in the points could be due to uncertainties in the calibration of the energy scales of the individual experiments, but overall a clear picture emerges of the general shape of the spectrum. Let us review some of the features of this spectrum.

- *The highest energy cosmic rays* It is immediately apparent from Fig. 15.19 that some cosmic rays of quite enormous energy are detected; a few showers containing up to 10^{11} particles have been observed in the largest arrays and, using the rule that there is about 1.4 GeV per particle at maximum development, the initiating particle must have had energy about 10^{20} eV, that is, 16 J in a single particle. The detection of such energetic

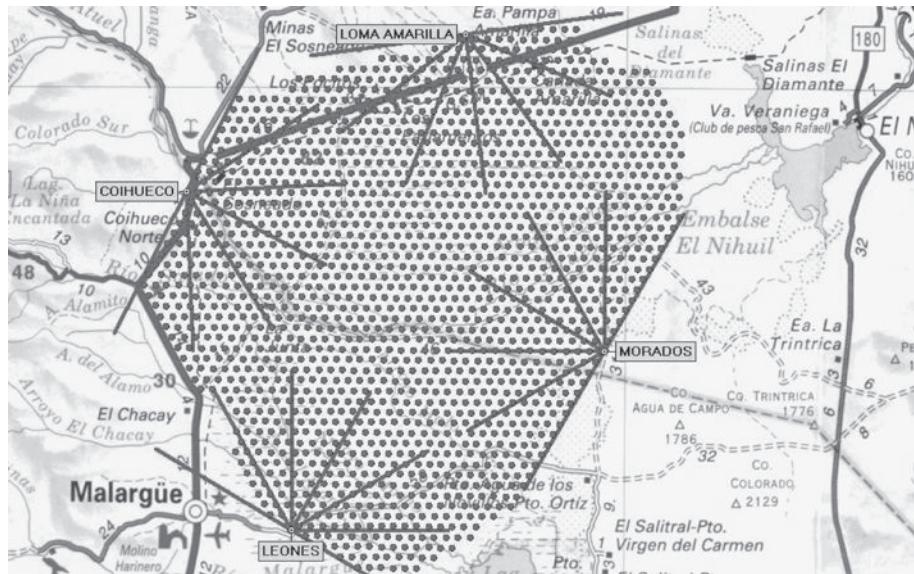


Fig. 15.18 A map showing the layout of the detectors of the Auger Observatory at the Pampa Amarilla (yellow prairie) in western Argentina.

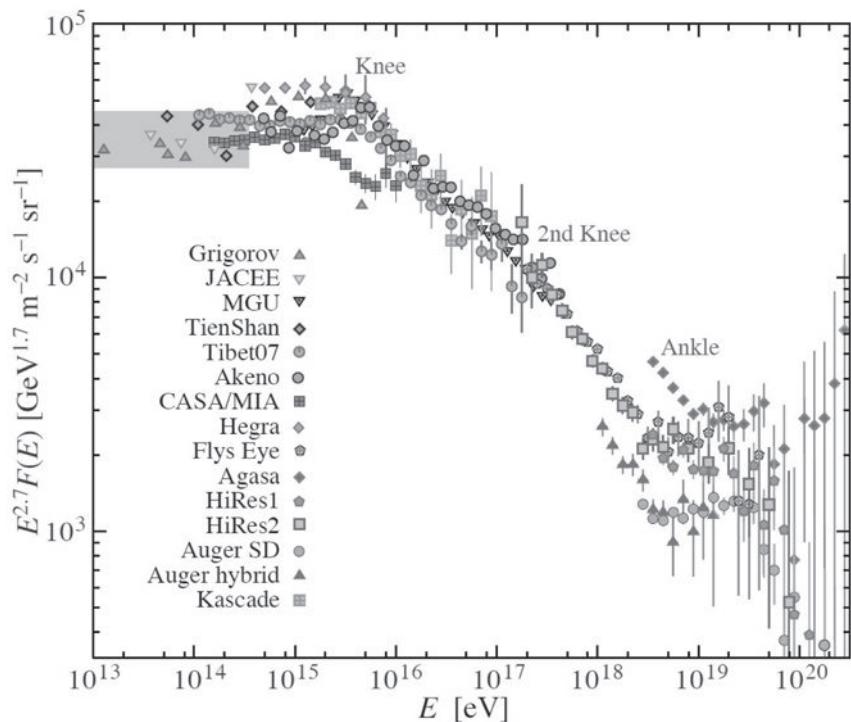


Fig. 15.19 The differential energy spectrum of high energy cosmic rays presented in the form $N(E) E^{2.7}$ (Amsler *et al.*, 2008).

particles has important implications for the origin of the very highest energy cosmic rays since their spectrum may be strongly attenuated because of interactions with photons of the Cosmic Microwave Background Radiation.

- *The energy spectrum at energies $10^{14} \lesssim E \lesssim 10^{18.5}$ eV* The differential spectrum shown in Fig. 15.19 spans the energy range 10^{14} – 10^{20} eV. There is reasonable agreement among the experiments about the shape of the energy spectrum in the energy range 10^{16} – $10^{18.5}$ eV. Overall, the spectrum can be approximated by the expression

$$j(E) = 1.5 \times 10^7 E^{-3.1} \text{ m}^{-2}\text{s}^{-1} \text{ sr}^{-1} \text{ GeV}^{-1}, \quad (15.28)$$

where the energy E is measured in GeV. The exponent $x = 3.1$ is significantly steeper than that found at lower energies, $E \lesssim 10^{15}$ eV, which are in the range $2.5 \leq x \leq 2.7$. There is thus a very clear steepening of the cosmic ray energy spectrum between the low and high energy regimes. The break in the overall spectrum at 10^{15} eV is often referred to as the ‘knee’ in the energy spectrum. There has been some debate about whether or not there is a ‘bump’ in the spectrum at $E \sim 10^{15}$ – 10^{16} eV.

- *The energy spectrum at energies $E \geq 10^{18.5}$ eV* At energies $E \geq 10^{18.5}$ eV, the data show a clear flattening of the energy spectrum. The scatter of the points at the highest energies in Fig. 15.19 reflects the great challenge of determining precisely the energy spectrum of particles which have arrival rates of only 1 particle km^{-2} century $^{-1}$ (Fig. 15.1). This has been the prime motivation for the construction of enormous air-shower arrays such as the Pierre Auger Observatory (Fig. 15.18).

There is agreement that the spectrum flattens above about $10^{18.5}$ eV, referred to as the ‘ankle’ in the energy spectrum. Before the HiRes and Auger projects, the AGASA team reported significant fluxes of particles with energies $E \geq 10^{20}$ eV (see Fig. 15.19). There are major challenges in the energy calibration of these data which have since been revised downwards in energy. The most recent results from the HiRes and Auger projects are shown in Fig. 15.20. The HiRes results were obtained from the Utah array of fly’s eye fluorescence telescopes whereas the Auger project had the advantage of using both fluorescence telescopes and an independent calibration of the particle energies from the 1600 water Cherenkov detectors of the surface array. Figure 15.20 shows that the flattening of the spectrum at energies greater than $10^{18.5}$ eV is observed in both the HiRes and Auger experiments, as well as evidence for a cut-off to the spectrum at energies greater than about 3×10^{19} eV. There is a displacement between the spectra found in the two experiments, but this might be associated with the problems of energy calibration. It is encouraging that both spectra have similar overall shapes.

- *Chemical composition* The air-shower technique does not allow direct estimates of the charges and masses of primary cosmic rays with energies $E \geq 10^{16}$ eV to be made. The observation of multi-cored showers has suggested that some of the primaries may be nuclei. Another approach involves estimating the path length X_{\max} at which the maximum development occurs as a function of the energy of the primary particles. If the primary particle of a given energy were an iron nucleus rather than a proton, it would have a Lorentz factor 56 less than the proton. Therefore the secondary and subsequent generations of charged particles would have on average much smaller Lorentz factors and the maximum development would occur higher in the atmosphere than the corresponding

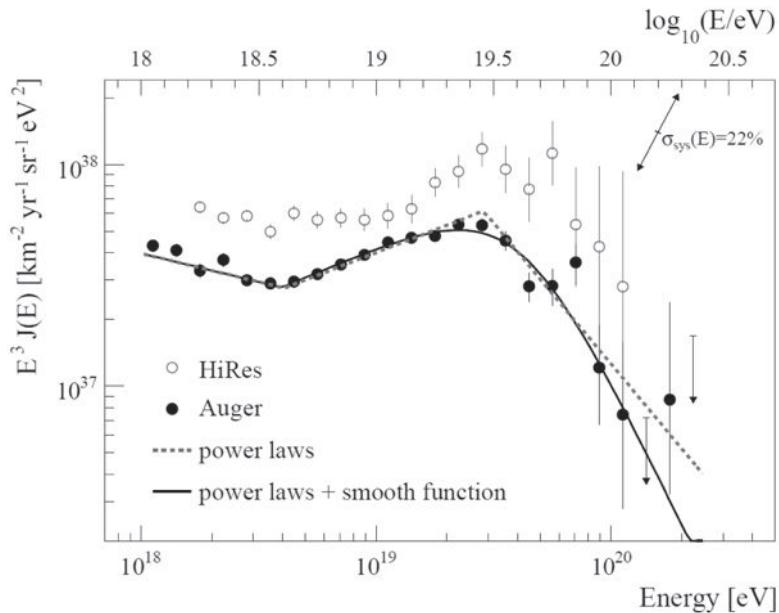


Fig. 15.20 The differential energy spectrum of high energy cosmic rays presented in the form $N(E) E^3$ for energies greater than 10^{18} eV (Abraham *et al.*, 2010b). The data are from the HiRes and Auger experiments. The combined energy spectrum is fitted with two functions and compared to data from the HiRes instrument. The effects of systematic uncertainties of the energy scale of 22% are indicated by arrows.

shower associated with a proton of the same energy. The predicted relations for protons and iron nuclei from numerical simulations of these showers are shown in the left-hand panels of Fig. 15.21 for comparison with observations with the HiRes and Auger experiments.

The development of the numbers of charged particles in the shower $N(X)$ at slant path length X can be fitted by the Gaisser–Hillas distribution,

$$N(X) = N_{\max} \left(\frac{X - X_0}{X_{\max} - X_0} \right)^{\frac{X_{\max} - X_0}{\Lambda}} \exp \left(\frac{X_{\max} - X}{\Lambda} \right), \quad (15.29)$$

where N_{\max} is the number of particles at shower maximum, X_{\max} is the slant depth of the maximum, and λ is a shower-development parameter. It turns out that the Gaisser–Hillas function provides a very good fit to the observed development of the numbers of charged particles in extensive air-showers as a function of path length (Fig. 15.17). Detailed simulations of the development of extensive air showers in the atmosphere for protons and heavier nuclei demonstrate how X_{\max} of the showers and the spread in the values of X_{\max} depend upon the path length through the atmosphere. The results are compared with the observations from the HiRes and Auger experiments in Fig. 15.21 (Aloisio *et al.*, 2009); the root mean square dispersion in the values of X_{\max} is denoted $\text{RMS}(X_{\max})$ in the right-hand panels.

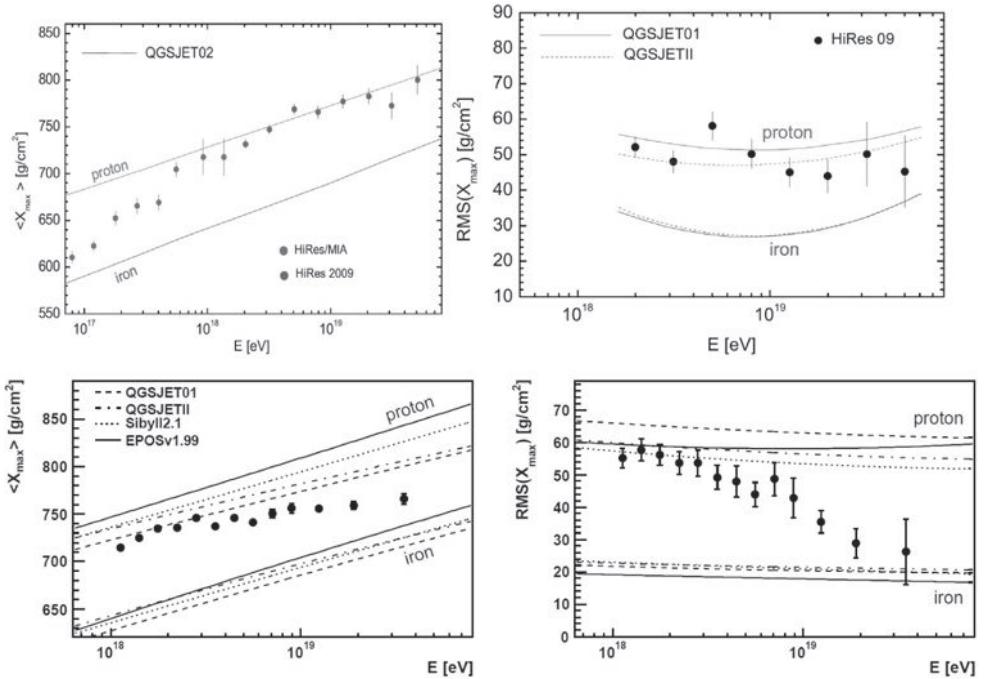


Fig. 15.21

The mean values of X_{\max} and $\text{RMS}(X_{\max})$ as a function of particle energy as measured by the HiRes (top panels) and Auger (lower panels) experiments compared with the predicted relations assuming the ultra-high energy cosmic rays are protons and iron nuclei (Aloisio *et al.*, 2009; Abraham *et al.*, 2010a).

The variation of $\text{RMS}(X_{\max})$ with energy is a particularly important relation since it is not as sensitive to models of the air-showers as the variation of X_{\max} with path length. The reason for the dependence of $\text{RMS}(X_{\max})$ upon the mass of the particle can be appreciated from the following argument. Consider the case of an iron nucleus with the same total energy as a proton. The energy per nucleon and the Lorentz factor γ of the iron particle are 56 times less than that of the proton. The iron nucleus can be considered to consist of 56 very small nucleons, according to the arguments given in Sect. 10.1. Each of these can be considered to be the initiator of a nucleonic cascade and will have a different value of X_{\max} because of the stochastic nature of the nucleonic cascade. Each cascade provides an independent sample of the value of X_{\max} and so according to the central limit theorem, when these are added together, the value of X_{\max} is determined with a standard deviation which is $\sqrt{56}$ smaller than that of an individual event. The value of $\text{RMS}(X_{\max})$ for iron nuclei would be expected to be about seven times narrower than that of protons. This effect is referred to in the literature as the *superposition principle*. The differences between the theoretical loci for iron nuclei and protons in the right-hand panels of Fig. 15.21 are largely attributed to this principle. The numbers are small and so we would not expect the result to be exact and, in addition, the predictions are sensitive to the way in which heavy nuclei are fragmented in the atmosphere. Nonetheless, the

argument illustrates why $\text{RMS}(X_{\max})$ is an important discriminant of the average masses of the particles.

At energies less than 10^{18} eV, the HiRes data in Fig. 15.21 indicate that the particles are largely heavy elements but at higher energies, the data suggest that the highest energy particles are protons. The Auger data in Fig. 15.21 only refer to particles with energies greater than 10^{18} eV and show that an increasing fraction of the ultra-high energy cosmic rays are heavy nuclei. The breadth of the distribution of values of X_{\max} in the Auger data in particular shows clear evidence of narrowing with increasing particle energy. The importance of these results for astrophysics and cosmology are discussed in Sect. 15.12.

15.11 The isotropy of ultra-high energy cosmic rays

In air-shower experiments, the incoming directions of the cosmic rays are determined and hence the isotropy of their distribution on the sky can be estimated. The data from underground muon studies are summarised in Fig. 15.7 which is due to Hillas (1984), who showed that there is agreement among the observers about the magnitude of the anisotropies and their directions on the sky.

At energies less than about 10^{16} eV, an anisotropic flux of cosmic rays is observed with good statistical significance. In the energy range $10^{16} \leq E \leq 10^{19}$ eV, there are measurements and limits to the anisotropy which provide significant constraints on models of the volume within which the cosmic rays are confined. One interpretation of the break in the spectrum of the cosmic rays at 10^{15} eV is that it represents the energy at which cosmic rays can escape from the Galaxy. As shown in Sect. 15.7, it would be expected that the cosmic rays with greater energies would display significant anisotropy if they originated in the Galaxy.

Let us estimate the gyroradius of ultra-high energy cosmic rays in the interstellar magnetic field. In Sect. 7.1, it was shown that the gyroradius r_g of the particle's orbit is

$$r_g = \frac{\gamma m_0 v}{ze} \frac{\sin \theta}{B} = \left(\frac{pc}{ze} \right) \frac{\sin \theta}{Bc} = \frac{R \sin \theta}{Bc}, \quad (15.30)$$

where R , the rigidity, is pc/ze , p the particle's momentum and θ is the pitch angle of the particle's trajectory with respect to the magnetic field direction. For ultra-relativistic energies, the rigidity of a proton in volts is the same as its energy in electron-volts. Representative values of r_g for a Galactic magnetic field strength of 3×10^{-10} T are shown in Table 15.6. These figures can be compared with the thickness of the Galactic disc, about 300 pc, and the radius of the halo, about 3–10 kpc. Thus, the gyroradii of protons of energy 10^{18} eV are roughly equal to the thickness of the Galactic disc.

Figure 15.19 shows that the overall spectrum of cosmic rays steepens at about 10^{15} eV. Ginzburg and Syrovatskii (1964) noted that the gyroradius of a proton of energy 10^{15} eV is roughly equal to the scale of the irregularities, which they inferred were responsible for the scattering of the cosmic rays, $\lambda \sim 0.1$ pc (Sect. 15.7). For energies greater than 10^{15} eV, the

Table 15.6 The radii of curvature r_g of cosmic ray protons and iron nuclei in a magnetic field of magnetic flux density 3×10^{-10} T.

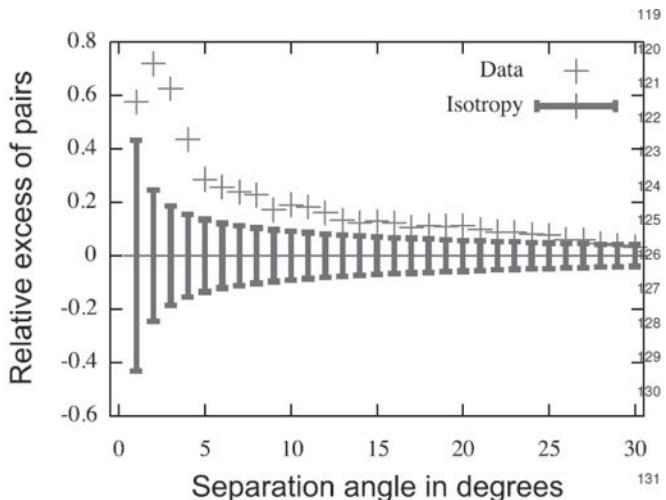
Rigidity, R (V)	Gyroradius of a proton	Gyroradius of an iron nucleus
10^{15}	0.36 pc	0.014 pc
10^{17}	36 pc	1.4 pc
10^{19}	3.6 kpc	140 pc
10^{21}	360 kpc	14 kpc

diffusion approximation would no longer be appropriate and particles could escape more readily from the Galaxy.

This argument was given simple quantitative form by Hillas (1984) who showed that the cosmic ray energy spectrum in the energy range $10^{14} < E < 10^{19}$ eV would be consistent with an power-law injection energy spectrum of the form $Q(E) \propto E^{-2.47}$, the anisotropy being due to the fact that the particles have shorter escape times at higher energies. In a simple picture, the anisotropy is expected to be inversely proportional to the time $T(E)$ which the particles remain within the Galaxy. The expected anisotropy can therefore be related to deviations of the observed energy spectrum from the injection spectrum since $N(E) \propto T(E)Q(E)$. This comparison is illustrated in Fig. 15.7 in which the solid line shows the predicted anisotropy as a function of energy derived from $T(E)$ which in turn is derived from the cosmic ray energy spectrum.

At the very highest energies $E > 10^{19}$ eV, protons with rigidities $R > 10^{19}$ V have gyroradii at least 10 times greater than the half-thickness of the disc. It is, therefore, impossible to confine them to the plane of the Galaxy. In the Galactic magnetic field, the very highest energy cosmic rays with $E \sim 10^{20}$ are barely deflected and consequently they should travel more or less directly from their sources to the Earth. The particles with energies greater than 10^{19} do not show any tendency to cluster towards the Galactic plane and many have arrival directions from high latitudes (Wdowczyk and Wolfendale, 1984). It is generally assumed that the very highest energy particles are of extragalactic origin.

The data accumulated by the Auger project provide the opportunity to test the hypothesis that the very highest energy cosmic rays originate from relatively nearby extragalactic objects. Fifty-eight events with energies $E \geq 6 \times 10^{19}$ eV were observed between 1 January 2004 and 31 March 2009, the positional accuracy of the arrival directions being better than 0.9° . These arrival directions have been correlated with the objects in various catalogues of nearby objects. Originally, the Pierre Auger Collaboration (2007) reported a correlation with nearby active galaxies and quasars contained in the compilation of Véron-Cetty and Véron (2006). More recent data, however, have not strengthened that correlation (Hague, 2009). Rather, the correlation with the objects in the 2MRS catalogue, which is the densest all-sky redshift survey to date and includes about 23 200 galaxies with extinction-corrected magnitudes brighter than $K = 11.25$ (Huchra *et al.*, 2005), has been shown to be just as strong as the correlation with active galaxies (Aublin, 2009). This limiting magnitude corresponds to a redshift $z \approx 0.02$, or a distance of roughly 80 Mpc. Figure 15.22 shows

**Fig. 15.22**

Cross-correlation between the 58 Auger events with energies $E \geq 6 \times 10^{19}$ eV and galaxies contained in the 2MRS catalogue of 23 200 objects to an infrared magnitude of $K = 11.25$ (Aublin, 2009). The diagram shows the cumulative excess of galaxies as a function of angular separation, the error bars indicating the dispersion in 68% of the isotropic realisation.

that there is an excess of pairs with angular separation less than 30° with the maximum excess occurring at an angular separation of about 3° . According to their simulations, such an excess was only found in 1.5×10^{-3} of their model reconstructions of the sky. The inference is that the arrival directions of those cosmic rays with energy $E \geq 6 \times 10^{19}$ eV are correlated with the large scale distribution of galaxies in the nearby Universe. Since active galaxies are part of that population, it still might be the case that the correlation is with the active systems, but other possibilities are perfectly plausible.

The strength of this correlation bears upon the issue of the masses of the highest energy cosmic rays. As shown in Table 15.6, if the particles were iron nuclei, they would have very much smaller gyroradii than protons of the same energy and so their arrival directions would be very much more strongly distorted by the Galactic and intergalactic magnetic fields. Little correlation with the positions of the sources of the cosmic rays would be expected. With a mixed population of protons and iron nuclei, the correlation would only be expected with the proton component.

15.12 The Greisen–Kuzmin–Zatsepin (GKZ) cut-off

The very highest energy cosmic rays have such large Lorentz factors that photons of the Cosmic Microwave Background Radiation have very high energies in the rest frame of the cosmic ray and so photo-pion and photo-pair production can take place, which degrade the energy of the cosmic ray. If a proton is bombarded with high energy γ -rays, pions

are created, the threshold for his process being $\epsilon_t = 200$ MeV and the cross-section about 250 microbarns. The reactions involved are

$$\gamma + p \rightarrow n + \pi^+, \quad (15.31)$$

$$\gamma + p \rightarrow p + \pi^0 \rightarrow p + \gamma + \gamma, \quad (15.32)$$

$$\gamma + p \rightarrow p + N\pi. \quad (15.33)$$

The charged pions then decay creating ultra-high energy muons and muon neutrinos (see Sect. 10.1). These high energy neutrinos could be detected by large ground-based neutrino detectors. The Cosmic Microwave Background Radiation permeates all space and therefore the cosmic rays cannot escape from it. The average energy of the photons of the background is $\varepsilon_0 = 6 \times 10^{-4}$ eV ($\nu = 1.5 \times 10^{11}$ Hz) and therefore, in the rest frame of the cosmic ray, their energies are

$$\varepsilon = \varepsilon_0 \gamma \left(1 + \frac{v}{c} \cos \theta\right). \quad (15.34)$$

Therefore, the threshold for pion production in the limit $v \rightarrow c$, $\cos \theta = 1$ corresponds to an energy $E = 2\gamma m_p c^2$ for protons where $\gamma = \epsilon_t/\epsilon_0$, that is $\gamma = 1.7 \times 10^{11}$, or $E = 1.7 \times 10^{20}$ eV. The proper calculation involves integration over the Planck spectrum of the Cosmic Microwave Background Radiation and over all angles. The threshold for the photo-pion production process then decreases to 5×10^{19} eV, well within the range of cosmic ray energies which have been observed in extensive air-showers.

The mean free path for a single scattering is $\lambda = (\sigma_{\pi p} N_{\text{photon}})^{-1}$. Taking $N_{\text{photon}} = 5 \times 10^8 \text{ m}^{-3}$ for the Cosmic Microwave Background Radiation and $\sigma_{\pi p} = 2.5 \times 10^{-32} \text{ m}^{-2}$, then $\lambda \approx 10^{23} \text{ m}$ corresponding to a propagation length of 3 Mpc or a propagation time of 10^7 years. The energy of the pion created in this process is $\gamma m_\pi c^2$ and therefore the fractional loss of energy of a cosmic ray proton is $\Delta E/E \approx m_\pi/m_p \approx 1/10$. Therefore, the total mean free path for the cosmic ray proton to lose all its energy corresponds to a propagation time of 10^8 years. If cosmic rays of this energy permeated all space and had been present for 10^{10} years, there should be a cut-off in the cosmic ray energy spectrum at about 5×10^{19} eV for protons. If the highest energy cosmic rays are protons, they cannot have originated from further than about 30 Mpc from our Galaxy. These arguments were first discussed by Greisen, Kuz'min and Zatsepin and the cut-off is known as the *GKZ cut-off* (Greisen, 1966; Zatsepin and Kuz'min, 1966). The existence of the cut-off depends upon the value of γ of the cosmic rays and so, if it turned out that the highest energy cosmic rays were iron nuclei rather than protons, the photo-pion production process would not be responsible for the cut-off at 5×10^{19} eV.

A similar calculation can be carried out for the electron–positron photo-pair production process. The threshold energy for this process is 1.02 MeV, about 200 times less than that for the photo-pion production mechanism. Therefore, the process is important for protons with Lorentz factors $\gamma \geq 10^9$, corresponding to proton energies of about 10^{18} eV. The cross-section for this process in the ultra-relativistic limit is $\sigma_{\text{pair}} = 10^{-30} \text{ m}^2$. Although this cross-section is 40 times larger than that for the production of pions, each photo-pair production event removes only 10^{-3} of the energy of the proton and so the time-scale for the protons to lose all their energy is 25 times longer, that is about 2.5×10^9 years.

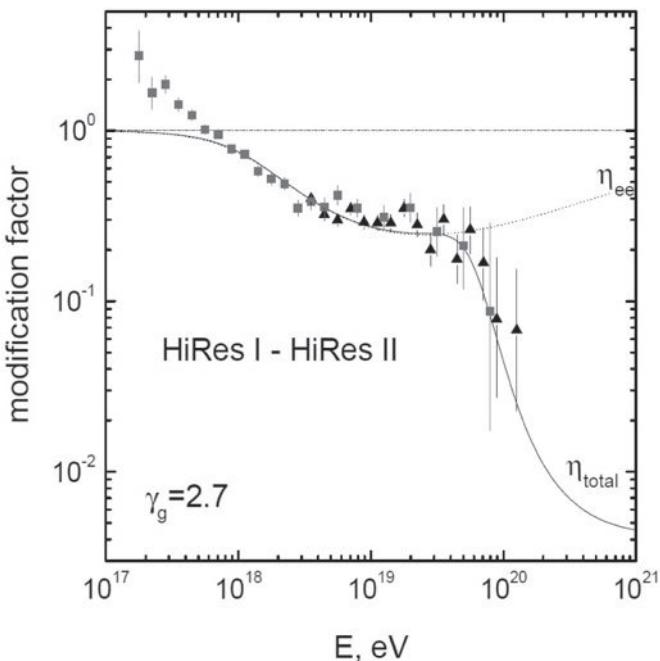


Fig. 15.23 Pair-production dip and GZK cut-off in terms of the modification factor $\eta(E)$ compared with the HiRes observational data. The curves labelled η_{tot} and η_{ee} show the total spectral modification and η_{ee} the spectral modifications due to adiabatic energy losses and pair production (Berezinsky *et al.*, 2006; Aloisio *et al.*, 2009).

Consequently this process is less important for cosmic ray protons of energy 5×10^{19} eV, but it does result in a distortion of the particle spectrum down to energies of about 10^{18} eV.

Detailed calculations of the shape of the GKZ cut-off and the photo-pair production loss mechanisms have been carried out by Berezinsky and his colleagues with the results shown in Fig. 15.23 (Berezinsky, 2007; Aloisio *et al.*, 2009). They present their results in terms of a modification factor $\eta(E)$ which is defined as

$$\eta(E) = J_p(E)/J_p^{\text{unm}}(E) \quad (15.35)$$

where $J_p(E)$ is the spectrum taking account of all the energy losses and $J_p^{\text{unm}}(E)$ excludes the above photo-pion and photo-pair processes. The sharp cut-off at about 5×10^{19} eV is the GKZ cut-off. The dotted line shows the ‘dip’ due to photo-pair production which extends down to about 10^{18} eV. The predictions are compared with the results of the HiRes observations. At energies less than 10^{18} eV, the slope of the observed spectrum increases to $x = 3.1$ and this is interpreted as the spectrum of Galactic cosmic rays. Berezinsky and his colleagues show that the predicted shape of the modification factor $\eta(E)$ seems to have been observed in the HiRes data, implying that the ultra-high energy cosmic rays are extragalactic and must be mainly protons.

At the time of writing, there appears to be a discrepancy between the results of the HiRes and Auger projects. The Auger Collaboration find clear evidence that the highest energy cosmic rays are very much heavier than protons, tending towards the iron locus shown

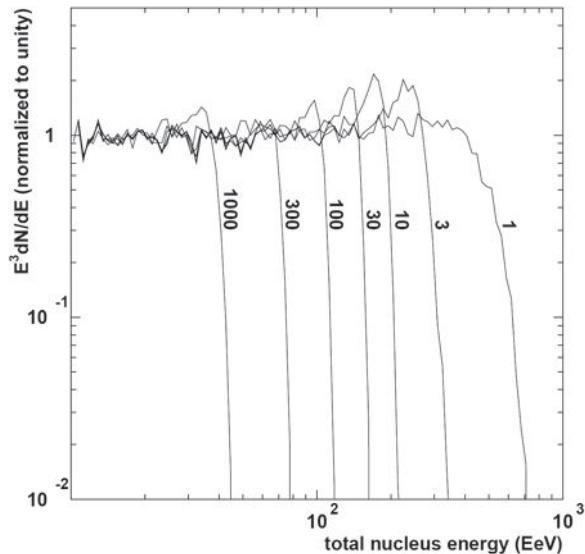


Fig. 15.24 The differential spectra of ultra-high energy cosmic ray iron nuclei (^{56}Fe) after propagating the distance in Mpc indicated on each curve. The source spectrum is assumed to be of power-law form $N(E) \propto E^{-3}$, over the interval $10\text{--}1000\text{ EeV}$, that is, $10^{19}\text{--}10^{21}\text{ eV}$ (Stecker and Salamon, 1999).

in Fig. 15.21. According to Berezinsky and his colleagues, the shape of the Auger energy spectrum also differs from the predicted total spectrum shown in Fig. 15.23.

If the highest energy particles were nuclei, there would still be a cut-off at energies $E \sim 5 \times 10^{19}\text{ eV}$ because of photonuclear interactions. This can be demonstrated by an order of magnitude estimate similar to those given above for pions and electrons. As above, the maximum energy of the photon in the centre of momentum frame of reference is $\varepsilon = 2\gamma\varepsilon_0 = 10^{-3}\gamma\text{ eV}$ for photons at the peak of the spectrum of the Cosmic Microwave Background Radiation. The binding energy per nucleon for nuclei is of the order of 10 MeV per nucleon and so it is not surprising that there is what is referred to as a *giant dipole resonance*, which peaks in the γ -ray energy range 10–30 MeV in the nuclear rest frame. The absorption of the photon leads to the excitation of one or two nucleons which are ejected from the nucleus, beginning the disintegration of the nucleus (Stecker and Salaman, 1999). The necessary value of the Lorentz factor γ to attain this energy in the centre of momentum frame of reference is therefore $\gamma \approx (1 - 3) \times 10^{10}\text{ eV}$. If A is the atomic mass number of the nucleus and E the total energy, the corresponding Lorentz factor is $E/A m_n c^2$. Therefore, at threshold, photo-disintegration of the nucleus takes place at energies $E \geq 10^{18} A\text{ eV}$. Thus, for the typical atomic mass numbers of cosmic ray nuclei, a photo-disintegration cut-off would also appear about $10^{19}\text{--}10^{20}\text{ eV}$. In addition, the path length over which the disintegration of the nuclei would take place is less than that of the photo-pion process. Stecker and Salaman give an approximate expression for the energy-weighted cross-section for the interaction of

$$\int_0^\infty \sigma(\varepsilon) d\varepsilon = 60 \frac{NZ}{A} \times 10^{-31} \text{ m}^2 \text{ MeV}. \quad (15.36)$$

Taking the width of the resonance to be of order 5–10 MeV, this cross-section is larger than the photo-pion and electron-pair production cross-sections. The subsequent disintegration of the nucleus has to be followed by Monte Carlo simulations and these have been carried out by Stecker and Salaman. The results of their detailed calculations are shown in Fig. 15.24 which shows that a cut-off is expected at energies less than 10^{20} eV, even if the highest energy cosmic rays are iron nuclei.

The situation remains to be resolved. The Auger statistics are now significantly greater than those of the HiRes observations and the value of $\text{RMS}(X_{\max})$ undoubtedly decreases with increasing energy. The implications are important. If the highest energy cosmic rays are iron nuclei, there would be no ultra high energy neutrinos associated with the decay of the π^+ particles created in the p-p collisions. Berezinsky and his colleagues refer to the latter possibility as the ‘disappointing model’. The continuing efforts of the Auger team and the construction of even larger air-shower arrays are the ways forward.

16.1 Introduction

In Sect. 8.9, a convincing case was made that the high energy electrons observed at the top of the atmosphere are a representative sample of those present throughout the interstellar medium and are responsible for the diffuse Galactic synchrotron radio emission. In Sect. 15.4, a similar exercise was carried out for cosmic ray protons. The spectrum and properties of the γ -ray emission of the Galaxy provide compelling evidence that a flux of cosmic ray protons, of similar properties to those observed in our vicinity in the Galaxy, permeates the plane of the Galaxy. In this chapter, these observations are interpreted in terms of the propagation of these particles from their sources through the interstellar medium and the energetics of potential sources in the Galaxy. Key diagnostic tools are provided by the aging processes which can result in observable features in the synchrotron spectra of relativistic electrons and by the energy requirements of sources of synchrotron radiation. The TeV γ -ray emission of supernova remnants is direct evidence for the presence of large fluxes of particles with cosmic ray energies in supernova remnants, although it is not yet clear if these are associated with high energy electrons or protons (Sect. 16.4.2). The tools are developed in the context of the origin of cosmic rays in supernovae explosions and are of applicability to the whole of high energy astrophysics. They will be further developed in the analysis of the physics of extragalactic radio sources and active galactic nuclei in Part IV.

16.2 Energy loss processes for high energy electrons

High energy electrons are subject to a number of energy loss processes which cause distortions of their injection energy spectra as they propagate through the interstellar medium from their sources and these potentially provide information about their life histories. The loss mechanisms involve interactions with matter, magnetic fields and radiation. First, the various loss mechanisms need to be written in forms suitable for inclusion in the diffusion loss equation under interstellar conditions. The dependence of the loss processes upon the energy of the electrons is the key factor.

16.2.1 Ionisation losses

From (5.36) and (5.37), the ionisation loss rate for the case of atomic hydrogen is

$$-\left(\frac{dE}{dt}\right)_i = 7.64 \times 10^{-15} n (3 \ln \gamma + 19.8) \text{ eV s}^{-1}, \quad (16.1)$$

where $\gamma = (1 - v^2/c^2)^{-1/2} = E/m_e c^2$ is the Lorentz factor of the electron and n is the number density of hydrogen atoms in particles m^{-3} . Thus, an electron of energy $E(\text{eV})$ loses all its energy in a time

$$\tau = \frac{E}{(dE/dt)_i} = \frac{E(\text{eV})}{7.64 \times 10^{-15} n (3 \ln \gamma + 19.8)} \text{ s}. \quad (16.2)$$

For example, a 3 GeV electron, for which $\gamma \approx 6000$, loses all its energy in $3 \times 10^{14}/n$ years so that, for a typical interstellar density of $n = 10^6 \text{ m}^{-3}$, its lifetime is about 3×10^8 years. Because the dependence upon energy is only logarithmic, a good approximation for the energy loss rate is that it amounts to about $10^{-5} n \text{ eV per year}$.

16.2.2 Bremsstrahlung

From the analyses of Sect. 6.6, the energy loss rate of ultra-relativistic electrons by bremsstrahlung can be written

$$-\frac{1}{E} \left(\frac{dE}{dt}\right)_{\text{brems}} = 4n Z^2 r_e^2 \alpha c \bar{g}, \quad (16.3)$$

where r_e is the classical electron radius, α is the fine structure constant and \bar{g} is a Gaunt factor (Heitler, 1954). When the nuclei are unscreened, as in a fully ionised plasma,

$$\bar{g} = \ln(2\gamma) - \frac{1}{3} = \ln \gamma + 0.36, \quad (16.4)$$

and the loss rate of the electron in a fully ionised hydrogen plasma is

$$-\frac{1}{E} \left(\frac{dE}{dt}\right)_{\text{brems}} = 7.0 \times 10^{-23} n (\ln \gamma + 0.36) \text{ s}^{-1}. \quad (16.5)$$

In the case of total screening,

$$\bar{g} = \ln(183 Z^{-1/3}) - \frac{1}{18}, \quad (16.6)$$

and so, for neutral hydrogen, the loss rate associated with interactions between the ultrarelativistic electron and the nuclei of the hydrogen atoms is

$$-\frac{1}{E} \left(\frac{dE}{dt}\right)_{\text{brems}} = 3.66 \times 10^{-22} n \text{ s}^{-1}. \quad (16.7)$$

If account is taken of the bremsstrahlung due to interactions with the bound electrons in the atoms as well, Z^2 should be replaced by $Z(Z + 1.3)$ in (16.3). For a wide range of electron energies, $100 \leq \gamma \leq 10^5$, the expressions (16.5) and (16.7) are the same within about a factor of 2, reflecting the fact that the differences in collision parameters occur within the logarithm. In both cases, the energy loss rate is proportional to the energy of the electron.

16.2.3 Adiabatic losses

If the electrons are confined within an expanding volume, they are subject to adiabatic losses as they do work which reduces the internal energy of the gas. It is convenient to develop expressions for the loss rate which can be used in different applications. First consider a non-relativistic Maxwellian gas. The loss of internal energy U when the gas does work in expanding from volume V to $V + dV$ is

$$dU = -p dV , \quad (16.8)$$

where p is the pressure of the gas. For a perfect monatomic gas, $U = \frac{3}{2}nkTV$ and $p = nkT$ where n is the number density of particles and T their temperature. Since the average energy per particle is $\frac{3}{2}kT$,

$$dU = nV dE = -\frac{2}{3}nE dV . \quad (16.9)$$

Therefore setting the total number of particles $nV = N$,

$$\frac{dE}{dt} = -\frac{2nE}{3N} \frac{dV}{dt} . \quad (16.10)$$

Now dV/dt is the rate of expansion of the volume V which is determined by the velocity field $\mathbf{v}(r)$. For a cube of sides dx, dy, dz , we add together the changes in volume due to the differential expansion of each of the three pairs of faces of the cube, that is,

$$\frac{dV}{dt} = (v_{x+dx} - v_x) dy dz + (v_{y+dy} - v_y) dx dz + (v_{z+dz} - v_z) dx dy . \quad (16.11)$$

Making a Taylor expansion for small dx, dy, dz ,

$$\frac{dV}{dt} = \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) dx dy dz = (\nabla \cdot \mathbf{v})V . \quad (16.12)$$

Substituting into (16.10),

$$\frac{dE}{dt} = -\frac{2}{3}(\nabla \cdot \mathbf{v})E . \quad (16.13)$$

This is the general expression for the energy loss rate due to adiabatic losses of a non-relativistic particle in an expanding flow. It can also be written in terms of the momentum p of the particle, since $E = p^2/2m$, in which case

$$\frac{dp}{dt} = -\frac{1}{3}(\nabla \cdot \mathbf{v})p . \quad (16.14)$$

The generalisation to the relativistic case is straightforward and follows exactly the same procedure but we write instead $U = 3nkTV$ and $p = \frac{1}{3}U$. In the ultra-relativistic limit,

$$dE/dt = -\frac{1}{3}(\nabla \cdot \mathbf{v})E ; \quad dp/dt = -\frac{1}{3}(\nabla \cdot \mathbf{v})p . \quad (16.15)$$

In the case of a uniformly expanding sphere, the velocity distribution inside the sphere is $v = v_0(r/R)$ where v_0 is the velocity of expansion of the outer radius R of the sphere.

Then, $\nabla \cdot \mathbf{v} = 3(v_0/R)$ and so, for relativistic particles within the sphere,

$$-\left(\frac{dE}{dt}\right)_{\text{ad}} = \left(\frac{v_0}{R}\right) E = \left(\frac{1}{R} \frac{dR}{dt}\right) E \quad \text{or} \quad -\frac{1}{E} \left(\frac{dE}{dt}\right)_{\text{ad}} = \left(\frac{1}{R} \frac{dR}{dt}\right). \quad (16.16)$$

The condition that adiabatic losses are important is that the time-scale of the expansion, $[(1/R)(dR/dt)]^{-1}$, is of the same order as the time the particles have been within the expanding volume. If the particles have always been confined within the source during its expansion, adiabatic losses are always important. Notice that the loss rate at any given time depends upon the instantaneous dynamics of the expansion through the term $(R^{-1} \dot{R})$.

16.2.4 Synchrotron radiation

The total energy loss rate by synchrotron radiation is given by the expression (8.8) which, in the ultra-relativistic limit $v \rightarrow c$, becomes

$$-\left(\frac{dE}{dt}\right)_{\text{synch}} = 2\sigma_T c \gamma^2 U_{\text{mag}} \sin^2 \theta, \quad (16.17)$$

or

$$-\left(\frac{dE}{dt}\right)_{\text{synch}} = 1.587 \times 10^{-14} \gamma^2 B_{\perp}^2 \text{ W} = 9.9 \times 10^4 \gamma^2 B_{\perp}^2 \text{ eV s}^{-1}, \quad (16.18)$$

where $B_{\perp} = B \sin \theta$ is measured in teslas. Averaging over an isotropic pitch angle distribution, the factor of 2 in (16.8) is replaced by $\frac{4}{3}$ and then

$$-\left(\frac{dE}{dt}\right)_{\text{synch}} = \frac{4}{3} \sigma_T c \gamma^2 U_{\text{mag}} = 6.6 \times 10^4 \gamma^2 B^2 \text{ eV s}^{-1}; \quad (16.19)$$

the total loss rate is proportional to E^2 .

16.2.5 Inverse Compton scattering

From the analysis of Sect. 9.3, the loss rate of an ultra-relativistic electron by inverse Compton scattering in a radiation field of energy density U_{rad} , given by (9.41), is

$$-\left(\frac{dE}{dt}\right) = \frac{4}{3} \sigma_T c \gamma^2 U_{\text{rad}}. \quad (16.20)$$

Inverse Compton scattering is important whenever high energy electrons propagate through a radiation field. At a typical point in the Galaxy, the average energy density of optical photons due to the light of the stars amounts to about $U_{\text{rad}} \approx 6 \times 10^5 \text{ eV m}^{-3}$ (see, for example, Toller, 1990). In addition, the Cosmic Microwave Background Radiation permeates all space with energy density at the present epoch $U_{\text{rad}} = 2.65 \times 10^5 \text{ eV m}^{-3}$, adopting a radiation temperature $T_r = 2.736 \text{ K}$.

The relative importance of inverse Compton and synchrotron losses for electrons of the same energy in our Galaxy is found by comparing (16.20) and (16.19),

$$\frac{(dE/dt)_{\text{IC}}}{(dE/dt)_{\text{synch}}} = \frac{U_{\text{rad}}}{U_{\text{mag}}}. \quad (16.21)$$

Putting in representative figures, $B = 3 \times 10^{-10}$ T and $U_{\text{rad}} = 6 \times 10^5$ eV m⁻³, $U_{\text{rad}}/U_{\text{mag}} = 3$. Thus, inverse Compton losses as well as synchrotron losses are likely to be important for high energy electrons in our Galaxy. Furthermore, suppose the electrons were to escape from the Galaxy into intergalactic space. Then, the energy loss due to synchrotron radiation is very small but the Cosmic Microwave Background Radiation is omnipresent and the electrons cannot escape from it. Therefore, they lose all their energy by inverse Compton scattering of the photons of the Cosmic Microwave Background. The maximum lifetime of high energy electrons anywhere in the Universe because of inverse Compton scattering is therefore

$$\tau = \frac{E}{dE/dt} = \frac{E}{\frac{4}{3}\sigma_{\text{TC}}c\gamma^2 U_{\text{CMBR}}} = \frac{2.3 \times 10^{12}}{\gamma} \text{ years}, \quad (16.22)$$

taking U_{CMBR} to be 2.65×10^5 eV m⁻³. Thus, for 100 GeV electrons, which are observed at the top of the atmosphere, $\tau \leq 1.15 \times 10^7$ years.

The typical energies of the high energy electrons which radiate in the radio waveband correspond to $\gamma = 10^3 - 10^4$ (Sect. 8.9) and so scattering of the photons of the Cosmic Microwave Background Radiation generates X-rays since

$$\nu \approx \gamma^2 \nu_0 \approx 10^6 \times 10^{11} \text{ Hz} = 10^{17} \text{ Hz}; \quad \varepsilon = 0.4 \text{ keV}. \quad (16.23)$$

The scattering of optical photons produces a flux of γ -rays with

$$\nu \approx \gamma^2 \nu_0 \approx 10^6 \times 10^{15} \text{ Hz}; \quad \varepsilon = 4 \text{ MeV}. \quad (16.24)$$

Thus, part of the Galactic γ -ray emission in the 1–100 MeV energy band may be due to inverse Compton scattering (Sect. 15.4 and Fig. 15.9). Inverse Compton scattering is important wherever relativistic electrons propagate through regions containing large energy densities of radiation, for example, in the central regions of quasars. In some quasars, the radiation density of synchrotron radiation is so high that inverse Compton scattering by the electrons which emitted the synchrotron radiation itself may be an important loss mechanism, the process known as *synchro-Compton radiation*. In extreme cases, it may lead to the *synchro-Compton catastrophe* (Sect. 22.3).

16.3 Diffusion-loss equation for high energy electrons

16.3.1 Distortions of the injection energy spectrum

The diffusion-loss equation was derived in Sect. 7.5 and has the form

$$\frac{dN(E)}{dt} = D \nabla^2 N(E) + \frac{\partial}{\partial E} [b(E)N(E)] + Q(E). \quad (16.25)$$

This expression describes the evolution of the energy spectrum as the particles diffuse from their sources under conditions of continuous energy losses described by $-(dE/dt) = b(E)$. Rather than solve the full diffusion-loss equation, let us simplify the problem by

neglecting diffusion and assume that the spectrum has reached a steady state under the continuous injection of particles described by $Q(E)$ and energy losses given by $b(E)$. First of all, we assume there is an infinite, uniform distribution of sources, each injecting high energy electrons with an injection spectrum $Q(E) = \kappa E^{-p}$. Then, all spatial dependence disappears from the solution and the diffusion-loss equation reduces to

$$\frac{d}{dE}[b(E)N(E)] = -Q(E) \quad \text{or} \quad \int d[b(E)N(E)] = - \int Q(E) dE. \quad (16.26)$$

Assuming $N(E) \rightarrow 0$ as $E \rightarrow \infty$, (16.26) can be integrated to give

$$N(E) = \frac{\kappa E^{-(p-1)}}{(p-1)b(E)}. \quad (16.27)$$

The form of $b(E)$ for high energy electrons under interstellar conditions is the sum of all the processes discussed in Sect. 16.2, namely,

$$b(E) = -\left(\frac{dE}{dt}\right) = A_1 \left(\ln \frac{E}{m_e c^2} + 19.8\right) + A_2 E + A_3 E^2. \quad (16.28)$$

The first term on the right-hand side, A_1 , describes ionisation losses and depends only weakly upon energy; the second term, A_2 , represents bremsstrahlung losses and adiabatic losses, if they are relevant; and the last term, A_3 , describes inverse Compton and synchrotron losses. This analysis enables us to understand the effect of continuous energy losses upon the initial spectrum of high energy electrons. Thus, from (16.27),

- (i) if ionisation losses dominate, $N(E) \propto E^{-(p-1)}$ – the energy spectrum is flatter by one power of E ;
- (ii) if bremsstrahlung or adiabatic losses dominate, $N(E) \propto E^{-p}$ – the spectrum is unchanged;
- (iii) if inverse Compton or synchrotron losses dominate, $N(E) \propto E^{-(p+1)}$ – the spectrum is steeper by one power of E .

These are also the equilibrium spectra expected whenever the continuous injection of electrons takes place over a time-scale longer than the lifetimes of the electrons involved. For example, if we inject electrons continuously with a spectrum E^{-p} into a source component for a time t and synchrotron losses are the only important loss process, an electron of energy E_s loses all its energy in a time t such that $-(dE/dt)t = E_s$. From (16.19), $E_s = 0.125/B^2 t$ eV, where B is measured in teslas and t in years. The electron energy spectrum in the source is different for electrons with energies greater and less than E_s ; for lower energies, the electrons do not lose a significant fraction of their energy and therefore the spectrum is the same as the injection spectrum, $N(E) \propto E^{-p}$. For energies greater than E_s , the particles have lifetimes less than t and we only observe those produced during the previous synchrotron lifetime τ_s of the particles of energy E , that is, $\tau_s \propto 1/E$. Therefore the spectrum of the electrons is one power of E steeper, $N(E) \propto E^{-(p+1)}$, in agreement with the analysis proceeding from the steady-state solution of the diffusion-loss equation.

There are two useful analytic solutions for the electron energy distribution under continuous energy losses due to synchrotron radiation and inverse Compton scattering. In the first case, it is assumed that there is continuous injection of electrons with a power-law energy

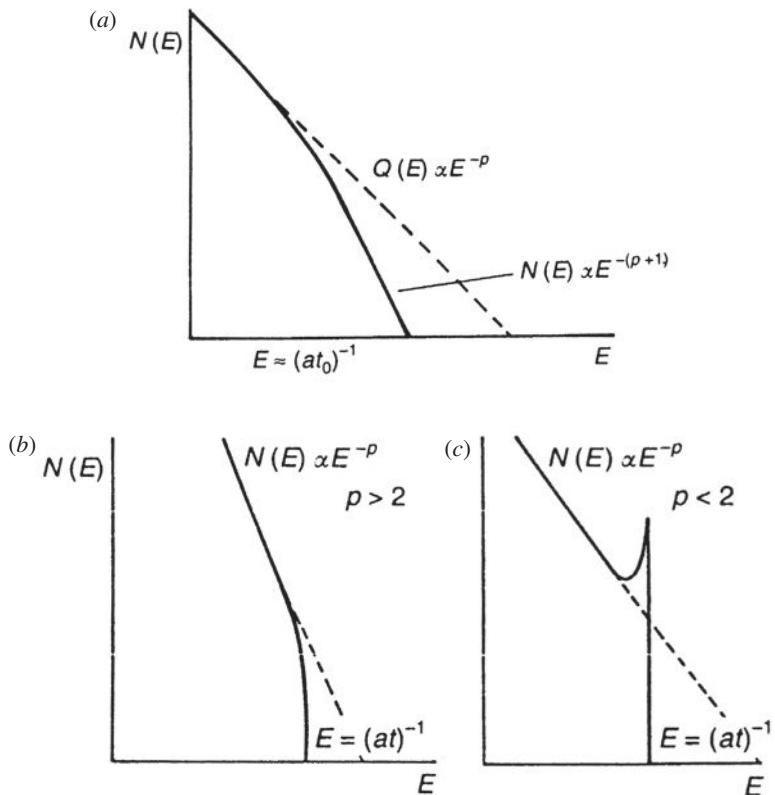


Fig. 16.1

(a) A solution of the diffusion-loss equation for steady-state injection of electrons with a power-law energy spectrum $Q(E) \propto E^{-p}$ in the presence of energy losses of the form $dE/dt = -aE^2$. (b) The time evolution of a power-law energy distribution injected at $t = 0$ with no subsequent injection of electrons. In this case $p > 2$. (c) As in case (b) but with $p < 2$.

spectrum $Q(E) = \kappa E^{-p}$ for a time t_0 . Writing the loss rate of the electrons in the form $b(E) = aE^2$, the energy spectrum after time t_0 is

$$N(E) = \begin{cases} \frac{\kappa E^{-(p+1)}}{a(p-1)} [1 - (1 - aEt_0)^{p-1}] & \text{if } aEt_0 \leq 1, \\ \frac{\kappa E^{-(p+1)}}{a(p-1)} & \text{if } aEt_0 > 1. \end{cases} \quad (16.29)$$

This form of spectrum is shown in Fig. 16.1a and agrees with the physical arguments given in the last paragraph.

A second useful case is that of the injection of electrons with a power-law energy spectrum at $t = 0$ with no subsequent injection of electrons. If $Q(E) = \kappa E^{-p} \delta(t)$, the solution of (16.25), again ignoring the diffusion term, is

$$N(E) = \kappa E^{-p} (1 - aEt)^{p-2}. \quad (16.31)$$

Thus, after time t , there are no electrons with energies greater than $(at)^{-1}$. If $p > 2$, the spectrum steepens smoothly to zero at $E = (at)^{-1}$ (Fig. 16.1b); if $p < 2$, there is a cusp in the energy spectrum at $E = (at)^{-1}$ (Fig. 16.1c). The number of electrons, however, remains finite and constant.

16.3.2 The high energy electron energy spectrum in the local interstellar medium

In a complete solution, we would take account of the diffusion of the electrons from their sources in the plane of the Galaxy to intergalactic space. It is simpler to work in an approximation in which the electrons only remain a certain time $\tau(E)$ within some Galactic confinement volume before escaping into intergalactic space. The diffusion term can then be replaced by a term of the form $N(E)/\tau(E)$ where $\tau(E)$ can be chosen to mimic diffusion if a Gaussian distribution of escape times is used or a ‘leaky box’ if an exponential distribution is adopted (see Sect. 15.5 and Fig. 15.10). If the escape time depends upon energy, simple solutions for the equilibrium energy spectrum can be found, if the energy dependence is chosen to be of power-law form (see Sect. 15.5.2).

As an example of how the results of Sect. 16.3.1 can be applied, a mean escape time of cosmic rays from the vicinity of the Solar System of 1.5×10^7 years was derived from the abundances of radioactive nuclei such as ^{10}Be in Sect. 15.6. Taking the ionisation loss rate to be $10^{-5}n$ eV per year and $n = 10^6 \text{ m}^{-3}$, 150 MeV electrons would lose all their energy in this time; adopting $B = 3 \times 10^{-10} \text{ T}$ for the interstellar magnetic flux density, 1.5×10^7 years would correspond to the lifetime of 20 GeV electrons due to synchrotron radiation losses. Since inverse Compton scattering losses are of at least comparable importance to synchrotron losses (see Sect. 16.2.5), a lifetime of 1.5×10^7 years would result in a spectral break at energies $E < 10 \text{ GeV}$. Thus, if electrons were continuously injected uniformly into the interstellar medium with energy spectrum $Q(E) \propto E^{-p}$, in the low energy regions in which ionisation losses dominate, $N(E) \propto E^{-(p-1)}$; in the high energy region, in which synchrotron losses dominate, $N(E) \propto E^{-(p+1)}$.

Breaks in the injection spectra are thus expected at energies at which the lifetimes of the electrons are equal to their escape times. There is a clear break in the spectrum of the Galactic radio emission, the radio spectral index at frequencies less than 200 MHz being about $\alpha = 0.4$, corresponding to $p = 1.8$ and about $\alpha = 0.9$, corresponding to $p = 2.8$ at higher frequencies (Fig. 8.13). The break in the radio spectrum at about 200 MHz can be associated with 3 GeV electrons radiating in a magnetic field of flux density $3 \times 10^{-10} \text{ T}$. There are, however, many reasons why the breaks in the energy spectra of the electrons and their corresponding radio spectra would be smoothed out over a wide range of energies and frequencies about these critical values. For example, Webster (1971) showed that the Galactic radio spectrum is barely consistent with a model in which the electron energy spectrum is a broken power law with $p = 1.8$ at energies less than 3 GeV and $p = 2.8$ at greater energies when the energy spectrum is convolved with the spectrum of synchrotron radiation (Fig. 16.2a).

A more serious concern involves solutions of the diffusion-loss equation if the electrons diffused from a distribution of sources in the plane of the Galaxy to the Solar

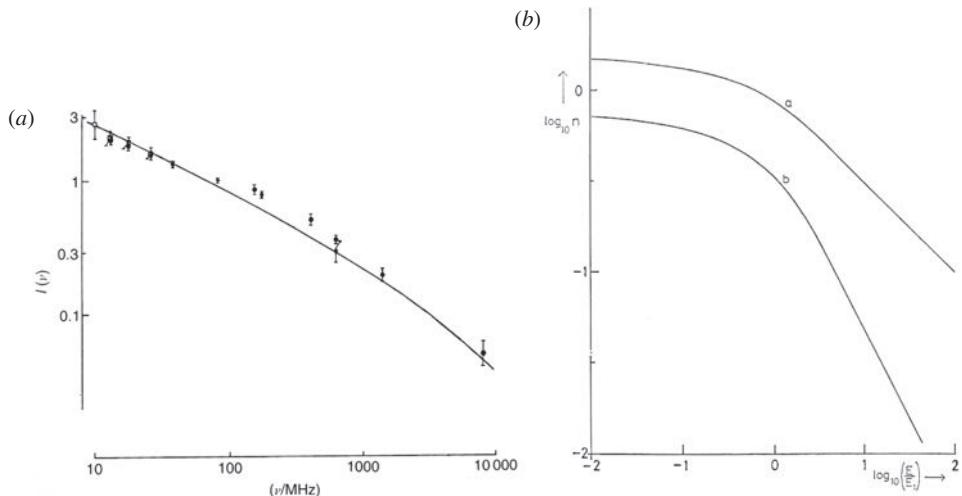


Fig. 16.2

(a) The spectrum of the radio emission of the Galaxy compared with the predicted synchrotron spectrum of a particle distribution consisting of a broken power-law distribution, convolved with the spectrum of synchrotron radiation (Webster, 1971). (b) The predicted cosmic ray electron spectrum at the centre of (a) a disc and (b) a sphere of sources with injection energy spectrum $N(E) \propto E^{-1.8}$. The ordinate is plotted as $\log N(E) E^{1.8}$ which tends to a constant value at low energies (Webster, 1970).

System. Webster (1970) showed that, if the break in the electron spectrum is attributed to the combination of inverse Compton and synchrotron losses and the observer is located at the centre of a uniform disc distribution of sources, the observed break in the spectral index p would be 0.5 rather than 1. A break in the spectral index p of 1 is obtained for a spherical distribution of sources about the observer. Even more serious is the fact that this change in spectral index would take place over a factor of 30 in energy (Fig. 16.2b). Convoluting this energy distribution with the synchrotron radiation spectrum, it is impossible to reproduce the observed change in radio spectrum between low and high radio frequencies. As Webster shows, this is the sharpest break which can be obtained from any diffusion model. This is a fatal argument for models in which the observed flux of cosmic rays reaches the Earth by diffusion from a distribution of Galactic sources.

For energies at which the loss time is of the same order as the synchrotron loss time, we might expect to see variations in the radio spectral index of the Galaxy in different directions. This is a notoriously difficult observation because whole sky surveys have been made with telescopes of very different angular resolutions and polar diagrams. Lawson and his colleagues (1987) attempted to do this for whole sky radio surveys in the frequency range 150 MHz to 5 GHz. They find little evidence for spectral variations at low radio frequencies but significant fluctuations are found at high frequencies. This would be consistent with the idea that significant variations occur at frequencies at which the loss time is of the same order as or less than the escape time.

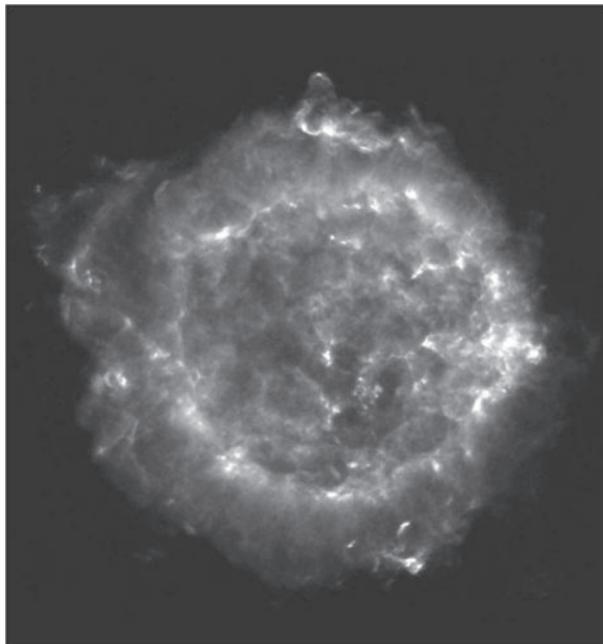


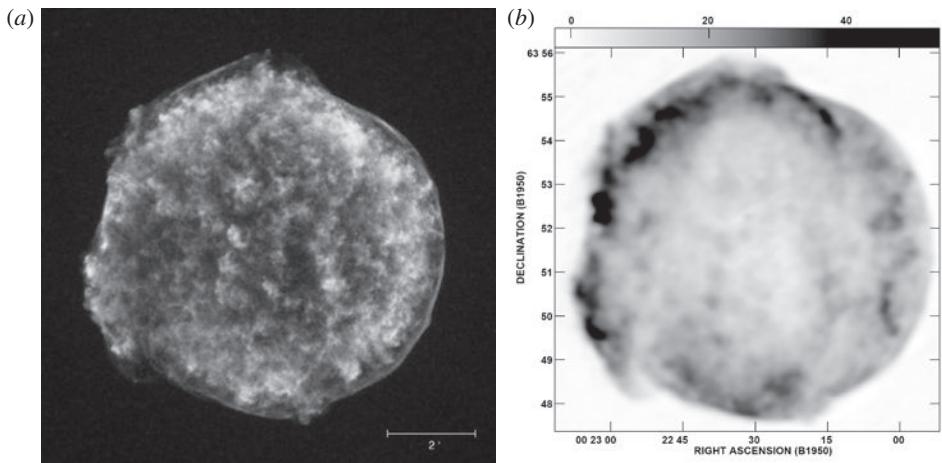
Fig. 16.3 The radio image of Cassiopeia A observed by the Very Large Array. (Courtesy of the National Radio Astronomy Observatory.) This image of the synchrotron emission of the supernova remnant should be compared with the composite infrared, optical and X-ray image shown in Fig. 13.3.

16.4 Supernova remnants as sources of high energy particles

The radio and γ -ray properties of supernova remnants provide convincing evidence that they are sources of high energy particles. Supernova remnants come in two varieties, the most common being the *shell-like* remnants observed in the radio, optical, X-ray and γ -ray wavebands in which the emission originates from an expanding shell. In contrast, the *filled-centre* or *Crab-like* remnants possess a central source in the form of a young pulsar and the brightness distribution is a maximum in the centre.

16.4.1 Radio observations of shell-like supernova remnants

The optical and infrared emission of shell-like supernova remnants is associated hot gas and dust while the X-ray emission is the bremsstrahlung of very hot gas associated with the heating of interstellar material by the supernova blast wave. In contrast, the radio emission is highly polarised, has a power-law energy spectrum and is identified with the synchrotron radiation of ultra-relativistic electrons accelerated in the shock waves associated with the blast waves. These features are illustrated by the spectacular high resolution images of the young supernova remnant Cassiopeia A in the optical, infrared, X-ray and radio wavebands (Figs 13.3 and 16.3). The supernova probably exploded about 275 years ago, but the event

**Fig. 16.4**

(a) A composite image of Tycho Brahe's supernova of 1572 from observations made by the Calar Alto telescope, the Chandra X-ray Observatory and the Spitzer Space Infrared Observatory. (Courtesy of NASA and the Space Telescope Science Institute.) (b) A radio image of Tycho Brahe's supernova at 1.4 GHz, as observed by the VLA (Katz-Stone *et al.*, 2000).

was not observed by astronomers at the time. The remarkable feature is how closely the relativistic particles and magnetic fields responsible for the radio emission coincide with the features in the optical, infrared and X-ray wavebands. We will argue that this coincidence is associated with the presence of strong shocks which are responsible for compressing the gas, dust and magnetic fields and, at the same time, result in first-order particle acceleration of energetic electrons to high energies.

Another example of a young shell-like supernova is Tycho's supernova (Fig. 16.4) which is also shell-like, possessing a very sharp, almost circular boundary. Again, there is very close coincidence between the thermal material and the synchrotron radio emission of ultra-relativistic electrons. In the case of the old supernova remnant the Cygnus Loop (Fig. 13.9), which is about 5000 years old, there is again a close correlation between regions of enhanced radio emissivity and the presence of optical emission lines, the diffuse X-ray emission filling a large fraction of the volume of the expanding shell.

In young supernova remnants, the intensity of the radio emission is so high that the energy densities of relativistic electrons and magnetic fields greatly exceed values which could be obtained by compressing the Galactic flux of high energy electrons and magnetic field in a strong shock. Therefore, the relativistic particles and fields must originate within the remnant itself. On the other hand, the radio emission of some old remnants may result from compression of the interstellar magnetic field and relativistic electron fluxes. High compressions are possible in the shells of old remnants because of the rapid cooling in the shocked region, when the temperature falls below about 10^6 K.

As discussed in Sects 13.1.2 and 13.1.3, the light-echo technique has been used to determine the types of supernovae which gave rise to these remnants. Cassiopeia A turned out to be a Type IIb supernova and Tycho's supernova a Type Ia. Despite the former being

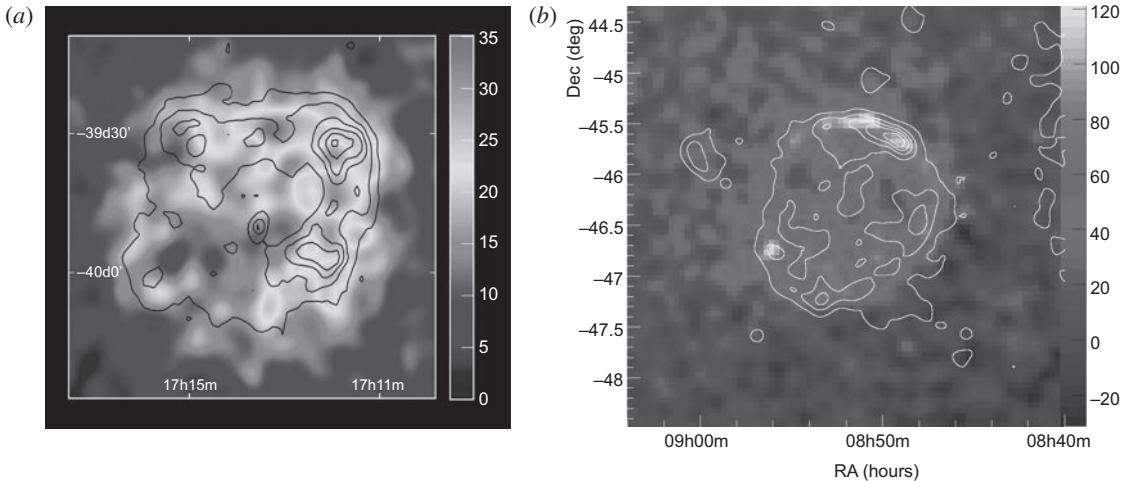


Fig. 16.5 (a) TeV gamma-ray image of RX J1713.7–3946 obtained by the HESS ultra-high energy Cherenkov γ -ray telescope in Namibia. The superimposed contours show the X-ray surface brightness as observed by the ASCA X-ray observatory in the 1–3 keV range (Aharonian *et al.*, 2004). (b) HESS observations of the ultra-high energy γ -rays from the shell-type supernova remnant RX J0852.0–4622 (Aharonian *et al.*, 2007b). Superimposed on the γ -ray image are contours of the X-ray emission at energies $\varepsilon \geq 1.3$ keV from the ROSAT All-Sky Survey.

a core-collapse supernova and the latter a white dwarf deflagration, both have ended up creating an expanding sphere of very high temperature gas.

16.4.2 γ -ray observations of shell-type supernovae

Direct evidence for the presence of particles with cosmic ray energies in supernova remnants has been provided by the remarkable γ -ray images obtained by imaging atmospheric Cherenkov γ -ray telescopes such as the HESS array in Namibia. The supernova remnant RX J1713.7–3946 (G347.3–0.5) was discovered by the ROSAT X-ray Observatory in the X-ray waveband. The HESS observations at TeV energies resolved the structure of the supernova remnant which has a shell morphology similar to that observed at X-ray energies (Fig. 16.5a). The γ -ray spectrum extends to energies greater than 100 TeV, providing direct evidence for the presence of high energy particles with at least this energy in the remnant.

More recently another shell-type supernova remnant, RX J0852.0–4622, was observed by the HESS γ -ray telescope, the emission originating from a remarkably thin shell, very similar to the structure of the remnant observed in the radio and X-ray wavebands (Fig. 16.5b). The contours show the X-ray image at energies $\varepsilon \geq 1.3$ keV from the ROSAT All-Sky Survey. Again, the coincidence of the γ -ray image with those at X-ray and radio wavelengths is direct evidence for the acceleration of particles to at least 100 TeV in the shells of supernova remnants. These observations are entirely consistent with the prevailing view that high energy particles with energies up to at least 20 TeV can be accelerated by first-order Fermi acceleration in the presence of strong shock waves (Chap. 17).

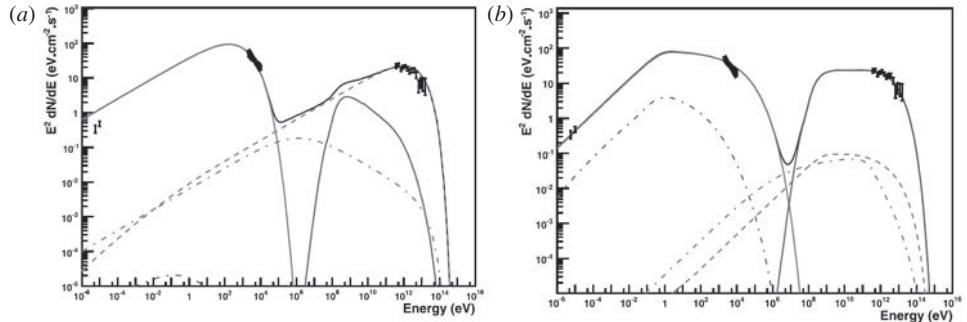


Fig. 16.6

Broad-band spectral energy distribution models for the supernova remnant RX J0852.0–4622 (Aharonian *et al.* 2007b). (a) In this inverse Compton scattering model, the assumed distance of the remnant is 200 pc, the injection spectrum is of power-law form with spectral index 2.4 with an exponential cut-off at 40 TeV and the electron–proton ratio is $\kappa = 1.7 \times 10^{-3}$. The magnetic field strength is 6×10^{-10} T and the density of the ambient medium 8×10^3 m $^{-3}$. (b) In this neutral pion decay model, the assumed distance of the remnant is 1 kpc, the injection spectrum is of power-law form with spectral index 2.0 with an exponential cut-off at 100 TeV and the electron–proton ratio is $\kappa = 4.5 \times 10^{-4}$. The magnetic field strength is 8.5×10^{-9} T and the density of the ambient medium 2×10^6 m $^{-3}$. The solid lines at low and high energies show the contributions of the synchrotron and pion production processes, respectively. The inverse Compton scattering and bremsstrahlung emissions are shown by dashed and dot-dashed lines, respectively, at high energies.

Aharonian and his colleagues (2007b) provide a detailed discussion of the nature of the γ -ray emission of RX J0852.0–4622, possible origins being inverse Compton scattering of the X-rays by high energy electrons, or the decay of neutral pions produced by p-p interactions between high energy protons and the thermal matter in the supernova shell. A complication is that the distance of the source, and consequently its expansion age, have not been established but it probably lies between about 200 pc and 1 kpc from the Solar System, the corresponding ages being several hundreds to several thousands of years. Examples of modelling the radio, X-ray and γ -ray emission of the remnant are shown in Fig. 16.6a and b. There is flexibility in the choice of parameters, particularly in view of the uncertainty about the distance of the remnant and the ambient thermal matter density. The adopted values for the case of inverse Compton scattering (Fig. 16.6a) and neutral pion decay (Fig. 16.6b) are given in the figure caption. It is evident that at this stage either process might be responsible for the TeV γ -ray emission.

If the TeV γ -rays were the result of the decay of neutral pions, the energy in high energy protons within the remnant can be evaluated. Aharonian and his colleagues (2007b) quote the result

$$W_p^{\text{tot}} \approx 10^{42} \left(\frac{D}{200 \text{ pc}} \right) \left(\frac{n}{10^6 \text{ m}^{-3}} \right)^{-1} \text{ J}, \quad (16.32)$$

extrapolating the energy spectrum of the protons to 1 GeV. This figure can be compared with the release of binding energy in the formation of a neutron star in a supernova explosion, $E \sim 10^{46}$ J, and the typical kinetic energy of expansion of the remnant $E \sim 10^{44}$ J.

16.5 The minimum energy requirements for synchrotron radiation

It is of importance to know how much relativistic particle energy there is present in supernova remnants and estimates of this can be found from the *minimum energy requirements* for synchrotron radiation. The following arguments can be applied to the synchrotron radiation emitted by the source at any frequency, be it radio, optical or X-ray wavelengths.

Suppose a source of synchrotron radiation has luminosity L_ν at frequency ν and its volume is V . The spectrum of the radiation is of power-law form, $L_\nu \propto \nu^{-\alpha}$. The radio luminosity can be related to the energy spectrum of the ultra-relativistic electrons and the magnetic flux density B in the source through the formula (8.130) for synchrotron radiation,

$$L_\nu = A(\alpha)V\kappa B^{1+\alpha}\nu^{-\alpha}, \quad (16.33)$$

where the electron energy spectrum per unit volume is $N(E) dE = \kappa E^{-p} dE$, $p = 2\alpha + 1$ and $A(\alpha)$ is a constant. From the expression (8.131), the constant $A(\alpha)$ is

$$A(\alpha) = 2.344 \times 10^{-25} (1.253 \times 10^{37})^\alpha a(2\alpha + 1) \quad (16.34)$$

where SI units are used and $a(2\alpha + 1)$ is given in Table 8.2. Writing the energy density in relativistic electrons as ε_e , the total energy present in the source responsible for the radio emission is

$$W_{\text{total}} = V\varepsilon_e + V \frac{B^2}{2\mu_0} = V \int \kappa E N(E) dE + V \frac{B^2}{2\mu_0}. \quad (16.35)$$

The luminosity of the source L_ν determines only the product $V\kappa B^{1+\alpha}$. If V is assumed to be known, the radio luminosity may be produced either by a large flux of relativistic electrons in a weak magnetic field, or vice versa. There is no way of deciding which combination of ε_e and B is appropriate from observations of L_ν . Between the extremes of dominant magnetic field and dominant particle energy, however, there is a minimum total energy requirement.

There will also be relativistic protons present in the radio source. In the interstellar medium of our Galaxy, there is about 100 times as much energy in relativistic protons as there is in electrons. In contrast, in the Crab Nebula, the energy in relativistic protons cannot be much greater than that in the electrons from the dynamics of its expansion. Therefore, to take account of the protons, it is customary to assume that they have energy β times that of the electrons, that is,

$$\varepsilon_{\text{protons}} = \beta\varepsilon_e; \quad \varepsilon_{\text{total}} = (1 + \beta)\varepsilon_e = \eta\varepsilon_e. \quad (16.36)$$

We therefore write

$$W_{\text{total}} = \eta V \int_{E_{\min}}^{E_{\max}} \kappa E N(E) dE + V \frac{B^2}{2\mu_0}. \quad (16.37)$$

The energy requirement (16.37) depends upon both κ and B but they are related through equation (16.33) for the observed luminosity of the source L_ν . We also require the relation between the frequency of emission of an ultra-relativistic electron of energy $E = \gamma m_e c^2 \gg m_e c^2$ in a magnetic field of flux density B . The maximum intensity of synchrotron radiation

occurs at a frequency

$$\nu = \nu_{\max} = 0.29\nu_c = 0.29 \frac{3}{2}\gamma^2\nu_g = 0.29 \times 4.199 \times 10^{10}\gamma^2B = CE^2B , \quad (16.38)$$

where ν_g is the non-relativistic gyrofrequency and $C = 1.22 \times 10^{10}/(m_e c^2)^2$. Therefore, the relevant range of electron energies in the integral (16.37) is related to the range of observable radio frequencies through

$$E_{\max} = \left(\frac{\nu_{\max}}{CB}\right)^{1/2}; \quad E_{\min} = \left(\frac{\nu_{\min}}{CB}\right)^{1/2}, \quad (16.39)$$

where ν_{\max} and ν_{\min} are the maximum and minimum frequencies for which the radio spectrum is known, or the range of frequencies relevant to the problem in hand. Then,

$$W_{\text{particles}} = \eta V \int_{E_{\min}}^{E_{\max}} E \kappa E^{-p} dE = \frac{\eta V \kappa}{(p-2)} (CB)^{(x-2)/2} [\nu_{\min}^{(2-p)/2} - \nu_{\max}^{(2-p)/2}] . \quad (16.40)$$

Substituting for κ in terms of L_ν and B from equation (16.33),

$$W_{\text{particles}} = \frac{\eta V}{(p-2)} \left[\frac{L_\nu}{A(\nu)VB^{1+\alpha}\nu^{-\alpha}} \right] (CB)^{(p-2)/2} [\nu_{\min}^{(2-p)/2} - \nu_{\max}^{(2-p)/2}] . \quad (16.41)$$

Preserving only the essential observables,

$$W_{\text{particles}} = G(\alpha)\eta L_\nu B^{-3/2} , \quad (16.42)$$

where $G(\alpha)$ is a constant which depends weakly on α , ν_{\max} and ν_{\min} if $\alpha \approx 0.75$. From the above analysis,

$$\begin{aligned} G(\alpha) &= \frac{1}{a(p)(p-2)} [\nu_{\min}^{-(p-2)/2} - \nu_{\max}^{-(p-2)/2}] \nu^{(p-1)/2} \\ &\times \frac{(7.4126 \times 10^{-19})^{-(p-2)}}{2.344 \times 10^{-25}} (1.253 \times 10^{37})^{-(p-1)/2} . \end{aligned} \quad (16.43)$$

Therefore

$$W_{\text{total}} = G(\alpha) \eta L_\nu B^{-3/2} + V \frac{B^2}{2\mu_0} . \quad (16.44)$$

The variations of the energies in particles and magnetic field are shown in Fig. 16.7 as a function of magnetic flux density B . There is a minimum total energy which can be found by minimising (16.44) with respect to B ,

$$B_{\min} = \left[\frac{3\mu_0}{2} \frac{G(\alpha)\eta L_\nu}{V} \right]^{2/7} . \quad (16.45)$$

This magnetic flux density B_{\min} corresponds approximately to equality of the energies in the relativistic particles and magnetic field. Substituting B_{\min} into (16.44),

$$W_{\text{mag}} = V \frac{B_{\min}^2}{2\mu_0} = \frac{3}{4} W_{\text{particles}} . \quad (16.46)$$

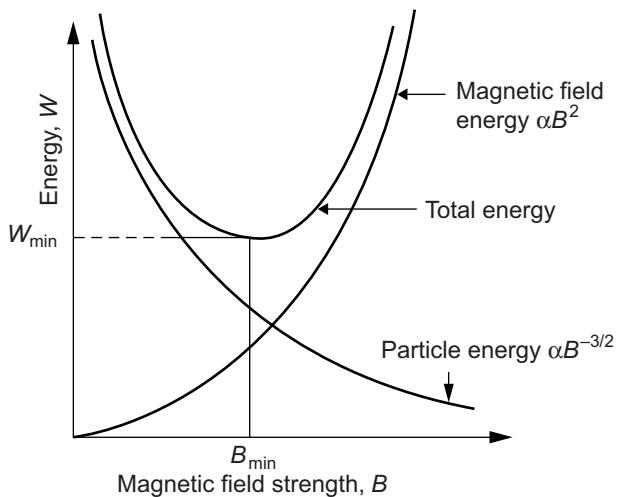


Fig. 16.7 Illustrating the origin of the minimum energy requirements of a radio source as a function of magnetic flux density B .

Thus, the minimum energy requirement corresponds closely to what is often referred to as *equipartition*. The minimum total energy is

$$W_{\text{total(min)}} = \frac{7}{6\mu_0} V^{3/7} \left[\frac{3\mu_0}{2} G(\alpha)\eta L_\nu \right]^{4/7}. \quad (16.47)$$

These results are frequently used in the study of the synchrotron radiation from radio, optical and X-ray sources but their limitations should be appreciated.

- (i) There is no physical justification for the source components being close to equipartition. It might be that the particle and magnetic field energies in the source components tend towards equipartition. For example, the magnetic field in the source components may be stretched and tangled by motions in the plasma so there might be rough equipartition between the magnetic energy density and the energy density in turbulent motions. The turbulent motions might also be responsible for accelerating the high energy particles and these particles might come into equipartition with the turbulent energy density if the acceleration mechanism were very efficient. In this way, there might be a physical justification for the source components being close to equipartition.
- (ii) The amount of energy present in the source component is sensitive to the value of η , that is, the amount of energy present in the form of protons and nuclei.
- (iii) The total amount of energy in relativistic particles is dependent upon the limits assumed to the energy spectrum of the particles. If $\alpha = 0.75$, we need only consider the dependence upon v_{\min} , which is quite weak, $W_{\min} \propto v_{\min}^{-0.25}$. There might, however, be large fluxes of low energy relativistic electrons present in the source components with a quite different energy spectrum and we would have no way of knowing whether or not they are present from the radio observations.
- (iv) Perhaps most important is the fact that the energy requirements depend upon the volume of the source. In the above calculation, it was assumed that the particles and

magnetic field fill the source volume uniformly. In fact, the emitting regions might occupy only a small fraction of the apparent volume of the source, for example, if the synchrotron emission originated in filaments or subcomponents within the overall volume V , as suggested by the radio image of Cassiopeia A (Fig. 16.3). In this case, the volume which should be used in the expressions (16.45) and (16.47) would be smaller than V . A *filling factor* f is used to describe the fraction of the volume occupied by radio emitting material. If f is small, the energy requirements are reduced.

- (v) On the other hand, we obtain a firm lower limit to the *energy density* within the source components since

$$U_{\min} = \frac{W_{\text{total}}(\min)}{V} = \frac{7}{6\mu_0} V^{-4/7} \left[\frac{3\mu_0}{2} G(\alpha)\eta L_v \right]^{4/7}. \quad (16.48)$$

For dynamical purposes, the energy density is more important than the total energy since $p = (\gamma - 1)U$ where γ is the ratio of specific heat capacities.

Thus, the values of the magnetic flux density and minimum energy found from these arguments should be considered only order of magnitude estimates. If the magnetic flux density and the energy in the relativistic gas depart radically from their equipartition values, the energy requirements are increased.

It is cumbersome to work out $G(\alpha)$ and so a simplified estimate of the energy requirements can be found in the following way. If the spectral index is $\alpha = 0.75$, the upper limit v_{\max} can be neglected in comparison with v_{\min} in evaluating $G(\alpha)$. If the luminosity $L(v)$ at a certain frequency v is known, a lower limit to the energy requirements is found if we set $v_{\min} = v$. Making these simplifications, the minimum energy requirement is

$$W_{\min} \approx 3.0 \times 10^6 \eta^{4/7} V^{3/7} v^{2/7} L_v^{4/7} \text{ J}, \quad (16.49)$$

where V is the volume of the source in m^3 , the luminosity $L(v)$ is in W Hz^{-1} and the frequency v is in Hz. In the same units, the minimum magnetic flux density is:

$$B_{\min} = 1.8 \left(\frac{\eta L_v}{V} \right)^{2/7} v^{1/7} \text{ T}. \quad (16.50)$$

This line of reasoning was very important in the late 1950s when the energy requirements of extragalactic radio sources were first estimated by Burbidge (1956; 1959). A good example was provided by the radio source Cygnus A, the brightest extragalactic radio source in the northern sky. At that time, it was thought that the source consisted of two components roughly 100 kpc in diameter. The source had luminosity roughly $8 \times 10^{28} \text{ W Hz}^{-1}$ at 178 MHz. Inserting these values into (16.49), the minimum total energy is $2 \times 10^{52} \eta^{4/7} \text{ J}$ corresponding to the rest-mass energy of $3 \times 10^5 \eta^{4/7} M_\odot$ of matter. The realisation of the enormous energy demands of extragalactic radio sources in the form of relativistic particles and magnetic fields stimulated the very rapid growth of high energy astrophysics in the 1960s. Evidently a very considerable amount of mass has to be converted into relativistic particle energy and ejected from the nucleus of the galaxy into enormous lobes well outside the body of the galaxy.

16.6 Supernova remnants as sources of high energy electrons

Let us estimate the minimum energy in the form of relativistic particles and magnetic fields present in the young supernova remnant Cassiopeia A (Fig. 16.3). Cassiopeiae A is the brightest radio source in the northern sky and has flux density $2720 \text{ Jy} = 2.72 \times 10^{-23} \text{ W m}^{-2} \text{ Hz}^{-1}$ at a frequency of 1 GHz. The bulk of the radiation originates within a roughly spherical volume of angular diameter 4 arcmin. For our present purposes, it is sufficient to assume that the radio source is a uniform sphere. Adopting a distance of 2.8 kpc, the radio luminosity of the source at 1 GHz is $L_{1\text{GHz}} = 2.6 \times 10^{18} \text{ W Hz}^{-1}$ and its volume is $V = 5.3 \times 10^{50} \text{ m}^3$. From (16.50), the minimum energy requirement is therefore $W_{\min} = 2 \times 10^{41} \eta^{4/7} \text{ J}$ and the equipartition magnetic field strength $B = 10\eta^{2/7} \text{ nT}$. This is a large amount of energy, recalling that the rest-mass energy of one solar mass of material is $M_{\odot}c^2 = 2 \times 10^{47} \text{ J}$. Another important comparison is with the binding energy released in the collapse of the inner regions of the pre-supernova star, $E \sim 10^{46} \text{ J}$. In addition, the radial motions and masses of the optical filaments from the centre of the explosion can be found from spectroscopic analyses of their intense emission lines. It is estimated that the total kinetic energy in optical filaments is about $2 \times 10^{44} \text{ J}$. When we recall that the minimum energy to produce the radio emission is $2 \times 10^{41} \eta^{4/7} \text{ J}$, it can be seen that there must exist a highly efficient means of converting the gravitational energy of collapse into high energy particle and magnetic field energy. We can therefore be confident that supernova remnants are very powerful sources of high energy electrons. If $\eta \approx 100$, a value appropriate to protons and electrons observed at the top of the atmosphere, the minimum energy requirements would be increased by a factor $\eta^{4/7}$, that is, about an order of magnitude.

We can estimate if supernova remnants such as Cassiopeia A can account for the energy density of high energy particles observed in our vicinity in the interstellar medium. Suppose the average time interval between Galactic supernovae is t_{SN} , the volume within which the high energy particles are confined is V , the characteristic escape time of particles from the confinement volume is t_c and the average energy release in high energy particles per supernova is E_0 . Then, the local energy density of high energy particles in a steady state is

$$\varepsilon_{\text{CR}} = \frac{t_c}{t_{\text{SN}}} \frac{E_0}{V}. \quad (16.51)$$

Suppose the lifetime of the high energy particles in the disc of the Galaxy is 1.5×10^7 years (Sect. 15.6). If the thickness of the disc is 700 pc and its radius 10 kpc, the volume is $6 \times 10^{60} \text{ m}^3$. The local energy density of high energy particles is roughly 10^6 eV m^{-3} (Sect. 15.3) and so, adopting a supernova rate of one per 30 years, the average energy release per supernova must be $E_0 \approx 10^{42} \text{ J}$. This should be compared with the above estimate for Cassiopeiae A of $2 \times 10^{41} \eta^{4/7} \text{ J}$. Thus, energetically, it is quite feasible to account for the total energy density of cosmic rays in terms of their origin in supernova explosions and their aftermath.

16.7 The evolution of supernova remnants

Many of the observed features of *shell-like supernova remnants* can be accounted for in terms of general considerations of the supersonic expansion of a sphere of very hot gas into the diffuse interstellar medium. The dynamical evolution can be divided into four stages.

- (i) The energy liberated in the supernova explosion is deposited in the envelope of the star which is heated to a very high temperature and ejected with velocity about $(10\text{--}20) \times 10^3 \text{ km s}^{-1}$. In a uniformly expanding sphere, the radial velocity of the expanding gas is proportional to the radial distance from its centre and the sphere remains undecelerated as long as the mass of interstellar gas swept up in the expansion is much less than the mass of the ejected gas. The density and pressure distributions are largely dependent upon the initial conditions. If the temperature inside the expanding sphere is uniform, it decreases adiabatically as the sphere expands according to the usual law $T \propto R^{-3(\gamma-1)}$ where γ is the ratio of specific heat capacities. Because the expansion velocity is highly supersonic, a shock front is formed ahead of the expanding sphere.

In the simplest picture, there is an abrupt discontinuity between the sphere of expanding gas and the swept-up material. The situation resembles the case of the supersonic piston analysed in Sect. 11.3.2 (see Fig. 11.7). The expanding sphere of hot gas is represented by the piston and the interface between the piston and the shocked gas immediately in front of it is called a *contact discontinuity*. In that example, it was shown that a shock wave runs ahead of the contact discontinuity and that the region between the piston and the shock is heated to a high temperature. In the limit of very strong shocks, meaning that the Mach number M_1 is very much greater than 1, the ratio of densities on either side of the shock wave is $\rho_2/\rho_1 = (\gamma + 1)/(\gamma - 1) = 4$ if the ratio of specific heat capacities is $\gamma = 5/3$ (see 11.73). In this limit, the ratio of temperatures behind and in front of the shock wave is $T_2/T_1 = 2\gamma(\gamma - 1)M_1^2/(\gamma + 1)^2 = (5/16)M_1^2$ if $\gamma = 5/3$, where M_1 is the Mach number of the shock wave with respect to the sound speed of the unshocked gas (11.74). Since supernova explosions are highly supersonic with respect to the sound speed in the interstellar gas, it is evident that the shocked gas must be heated to very high temperatures and therefore it is not at all surprising that young supernova remnants are intense X-ray emitters (see Fig. 7.14). The structure of the remnant during this phase is shown in Fig. 16.8a.

- (ii) When the swept-up mass becomes greater than the ejected mass, the dynamics are described by the adiabatic blast-wave similarity solution of Taylor (1950a; 1950b) and Sedov (1959). During this phase, the overall dynamics are determined by the total mass of expanding gas, which is now mostly swept up interstellar gas, and the energy released in the initial explosion. The dynamics can be derived using dimensional analysis.¹ The only variables which can enter the solution are the energy of the explosion E and the density of the ambient medium ρ_0 which provides all the swept-up mass. The

¹ The derivation of this result using the Buckingham Π theorem can be found in my book *Theoretical Concepts in Physics* (Longair, 2003).

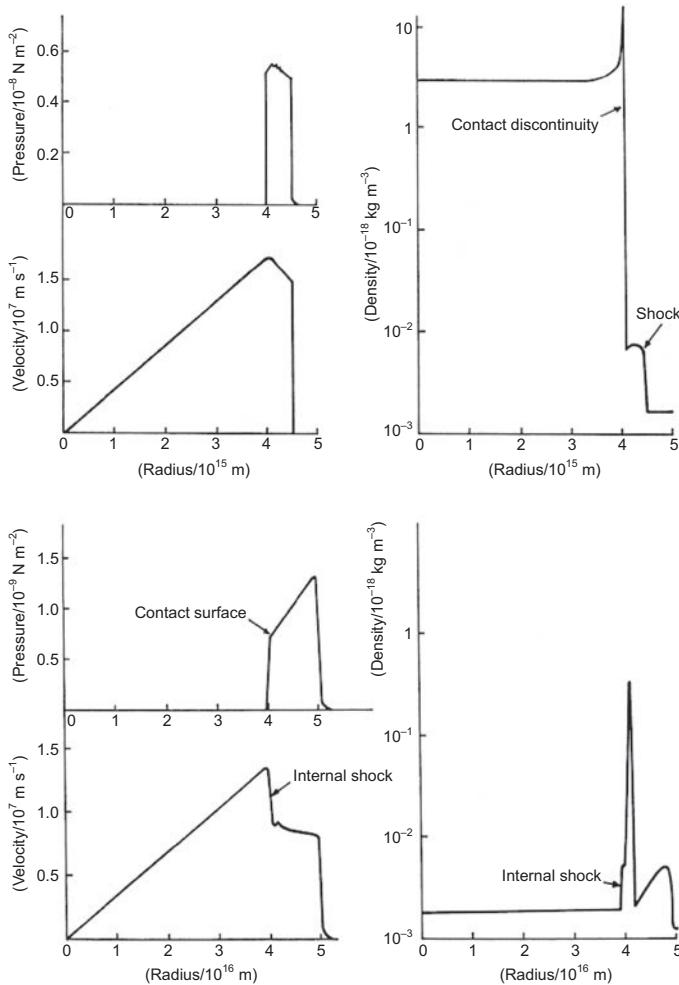


Fig. 16.8 The evolution of the internal properties of a young supernova remnant from the time when the mass ratio is very much less than 1 to the Sedov phase when the swept-up mass dominates the dynamics. Features of importance are the formation of the dense shell of material just inside the contact discontinuity and the role of the reverse inner shock in heating up the material inside the expanding sphere (Gull, 1975).

dynamical variables are the radius and time. The dimensions of E/ρ_0 are L^5T^{-2} and consequently, we can find a dimensionless quantity $(E/\rho_0)t^2/r^5$ which describes the dynamics of the expansion. Therefore, $r \propto (E/\rho_0)^{1/5}t^{2/5}$. Taylor's study of explosions, including his analysis of Mack's movie of the first Trinity atomic bomb test in the New Mexico desert, showed that this relation is closely obeyed.

When the deceleration of the expanding sphere becomes significant, there are important changes in the temperature, density and pressure distribution inside the sphere. The most important is that the outer shells of the expanding sphere are decelerated first and so the material inside the sphere catches up with the material in the outer layers.

Therefore, the matter density begins to increase at the boundary of the expanding sphere. As the deceleration continues, the flow of gas into the outer layers becomes supersonic relative to the sound speed inside the sphere itself and so a shock wave forms on the inner edge of the compressed outer layers (Fig. 16.8b). The formation of this internal shock wave has the effect of reheating strongly the matter in the outer shells. The net result is that, although the material inside the sphere cooled during the adiabatic phase, the gas is reheated by the conversion of a large fraction of the kinetic energy of expansion back into heat. The internal shock wave propagates back through the expanding gas towards the origin and in the process heats up all the ejected gas. Whereas in phase (i), the kinetic energy of expansion of the gas was being communicated to the swept up interstellar gas, in the deceleration phase the kinetic energy is fed back into the ejected gas itself. This heated gas is a strong soft X-ray emitter, entirely consistent with the X-ray images of supernova remnants (Figs. 13.3b and 16.4). As we will discuss in Sect. 16.8, this structure is unstable to Rayleigh–Taylor instabilities which are likely to be responsible for the irregular structures seen in the shells of young supernova remnants such as Cassiopeia A and Tycho’s supernova.

In the case of the young supernova remnant Cassiopeia A, it is estimated that the ratio of the swept up to the ejected gas is about 1 and so it is intermediate between phases (i) and (ii). Tycho’s supernova is already in phase (ii) and is expanding according to the Sedov relation $r \propto t^{2/5}$.

- (iii) As the remnant continues to expand, the temperature in the region behind the shock front, which now contains most of the expanding mass, drops below 10^6 K and cooling by line emission of heavy ions becomes important. The resulting compression to preserve pressure balance at the shock front increases and the shell forms a dense ‘snowplough’. Older remnants such as the Cygnus Loop, which is about 50 000 years old, are associated with stage (iii) and the optical line emission of oxygen and sulphur observed from the filaments of the shell are associated with cooling gas.
- (iv) The expansion eventually becomes subsonic, $v \leq 20 \text{ km s}^{-1}$, and the supernova remnant loses its identity; it is dispersed by random motions in the interstellar medium.

16.8 The adiabatic loss problem and the acceleration of high energy particles

Before we are lulled into a false sense of security, we need to take into account the *adiabatic loss problem* which pervades much of the astrophysics of clouds of relativistic plasma. If high energy particles are accelerated in the initial supernova explosion, they lose all their energy adiabatically during phases (i) and (ii) of the expansion described in Sect. 16.7 before they are released into the interstellar medium.

Suppose the relativistic particles were accelerated very soon after the catastrophic event which gave rise to the supernova. Then, the relativistic gas exerts a pressure on its surroundings and consequently suffers adiabatic losses, as described in Sect. 16.2.3. For

a relativistic particle, the energy decreases with increasing radius as $E \propto r^{-1}$. This also applies to the total relativistic particle energy and thus, if the total energy is W_0 at radius r_0 , when the remnant has expanded to radius r , the internal energy of the relativistic gas would be only $(r_0/r)W_0$. The energy of the particles has gone into the kinetic energy of expansion of the supernova remnant. There are two ways of looking at this result. On the one hand, unless we find a mechanism for regenerating the energy in the relativistic gas, we will not obtain adequate energy to account for the local energy density of cosmic rays from sources such as Cassiopeia A by the time they have expanded to phases (iii) and (iv). Alternatively, we can say that, if r_0 were too small, the energy requirements in relativistic particles would become enormously large in the early stages of the expansion.

There are also consequences for the radio emission of the remnant. If the expansion were purely adiabatic and the magnetic flux density decreased as $B \propto r^{-2}$, as is expected according to magnetic flux freezing (Sect. 11.2), the radio luminosity would decrease rapidly as the remnant expands. From (16.25), the evolution of the electron energy spectrum with radius can be evaluated in an adiabatic expansion. The diffusion-loss equation reduces to

$$\frac{dN(E)}{dt} = \frac{\partial}{\partial E} [b(E)N(E)] , \quad (16.52)$$

where $b(E) = (1/r)(dr/dt)E$. Note that $N(E)$ now refers to all the particles in the remnant rather than the number per unit volume.² Therefore, during the expansion, we can write $N(E) = V\kappa(r)E^{-p}$ since, as we have shown in Sect. 16.3.1, the spectral index does not change under adiabatic losses. Therefore, writing $N(E) = V\kappa(r)E^{-p} = KE^{-p}$,

$$\frac{dN(E)}{dt} = \frac{\partial}{\partial E} \left(\frac{1}{r} \frac{dr}{dt} KE^{-(p-1)} \right) , \quad \frac{dN(E)}{N(E)} = -(p-1) \frac{dr}{r} , \quad \frac{N(E, r)}{N(E, r_0)} = \left(\frac{r}{r_0} \right)^{-(p-1)} , \quad (16.53)$$

that is,

$$\frac{K(r)}{K(r_0)} = \left(\frac{r}{r_0} \right)^{-(p-1)} . \quad (16.54)$$

The synchrotron radio luminosity therefore varies with radius as

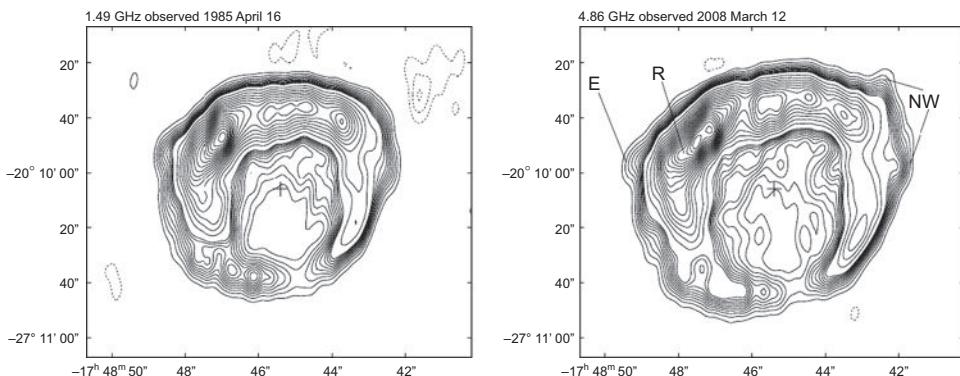
$$I_\nu = A(\alpha)K(R)B^{(1+\alpha)}\nu^{-\alpha} \propto r^{-(p-1)}r^{-2(1+\alpha)} , \quad (16.55)$$

where $p = 2\alpha + 1$ and so

$$I_\nu(r) \propto r^{-2(2\alpha+1)} = r^{-2p} . \quad (16.56)$$

Whilst it is observed that Cassiopeia A is decreasing in radio luminosity at roughly the rate predicted by this expression, it cannot be extrapolated back to the earliest phases of the expansion because young supernovae in external galaxies would then be very intense radio sources which is contrary to observation. Furthermore, the magnetic field strength in Cassiopeia A at radius $r \approx 5 \times 10^{16}$ m is $\approx 10^{-8}$ T which, if compressed to the dimensions of a giant star which has radius $r \approx 10^{11}$ m, would far exceed conceivable field strengths.

² If $N(E)$ were the number per unit volume, we would have to add the term $-N(\nabla \cdot \mathbf{v})$ to the right-hand side (17.32).

**Fig. 16.9**

VLA observations of the young supernova remnant G1.9+0.3 obtained (a) at 1.46 GHz in 1985 and (b) at 4.86 GHz in 2008 (Green *et al.*, 2008). Not only is the remnant expanding, but its flux density is increasing at a rate of about 2% per year.

All these arguments suggest that the relativistic particles and the magnetic field cannot have originated in the initial explosion of the supernova. Rather, the acceleration of the high energy particles and the generation of strong magnetic fields must take place during the expansion phase itself. This process has been observed to take place in the youngest radio supernova remnant known in the Galaxy, G1.9+03, which must have exploded only about 100 years ago, but was not observed because it lies close to the Galactic centre – it would have suffered about 50 magnitudes of extinction in the optical waveband. The expansion of the radio shell was observed between 1985 and 2008, the corresponding mean expansion speed being $v \geq 12\,000 \text{ km s}^{-1}$. In addition, Green and his colleagues (2008) estimate that the radio luminosity of the source is increasing at a rate of about 2% per year. Supernovae exhibit an initial burst of radio emission which normally declines after about a year. Particle acceleration and magnetic field amplification must take place subsequently in order to explain the luminosities of supernova remnants such as Cassopeia A and Tycho's supernova remnant. G1.9+03 shows this process taking place in the youngest known supernova remnant which has expanded into the diffuse interstellar medium.

Fortunately, there exist physical mechanisms by which this can be achieved. The origin of the strong magnetic fields in young supernova remnants has been elegantly explained by Gull (1975) who has studied the dynamics of young supernova shells during phases (i) and (ii) of the scheme outlined in Sect. 16.7. The deceleration of the expanding sphere causes matter to accumulate in a dense shell just inside the contact discontinuity between the sphere and the interstellar medium (see Fig. 16.8b). The force of deceleration acting upon this dense shell is in pressure balance with the much less dense shocked interstellar gas and this results in a *Rayleigh–Taylor instability*. This is the same instability which occurs when a heavy fluid is supported by a light fluid in a gravitational field. In these instabilities, magnetic field amplification is possible as a ‘seed’ magnetic field is stretched and sheared in this turbulent unstable zone. Calculations of the amount of energy which might be channelled into the magnetic field showed that magnetic flux densities up to

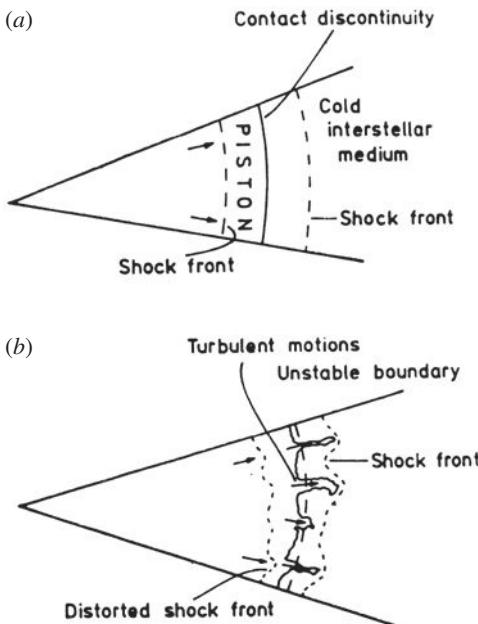


Fig. 16.10

(a) Illustrating the structure of a supersonically expanding sphere of hot gas. The sphere of gas acts as a piston and ahead of it runs a shock front which heats and compresses the cold interstellar gas. As the sphere expands into the interstellar gas, it is decelerated and matter inside the hot sphere begins to pile up at the interface (or contact discontinuity) between the shocked interstellar gas and the interior of the sphere. An internal shock front forms as matter piles up at the interface and the temperature of the gas inside the sphere cools. (b) Because of the piling up of matter at the interface and the deceleration of the shell, there is an effective force of gravity acting on the dense shell of gas which is supported by the much more rarified shocked interstellar gas. This results in a Raleigh–Taylor instability as illustrated (Gull, 1975).

10^{-7} T might be obtained for a supernova such as Cassiopeia A. The instability stretches the field in the radial direction and this is in agreement with the observed magnetic field configuration in young shell-like supernovae. The subsequent evolution of these instabilities may lead to the formation of transient optical and X-ray filaments as observed in young remnants. More recently, other plasma instabilities have been found which enable the magnetic flux density to be amplified to values of order 10^{-7} T, as discussed in more detail in Sect. 17.5.

There remains the problem of accelerating the high energy particles which are present in the shells of supernova remnants. The break-through came in the late 1970s when four independent papers appeared which showed how high energy particles can be accelerated efficiently in strong shock waves. The papers by Krymsky (1977), Axford, Lear and Skadron (1977), Bell (1978) and Blandford and Ostriker (1978) were the seeds of a remarkable development in the understanding of the processes of particle acceleration in strong shock waves. This key topic in high energy astrophysics is discussed in some detail in Chap. 17.

The important point of principle is that it is possible to overcome the problem of adiabatic losses by using the kinetic energy of expansion of the remnant. When the sphere expands into the interstellar medium, the internal energy of the expanding sphere is not lost, but is stored as the kinetic energy of the expanding material. This energy can then be recycled through the action of the strong shock waves and instabilities associated with the decelerating shell of the supernova remnant.

Observations of cosmic rays and sources of non-thermal radiation indicate that the process of acceleration of high energy particles must account for the following features:

- (i) The formation of a power-law energy spectrum for all types of charged particles. The energy spectrum of cosmic rays and the electron energy spectrum of non-thermal sources have the form $dN(E) \propto E^{-x} dE$, where the exponent x typically lies in the range 2–3.
- (ii) The acceleration of cosmic rays to energies $E \sim 10^{20}$ eV.
- (iii) In the process of acceleration, the chemical abundances of the primary cosmic rays should be similar to the cosmic abundances of the elements.

It would be helpful if we could appeal to the physics of laboratory plasmas for some guidance, but the evidence is somewhat ambivalent. On the one hand, if we want to accelerate particles to very high energies, we need to go to a great deal of trouble to ensure that the particles remain within the region of the accelerating field, for example, in machines such as betatrons, synchrotrons, cyclotrons, and so on. Nature does not go to all this trouble to accelerate high energy particles. On the other hand, as soon as we try to build machines to store high temperature plasmas, such as tokamaks, the configurations are usually grossly unstable and, in the instability, particles are accelerated to suprathermal energies.

17.1 General principles of acceleration

The acceleration mechanisms may be classified as *dynamic*, *hydrodynamic* and *electromagnetic*. Often, there is no clear distinction between these because, being charged particles, they are closely tied to magnetic field lines. In some models, the acceleration is purely dynamical, for example, in those cases where acceleration takes place through the collision of particles with clouds. Hydrodynamic models can involve the acceleration of whole layers of plasma to high velocities. The electromagnetic processes include those in which particles are accelerated by electric fields, for example, in neutral sheets, in electromagnetic or plasma waves and in the magnetospheres of neutron stars.

The general expression for the acceleration of a charged particle in electric and magnetic fields is

$$\frac{d}{dt}(\gamma m \mathbf{v}) = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (17.1)$$

In most astrophysical environments, static electric fields cannot be maintained because of the very high electrical conductivity of ionised gases – any electric field is very rapidly short-circuited by the motion of free charges (Sect. 11.1.3). Therefore, electromagnetic mechanisms of acceleration can only be associated either with non-stationary electric fields, for example, strong electromagnetic waves, or with time-varying magnetic fields. In a static magnetic field, no work is done on the particle but, if the magnetic field is time-varying, work can be done by the induced electric field, that is, the electric field \mathbf{E} given by Maxwell's equation $\text{curl } \mathbf{E} = -\partial \mathbf{B}/\partial t$. It has been suggested that phenomena such as the betatron effect might be applicable in some astrophysical environments. For example, the collapse of a cloud of ionised gas with a frozen-in magnetic field could lead to the acceleration of charged particles since they conserve their adiabatic invariants in a time-varying magnetic field (see Sect. 7.2). There is no obvious way in which this mechanism could play a role in the regions where particles are known to be accelerated to high energies, for example, in the shells of supernova remnants.

17.2 The acceleration of particles in solar flares

An important example of particle acceleration by induced electric fields occurs in neutral sheets, the physics of which was discussed in some detail in the context of solar flares in Sect. 11.6. The source of energy of the flare is the reconnection of magnetic field lines and strong electric fields are generated because $\text{curl } \mathbf{B}$ and $\partial \mathbf{B}/\partial t$ are not zero in these regions. The issue is how effective these processes are in creating electric fields which can accelerate charged particles to high energies.

Suppose a DC electric field \mathbf{E} is created in the reconnection region. We need the equation of motion of a charged particle in this field, taking account of the fact the particle collides with other particles in the plasma in stochastic electrostatic encounters (Sect. 11.1.2). The equation of motion of the electron is

$$m_e \frac{d\mathbf{v}}{dt} = -e\mathbf{E} - v_c m_e \mathbf{v}, \quad (17.2)$$

where v_c is the collision frequency for electrostatic collisions given by (11.17) and (11.18). For simplicity, considering the electric field to be parallel to the velocity of the electron,

$$m_e \frac{dv}{dt} = eE - v_c m_e v. \quad (17.3)$$

Adopting (11.17) for the self-collision frequency of the electrons, $v_c = 1/t_c$, we set $m = m_e$, $N = n_e$ and $Z = 1$ and so

$$v_c = \frac{e^4 n_e \ln \Lambda}{2\pi \epsilon_0^2 m_e^2 v^3} = \frac{e^4 n_e \ln \Lambda}{2\pi \epsilon_0^2 m_e^{1/2} (3kT)^{3/2}}. \quad (17.4)$$

In the last equality of (17.4), the mean squared velocity of the electrons is set equal to $3kT/m_e$ for a Maxwellian velocity distribution at temperature T . Because of the v^{-3} dependence of the collision frequency upon the velocity of the electron, once the electron's

velocity becomes greater than a critical velocity v_c , the effect of collisions becomes less and less important and hence the electrons are accelerated without any impediment under the influence of the electric field. This process is known as *electron runaway*. For velocities less than the critical velocity, the particles are not accelerated. The critical velocity is found by setting the right-hand side of (17.3) equal to zero:

$$v_c = \left(\frac{e^3 n_e \ln \Lambda}{2\pi \epsilon_0^2 m_e E} \right)^{1/2}. \quad (17.5)$$

Correspondingly, for a thermal plasma, there is a critical electric field associated with this process which is known as the *Dreicer field* E_D ,

$$E_D = \frac{e^3 n_e \ln \Lambda}{6\pi \epsilon_0^2 k T}, \quad (17.6)$$

at which electron runaway occurs for all the electrons. Writing this relation in terms of the Debye length of the plasma $\lambda_D = (\epsilon_0 k T / n e^2)^{1/2}$,

$$E_D = \frac{e \ln \Lambda}{6\pi \epsilon_0 \lambda_D^2}. \quad (17.7)$$

This result is remarkably similar to that quoted by Heyvaerts (1981). Inserting values of the numerical constants,

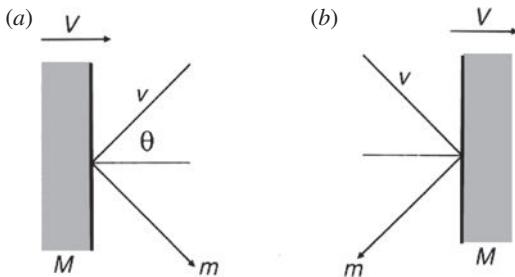
$$E_D = 2 \times 10^{-13} \frac{n_e \ln \Lambda}{T} \text{ V m}^{-1}, \quad (17.8)$$

where n_e is measured in electrons m^{-3} and T in kelvin.

Adopting the values of these quantities quoted in Sect. 11.6 for solar flares, $n_e = 10^{16} \text{ m}^{-3}$ and $T = 2 \times 10^6 \text{ K}$, the Dreicer field is $E_D = 2 \times 10^{-2} \text{ V m}^{-1}$. The strength of the electric field in the neutral sheet may be estimated by writing $E \sim v_R B \approx 0.02 \text{ V m}^{-1}$. Thus, this process may well be important in particle acceleration in neutral sheets in solar flares.

There are, however, a number of problems. First of all, there is a limit to the flow of charge because the beam of particles creates a magnetic field which cannot exceed the initial magnetic flux density. The result is an induced return current which tends to neutralise the effect of the electric field. Second, the beam of accelerated particles is subject to a variety of streaming instabilities. In particular, the beam may excite ion sound waves which enhance the resistivity of the plasma. If, however, only a small fraction of the electrons have velocities exceeding the critical velocity, these may be accelerated without exciting plasma instabilities. This process may also be important as part of the stochastic acceleration of particles. Heyvaerts (1981) gives a comprehensive description of the plasma instabilities which are likely to be associated with electron runaway.

As discussed in Sect. 13.9, ‘gaps’ are expected in the structure of the electromagnetic fields in pulsar magnetospheres. These are regions in which induced electric fields are maintained and are potential sites for the acceleration of charged particles. The acceleration in the polar cap gaps is likely to be limited by pair-production processes, but the outer gaps, which have scale of roughly that of the light cylinder, offer the possibility of particle acceleration to high energies (Sect. 17.6).

**Fig. 17.1**

Illustrating the collision between a particle of mass m and a cloud of mass M : (a) a head-on collision; (b) a following collision. The probabilities of head-on and following collisions are proportional to the relative velocities of approach of the particle and the cloud, namely, $v + V \cos \theta$ for (a) and $v - V \cos \theta$ for (b). Since $v \approx c$, the probabilities are proportional to $1 + (V/c) \cos \theta$ where $0 < \theta < \pi$.

17.3 Fermi acceleration – original version

The Fermi mechanism was first proposed in 1949 as a stochastic means by which particles colliding with clouds in the interstellar medium could be accelerated to high energies (Fermi, 1949). We first consider Fermi's original version of the theory, the problems it encounters and how it can be reincarnated in a modern guise. The analysis contains a number of features which are important for particle acceleration in general. In Sect. 17.4, the modern version of first-order Fermi acceleration is described.

In Fermi's original picture, charged particles are reflected from 'magnetic mirrors' associated with irregularities in the Galactic magnetic field. The mirrors are assumed to move randomly with typical velocity V and Fermi showed that the particles gain energy stochastically in these reflections. If the particles remain within the acceleration region for some characteristic time τ_{esc} , a power-law distribution of particle energies is obtained.

Let us repeat Fermi's calculation in which the collision between the particle and the mirror takes place such that the angle between the initial direction of the particle and the normal to the surface of the mirror is θ (Fig. 17.1a). We carry out a relativistic analysis of the change in energy of the particle in a single collision.

The mirror is taken to be infinitely massive and so its velocity is unchanged in the collision. The centre of momentum frame of reference is therefore that of the cloud moving at velocity V . The energy of the particle in this frame is

$$E' = \gamma_V(E + V p \cos \theta), \quad \text{where} \quad \gamma_V = \left(1 - \frac{V^2}{c^2}\right)^{-1/2}. \quad (17.9)$$

The x -component of the relativistic three-momentum in the centre of momentum frame is

$$p'_x = p' \cos \theta' = \gamma_V \left(p \cos \theta + \frac{V E}{c^2} \right). \quad (17.10)$$

In the collision, the particle's energy is conserved, $E'_{\text{before}} = E'_{\text{after}}$, and its momentum in the x -direction is reversed, $p'_x \rightarrow -p'_x$. Therefore, transforming back to the observer's frame,

$$E'' = \gamma_V(E' + Vp'_x). \quad (17.11)$$

Substituting (17.9) and (17.10) into (17.11) and recalling that $p_x/E = v \cos \theta/c^2$, the change in energy of the particle is

$$E'' = \gamma_V^2 E \left[1 + \frac{2Vv \cos \theta}{c^2} + \left(\frac{V}{c} \right)^2 \right]. \quad (17.12)$$

Expanding to second order in V/c , we find

$$\Delta E = E'' - E = E \left[\frac{2Vv \cos \theta}{c^2} + 2 \left(\frac{V}{c} \right)^2 \right]. \quad (17.13)$$

We now average over the angle θ . Because of scattering by hydromagnetic waves or irregularities in the magnetic field, it is likely that the particle is randomly scattered in pitch angle between encounters with the clouds and we can therefore work out the mean increase in energy by averaging over a random distribution of angles θ in (17.13). There is a slightly greater probability of head-on encounters as opposed to following collisions, as illustrated in Fig. 17.1. The probability of encounters taking place at an angle of incidence θ is given by exactly the same reasoning which led to the rate of arrival of photons at an angle θ in our analysis of inverse Compton scattering (see the discussion of Sect. 9.3). The only difference is that the particles move at a velocity v rather than c . For simplicity, let us consider the case of a relativistic particle with $v \approx c$ in which case the probability of a collision at angle θ is proportional to $\gamma_V [1 + (V/c) \cos \theta]$. Recalling that the probability of the angle lying in the angular range θ to $\theta + d\theta$ is proportional to $\sin \theta d\theta$, on averaging over all angles in the range 0 to π , the first term in (17.13) in the limit $v \rightarrow c$ becomes

$$\left\langle \frac{\Delta E}{E} \right\rangle = \left(\frac{2V}{c} \right) \frac{\int_{-1}^1 x [1 + (V/c)x] dx}{\int_{-1}^1 [1 + (V/c)x] dx} = \frac{2}{3} \left(\frac{V}{c} \right)^2, \quad (17.14)$$

where $x = \cos \theta$. Including the last term in (17.13), the average energy gain per collision is

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3} \left(\frac{V}{c} \right)^2. \quad (17.15)$$

This illustrates the famous result derived by Fermi that the average increase in energy is only second-order in V/c . This result leads to an exponential increase in the energy of the particle since the same fractional increase occurs per collision. If the mean free path between clouds along a field line is L , the time between collisions is $L/(c \cos \phi)$ where ϕ is the pitch angle of the particle with respect to the magnetic field direction. Averaging $\cos \phi$ over the pitch angle ϕ , the average time between collisions is $2L/c$. Therefore, the average

rate of energy increase is

$$\frac{dE}{dt} = \frac{4}{3} \left(\frac{V^2}{cL} \right) E = \alpha E . \quad (17.16)$$

The particle is assumed to remain within the accelerating region for a time τ_{esc} . The resulting spectrum can be found from the diffusion-loss equation (7.41),

$$\frac{dN}{dt} = D\nabla^2 N + \frac{\partial}{\partial E} [b(E)N(E)] - \frac{N}{\tau_{\text{esc}}} + Q(E) . \quad (17.17)$$

We are interested in the steady-state solution in the absence of diffusion and so $dN/dt = 0$ and $D\nabla^2 N = 0$. It is also assumed that there are no sources, $Q(E) = 0$. The energy loss term $b(E) = -dE/dt$ now becomes an energy gain term, $b(E) = -\alpha E$. Therefore, (17.17) is reduced to

$$-\frac{d}{dE} [\alpha E N(E)] - \frac{N(E)}{\tau_{\text{esc}}} = 0 . \quad (17.18)$$

Differentiating and rearranging this equation,

$$\frac{dN(E)}{dE} = - \left(1 + \frac{1}{\alpha \tau_{\text{esc}}} \right) \frac{N(E)}{E} , \quad (17.19)$$

and so

$$N(E) = \text{constant} \times E^{-x} , \quad (17.20)$$

where $x = 1 + (\alpha \tau_{\text{esc}})^{-1}$. The Fermi acceleration mechanism results in a power-law energy spectrum.

In Fermi's paper of 1949, it was assumed that collisions with interstellar clouds were the main source of energy for the particles. There were, however, problems with this picture.

- (i) The random velocities of interstellar clouds in the Galaxy are very small in comparison with the velocity of light, $V/c \leq 10^{-4}$. Furthermore, the mean free path for the scattering of cosmic rays in the interstellar medium is of the order of 0.1 pc (Sect. 15.7) and so the number of collisions would amount to about a few per year, resulting in a very slow gain of energy by the particles. We might do rather better if attention was restricted to regions where there is small scale turbulence, for example, in the shells of young supernova remnants where there is certainly a great deal of small scale structure present and the velocities are very much greater than in the general interstellar medium.
- (ii) Ionisation losses hamper the acceleration of particles from low energies. The form of the ionisation loss rate as a function of kinetic energy is compared qualitatively with a uniform acceleration rate in Fig. 17.2. If the acceleration mechanism is to be effective, the particles must either be injected into the acceleration region with energies greater than that corresponding to the maximum energy loss rate or else the initial acceleration process must be sufficiently rapid to overcome the ionisation energy losses. This is known as the *injection problem* and is present in all acceleration mechanisms.
- (iii) There is nothing in the theory which tells us why the exponent x of the energy spectrum should be roughly 2.5. It would be remarkable if the mechanism of acceleration in

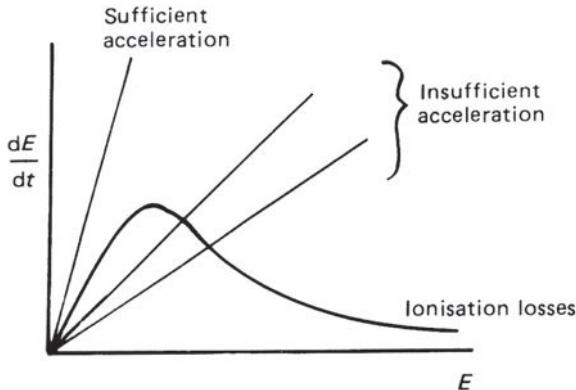


Fig. 17.2 Comparison of the acceleration rate and energy loss rate due to ionisation losses for a high energy particle.

very diverse types of source were such that the product of the characteristic escape time τ_{esc} and the rate of energy gain, as represented by the constant α , conspired to give the same value of x .

This version of second-order Fermi acceleration **disguises a key aspect of the acceleration process**. The particle's energy changes all the time stochastically and so, if particles were injected with a single energy, the energy distribution would be broadened by random encounters with interstellar clouds. On average, the root mean square change of energy of the particle is $O(V/c)$, whereas the systematic increase in energy is only $O(V/c)^2$. **In the full calculation, we have to take account of the statistical nature of the acceleration process as well as the average systematic increase in energy.**

The problem is exactly the same as that discussed in the context of Comptonisation and the Sunyaev–Zeldovich effect in Sect. 9.5 and illustrated in Fig. 9.12. That figure shows the broadening of the photon spectrum which necessarily accompanies the second-order mean increase in energy. The full treatment which takes account of the stochastic nature of the acceleration process and the spreading of the energy spectrum by scattering starts from the Fokker–Planck equation for the diffusion of the particles in momentum space. This approach is described in the review by Blandford and Eichler (1987) who show that, for an isotropic distribution of scatterers, the differential energy spectral index is

$$x = \frac{3}{2} \left(1 + \frac{4cL}{3\langle V^2 \rangle \tau_{\text{esc}}} \right)^{1/2} - \frac{1}{2}. \quad (17.21)$$

This result is slightly different from that found in the simple Fermi argument, except in the limit of small accelerations.

The Fokker–Planck equation can be reduced to a diffusion-loss equation with an additional term which describes the diffusion of the particles in momentum space, or energy space in the case of **isotropic scatterers**. From (7.42),

$$\frac{dN}{dt} = D\nabla^2 N + \frac{\partial}{\partial E} [b(E)N(E)] - \frac{N}{\tau_{\text{esc}}} + Q(E) + \frac{1}{2} \frac{\partial^2}{\partial E^2} [d(E)N(E)], \quad (17.22)$$

where $d(E)$ is the mean square change in energy per unit time,

$$d(E) = \frac{d}{dt} \langle (\Delta E)^2 \rangle . \quad (17.23)$$

As before, an expression for the average value of $(\Delta E)^2$ can be found from (17.13). To second order in V/c , the only term which survives is $(\Delta E)^2 = 4E^2(V/c)^2 \cos^2 \theta$ and, averaging over angle as before,

$$\langle (\Delta E)^2 \rangle = 4E^2 \left(\frac{V}{c} \right)^2 \frac{1}{2} \int_0^\pi \cos^2 \theta \sin \theta d\theta = \frac{4}{3} E^2 \left(\frac{V}{c} \right)^2 . \quad (17.24)$$

Thus, there is a very close relation between the mean square energy change and the average increase in energy per collision. Comparing (17.16) and (17.23), $d(E) = -Eb(E)/2 = \alpha E^2/2$, recalling that $b(E)$ is defined to be the rate of loss of energy. We seek steady-state solutions of (17.22), including the stochastic acceleration term, that is,

$$\frac{\partial}{\partial E} [b(E)N(E)] - \frac{N}{\tau_{\text{esc}}} + \frac{1}{2} \frac{\partial^2}{\partial E^2} [d(E)N(E)] = 0 . \quad (17.25)$$

Substituting for $b(E)$ and $d(E)$ and seeking solutions of power-law form, $N \propto E^{-x}$, the partial differential equation reduces to the quadratic equation

$$x^2 + x - \left(2 + \frac{4}{\alpha \tau_{\text{esc}}} \right) = 0 , \quad (17.26)$$

which has solution

$$x = \frac{3}{2} \left(1 + \frac{16}{9\alpha\tau_{\text{esc}}} \right)^{1/2} - \frac{1}{2} , \quad (17.27)$$

exactly the result derived by Blandford and Eichler since $\alpha = (4/3)(V^2/cL)$. In the second-order acceleration mechanism, the values of α and τ_{esc} to be used in the formulae for $b(E)$ and $d(E)$ are model dependent.

In the modern version of second-order Fermi acceleration, the particles interact with various types of plasma wave and gain energy by being scattered stochastically by them. This process is similar to that described in Sect. 7.4 concerning the interaction of high energy particles with waves or irregularities in the interstellar magnetic field.

17.4 Diffusive shock acceleration in strong shock waves

The acceleration mechanism which has dominated much astrophysical thinking since the late 1970s is associated with particle acceleration in strong shock waves, often referred to as *diffusive shock acceleration*. The key feature of this process is that the acceleration is first order in the shock velocity and automatically results in a power-law spectrum with energy spectral index $x \approx 2$.

Let us first write the essence of the original Fermi mechanism in a slightly different way. We define the constants β and P as follows: $E = \beta E_0$ is the average energy of the particle

after one collision and P is the probability that the particle remains within the accelerating region after one collision. Then, after k collisions, there are $N = N_0 P^k$ particles with energies $E = E_0 \beta^k$. Eliminating k between these quantities,

$$\frac{\ln(N/N_0)}{\ln(E/E_0)} = \frac{\ln P}{\ln \beta}, \quad (17.28)$$

and hence

$$\frac{N}{N_0} = \left(\frac{E}{E_0} \right)^{\ln P / \ln \beta}. \quad (17.29)$$

In fact, this value of N is $N(\geq E)$ since this is the number which reach energy E and some fraction of them continue to be accelerated to higher energies. Therefore

$$N(E) dE = \text{constant} \times E^{-1+(\ln P / \ln \beta)} dE. \quad (17.30)$$

In this reformulation, we have again recovered a power-law energy distribution. The equivalence with the derivation of Fermi acceleration in Sect. 17.3 follows from equation (17.11), namely, $(\ln P / \ln \beta) \equiv -(\alpha \tau_{\text{esc}})^{-1}$. Evidently, β is related to α and P to τ_{esc} .

In the original version of the Fermi acceleration described in Sect. 17.3, α is proportional to $(V/c)^2$ and is referred to as *second-order Fermi acceleration*. We would do much better if there were only head-on collisions. Inspection of (17.13) shows that the fractional energy increase is $\Delta E/E \propto 2V/c$, that is, first order in V/c and, appropriately, this is referred to as *first-order Fermi acceleration*.

The diffusive shock acceleration process is first-order Fermi acceleration in the presence of strong shock waves and was discovered independently by a number of workers in the late 1970s. The papers by Axford, Leer and Skadron (1977), Krymsky (1977), Bell (1978) and Blandford and Ostriker (1978) stimulated an enormous amount of interest in this process in many diverse astrophysical environments. There are two different ways of tackling the problem, the first starting from the Fokker–Planck equation for the evolution of the momentum distribution of high energy particles in the vicinity of a strong shock, and the second, a more physical approach in which the behaviour of individual particles is followed. Bell (1978) adopts the latter approach which we will follow here.

The model involves a strong shock propagating through a diffuse medium, for example, the shock waves which propagate through the interstellar medium ahead of the supersonic shells of supernova remnants (Fig. 16.8). A flux of high energy particles is assumed to be present both in front of and behind the shock front. The particles are assumed to be propagating at speeds close to that of light and so the velocity of the shock is very much less than those of the high energy particles. The high energy particles scarcely notice the shock at all since its thickness is normally very much smaller than the gyroradius of the high energy particle. Because of scattering by streaming instabilities or turbulent motions on either side of the shock wave, when the particles pass though the shock in either direction, they are scattered so that their velocity distribution rapidly becomes isotropic in the frame of reference of the moving fluid on either side of the shock.

The physics of strong shock waves was analysed in Sect. 11.3. The material ejected in supernova explosions attains velocities of order 10^4 km s^{-1} , which is very much greater

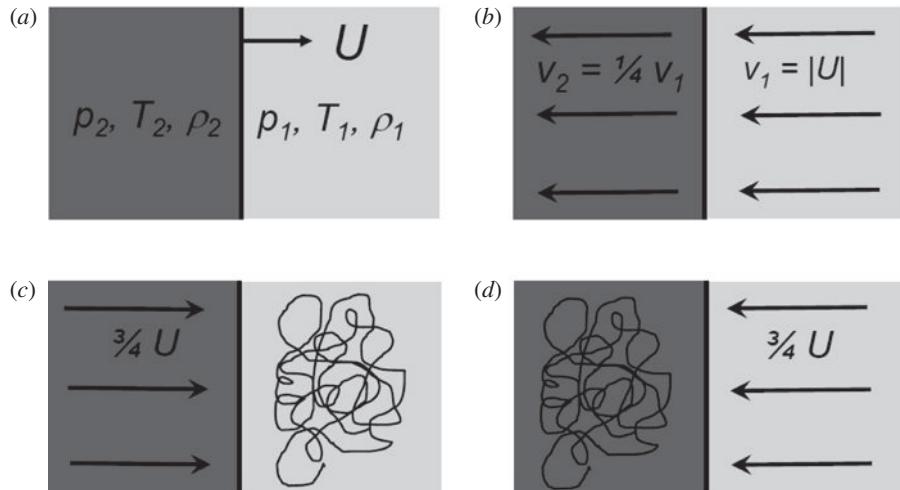


Fig. 17.3

The dynamics of high energy particles in the vicinity of a strong shock wave. (a) A strong shock wave propagating at a supersonic velocity U through stationary interstellar gas with density ρ_1 , pressure p_1 and temperature T_1 . The density, pressure and temperature behind the shock are ρ_2 , p_2 and T_2 , respectively. The relations between the variables on either side of the shock front are given by the relations (11.72)–(11.74). (b) The flow of interstellar gas in the vicinity of the shock front in the reference frame in which the shock front is at rest. In this frame of reference, the ratio of the upstream to the downstream velocity is $v_1/v_2 = (\gamma + 1)/(\gamma - 1)$. For a fully ionised plasma, $\gamma = 5/3$ and the ratio of these velocities is $v_1/v_2 = 4$ as shown in the figure. (c) The flow of gas as observed in the frame of reference in which the upstream gas is stationary and the velocity distribution of the high energy particles is isotropic. (d) The flow of gas as observed in the frame of reference in which the downstream gas is stationary and the velocity distribution of high energy particles is isotropic.

than the sound and Alfvén speeds of the interstellar medium, which are at most about 10 km s^{-1} . A strong shock wave travels at a highly supersonic velocity $U \gg c_s$, where c_s is the sound speed in the ambient medium (Fig. 17.3a), the Mach number M being $U/c_s \gg 1$. It is often convenient to transform into the frame of reference in which the shock front is at rest and then the upstream gas flows into the shock front at velocity $v_1 = U$ and leaves the shock with a downstream velocity v_2 (Fig. 17.3b). The equation of continuity requires mass to be conserved through the shock and so

$$\rho_1 v_1 = \rho_2 U = \rho_2 v_2 . \quad (17.31)$$

In the case of a strong shock, $\rho_2/\rho_1 = (\gamma + 1)/(\gamma - 1)$ where γ is the ratio of specific heat capacities of the gas (Sect. 11.3.1). Taking $\gamma = 5/3$ for a monatomic or fully ionised gas, $\rho_2/\rho_1 = 4$ and so $v_2 = (1/4)v_1$ (Fig. 17.3b).

Now consider high energy particles ahead of the shock. Scattering ensures that the particle distribution is isotropic in the frame of reference in which the gas is at rest. It is instructive to draw diagrams illustrating the situation so far as typical high energy particles upstream and downstream of the shock are concerned. The shock advances through the medium at velocity U but the gas behind the shock travels at a velocity $(3/4)U$ relative to the upstream gas (Fig. 17.3c). When a high energy particle crosses the shock front, it obtains

a small increase in energy, of order $\Delta E/E \sim U/c$, as we will show below. The particles are then scattered in the region behind the shock front so that their velocity distributions become isotropic with respect to that flow.

Now consider the opposite process of the particle diffusing from behind the shock to the upstream region in front of the shock (Fig. 17.3d). Now the velocity distribution of the particles is isotropic behind the shock and, when they cross the shock front, they encounter gas moving towards the shock front, again with the same velocity $(3/4)U$. In other words, the particle undergoes exactly the same process of receiving a small increase in energy ΔE on crossing the shock from the downstream to the upstream flow as it did in travelling from upstream to downstream. This is the clever aspect of the acceleration mechanism. Every time the particle crosses the shock front it receives an increase of energy – there are never crossings in which the particles lose energy – and the increment in energy is the same going in both directions. Thus, unlike the original Fermi mechanism in which there are both head-on and following collisions, in the case of the shock front, the collisions are always head-on and energy is transferred to the particles.

By simple arguments, due to Bell (1978), both β and P can be determined quantitatively for a complete acceleration cycle. First, we evaluate the average increase in energy of the particle on crossing from the upstream to the downstream sides of the shock. The gas on the downstream side approaches the particle at a velocity $V = (3/4)U$ and so, performing a Lorentz transformation, the particle's energy when it passes into the downstream region is

$$E' = \gamma_V(E + p_x V), \quad (17.32)$$

where the x -coordinate is taken to be perpendicular to the shock. The shock is assumed to be non-relativistic, $V \ll c$, $\gamma_V = 1$, but the particles are relativistic and so $E = pc$, $p_x = (E/c) \cos \theta$. Therefore,

$$\Delta E = pV \cos \theta \quad \frac{\Delta E}{E} = \frac{V}{c} \cos \theta. \quad (17.33)$$

The probability that the particles which cross the shock arrive within the angles θ to $\theta + d\theta$ is proportional to $\sin \theta d\theta$ and the rate at which they approach the shock front is proportional to the x -component of their velocities, $c \cos \theta$. Therefore the probability of the particle crossing the shock is proportional to $\sin \theta \cos \theta d\theta$. Normalising so that the integral of the probability distribution over all the particles approaching the shock is equal to 1, that is, those with θ in the range 0 to $\pi/2$,

$$p(\theta) = 2 \sin \theta \cos \theta d\theta. \quad (17.34)$$

Therefore, the average gain in energy on crossing the shock is

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{V}{c} \int_0^{\pi/2} 2 \cos^2 \theta \sin \theta d\theta = \frac{2}{3} \frac{V}{c}. \quad (17.35)$$

The particle's velocity vector is randomised without energy loss by scattering in the downstream region and it then recrosses the shock, as illustrated in Fig. 17.3d, when it gains another fractional increase in energy $(2/3)(V/c)$. Therefore, in making one round trip across

the shock and back again, the fractional energy increase is, on average,

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{4}{3} \frac{V}{c} . \quad (17.36)$$

Consequently,

$$\beta = \frac{E}{E_0} = 1 + \frac{4V}{3c} , \quad (17.37)$$

in one round trip.

To work out the escape probability P , we use the argument due to Bell (1978). According to classical kinetic theory, the number of particles crossing the shock is $\frac{1}{4}Nc$ where N is the number density of particles. This is the average number of particles crossing the shock in either direction since, as noted above, the particles scarcely notice the shock. Downstream, however, the particles are swept away, or ‘adverted’, from the shock because the particles are isotropic in that frame. Referring to Fig. 17.3b, it can be seen that particles are removed from the region of the shock at a rate $NV = \frac{1}{4}NU$. Thus, the fraction of the particles lost per unit time is $\frac{1}{4}NU/\frac{1}{4}Nc = U/c$. Since the shock is assumed to be non-relativistic, only a very small fraction of the particles is lost per cycle. Thus, $P = 1 - (U/c)$. Therefore,

$$\ln P = \ln \left(1 - \frac{U}{c} \right) = -\frac{U}{c} \quad \text{and} \quad \ln \beta = \ln \left(1 + \frac{4V}{3c} \right) = \frac{4V}{3c} = \frac{U}{c} . \quad (17.38)$$

Inserting $\ln \beta$ and $\ln P$ into (17.32),

$$\frac{\ln P}{\ln \beta} = -1 . \quad (17.39)$$

Therefore, the differential energy spectrum of the high energy particles is

$$N(E) dE \propto E^{-2} dE . \quad (17.40)$$

This is the result we have been seeking. It may be objected that we have obtained a value of 2 rather than 2.5 for the exponent of the differential energy spectrum and that problem cannot be neglected. The mechanism excited a great deal of interest however because it provided an excellent physical reason why power-law energy spectra with a unique spectral index should be found in diverse astrophysical environments. The only requirements are the presence of strong shock waves and that the velocity vectors of the high energy particles are randomised on either side of the shock. There are undoubtedly strong shocks in sources of high energy particles, for example, in supernova remnants, active galactic nuclei and the extended components of extragalactic radio sources.

An important feature of the model is that the particles are scattered in both the up-stream and down-stream regions. In Bell’s original proposal, the particles which recross the shock from downstream to upstream result in bulk streaming of the relativistic particles through the unperturbed interstellar medium and, consequently, the particles are scattered by the generation of Alfvén and hydromagnetic waves which grow to large amplitude, as described in Sect. 7.4. As the growing waves are advected into the region behind the shock, they create turbulent motions which can scatter the particles as described in Sect. 7.3. It is therefore expected that the high energy particles will be confined within some characteristic

distance in front of the shock. The number of high energy particles is expected to decrease exponentially ahead of the shock wave. In consequence, the acceleration process can be modelled by the ‘box-model’ of Drury and his colleague in which the acceleration occurs within one diffusion length on either side of the shock front (Drury *et al.*, 1999).

There is an upper limit to the energy to which particles can be accelerated by this mechanism. Lagage and Cesarsky (1983) describe in detail the processes which can limit the acceleration process. The basic problem is that, although the first-order acceleration mechanism is a distinct improvement upon the original Fermi mechanism, it is still not a rapid process. The particles have to diffuse back and forth across the shock wave many times and, in the case of the shells of supernova remnants, their energies increase by about one part in 100 at each crossing. The supernova blast wave is decelerated once the remnant has swept up roughly its own mass of interstellar gas and then enters the Sedov phase of evolution (Sect. 16.7). Although the acceleration mechanism continues throughout the life of a supernova remnant until it dissolves into the interstellar medium after about 10^5 – 10^6 years, most of the particle acceleration occurs during the undecelerated blast wave phase which lasts less than about 10^3 years, as illustrated by the computations of Lagage and Cesarsky. Assuming the magnetic flux density of the interstellar magnetic field to be 10^{-10} T, they estimated that the maximum upper limit to the energy of particles is about 10^5 GeV nucleon $^{-1}$ = 10^{14} eV nucleon $^{-1}$, their more realistic estimates corresponding to about an order of magnitude less than this value. The figure of 10^{14} eV corresponds roughly to a radius of gyration of a proton of about 1 pc if the magnetic flux density is $B = 10^{-10}$ T (see Table 15.6). The energy spectrum of cosmic rays extends well beyond this upper limit to 10^{20} eV.

The origin of this problem can be appreciated from an argument due to Syrovatskii for the maximum attainable energy of a particle in a magnetic field of flux density B and scale L . The first Maxwell equation, or Faraday’s law, can be used to estimate the induced electric field in the region,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (17.41)$$

If U is the speed of the shock, this equation can be rewritten to order of magnitude as

$$\frac{E}{L} \sim \frac{B}{L/U}; \quad E \sim BU. \quad (17.42)$$

The energy of a particle of charge ze accelerated by the induced electric field is then

$$E_{\max} = \int zeE dx = zeBUL. \quad (17.43)$$

Hence, to order of magnitude, the energy of the accelerated particle is roughly $eBUL$ per nucleon. For the parameters adopted for young supernova remnants above, $B = 10^{-10}$ T, $U = 10^4$ km s $^{-1}$ and $t \approx 10^3$ years, we recover roughly the upper limit derived by Lagage and Cesarsky, $E_{\max} \sim 10^{14}$ eV.

There is one further important aspect of this acceleration process. The particles are accelerated where they are needed and this enables the adiabatic loss problem (Sect. 16.8) to be overcome. In supernova remnants such as Cassiopeia A and Tycho’s supernova, the

particles are accelerated *in situ* and the energy for accelerating them is extracted from the kinetic energy of the expanding supernova shell. As suggested in Sect. 16.8, there are therefore good reasons why both the magnetic field energy *and* the particle energy in shell-like supernova remnants are derived from the kinetic energy of expansion of the supernova.

17.5 Beyond the standard model

The simple physical model of diffusive particle acceleration described in Sect. 17.4 contains the essence of the physics of the acceleration process. That analysis has an almost ‘thermodynamic’ flavour to it in that it omits all mention of the facts that the strong shock waves are collisionless and that there are magnetic fields in the plasma which mediate the effective viscosity which transports energy and momentum through the shock wave. A full treatment begins with a Fokker–Planck type equation for the evolution of the energy spectrum of the particles, the approach taken by Axford, Leer and Skadron (1977), Krymsky (1977) and Blandford and Ostriker (1978).

Since the pioneering researches of the 1970s, a huge amount of effort has been devoted to the detailed physics of diffusive particle acceleration. The motivation has been the need to understand how the prediction of the standard model, $N(E) \propto E^{-2}$, is modified when the effects of the magnetic field are included in the model, as well as considering the full range of weak, strong and relativistic shocks. The shocks may be parallel, perpendicular or inclined to the magnetic field direction, all resulting in slightly different predictions for the particle energy spectrum. For our present purposes, it is sufficient to note that, for non-relativistic shocks, the same calculation performed in Sect. 17.4 can be repeated for the case $r = v_1/v_2 < 4$ and then the following relations are found for the spectrum of the particles:

$$r = v_1/v_2 = \rho_2/\rho_1 ; \quad q = 3r/(r - 1) ; \quad N(E) \propto E^{-q+2} ; \quad f(p) \propto p^{-q} , \quad (17.44)$$

where the last expression describes an isotropic particle distribution in momentum space. Steeper spectra are expected for smaller velocity or density ratios.

The mechanism has been the subject of detailed numerical simulation. These studies have shown that diffusive particle acceleration can be remarkably effective, the conversion of kinetic energy of the shock into high energy particles amounting to about 50%. As a result, the influence of the energy density of the accelerated particles upon the structure of the shock and the stability of the flows cannot be neglected. The result also means that the process of acceleration in intrinsically nonlinear. Let us first review observational evidence and theoretical developments concerning the strengths of the magnetic fields present in the shock fronts.

17.5.1 Magnetic fields in supernova shock fronts

In a strong shock wave in an fully ionised gas, the increase in density behind a strong shock corresponds to a factor of $(\gamma + 1)/(\gamma - 1) = 4$ greater than the ambient gas density.

Consequently, if magnetic flux freezing is applied to the ionised gas on either side of the shock wave, an increase by only a factor of 4 would be obtained over the interstellar magnetic field. This is a contributory factor to the problem of accelerating particles to energies greater than about 10^{14} eV discussed in Sect. 17.4 (Lagage and Cesarsky, 1983). Two results have suggested that the magnetic fields in the strong shocks in supernova remnants undergo a much greater flux density increase.

The first piece of evidence is derived from observations of the X-ray emission from the sharp boundaries of young supernova remnants such as those observed in the images of Cassiopeia A (Fig. 13.3a) and Tycho's supernova remnant (Fig. 16.4a). These narrow shells are interpreted as the shock waves about the supernova remnant. Vink and Laming (2003) interpreted the narrow non-thermal X-ray emission bounding Cassiopeia A as the synchrotron emission of very high energy electrons accelerated in the shock itself. These very high energy electrons suffer such strong synchrotron radiation losses that they lose all their energy in the vicinity of the shock. The transient nature of some of these very narrow filaments is compelling evidence that the process of diffuse shock acceleration of ultra-relativistic electrons is indeed taking place in these shock waves. Völk and his colleagues showed that the X-ray images of Cassiopeia A, Tycho and four other young supernova remnants all possess very narrow X-ray filaments which define the location of the shock wave about the supernova remnant and that these are all consistent with a similar model (Völk *et al.*, 2005). In all cases, very strong magnetic fields were inferred to be present in the shock wave, typical values being about $(3-10) \times 10^{-8}$ T. These are very much greater values than could be accounted for by compressing the interstellar magnetic field in a strong shock.

The second piece of evidence is that physical mechanisms have been discovered by which such strong magnetic fields can be generated in strong shock waves. Bell and Lucek (2000; 2001) carried out numerical calculations in which they found that the magnetic flux densities behind strong shocks which were accelerating particles to high energies could be very considerably enhanced over the standard compressed values as a result of instabilities which tangled and amplified the magnetic field. In subsequent papers, Bell (2004; 2005) demonstrated that these simulations lead to a non-resonant instability driven by the cosmic ray current itself. In a simple picture, cosmic ray currents interact with loops of magnetic field resulting in a $\mathbf{j} \times \mathbf{B}$ force which expands the loops. Since the mass density of particles on the loop decreases as the loop expands, the acceleration increases resulting in an instability which produces strong loops of magnetic field. He showed that the energy density in the magnetic field could approach the energy density in the high energy particles. The pressure of the accelerated particles can approach the ram pressure of the swept-up gas and so the amplification of the magnetic flux density is of the order of the Alfvén Mach number, of the order 100–1000.

This is, however, only one of a number of mechanisms by which strong fields can be created. In the upstream region, other mechanisms include the streaming instability proposed by Bell (1978), the Drury instability (Drury and Falde, 1986), the firehose instability (Quest and Shapiro, 1996) and small-scale dynamos (Beresnyak *et al.*, 2009). Furthermore, instabilities can also operate downstream, for example, the nonlinear Richtmeyer–Meshkov instability (Giacalone and Jokipii, 2007) and charge exchange current-driven instabilities

(Ohira *et al.*, 2009). The inference is that the magnetic flux density can be amplified to very high values by these instabilities, accounting for the strong fields now known to be present in supernova shock waves. These arguments suggest that the values of B which should be used in (17.43) could be two or three orders of magnitude greater than the values used by Lagage and Cesarsky. In consequence, there would be no problem in accelerating particles up to the knee at $10^{15} - 10^{16}$ eV observed in the cosmic ray spectrum.

17.5.2 Nonlinear diffuse shock acceleration

Because of the strong nonlinearity of the acceleration process, numerical methods have to be employed to understand the process of acceleration in more detail. The enhancements needed to construct more realistic models of the process of shock acceleration are illustrated schematically in Fig. 17.4. Going round the diagram,

- ① The basic picture of diffusive shock acceleration is shown, taken from the single particle simulations of Ballard and Heavens (1992). In fact, this diagram is for a relativistic shock, but the principles of diffusive shock acceleration are the same.
- ② The high energy particles diffuse from downstream to upstream.
- ③ There must be some source of injection of seed particles and in this diagram it is assumed that these are thermal leakage particles.
- ④ The streaming of the particles gives rise to the generation of resonant Alfvén waves, non-resonant waves and turbulence.
- ⑤ The results of these instabilities are: (i) to enhance greatly the strength of the magnetic field, (ii) to accelerate particles to the ‘knee’ in the cosmic ray energy spectrum, (iii) to cause further scattering of particles and the dissipation of waves, and (iv) the diffusion of particles in physical and momentum space.

Descriptions of various approaches to the numerical modelling of the acceleration process and of the many diverse aspects of the physics involved are given by Kang and Jones (2006). The important difference from the simple model of Sect. 17.4 is the influence of the diffusion of the high energy particles back upstream upon the structure of the shock. The resulting *precursor* has a roughly exponential pressure distribution ahead of the shock and the streaming of the particles results in a decrease in the velocity jump across it, as is illustrated in Fig. 17.5a. Of particular interest is the fact that Kang and her colleagues have shown that the resulting momentum distribution of the accelerated particles has self-similar form as a function of time, as illustrated in Fig. 17.5b (Kang *et al.*, 2009).

The momentum spectrum of the high energy particles has the form $f(p) \propto g_s(p) p^{-4}$ and so if $g_s(p)$ is independent of momentum $f(p) \propto p^{-4}$ and $N(E) \propto E^{-2}$. The results shown in Fig. 17.5b display a number of important features. The fact that the solution is self-similar suggests that the form of the energy spectrum is stationary as the shock expands into the surrounding interstellar medium. The diagram shows that the self-similar spectrum is steepened relative to $N(E) \propto E^{-2}$ at intermediate energies and then flattens at higher energies. This can be understood in terms of the decreased value of the velocity, or density, ratio $r = u_1/u_2$ across the shock, which according to (17.44) leads to a steeper energy spectrum. The flattening at higher energies results from the fact that the higher energy

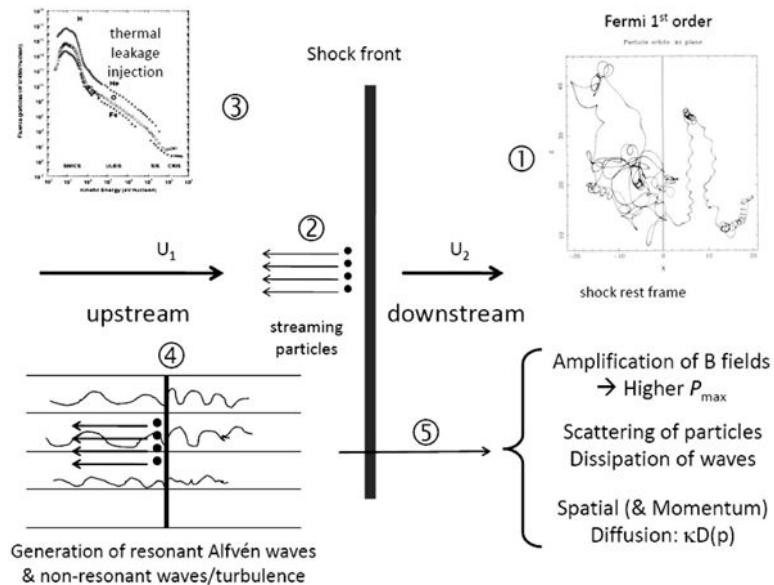


Fig. 17.4 Illustrating some of the physical processes to be included in a fully self-consistent model of diffusive shock acceleration. (Courtesy of Professor Hyesung Kang.)

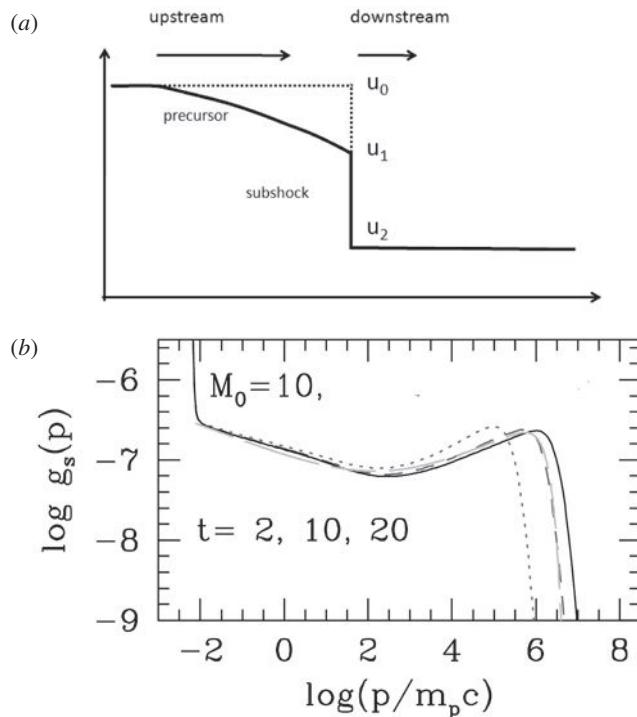


Fig. 17.5 (a) Illustrating the modification to the velocity distribution in the shock precursor due to the streaming of high energy particles back upstream through the shock. (b) The self-similar solution for the momentum distribution of the particles for a shock with Mach number $M = 10$ over a factor of 10 in dimensionless time (Kang *et al.*, 2009). The momentum distribution of the high energy particles is $N(p) \propto g_s(p) p^{-4}$.

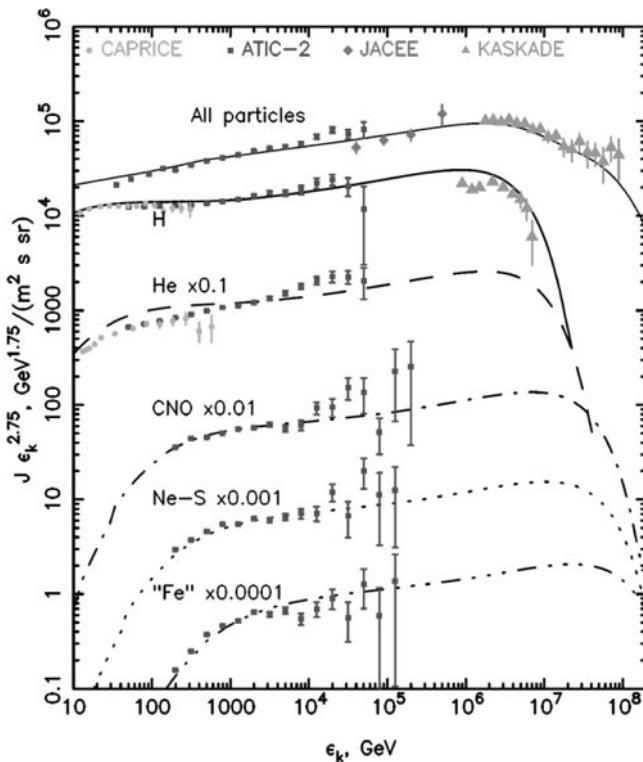


Fig. 17.6 The energy spectra of individual species fitted to the results of diffusive shock acceleration by Berezhko and Völk (2007).

particles have greater gyroradii and experience the acceleration associated with the larger ratio of u_0/u_2 rather than u_1/u_2 . At the highest energies, the exponential decrease in the energy spectrum is associated with the Lagage and Cesarsky cut-off. The predicted energy spectra for different cosmic ray species from calculations by Berezhko and Völk (2007) are shown in Fig. 17.6 which displays a number of the features of the spectrum discussed above.

An important recent addition to this story has been the detection of H α emission from the shock wave and from the precursor region (Lee *et al.*, 2007). Figure 17.7a shows a recent image of the shock front in knot b of Tycho's supernova remnant, as observed by the Hubble Space Telescope. The narrow emission coincides with the location of the synchrotron X-ray emission of the shock front. To the left of the narrow emission feature, diffuse H α can be observed which falls off in intensity with distance from the shock front. The origin of the H α emission is charge exchange interactions between the neutral interstellar gas and the hot plasma on either side of the shock. The atomic physics of this process enables information about physical conditions in the shock front and its vicinity to be obtained. For example, it provides a direct estimate of the ratio of the electron to ion temperatures in these regions. In Fig. 17.7b, the intensity of H α emission from the shock itself and the precursor are clearly

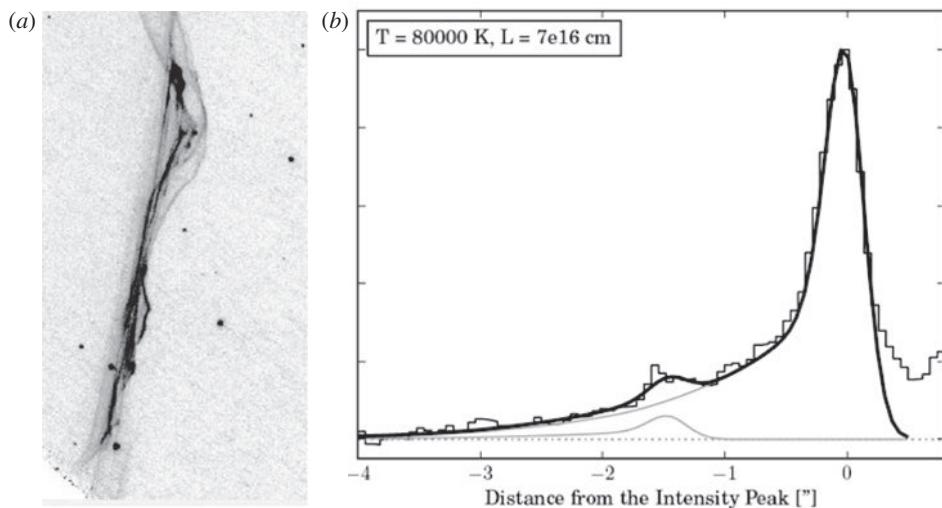


Fig. 17.7 (a) The shock wave in knot b in Tycho’s supernova remnant as observed by the HST in the H α line of hydrogen. (b) A fit of the profile of the H α intensity in the vicinity of knot b, showing the emission from the shock itself and from the precursor region. (Lee *et al.* 2010).

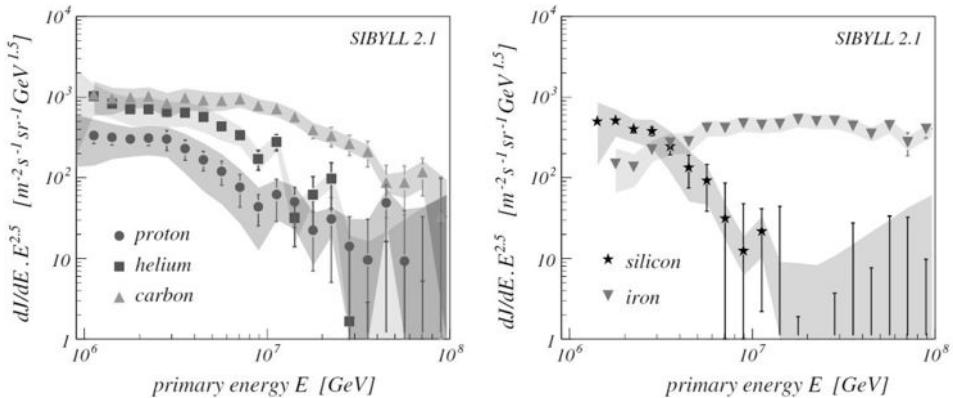
distinguished. The H α emission provides a new diagnostic tool for probing the shock and precursor regions and are consistent with the general picture developed above.

The conclusions of these studies are that high energy particles and turbulent magnetic fields are the natural products of the formation of collisionless shocks in cosmic environments with a more or less universal power-law energy spectrum which is a good match to the observed spectra. About 50% of the shock kinetic energy is transferred to the cosmic rays if the strong shocks have Mach numbers $M_s \geq 30$.

17.5.3 The energy spectrum of cosmic rays at and above the ‘knee’

The results of Sect. 17.5.2 provide an explanation for the origin of the energy spectrum of cosmic rays beyond the ‘knee’. Protons with energies up to 10^{16} – 10^{17} eV can be accelerated by diffusive shock acceleration in the strong magnetic fields now known to be present in the shocks present in young supernova remnants and, since the process accelerates particles to the same energy per nucleon, helium, carbon, oxygen and iron nuclei are accelerated to even greater total energies.

Support for this picture is provided by observations of cosmic ray air-showers by the KASCADE experiment. The Karlsruhe air-shower array is dedicated to the precise characterisation of cosmic ray showers with energies in the range 10^{16} – 10^{18} eV and uses a number of different types of detector to achieve this. The development of cosmic ray showers is modelled by two independent codes using the best understanding of nucleon–nucleon interactions at very high energies. A variety of diagnostic tools are used to separate the showers into those initiated by protons, helium nuclei and heavier species.

**Fig. 17.8**

The energy spectra of different cosmic ray species with energies in the range 10^{16} – 10^{18} eV as deduced from the Karlsruhe KASCADE experiment. The characterisation of different species was determined by the SYBILL 2.1 simulations of the development of cosmic rays showers. Energy spectra for protons, helium nuclei and carbon are shown in the left panel and silicon and iron in the right panel based on SIBYLL simulations. The shaded bands are estimates of the systematic uncertainties in the modelling procedure (Antoni *et al.*, 2005).

Figure 17.8 shows the results of the KASCADE experiment. The knee in the energy spectrum of helium nuclei is observed at significantly higher energy than that of the protons and this trend continues for species such as carbon, silicon and iron nuclei. As the authors state in their paper (Antoni *et al.*, 2005): ‘For both models the position of the knees in these spectra is shifted towards higher energy with increasing element number.’ They also caution, however, that: ‘At present, the limiting factors of the analysis are the properties of the high energy interaction models used and not the quality or the understanding of the KASCADE data.’ Note that the separate determinations of the energy spectra for different species sum to the total energy spectrum shown in Fig. 15.19.

Support for this result is provided by the HiRes 1 and 2 experiments which suggested that the composition of the cosmic rays at energies $E \sim 10^{18}$ eV is significantly heavier than that of protons alone (Fig. 15.21). Figure 17.9 shows schematically how the overall spectrum of cosmic rays up to 10^{18} eV and possibly higher could be accounted for by this process.

17.6 The highest energy cosmic rays

It seems highly probable that cosmic rays with energies up to $E \sim 10^{18}$ eV originate in our Galaxy and supernovae remnants are their likely source. There remain, however, the problems of accounting for the very highest energy cosmic rays with $E \sim 10^{19}$ – 10^{20} eV. It is convenient to introduce the diagram first presented by Hillas (1984) which indicates potential sites for their acceleration (Fig. 17.9). The original Hillas diagram has been

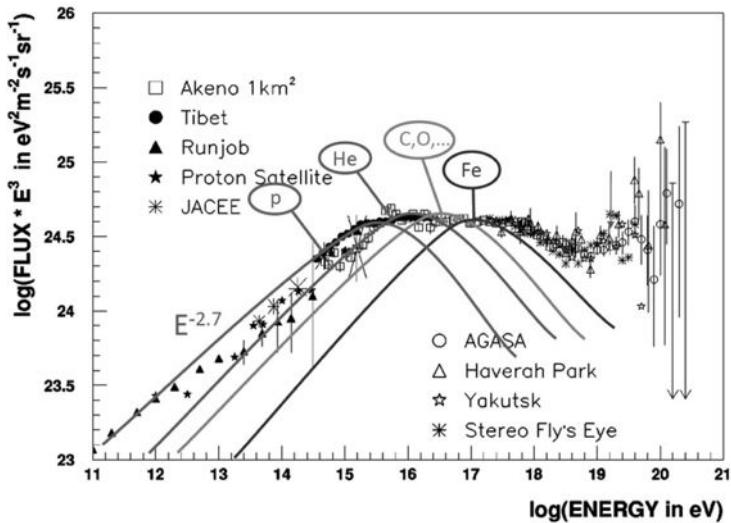


Fig. 17.9 Illustrating how the overall energy spectrum of cosmic rays could be accounted for as the superposition of the energy spectra of different species accelerated in the strong magnetic fields in supernova shock waves. The data are taken from the survey by Nagano and Watson (2000). (Courtesy of Professors Angela Olinto and Hyesung Kang.)

adapted in the light of new astrophysical understanding since 1984. The expression for the maximum energy of a cosmic ray (17.41) can be written as follows:

$$E_{\max} = z e B U L; \quad \frac{E_{\max}}{z\beta} = e B c L, \quad (17.45)$$

where $\beta = U/c$. If the particles have energies $\sim 10^{20}$ eV, the necessary combination of scale L and magnetic flux density B is given by

$$\frac{3 \times 10^{11}}{z\beta} = B L, \quad (17.46)$$

where L is in metres and B in tesla. This relation is plotted in Fig. 17.10. The diagram needs to be interpreted judiciously. If potential sources lie to the left of the locus (17.45), protons cannot be accelerated to 10^{20} eV in these objects. The diagram indicates that there are many potential sites of particle acceleration to $E \sim 10^{20}$ eV in a wide range of high energy astrophysical objects. Note, however, that the denominator of the left-hand side of (17.45) contains the factor $\beta = U/c$. Therefore, the potential sites of acceleration are only properly represented on the diagram if the velocities involved are $U \sim c$. This may indeed be the case for pulsar magnetospheres, γ -ray bursts, active galactic nuclei and radio jets, but it is unlikely to be so in the cases of the shocks in clusters of galaxies and supernova remnants. The magnetic flux densities would have to be correspondingly stronger for acceleration to 10^{20} eV to be feasible.

The crucial issues concern the masses and charges of the highest energy cosmic rays. If they are protons, an origin in active galaxies seems the most natural explanation, given

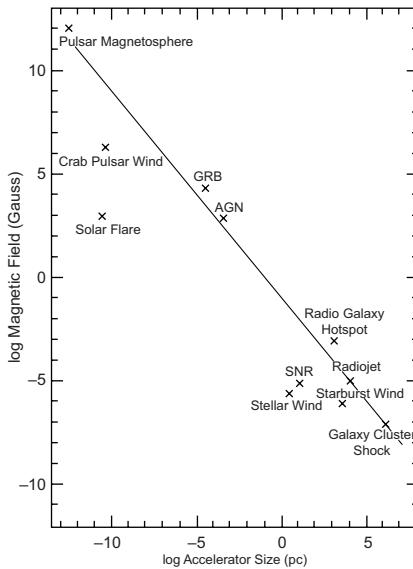


Fig. 17.10 The combinations of length-scale and magnetic flux density necessary to accelerate particles to energy $10^{20}/z\beta$ eV where $\beta = U/c$ (after Hillas, 1984). The typical magnetic flux densities found in different classes of astronomical objects on different scales are indicated by crosses.

their isotropy and correlation with relatively nearby large scale structures. On the other hand, if they are iron nuclei, energies of 10^{20} eV can be attained even if the velocities U are significantly less than the speed of light.



PART IV

EXTRAGALACTIC HIGH ENERGY ASTROPHYSICS

18.1 Introduction

The objective of this chapter is to set the scene for the later chapters, in which the physics of many different aspects of active galactic nuclei and their interactions with their surroundings are studied in some detail. Much of the lore and terminology of the active galactic nuclei are the product of the historical development of the subject. The history of the discovery of different types of active galaxy and the techniques used to find them are described briefly in this chapter.¹ A key feature of active galactic nuclei is that they are intrinsically broad-band, indeed multi-waveband, objects, each waveband providing complementary information, as well as possessing their own terminology and astrophysical infrastructure. There are several excellent books on different aspects of active galaxies. *Active Galactic Nuclei* by Robson (1999), *An Introduction to Active Galactic Nuclei* by Peterson (1997), *Quasars and Active Galactic Nuclei – an Introduction* by Kembhavi and Narlikar (1999), and *Active Galactic Nuclei* by Krolik (1999) can be recommended as providing a range of varied approaches to putting some order into their study.

18.2 Radio galaxies and high energy astrophysics

Cosmic rays, discovered by Victor Hess in 1913, provided the first evidence for the existence of relativistic matter originating from extraterrestrial sources (Sect. 1.10). It was, however, only after the Second World War and the development of the new astronomies that the astrophysical role of high energy particles and cosmic magnetic field could be addressed on the basis of astronomical observation. The birth of modern high energy astrophysics can be plausibly associated with Karl Jansky's announcement of the discovery of the radio emission of our Galaxy in May 1933. Following the pioneering studies by Alfvèn and Herlofson (1950), Kiepenheuer (1950) and Ginzburg (1951), the radio emission of the Galaxy was identified with synchrotron radiation, because of the power-law form of its radio spectrum and its high degree of polarisation. Subsequent radio surveys revealed large populations of Galactic and extragalactic radio sources, key discoveries being the association of the

¹ Many more details of the history are included in my book *The Cosmic Centre: A History of Astrophysics and Cosmology* (Longair, 2006).

radio source Cassiopeia A with a young supernova remnant and the radio source Cygnus A with a galaxy at a redshift of 0.057 (Baade and Minkowski, 1954). The observation of synchrotron radiation from Cassiopeia A demonstrated that supernova remnants are sources of relativistic electrons, with the consequences discussed in Sect. 16.6. The extragalactic radio source Cygnus A turned out to be more than a million times more powerful as a radio emitter than our own Galaxy. Just as remarkable was the fact that the radio emission did not originate from the galaxy itself. In 1953, Jennison and Das Gupta at Jodrell Bank used radio interferometric techniques to show that the radio emission originated from two huge lobes, more or less symmetrically disposed with respect to a luminous radio galaxy, which is the brightest member of a rich cluster of galaxies (Jennison and Das Gupta, 1953).

In 1953, Shklovsky proposed that the radio and optical continuum emission of the supernova remnant, the Crab Nebula, was synchrotron radiation (Shklovsky, 1953). A consequence of this proposal was that the optical continuum emission of the nebula should be linearly polarised and this was discovered to be the case by Dombrovski (1954), Vashakidze (1954) and by Oort and Walraven (1956). The famous optical jet in the radio galaxy M87 was also found to be linearly polarised and its radiation was also identified with the synchrotron emission of ultra-relativistic electrons.

In 1956, Geoffrey Burbidge estimated the minimum energy requirements in relativistic particles and magnetic fields for synchrotron radiation sources such as Cygnus A. He found that there must be at least 2×10^{52} J of energy in the form of relativistic particle and magnetic field energy, corresponding to a rest-mass energy of $3 \times 10^5 M_\odot$ of matter, assuming that only relativistic electrons are present in the sources, that is, $\eta = 1$ in the notation of Sect. 16.5 (Burbidge, 1956). It was inferred that there must exist some efficient mechanism by which rest-mass energy can be converted into ultra-high energy electrons and magnetic fields and then ejected from the host galaxy into intergalactic space, presumably through a process similar to that responsible for the jet in M87.

The galaxies associated with these bright radio sources are known as *radio galaxies* and their radio luminosities ranged from about 1000 to 100 million times that of our Galaxy. All galaxies emit some radio emission, in many cases similar to that of our own Galaxy which can be accounted for by the acceleration of electrons in supernova explosions. But something very different must be occurring in the radio galaxies. The term radio galaxy is conventionally taken to refer to those galaxies which are very much more powerful radio emitters than galaxies such as our own and are powered by jets of relativistic material originating in their active galactic nuclei.

18.3 The quasars

18.3.1 The discovery of quasars

These unexpected discoveries led to an upsurge in the construction of radio telescopes to unravel the astrophysics of extragalactic radio sources and to study their use as cosmological

probes. Their distances were found by first determining accurate radio positions for the sources and then identifying the associated radio galaxies, for which redshifts could be measured. The extragalactic radio sources present in bright source catalogues are associated with very luminous galaxies and so, even in the 1950s, they could be observed to significant cosmological distances. In 1960, Minkowski measured a redshift 0.46 for the radio galaxy associated with 3C 295, for many years the galaxy with the largest measured redshift (Minkowski, 1960).

In 1958, Morgan introduced a morphological classification scheme, often called the Yerkes scheme, which was based upon the central concentration of the light in galaxies (Morgan, 1958). Most normal galaxies could be accommodated in a sequence *a-f-g-k*, objects of class *a* having the weakest central concentration and those of class *f* the strongest. There were some galaxies, however, which Morgan found lay outside the standard morphological classifications:

- *D galaxies* These were originally defined as ‘systems showing rotational symmetry without pronounced spiral or elliptical structure’. The definition was revised by Matthews, Morgan and Schmidt (1964) in order to exclude spirals and ellipticals. ‘These galaxies have an elliptical-like nucleus surrounded by an extensive envelope . . . no D galaxies as defined here are observed to be highly flattened.’ The defining feature is the extensive stellar envelope. Among the radio galaxies they discussed were a few *supergiant D galaxies* – these were observed close to the centres of Abell clusters and were denoted cD galaxies. Their diameters were 3–4 times as great as those of lenticular galaxies in the same cluster.
- *N-galaxies* These galaxies were originally defined to be ‘systems having small brilliant nuclei superimposed upon a considerably fainter background.’ In the survey of Matthews, Morgan and Schmidt (1964), they noted that they possessed ‘brilliant, star-like nuclei containing most of the luminosity of the system. A faint, nebulous envelope of small visible extent is observed.’ Examples included the galaxies associated with radio sources 3C 227, 234 and 445.

By the early 1960s, the positions of radio sources could be measured with sufficient accuracy for identifications to be made with faint galaxies. Among the identifications were three strange cases, 3C 48, 196 and 286, which were identified with stars of an unknown type with strange optical spectra (Matthews and Sandage, 1963). The breakthrough came in 1962 when Hazard, Mackey and Shimmins measured the structure and position of the radio source 3C 273 by the method of lunar occultation (Hazard *et al.*, 1963). One of the components was coincident with a thirteenth magnitude star, while the other coincided with a faint optical jet (Fig. 1.10). The spectrum of the star was measured by Schmidt in December 1962 who realised that the strong emission line spectrum was the Balmer series of hydrogen, but shifted to longer wavelengths than their rest wavelengths – the redshift of 3C 273 was found to be $z = 0.158$ (Schmidt, 1963). The other members of this class were also found to have large redshifts, 3C 48, 196 and 286 having redshifts of 0.367, 0.871 and 0.849, respectively. They were named *quasi-stellar radio sources* and within a year this name was abbreviated to the word *quasar*. In 1965, the redshift of the quasar 3C 9 was found to be $z = 2.016$, by far the largest redshift for any object at that time (Schmidt, 1965).

The discovery of quasars came as a complete surprise. The optical luminosity of 3C 273 is enormous. Its absolute magnitude in the V waveband was $M_V = -25.5 + 5 \log h$, compared with a typical galaxy luminosity of $M_V^* = -21 + 5 \log h$, in other words, about 100 times more luminous than a galaxy such as our own. Galaxies at the same redshift as 3C 273 are the faint objects seen towards the bottom of the image shown in Fig. 1.10. Searches through the Harvard plate archives then showed that the enormous luminosity of 3C 273 was variable on the time-scales of a year and greater as far back as the plates had been taken (Smith and Hoffleit, 1963). These properties were quite unprecedented and a few astronomers including Hoyle, Burbidge and Arp wondered whether or not the redshifts of the quasars really were cosmological. It was not long, however, before similar properties were found in related classes of active galaxies, which were certainly at the cosmological distance implied by their redshifts. It was not long before radio-quiet counterparts of the quasi-stellar radio sources were discovered. The first examples were announced by Sandage in 1965 and these *radio-quiet quasars* are about 100 times more common than their radio-loud counterparts (Sandage, 1965).

These discoveries changed the astronomical scene for good.² For the first time, optical and radio astronomers met with general relativists to understand the role general relativity might play in the study of these active systems. Perhaps the most important conclusion was the realisation that the quasars must involve strong gravitational fields and so general relativity must play a central role in understanding their physical properties.

18.3.2 Finding radio-quiet quasars

The construction of complete samples of radio-loud quasars was relatively straightforward, but the compilation of corresponding samples of optically selected quasars was more challenging, because they are rare objects which must be distinguished among the very much more populous normal stars. A composite quasar spectrum spanning the wavelength range 100–600 nm is presented in Fig. 18.1, showing the underlying broad-band continuum spectrum, as well as the presence of the strong broad emission lines, which can strongly influence their colours (Francis *et al.*, 1991). Woltjer (1990) has given a concise summary of the relative merits of the different procedures for defining statistically complete quasar samples.

- One of the most successful techniques for finding radio-quiet quasars has been the search for star-like objects with ultraviolet excesses, one of the characteristics of the first samples of radio quasars. This technique relies upon the fact that the UV-optical continuum spectrum of quasars is, to a first approximation, a power law and so they are relatively more intense ultraviolet and infrared emitters than normal stars. The pioneering studies of Bracessi and his colleagues first demonstrated convincingly that the UV-excess objects have a much steeper number count than that expected in a uniform

² Some impression of the excitement in these early days of the study of active galactic nuclei can be gained from the volume *Quasi-stellar Radio Sources and Gravitational Collapse*, the proceedings of the First Texas Symposium on Relativistic Astrophysics held in Dallas in December 1963, which also includes reprints of many of the important early papers (Robinson *et al.*, 1965).

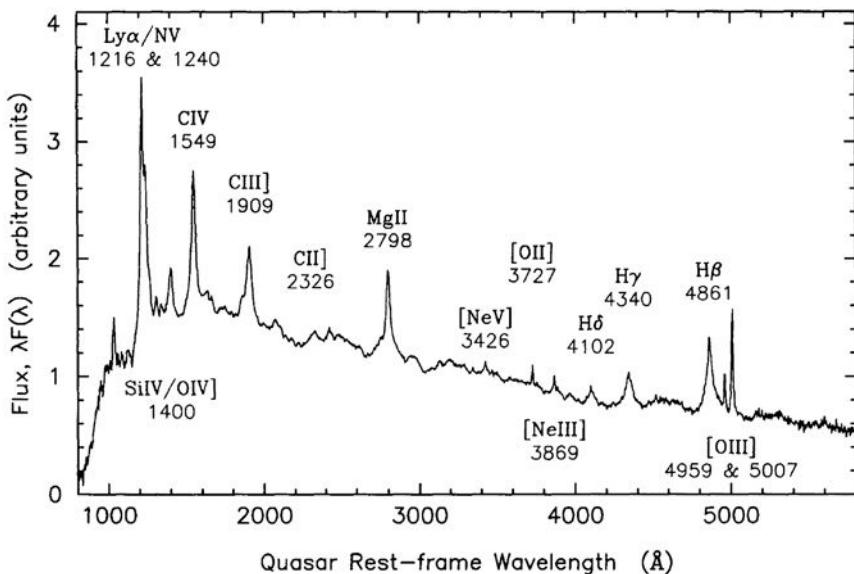
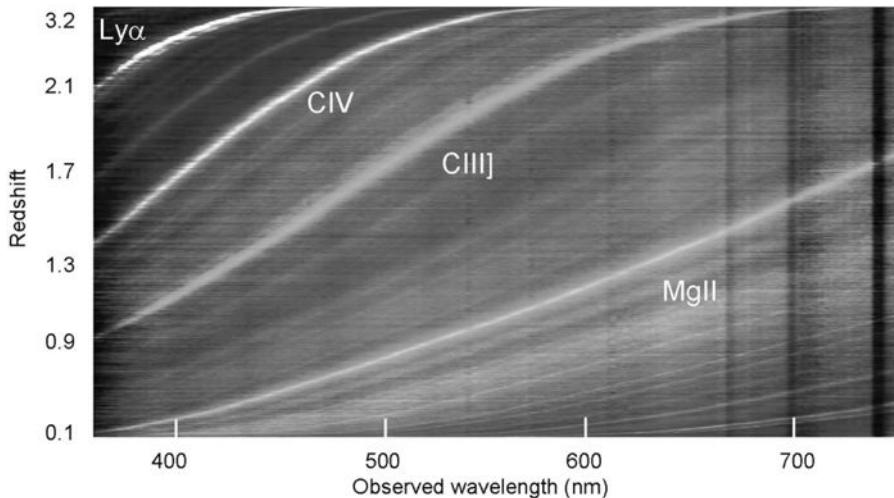


Fig. 18.1

A composite mean quasar spectrum extending from about 100 to 600 nm, found by averaging over 700 quasars from the Large Bright Quasar Survey of Francis and his colleagues (1991). The spectrum is plotted in units of λF_λ . The spectrum contains a number of the common emission lines found in quasars superimposed upon a strong non-thermal continuum. The flux scale is arbitrary.

Euclidean world model (Braceggi *et al.*, 1970). This technique is very successful in discovering quasars with redshifts $z \leq 2.2$. At this redshift, the Lyman- α emission line is redshifted into the B filter and so the quasars no longer exhibit ultraviolet excesses in the $(U-B)$ colour index. At larger redshifts, absorption lines associated with the Lyman- α forest depress the ultraviolet emission beyond the Lyman- α line so that these quasars do not exhibit such a strong ultraviolet excess. The ultraviolet excess technique was exploited by Schmidt and Green who derived a complete sample of 114 bright radio-quiet quasars with B , on average, less than 16.16 in a survey which covered about a quarter of the whole sky (Schmidt and Green, 1983). Interestingly, only 114 quasars were confirmed spectroscopically among the 1874 ultraviolet excess objects which satisfied the ultraviolet excess selection criteria. Most of the 1874 objects were hot, hydrogen-atmosphere subdwarf stars, the sdB stars, and white dwarfs.

The ultraviolet excess technique was exploited in the 2dF Quasar Redshift Survey carried out at the Anglo-Australian telescope. The ultraviolet excesses were defined from machine scans of plates taken by the UK Schmidt Telescope in the (u, b_J, r) wavebands. 25 000 quasars were discovered by this technique. The stacked spectra in the observer's rest frame are shown in Fig. 18.2. The observed wavelength is plotted on the abscissa and the redshift on the vertical axis. The plot shows the characteristic strong emission lines of Mg II, C III], C IV and Ly α seen in quasar spectra as they are redshifted through the optical waveband. At the largest redshifts, the Lyman- α line is redshifted into the optical window.

**Fig. 18.2**

The observed spectra of 20 000 quasars from the 2dF Quasar Survey. The spectra have been stacked in order of redshift from $z = 0.1$ to $z = 3.2$. The prominent emission lines seen in quasar spectra are labelled. (Courtesy of the 2dF Quasar Survey Team.)

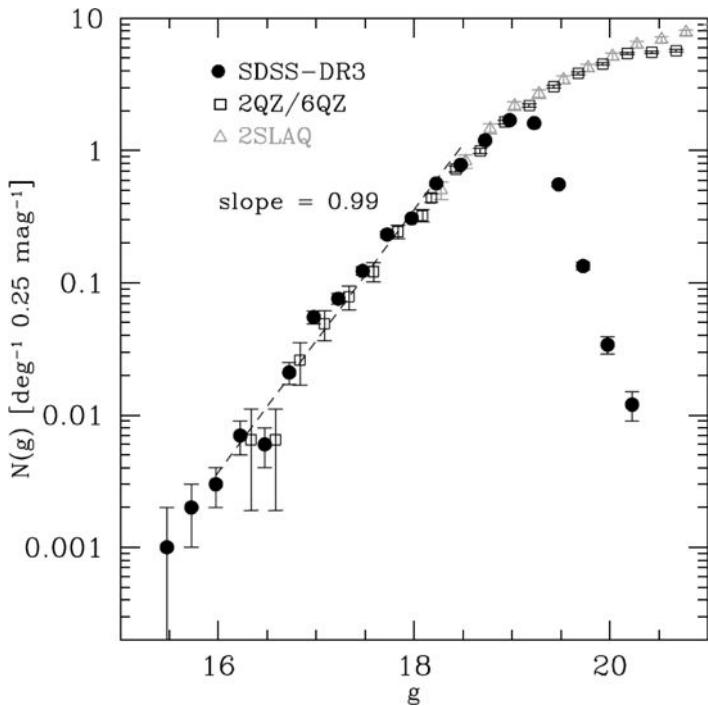
- The extension of this technique involved the use of multicolour photometry to longer optical wavelengths in order to distinguish objects with the typical spectra of large redshift quasars from stars. Koo and Kron used (U, J, F, N) photometry to find radio-quiet quasars with ultraviolet excesses to $B = 23$ (Koo and Kron, 1982) and found the first evidence for the convergence of the counts of radio-quiet quasars. The technique was extended by Warren and his colleagues to four-colour photometry using observations in the U , J , V , R and I wavebands, providing four colours ($U-J, J-V, V-R, R-I$) (Warren *et al.*, 1987). Stars lie along a rather narrow locus in this four-dimensional colour space. By searching for objects which lay well away from that locus, Warren and his colleagues found the first quasar with redshift $z > 4$. This technique for searching for quasars with redshifts greater than 4 was further refined by Irwin and his colleagues, who realised that they could be found by means of two-colour photometry from observations in the (B_J, R, I) wavebands (Irwin *et al.*, 1991). The colours of these large redshift quasars are very different from those of stars because the redshifted Lyman- α forest enters the B_J waveband and so strongly depresses the redshifted continuum intensities of the quasars at blue wavelengths.

The largest sample of quasars to date has been found from images taken through the *ugriz* filters as part of the Sloan Digital Sky Survey (Richards *et al.*, 2006). The five filters spanned the complete optical spectrum between the sky cut-off at ultraviolet wavelengths to the cut-off wavelength of silicon CCD detectors at about 1.1 μm . These broad-band filters were centred on the wavelengths $u - 350$ nm, $g - 480$ nm, $r - 625$ nm, $i - 770$ nm and $z - 910$ nm. From the third release of the SDSS data, 46 420 spectroscopically confirmed broad-line quasars were discovered, of which 15 343 formed

a complete statistical sample which has been used to study the evolution of the luminosity function of optically selected quasars out to redshift $z = 5$.

- The extension of this colour selection technique to even larger redshifts has been carried out using SDSS data by Fan and his colleagues (Fan *et al.*, 2001, 2004). They searched for ‘*i*-band drop-outs’, meaning that the *i*-band intensity was very significantly depressed relative to the *z*-band intensity. In these cases, the depression of the continuum intensity to the short wavelength side of the Lyman- α line is shifted to the extreme red end of the optical spectrum. They discovered nine quasars with redshifts $z > 5.7$, the largest redshift being 6.28. These observations constrain the evolution of the luminosity function of the quasars at the very largest redshifts.
- Another approach is to make use of the fact that the Lyman- α and C IV emission lines are always very strong in the spectra of quasars and are superimposed upon a roughly power-law continuum energy distribution (Fig. 18.1). Osmer and his colleagues (1982) pioneered the use of a dispersion prism, or grating, in conjunction with a wide-field telescope to discover quasars with redshifts $z > 2$, at which these lines are redshifted into the optical waveband. Perhaps the most remarkable use of this technique has been the survey of Schmidt and his colleagues (Schneider *et al.*, 1991; Schmidt *et al.*, 1995) who used the Palomar 200-inch telescope as a fixed transit instrument in conjunction with a grism and a large area CCD camera, which was clocked at the sidereal rate. In this way, six narrow bands across the sky were scanned both photometrically in the *v* and *i* wavebands, as well as spectroscopically, resulting in a total scanned area of 62 degree². Of 1660 candidate emission line objects, 141 were found to be quasars in the redshift interval $2.0 < z < 4.7$ (Schneider *et al.*, 1991; Schmidt *et al.*, 1995).
- Finally, one of the most important characteristics of the quasars is that they are variable over time-scales from days to decades. If a sufficiently long baseline is used, say, of the order of 10 years, all quasars are found to be variable. Following the pioneering efforts of Hawkins (1986), it has been confirmed that this is a successful approach for selecting complete samples of quasars. The degree of variability of the quasar is correlated with its luminosity, but not with redshift (Hook *et al.*, 1991). In one variant of this approach, Majewski and his colleagues used a combination of variability and the lack of proper motions to estimate the completeness of various approaches to the definition of complete quasar samples (Majewski *et al.*, 1991). They found the important result that the multicolour surveys miss at most up to 34% of the quasars and probably the actual percentage is far less than this figure.

The upshot of all these studies is that the quasars display a steep number count which converges at faint magnitudes (Fig. 18.3). The changes of the optical luminosity function of optically selected quasars with cosmic epoch was derived by Boyle and his colleagues (2000) out to redshift $z \sim 2.3$ using a complete sample of 6000 quasars from the AAT 2dF Quasar Redshift survey (Fig. 18.4). At zero redshift, the luminosity function of the overall quasar population joins smoothly on to that of the Seyfert galaxies, an important but natural continuity of the properties of active galactic nuclei. It is interesting that Schmidt and Green classified only 92 of their 114 ultraviolet excess objects as quasars since they had absolute magnitudes brighter than $M_B = -23$. The 22 lower luminosity

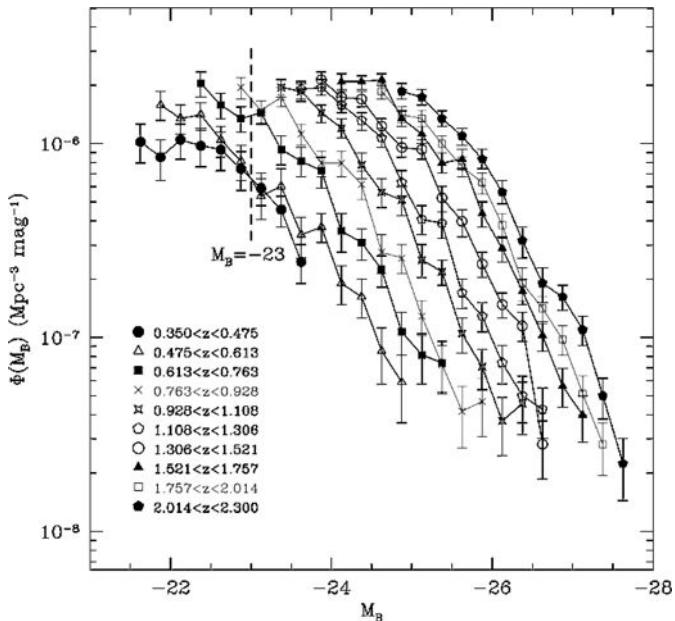
**Fig. 18.3**

Differential g -band number counts for quasars from the third data release of the Sloan Digital Sky Survey. The colour selection criteria have been matched to those of the 2dF Quasar Redshift Survey. The cut-off at faint magnitudes in the SDSS sample is due to the i -band limiting magnitude of the survey. Also shown is a power-law fit to the bright end of the SDSS-DR3 sample; it has slope 0.99 ± 0.12 , compared with the Euclidean expectation of 0.6 (Richards *et al.*, 2006).

objects were classified either as Seyfert 1 nuclei or what they termed low-luminosity quasars.

18.4 Seyfert galaxies

In many ways, the quasars were discovered too early. It was soon realised that the quasars are only the most extreme examples of what are now termed *active galactic nuclei*. Individual cases of galaxies with compact luminous nuclei had been noted much earlier: NGC 1068 by Fath in 1908, NGC 5236 by Slipher in 1917, NGC 4151 by Campbell and Moore in 1918 and NGC 4051 by Hubble in 1926. The first systematic study of these galaxies was carried out by Carl Seyfert in 1943. His list of galaxies was selected from the plate archives of the Mount Wilson Observatory, their distinctive features being that they contained an ‘exceedingly luminous stellar or semi-stellar nucleus which contains a relatively large percentage of the total light’. In addition, the spectra of these nuclei showed many high excitation emission lines, specifically, very intense and broad Balmer and forbidden lines.

**Fig. 18.4**

The evolution of the optical luminosity function for 6000 optically selected quasars in the redshift range $0.35 \leq z \leq 2.3$ observed in the 2dF Quasar Redshift Survey carried out at the Anglo-Australian Telescope (Boyle *et al.*, 2000).

The Doppler broadening of the emission lines corresponded to velocities up to 8500 km s⁻¹, quite different from those of normal galaxies, which generally display only stellar absorption lines or narrow emission lines associated with regions of star formation. In addition, the spectra of the intense continuum radiation from their nuclei were smooth, unlike the spectrum of starlight. The majority of Seyfert's galaxies were spirals.

Seyfert's pioneering efforts were largely neglected until the 1960s. In 1963, just before the discovery of the quasars, Burbidge, Burbidge and Sandage surveyed a wide range of evidence for violent activity originating in the nuclei of galaxies in a prescient paper entitled *Evidence for the occurrence of violent events in the nuclei of galaxies* (Burbidge *et al.*, 1963). The similarity between the properties of the quasars and the Seyfert galaxies was reinforced in 1968 when it was shown that the nuclear continuum emission of Seyfert galaxies was variable (Fitch *et al.*, 1967).

Markarian had initiated a survey of galaxies with strong blue and ultraviolet continuum spectra which were found in objective prism surveys of galaxies brighter than about sixteenth magnitude using the 1-m Schmidt telescopes at the Byurakan Observatory (Markarian, 1967). Over a period of 10 years, Markarian and his colleagues published catalogues of galaxies with blue and ultraviolet excesses, the final catalogue containing about 1500 galaxies (Markarian *et al.*, 1981). The objects in Markarian's catalogue are referred to as *Markarian galaxies* and the majority of them are star-forming galaxies which have intense blue and ultraviolet continua because of the presence of hot young massive stars. About 10% of the Markarian galaxies, however, possess active galactic nuclei with intense

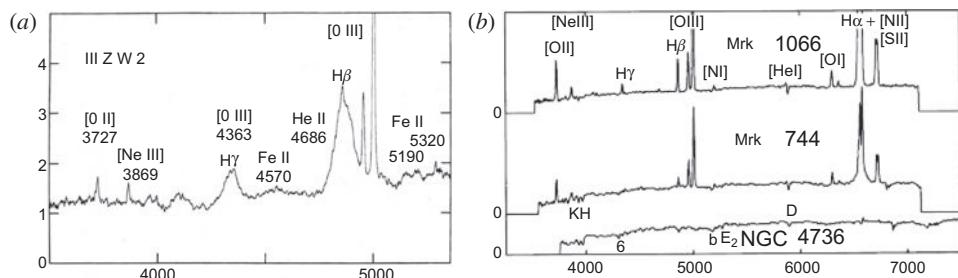


Fig. 18.5

Typical spectra of Seyfert 1 and Seyfert 2 galaxies. (a) The spectrum of the Seyfert 1 galaxy III Zw 2, exhibiting the characteristic broad emission lines of the resonance lines of hydrogen while the forbidden lines of oxygen and neon are narrow (Osterbrock, 1978). (b) The spectra of the Seyfert galaxies Mkn 1066 (Type 2) and Mkn 744 (Type 1.8) compared with that of the galaxy NGC 4736. The permitted emission lines are much narrower than in the case of the Seyfert 1 galaxy and have the same linewidths as the forbidden lines (Goodrich and Osterbrock, 1983).

continuum emission and strong broad emission lines, exactly the same properties as the Seyfert galaxies. Seyfert galaxies were also found in Zwicky's lists of compact galaxies which were discovered during the compilation of his *Catalogue of selected compact galaxies and of post-eruptive galaxies* (Zwicky and Zwicky, 1971). In his words, 'We call compact galaxies those which can just be distinguished from stars on plates taken with the Palomar 48-inch telescopes and which have diameters of 2–5 arcsec ...'. Sargent (1970) showed that, among the 2000 compact galaxies in the catalogue, there is a diverse range of types, the common characteristic being that they are all of high surface brightness. Among these are Seyfert galaxies, N-galaxies and quasar-like objects.

Spectroscopic studies of these samples of Seyfert galaxies revealed that they come in two varieties. Khachikian and Weedman (1971; 1974) showed that there are significant differences in the widths of the emission lines among the Seyfert galaxies. The *Seyfert 1 galaxies* have permitted emission lines which are very much broader than the forbidden lines. Specifically, permitted lines, such as the Balmer series of hydrogen, neutral and ionised helium and singly ionised iron, have full widths at half maximum corresponding to velocities up to about 10 000 km s $^{-1}$ and are often variable in intensity, whereas the forbidden lines, such as [O III], are very much narrower, with velocity widths corresponding to only about 1000 km s $^{-1}$. In the *Seyfert 2 galaxies*, the permitted and forbidden lines have similar velocity widths, about 1000 km s $^{-1}$. Examples of the spectra of Seyfert 1 and Seyfert 2 galaxies are shown in Fig. 18.5. It is inferred that, in the Seyfert 2 galaxies, the narrow permitted and forbidden emission lines originate from the same clouds located up to about 1 kpc from the nucleus, whereas in the Seyfert 1 galaxies, in addition to the narrow line regions, the broad emission lines originate from very much more compact regions, within about 1 pc of the nucleus itself. The broad linewidths are assumed to be associated with the motions of gas clouds close to the nucleus.

As the quality of spectra continued to improve, it became apparent that there are Seyfert galaxies which are intermediate between the two classes. The permitted lines can consist of two components, the standard broad permitted lines of the Seyfert 1 spectrum superimposed

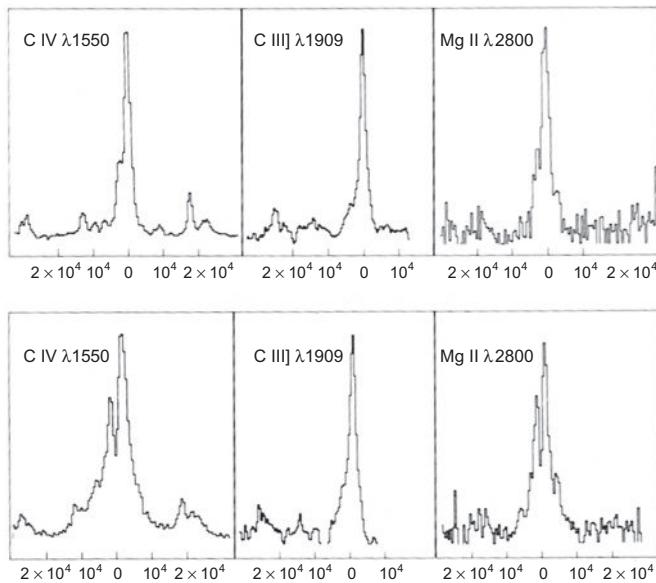


Fig. 18.6 Illustrating the dramatic changes in the line profiles of the permitted lines in the spectrum of Seyfert 1 galaxy NGC 4151 as observed by the International Ultraviolet Observatory. The upper series of line profiles was taken in February 1978 and the lower set in June 1980. In the upper series, the ultraviolet continuum intensity is relatively weak while in the lower set, the continuum is bright. The C IV and Mg II lines change dramatically between the two epochs while the partially forbidden line C IIIJ is more or less unchanged (from Ulrich *et al.*, 1984).

upon which is a narrow line, corresponding to the Seyfert 2 spectrum. These galaxies have been designated Seyfert 1, 1.5, 1.8, 1.9, 2, reflecting the relative strengths of the broad and narrow-line components. Regular observations of NGC 4151 by the International Ultraviolet Observatory showed that the broad line component is highly variable so that the spectral properties changed it from being a Seyfert 1 to a Seyfert 2 galaxy and vice versa (Fig. 18.6). The most extreme Seyfert 1 galaxies form a continuous sequence with the radio-quiet quasars. In their survey of bright radio-quiet quasars, Schmidt and Green (1983) arbitrarily chose to distinguish between the Seyfert 1 galaxies and radio-quiet quasars at a nuclear luminosity of $M_B = -21.5 + 5 \log h$.

As detailed studies of the spectra of radio galaxies advanced during the 1970s, it became apparent that they too could be divided into two classes, the *broad-line radio galaxies*, BLRGs, which have spectra similar to the Seyfert 1 galaxies, and the *narrow-line radio galaxies*, NLRGs, which are similar to the Seyfert 2 galaxies (Osterbrock and Ferland, 2005). All the N-galaxies turned out to be broad-line radio galaxies. The differences between the optical spectra of the Seyfert 2 and narrow-line radio galaxies are small, but there are significant differences between those of the Seyfert 1 and the broad-line radio galaxies. As described by Osterbrock and Ferland, these are: (i) the line profiles of the hydrogen lines are better described as Seyfert 1.5 type spectra rather than Seyfert 1; (ii) the broad components are typically broader than those observed in Seyfert 1 spectra; (iii) the Fe II lines are significantly stronger in the Seyfert galaxies as compared with the

broad-line radio galaxies. Perhaps the most striking difference between the Seyferts and the radio galaxies is that practically all the Seyferts are spiral galaxies whereas the radio galaxies are all giant elliptical, D or cD galaxies.

This is far from the end of the story, however, because, as the identifications of the radio galaxies extended to redshifts $z \sim 1$, some of the radio galaxies were found to possess very strong but narrow emission lines, which turned out to be associated with large emission line regions excited by radio jets far from the active galactic nucleus. This topic is taken up in Sect. 20.5.

18.5 Blazars, superluminal sources and γ -ray sources

Among the sources found in low frequency radio surveys were a number of objects with flat radio spectra, $I_\nu \propto \nu^0$. As radio technology developed, surveys were carried out at shorter wavelengths and many more of these ‘flat-spectrum’ objects were discovered. They were all found to be of very small angular size and long-term monitoring of these compact objects showed that they were often variable on the time-scale of months or years. The compact nature of these sources was confirmed by very long baseline interferometry (VLBI) which showed that the most compact of these had brightness temperatures $T_b = (\lambda^2/2k)(S_\nu/\theta^2) \geq 10^{10}$ K, indicating that the emitting electrons had to be relativistic. The natural interpretation of these observations was that the flat spectra resulted from the superposition of a number of compact source components in which synchrotron self-absorption took place at high radio frequencies (Sect. 8.7). Many of these compact sources were associated with quasars.

In the course of these studies, the compact and highly variable radio source VRO 42.22.01 was found to be associated with the ‘variable star’ BL-Lacertae, or BL-Lac (McLeod and Andrew, 1968). The optical spectrum of BL-Lac was found to resemble a power law and was featureless. This source became the prototype for the objects known as *BL-Lac objects*. The spectrum of the BL-Lac object OJ 287 from radio to X-ray wavelengths is shown in Fig. 18.7, illustrating the broad-band continuum nature of the emission. Because of the absence of spectral features, it is a challenge to measure their redshifts, but this has been achieved, either by making observations with very high signal-to-noise ratio to detect faint emission lines, or by making observations at periods when the active nucleus is in a state of low activity. These observations have shown that the BL-Lac objects are generally relatively low redshift objects, $z \leq 0.2$. They are relatively rare objects, a few hundred of them having been identified. The BL-Lac objects are all highly variable radio sources.

A closely related family of objects are the *optically, violently variable* objects, or OVV. These are highly variable objects at both radio and optical wavelengths with similar properties to the BL-Lac objects, but with the distinctive feature that they possess strong, broad emission lines, exactly as if they were classical quasars. They span a somewhat larger redshift range than the BL-Lac objects, $0.1 \leq z \leq 2$. The term *blazar* is used to encompass both the BL-Lacs and the OVV within a single family.

Two other characteristic properties of blazars are first that they display high degrees of linear polarisation and second that most of these compact radio sources exhibit the

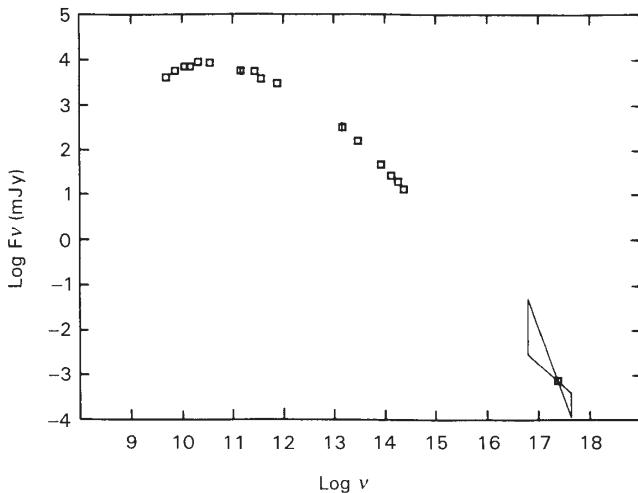


Fig. 18.7 The millimetre to X-ray spectrum of the BL-Lac object OJ287, showing that it consists of a single smooth continuous spectrum which is interpreted as the spectrum of synchrotron radiation which becomes self-absorbed at far-infrared and millimetre wavelengths (Robson, 1999).

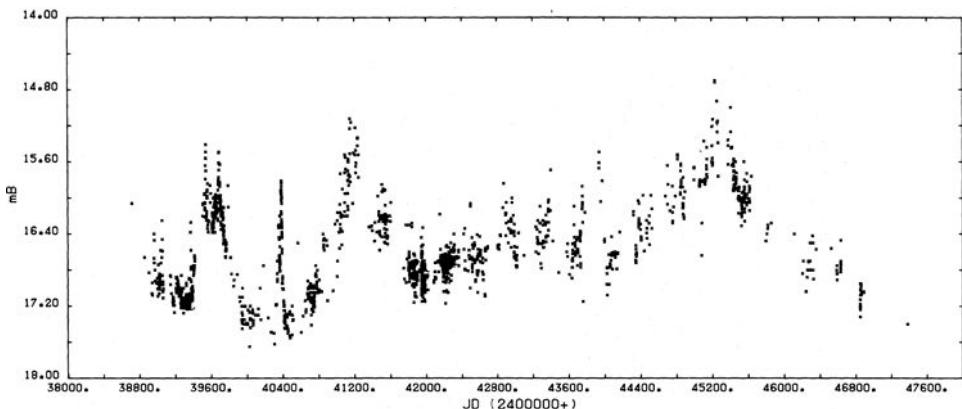


Fig. 18.8 The variability of the optical continuum emission of the blazar 3C 345 (Kidger, 1989).

phenomenon of *superluminal motion*, which was discovered by VLBI observations of compact radio sources in the 1970s. In the superluminal sources, the components of the compact radio source are observed to separate at speeds greater than the speed of light. This topic is dealt with in some detail in Chap. 22, where it is shown that the phenomenon can be attributed to the relativistic motion of the radio source components ejected from the active nucleus. The components have to be ejected at an angle quite close to the line of sight for the largest superluminal speeds to be observed. As shown in that chapter, the BL-Lac objects are probably associated with low-luminosity radio galaxies of Fanaroff–Riley class 1, while the OVV objects are associated with the Fanaroff–Riley class 2 sources.

One of the most important discoveries of the Compton Gamma-Ray Observatory has been that virtually all the intense variable extragalactic γ -ray sources are associated with blazars which exhibited superluminal motion of their compact nuclear radio components. As discussed in Chap. 22, the extreme luminosities and short time-scale variability of the γ -ray emission can be accounted for if it is assumed that the γ -ray source moves at a velocity close to the speed of light towards the observer, consistent with the observation of superluminal motion of the compact radio components.

18.6 Low Ionisation Nuclear Emission Regions – LINERS

The quasars and blazars are among the most extreme examples of high energy astrophysical activity in the nuclei of galaxies, but it is important to know whether or not similar phenomena occur at more modest levels in all galaxies. The central regions of galaxies are often regions of intense star formation and these give rise to the characteristic emission spectra of H II regions with emission lines of [O II], for example. In 1980, Heckman showed that there are in addition, many cases in which an emission line spectrum originates from within the nuclear regions, similar in some ways to the Seyfert spectra, but with narrower and weaker lines (Heckman, 1980). For example, neutral oxygen [O I] can be strong but high excitations lines such as [Ne V] and [O III] are relatively weak. These weak active galactic nuclei are known as *Low Ionisation Nuclear Emission Regions* or LINERs. These phenomena are quite common, approximately one-third of all galaxies within about 20–40 Mpc displaying the characteristics of LINERs. They are also commonly found among the luminous infrared galaxies. In some cases, these objects possess weak compact nuclear radio sources. While a convincing case can be made that the strong, broad emission lines in the nuclei of quasars and Seyfert galaxies are the result of photoionisation of clouds in the vicinity of an intense source of ultraviolet ionising radiation, there is an ongoing debate about whether or not this is the case for the LINERs. Collisional ionisation by shock waves may also be responsible for the excitation of the emission line regions.

18.7 Ultra-Luminous Infrared Galaxies – ULIRGs

The IRAS satellite carried out the first complete sky survey in the far-infrared waveband in the early 1980s (Sect. 1.4 and Fig. 1.5), the broad-band detectors having central wavelengths of 12, 25, 60 and 100 μm , wavebands which are inaccessible from ground-based observations. Of particular interest for extragalactic astrophysics was the discovery that many galaxies, particularly those in which there is evidence for large amounts of star formation, are very powerful far-infrared sources. The cause of the far-infrared emission is the emission of dust heated by intense optical and ultraviolet emission. In many cases, it is certain that the optical and ultraviolet emission is due to the presence of star-forming regions within the galaxy. This is not necessarily the case, however, for a number of the

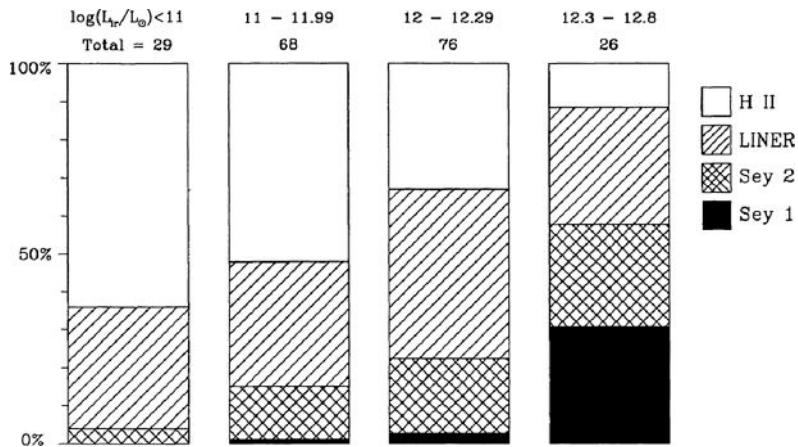


Fig. 18.9 The fraction of galaxies with different types of spectra in complete samples of ultra-luminous infrared galaxies (ULIRGs) as a function of far-infrared luminosity. The fraction of Seyfert 1 and 2 galaxies increases to over 50% in the most luminous samples (Veilleux, 1999).

most luminous IRAS galaxies. Those with far-infrared luminosities $L_{\text{ir}} \geq 10^{12} L_{\odot}$ are as luminous as the quasars and are known as Ultra-Luminous InfraRed Galaxies, or ULIRGs. Soifer and his colleagues (1987) showed that the space density of ULIRGs is greater than that of radio-quiet quasars of the same luminosity. The properties of these galaxies have been surveyed by Sanders and Mirabel (1996), who addressed the question of the origin of the intense ultraviolet and optical emission which heat the dust.

Optical and near-infrared spectroscopy has shown that some of these hyperluminous infrared sources display the characteristic strong narrow emission lines of a star-bursting stellar population, while others have broad or high ionisation lines, characteristic of the Seyfert galaxies and quasars. Veilleux (1999) and his colleagues studied complete samples of the more luminous IRAS sources and classified them according to the features observed in their optical and infrared spectra (Fig. 18.9). In the most luminous sample of ULIRGs, $12.3 \leq \log(L_{\text{ir}}/L_{\odot}) \leq 12.8$, more than 50% of the galaxies display Seyfert galaxy characteristics, roughly half of them being Seyfert 1 galaxies and half Seyfert 2s. The dereddened luminosities of the broad-line emission seen in those classified as Seyfert 1 galaxies are similar to those of optically selected quasars of comparable bolometric luminosity. Thus, some of the most luminous ULIRGs are powered by the optical-ultraviolet emission of a quasar-like active galactic nucleus.

The intense far-infrared emission is concentrated towards the centres of these galaxies. Veilleux noted that, even in cases in which the presence of an active galactic nucleus is clearly established, star-burst activity may also contribute a large fraction of the bolometric luminosity. In the case of the ULIRGs, there is convincing evidence that many of them are associated with strong interactions between galaxies. As in the case of systems such as the Antennae (Fig. 3.3a), such collisions between gas-rich spiral galaxies can result in powerful starbursts and the disruption of the interstellar gas within the galaxies. The resulting collisions between gas clouds can lead to the channelling of gas into the nuclear

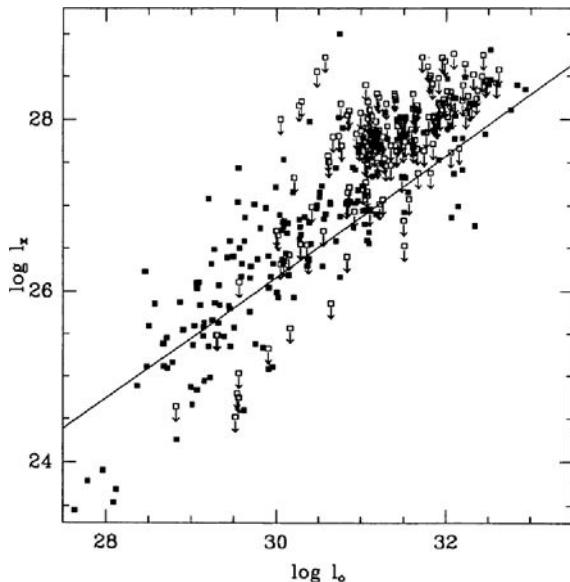
**Fig. 18.10**

High resolution optical images of six quasars observed by the Hubble Space Telescope. Top row: PG 0052+251, IRAS 04505–2958 and QSO 0316–346; bottom row: PHL 909, PG 1012+008 and IRAS 13218+0552. Details of the analysis of these images and 14 others are discussed by Bahcall and his colleagues (1997a). (Courtesy of Michael Disney, NASA and the Space Telescope Science Institute.)

regions and so provide a means of feeding the black hole (Sanders *et al.*, 1988). As discussed in Chap. 19, there is compelling evidence that most massive galaxies possess massive black holes in their nuclei and so can become active once they are supplied with fuel. The observation of highly distorted structures in the galaxies underlying the quasars support this picture (Fig. 18.10). This evidence suggests a close relation between star-burst phenomena and the formation of active galactic nuclei.

18.8 X-ray surveys of active galaxies

The UHURU Sky Survey established that, away from the Galactic Plane, there is a large population of extragalactic X-ray sources (Fig. 1.11a). All classes of active galactic nuclei described in the previous sections are also members of this population of X-ray sources. In addition, the X-ray and optical luminosities of individual active galactic nuclei are strongly correlated. This is illustrated by the results of the survey of quasars and Seyfert 1 galaxies observed by the Einstein X-ray Observatory. Wilkes and her colleagues (1994) plotted the X-ray against the optical luminosities of a sample of 343 such objects, the best-fit relation being $L_x \propto L_{\text{opt}}^{0.71}$, although there is some scatter about this relation. A similar result is found if the analysis is restricted to radio-loud quasars, the only difference being that the X-ray luminosities are about a factor of 3 greater than those of radio-quiet quasars of the same optical luminosity (Ku *et al.*, 1980; Zamorani *et al.*, 1981) – the presence of radio emission influences the soft X-ray luminosity, over and above that expected from the optical-X-ray

**Fig. 18.11**

A plot of the X-ray luminosity of a sample of 343 radio-quiet quasars and Seyfert 1 galaxies at 2 keV against their optical luminosities at 250 nm. The units of luminosity L_X and L_0 are in erg s^{-1} . The X-ray data are taken from those active galaxies targeted by the Einstein X-ray Observatory and includes both X-ray detections and upper limits. The straight line shows the underlying correlation when account is taken of the many sources for which only upper limits are available (Wilkes *et al.*, 1994).

correlation. The strongest radio-X-ray correlation is between the luminosity of the compact radio cores and the X-ray emission, rather than with the extended radio emission, a result found for both radio quasars and radio galaxies (Fabbiano *et al.*, 1984; Kembhavi *et al.*, 1986).

The largest survey of the complete X-ray sky was carried out by the ROSAT mission in the period 1990–99 in the 0.1–2.4 keV X-ray energy band. The survey contains about 60 000 sources and, in addition, very deep observations were made of small regions of sky to define the X-ray source counts to the faintest achievable flux densities. There is a large excess of faint sources with a steep number count which converges at the lowest X-ray flux densities (Hasinger *et al.*, 1993). Boyle and his colleagues (1993) have determined the evolution of the X-ray luminosity function of these sources and shown that it is similar to that of optically selected quasars shown in Fig. 18.4. The similarity of their behaviour can be appreciated from the fact that the evolving luminosity functions can be related by the scaling relation $L_X \propto L_{\text{opt}}^{0.88 \pm 0.08}$.

It is straightforward to determine how much of the X-ray background intensity in the 0.5–2 keV energy band can be attributed to discrete X-ray sources by integrating the X-ray number counts. There is general agreement that essentially all the background intensity in this waveband is associated with discrete sources and that this population is dominated by active galactic nuclei. A major fraction of these sources are Type 1 active galactic

nuclei in which the characteristic broad emission lines and non-thermal continuum are observed. Therefore, the majority contributors to the X-ray background at these energies are unobscured sources in which the nuclear regions of active galaxies are observed directly. The main contributors to the soft X-ray waveband have typical photon number spectral index about 1.9 with a dispersion of about 0.2–0.3 about this value. This immediately poses a problem since the spectral index of the X-ray background spectrum is significantly smaller than this value in the 1–10 keV waveband, roughly $I(\varepsilon) \propto \varepsilon^{-1.4}$ in terms of photon number spectral index. Therefore, the same population cannot explain the X-ray background intensity at harder X-ray energies, $\varepsilon \approx 10$ keV.

A similar analysis to that of the last paragraph can be carried out for the sources detected at hard X-ray energies ($\varepsilon \sim 10$ keV). At least 50% and probably more of the background is associated the discrete sources already detected, but their spectral mix is somewhat different. The most important difference is that there is a major population of sources which are strongly absorbed at soft X-ray energies. Sources have been observed in which photoelectric absorption strongly suppresses the soft X-ray flux density and so ‘hardens’ the X-ray spectra. The inferred column densities can be up to 10^{24} cm $^{-2}$ and greater. For column depths greater than 10^{24} cm $^{-2}$, the effects of Comptonisation need to be taken into account and this process causes major distortions of the X-ray source spectra whatever the initial input photon spectrum. Gilli and his colleagues provide template spectra for this wide range of source spectra which include the effects of iron lines and reflected X-ray components from the inner regions of the active galactic nucleus (Gilli *et al.*, 2007). At energies greater than 10 keV, the number count data are much sparser, but the background spectrum continues to be remarkably flat up to about 30 keV beyond which it changes slope rather dramatically and continues more or less as a power law to γ -ray energies. It is assumed that Comptonised sources make a significant contribution to the background spectrum at energies $\varepsilon \geq 10$ keV.

The upshot of this discussion is that quasars and active galactic nuclei are the dominant extragalactic populations in the X-ray waveband.

18.9 Unification schemes for active galaxies

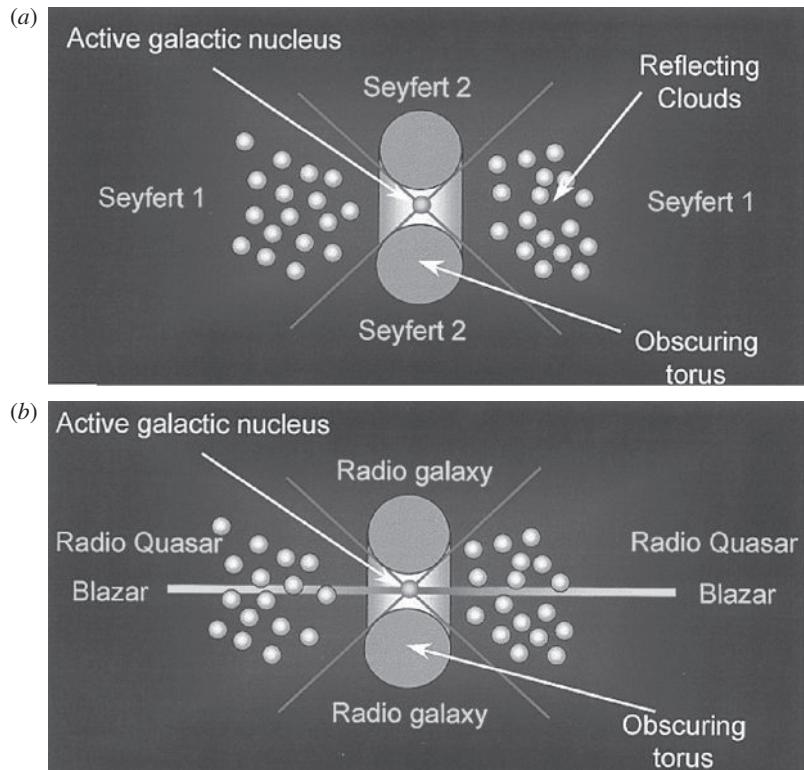
The concept of unification resulted from the realisation that projection effects are important in determining the observed properties of active galactic nuclei. There are several effects which can account for the differences between the various classes of active galaxy. For example, the presence of an obscuring torus surrounding the active nucleus allows the nuclear regions to be observed directly only along lines of sight at relatively small angles to the axis of the torus. Another example involves relativistic effects associated with the presence of relativistic beams or jets emitted from the active nucleus – the observed properties of the emission depend strongly upon the angle at which the jet is observed. It is therefore certain that projection effects are important, but it is much less clear the extent to which all types of active galactic nuclei can be unified within a single scheme.

18.9.1 Polarisation studies of Seyfert 1 and Seyfert 2 galaxies

The importance of projection effects in distinguishing between Seyfert 1 and Seyfert 2 galaxies was first convincingly demonstrated by Antonucci and Miller (1985) in their study of the nearby Seyfert 2 galaxy NGC 1068. In their spectropolarimetric observations, they found that the polarised line emission is as broad as the broad permitted lines observed in Seyfert 1 galaxies. They interpreted these observations in terms of what became the standard model for the unification of Seyfert 1 and Seyfert 2 galaxies. There is some form of ‘obscuring torus’ about the nucleus so that, when the galaxy is observed at a small angle to the axis of the torus, the nuclear regions, which contain the active nucleus, the characteristic Seyfert 1 broad-line regions and strong blue and UV continuum, are observed. If the axis of the torus is observed at a large angle to the line of sight, the nuclear regions are obscured and so only the narrow-line regions, which are located much further from the nucleus than the broad-line regions, are observed in direct light. The obscured central regions can, however, be observed in light reflected from clouds outside the region of the torus by the gas and dust in the narrow-line regions. The continuum and broad-line emission from the nuclear regions are polarised in the process of reflection by Thomson or Rayleigh scattering. The unification scheme is illustrated in Fig. 18.12a, which is a highly schematic cross-section through the nuclear regions of a Seyfert galaxy.

Surveys of the unification hypothesis for Seyfert 1 and 2 galaxies are given by Antonucci (1993) and Miller (1994). The principal pieces of evidence which support the picture are as follows:

- (i) Spectropolarimetric observations have been successfully carried out for a number of other Seyfert 2 galaxies and narrow-line radio galaxies, confirming the presence of reflected broad-line emission from the nuclear regions. These are demanding observations, the percentage polarisations normally amounting to only a few percent.
- (ii) Further evidence for the obscuration of the ionising continuum emission from the nucleus is found by comparing the flux of ionising photons observed from the nucleus with the flux required to excite the emission line regions. In a number of cases, the observed flux of ionising radiation is significantly less than that required to account for the ionisation and excitation of the gas clouds, suggesting that they are exposed to a more intense radiation field than that which would be inferred if the continuum emission were emitted isotropically.
- (iii) This line of reasoning indicates that the radiation from the nucleus is emitted in an ‘ionisation cone’. Figure 18.13 shows two examples observed by the Hubble Space Telescope in which the ionisation from the nucleus take places in an ‘ionisation cone’. In the case of NGC 1068 (Fig. 18.13a), the cone structure is consistent with the geometry inferred by Antonucci and Miller (1985) and with photon counting analyses. In the case of the radio galaxy 3C 270 (NGC 4261), there is a cone-like structure within the nuclear regions of the galaxy and, in addition, there is a ring, or torus, of emission about the nuclear regions (Fig. 18.13b) – this torus is on a very much greater physical scale than that which would be needed for the unified picture, but it is indicative of the types of structures which can form in the central regions of galaxies. The radio

**Fig. 18.12**

(a) A schematic diagram illustrating the unified picture for Seyfert 1 and Seyfert 2 galaxies. The nuclear regions contain the source of continuum radiation and the broad-line regions which can be observed in scattered light, if the axis of the torus is observed at a large angle to the line of sight. The reflecting clouds occupy more or less the same volume as the clouds responsible for the narrow-line emission. (b) Illustrating the unified model for radio galaxies and radio quasars observed in samples of bright radio sources. Radio quasars are observed when the axis of the radio source lies within about 45° of the line of sight. When observations are made almost along the axis of the radio jet, superluminal radio sources and blazars are observed.

jets are collimated along the axis of the torus. NGC 4261 is a radio galaxy rather than a Seyfert galaxy, but it shows the type of geometry expected in unified pictures of Seyfert 1 and Seyfert 2 galaxies.

- (iv) If the unified picture of Seyfert 1 and 2 galaxies is correct, there should be a difference between the X-ray absorption properties of Seyfert 1 and Seyfert 2 galaxies. Observations made with the *Ginga* and ASCA satellites have shown that Seyfert 2 galaxies have the same properties as Seyfert 1 galaxies at hard X-ray energies, 1–10 keV, at which the effects of absorption would be expected to be negligible (Wilkes, 1999). These similarities include a steep X-ray spectral index, as well as the presence of a reflected X-ray component and the iron K α emission feature. It is found that the Seyfert 2 galaxies generally show stronger low-energy X-ray absorption with typical

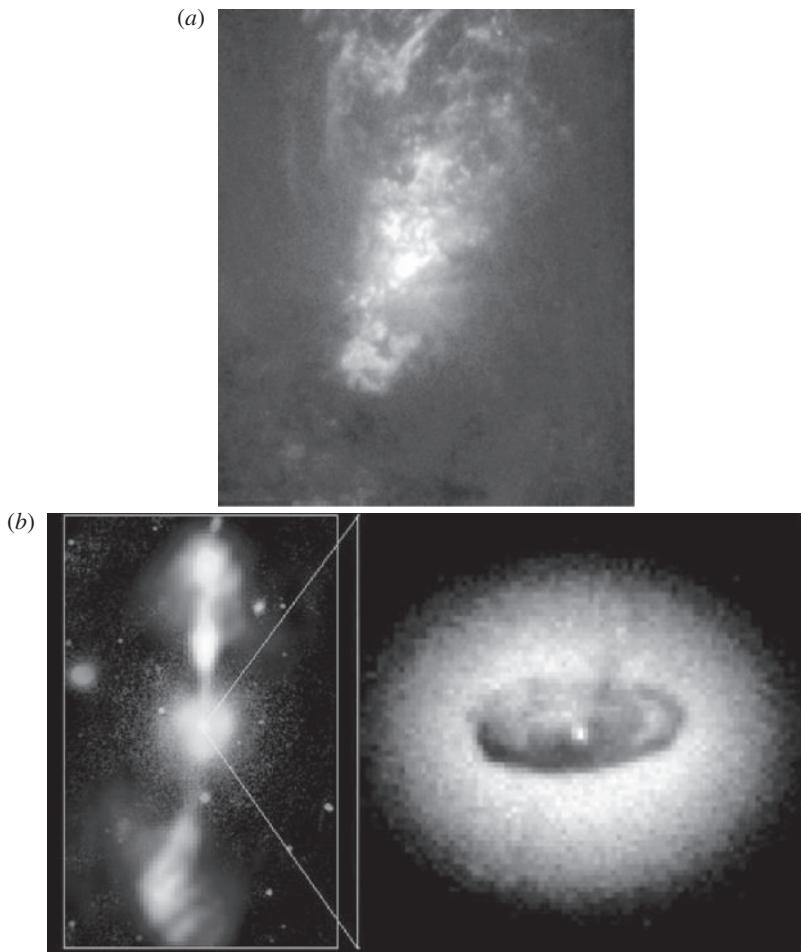


Fig. 18.13 (a) The structure of the line-emitting regions close to the nucleus of the Seyfert 2 galaxy NGC 1068. The nucleus of the galaxy lies towards the bottom left of the picture and the emission line regions lie within a cone with apex centred on this nucleus. (b) The central regions of the radio galaxy NGC 4261 (3C 270) showing the central ring and evidence for a cone-shaped emission region, the axis of which is aligned with the radio jet. In the left-hand panel, the collimated jets to the north and south of the galaxy show the distribution of the radio emission, which are powered by jets from the active nucleus. (Courtesy of NASA and the Space Telescope Science Institute.)

column densities $\geq 10^{22} \text{ cm}^{-2}$, compared with $\sim 10^{21} \text{ cm}^{-2}$ for Seyfert 1 galaxies and quasars, supporting the unification picture advocated by Antonucci and Miller.

18.9.2 Radio quasars and radio galaxies

The role of projection effects in determining the observed properties of extragalactic radio sources has been advocated by Barthel (1989; 1994), who discussed the unification of *radio galaxies* with *radio quasars*. By these terms are meant those objects which appear

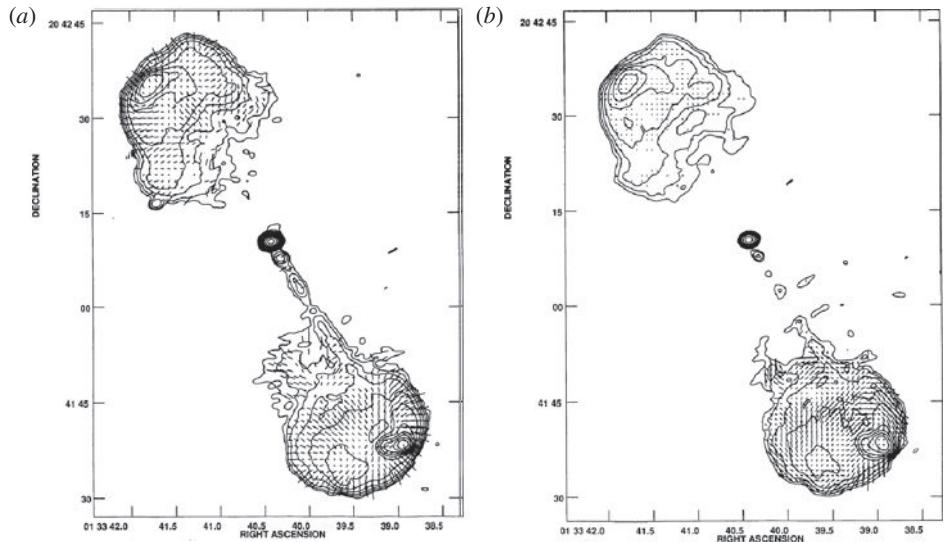


Fig. 18.14

The radio structure of the radio quasar 3C 47 at (a) 5 GHz and (b) 1.4 GHz, showing the one-sided radio jet emanating from the nucleus and the polarisation vectors at the two frequencies (Fernini *et al.*, 1991). The extended radio component on the counter-jet side is much more depolarised at the lower frequency than that on the jet side.

in catalogues of bright radio sources, such the 3CRR catalogue and other bright radio source catalogues (Laing *et al.*, 1983; Wall and Peacock, 1985). These objects are very much rarer than the Seyfert galaxies, the sources in these catalogues having redshifts in the range $0.1 < z < 2$. They have the advantage that it is straightforward to define statistically complete samples of extragalactic radio sources and, furthermore, their intense radio emission is generally in the form of highly collimated, linear structures so that projection effects can be studied. Barthel's discussion centred upon the properties of the classical double radio sources, those referred to as FR2 radio sources (see Chap. 21). The evidence for the importance of projection effects and the unification of the radio galaxies with the radio quasars can be summarised as follows (Barthel, 1994):

- (i) Superluminal motions are found in many compact radio quasars. According to the preferred ballistic model for these sources (Sect. 22.2), the axes along which the relativistic jet propagates must lie at a small angle to the line of sight.
- (ii) In the radio quasars, one-sided radio jets are often observed emanating from their nuclei along the axes of their large-scale radio structures. It is natural to attribute the one-sidedness of the radio jets to the combination of relativistic Doppler and aberration effects, the intensity of the jet being enhanced by the blueshift of the emission while the counter-jet moving away from the observer is too faint to be detected because of its large redshift. A good example of such a one-sided jet is observed in the radio quasar 3C 47 (Fig. 18.14).
- (iii) This picture is supported by the depolarisation asymmetries of the radio lobes, which are fed by the radio jets. The term depolarisation means the decrease in the percentage polarisation of the source with increasing wavelength (Sect. 12.3.4). The greater the

depolarisation, the greater the path length through the depolarising medium, which is assumed to be the gaseous halo enveloping the radio source. It is found that the depolarisation is greater on the invisible counter-jet side, as is expected if the jet-side is pointing towards the observer. This phenomenon is clearly displayed by the radio quasar 3C 47 (Fig. 18.14). This correlation of counter jet side with greater depolarisation was discovered by Laing (1988) and Garrington and his colleagues (1988) and is referred to as the Laing–Garrington effect. In Laing’s survey of 1993, the correlation was found to hold good for a sample of 40 quasars and seven radio galaxies (Laing, 1993).

- (iv) The projected linear sizes of the radio galaxies are greater than those of the radio quasars at the same redshifts, as would be expected if the radio quasars are observed at a smaller angle to the line of sight as compared with the radio galaxies.
- (v) The optical emission observed about a number of radio galaxies is linearly polarised, as expected if the radiation of an obscured nucleus were being scattered into the line of sight, as in the case of the Seyfert 2 radio galaxies.

These pieces of evidence are consistent with the unification scenario for radio galaxies and radio quasars illustrated in Fig. 18.12b.

The picture bears a strong resemblance to the unification scenario for Seyfert galaxies, but now radio quasars are observed when the nucleus is observed within a cone of half-angle roughly 45° and a radio galaxy observed when the nuclear regions are hidden by the obscuring torus. Highly collimated radio jets are assumed to be emitted by the nucleus along the axis of the torus and these jets are responsible for powering the outer radio hot-spots and extended radio lobes. If the radio jet is observed at an angle close to the line of sight, the emission of the relativistic jet is strongly enhanced by aberration and Doppler effects and superluminal motions can be observed. It is natural to attribute the *blazar* phenomenon to radio sources in which the relativistic jets point almost precisely along the line of sight to the observer. In a few extreme blazars, high dynamic range observations have enabled the underlying double radio source to be observed, a good example being the blazar 3C 371, which has the appearance of an FR2 double radio source observed more or less end-on (Fig. 18.15).

Another unification scenario concerns the relation between the BL-Lac objects and radio galaxies. A good case can be made for BL-Lac objects being the relativistically beamed counterparts of the Fanaroff–Riley class 1 radio sources in which the maximum brightnesses of the source components lie less than half the distance to the outer limit of the radio source structure (Chap. 21). When radio observations with very high dynamic range are made of BL-Lac objects, weak extended radio sources are observed about the intense radio core. Their low luminosity radio structures are similar to those of FR1 sources. It is natural to interpret the radio emission of the BL-Lac objects as the relativistically beamed radio emission of jets emerging from the nucleus at angles very close to the line of sight. There are, however, differences between the FR1 and FR2 radio jets on larger scales, in that normally the radio jets in the FR1 sources are double-sided. Thus, unlike those in the radio quasars, the radio jets are not relativistically beamed once they leave the nuclear regions. It is assumed that the jets in the FR1 sources must be rapidly decelerated between the scales on

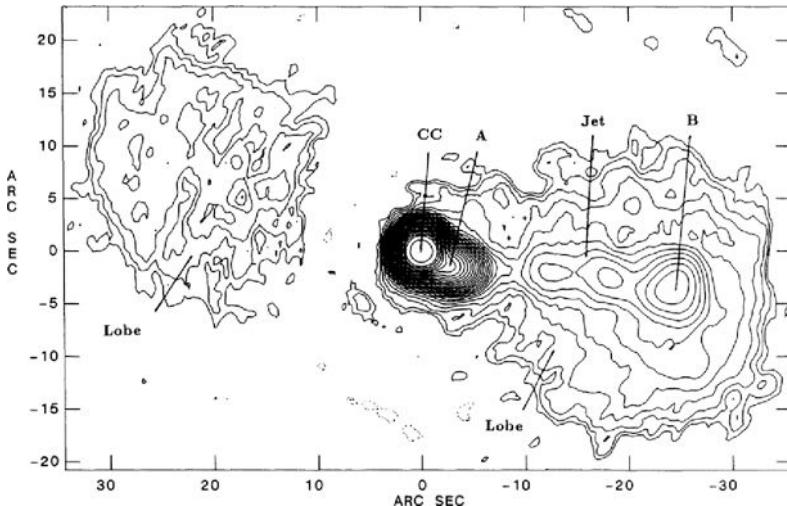


Fig. 18.15 The highly variable radio source 3C 371 observed with high dynamic range (Wrobel and Lind, 1990). The intense compact core (CC) is the highly variable component and very precise data analysis is required to reveal the underlying extended double radio source which has the characteristics of an FR2 double radio source viewed more or less end-on.

which the BL-Lac phenomenon is observed and that which can be observed by instruments such as the Very Large Array. The differences between the two types of radio source is assumed to be correlated with the power of the jets or beams powering them.

There are many other possible applications of these concepts. For example, Orr and Browne (1982) studied the issue of how much of the dispersion in the radio luminosities of the compact radio cores of extended double radio quasars could be attributed to relativistic beaming effects. Their analysis made use of the fact that, since the 3CR sample of radio sources is selected according to the flux densities of their extended radio structure, there should be no orientation bias in the selection of their radio cores, which are generally very weak at the low radio selection frequency of 178 MHz. By working in terms of the ratio of the core to total flux density $R = S_{\text{core}}/S_{\text{total}}$ at high frequencies, they were able to show that a considerable part of the dispersion in R could be due to the effects of observing the beamed core emission at different angles to the line of sight. Support for this picture was provided by the analysis of Kapahi and Saikia (1982) who found that other properties of the sources correlated with R , as might be expected for this model. For example, on average, the double radio structures of sources with larger values of R would be expected to be of smaller physical size, since they are observed closer to the line of sight, and so on. Browne and Murphy (1987) went on to apply these ideas to the beaming of the X-ray emission from compact nuclei.

18.9.3 Conclusions

Orientation and beaming effects clearly play an important role in the observed properties of active galactic nuclei. What is not so clear is the extent to which all classes of active galaxy can be unified with every other sort. For example, how are the radio-loud and the

radio-quiet quasars related? Scheuer and Readhead (1979) proposed that the luminous compact, flat-spectrum radio sources might be the same class of object as the radio-quiet quasars, the radio-loud objects simply being those in which the radio emission is beamed towards the observer. This type of unification does not seem to work, because the predicted flux density distribution of the radio-quiet quasars is different from that which was subsequently observed. Furthermore, with high dynamic range radio observations, double radio sources have been discovered underlying the strong compact radio sources, similar to the extended radio sources, whereas, where the structures the radio-quiet quasars can be determined, they are found to be much more compact objects. There also seems to be a distinction between the parent galaxies of the radio-loud and radio-quiet objects. The most powerful radio sources are usually associated with giant elliptical galaxies whereas the spirals are very rarely associated with the radio luminosities or radio structures found in the classical powerful radio galaxies. We will follow up some of these issues in the succeeding chapters.

19.1 The properties of black holes

The properties of Schwarzschild and Kerr black holes were described in some detail in Sect. 13.11 and compelling evidence for stellar-mass black holes in X-ray binary systems was presented. These concepts are now applied to the study of black holes in the nuclei of galaxies. Let us recall some of their key properties.

- (i) Dead stars with masses greater than $3M_{\odot}$ must be black holes.
- (ii) The only properties which isolated black holes can possess are *mass, angular momentum and electric charge*.
- (iii) For a non-rotating black hole, a *Schwarzschild black hole*, there is a spherical surface about the black hole from which electromagnetic radiation suffers an infinite gravitational redshift, as observed from outside this surface. This *surface of infinite redshift* has radius

$$r_g = \frac{2GM}{c^2} = 3 \left(\frac{M}{M_{\odot}} \right) \text{ km}, \quad (19.1)$$

and is known as the *Schwarzschild radius*. Radiation with frequency ν_0 emitted at radius r from the black hole suffers a *gravitational redshift*, so that the frequency of the radiation as observed at an infinite distance from the black hole ν_{∞} is

$$\nu_{\infty} = \nu_0 \left(1 - \frac{2GM}{rc^2} \right)^{1/2} = \nu_0 \left(1 - \frac{r_g}{r} \right)^{1/2}. \quad (19.2)$$

- (iv) There is a *last stable circular orbit* about a Schwarzschild black hole at radius $r = 3r_g$. Within this radius, test particles spiral inevitably into the black hole, contributing to its mass and angular momentum. As will be shown below, the speed of a test particle on the last stable circular orbit of a Schwarzschild black hole is $v_{\phi} = c/2$.
- (v) In the case of black holes with finite angular momentum J , the *Kerr black holes*, the surface of infinite redshift occurs at radius

$$r_{\infty} = \frac{GM}{c^2} + \left[\left(\frac{GM}{c^2} \right)^2 - \left(\frac{J}{Mc} \right)^2 \right]^{1/2}. \quad (19.3)$$

There is a maximum angular momentum which a rotating black hole can possess, $J_{\max} = GM^2/c$. The radius of the surface of infinite redshift for a maximally rotating

black hole then occurs at $r_\infty = GM/c^2 = \frac{1}{2}r_g$, that is, half the Schwarzschild radius of a non-rotating black hole.

- (vi) There is a last stable orbit about a Kerr black hole, but now test particles can orbit in either the corotating or counter-rotating directions with respect to the angular momentum axis of the black hole. For a maximally rotating Kerr black hole, the last stable circular orbit for corotating test particles coincides with r_∞ , that is $r = GM/c^2$, one sixth of the corresponding radius for a non-rotating, Schwarzschild black hole.
- (vii) The binding energies of particles on the last stable orbit for Schwarzschild and maximally rotating Kerr black holes, relative to their rest mass energies, are

$$\text{Schwarzschild} \quad \left[1 - \sqrt{\frac{8}{9}} \right] \quad \text{Kerr} \quad \left[1 - \sqrt{\frac{1}{3}} \right] \quad (19.4)$$

corresponding to 5.72% and 42.3% of their rest mass energies, respectively. A fraction of the rotational energy of a rotating black hole can also be made available to the external Universe. In terms of the rest-mass energy of the black hole, this fraction is

$$1 - 2^{-1/2} \{ 1 + [1 - (J/J_{\max})^2]^{1/2} \}^{1/2}, \quad (19.5)$$

amounting to 29% for a maximally rotating Kerr black hole, $J = J_{\max}$.

The key points are:

- Black holes are the most compact objects which can exist with mass M . The Schwarzschild radius r_g is the typical dimension of a black hole and is a convenient length-scale for describing phenomena on the scale of the surface of infinite redshift and the last stable orbit of the black hole.
- Black holes are the most powerful energy sources in astrophysics. The 42% of the rest-mass energy of matter which can be released when it falls into a black hole can be compared with the 0.7% released in the nucleosynthesis of helium from hydrogen in stars. The energy is released on the most compact scale which an object of mass M can have.

19.2 Elementary considerations

The key pieces of observational evidence which forced astrophysicists to take black holes very seriously as the energy sources for active galactic nuclei were their extreme luminosities and, even more important, the fact that these were observed to vary rapidly over short time-scales. Essentially all the classes of active galactic nuclei described in Chap. 18 possess quasar-like phenomena in their nuclei, in the sense that there are point-like sources of non-thermal radiation which vary in intensity on time-scales of years, months or days. Some of the most rapid variations are found in the blazars and in the X-ray variability of some Seyfert galaxies. The light curve of the blazar 3C 345 (Fig. 18.8) and the X-ray variability of the Seyfert galaxy MCG -6-30-15 (Fig. 14.3b) are good examples of these phenomena. In the latter case, significant variations in the X-ray luminosity can take place on the time-scale

of about one hour in a source which lies at a cosmological distance. Equally impressive are the variations observed in the hyperluminous γ -ray sources observed by the Compton Gamma-Ray Observatory. Sources such as 3C 279 have been observed to vary by roughly an order of magnitude in γ -ray luminosity in a matter of days.

The case for black holes in these active galactic nuclei can be appreciated by combining two results already derived. The first is the expression for the *Eddington limiting luminosity*, L_{Edd} , derived in Sect. 14.2.2,

$$L_{\text{Edd}} = 1.3 \times 10^{31} \left(\frac{M}{M_{\odot}} \right) \text{ W}. \quad (19.6)$$

L_{Edd} is the upper limit to the luminosity which a source of mass M can have because, if the source had greater luminosity, the radiation pressure of the source would blow the star apart, or equivalently, infalling matter would be blown away. This limit can be exceeded if the source is non-stationary and for special geometries, but it provides a useful upper bound to the steady luminosity of a source of mass M .

The second result is the *causality* relation, which states that variations in the intensity of a source of size r cannot be observed to take place on time-scales less than r/c , since this is the time it takes electromagnetic waves to cross the source. The smallest size which a source of mass M can have is roughly the Schwarzschild radius $r_g = 2GM/c^2$ and hence a lower limit to the time-scale of variations from such a source is the Schwarzschild radius divided by the speed of light,

$$T \geq \frac{r_g}{c} = \frac{2GM}{c^3} \approx 10^{-5} \left(\frac{M}{M_{\odot}} \right) \text{ s}. \quad (19.7)$$

The results (19.6) and (19.7) should be used with caution. Specifically, we need to worry about the impact of unification and relativistic beaming upon the observed properties of active galaxies. It is certain that relativistic aberration effects can significantly alter the conclusions discussed below, and are essential in accounting for the superluminal radio sources and the hyperluminous γ -ray sources. Nonetheless, the combination of the causality limit (19.7) with the Eddington luminosity (19.6) can be used to illustrate how black holes can account naturally for rapid variations in the most extreme active galactic nuclei.

As an example, the expression for the Eddington luminosity (19.6) indicates that a $10^9 M_{\odot}$ black hole can have luminosity up to $\sim 10^{40}$ W and, from the causality relation (19.7), variations in this luminosity could occur on time-scales as short as about 5 hours. Except for sources in which relativistic beaming is known to be important, these constraints can accommodate most active galactic nuclei. A good example of the type of analysis which supports this picture is that of Wandel and Mushotzky (1986), referred to in Sect. 14.2.3 and illustrated by Figs 14.4a and b. For the sample of quasars and Seyfert 1 and 2 galaxies available to them at that time, the masses inferred from the causality argument, as applied to the X-ray variability of the active galactic nuclei, were in reasonable agreement over a considerable range of black hole masses, $10^6 \leq M/M_{\odot} \leq 10^{10}$, with dynamical estimates based upon the breadth of the emission lines in their optical spectra. None of the active galactic nuclei fell within the forbidden region, according to the Eddington criterion. Typically, the luminosities amount to about 1–10% of the Eddington limit (Fig. 14.4b).

Thus, supermassive black holes can account for many of the most extreme properties of active galactic nuclei and most astrophysicists are persuaded by such arguments that there are supermassive black holes in active galactic nuclei. It is important, however, to see how well the black hole picture survives closer scrutiny.

19.3 Dynamical evidence for supermassive black holes in galactic nuclei

19.3.1 Dynamical estimates of the masses of galactic nuclei

The simplest approach to making estimates of black hole masses is to use the Keplerian arguments discussed in Sect. 3.5.2 in which test objects are assumed to be in bound circular orbits with velocity v about the nucleus at radius r . Assuming a spherically symmetric mass distribution, $M(\leq r) = v^2 r / G$. There are, however, many pitfalls in attempting to make unambiguous estimates of the masses and dimensions of active galactic nuclei, as discussed by Kormendy and Turnrose (1995). There are two parts to the argument. First of all, reliable estimates are needed of the mass contained within the nuclear regions of galaxies. It is simplest to express the results in terms of the mass-to-luminosity ratio M/L of the central regions as a function of radius. These estimates can then be compared with the values typically found in galaxies to find out if there is evidence for a dark massive object in the nucleus. The second step is to consider whether or not the dark massive object is a black hole, rather than, say, a dense cluster of stars. Neither of these steps is trivial.

The velocity distributions of stars and gas clouds act as tracers of the gravitational potential within the central regions of galaxies. In the case of stars, the problem is that, in general, their velocity distributions are triaxial, meaning that we cannot assume that the components of the velocity dispersion σ_r , σ_θ and σ_ϕ are the same in the three orthogonal directions. As discussed in Sect. 3.5.3, there is compelling evidence that, in general, the velocity ellipsoid (σ_r , σ_θ , σ_ϕ) of stars in elliptical galaxies is triaxial. In the case of gas clouds, there are reasons why they may not move in circular orbits. In the vicinity of intense sources of optical-ultraviolet radiation, the gas clouds are subject to the effects of radiation pressure as well as being buffeted by shock waves and jets of relativistic material ejected from the nucleus. It is necessary to be convinced that the tracers of the velocity field provide reliable evidence about the gravitational potential distribution in the vicinity of the galactic nucleus.

Kormendy and Turnrose (1995) and Faber (1999) discuss the problems of determining the mass distribution $M(< r)$ of a spherically symmetric stellar distribution in terms of the mean circular velocity about the centre $V(r)$ and the velocity dispersions in the three spherical polar directions with respect to the nucleus, σ_r , σ_θ and σ_ϕ at radius r ,

$$M(< r) = \frac{V(r)r^2}{G} + \frac{\sigma_r^2 r}{G} \left[-\frac{d(\log \varrho)}{d(\log r)} - \frac{d(\log \sigma_r^2)}{d(\log r)} - \left(1 - \frac{\sigma_\theta^2}{\sigma_r^2}\right) - \left(1 - \frac{\sigma_\phi^2}{\sigma_r^2}\right) \right]. \quad (19.8)$$

This expression is derived by Binney and Tremaine (2008) from the first moment of the collisionless Boltzmann equation, which describes the radial momentum balance of a self-gravitating equilibrium stellar system. It is assumed that the axes of the velocity ellipsoid point along the r, θ, ϕ directions. ϱ is the density of the population of objects which are used to trace the gravitational potential and it is assumed that they do not contribute to the mass of the Galaxy. For example, ϱ could refer to the density of hot gas or stars. It is assumed that these tracers take up an equilibrium distribution within the gravitational potential defined by $M(< r)$.

We can understand the significance of the various terms in (19.8) by considering some special cases.

- (i) If the dynamics are dominated by rotation, the first term on the right-hand side of (19.8) is dominant and we recover the usual Keplerian relation discussed above.
- (ii) If there is no rotation, $V = 0$, (19.8) bears a striking resemblance to (4.19) derived in Sect. 4.4 for estimating the overall mass distribution of a cluster of galaxies from the distribution of its bremsstrahlung X-ray emission. In that case, we began with the equation of hydrostatic support for an atmosphere with density and temperature distributions $\varrho(r)$ and $T(r)$, respectively, the hot gas acting as a tracer of the gravitational potential. The equation (4.19) is

$$M(< r) = -\frac{kTr^2}{G\mu m_H} \left[\frac{d(\log \varrho)}{dr} + \frac{d(\log T)}{dr} \right]. \quad (19.9)$$

In the case of a gas, the pressure is isotropic and so the velocity dispersion of the particles is also isotropic, $\sigma_r = \sigma_\theta = \sigma_\phi$ and $\sigma^2 = 3\sigma_r^2$. Furthermore, for a perfect gas, $\frac{1}{2}\mu m_H \sigma^2 = \frac{3}{2}\mu m_H \sigma_r^2 = \frac{3}{2}kT$ and $d(\log \sigma_r^2) = d(\log T)$. Therefore, (19.9) can be rewritten

$$M(< r) = -\frac{\sigma_r^2 r}{G} \left[\frac{d(\log \varrho)}{d(\log r)} + \frac{d(\log \sigma_r^2)}{d(\log r)} \right]. \quad (19.10)$$

This expression is identical with (19.8) for an isotropic velocity distribution.

- (iii) The last two terms in round brackets in (19.8) represent the anisotropy of the velocity ellipsoid. If the orbits of the particles were purely radial, $\sigma_\theta = \sigma_\phi = 0$ with no bulk rotation, $V = 0$, (19.8) can be rewritten as follows,

$$M(< r) = \frac{\sigma_r^2 r}{G} \left[-\frac{r}{\varrho} \frac{d\varrho}{dr} - \frac{d(\log \sigma_r^2)}{d(\log r)} - 2 \right] = -\frac{1}{G\varrho} \frac{d}{dr} (\varrho r^2 \sigma_r^2), \quad (19.11)$$

where ϱ represents the radial equilibrium distribution of the *tracers* of the mass distribution as a function of r . Multiplying both sides by $4\pi G\varrho$, we find

$$\frac{4\pi r^2 \varrho GM(< r)}{r^2} dr = \frac{GM(< r)}{r^2} dM_t = -d(4\pi r^2 \varrho \sigma_r^2). \quad (19.12)$$

The expressions in the first two parts of the equality represent the inward gravitational force acting on a shell of mass $dM_t = 4\pi r^2 \varrho dr$ of the tracers of the gravitational field. The third part represents the outward pressure gradient associated with the velocity dispersion in the radial direction. In other words (19.12) is the one-dimensional radial

analogue of (19.9). In the one-dimensional case, the pressure only acts in the radial direction and so, using the equipartition theorem, $\frac{1}{2}Nm\sigma_r^2 = \frac{1}{2}NkT$, where N is the number density of the trace particles of mass m . Since $p = NkT$, it follows that $p = \rho\sigma_r^2$.

The above analysis refers to the actual space densities and velocity dispersions within the galaxy. In making observations of active galaxies, these quantities are not directly observable, but have to be inferred by deprojecting and deconvolving the observed velocity profiles along the line of sight, which is not a trivial task.

19.3.2 Examples of estimates of the masses of active galactic nuclei

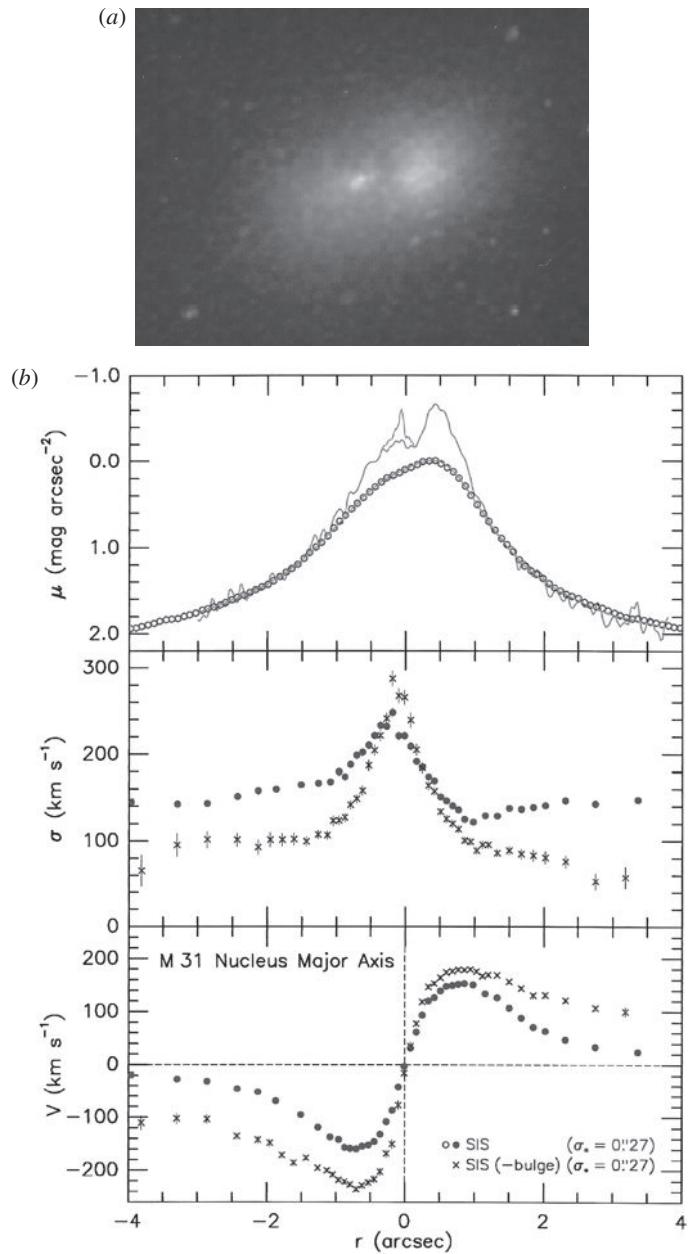
The problems of interpretation are best illustrated by considering a few examples of galaxies which have been studied in some detail. As pointed out by Kormendy and Richstone (1995), there is a dichotomy, in that the galaxies for which the highest spatial resolution is available are nearby objects, whereas the most extreme active galactic nuclei are distant objects. The study of nearby systems is, however, of the greatest importance since they can be used to determine the general prevalence of dark massive objects in the nuclei of galaxies in general, whether or not they contain active galactic nuclei.

M31

The nearest giant spiral galaxy to our own Galaxy is M31, the central regions of which, as observed by the Hubble Space Telescope, are shown in Fig. 19.1a. Originally, it was thought that the structure consisted of a double nucleus, but in fact the stellar distribution is in the form of a disc with somewhat enhanced stellar density in its right-hand portion. The HST images show that there is a compact ultraviolet star cluster close to the centre of the disc. Kormendy and Bender (1999) carried out a detailed photometric and kinematic study of the disc with the results shown in Fig. 19.1b in which the surface brightness μ , velocity dispersion σ and rotation velocity V are all measured along the major axis. As they express it, ‘all spectra confirm the steep rotation and velocity dispersion gradients that imply that M31 contains a $3.3 \times 10^7 M_\odot$ central dark object’. If the region about the dynamical centre of the disc is modelled as a cylinder of radius 0.13 arcsec, the observed mass-to-light ratio is $M/L_V \approx 300$; if it is taken to be a sphere of radius 0.13 arcsec, the value of $M/L_V \approx 2200$. The dynamical centre of the disc where these very large values of M/L_V are found is close to the centre of the ultraviolet star cluster. The most natural explanation is that a black hole of mass $3.3 \times 10^7 M_\odot$ is present in the nucleus of M31.

M87 or NGC 4486

M87 is one of the nearest giant elliptical galaxies which displays many of the characteristics of the most extreme active galaxies (Fig 3.2a). These features include a source of intense non-thermal radiation in its nucleus, a one-sided optical/radio jet with knots which move out

**Fig. 19.1**

- (a) An HST image of the nuclear regions of M31 showing the disk of stars and the nuclear ultraviolet star cluster.
- (b) The variation of surface brightness μ , velocity dispersion σ and rotational velocity V along the major axis of the nuclear regions of M31 (Kormendy and Bender, 1999). The dynamical centre coincides closely with the location of the nuclear ultraviolet star cluster.

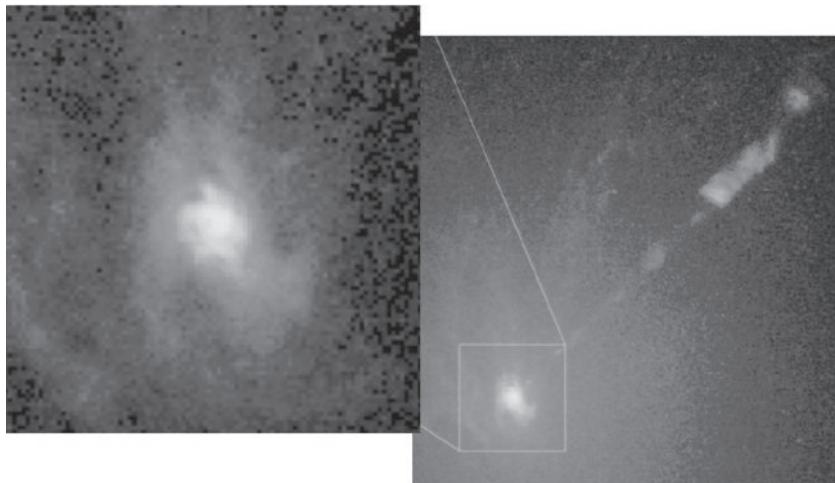


Fig. 19.2 An HST image of the nuclear regions of M87 showing the central gaseous disc with axis parallel to the optical jet (Ford *et al.*, 1994).

from the nucleus at a relativistic speed, and a double-lobed radio structure with characteristic FR1 morphology.

Pioneering studies of the velocity dispersion of the stars in the nuclear regions were carried out by Young, Sargent and their colleagues (Young *et al.*, 1978; Sargent *et al.*, 1978) who found that it increases towards the nucleus. Assuming that the velocity dispersion is isotropic, a central mass of $3 \times 10^9 M_\odot$ was inferred. This proved to be a controversial result, however, as there is negligible rotation of the stars about the nucleus and so the mass estimate depended strongly upon the assumptions made about the isotropy of the velocity ellipsoid of the stars. Illingworth (1977), Binney (1978) and others showed how the observations could be accounted for in terms of an anisotropic velocity ellipsoid, in which the radial velocity dispersion becomes greater close to the nucleus. This model removed the necessity of a massive black hole in the nucleus.

A complementary approach was taken by Ford and his colleagues (1994) who discovered a disc of ionised gas about the nucleus of M87 from HST imaging observations (Fig. 19.2). As can be seen from that image, the optical jet lies along the axis of the nuclear disc. Spectroscopy by Harms and his colleagues (1994) established that the disc is in Keplerian rotation about the nucleus, the inferred mass of the nuclear regions being $3 \times 10^9 M_\odot$. The agreement between the mass inferred from the motions in the ionised disc and the estimates of Sargent and his colleagues (1978) is reassuring – it would be surprising if this agreement were fortuitous.

M106 or NGC 4258

One of the problems with spectroscopic studies of active galactic nuclei is finding tracers which can probe as closely as possible to the black hole itself. The best that can be achieved

currently in the optical waveband is the diffraction limit of the Hubble Space Telescope, about 0.1 arcsec. A key discovery for achieving up to two orders of magnitude greater angular resolution has been the use of water-maser emission from the nuclear regions of active galaxies. A case of special importance is the nearby galaxy NGC 4258 which is classed as a LINER and which possesses jets of synchrotron radio emission. It is a member of a class of extragalactic maser sources known as *megamasers*, which can have luminosities up to about a million times those of Galactic maser sources. Within our own Galaxy, water masers are formed in relatively dense gas clouds containing molecules and dust, either in star-forming regions, or in the envelopes of late-type stars.

In a remarkable set of observations made with the Very Long Baseline Array (VLBA) operating at the H₂O maser wavelength of 1.3 cm, Miyoshi and his colleagues (1995) mapped the location and velocities of the H₂O masers with an angular resolution better than a milliarcsecond. A great advantage of these maser lines is that they are very narrow, typically only about 1 km s⁻¹ in width. Maser emission is only observed if the radiation passes through a sufficiently long path length of the masing medium without the Doppler shift of the moving clouds shifting the wavelength of the radiation from the masing wavelength. Hence, the maser emission must originate from regions in which the velocity gradient is zero. This occurs in regions in which the molecular ring is observed tangentially, and also along the line of sight directly towards the nucleus. A cartoon of the location of the maser sources in the molecular ring about the nucleus is shown in Fig. 19.3. The rotation curve derived from these observations is shown in Fig. 19.4.

If θ is the angle between a line drawn from the nucleus to a cloud in the ring and the line of sight to the observer, the projected velocity of the cloud, shown as Doppler shifted features in the spectrum at the bottom of Fig. 19.3, is $v_0 \sin \theta$, where v_0 is the speed of rotation of the ring. The high velocity masers observed on either side of the nucleus correspond to $\theta = \pm\pi/2$ and so measure directly the rotational velocity of the disc at different radii from the centre. In the case of the masers observed in the direction of the radio continuum source in the nucleus, for small θ , the projected velocity is $v_0\theta = v_0 l/R$, where R is the radius of the ring and l is the distance measured round the ring. This expression accounts for the linear proportionality of the projected velocity with l seen in the insert in Fig. 19.4.

Another way of determining the kinematics of the ring is to follow the change in velocity of individual spectral features with time. This has been achieved by long term monitoring of the zero-velocity features in the H₂O spectrum with the Effelsberg 100-metre Telescope (Greenhill *et al.*, 1995a). If the clouds responsible for the maser emission are in circular orbits about the nucleus, they are centripetally accelerated as they pass in front of the nucleus. The individual maser velocities have been observed to drift by about 9 km s⁻¹ y⁻¹, corresponding to the *centripetal acceleration* of the molecular clouds in the ring towards the nucleus and this acceleration is consistent with the rotation curve shown in Fig. 19.4 and the structure shown in Fig. 19.3. It was unquestionably a stroke of good fortune that the nuclear molecular disc in NGC 4258 was discovered since the disc has to be almost precisely in the correct orientation with respect to the line of sight before megamasers could be observed. Miyoshi and his colleagues (1995) estimate that there is only about a 6% chance that the orientation of the molecular ring would be such that maser emission would be observable.

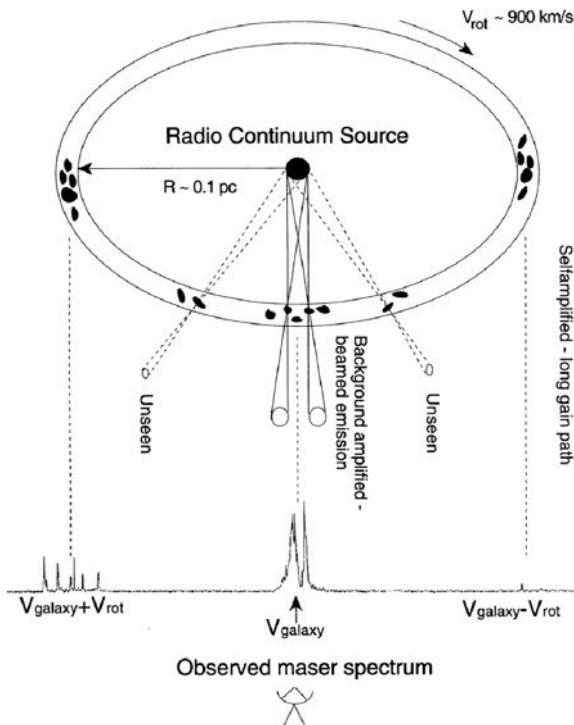
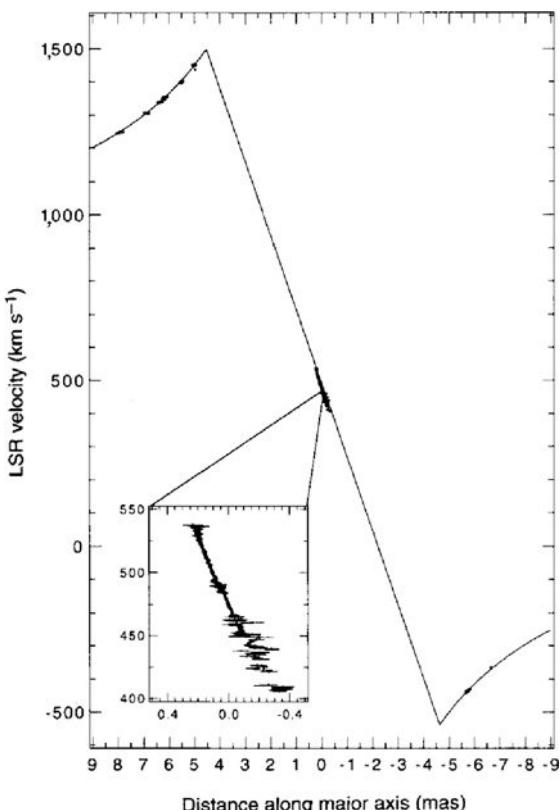


Fig. 19.3 The location of the H_2O megamasers in the nuclear regions of NGC 4258 (Greenhill *et al.*, 1995b), illustrating the origin of the high velocity components observed in the molecular line spectrum of the source, which is shown at the bottom of the diagram. The view is from slightly above the plane containing the disc and the Earth. The masing region occupies a thin annulus with fractional thickness of 0.3 or less. The masers with velocities roughly equal to the recession velocity of the galaxy become visible when they pass in front of the continuum source, the diameter of which is estimated to be about 0.014 pc.

The mass of the central dark object is found to be about $3.7 \times 10^7 M_\odot$, the corresponding mass density being greater than $4 \times 10^9 M_\odot \text{ pc}^{-3}$. This figure is about 40 000 times greater than the densities of globular clusters, so high that it is unlikely that the mass could be a cluster of known types of stellar object. If the dark mass consisted of a cluster of stars, the number density would be so great that it would evaporate by dynamical friction as described in Sect. 5.7. According to Faber (1999), the evaporation time for a cluster of stars each of mass M_* from the nuclear region in NGC 4258 would be

$$t_{\text{evap}} = 1.5 \times 10^8 \left(\frac{M_*}{1.4 M_\odot} \right)^{-1} \text{ years}, \quad (19.13)$$

assuming the observed mass and size of the dark nucleus. Thus, if the dark mass consisted of, say, neutron stars with mass $1.4 M_\odot$, the cluster would have evaporated long ago. Only if the objects were of low mass $M \leq 0.03 M_\odot$ could they survive for the age of the galaxy. If these objects were, say, planets or brown dwarfs, the collision time between them would be very short, about 10^5 years, and so the cluster would rapidly evolve towards the formation

**Fig. 19.4**

The rotation curve of the nuclear regions of the Galaxy NGC 4258 from VLBI observations of the water vapour maser line at 1.3 cm. Outside the inner ± 3 milliarcseconds, the radial velocities follow a Keplerian rotation law, indicating the presence of a massive compact object in the nucleus (Miyoshi *et al.*, 1995).

of a massive black hole in any case. It seems inevitable that there is a black hole with mass $3.7 \times 10^7 M_\odot$ in the centre of NGC 4258.

Another intriguing feature of these observations is that they are direct observational evidence for a dusty molecular ring, or torus, orbiting an active galactic nucleus. Furthermore, the radio jets are oriented perpendicular to the plane of the ring. In this case, the disc, or torus, is very thin and does not obscure the nuclear regions, despite the fact that it is observed almost edge-on.

The black hole in the Galactic centre

The centre of our own Galaxy is the closest example we have of a galactic nucleus. The optical extinction to the nuclear regions amounts to about 30 magnitudes but, because of the strong dependence of extinction upon wavelength, the central regions of the Galaxy can be observed in the infrared waveband (Sect. 12.4). A key observation was the discovery of the compact variable non-thermal radio source Sagittarius A* (Sgr A*) in the very centre of the Galaxy. VLBI observations at 1.3 mm wavelength have established that the source

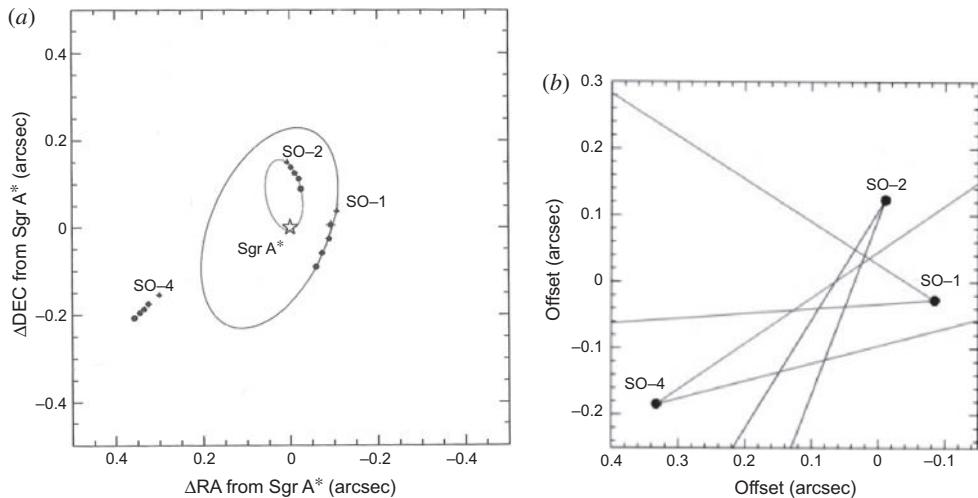


Fig. 19.5

- (a) The orbits of stars about the Galactic Centre determined by measurements of their proper motions over a period of four years using near-infrared speckle techniques at $2 \mu\text{m}$ with the Keck 10 metre optical-infrared telescope.
- (b) Estimates of the directions of the acceleration vectors of the infrared stars SO-1, SO-2 and SO-4, the opening angles indicating the uncertainty in the measured directions. The most likely location of the central attractor includes the non-thermal radio source Sgr A* (Ghez *et al.*, 2000).

is 37 microarcseconds in angular size, assuming a circular Gaussian model for the source (Doeleman *et al.*, 2008). Taking the distance of the Galactic Centre to be 8 kpc, this angular scale corresponds to 1.5×10^{-6} pc, or 0.1 AU in physical size. Infrared observations of the distribution of stars in the vicinity of Sgr A* have shown that the radio source coincides with the dynamical centre of this mass distribution and that the velocity dispersion increases towards Sgr A* according to a Keplerian law (see the survey of Melia and Falcke, 2001).

Equally impressive are the observations in the infrared waveband at $2 \mu\text{m}$ of the motions of stars about the Galactic nucleus. The first observations to be published were those made with the Keck 10 metre telescope by Ghez and her colleagues with a very high speed infrared camera designed for speckle observations (Ghez *et al.*, 2000). Those frames which were diffraction limited were selected and co-added. From observations over a period four years, the orbits of stars within half an arcsec of the compact radio source Sgr A*, which is assumed to define the Galactic nucleus, were measured (Fig. 19.5a). The acceleration vectors of three infrared stars were determined and, within the uncertainties with which these were determined, the centre of gravitational attraction coincided with the position of Sgr A* (Fig. 19.5b) (Ghez *et al.*, 2000). These data implied that there is a dark object of mass between $(2.3 \text{ and } 3.3) \times 10^6 M_\odot$ within a radius of 0.015 pc of the nucleus of our Galaxy. The lower limit to the mass density within this region was $8 \times 10^{12} M_\odot \text{ pc}^{-3}$, far exceeding the mass density inferred for the nucleus of NGC 4258. The Schwarzschild radius of a black hole of mass $2.6 \times 10^6 M_\odot$ is 2.6×10^{-7} pc and so the size of the radio source Sgr A* as measured by VLBI is about six times the Schwarzschild radius of the central black hole.

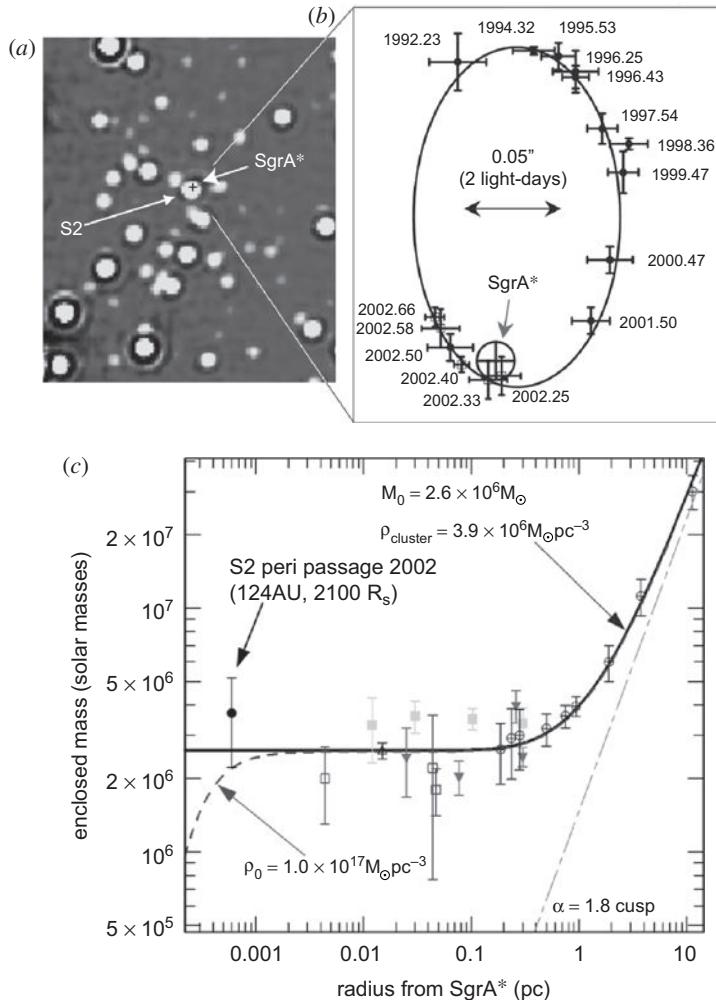


Fig. 19.6

(a) An image of the 2×2 arcsec² field centred on the compact radio source Sgr A* which is assumed to be located in the nucleus of our Galaxy. (b) The orbit of the infrared star S2 (the same as SO-2 in Fig. 11.5a), as determined by infrared observations taken with the 3.5 m New Technology Telescope at the ESO La Silla Observatory and the 8.2 m YEPUN telescope of the ESO VLT at Cerro Paranal. (c) The mass distribution in the Galactic Centre determined from the kinematics of infrared stars orbiting the central black hole (Schödel *et al.*, 2002).

Even more remarkable is the series of observations by Genzel and his colleagues using the facilities of the European Southern Observatory. Ten years of high resolution astrometric imaging enabled Genzel's team to trace two-thirds of the orbit of the star labelled SO-2 in Fig. 19.6a about Sgr A* (Schödel *et al.*, 2002). These observations show that the star is moving in a bound, highly elliptical orbit with an orbital period of 15.2 years and that Sgr A* lies in one focus, as expected from Kepler's laws of planetary motion. The distance of closest approach of the star SO-2 to the nucleus is only 17 light hours, or 5×10^{-4} pc,

corresponding to about 100 times the Schwarzschild radius of the black hole. The motions of numerous infrared stars have enabled the mass distribution in the Galactic nucleus to be determined (Fig. 19.6c). The inferred mass of the black hole is $2.6 \times 10^6 M_\odot$ and the lower limit to the mass density from observations of SO-2 is $10^{17} M_\odot \text{ pc}^{-3}$. The same set of infrared observations revealed the very faint infrared counterpart of Sgr A*. ‘Flares’ have been discovered, probably associated with the accretion of mass onto the central black hole (Genzel *et al.*, 2003).

19.4 The Soltan argument

How common are black holes in the nuclei of galaxies? Constraints on the typical masses of the black holes in galaxies can be found from an important argument due to Soltan (1982). The analysis begins with the expression for the maximum luminosity of a black hole powered by accretion (see Sect. 14.2.1), $L = \xi \dot{m} c^2$, where \dot{m} is the rate at which mass is accreted by the black hole and ξ is the efficiency of the conversion of rest-mass energy into radiation as the matter falls into the black hole. The maximum values of ξ are 0.06 for a Schwarzschild black hole and 0.42 for a maximally rotating Kerr black hole. Soltan realised that the total luminosity emitted by all quasars in the Universe provides an estimate of the amount of mass which must have collapsed to form massive black holes over the range of redshifts at which quasars are observed.

Suppose the *bolometric luminosities* L of quasars are wholly the result of the accretion of matter onto black holes – the same process also results in the growth of their masses. If the source radiates at luminosity L for a time T , the total amount of mass consumed by the black hole is

$$m = \int \dot{m} dt = \frac{1}{\xi c^2} \int L dt = \frac{1}{\xi c^2} LT . \quad (19.14)$$

We now integrate LT over the luminosity function of all quasars at all epochs. The luminosity function of quasars evolves strongly with cosmic epoch and it is simplest to write this evolution in *comoving coordinates* as $n(L, z) dL$, meaning the number density of sources of luminosity L in a coordinate system which removes the expansion of the system of coordinates.¹ Consequently, the total mass per unit volume which must have collapsed into black holes is at least

$$\varrho_{\text{BH}} = \frac{1}{\xi c^2} \int \int L n(L, z) dL dt , \quad (19.15)$$

where t is cosmic time and the relation between t and redshift z depends upon the cosmological model.

Next, we relate the double integral (19.15) to the number counts of quasars. Using the standard relations between intrinsic properties and observables, the number of quasars with

¹ For details of the calculations and notation used in this section, see my book *Galaxy Formation* (Longair, 2008).

luminosity L in the redshift interval z to $z + dz$ with flux densities in the range S to $S + dS$ is

$$N(S, z) dS dz = n(L, z) dL D^2 dr , \quad (19.16)$$

where r is the comoving radial distance coordinate and D is the ‘distance measure’, which is related to the luminosity distance D_L by $D(1+z) = D_L$. For bolometric luminosities and flux densities, the flux density–luminosity relation is

$$L = 4\pi D_L^2 S = 4\pi D^2 (1+z)^2 S . \quad (19.17)$$

Substituting for $n(L, z) dL$ from (19.16) and for L from (19.17), and using the relation between the increment of cosmic time dt and the increment of comoving distance coordinate dr , $dr = c dt (1+z)$, we find

$$\begin{aligned} \varrho_{\text{BH}} &= \frac{1}{\xi c^2} \int \int 4\pi D^2 (1+z)^2 S \frac{N(S, z) dS}{D^2} \left(\frac{dt}{dr} \right) dz \\ &= \frac{4\pi}{\xi c^3} \int \int (1+z) S N(S, z) dS dz . \end{aligned} \quad (19.18)$$

This is the result derived by Soltan (1982). The average density of mass in the Universe in the form of massive black holes is determined by integrals over the *observed* number–flux density relation for quasars and the *observed* redshift distribution in each flux density interval.

This is the easy part of the exercise – the more difficult part is to estimate the bolometric luminosity function for quasars and their redshift distribution, or mean redshift, at each flux density. The more difficult of these is the conversion from the number counts of quasars in the form of a number–apparent magnitude relation in, say, the B waveband $N(B) dB$ to the bolometric luminosity emitted over all wavelengths. We need to make a number of rough approximations. First of all, most of the quasars in complete samples lie in the redshift interval $1 \lesssim z \lesssim 3$. The factor $(1+z)$ is therefore typically about 3 and it is not a bad approximation to take the $(1+z)$ factor outside the integral. Next, the function $\int S N(S) dS$ can be found from counts of radio-quiet quasars. Fortunately, the counts of quasars converge at magnitudes $B \geq 21$ and so this integral can be worked out directly from the number counts. There only remains the problem of converting luminosities in the B waveband to bolometric luminosities over all wavebands. Since the continuum spectra of quasars extend from the millimetre to the hard X-ray and γ -ray wavebands, this bolometric correction can be quite large. Krolik (1999) adopts the device of guessing that a factor of 10 is not implausible and then including a factor of $(F_{\text{bol}}/10F_B)$ in the final result, where F_{bol}/F_B is the bolometric correction for the optical waveband. Finally, it is useful to normalise the mass density in black holes to the number density of L^* galaxies, thus obtaining an estimate of the typical black hole mass in an L^* galaxy.

Krolik (1999) provides a convenient relation which can be derived from Soltan’s analysis:

$$\langle M_{\text{bh}} \rangle = 1.6 \times 10^7 \left(\frac{F_{\text{bol}}}{10F_B} \right) \left(\frac{\langle 1+z \rangle}{3} \right) \times \left(\frac{h}{0.75} \right)^{-3} \left(\frac{\xi}{0.1} \right)^{-1} M_{\odot} \text{ per } L^* \text{ galaxy} . \quad (19.19)$$

For reasonable values of the quantities in round brackets in (19.19), that is, setting them all equal to unity, the typical L^* galaxy should possess a black hole remnant of mass about $10^7 M_\odot$. This argument suggests that, on average there must be massive black holes in most massive galaxies if we are to explain the observed numbers of quasars. This result is consistent with evidence on the masses of black holes in nearby galaxies with $L \sim L^*$ discussed in Sect. 19.3.

There is one other remarkable feature of (19.19) – no assumption was made about the distances of the objects counted since the integrals are taken over the flux densities and number counts of the quasars in the samples. This argument shows that it is very natural that there should be massive black holes in the nuclei of massive galaxies.

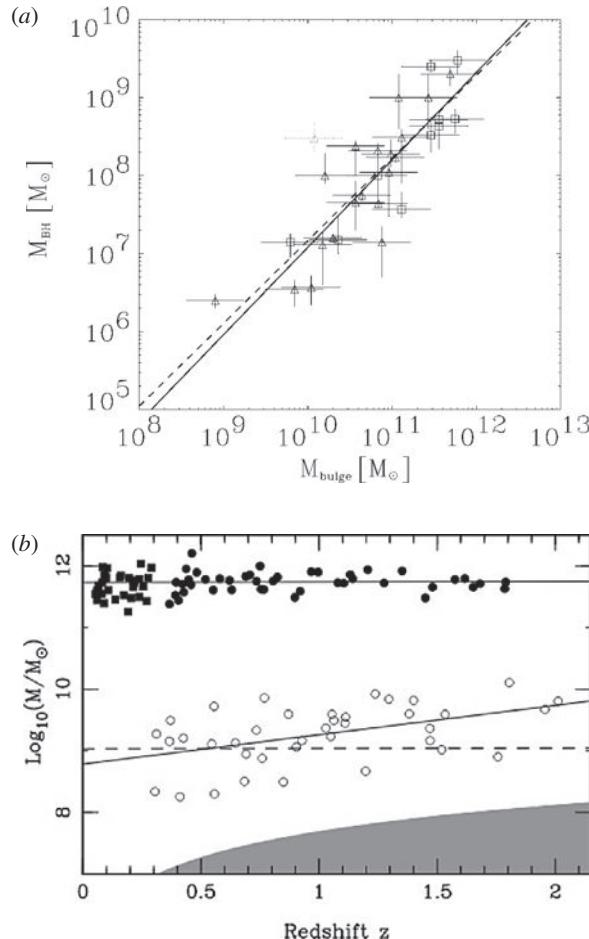
19.5 Black holes and spheroid masses

The argument of the last section indicates that there must be massive black holes in the nuclei of luminous galaxies. Granted the many caveats which they made about the precise determination of black hole masses in galactic nuclei, Kormendy and Richstone (1995) plotted their best estimates of the black hole masses against the absolute magnitude of their spheroids, a measure of their stellar masses, for the nearby galaxies for which the mass estimates were reasonably secure. They found a correlation between these quantities, but warned of the many selection effects which might induce such a correlation.

Magorrian and his colleagues (1998) carried out a similar analysis using a larger sample of nearby galaxies and found a strong correlation. The mass estimates of the black holes and spheroid stellar masses were improved by Häring and Rix (2004) who showed that there is an almost linear relation between these quantities, $M_{\text{bh}} \propto M_{\text{bulge}}^{1.12 \pm 0.06}$ (Fig. 19.7a). Typically, the black hole mass amounted to about 0.1–0.2% of the stellar mass of the bulge.

Another approach is to adopt the unification scenario for radio galaxies and radio quasars in bright radio source samples. The success of the orientation-based unification scheme for 3CR radio galaxies and radio quasars means that they can be considered members of the same parent population. McLure and his colleagues (2006) derived the black hole masses from the breadths of the quasar emission lines, assuming the gas clouds to be in virial equilibrium, and the stellar masses of the bulges were found from modelling the ultraviolet-optical-infrared spectral energy distributions of the radio galaxies. The results of this analysis are shown in Fig. 19.7b, in which the $M_{\text{bh}}/M_{\text{sph}}$ ratio is again about 0.2%. There is some suggestion that the ratio decreases with decreasing redshift, as if the black holes had stopped growing while the bulges continued to increase in mass.

Understanding these relations is a challenging problem in the theory of galaxy evolution. Apparently, the growth of black holes cannot be dissociated from the formation and evolution of the galaxies themselves. In particular, active galactic nuclei must be involved in the feedback processes which are necessary to prevent the continued growth of giant elliptical galaxies to small redshifts, the problem of ‘galaxy downsizing’. Several other aspects of the evolution of galaxies and clusters seem to involve feedback in the sense that the formation of the central black hole influences the large scale properties of these systems.

**Fig. 19.7**

(a) A plot of the mass of black holes in the nuclei of nearby galaxies against the stellar masses of their bulges, or spheroids, from the analysis of Häring and Rix (2004). (b) A plot of the 3CRR radio galaxy stellar masses (filled circles) and the 3CRR quasar black hole masses (open circles) as a function of redshift. The solid lines are least squares fits to the $M_{\text{sph}} - z$ and $M_{\text{bh}} - z$ relations. The dashed line shows the 3CRR quasar black hole masses assuming that the $M_{\text{bh}}/M_{\text{sph}}$ ratio has the same value as at the present epoch. In the shaded grey area, 3CRR quasar black hole masses are excluded by the radio flux limit (McLure *et al.*, 2006).

19.6 X-ray observations of fluorescence lines in active galactic nuclei

One of the great observational challenges is to study the behaviour of matter in the strong gravitational fields close to the black hole. There is now compelling evidence that this can be achieved using X-ray spectral observations of active galactic nuclei. The key observations concern the asymmetric iron fluorescence lines observed in the X-ray spectra of Seyfert 1 galaxies. The most convincing case involves the Seyfert 1 galaxy MCG -6-30-15 which

has already been discussed in the context of the rapid variability of its X-ray emission (Fig. 14.3b). There are significant variations in its X-ray flux density on the time-scale of hours, indicating that the emission originates very close to the nucleus itself.

It is simplest to discuss the observations in the context of the thin accretion disc model for the fuelling of the active galactic nucleus. As discussed in Sect. 14.2.3., according to the standard picture of a thin accretion disc, the temperature at its inner edge is roughly

$$T \approx 2 \times 10^7 \left(\frac{M}{M_\odot} \right)^{-1/4} \text{ K}. \quad (19.20)$$

To account for the luminosities observed in active galactic nuclei, the masses of the black holes must be greater than about $10^7 M_\odot$ and so the temperature of the thin discs would be at most $T \sim 10^6$, meaning that the disc itself cannot be the source of the X-rays. Rather, the X-ray emission must be created by some other process and a popular picture is one in which the accretion disc is embedded in an X-ray emitting halo which is the source of the X-ray continuum radiation. The observed X-ray spectrum can be taken to be of power-law form $I(\varepsilon) \propto \varepsilon^{-1}$ in the 1–10 keV energy interval, where the intensity is expressed in power per unit energy interval. The X-ray line emission observed in the spectra of Seyfert 1 galaxies is attributed to the fluorescent emission of the ions present in the accretion disc.

19.6.1 Fluorescent X-ray lines

Details of the fluorescent emission process are given by Fabian (1998) and a more detailed treatment by Liedahl (1999). Since the material of the disc is much cooler than the surrounding X-ray emission, elements such as iron are not fully ionised and, although they may be partially ionised, these ions still possess filled K- and L-shell electrons. The first part of the process is the absorption of an X-ray by the iron ion, the greatest cross-section being for removal of one of the two K-shell electrons. The energy threshold for this process is 7.1 keV for neutral iron and increases as the iron ions become more and more stripped of electrons. There are then two outcomes to the process by which the vacancy in the K-shell is filled by an L-electron. Either an X-ray photon of energy 6.4 keV is emitted in a permitted electromagnetic transition, or the energy is used to eject an electron from the L-shell, a process known as the *Auger effect*. The probabilities of these processes occurring are 0.34 for the emission of a 6.4 keV photon and 0.66 for the ejection of an Auger electron.

Matt, Fabian and Reynolds (1997) have computed the predicted reflection spectrum from the surface layers of a cool accretion disc which is shown in Fig. 19.8, from which it can be seen that the 6.4 keV fluorescent line of iron is by far the strongest line. The great strength of the iron line is due to the combination of the large cross-section for photoelectric absorption by the K-shell of iron and iron's large cosmic abundance. The importance of these considerations for black hole physics is that the 6.4 keV fluorescent iron line acts as a tracer of the velocity field in the accretion disc. In particular, the accretion disc extends into regions of the disc in which special and general relativistic effects are large and these can strongly influence the shape of the line profile.

The best example to date of the relativistic broadening of the 6.4 keV line is found in the spectrum of the Seyfert 1 galaxy MCG -6-30-15. Figure 19.9 shows the spectrum of

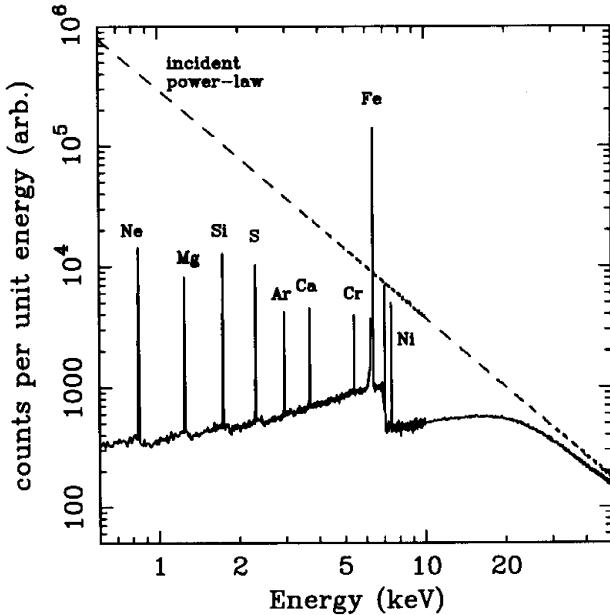


Fig. 19.8 The reflection spectrum produced by a power law of photon index 2, corresponding to $I(\varepsilon) \propto \varepsilon^{-1}$, incident upon a slab of material with the cosmic abundances of the elements (Fabian, 1998, after Matt, Fabian and Reynolds, 1997).

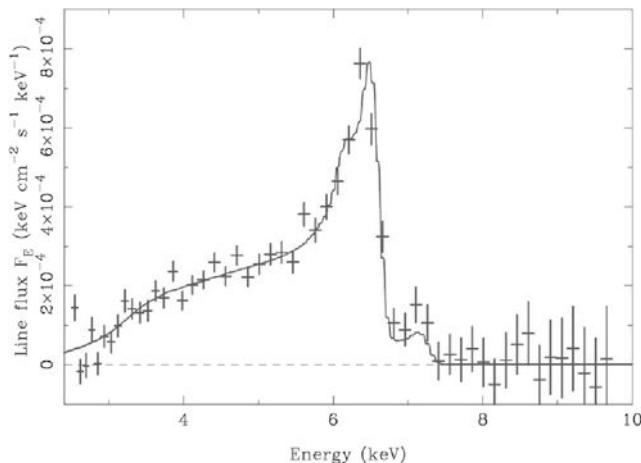


Fig. 19.9 The broad asymmetric iron fluorescent line as observed in a long observation of the Seyfert 1 galaxy MCG -6-30-15 by the EPIC instrument of the XMM-Newton Observatory (Fabian *et al.*, 2002). A smooth continuum spectrum has been subtracted from the overall spectrum to reveal the profile of the line. The best-fit spectrum corresponds to the disc having an angle of inclination $i = 30^\circ$ and an inner radius $r = 2GM/c^2$. The disc emissivity is described by a broken power law in radius, where the break radius occurs at $r = 6GM/c^2 = 3r_g$. Beyond this radius, the disc has an emissivity profile $q \propto r^{-2.5}$ resulting in an equivalent width of 200 eV, corresponding to the core of the line. Within $3r_g$ the emissivity steepens, resulting in the strong low energy tail to the line.

the line as observed in a long integration by the EPIC instrument of the XMM-Newton satellite (Fabian *et al.*, 2002). The profile of the line is found by subtracting a smooth continuum spectrum from the overall spectrum. The key features are the rather abrupt cut-off at energies greater than 6.7 keV and the extension of the line to energies less than 4 keV to the low energy side of the line. This asymmetry occurs naturally if the thin accretion disc extends inwards towards the last stable orbit about the black hole. There are two relativistic effects which lead to the broadening of the line. The first is the gravitational redshift which shifts the spectrum to lower X-ray energies and the second is the transverse Doppler effect of photons emitted from a source moving at an angle θ to the line of sight (see Sect. 9.3),

$$\varepsilon_{\text{obs}} = \frac{\varepsilon_0}{\gamma \left(1 - \frac{v}{c} \cos \theta\right)}. \quad (19.21)$$

Thus, if the plane of the accretion disc lies at a large angle to the line of sight, $\theta \rightarrow \pi/2$, the transverse Doppler shift associated with the γ in the denominator dominates and redshifts the line to lower energies. Thus, the asymmetric profile of the 6.4 keV line has a natural interpretation if the lowest energy parts of the line originate close to the last stable orbit of an accretion disc about a massive black hole and the disc is observed more or less face-on. This would be in accord with the unified picture of active galactic nuclei since, in Seyfert 1 galaxies, the nuclear regions are observed directly and, if the accretion disc is in the same plane as the torus, the plane of the accretion disc would be observed quite close to the plane of the sky.

19.6.2 Pedagogical interlude – circular velocities about black holes

As an example, let us work out the Doppler and gravitational redshifts from the last stable orbit of a Schwarzschild black hole, which lies at $r = 3r_g$, where $r_g = 2GM/c^2$ is the Schwarzschild radius. We recall that the Schwarzschild metric is

$$ds^2 = dt^2 = dt'^2 \left(1 - \frac{2GM}{rc^2}\right) - \frac{1}{c^2} \left[\frac{dr^2}{\left(1 - \frac{2GM}{rc^2}\right)} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (19.22)$$

The meaning and interpretation of the metric was discussed in some detail in Sect. 13.11. As shown in that section, dt is the time increment of an observer at rest at an infinite distance from the black hole, what is referred to as *coordinate time*. The first term in the metric on the right-hand side of (19.22) is the time increment dt' measured at radial coordinate r ,

$$dt' = dt \left(1 - \frac{2GM}{rc^2}\right)^{1/2}. \quad (19.23)$$

This expression describes quantitatively how the rate at which time passes depends upon the gravitational potential. Thus, the period T_0 of an electromagnetic wave emitted at the

point r in the gravitational field becomes T_∞ when the wave propagates to $r = \infty$ and so

$$T_0 = T_\infty \left(1 - \frac{2GM}{rc^2}\right)^{1/2}; \quad v_\infty = v_0 \left(1 - \frac{2GM}{rc^2}\right)^{1/2}. \quad (19.24)$$

The second expression in (19.24) is the *gravitational redshift* from any radius $r \geq r_g$. Since the last stable orbit lies at $r = 3r_g$, it follows that the gravitational redshift amounts to $v_\infty/v_0 = \sqrt{2/3} = 0.816$.

The second effect is the relativistic Doppler shift associated with circular motions in the accretion disc about the black hole. The simplifying feature of the thin accretion discs discussed in Sect. 14.3 is that the particles move in Keplerian orbits about the centre. Therefore, we need only find the relativistic equivalents of Kepler's laws for test particles in circular orbits about the black hole from (19.22). For motion in a plane, we set $d\phi = 0$.

The simplest way of working out the dynamics of test particles is to use the invariants which can be derived by applying the Euler–Lagrange equations to the metric (19.22). It is straightforward to show that the equivalent of the law of conservation of angular momentum is the expression

$$r^2 \frac{d\theta}{d\tau} = h = \text{specific angular momentum}, \quad (19.25)$$

and the expression for the radial motion of the test particle is

$$\frac{d^2u}{d\theta^2} + u = \frac{GM}{h^2} + \frac{3GM}{c^2}u^2, \quad (19.26)$$

where $u = 1/r$.² In these expressions, θ is the angle measured in the plane of the accretion disc, r the angular diameter distance from the centre and $d\tau$ is the increment of proper time $ds = d\tau$. For circular motion about the nucleus $d^2u/d\theta^2 = 0$ and so

$$u = \frac{GM}{h^2} + \frac{3GM}{c^2}u^2. \quad (19.27)$$

From this expression we can find the value of h for the motion of a particle in a circular orbit at any radius $r \geq 3r_g$. Inserting the value $r = 3r_g$, we find $h = \sqrt{12}GM/c$.

Inspection of the $d\theta$ term of the metric (19.22) shows that the component of the four-velocity in the circumferential direction is

$$\gamma v = r \frac{d\theta}{d\tau}, \quad (19.28)$$

and so, using the invariant (19.25),

$$\gamma v = \frac{h}{r} = \frac{c}{\sqrt{3}} \quad \text{for } r = 3r_g. \quad (19.29)$$

Therefore, $v = c/2$, that is, test particles on the last stable orbit about a Schwarzschild black hole orbit at half the speed of light. In the case of a face-on accretion disc, the

² I have given simple derivations of these expressions in Chap. 17 of *Theoretical Concepts in Physics* (Longair, 2003).

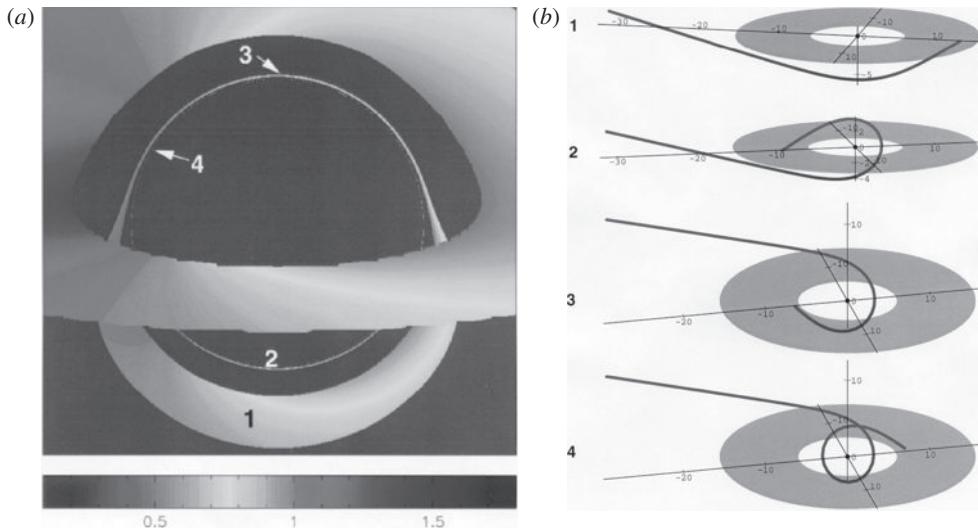


Fig. 19.10 The expected red- and blueshift distribution of radiation from the innermost regions of a thin accretion disc about a Schwarzschild black hole which extends into the last stable orbit at $r = 3r_g$. The grey-scale shading shows the energy shift of the spectral line relative to its rest wavelength. The sketches in the right-hand column show the trajectories of the photons from different parts of the disc to the distant observer. The plane of the disc is inclined at an angle $i = 75^\circ$ to the plane of the sky. (Courtesy of Youri Dabrowski.)

transverse Doppler shift from $r = 3r_g$ is $v_{\text{obs}}/v_0 = 1/\gamma = \sqrt{3}/2 = 0.866$. Combining the gravitational and transverse Doppler shifts, the maximum redshift of a face-on accretion disc about a Schwarzschild black hole is $v_\infty = \sqrt{2/3} \times \sqrt{3}/2 v_0 = v_0/\sqrt{2}$. In the case of the fluorescent line of iron at 6.4 keV, the line could be redshifted to 4.5 keV and there would be no blue shifted component of the line.

To work out the shape of the spectral line at an arbitrary angle to the line of sight is a much more complex calculation because of the strong curvature of space-time close to the black hole. The trajectories of light rays which can reach the observer can be worked out by ray tracing in the strongly curved space-time about the black hole. An example of the distribution of red- and blueshifts from a thin accretion disc inclined at an angle of 75° to the plane of the sky, as seen by a distant observer, is shown in Fig. 19.10a. In Fig. 19.10b the light paths from different parts of the disc to the observer are shown. Some of the trajectories are quite complex, in the fourth example, the light path making a complete circuit about the black hole before setting off towards the distant observer.

Examples of the predicted shapes of the 6.4 keV line are shown in Fig. 19.11a and b. In the case of a Schwarzschild black hole, we can understand the profile of the face-on disc C from the above considerations. In these computations, the discs extended from the last stable orbit at $r = 3r_g$ to $r = 7.5r_g$. The maximum redshift, corresponding to emission from the last stable orbit, is $v_\infty = v_0/\sqrt{2}$. Exactly the same calculation can be performed for the outer radius of the disc and then the gravitational redshift is $v_\infty = v_0\sqrt{13/15}$ and the transverse Doppler shift $v_\infty = v_0\sqrt{12/13}$. Therefore, the combined redshift from the outer

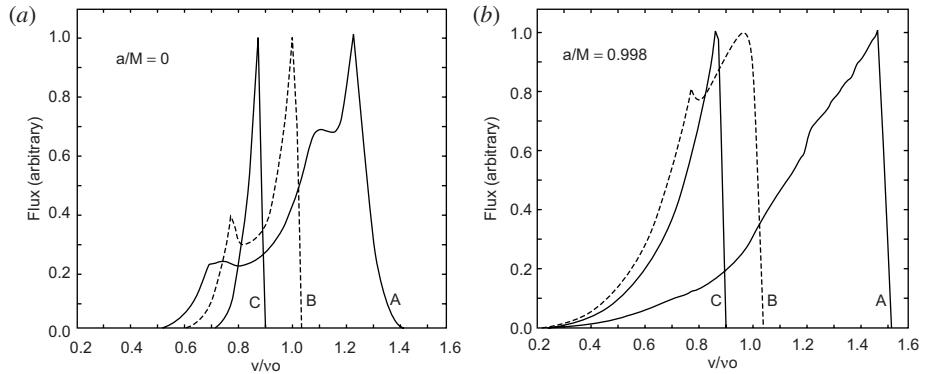


Fig. 19.11

The predicted line shapes of the fluorescent 6.4 keV line for (a) Schwarzschild and (b) Kerr black holes (Dabrowski *et al.*, 1997). In both cases, the inclinations of the plane of the accretion disc to the plane of the sky are: A $i = 85^\circ$; B $i = 30^\circ$; C $i = 0^\circ$. The extreme Kerr black hole shown in the right-hand panel has $a/M = 0.998$.

radius of the disc is $v_\infty = v_0/\sqrt{12/13} \times \sqrt{13/15} = v_0/\sqrt{4/5} = 0.89v_0$, in agreement with the sharp cut-off observed in the line profile C in Fig. 19.11a.

As the inclination of the disc to the plane of the sky increases, more extensive blue and red wings are observed, as expected from (19.21). The profiles have a characteristic ‘double-horned’ appearance and this is a characteristic feature of the velocity profiles of rotating discs observed at an angle to the line of sight. It is a simple calculation to show that, in the non-relativistic limit, an isotropically emitting annulus rotating at speed v_0 , inclined at an inclination i to the plane of the sky, has an observed velocity profile

$$I(v_{\parallel}) dv_{\parallel} \propto \frac{dv_{\parallel}}{\left[1 - \frac{v_{\parallel}^2}{(v_0 \sin i)^2} \right]^{1/2}} \quad \text{for } -v_0 \sin i \leq v_{\parallel} \leq v_0 \sin i . \quad (19.30)$$

The intensity is greatest at the points at which the disc is observed edge-on. The same behaviour is observed in the models A and B shown in Fig. 19.11a, the greater the angle of inclination, the broader the wings. However, in the case of accretion discs about black holes, relativistic aberration effects and the gravitational redshift distort the symmetric profile. The gravitational redshift shifts the whole spectrum to lower energies, while the blueshifted ‘horn’ has greater intensity than the redshifted ‘horn’. The complexity of the paths of the light rays shown in Fig. 19.10 means that the predicted line profiles can only be derived by ray tracing. One of these profiles is fitted to the observed profile of the line in Fig. 19.9, the angle of inclination being $i = 30^\circ$ and the inner radius $r = 2GM/c^2$. The relativistic ‘double-horned’ signature of a relativistic disc has been somewhat smoothed out but the model still provides a convincing explanation for the shape of the line profile. As expressed by Fabian (2009), the high energy cut-off is primarily determined by the inclination of the disc and the extent of the line to the low energy side of the line by the magnitude of the gravitational redshift.

We can estimate the maximum redshift of a Schwarzschild black hole for the case in which the angle of inclination is $i = \pi/2$. The gravitational redshift remains the same as

before, $v_\infty/v_0 = \sqrt{2/3}$, but, from (19.21), the Doppler shift is $v_\infty/v_0 = [\gamma(1 + v/c)]^{-1}$, where $v/c = 0.5$. Hence, the maximum combined redshift is $v_\infty/v_0 = (1/(\sqrt{3} \times \sqrt{2/3})) = \sqrt{2}/3 = 0.471$. Thus, if the line profile extends to energies less than about $0.5v_0$, a Schwarzschild black hole does not provide a large enough combined redshift.

Larger redshifts are found in the case of Kerr black holes. As discussed in Sect. 13.11.2, for a maximally corotating black hole, the last stable orbit coincides with the surface of infinite redshift. It turns out that the speed of a particle on the last stable orbit is again $v/c = 0.5$, but the radiation suffers a very much larger gravitational redshift, in fact, it would be infinite at $r = GM/c^2 = r_g/2$ for a maximally rotating black hole. The computations of Dabrowski and his colleagues (1997) illustrate the expected line profiles for accretion discs which extend through the static radius to the last stable orbit about the rotating black hole (Fig. 19.10b) – these examples are for an extreme Kerr black hole with $a/M = 0.998$. Many of the same considerations apply as in the case of the Schwarzschild black hole, but the effects of the greater gravitational redshifts mean that the redshifted tail of the line profile can extend to very much smaller values of v_∞/v_0 than in the Schwarzschild case. In Fig. 19.8, a fit of the observed profile to that expected from an accretion disc about a Kerr black hole with $i = 30^\circ$ and an inner radius $r = 2GM/c^2$ is shown. The disc emissivity is described by a broken power-law in radius, where the break radius occurs at $r = 6GM/c^2 = 3r_g$. Beyond this radius, the disc emissivity profile is $q \propto r^{-2.5}$ resulting in an equivalent width of 200 eV, corresponding to the core of the line seen in Fig. 19.8. Within $3r_g$ the emissivity steepens, resulting in the strong low energy tail to the line.

Fabian and his colleagues (2009) describe observations of Seyfert 1 galaxies made with X-ray satellites such as XMM-Newton, Suzaku and RXTE and these show that broad iron lines are commonly found among Seyfert 1 galaxies, the widths generally being so great that the black holes must be spinning rapidly. Similar broad lines are observed in the X-ray spectra of the Galactic black holes discussed in Sect. 14.7.5 and illustrated in Fig. 14.28. The physics of these spectra is exactly the same as discussed above.

19.7 The growth of black holes in the nuclei of galaxies

Is it plausible that supermassive black holes grow to masses $M \sim 10^9 M_\odot$ over cosmological time-scales? Let us derive the formula for the rate of growth of black hole masses first described by Salpeter.

19.7.1 The Salpeter time-scale

Salpeter (1964) addressed the question of the maximum rate at which the mass of a black hole can grow by accretion. The maximum luminosity of a black hole powered by accretion is

$$L = \xi \dot{m} c^2 \left(\frac{r_g}{r} \right) , \quad (19.31)$$

where \dot{m} is the rate at which mass is accreted by the black hole and ξ is the efficiency of the accretion process. At first sight, it might seem that the black hole mass could grow arbitrarily rapidly, but this is not the case because the accretion rate is limited by the radiation pressure associated with the luminosity L . As shown in Sect. 14.2, the maximum luminosity of a compact object of mass M is given by the Eddington luminosity,

$$L_E = 1.3 \times 10^{31} \left(\frac{M}{M_\odot} \right) W = A \left(\frac{M}{M_\odot} \right). \quad (19.32)$$

Thus, for an object of mass M , there is an upper limit to the mass accretion rate \dot{m} . If the luminosity of the system is less than the Eddington luminosity, this can be attributed to a lower mass accretion rate, which is therefore

$$\dot{m} = \frac{L}{\xi c^2} \left(\frac{r}{r_g} \right) = \frac{L}{L_E} \frac{L_E}{\xi c^2} \left(\frac{r}{r_g} \right) = \frac{L}{L_E} \left(\frac{r}{r_g} \right) \frac{AM}{\xi M_\odot c^2}, \quad (19.33)$$

corresponding to exponential growth of the mass of the black hole. The characteristic time-scale for the exponential growth is

$$\tau = \frac{M}{\dot{m}} = \frac{\xi M_\odot c^2}{A} \left(\frac{r_g}{r} \right) \frac{L_E}{L}. \quad (19.34)$$

Setting $r = r_g$, this time-scale is

$$\tau = \frac{M}{\dot{m}} = 4.5 \times 10^7 \left(\frac{\xi}{0.1} \right) \left(\frac{L_E}{L} \right) \text{ years}. \quad (19.35)$$

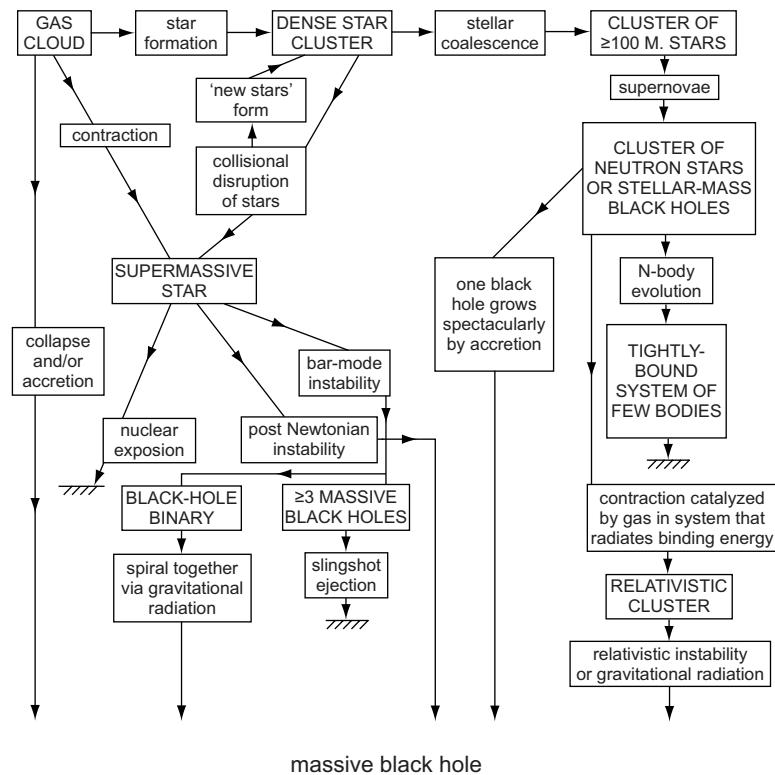
If a black hole accretes mass at this limiting rate, the time-scale for the exponential growth of its mass is only about 10^8 years and so there is no problem in principle in accounting for the existence of supermassive black holes by growth from stellar-mass black holes. It does, however, require the accretion process to take place throughout a substantial fraction of the lifetime of the galaxy. Note also that the time-scale is shorter if the value of ξ is small – if the dissipation process is inefficient, more mass is deposited into the black hole for a given luminosity. Low efficiency of the conversion of accreted mass into radiation clearly helps the growth of the black hole and this occurs in the advective mode of mass transfer into the black hole discussed in Sect. 20.2.

19.7.2 The Rees diagram

Many possible routes for the formation and growth of black holes in active galactic nuclei are pleasantly summarised in Fig. 19.12, which is often referred to as the *Rees diagram* (Rees, 1984) and which shows various scenarios, or pathways, by which supermassive black holes might be formed. It is an interesting challenge to work out the physical processes which play a role along each segment of the diagram.

There are some common features to the various pathways through the diagram.

1. Many of the routes begin with all the matter in gaseous form on a galactic scale and effective dissipation processes are needed to enable the matter to get rid of the gravitational potential energy which it necessarily possesses as it makes its way into

**Fig. 19.12**

The Rees diagram showing the many pathways by which a supermassive black hole might form in the nucleus of a galaxy. Those pathways which end without the formation of a supermassive black hole are shown with hatched terminators. (From Robson, 1999, after Rees, 1984.)

closer and closer orbits about the nucleus. Closely related to this problem is the necessity of getting rid of the angular momentum which the material acquires as it moves into more tightly bound orbits about the centre. One of the challenges is to ensure that a significant fraction of the energy is dissipated in the non-thermal forms which are the signature of the emission mechanisms from active galactic nuclei.

2. Each of the pathways involves star formation and stellar evolution. Some parts of the story are compelling, for example, the fact that stellar-mass black holes are observed to be present in binary systems shows that this is a route that certainly occurs in nature. What is less clear is the role of feedback mechanisms between the processes of star formation and evolution and the evolution of physical conditions in the nuclear regions of the galaxy. Do stellar winds and supernovae blow away the mass needed to fuel the black holes?
3. Some routes involve pieces of physics which may be crucial in the initial formation process of the black hole. For example, the conventional upper limit to the masses of main sequence stars is about $100 M_{\odot}$. There is, however, the possibility that supermassive stars might form, as proposed by Fowler and Hoyle (1963), which would be highly

unstable, and so be the progenitors of black holes with masses significantly greater than $10 M_{\odot}$.

4. Many of the pathways involve the evolution of compact star clusters and give rise to the issue of whether or not they can evolve rapidly enough to form black holes over cosmological time-scales. The considerations of Sect. 19.3.2 on the evolution of clusters of compact stars and neutron stars in galactic nuclei are relevant to this issue. Krolik (1999) gives a good discussion of the issues involved in interpreting this aspect of the Rees diagram.
5. Many important clues are provided by observation. For example, in a number of the most luminous ultraluminous infrared galaxies (ULIRGs), the far-infrared emission originates from their nuclear regions. As discussed in Sect. 18.7, the source of energy may involve star-formation activity associated with a nuclear starburst or with dust clouds heated by an active galactic nucleus. Many of the ultra-luminous far-infrared sources are associated with interacting galaxies. An appealing interpretation of these observations is that, in the collision of the interstellar media of the galaxies, some of the interstellar gas goes directly into the formation of stars throughout the interaction region, as in the case of the Antennae (Fig. 3.3a), but some of the gas loses its angular momentum and drifts into the nuclear regions of the galaxies either to form a nuclear star cluster or provide fuel for a pre-existing black hole. The necessary mass accretion rate onto a $10^9 M_{\odot}$ black hole radiating with 10% efficiency at the Eddington luminosity amounts to only $100 M_{\odot} \text{ year}^{-1}$, which is not excessive. The importance of these studies for the formation of black holes in the nuclei of galaxies is that they demonstrate that compact nuclear star clusters can be formed rather naturally in the course of galactic evolution.

It seems entirely plausible that many of the pathways shown in Fig. 19.12 result in the formation of black holes in the nuclei of galaxies and these could then grow in mass by the processes of accretion and/or coalescence.

20.1 The prime ingredients of active galactic nuclei

It is convenient to divide the necessary ingredients of active galactic nuclei into two types – the *primary ingredients*, which originate close to the black hole and its associated accretion disc, and *secondary phenomena*, which result from the interaction of the primary ingredients with the environment of the black hole. Figure 20.1 is a schematic diagram showing some of the components of typical models. The primary ingredients are *intense non-thermal continuum radiation* and *fluxes of relativistic material in the form of highly collimated jets*. The secondary phenomena result from the interaction of the primary components with the surrounding medium, in particular, gas clouds in the vicinity of the nucleus and the ambient interstellar and intergalactic gas. The former gives rise to the strong emission line spectrum observed at optical, ultraviolet and infrared wavelengths whilst the interactions of the relativistic jets with the interstellar and intergalactic gas give rise to the structures observed in extragalactic radio sources and in intense γ -ray emission. We study the physics of high energy particles in extragalactic radio sources and galactic nuclei and the role of relativistic beaming in the following chapters.

20.2 The continuum spectrum

As discussed in Chap. 18, active galactic nuclei contain intense continuum emission with non-thermal spectra. The examples illustrated in that chapter include typical spectra of Types 1 and 2 Seyfert galaxies (Fig. 18.5a and b), a composite quasar spectrum (Fig. 18.1) and a multi-waveband spectrum of the BL-Lac object OJ287 (Fig. 18.7). In these cases, most of the luminosity is in the continuum spectrum rather than in the lines; furthermore, the variability of the source is almost entirely associated with the continuum spectrum. The continuum spectrum is unlike any stellar or galaxy spectrum, the latter being the integrated light of many different types of star, displaying characteristic features such as the atomic and ionic absorption lines, the Balmer decrement and narrow absorption lines associated with regions of star formation. Very often the continuum spectrum can be represented by a power law in the optical and infrared wavebands, most objects having spectra which can be described by $I_\nu \propto \nu^{-\alpha}$ with the average value of the spectral index α close to unity but with considerable dispersion about this value. For a number of bright objects, a more

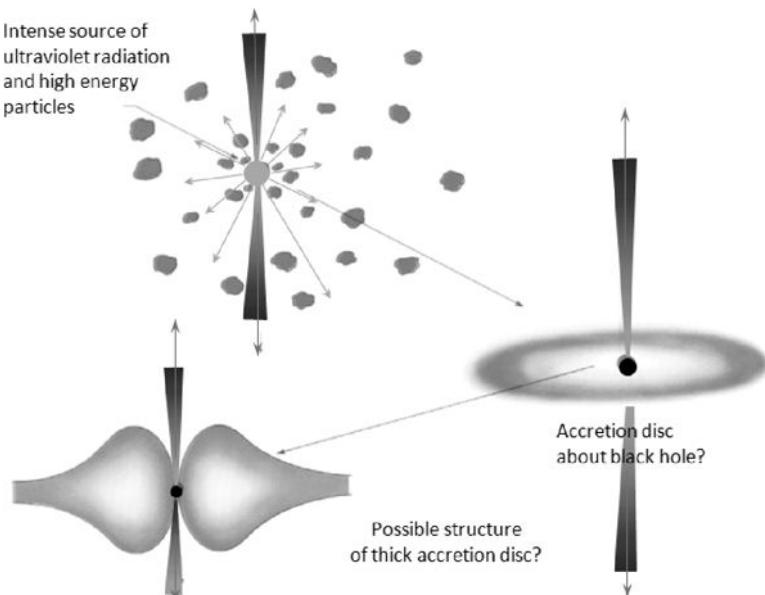


Fig. 20.1 A schematic diagram showing the necessary ingredients of a model of an active galactic nucleus. There must be a compact source of high energy particles and intense ultraviolet ionising radiation in the very centre. The nucleus is surrounded by gas clouds, which are heated and excited by the ionising radiation from the nucleus. The clouds closest to the centre have high particle densities, values of about 10^{16} m^{-3} being necessary to de-excite the forbidden line radiation from these regions. The broad-line emission observed in quasars, Type 1 Seyferts and broad-line radio galaxies originate in these regions. Further out, the clouds are less dense with densities of the order of $10^{10}-10^{12} \text{ m}^{-3}$, and these are the source of the narrow-line regions observed in Type 2 Seyfert galaxies and some of the narrow-line radio galaxies. The inserts show an accretion disc which thickens close to the black hole. These may be responsible for the collimation of the relativistic jets.

detailed decomposition of the continuum spectrum in the optical and ultraviolet wavebands can be carried out and these show evidence for excess emission at ultraviolet wavelengths as compared with the extrapolation of a power-law spectrum. The best example is the ‘blue bump’ observed in the ultraviolet spectrum of 3C 273 (Fig. 14.5). In addition, observations of active galaxies in the soft X-ray waveband $\varepsilon \lesssim 1 \text{ keV}$ often provide evidence for a steep spectrum component which may represent the high energy tail of the blue bump. A number of cases are known in which extremely steep spectra are found in the infrared waveband, optical to infrared spectral indices as steep as $\alpha \simeq 6$ being found. These are sometimes referred to as ‘infrared quasars’ since essentially all their energy is emitted in that waveband.

It is important to know the continuum spectra of active nuclei in the ultraviolet waveband and beyond because this emission is responsible for the photoexcitation and photoionisation of diffuse gas clouds in the vicinity of the nucleus, as is convincingly demonstrated from the considerations of Sects 20.3 and 20.4. Whilst excellent observations in the wavelength range 120–300 nm have been made with the IUE, few observations of active galaxies or

quasars are available from about 120 nm to the X-ray waveband. As discussed in Sect. 18.8, quasars and active nuclei dominate the extragalactic X-ray sky and their X-ray spectra form a natural extension of the optical spectrum into the X-ray waveband (Fig. 18.11). The optical-to-X-ray spectra of most quasars can be described by a power law with spectral index $\alpha \simeq 1$, this form of spectrum indicating that roughly as much energy is radiated per decade of frequency in the optical as in the X-ray waveband. Information about the shape of the continuum spectrum in the Lyman continuum and far-ultraviolet waveband can be obtained from the presence of high excitation lines in the emission line spectra. As discussed below, one of the attractions of the photoexcitation picture is that a power-law continuum spectrum, containing photons with a very wide range of energies, can account for the observation of a very wide range of different ionisation states of elements in a single spectrum.

The continuum radiation is also known to be polarised. The strongest polarisation is found in the cases of the BL-Lac objects in which up to 30–40% linear polarisation has been observed. In a few important cases, polarisation changes have been followed through strong radio and infrared outbursts, and these are important in understanding the magnetic field geometry of the source regions. It appears that the radiation associated with the ‘blue bump’ is not polarised since the total percentage polarisation decreases when it contributes significantly to the total continuum intensity. This is consistent with the blue bump being mostly thermal radiation.

The nature of the continuum spectra of active galactic nuclei is likely to be associated with a number of the processes already considered in some detail.

- A natural explanation of the polarised power-law energy spectrum is that it is the *synchrotron radiation* of ultra-relativistic electrons accelerated close to the active nucleus. As discussed in Chap. 17, all that is required is the presence of strong shocks in the nucleus and acceleration of charged particles is inevitable.
- An ultraviolet thermal spectrum can be naturally explained as the *thermal emission* of an optically thick thin accretion disc, following the considerations of Sect. 14.3. There, it was shown that the simplest prediction of the model is that the temperature of the thermal emission changes as $M^{-1/4}$. Thus, scaling from the case of the thermal emission in the X-ray waveband for accretion discs about solar mass black holes, the emission from the supermassive black holes in the nuclei of galaxies would be expected to lie in the ultraviolet waveband. The case of supermassive black holes is, however, highly non-trivial since the optically thick assumption and the approximation that the disc radiates like a black-body are evidently crude assumptions.
- Comptonisation is undoubtedly a key process in the formation of the hard X-ray and γ -ray spectra of active galactic nuclei (Sect. 9.4). As demonstrated in that section, the process has the feature of creating power-law spectra by purely thermal means.
- The evidence of the time-lag between variations in the X-ray emission of Seyfert 1 galaxies and the optical continuum (Sect. 20.5.2) suggests that reprocessing of the X-ray emission plays an important role in the formation of the optical continuum emission of these active galactic nuclei.

The reader is referred to the text by Krolik (1999) for more detailed consideration of continuum emission processes.

20.3 The emission line regions – the overall picture

The emission line spectra of active galaxies provide evidence for a wide range of physical conditions in gas clouds at different distances from the nucleus, the most important parameters being the intensity and spectrum of the ionising radiation and the particle densities within the gas clouds. The full apparatus of atomic and ionic spectroscopic analysis has been employed to determine these parameters, the diagnostic tools being described in detail in the excellent text *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* by Osterbrock and Ferland (2005). From these results, one may then proceed to develop models for the distribution of the gas clouds about the black hole which may be taken to be a point source of ultraviolet ionising radiation.

The general picture which emerges is illustrated in Fig. 20.1. The emission line regions which originate closest to the black hole are responsible for the broad-line emission observed in the most active nuclei, the Seyfert 1 galaxies and the quasars. The most direct evidence for this is the variability of the emission line profiles on the time-scale of months and less, as illustrated by the case of the spectrum of NGC 4151 (Fig. 18.6), and so they must originate very close to the nuclei. They must also be rather dense regions because the strong broad emission lines are not accompanied by the corresponding forbidden lines. This observation is attributed to the fact that forbidden lines, such as those of doubly ionised oxygen [O III], are suppressed at electron densities at which collisional de-excitation depopulates the upper levels of the transitions more rapidly than radiative transitions as discussed in Sect. 12.3.2; a list of de-excitation densities is given in Table 12.2 for a number of common ionic species observed in the spectra of active galactic nuclei. For the broad-line regions, the electron densities must be greater than about 10^{14} m^{-3} . Thus, in the nuclear regions within about 0.1 pc or less of the black hole, there are dense clouds with a large velocity dispersion. The broadening of the lines might be the result of turbulent, rotational, expansion or contraction velocities and careful study of the time evolution of the line profiles during an outburst from the nucleus can be used to distinguish between these possibilities.

One important feature of these analyses is that the gas must be confined to clouds which have a very small *filling factor*, meaning the fraction of the volume about the nucleus which the clouds occupy. The reason for this is that, if the gas were uniformly distributed with the density of the broad-line regions, the optical depth for Thomson scattering would be very much greater than 1 and so the ultraviolet and optical emission from the nucleus would not be observed. Taking as an example the parameters of the broad-line regions discussed in the last paragraph, the optical depth to Thomson scattering of a region of density 10^{14} m^{-3} and scale 0.1 pc would be $\tau = \sigma_T N_e l \sim 20$. This is a conservative lower limit to the opacity. Consequently, the emission line regions must be confined to small clouds which occupy only a small fraction of the volume and allow the nuclear regions to be observed directly. Typical filling factors for models of active galactic nuclei are

$f \sim 10^{-4}\text{--}10^{-6}$. Because the optical depths in the individual clouds are large, only the regions of the clouds facing towards the source of the ultraviolet radiation are illuminated. This explains why clouds, rather than a diffuse region surrounding the nucleus, are shown in Fig. 20.1.

Further out from the nucleus are the clouds responsible for the narrow emission line spectrum. These emit strong permitted and forbidden lines with similar linewidths and also exhibit a wide range of ionisation states. This is consistent with a picture in which the regions have particle densities $\sim 10^9\text{--}10^{11} \text{ m}^{-3}$ and are photoionised by the ultraviolet radiation from the nucleus. The dimensions of the regions can be roughly estimated from the strengths of the lines, the inferred particle densities and the fraction of the total volume occupied by clouds. In many Seyfert nuclei, there is often evidence for both types of gas cloud being present, the relative strengths of the broad and narrow line components varying from Type 1 Seyfert galaxies, which are pure broad-line systems, to pure Type 2 Seyferts in which the narrow lines are dominant. The radio galaxies which are powerful radio sources have similar Type 1 and Type 2 spectra at low redshifts but there are important differences at redshifts $z \sim 1$. The luminous 3CR radio galaxies at $z \approx 1$ possess very strong, narrow emission lines which can extend to 50–100 kpc, greater than the dimensions of the galaxy itself. The nature of the line emission from these galaxies is discussed in Sect. 20.5. Finally, there are even weaker nuclei in which only a low ionisation spectrum is observed. In these ‘low ionisation, narrow emission line region’ galaxies, the excitation mechanism for the clouds could be either collisions or photoexcitation by ultraviolet radiation.

The densities and temperatures which can be studied are largely determined by the lines available for study in the optical, ultraviolet and X-ray spectral regions. There may well be a wide range of densities and temperatures present in the gas clouds surrounding active nuclei but the diagnostic tools are only sensitive within rather specific ranges of density and temperature.

20.4 The narrow-line regions – the example of Cygnus A

An excellent example of the types of information which can be gleaned from optical spectrophotometric observations of the emission line regions of active galactic nuclei is provided by the radio galaxy Cygnus A. The radio source is the brightest extragalactic radio source in the northern sky and it has a quite anomalously large flux density for its luminosity. As pointed out by Stockton and Ridgway (1996), at a redshift $z = 0.0561$, the radio source is about 10 times closer than typical radio sources of the same intrinsic radio luminosity, the 3CR radio sources of the same luminosity lying at redshifts $z \sim 1$.

The rich optical spectrum of Cygnus A is shown in Fig. 20.2, displaying all the characteristic features of a narrow-line radio galaxy or Seyfert 2 galaxy. This is a particularly useful example because it has been the subject of detailed study and provides an excellent example of what can be learned from these studies. In Table 20.1, the line intensities are listed, taken from Osterbrock and Ferland’s book which provides a clear discussion of how

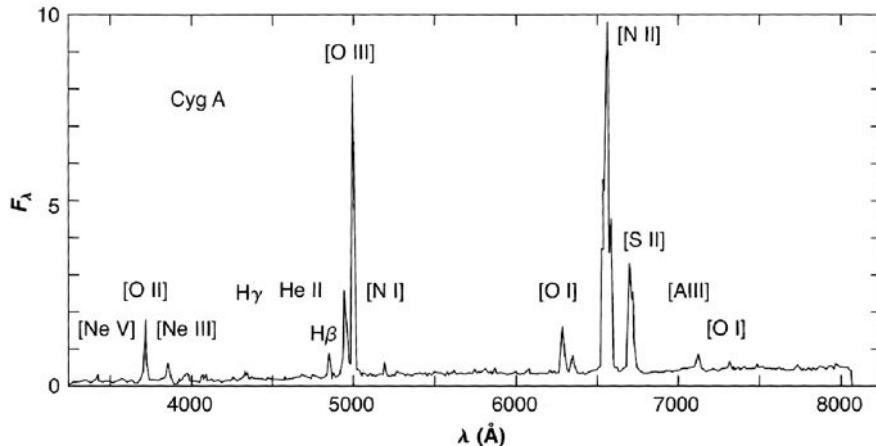


Fig. 20.2 The optical spectrum of the narrow-line radio galaxy Cygnus A showing the prominent narrow emission lines (Osterbrock and Ferland, 2005).

such a spectrum can be interpreted and the uncertainties involved. We follow their careful presentation in the following paragraphs.

- What is immediately striking from Fig. 20.2 and Table 20.1 is the wealth of emission lines, particularly of the forbidden lines of ions in a wide range of states of ionisation. Thus, oxygen, neon and iron are all present in a variety of different ionisation states.
- The Balmer hydrogen series from H α to H δ , known as the *Balmer decrement*, is present and these lines are of special importance since their relative intensities can be predicted from atomic physics. Following the detailed analysis of Osterbrock and Ferland, the predicted ratio of intensities are almost independent of temperature and density over a wide range of physical conditions and typically have ratios H α : H β : H γ : H δ \equiv 2.847 : 1 : 0.469 : 0.260. Inspection of Table 20.1 shows that the observed Balmer decrement is 6.61 : 1 : 0.32 : 0.17, much steeper than expected values and similar to the decrements observed in H II regions and planetary nebulae. The cause of this discrepancy is interstellar dust extinction which, as discussed in Sect. 12.4, is strongly wavelength dependent. Anticipating the discussion of subsequent paragraphs, if the regions are assumed to have temperature $T \approx 10^4$ K and electron number density $n_e \approx 10^{10}$ m $^{-3}$, the extinction amounts to $E(B - V) = 0.69 \pm 0.04$. Assuming the other lines in the narrow-line spectrum originate from the same regions, their intensities can be corrected for interstellar extinction, with the results shown in column 4 of Table 20.1. It can be seen that the corrected Balmer decrement is in good agreement with theoretical expectations.
- This is a substantial correction for dust extinction, but not all of it originates in the Cygnus A host galaxy. Cygnus A lies at a low galactic latitude and it is estimated that about half of the extinction arises in our own Galaxy. This correction not only affects the line ratios,

Table 20.1 Observed and calculated relative line fluxes of Cygnus A (Osterbrock and Ferland, 2005). The intensities have been normalised to that of H β .

Ion	Wavelength (nm)	Measured flux	Corrected flux	Crab Nebula	Photoionisation model
[Ne V]	334.6	0.14	0.38	–	0.12
[Ne V]	342.6	0.36	0.95	0.46	0.34
[O II]	372.7	2.44	5.00	10.3	0.24
[Ne III]	386.9	0.66	1.23	1.56	0.53
[Ne III]	396.7	0.22	0.40	0.47	0.16
[S II]	407.2	0.14	0.23	0.31	
H δ	410.1	0.17	0.28	0.31	0.26
H γ	434.0	0.32	0.46	0.61	0.47
[O III]	436.3	0.16	0.21	0.19	0.19
He I	447.1	≤ 0.07	≤ 0.09	0.28	0.02
He II	468.6	0.25	0.28	0.53	0.18
H β	486.5	1	1	1	1
[O III]	495.9	4.08	3.88	2.81	6.3
[O III]	500.7	13.11	12.30	8.43	18.1
[N I]	519.9	0.40	0.32	–	–
[Fe XIV]	530.3	≤ 0.10	≤ 0.08	–	0.01
[Fe VII]	572.1	≤ 0.10	≤ 0.06	–	0.03
[N II]	575.5	0.14	0.09	0.11	–
He I	587.6	0.13	0.08	0.79	0.06
[Fe VII]	608.7	≤ 0.07	≤ 0.04	–	0.04
[O I]	630.0	2.10	1.10	1.20	1.24
[O I]	636.4	0.69	0.35	0.33	0.41
[Fe X]	637.5	0.10	0.05	–	0.07
[N II]	654.8	3.94	1.90	1.56	0.29
H α	656.3	6.61	3.08	3.28	2.85
[N II]	658.3	13.07	6.15	4.69	0.86
[S II]	671.6	3.65	1.66	5.00	
[S II]	673.1	3.29	1.51	–	
[Ar III]	713.6	0.64	0.25	0.38	–
[O II]	732.5	0.35	0.13	0.50	–
[Ar III]	775.1	0.13	0.043	–	–

but also the ionising continuum radiation. The continuum radiation consists of both the stellar radiation of the giant elliptical galaxy and the ionising continuum. Once the former had been subtracted, the remaining featureless continuum has an intensity spectrum of the form $L(\nu) \propto \nu^{-\alpha}$, where $\alpha = 3.8$. Applying the correction for dust extinction, the value of α is reduced to 1.6. As discussed by Osterbrock and Ferland, such a continuum

spectrum extrapolated into the ultraviolet region of the spectrum can account for as diverse emission lines as [O I], [S II], [Ne V] and [Fe VII].

- The ratios of the strengths of different lines can be used to estimate the temperatures and densities of the emitting line regions. For example, the [O III] intensity ratio $(\lambda 372.9 + \lambda 500.7)/\lambda 436.3$ and the [N II] ratio $(\lambda 654.8 + \lambda 658.3)/\lambda 575.5$ are consistent with temperatures $T \approx 10^4$ K in the low density limit $n_e < 10^7 \text{ m}^{-3}$. While the [OIII] $\lambda 372.9/\lambda 372.6$ ratio is a good electron density diagnostic in regions of ionised hydrogen in the Galaxy, the linewidths of the lines in the Cygnus A spectrum are too broad to make this test useable. The [S II] $\lambda 671.6/\lambda 673.1$ ratio would be consistent with electron densities $(3-4) \times 10^6 \text{ m}^{-3}$.
- Once assumptions are made about the temperatures of the regions in which the different ionic species are found, estimates of the element abundances in these regions can be made and it is found that they are not so different from the Galactic abundances of the elements listed in Table 15.1. A significant difference is the lower abundance of iron, Fe, which is probably because of depletion onto interstellar grains, as is found for the heavier elements in the interstellar gas.
- A compelling case can be made that photoionisation is the process responsible for the excitation of the emission lines. The temperature-sensitive forbidden lines all result in estimates of the electron temperature of 10^4 K, consistent with the operation of the ‘thermostat’ mechanism which results from the rapid increase in cooling by collisionally excited line radiation with increasing temperature. Another check on the photoionisation model is whether or not there are sufficient ionising photons from the active nucleus to balance the recombination rate in the ionised gas. The intensity of the H β line can be used to estimate the number of recombinations using the procedures developed in the 1930s by Zanstra. Osterbrock and Ferland carry out this calculation and have shown that there are indeed sufficient ultraviolet photons in the extrapolated continuum spectrum of Cygnus A to account for the recombination spectrum.
- Another approach is to compare the emission line spectrum of Cygnus A with that of the Crab Nebula which is known to be photoionised by the optical and ultraviolet synchrotron radiation which has a spectrum similar to that inferred to be present in Cygnus A. The intensities of the emission lines, corrected for the effects of dust extinction, are listed in column 5 of Table 20.1. This approach automatically takes account of such complicating factors as inhomogeneities in the emitting clouds, stratification, extinction, and so on. Comparing columns 4 and 5 of Table 20.1 shows that there is remarkably good general agreement between the intensities of the emission lines.
- Having concluded that the emission line regions are excited by photoionisation, the physical properties of the clouds can be deduced from the luminosity of the H β line using (12.11). Assuming $T = 10^4$ K and $n_e = 10^{10} \text{ m}^{-3}$, typical values for the clouds are $r \sim 100 \text{ pc}$, the filling factor $\varepsilon \approx 10^{-2}$ and mass $M \approx 10^6 M_\odot$. These figure are consistent with what is observed in nearby Seyfert 2 galaxies, such as NGC 1068, which have been observed by narrow-band imaging.
- Next, the observations can be compared with the results of modelling of the predicted emission-line spectrum of clouds excited by the observed ultraviolet continuum radiation extrapolated into the ionising ultraviolet continuum. In the simplest models, the region

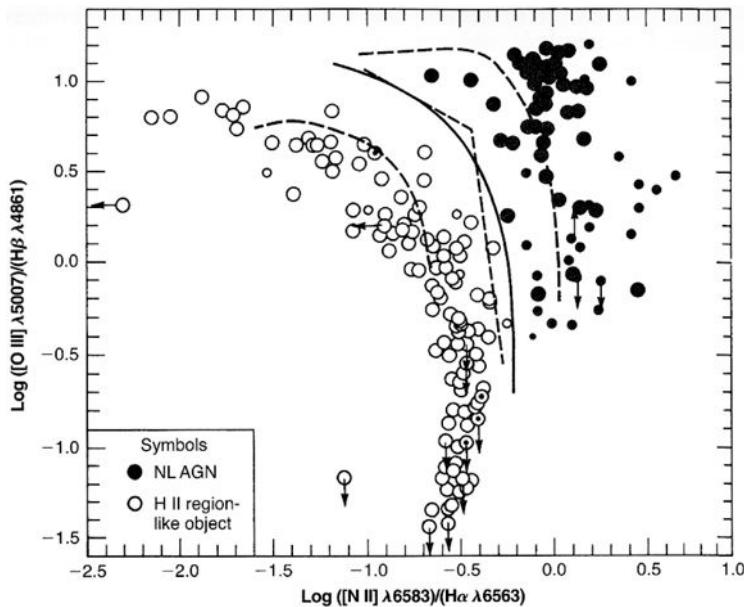


Fig. 20.3 The O III/H β versus [N II]/H α diagnostic diagram for emission line galaxies (Osterbrock and Ferland, 2005). The data have been corrected for the effects of reddening. The active galaxies in the upper right of the diagram form a quite distinct grouping from those galaxies whose emission lines originate in regions of ionised hydrogen. The solid line shows the dividing line between the two classes of galaxies and the dashed lines the results of various model calculations for photoionisation by hot stars (the H II galaxies) and by a non-thermal continuum from the active galactic nucleus (Osterbrock and Ferland, 2005).

can be taken to be spherical or a plane-parallel slab. The modelling proceeds by exactly the same route as that of H II regions and planetary nebulae, but with a non-thermal source of excitation. The emission spectra depend primarily upon the *ionisation parameter* U , which is defined by the relation,

$$U = \frac{1}{4\pi r^2 c n_{\text{H}}} \int_0^\infty \frac{L_\nu}{h\nu} d\nu , \quad (20.1)$$

where L_ν is the luminosity of the source per unit frequency interval and r is the distance from the source. Inspection of (20.1) shows that U is the dimensionless ratio of the number density of ionising photons to the electron density in the emission line region. For an electron density $n_e = 10^{10} \text{ m}^{-3}$ and a filling factor $\varepsilon = 10^{-2}$, this simple model results in the predicted line intensities shown in Table 20.1, column 6. Osterbrock and Ferland remark that ‘the model gives a good representation of the observed spectrum of Cygnus A’. The agreement with observation can be improved by adjusting the assumed values of the abundances of the elements on plausible physical grounds.

- Various line ratios can be used to demonstrate the difference between excitation by hot stars and by a non-thermal ultraviolet continuum. In the example shown in Fig. 20.3, the O III/H β ratio is plotted against the [N II]/H α ratio for a variety of emission line galaxies, including both active galaxies and those with large numbers of H II regions

(Osterbrock and Ferland, 2005). The coding shows that the active galaxies in the upper right of the diagram form a quite distinct grouping from those galaxies whose emission lines originate in regions of ionised hydrogen. The solid line shows an empirical dividing line between the two classes of galaxies. The dashed lines show the results of model calculations for photoionisation by hot stars (the H II galaxies) and by a non-thermal continuum from the active galactic nucleus. In the latter case, the spectral index is taken to be $\alpha = 1.5$ and the electron density $n_e = 10^3 \text{ cm}^{-3}$; the upper dashed line assumes the standard abundance of the elements while that to the left assumes the metal abundance is only 10% of the cosmic abundance. The significance of the ionisation parameter can be appreciated from the fact that at the top end of each dashed locus, the value is $U = 10^{-2}$ while at the bottom end it is 10^{-4} .

The above somewhat lengthy description of the analysis of spectrophotometric observations of Cygnus A and related objects indicates the steps necessary to interpret the observed narrow emission line spectrum of an active galactic nucleus. There are complexities and uncertainties associated with each step in the analysis. For example, no account is taken of the detailed shapes of the emission line clouds which are not known and a number of approximations and simplifications were needed to make progress. Nonetheless, the encouraging aspect of the analysis is that plausible parameters have been derived for the structures and properties of the clouds. The properties of the clouds are similar to those of other Type 2 systems which have been observed spectrophotometrically.

20.5 The broad-line regions and reverberation mapping

20.5.1 The physical properties of the broad-line regions

The broad-line regions are interpreted as originating in very much denser gas clouds than in the narrow-line regions. The evidence for this is the absence of forbidden lines such as [O III] which have intensities less than 1% of those observed in the narrow-line regions. On the other hand, the semi-forbidden line of C III] $\lambda 1909$ is observed with the same broad linewidth as the permitted lines, for which the critical density is $n_e \approx 10^{16} \text{ m}^{-3}$. Taken together, these constraints imply that the electron density in the broad-line regions is of the order of $10^{15}\text{--}10^{16} \text{ m}^{-3}$.

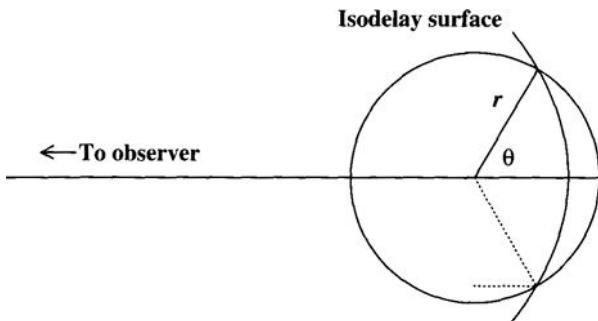
One striking feature of the broad-line regions is the presence of broad features associated with singly ionised iron, Fe II, which provide important diagnostic tools for their properties. The ionisation potential of Fe⁰ is 7.9 eV and that of Fe⁺, 16.2 eV. There is no direct temperature diagnostic for the broad-line regions because of the absence of the forbidden lines, but the temperature of the region must be less than 35 000 K or all the Fe⁺ would be converted into Fe²⁺. The great strengths of the iron lines can be attributed to the large number of permitted transition in the Fe II spectrum. As expressed by Osterbrock and Ferland, the optical Fe II lines are similar to the Balmer series of H I whilst the ultraviolet resonance lines are similar to the Lyman series.

Evidence that photoexcitation and ionisation are the dominant processes in the broad-line regions is provided by the strong correlation between the intensities of the hydrogen recombination lines and the featureless continuum. Furthermore, the same ions observed in the narrow-line spectra are observed with similar relative intensities, the implication being that the ionisation parameters for the broad and narrow lines are similar. Thus, if the electron density in the broad-line region is of the order of 10^{16} m^{-3} , the dimension of the regions about the active nucleus must be of order 0.1 pc, since $U \propto (n_e r^2)^{-1}$. Models of the structure of the broad-line clouds with these properties in active galactic nuclei suggest that their temperatures are of the order of 15 000 K, consistent with the presence of the intense Fe II lines. Specifically, Osterbrock and Ferland quote the following figures for Seyfert 1 and broad-line radio galaxies: if the H β luminosity is $10^9 L_\odot$, the mass of ionised gas is $M_{\text{ion}} = 36(10^{15}/n_e) M_\odot$ and the dimension of the region $R = 0.015\varepsilon^{-1/3}(10^{15}/n_e)^{2/3}$ pc. As a result, it is not surprising that the emission lines in the Seyfert 1 and broad-line radio galaxies are often observed to be variable. More detailed information about the structure of these regions can be found from the reverberation mapping technique discussed in Sect. 20.5.2. Performing the same calculation for the intense H β line observed in quasars with $L(\text{H}\beta) \approx 10^{10} L_\odot$, masses of the clouds are of order $2 \times 10^3 M_\odot$ and their dimensions $r \sim 0.25 \text{ pc} \approx 0.8 \text{ light years}$. Hence, as observed the quasar emission lines are expected to be less variable than those observed in the Seyfert 1 galaxies.

The intensities of the lines have to be corrected for the effects of interstellar dust extinction, but this is not as straightforward as in the case of the narrow-line regions because of the very large opacity in the Lyman- α line. For the parameters of the clouds in Seyfert 1 galaxies given in the last paragraph, the Lyman- α opacity is of the order of 10^6 and so the 2^2 P states are highly populated. Consequently, the line intensities of the Balmer series differ from the Balmer decrement used to estimate the extinction in the narrow-line regions – the corrected relative line strengths need to be used to estimate the interstellar extinction. It is unlikely, however, that there is dust within the broad-line regions themselves for two reasons. First, if the temperature of the gas in the broad-line regions is 15 000 K, dust grains would be evaporated in these high density environments. Second, strong permitted lines of Al, Ca, Si and Fe are observed in the spectra of the broad-line regions. These lines are normally strongly depleted if dust is present in these regions.

20.5.2 Reverberation mapping

One of the most powerful techniques for studying the physical properties of the emission line regions in the vicinity of active galactic nuclei is the technique of *reverberation mapping*. This approach involves the long-term monitoring of the optical-ultraviolet spectra and continuum intensities of active galactic nuclei and has been successfully applied to Seyfert 1 galaxies. In the late 1970s, a consortium of European astronomers used the IUE satellite to undertake systematic long-term observations of the bright Seyfert galaxy NGC 4151 and found that there are correlated variations between the strength of the C IV emission line and the intensity of the ultraviolet continuum (Fig. 18.6). The key discovery was a time delay between variations in the ionising ultraviolet continuum and the corresponding variations in the broad-line spectrum (Ulrich *et al.*, 1984a). This result demonstrated conclusively that

**Fig. 20.4**

Illustrating the parabolic isodensity surface expected from the illumination of a spherical shell about an active galactic nucleus by a pulse of ultraviolet radiation (Peterson, 1997).

the broad-line regions are excited and photoionised by the ultraviolet continuum radiation from the nucleus and also enabled a direct estimate of the distance of these clouds from the nucleus to be made, since the ultraviolet photons travel from the nucleus to the line-emitting regions at the speed of light. In the case of NGC 4151, the time delay amounted to about 13 days for the C IV line and so the velocity field about the nucleus could be probed on a scale of about 0.01 pc. A problem in interpreting these observations is that the kinematics of the line-emitting clouds in the nuclear regions is not known. The clouds could be in bound orbits about the nucleus, or they could be part of an inflow or outflow from it. In principle, these alternatives can be distinguished by studying the time evolution of the profiles of the emission lines. Ulrich and her colleagues argued that the variations of the breadths of the lines could be interpreted as orbital motions about the nucleus and then the combination of the distances of the clouds from the nucleus with their velocities enabled the mass of the central regions to be estimated – this turned out to be $M \sim (0.5 - 1) \times 10^9 M_\odot$ (Ulrich *et al.*, 1984b).

The natural extension of this technique is to study the time variations of a variety of different emission lines which originate at different distances from the source of ionising radiation. Excellent examples of the results of this type of study and the problems of interpretation are given by Peterson (1997). The technique depends upon the facts that, when the clouds are excited by the incident ultraviolet flux, they respond essentially instantaneously and also that the line-emitting clouds have a small filling factor, so that the ultraviolet ionising photons propagate unhindered from their source to the clouds. Consequently, the time delay between variations in the continuum intensity and those in the line-emitting clouds depends only upon the light travel time from the nucleus and the geometry of the clouds.

The time delay expected for the excitation of a spherical shell by a burst of ultraviolet emission in the nucleus can be worked out from Fig. 20.4. The locus of points on the shell with constant time delay τ from the nucleus to the observer via the dotted line shown on the diagram is a parabola, which is known as the *isodelay surface*. Geometrically, the time delay at angle θ from the nucleus to the line of sight, relative to the time the first signal from the near-side of the shell arrives at the observer, is $\tau = (1 + \cos \theta)(r/c)$, where r is the radius of the shell. As Peterson shows, for the case of a single shell, there is a simple

solution for the observed response to a short burst of ultraviolet radiation. The surface area of an annulus of angular thickness $d\theta$ is $2\pi r^2 \sin \theta d\theta$ and so, if a flux ξ per unit area is reradiated in the emission line for a given incident continuum intensity, the observed flux density is proportional to

$$\Psi(\theta) d\theta = 2\pi r^2 \xi \sin \theta d\theta . \quad (20.2)$$

But, $d\tau = -(r/c) \sin \theta d\theta$ and so

$$\Psi(\tau) d\tau = 2\pi r c \xi d\tau \quad \text{for } 0 \leq \tau \leq 2r/c . \quad (20.3)$$

Thus, to the distant observer, the observed response of the ring is that the luminosity of the line rises to the value given by (20.3) at $\tau = 0$ and remains at that constant intensity for a time $\tau = 2r/c$, after which it falls to zero. This is the simplest example of a *transfer function* which relates variations in the nuclear flux to time variations in the emission line as observed by a distant observer. Transfer functions for other geometries are described by Blandford and McKee (1982). In general, when integrated over a distribution of clouds at different radii, $\Psi(\tau)$ is a more complex function than that given by (20.3).

If the intensity of the central continuum source varies with time, the distant observer measures contributions to the line intensity at time t from a range of different times in the past, as described by the transfer function $\Psi_i(\tau)$, for the ionic species i . We can write the line intensity as

$$L_i(t) = \int_{-\infty}^{\infty} \Psi_i(\tau) C(t - \tau) d\tau , \quad (20.4)$$

where τ represents the time delay between variations in the continuum intensity and what the distant observer measures. $C(t)$ describes the time evolution of the burst of ionising radiation from the nucleus and $\Psi_i(\tau)$ the time response of the clouds about the nucleus. L_i is the observed time variation of the emission line. In the ideal case, precise observations would be made of $C(t)$ and $L_i(t)$ and the transfer function $\Psi_i(\tau)$ determined for each emission line and so information would be derived about the spatial distribution of different ions in the vicinity of the nucleus. To carry through this programme, spectroscopic observations of very high quality are needed and generally this is not possible. It is, however, feasible to estimate the time delays, or *lags*, between variations in the nucleus and the response of the line by cross-correlation techniques.

Following Peterson's analysis, a cross-correlation function is defined to be

$$F_{CCF}(\tau) = \int_{-\infty}^{\infty} L_i(t) C(t - \tau) dt , \quad (20.5)$$

while the autocorrelation function for the variation of the nuclear continuum intensity F_{ACF} is

$$F_{ACF}(\tau) = \int_{-\infty}^{\infty} C(t) C(t - \tau) dt . \quad (20.6)$$

We now insert (20.3) into (20.4) and find

$$F_{CCF}(\tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Psi_i(\tau') C(t - \tau') d\tau' \right] C(t - \tau) dt . \quad (20.7)$$

Now change the order of integration and then

$$F_{CCF}(\tau) = \int_{-\infty}^{\infty} \Psi_i(\tau') \left[\int_{-\infty}^{\infty} C(t - \tau') C(t - \tau) dt \right] d\tau'. \quad (20.8)$$

$$= \int_{-\infty}^{\infty} \Psi_i(\tau') F_{ACF}(\tau - \tau') d\tau'. \quad (20.9)$$

This calculation demonstrates the pleasant result that the cross-correlation function $F_{CCF}(\tau)$ is the convolution of the autocorrelation function $F_{ACF}(\tau)$ and the transfer function $\Psi_i(\tau)$. This means that, in principle, we can now use the convolution theorem to recover the Fourier transform of the transfer function, but this procedure is only successful if the data are of very high signal-to noise ratio.

An example of the application of the cross-correlation procedure to a beautiful set of optical and ultraviolet observations of the Seyfert 1 galaxy NGC 5548 is shown in Fig. 20.5 and the lags associated with various emission lines are tabulated in Table 20.2. Notice that these observations probe structures in the nuclear regions on the scale of light-days. It is evident that the transfer functions are different for species in different states of ionisation. As expressed by Peterson (1991), lines which are prominent in highly ionised gases, He II, N V and C IV, have shorter lag times than those prominent at lower ionisation levels, for example, the Balmer lines. There must therefore be stratification in the ionisation structure of the clouds about the active nucleus. For example, the C IV lines originate from quite different regions from those of C III].

The data can be used to estimate the mass of the central dark massive object. The light travel times from the nucleus to the emitting clouds are well determined by the cross-correlation analysis, but a good mass estimate depends upon being convinced that the virial theorem can be applied using the observed velocity widths of the lines as a measure of the orbital motions of the clouds about the nucleus. In principle, this can be determined by studying the cross-correlation function of the intensity variations of the line luminosities at different velocities, but this requires very high quality data. According to Peterson (1997), for NGC 5548, predominantly radial outflow can be excluded and so the virial theorem can be used to make an estimate of the nuclear mass. For the intense C IV line, appropriate values are $\tau = 10$ days and $\Delta v_{FWHM} = 4,500 \text{ km s}^{-1}$, and Peterson finds $M = rv^2/G \sim 4 \times 10^7 M_\odot$.

If the broad-line regions are associated with clouds orbiting an active galactic nucleus and the broadening of the lines is attributed to these motions, it is expected that there should be a correlation between linewidth and the lag since, according to Kepler's laws of planetary motion, $v \propto r^{-1/2}$ and so $\Delta v \propto \tau^{-1/2}$. Figure 20.6 shows this correlation for Seyfert galaxies NGC 7469 and 5548 and the broad-line radio galaxy 3C390.3 (Osterbrock and Ferland, 2005), the lighter line showing the best-fit line of each galaxy and the solid line the expected relation $\Delta v \propto \tau^{-1/2}$. The agreement with the simple theory is satisfactory, the masses of the active nuclei in NGC 7469 and 3C 390.3 being found to be $8.4 \times 10^6 M_\odot$ and $3.2 \times 10^8 M_\odot$, respectively.

The same technique has been used by Suganuma and his colleagues (2006) to determine the optical-infrared time lags for four Seyfert 1 galaxies as part of their MAGNUM programme. Infrared continuum observations in the J , H and K wavebands for the galaxies

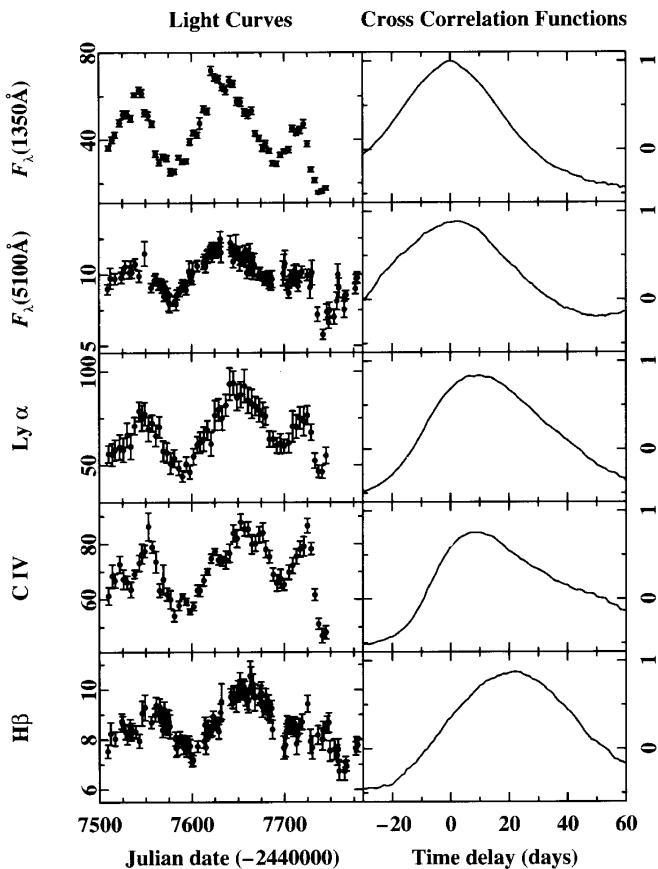


Fig. 20.5 Light curves showing the time variation of the ultraviolet (135 nm) and optical (510 nm) continuum intensity, as well as the fluxes of three prominent emission lines, Lyman- α (121.6 nm), C IV (154.9 nm) and H β (510 nm) in the Seyfert 1 galaxy NGC 5548 (Clavel *et al.*, 1991; Peterson *et al.*, 1991). The ultraviolet observations were made with the IUE observatory and the optical observations with ground-based telescopes. In the right-hand column, the cross-correlation functions of the time variation of the ultraviolet continuum intensity with the optical continuum and the different emission lines are shown, the first panel being the autocorrelation function of the ultraviolet continuum. The emission lines show the same pattern of variation as the ultraviolet and optical continuum, but with a time delay due to the light-travel time from the nucleus to the broad-line regions (see Table 20.1). (From Peterson (1997).)

NGC 5548, NGC 4051, NGC 3227 and NGC 7469 complemented existing observations which established time lags for the strong permitted lines in the ultraviolet-optical wavebands. The colour temperatures of these Seyfert 1 nuclei in the near-infrared waveband are all about 1500–1800 K, similar to the values found for many similar Seyfert 1 galaxies. This radiation is interpreted as the emission of heated dust in a dusty torus surrounding the active nucleus, as inferred to be present according to the unified picture of Seyfert 1 and 2 galaxies illustrated in Fig. 18.12a. A thermal temperature of 1500–1800 K corresponds to the sublimation temperature for heated dust grains and provides a natural explanation for the constancy of the near-infrared colour temperature. If this interpretation is correct, the

Table 20.2 The cross-correlation lags between the ultraviolet continuum intensity and the emission lines observed in the spectrum of NGC 5548 (Peterson, 1997).

Feature	Time lag (days)	Feature	Time lag (days)
N v $\lambda 1240$	2	C IV $\lambda 1549$	10
He II $\lambda 1640$	2	H γ $\lambda 4340$	13
'Small blue bump'	6	H α $\lambda 6563$	17
He II $\lambda 4686$	7	H β $\lambda 4861$	20
He I $\lambda 5876$	9	C III] $\lambda 1909$	22
Lya $\lambda 1216$	10		

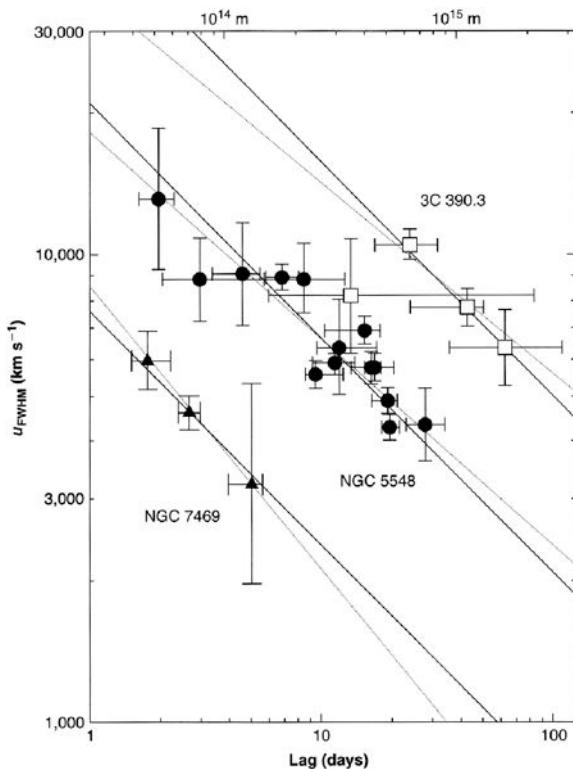


Fig. 20.6 The linewidth–time-lag correlation diagram for the Seyfert galaxies NGC 7469 and 5548 and the broad-line radio galaxies 3C 390.3 (Osterbrock and Ferland, 2005).

time lags of these active nuclei in the infrared waveband should be greater than those of the broad-line emission regions. The observed V and K intensity variations of NGC 5548 are shown in Fig. 20.7, the dotted line showing the intensity variation in the K -waveband advanced by 60 days, very significantly greater than the lags listed in Table 20.2. The results of a cross-correlation analysis found that the lags for the four Seyfert 1 galaxies are

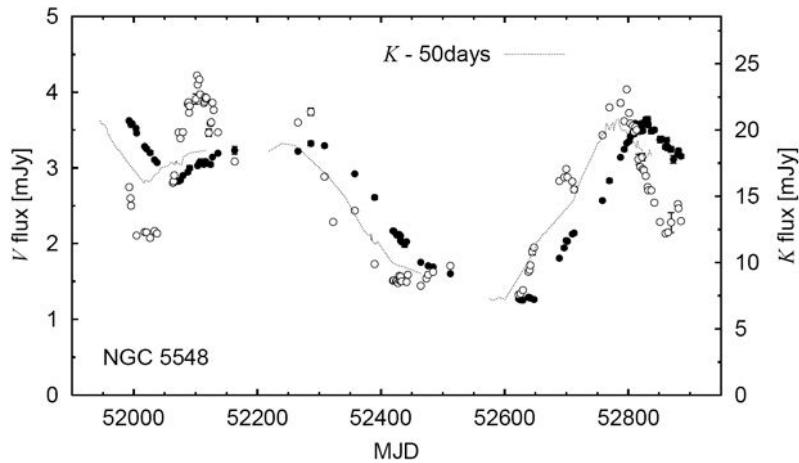
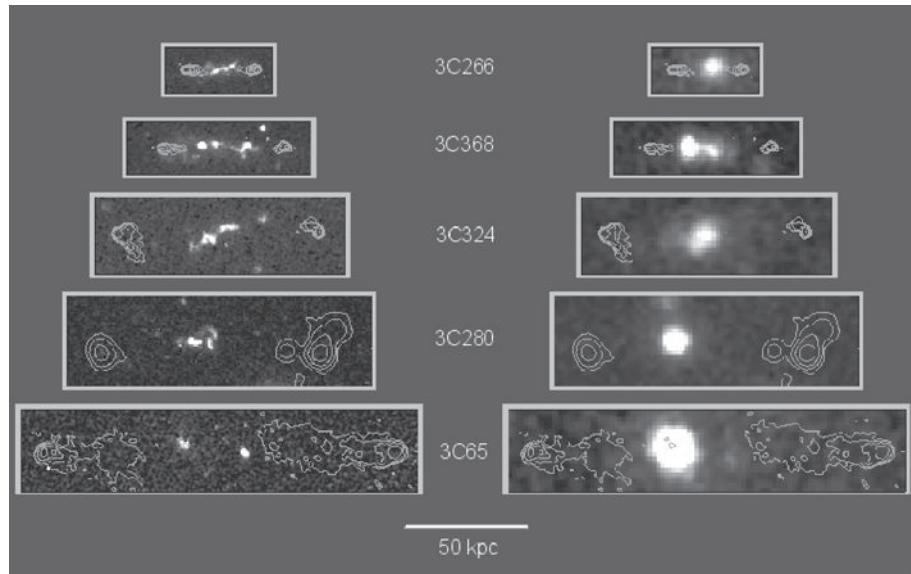


Fig. 20.7 The V and K light curves of the nucleus of NGC 5548. Open and filled circles show the V and K magnitudes, respectively. The dotted line is the K light curve shifted backwards by 50 days (Suganuma *et al.*, 2006).

47–53 days for NGC 5548, 11–18 days for NGC 4051, about 20 days for NGC 3227, and 65–87 days for NGC 7469. Furthermore, the time lag would be expected to be correlated with the continuum luminosity of the active nucleus. The flux density at the inner edge of the torus is $L/4\pi r^2$ and, in a simple approximation, this can be equated to the black-body emission per unit area of the torus, σT_{dust}^4 . Since the sublimation temperature T_{dust} is assumed to be fixed and the time lag τ is proportional r , it follows that we would expect $\tau \propto L^{1/2}$. This is the result found by Suganuma and his colleagues for their four Seyfert 1 galaxies and other examples from the literature. Finally, the cross-correlation between the X-ray variability measured by the Rossi X-ray Timing Explorer and the optical-ultraviolet continuum radiation of NGC 5548 has shown that the optical-ultraviolet emission lags behind the X-ray emission by $\tau = 1.6^{+1.0}_{-0.5}$ days. This observation supports the view that the optical-ultraviolet continuum radiation results from the thermal reprocessing of the X-ray emission.

20.6 The alignment effect and shock excitation of emission line regions

Photoexcitation is the dominant process of line emission in active galactic nuclei but in some circumstances shock excitation may play a significant role. One example is associated with the *alignment effect* discovered by McCarthy, Chambers and their colleagues (McCarthy *et al.*, 1987; Chambers *et al.*, 1987). They showed that the optical images of the luminous 3CR radio galaxies at redshifts $z \approx 1$ are aligned with their extended radio structures. As discussed in the next chapter, the radio hot spots and extended radio lobes observed in FR2 radio galaxies are powered by jets of relativistic material originating in the vicinity of the

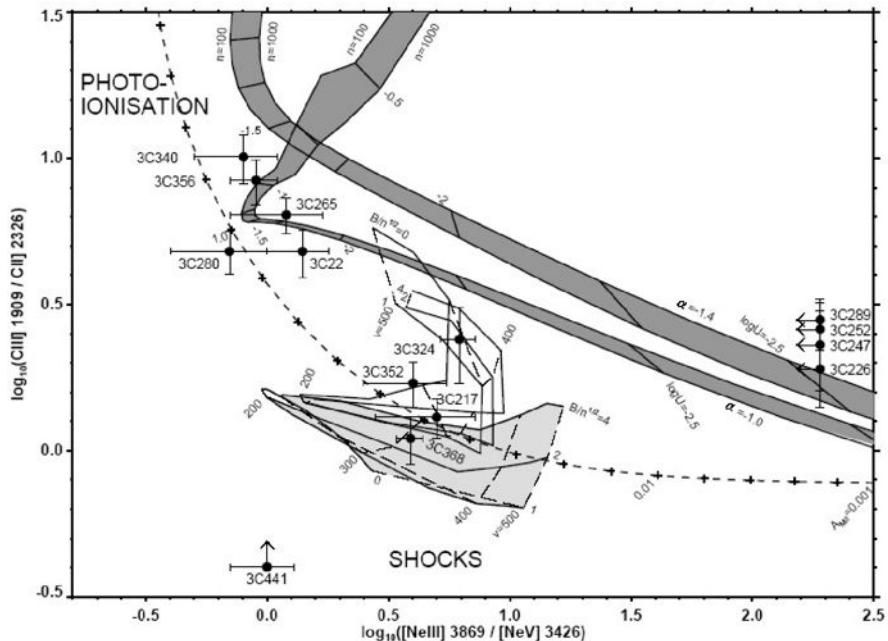
**Fig. 20.8**

HST images (left column) and the corresponding UKIRT images (right column) of the radio galaxies 3C 266, 368, 324, 280 and 65 with the VLA radio contours superimposed (Best *et al.*, 1996). The images are drawn on the same physical scale. The angular resolution of the HST images is 0.1 arcsec while that of the ground-based infrared images is about 1 arcsec.

black hole in the active galactic nucleus. These jets are highly supersonic with respect to the sound speed in the surrounding interstellar and intergalactic medium and so strong shocks are driven through the ambient gas.

The nature of the alignment effect was clarified by images of the optical structures of the radio galaxies taken with the Hubble Space Telescope by my colleagues Best and Röttgering and me (Best *et al.*, 1997). These observations were complemented by infrared images of the galaxies at $2 \mu\text{m}$ observed with the UK Infrared Telescope and by high resolution radio images obtained using the VLA. A sequence of radio galaxies at $z \approx 1$, all displaying the alignment effect, is shown in Fig. 20.8, the images all being drawn on the same physical scale. The contour lines show the structure of the radio lobes. In each of the five examples shown in Fig. 20.8, the infrared images to the right of the diagram show little evidence of the alignment effect. Modelling of the stellar energy distribution of these galaxies in the optical and infrared wavebands indicates that the infrared images are dominated by the light of the underlying giant elliptical galaxies which are the host galaxy bodies of the powerful radio galaxies observed at small redshifts. The alignment effect seen in the optical waveband in the left-hand column of Fig. 20.8 is primarily associated with strong narrow emission lines, as was demonstrated by long-slit optical spectroscopy.

The line ratios observed in the spectra of these aligned structures enable the excitation mechanism to be determined. The alternatives of photoionisation and shock excitation result in somewhat different predictions for the line ratios of certain elements in different states of ionisation. The reason for these differences is that, in the case of photoexcitation, the broadband nature of the ionising continuum spectrum results in a wide range of different ionisation

**Fig. 20.9**

The diagnostic emission line diagram for the line emitting regions associated with the alignment effect in 3CR radio galaxies at redshifts $z \simeq 1$ (Best *et al.*, 2000). The diagnostic tools used to determine the excitation of the emission lines are the line ratios $\text{CIII]}(1909)/\text{CII]}(2326)$ against $[\text{NeIII}](3869)/[\text{NeV}](3426)$. The areas of the diagram occupied by emission line clouds excited by photoionisation and shock excitation are shown. The upper shaded regions correspond to simple photoionisation models ($\alpha = 1.0$ and $\alpha = 1.4$), as described in Section 2.2. The dashed line corresponds to the sequence for photoionisation models including matter bounded clouds described in Section 2.3. The lower shaded region covers the ratios predicted by the shock models described in Section 2.4; the unshaded region just above this corresponds to the shock models including a precursor region (see Section 2.4). The five galaxies plotted towards the edge of the diagram have no data available for one of their emission line ratios.

states for any particular element, whereas in the shock excitation process, the emission line regions behind the shock waves tend towards roughly the same temperature regime and so a much narrower range of ionisation states is expected from the same element. Our observations of the line ratios $\text{CIII]}(1909)/\text{CII]}(2326)$ against $[\text{NeIII}](3869)/[\text{NeV}](3426)$ are compared with the expectations of the models of the excitation process in Fig. 20.9, in which the predictions of the shock excitation models are derived from the analyses of Dopita and Sutherland (1996). The diagram shows clearly how the two different models give rise to clustering of the radio galaxies in different locations on the diagnostic line diagram (Best *et al.*, 2000). It is intriguing that the excitation mechanism for the radio sources with physical scales less than 120 kpc, those illustrated in Fig. 20.8, is shock emission, while in the larger aligned sources can be associated with the passage of the strong shock associated with the supersonic expansion of the lobes of the radio source components through ambient gas clouds in the interstellar and intergalactic medium about the radio galaxies.

Dopita and Sutherland argue that shock excitation is likely to be important wherever strong shocks are expected, for example, in the nuclear regions of starburst and ultra-luminous far-infrared galaxies (ULIRGs)

20.7 Accretion discs about supermassive black holes

Much of the formalism developed for the study of accretion and accretion discs about compact stars developed in Chap. 14 can be adapted to the study of accretion discs about supermassive black holes in active galactic nuclei. The study of thin accretion discs illustrated the effectiveness of the accretion process in understanding many of the key features of high energy phenomena associated with white dwarfs and neutron stars, particularly when the role of magnetic fields is included.

In the case of supermassive black holes, the key difference is that they do not possess a solid surface. In the cases of white dwarfs and neutron stars, the binding energy of the accreted matter can be fully recovered by dissipation in the vicinity of that surface. In the case of black holes, there exist solutions of the equations of fluid flow in their vicinity which result in the *advection* of the energy of the flow through the last stable orbit and the event horizon into the black hole without the dissipation of energy. These stable flows are known as *advection dominated accretion flows* or ADAFs. They can have very low radiative efficiency η relative to the standard results quoted in Sect. 13.11. As a result, these flows can result in accretion rates onto black holes which exceed the rate inferred from the limits derived from the Eddington limiting luminosity. An implication of these types of accretion flow is that the presence of an accreting supermassive black hole in an active galactic nucleus does not necessarily mean that all the binding energy of the accreted material is radiated away, so that even if there are accreting massive black holes in all active galactic nuclei, they are not necessarily hyperluminous sources. In the pioneering paper by Abramowicz and his colleagues (1978), the objective was to explain why some of the radio galaxies which were inferred to be generating powerful beams of relativistic material were not hyperluminous optically. The development of these models was further stimulated by the fact that most elliptical galaxies are embedded in cooling flows which would provide large inflow rates onto the black holes in their nuclei. Yet, few of these giant elliptical galaxies are observed to be intense X-ray sources with the luminosities expected from the observed rate of mass inflow (Fabian and Rees, 1995).

The analysis of accretion flows in the general case is a major undertaking and some flavour for what is involved is conveniently summarised in Chap. 11 of the book *Accretion Power in Astrophysics* by Frank, King and Raine (2002). They point out that low radiative efficiencies may result either because the accretion rate is low and so the accreting gas is unable to cool, or because the accretion rate and particle density are so high that the flow is extremely optically thick with the result that the radiation is trapped and is dragged with the plasma into the black hole.

The thin discs studied in Sect. 14.3 were analysed assuming that the dissipation rate by radiation is exactly balanced by the rate of energy generation by viscosity in the accretion

disc and, although we did not prove this, the resulting accretion disc is stable. This energy balance is not necessarily maintained under the conditions discussed in the last paragraph. The challenge is therefore to find steady-state solutions in which the energy input and dissipation rate are not necessarily in balance. In the general case of three-dimensional gas dynamics, this is a highly non-trivial problem. Frank and his colleagues show how the problem can be reduced to manageable proportions by considering discs in which there are small velocity components perpendicular to the disc. They write the energy equation for flows in such discs as follows:

$$q_{\text{adv}} = \varrho v_r T \frac{ds}{dr} = q_+ - q_- , \quad (20.10)$$

where s is the specific entropy of the flow and q_+ and q_- are the volume heating and cooling rates respectively – the vertical advection of energy has been neglected. When $q_{\text{adv}} \ll q_+ \approx q_-$, the standard thin disc solutions are recovered. The case $q_- \ll q_+ \approx q_{\text{adv}}$ corresponds to the advection-dominated solutions we are interested in here. The energy generation rate is the same as that deduced in Sect. 14.3.3, namely

$$q_+ = \varrho v r^2 \left(\frac{d\Omega}{dr} \right)^2 , \quad (20.11)$$

and the energy loss rate by the generalised cooling function by two-body interactions,

$$q_- = n_e n_i \Lambda(T_e) . \quad (20.12)$$

These are the heating, cooling and advection rates which are included in the equations of conservation of mass, angular momentum and energy, which can be written

$$\begin{aligned} \dot{M} &= -2\pi r \Sigma v_r , \\ \dot{M} &= 3\pi v \Sigma g / f_* , \\ Q_+ &= Q_- + Q_{\text{adv}} . \end{aligned}$$

In these equations, Σ is the surface density integrated through the width of the disc; the local kinematic viscosity v is parameterised in terms of the α parameter introduced in Sect. 14.3.2, $v = \alpha c_s H$; f_* is a function which is close to unity except near the inner boundary and $g = -\frac{2}{3}(\frac{d \ln \Omega}{d \ln r})$.

The steady-state solutions at a fixed radius r can be displayed on a diagram in which the accretion rate \dot{m} is plotted against surface density Σ . In Fig. 20.10a, the accretion rate has been normalised to the critical Eddington rate, which corresponds to the Eddington critical luminosity

$$L_E = \frac{2\pi r_g m_p c^3}{\sigma_T} = \frac{1}{2} \dot{m}_{\text{crit}} c^2 \left(\frac{r_g}{r} \right) , \quad (20.13)$$

where $r \approx r_g$. The diagram illustrates the properties of typical discs at radii $r \lesssim 10^2 r_g$ and spans a range of about 10^6 along each axis. Specific scales are not shown because they are dependent upon the detailed physics of the model – the aim of the diagram is to classify the solutions without going into the somewhat complex details. The steady state solutions are

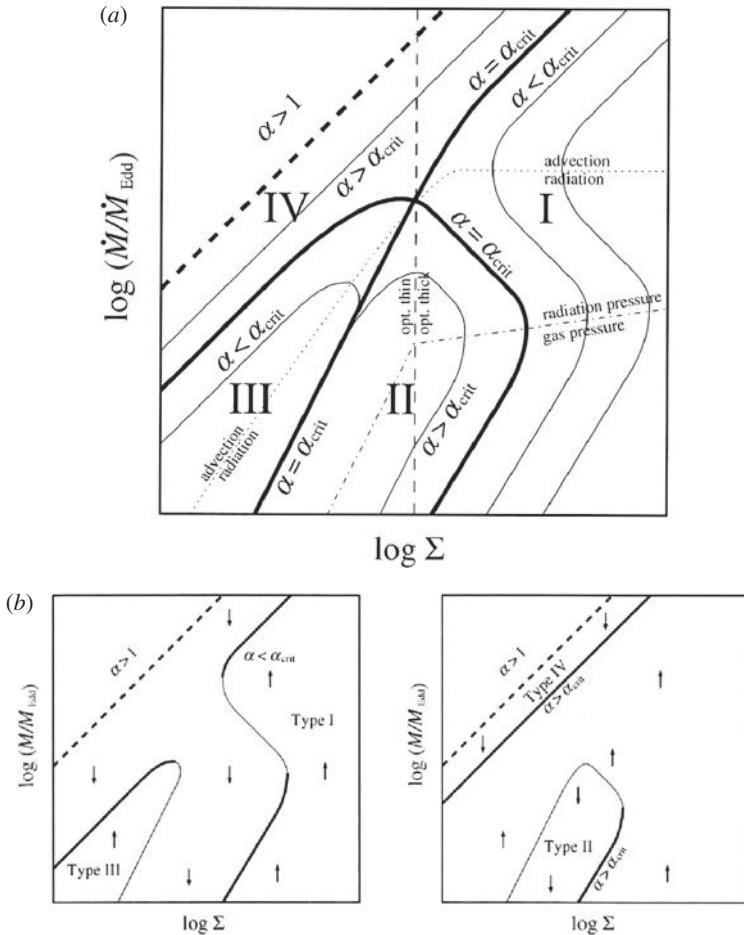


Fig. 20.10 (a) A plot of the accretion rate in terms of the critical Eddington rate against the surface density of the disc showing the loci of solutions of the equations of gas dynamics for accretion discs. (b) Possible local disc solutions at some radius r for $\alpha < \alpha_{\text{crit}}$ (left) and $\alpha > \alpha_{\text{crit}}$ (right). The thick lines indicate those branches which are thermally and viscously stable. The arrows show the direction of evolution of the disc annuli which are not on the local equilibrium curve (Frank *et al.*, 2002).

indicated schematically in Fig. 20.10a where they are parameterised in terms of the value of the α -parameter. The heavy lines labelled α_{crit} divide the plane into four regions labelled I to IV which correspond to different classes of solution of the dynamical equations. The intersection point of the two heavy loci corresponds to accretion rates $0.1 \lesssim \dot{m}/\dot{m}_{\text{crit}} \lesssim 10$ and surface densities $1 \lesssim \Sigma(\text{g/cm}^2) \lesssim 10^2$.

A number of lines divide the plane into regions in which different physical processes are dominant. These are:

- The vertical dashed line passing through the point at which the α_{crit} loci intersect divides optically thin from optically thick regions.

- The dotted-dashed line which partitions off the lower right region of the diagram separates those regions in which the gas pressure is dominant from those to the top left where radiation pressure is dominant.
- Further towards the top left, the dotted line divides solutions in which radiation is dominant from those in which advection dominates.
- Finally, to the top left is a heavy line to the left of which $\alpha > 1$. These solutions are unphysical since the scale height of the turbulent eddies would be greater than the thickness of the disc.

It can be seen that the changes of slope of the solutions on the diagram correspond roughly to changes from one set of physical conditions to another. For a given accretion rate, there are a number of possible solutions because the energy equation can be satisfied by different combinations of radiation and advective energy loss in the optical thin and thick regimes. As shown by Frank and his colleagues, the various loci for a given value of α can be approximated by power laws in the different physical regimes. Thus, for the standard optically thick disc in which the gas pressure is dominant and the cooling by electron scattering, $\dot{m} \propto \Sigma^{5/3}$, whereas for advective cooling $\dot{m} \propto \Sigma$. In the case of a radiation-dominated, optically thick disc, the relation becomes $\dot{m} \propto \Sigma^{-1}$ and is unstable according to the analysis of Lightman and Eardley (1974).

The key point is that the discs can be thermally or viscously unstable. In the former case, the material of the disc would cool catastrophically, resulting in the formation of dense clouds within the flow – in the latter case, the accretion disc would fragment into discrete rings. An advantage of the presentation in Fig. 20.10a is that the slopes of the solutions of the equations for a given value of α indicate regions where viscous instabilities are important and the role of thermal instabilities can also be illustrated. Thus, all the loci with positive slopes on the diagram are viscously stable. In addition, the thermal stability of the solutions need to be considered. The vertical arrows in Fig. 20.10b indicate the direction of evolution of the thermal instability. Taking these considerations together, Frank, King and Raine and his colleagues show that the loci drawn as heavy lines in Figs. 20.10b are both viscously and thermally stable, the left-hand diagram for accretion with $\alpha \lesssim \alpha_{\text{crit}}$ and the right-hand diagram for $\alpha \gtrsim \alpha_{\text{crit}}$.

The importance of Fig. 20.10b is that it shows that, in addition to the standard solutions, there exist ADAF solutions at super-Eddington accretion rates and these have found application in those active systems which must be accreting at a high rate to account for their radio properties and yet were observed to be only weak active galactic nuclei at optical wavelengths. Detailed models have been developed for the emission spectra of such flows and these have been successfully fitted to the observations.

This exposition of the physics of accretion discs is only the beginning of a long and complex story. Frank and his colleagues emphasise the very large amount of analysis which has been carried out on various developments of the basic model and the care needed in interpreting the solutions. For example, in one development, outflow from the disc can be included so that the disc gives rise both to the fuelling of the black hole from the accretion disc, but also outflow from the disc in the form of winds, so that not all the accreted matter ends up being swallowed by the black hole. Such models in which the mass accretion rate

varies as $\dot{m} \propto r^p$ with $0 \lesssim p \lesssim 1$ are known as adiabatic inflow–outflow solutions (ADIOS) and have been developed by Blandford and Begelman (1999).

Another variant developed out of the realisation by Narayan and Yi (1994) that ADAFs are subject to a convective instability at low values of the viscosity parameter α . The convective motions redistribute the matter in the disc so that, rather than the standard ADAF density distribution $\varrho \propto r^{-3/2}$, the density profile is significantly flatter $\varrho \propto r^{-1/2}$ (Narayan *et al.*, 2000). In these convection dominated accretion flows (CDAFs), the angular momentum per unit volume turns out to be more or less constant, the inward transfer of angular momentum by convection being balanced by the outward transport by viscous forces.

21.1 Extended radio sources – Fanaroff–Riley types

All galaxies are sources of radio emission – high energy electrons are accelerated in supernova remnants and these are dispersed throughout the interstellar medium where they radiate radio waves by synchrotron radiation. These are, however, very weak radio emitters indeed compared with what are conventionally referred to by the terms *radio galaxy* or *radio quasar* in which the radio luminosity can exceed that of our own Galaxy by factors of 10^8 or more. The big surprise was the discovery that the most luminous radio sources contain jets of relativistic material which give rise to a wide variety of large scale radio structures. The example of the brightest extragalactic radio source in the northern sky, Cygnus A, as observed by the Very Large Array, illustrates a number of the characteristic features of these sources (Fig. 21.1). Many more details are contained in the review by Carilli and Barthel (1996).

- The radio spectra of all regions of the radio structure have non-thermal spectra and the radiation is linearly polarised. These features make the identification of the radio emission as synchrotron radiation wholly convincing.
- The huge radio lobes are symmetrically disposed on either side of the active galactic nucleus but they extend far beyond the confines of the host galaxy as can be seen in Fig. 21.1a and b.
- At the leading edges of the radio lobes there are ‘hot-spots’ of very high radio surface brightness. These regions are powered by narrow radio jets which can be seen emanating from the nucleus of the radio galaxy in Fig. 21.1a and which are shown schematically in Fig. 21.1b.
- The spectral index α of the radio emission, defined by $I_\nu \propto \nu^{-\alpha}$, varies systematically from the leading edge where the hot spots are located to the inner edge of the lobe. The correlation is in the sense that $\alpha \approx 0.5$ at the hot spots and increases to values as large as 2.5 at the inner part of the lobe.
- There is a compact radio source in the nucleus of the galaxy which has physical size on the scale of parsecs. VLBI observations have shown that this radio source is elongated in the direction of the radio jets emanating from the nuclear regions of the galaxy.
- The galaxy is the most luminous member of a rich cluster of galaxies (Owen *et al.*, 1997). They find that the velocity histogram has two peaks, one centred on Cygnus A and a more significant peak redshifted by about 2060 km s^{-1} from the velocity of Cygnus A. The dynamical centroid of the spatial distribution of the galaxies is also shifted somewhat to

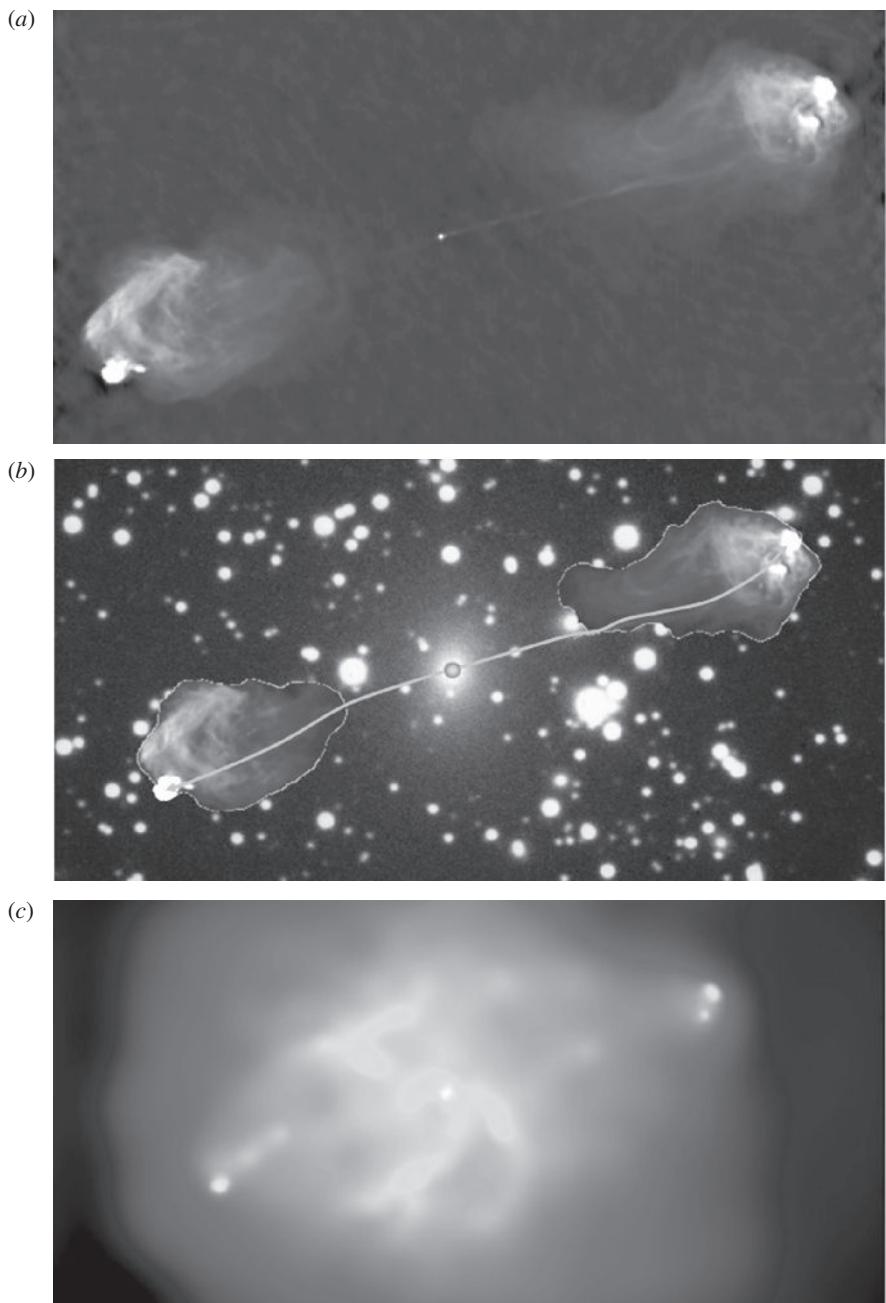


Fig. 21.1 (a) The radio source Cygnus A as observed by the US Very Large Array (VLA) (Perley *et al.*, 1984). There is a compact radio source coincident with the nucleus of the galaxy. A narrow collimated jet extends from the nucleus to the extended radio lobe to the right of the image and a similar fainter jet is present on the opposite side of the nucleus. The hot spots at the outer edges of the radio lobes are sites of particle acceleration. There is a great deal of structure in the lobes, probably associated with the escape of the particles from the hot spots and the interaction of this relativistic gas with the surrounding medium. (b) The radio structure of Cygnus A superimposed upon an optical image of the field showing relative scales of the host giant elliptical galaxy and the radio structure (optical image courtesy of Dr. Alan Stockton). The nucleus and the trajectory of the radio jets are shown schematically. (c) Cygnus A as observed by the Chandra X-ray Observatory. The diffuse X-ray emission consists of diffuse thermal bremsstrahlung as well as non-thermal radiation from the hot spots. (Courtesy of NASA and the Chandra Science Team.)

the north-west. The overall velocity dispersion of 41 galaxies for which radial velocities have been measured is 1581 km s^{-1} , consistent with the dispersions found in rich clusters.

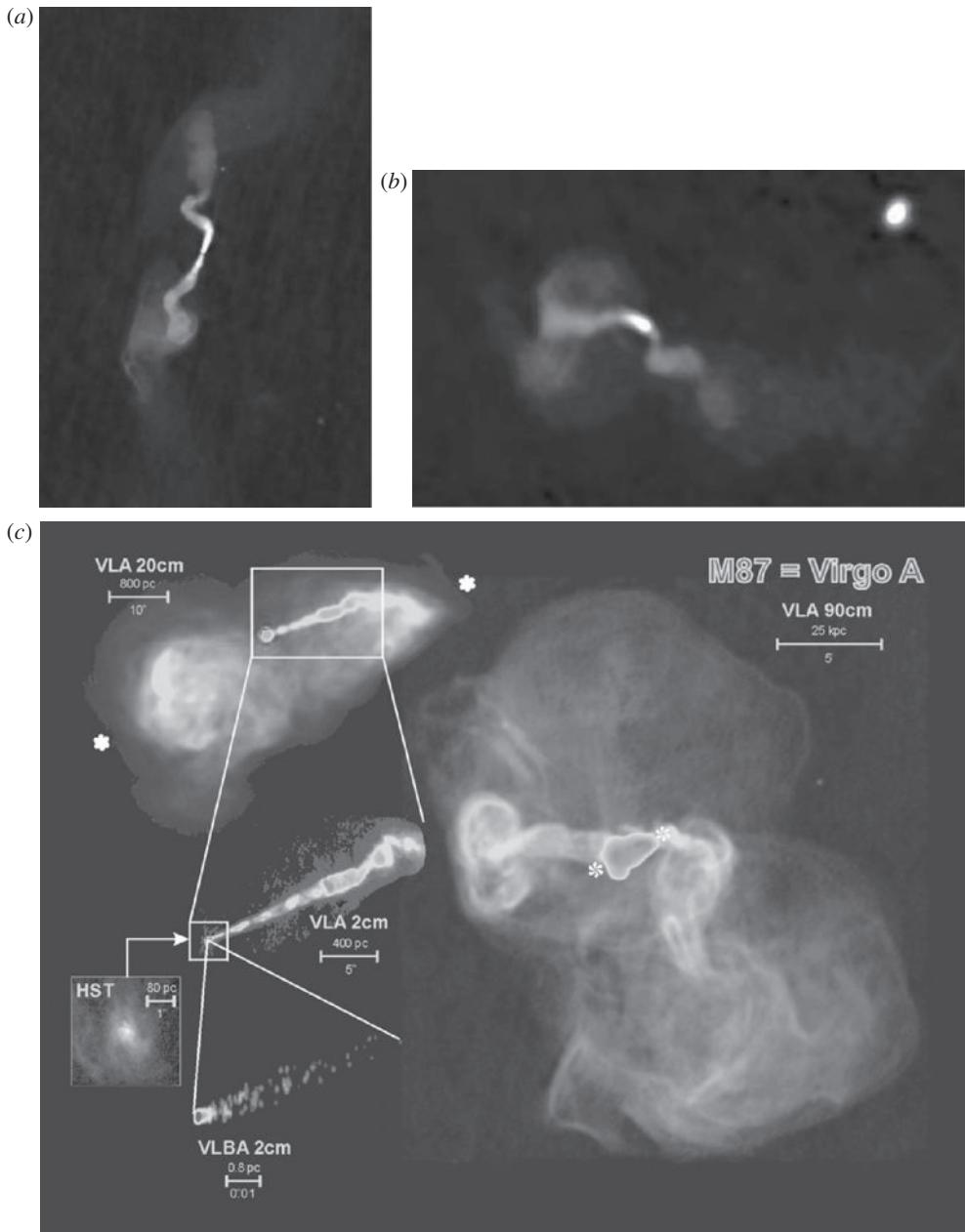
- The whole radio structure is embedded in the hot diffuse intergalactic gas in the cluster of galaxies. Figure 21.1c shows the distribution of X-ray emission as observed by the Chandra X-ray Observatory. Most of the X-ray emission is the bremsstrahlung of the hot intracluster gas, but that associated with the hot spots has a non-thermal spectrum.

The most luminous extragalactic radio sources selected at a low radio frequency, meaning at less than about 1 GHz, have this type of extended double structure on scales from tens of kpc to a Mpc or greater, the largest source having scale 6 Mpc. It is striking that the sources are self-similar in appearance, meaning that there are no clues from their structures whether they are physically large or small. The most luminous radio galaxies and essentially all the radio quasars selected at low frequencies have the same appearance, except that the radio sizes of the radio quasars are about half those of the radio galaxies. It was argued in Sect. 18.9.2 that the radio galaxies and radio quasars observed in bright source catalogues are consistent with a unification scenario in which the radio quasars are double radio sources in which the angle between the axis of the jet and the line of sight to the active galactic nucleus is less than about 45° while in the case of the radio galaxies, the angle is greater than 45° (Barthel, 1989, 1994).

The features of extended double radio structures with hot spots at the outer edges of the radio lobes are characteristic of the most luminous radio sources in the sense to be defined below. Less luminous sources have somewhat different radio structures with the regions of highest surface brightness lying much closer to the nucleus as is illustrated by the examples of 3C 31, 3C 66B and 3C 274 (Virgo A) shown in Fig. 21.2. In these example, the spectral index increases with distance from the active galactic nucleus. The high surface brightness regions are associated with intense jets of radio emission originating in the active galactic nucleus. In the spectacular case of the active galaxy M87 in the nearby Virgo cluster of galaxies (Fig. 21.2c), radio structure is observed on a very wide range of scales from sub-pc structure as observed by VLBI, through the well-known jet to the diffuse extended radio structures on the scale of 100–200 kpc. The intense optical jet coincides with the jet seen in the optical image (Fig. 3.2a). For comparison, the central regions observed by the Hubble Space Telescope are shown as an inset in Fig. 3.2c.

Among this class of lower luminosity radio sources are the ‘radio-trail’ sources observed in clusters of galaxies. In these cases, the motion of the galaxies through the intracluster gas results in ram-pressure distortion of the jets resulting in the galaxies leaving a trail of radio emission behind them. In the case of the weak radio source 3C 83.1B associated with the galaxy NGC 1265, the galaxy has a projected velocity of about $2\,000 \text{ km s}^{-1}$ with respect to the frame of reference of the cluster and so the radio jets are swept backwards by the ram pressure of the intracluster gas (Fig. 21.3a). In the case of the weak radio galaxy NGC1128 in the cluster Abell 400, both components of the dumbbell galaxies are radio sources with supermassive black holes in their nuclei and both radio source show distortions associated with their motion within their common envelope in the central regions of the cluster (Fig. 21.3b).

Fanaroff and Riley (1974) pointed out that there is a remarkably abrupt division in radio luminosity between what are often termed the classical double radio sources such as

**Fig. 21.2**

Examples of Fanaroff–Riley class 1 radio sources. (a) The radio source 3C 31 as observed by the US Very Large Array (VLA). (b) The radio structure of 3C 66B as observed by the VLA. (c) The radio structure of the FR1 radio source 3C 274 associated with the giant elliptical galaxies M87 in the nearby Virgo Cluster of galaxies. The inserts show the features observed on different physical scales from jets observed on sub-parsec scales in the nucleus by VLBI through the prominent radio jet which coincides with the optical jet (Fig. 3.2a) to the huge diffuse structure observed on the scale of 100 kpc. (Images courtesy of NRAO and AUI.)

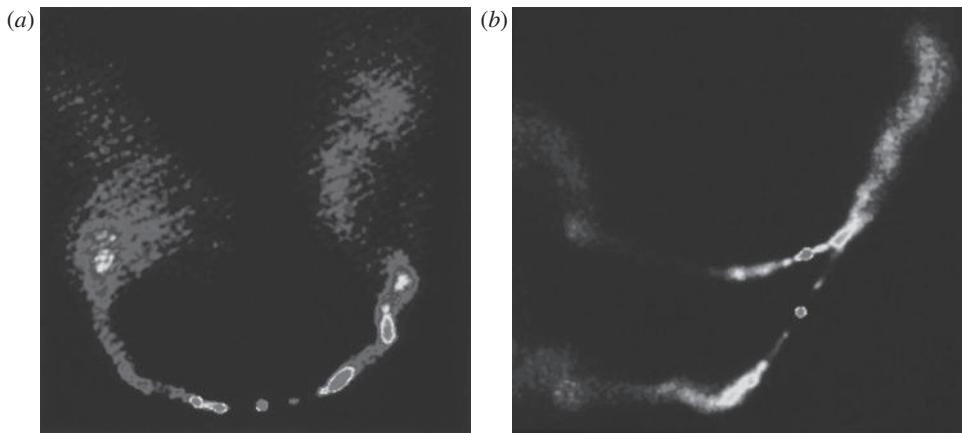


Fig. 21.3 Examples of ‘radio-trail’ radio sources. (a) The radio source 3C 83.1B associated with the galaxy NGC 1265 in the Perseus Cluster of galaxies as observed by the US Very Large Array (VLA). (b) The radio structure of 3C 75 as observed by the VLA. The radio image shows two radio trail sources associated with supermassive black holes in the dumbbell galaxy NGC 1128 in the central regions of the cluster Abell 400. (Images courtesy of NRAO and AUI.)

Cygnus A with radio lobes and radio hot spots at the advancing edge and those in which the maximum surface brightness occurs much closer to the nucleus, as in the cases of 3C 33, 3C 66.1A and the radio trail sources. The latter sources are referred to as Fanaroff–Riley class 1 (FR1) sources and the former as class 2 (FR2) sources. The abruptness of the change in character with radio luminosity is best illustrated by the diagram presented by Owen and Ledlow (1994) (Fig. 21.4). The dividing line between the FR1 and FR2 radio galaxies occurs about a radio luminosity of $10^{25} \text{ W Hz}^{-1}$ but Owen and Ledlow note that there is evidence that the change-over luminosity is a function of the absolute optical magnitude of the host galaxy.

All the radio galaxies plotted in Fig. 21.4 are giant elliptical galaxies with radio luminosities $L_{1.4\text{GHz}} \gtrsim 10^{23} \text{ W Hz}^{-1}$ at 1.4 GHz. Below this radio luminosity the character of the radio source population changes, as illustrated by the radio luminosity function presented by Condon (1989) (Fig. 21.5). The luminosity function of the extended double radio sources associated with giant elliptical galaxies flattens and the low radio luminosity region of the function is dominated by the emission of spiral galaxies. The majority of these sources are likely to be powered by supernova events, similar to those which give rise to the radio emission of our own Galaxy.

Support for this picture is provided by the strong correlation between the far-infrared luminosity and radio luminosity found for the brightest far-infrared sources in the IRAS survey. Figure 21.6 shows this correlation for the brightest far-infrared emitters selected at $60 \mu\text{m}$ from the IRAS survey (Condon, 1992). In selecting the sources shown in Fig. 21.6, those galaxies known to contain active galactic nuclei, such as the Seyfert galaxies, have been excluded. This tight correlation has been confirmed in larger deeper samples of star-forming galaxies. It finds a natural explanation in the association of star-forming regions. The progenitors of Type II supernovae, which undergo core collapse to form a neutron

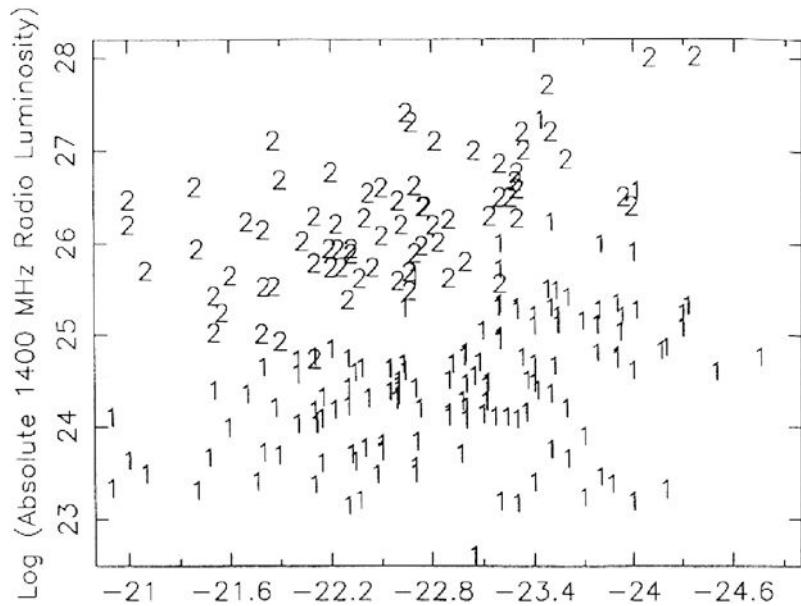


Fig. 21.4 A plot of radio luminosity at 1.4 GHz against optical absolute magnitude for bright radio galaxies and radio quasars, illustrating the distinction between Fanaroff Riley class 1 and 2 double radio sources (Owen and Ledlow, 1994).

star or black hole, are massive stars with relatively short lifetimes and so they release the high energy particles accelerated in supernova remnants while the star-forming region is still in an active phase. The strong far-infrared luminosities of these galaxies are taken as a measure of star-forming activity, often stimulated by strong gravitational interactions between galaxies (see Sect. 18.7).

21.2 The astrophysics of FR2 radio sources

The analysis of the astrophysical properties of the FR1 and FR2 radio sources involves the theory of synchrotron radiation developed in Chap. 8, as well as the various techniques for investigating the properties of interstellar and intergalactic gas and magnetic fields described in Chap. 12. A schematic model for FR2 radio sources is shown in Fig. 21.7a. A relativistic jet is emitted from the nucleus of the active galaxy which ‘burns’ its way out through the interstellar and intergalactic gas much like a laser beam. At the interface between the beam and the intergalactic gas there is a shocked region, similar in shape to the magnetospheric cavity about the Earth’s magnetic field (see Sects 11.4). Particles are accelerated in the shock waves at the advancing edge of the jet, just as in the case of the supernova remnants discussed in Sects 17.4 and 17.5. The acceleration of particles and the amplification of magnetic fields in the region where the jet encounters the interstellar or intergalactic medium gives rise to the hot spots at the advancing edges of the radio lobes. The relativistic gas and magnetic field expand forming a cavity, or cocoon, of mixed relativistic

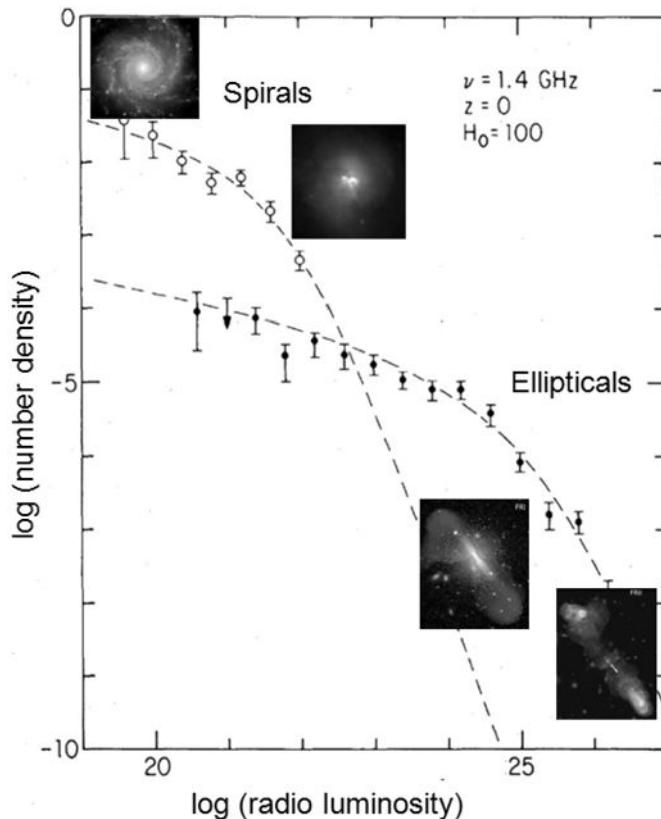
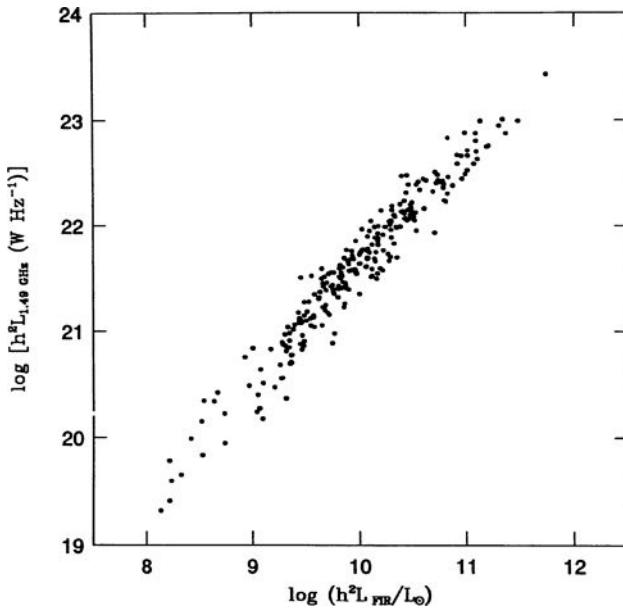


Fig. 21.5 The radio luminosity function at the present epoch, showing the functions separately for spiral and giant elliptical galaxies (Condon, 1989). The types of radio source associated with different radio luminosities are illustrated schematically: $L_{1.4\text{ GHz}} \gtrsim 10^{25} \text{ W Hz}^{-1}$, FR2 sources; $10^{25} \gtrsim L_{1.4\text{ GHz}} \gtrsim 10^{23} \text{ W Hz}^{-1}$, FR1 sources; $10^{23} \gtrsim L_{1.4\text{ GHz}} \gtrsim 10^{21} \text{ W Hz}^{-1}$, star-forming galaxies; $L_{1.4\text{ GHz}} \lesssim 10^{21} \text{ W Hz}^{-1}$, spiral galaxies. Note that, because of the effects of cosmological evolution, the relative contributions of spirals and ellipticals changes with redshift.

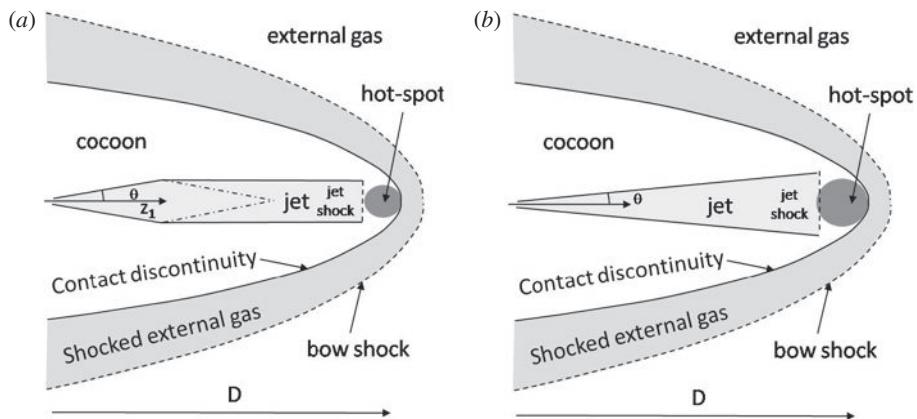
plasma and magnetic fields which give rise to the diffuse radio lobes observed in FR2 sources. When the jet switches off, the diffuse lobes dissipate their energy by synchrotron radiation or inverse Compton scattering of the Microwave Background Radiation or by adiabatic losses. Let us show how this picture comes about more quantitatively.

21.2.1 Energetics and energy densities

We begin with an assessment of the energies and energy densities involved in FR2 radio sources, taking Cygnus A (Fig. 21.1) as the archetype for such sources. It is simplest to begin with the expression for the minimum energy requirements for synchrotron radiation given in Sect. 16.5. It was emphasised in that section that these are convenient reference values, because the results depend upon the filling factor and the ratio of energy densities of protons to electrons, η . The minimum energy requirements are also close to the condition

**Fig. 21.6**

The far-infrared–radio correlation for bright galaxies selected at $60\text{ }\mu\text{m}$ from the IRAS sky survey (Condon, 1992). All classes of active galaxy have been excluded from this plot. The abscissa is the far-infrared luminosity of the galaxy measured in solar units and the ordinate the radio luminosity in W Hz^{-1} at 1.49 GHz .

**Fig. 21.7**

Models for one side of FR2 double radio sources. (a) A typical FR2 source such as Cygnus A including self-collimation of the jet of relativistic material. (b) A model for compact symmetric sources observed in the nuclei of galaxies in which the jet propagates ballistically from the nucleus (after Alexander, 2006).

that there are equal energy densities in the high energy particles and the magnetic field (see (16.46)). While it is plausible that the source components may well be close to equipartition following the arguments of Sect. 17.5.1, there is no guarantee that this condition actually holds throughout the volume of the radio emitting regions.

Granted these caveats, the energy requirements for the extended lobes of Cygnus A are of the order of $2 \times 10^{52} \eta^{4/7}$ J corresponding to the rest-mass energy of $3 \times 10^5 \eta^{4/7} M_\odot$ of matter. The corresponding equipartition magnetic field strengths in the extended radio lobes are of the order of 7 nT, corresponding to a minimum energy density of $\varepsilon = 5 \times 10^{-11} \text{ J m}^{-3}$ and a minimum pressure $p = \frac{1}{3} \varepsilon$ of $10^{-11} \text{ N m}^{-2}$. Within the hot spots, the energy densities and pressures are greater, the equipartition magnetic field strengths being of order 25 nT, corresponding to a minimum energy of 10^{-9} J m^{-3} and a minimum pressure $p \sim 3 \times 10^{-10} \text{ N m}^{-2}$. The energy densities and pressures in the synchrotron jet are of the same order as those in the hot spots.

These figures can be compared with the pressure of the hot intergalactic gas surrounding the Cygnus A galaxy which is shown in Fig. 21.1a. In the vicinity of the radio lobes, the electron number density is $n_e \approx 10^4 \text{ m}^{-3}$ and the temperature of the gas $kT = 7.7 \text{ keV}$. Therefore, the pressure of the intergalactic gas in the vicinity of the lobes is $p = n_e kT \approx 10^{-11} \text{ N m}^{-2}$. Thus, while the lobes may be approaching pressure balance with the intergalactic gas, the pressure would be inadequate to confine the hot spots.

Another way of constraining the energy density in the magnetic field in the hot spots is to compare their X-ray and radio luminosities (Wilson *et al.*, 2000). The hot spots in the radio and X-ray images in Figs 21.1a and c coincide and a natural interpretation of this observation is that the X-ray emission is the synchrotron self-Compton radiation of the electrons responsible for the radio emission (Sect. 9.6). As shown in that section, the ratio of the total luminosities in the X-ray and radio wavebands depends upon the ratio of the energy density in radio photons to the magnetic energy density. Wilson and his colleagues (Wilson *et al.*, 2000) show that the energy density of radio emission in the hot spots is 100 times the energy density in the Cosmic Microwave Background Radiation but still about 100 times less than the energy density in the equipartition magnetic field. They modelled the radio–X-ray emission spectrum using the codes of Band and Grindlay discussed in Sect. 9.6 for the spherically symmetric model with the results shown in Fig. 21.8. It can be seen that the total energy emitted in the X-ray waveband is about two orders of magnitude less than that emitted in the radio waveband. The inferred magnetic flux density in the bright hot spots A and D are 15 nT, close to the equipartition fields discussed above. They make the point that this is a lower limit to the magnetic flux density within the source regions – if the magnetic flux density were any lower, the X-ray intensity would exceed that observed.

21.2.2 Synchrotron losses and time-scales

Cygnus A is such a bright radio source that high resolution spectral mapping of its radio structure has enabled time-scales to be estimated for different regions of the source. The technique is illustrated in Fig. 16.1 which shows the steepening of the radio spectrum as a result of synchrotron losses. In a simple picture, it would be expected that as the source ages, the spectral break seen in Fig. 16.1a would move to lower frequencies. Figure 21.9 shows that the radio spectral index α changes from about 0.5 in the hot spots to much larger values towards the regions of the lobes further from the hot spots. In fact, in all regions of the source, the spectral index at low radio frequencies tends to the value 0.5, consistent with the spectral break moving to lower frequencies further from the hot spots.

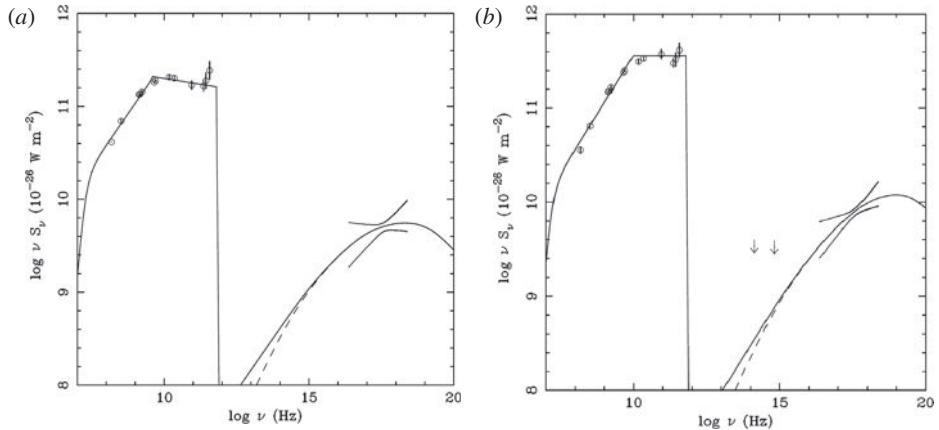


Fig. 21.8 (a) The radio–X-ray spectrum of hot spot A of Cygnus A. The points show the radio flux densities and the line through them the model of the synchrotron radiation. The X-ray limits show the 90% confidence limits for the X-ray spectrum as observed by the Chandra X-ray Observatory. The solid line is the predicted synchrotron self-Compton spectrum for lower limits to the electron spectrum of $\gamma_{\min} = 1$ and the dashed line for $\gamma_{\min} = 100$. (b) The same as (a), but for hot spot D (Wilson *et al.*, 2000).

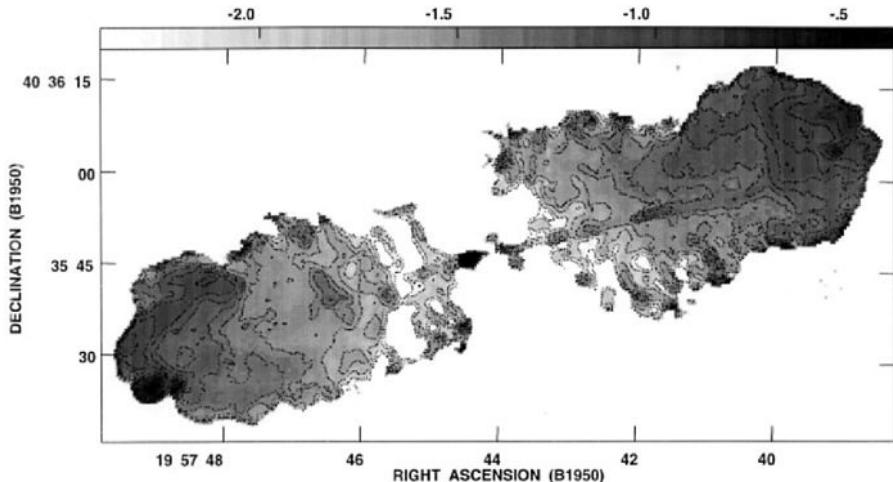


Fig. 21.9 The spectral index distribution across Cygnus A between 1.5 and 5 GHz at an angular resolution of 1.1 arcsec (Carilli and Barthel, 1996). The grey scale runs from 2.4 (white) to 0.4 (black), using the convention that the spectral index is defined by $S_\nu \propto \nu^{-\alpha}$.

To convert these data into typical ages for the electrons in different regions of the source, equipartition magnetic flux densities have been adopted at each point in the radio lobes. The ages for the electrons in different regions of the source can then be found from the frequency at which the break in the radio spectrum is observed through the relations which can be readily derived from (8.124) and (8.127). The maximum of the synchrotron spectrum

occurs close to frequency $\nu_{\text{crit}} = 2.199 \times 10^{10} \gamma^2 B$. Therefore the synchrotron loss-time for an electron of energy $E = \gamma m_e c^2$ is

$$t_{\text{sync}} = \frac{E}{dE/dt} = 1.6 \times 10^6 B^{-3/2} \nu_{\text{crit}}^{-1/2} \text{ s}, \quad (21.1)$$

where B is measured in tesla and ν_{crit} in Hz. Many authors who have analysed this problem agree that, according to this approach, the lifetimes of the ultra-relativistic electrons in the vicinity of the hot spots are of the order of 10^6 years and that this increases to about 5×10^6 years at the regions of the lobes furthest from the hot spots (Alexander *et al.*, 1984; Carilli *et al.*, 1991). Using the same formalism, the lifetimes of the electrons within the hot spots are only about 10^5 years.

A number of conclusions follow from these results.

- The light travel time from the nucleus of Cygnus A to the hot spots is at least 200 000 light-years, but the lifetimes of the electrons within the hot spots are less than 10^5 years. Consequently, the hot spots need to be continually supplied with energy and the obvious way in which this takes place is through the continuous supply of energy from the nucleus by the jets. As a result, the particles must be accelerated in shock waves in the vicinity of the hot spot, either at the interface to the undisturbed intergalactic medium, or in a reverse shock at the end of the jet which supplies the hot spot with a continuous supply of energy.
- The overall age of the source must be only about 10^7 years and so the rate at which the interface between the jet and the intergalactic medium advances must be at least $0.02c$. Taking the temperature of the intergalactic gas to be 7.7 keV, the speed of sound is about $0.004c$ and so the hot spot must advance supersonically through the intergalactic medium, with Mach number 5 or greater.
- The rate of supply of energy from the active galactic nucleus during the lifetime of the source can be found from the energy required to account for the synchrotron radio emission and the age of the source, say, 10^7 years. Taking the minimum energy requirement to be 2×10^{52} J, the rate of supply of energy to create the observed relativistic particles and magnetic fields must be at least 10^{38} W. The Eddington luminosity is $1.4 \times 10^{31} (M/M_\odot)$ W and so the mass of the central black hole must be at least $10^7 M_\odot$. Thus, although the radiated energy in the form of radio waves is significantly less than the energy emitted in the optical and X-ray wavebands, the amount of energy which has to be supplied by the jets originating in the active nucleus is of the same order as the non-thermal luminosities of some of the most extreme active galactic nuclei.

This analysis explains the origin for the structures shown in Fig. 21.7a. Other features related to it can be seen in the X-ray image of the sources (Fig. 21.1c). Except for the hot spots, the features seen in the image are all the bremsstrahlung of hot intracluster gas. The analysis of Wilson and his colleagues (Wilson *et al.*, 2002) shows that the radio structure is embedded in hot intergalactic gas from which the masses of the galaxy and of the intracluster gas can be estimated (Sect. 4.4). Within 500 kpc, the mass of the cluster is $(2.0 - 2.8) \times 10^{14} M_\odot$. Within the same radius, the mass of intergalactic gas is $1.1 \times 10^{13} M_\odot$. In addition, features in the map may be identified with the regions of shocked intergalactic

gas heated by the passage of the shock wave which runs ahead of the cocoon. Wilson and his colleagues find that the ‘limb-brightened’ regions which outline the cocoon have slightly higher temperatures than the diffuse gas, while the features which are observed perpendicular to the axis of the source are slightly cooler. We note that, for the diffuse gas, the cooling time of the gas within 200 kpc of the core of the cluster is less than a Hubble time.

A remarkable feature of Cygnus A is that the rotation measure of its polarised radio emission is anomalously large, the values varying systematically along the radio structure from -4000 rad m^{-2} to $+3000 \text{ rad m}^{-2}$ (Dreher *et al.*, 1987). In addition, Dreher and his colleagues find no evidence for internal depolarisation within the radio lobes. Adopting the equipartition magnetic flux density within the lobes, they find an upper limit of $2 \times 10^2 \text{ m}^{-3}$ for the thermal electron density, significantly less than the electron density in the surrounding intergalactic medium and implying that these regions are essentially free of any thermal gas. The large rotation measures are attributed to the presence of a large scale magnetic field in the intergalactic gas in the central regions of the cluster. Adopting the radial number density distribution found from the X-ray observations, they derive magnetic flux densities of between 0.2 and 1 nT for the intracluster magnetic field. Similar large rotation measures have been discovered in other extended radio sources in dense, X-ray emitting clusters.

Although Cygnus A is anomalously nearby for its radio luminosity and so can be studied in far greater detail than comparable sources, the properties described above are typical of the FR2 radio sources. The energy requirements are greatest for the largest and most diffuse sources and can be $\geq 10^{54} \text{ J}$, corresponding to the conversion of at least $5 \times 10^5 M_\odot$ of matter into high energy particles and magnetic field energy over lifetimes of the order of 10^7 – 10^8 years.

21.2.3 The gas dynamics of FR2 radio sources

The schematic model shown in Fig. 21.7a has been the subject of detailed gas-dynamical studies. In finding realistic solutions for the dynamics of such sources, a crucial role is played by the cocoon in stabilising the jet. The jet transfers energy at a relativistic velocity to the hot spots in order to maintain their luminosities, most of the energy being in the form of the kinetic energy of the material of the jet. The nature of the material of the jet is not certain – it could consist of protons and electrons or possibly electron–positron plasma. For the purposes of modelling, it is only necessary to assume that the jet consists of relativistic material. At the point at which the jet encounters the intergalactic medium, a stand-off shock is formed resulting in the bow shock–contact discontinuity structure illustrated in Fig. 21.7a. Along the axis of the source, the pressure of the hot spot is balanced by the ram pressure of the intergalactic medium. If the rate of advance of the hot spot is v , ram pressure balance requires $\rho v^2 = p = \frac{1}{3}\varepsilon$, where ε is the energy density in the hot spot. Adopting $n_e = 10^4 \text{ m}^{-3}$ for the ambient density of the intergalactic gas and the minimum energy density $\varepsilon = 10^{-9} \text{ J m}^{-3}$ in the hot spots, we find $v \approx 0.02c$, consistent with the figures derived from the synchrotron aging argument. This value is also consistent with the upper limits for the speeds of advance of the hot spots which can be derived from the symmetry

of their locations with respect to the host galaxy. Our analysis of a complete sample of 3CR FR2 radio sources showed that the speeds of advance had to be less than $0.2c$ (Longair and Riley, 1979).

The ‘waste energy’ problem was recognised in early studies of the physics of FR2 radio sources in the sense that not all the energy which is needed to supply the hot spots is lost by radiation – rather, the energy must escape from the hot spots and so form a ‘back-flow’ into the volume behind the hot spot (Longair *et al.*, 1973; Scheuer, 1974). In the process, the formation of a cocoon, identified with the extended radio lobes seen in Cygnus A and other FR2 sources, is inevitable. For FR2 sources such as Cygnus A, Begelman (1996) conjectured on the basis of numerical simulations that self-similar evolution of the structure of the radio sources would take place if the ratio of the hot spot to cocoon pressure was a constant. He found self-similar solutions in which the density of the ambient intergalactic medium decreased as a power law of distance from the host galaxy.

This concept was taken further by Kaiser and Alexander (1997) who showed that fully self-consistent models for the evolution of FR2 sources are obtained if the jet itself is confined by the pressure of the cocoon. The condition advocated by Begelman then comes about naturally since the jet power, the luminosity of the hot spots and the pressure of the cocoon are all proportional to each other. Kaiser and Alexander found self-similar solutions, again for the case in which the number density in the intracluster gas decreases as a power law of distance from the active galactic nucleus. The cocoon is then able to collimate the jet since the cocoon pressure is balanced by the sideways ram pressure of the jet. As a consequence, an oblique, or re-collimation, shock is driven into the jet, as indicated schematically by the dotted lines within the jet in Fig. 21.7a.

The stability of the jets has been the subject of numerous studies. The question is whether or not instabilities such as the Kelvin–Helmholtz instability at the interface between the jet and the cocoon disrupt the flow. Initial indications were that the relativistic version of the Kelvin–Helmholtz instability for radio jets is unstable and has the potential to destroy the jet (Turland and Scheuer, 1976), but may be stabilised by the presence of shear layers about the jets (Hughes, 1991). Numerical simulations have shown that such structures can be maintained and result in convincing representations of the structures of FR2 radio sources (Jones, 2008; Krause, 2005; Huerte – Espinosa *et al.*, 2010).

21.2.4 Young FR2 radio sources

Young counterparts of the FR2 sources have been identified with various classes of compact sources. Examples of these are the gigahertz-peaked spectrum sources (GPS) and compact steep spectrum sources (CSS) (Snellen *et al.*, 2004). Members of this class include the compact symmetric objects (CSO) discovered in VLBI observations by Wilkinson and his colleagues. These have all the appearance of FR2 double radio sources as seen in the example shown in Fig. 21.10, but on a scale about 1000 times smaller than the classical FR2 sources (Wilkinson *et al.*, 1994). Their radio luminosities are much greater than the FR1/FR2 transition luminosity. The ages of the compact symmetric objects have been determined from the proper motions of the hot spots, typical ages ranging from hundreds to

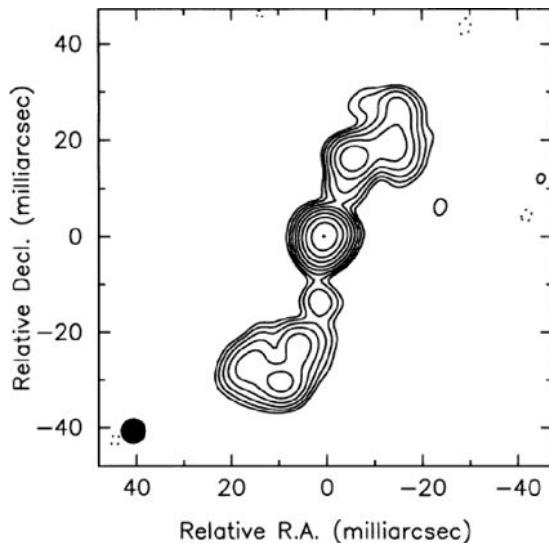


Fig. 21.10 The compact symmetric object (CSO) 2352+495 observed by the VLBA, the separation between the outer radio components being about 220 pc (Wilkinson *et al.*, 1994).

thousands of years, the mean value for the sample studied by Gugliucci and his colleagues being about 500 years (Gugliucci *et al.*, 2005).

The compact symmetric sources are located within the central regions of their host galaxies and, although the number density of electrons is greater than in the intracluster gas, there are important differences in the astrophysics of their evolution as compared with the similarity solutions discussed in Sect. 21.2.3. Alexander (2006) pointed out that the FR2 similarity solution depends upon neglecting the inertial mass density in the jet and cocoon relative to that in the surrounding intergalactic gas. At an early enough stage in the evolution of the jet, the inertial mass density in the jet exceeds the mass density in the central regions of the galaxy and then a different solution is needed. Because the inertial mass of the material of the jet exceeds that of the surrounding medium, the jet can be considered to expand *ballistically*, meaning that the external medium can be ignored and the jet expands with constant opening angle θ .

If the density distribution in the ambient medium is described by $\varrho = \varrho_0(r/a_0)^{-\beta}$, the inertial mass flow rate by \dot{M} and the jet kinetic power by Q , a characteristic length-scale L_1 can be defined by

$$L_1 = \left(\frac{\varrho_0^2 a_0^{2\beta} Q}{\dot{M}^3} \right)^{1/2(\beta-2)} = 25 \left(\frac{Q}{10^{39} \text{ W}} \right)^{1/2} \left(\frac{n_{\text{ext}}}{3 \times 10^5 \text{ m}^{-3}} \right)^{-1/2} \left(\frac{v_{\text{jet}}}{c} \right)^{-3/2} \text{ pc}. \quad (21.2)$$

Alexander shows that this scale corresponds roughly to the condition that the internal mass density is greater than the external density. For scales less than this value, a jet shock forms and almost all the energy of the jet goes into the work done by the bow shock and the internal energy stored in the hot spot. For scales of the order L_1 , the cocoon begins to

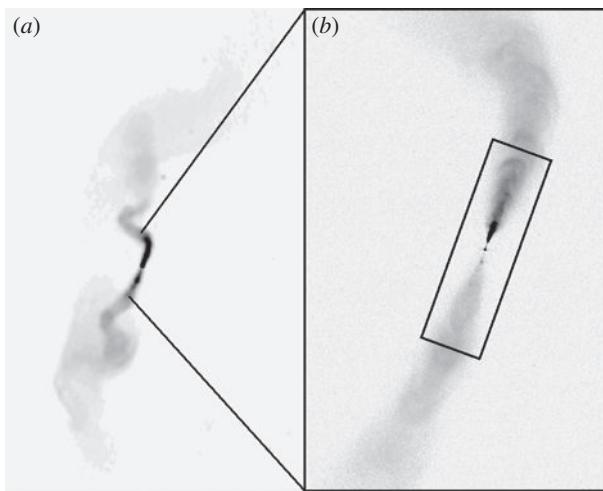


Fig. 21.11 (a) The radio structure of the FR1 radio galaxy 3C 31 at 1.4 GHz at an angular resolution of 5.5 arcsec (1.9 kpc) as observed by the VLA (see also Fig. 21.2a). (b) The central region of 3C 31 observed by the VLA at 8.4 GHz with angular resolution 0.25 arcsec (85 pc). The radio source is associated with the giant elliptical galaxy NGC 383 which is the brightest member of a group of galaxies (Laing and Bridle, 2002).

form and the structure shown in Fig. 21.7b is more appropriate. Alexander found solutions corresponding to the transition from the structure shown in Fig. 21.7b to Fig. 21.7a and the compact symmetric sources are identified with this transition phase in the early evolution of FR2 sources.

21.3 The FR1 radio sources

The FR2 sources are much less luminous radio sources than the FR1s and they do not contain hot spots at the leading edges of the radio lobes. Many of these sources are members of rich clusters of galaxies and the diffuse FR1 radio lobes are assumed to be in pressure balance with the hot intergalactic gas. As first pointed out by Gull and Northover (1973), these large lobes of relativistic gas form a low density bubble in the gravitational potential of the cluster and so rise up the gravitational potential gradient by buoyancy. Direct evidence for this process occurring in low luminosity FR1 radio sources is provided by the X-ray-radio images of clusters such as the Perseus cluster where the cavity inflated by the buoyant relativistic gas is observed (Fabian *et al.*, 2000) (Fig. 4.9).

An important feature of the jets observed in the FR1 sources is that, although they are one-sided as observed on small scales, the jets are double when observed on much larger scales. The example of 3C 31 is illustrated in Fig. 21.11 showing the asymmetric structure in the central regions of the galaxy and the diverging jet on kpc scales. VLBI observations of the core of 3C 31 show the characteristic one-sided structure observed in

many compact sources, the absence of a counter-jet being attributed to the beaming of the emission due to its bulk motion at a speed close to that of light (Venturi *et al.*, 1996).

The interpretation of these observations is that there are indeed relativistic jets present in the FR1 radio sources, but they have much lower luminosities than the FR2 radio galaxies. The combination of weaker jet power and perhaps different environmental conditions in the central regions of the galaxy lead to the rapid deceleration of the jet to non-relativistic velocities. The source 3C 31 is an ideal case for study because it is a bright source and so can be observed with very high sensitivity, both in total intensity and in polarisation. The detailed study by Laing and Bridle (2002) illustrates how much information can be obtained about the dynamics and magnetic field structure from such observations. In their simulations, the on-axis velocity of the jet decreases from values of order 0.8–0.9c within the innermost 1 kpc to values less than 0.2c at a distance of 15 kpc from the nucleus. The deceleration and flaring of the jet is associated with the entrainment of matter by the jet. With these assumptions, they were able to model in considerable detail the observed asymmetry of the radio emission seen in Fig. 21.11 while assuming that the jets are symmetrically ejected from the nucleus. As Laing and Bridle point out, the success of their simulations raises a number of important issues about the physics and structure of relativistic jets.

21.4 The microquasars

The problems of understanding the physics of the double radio structures observed in extragalactic radio sources was for a long time the province of extragalactic radio astronomers. In 1994, however, evidence was found for relativistic jets in Galactic binary X-ray sources in which there is evidence for the presence of a black hole as the invisible companion. Mirabel and Rodrigues observed the intense hard X-ray source GRS 1915+105 during a radio outburst and discovered that its radio structure was double and that the components were separating from the radio core at 1.25 times the speed of light (Mirabel and Rodrigues, 1994) (Fig. 21.12a). The X-ray properties of the source strongly suggested that the energy source in the binary system is a black hole. Since their discovery in 1994, several other examples have been discovered. Mirabel and Rodrigues pointed out the remarkable similarities between the radio properties of these source, which they term *microquasars*, and the radio galaxies and radio quasars (Mirabel and Rodrigues, 1998) (Fig. 21.12b). The big advantage of studying the microquasars is that the time-scales of the phenomena associated with their variability and energy production scale as the mass of the black hole, and so phenomena which would take thousands or millions of years for a $10^9 M_\odot$ would be expected to take place on the time-scale of minutes or days in a $10 M_\odot$ Galactic black hole. Such rapid changes in the radio and X-ray emission of these sources have been observed.

Other small scale Galactic counterparts of these sources include the bright X-ray source Sco X-1 and the system SS 433. Both are associated with compact objects within our Galaxy and both exhibit double radio structure similar in many ways to the extended

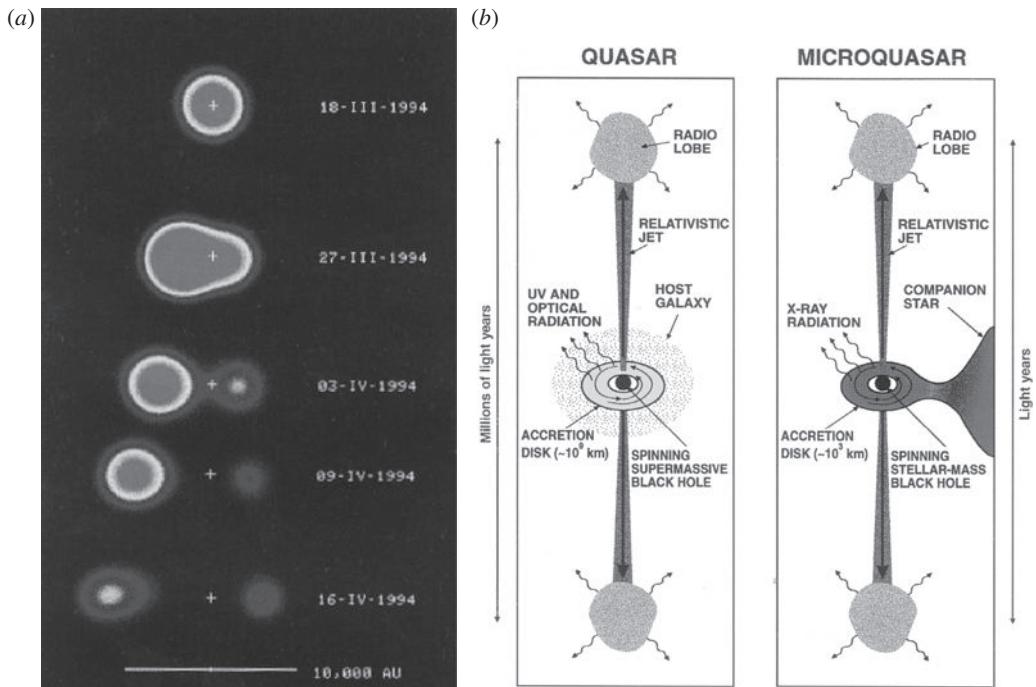
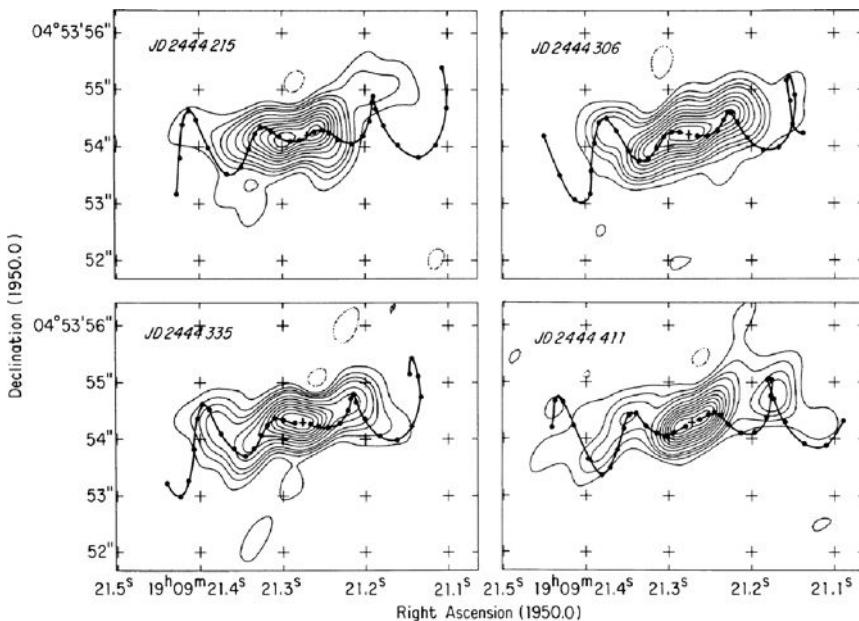


Fig. 21.12 (a) The evolution of the radio structure of the binary X-ray source GRS 1915+105 over a three week period in 1994 (Mirabel and Rodrigues, 1994). (b) A comparison between the properties of the radio jets in quasars and microquasars (Mirabel and Rodrigues, 1998).

extragalactic radio sources. SS 433 is associated with the supernova remnant W50 which exploded about 10 000 years ago, and is an eclipsing X-ray binary system. The primary star is probably a black hole, but it could be a neutron star. The spectrum of the secondary companion star suggests that it is a late A-type star. The case of SS 433 is of particular interest because the radio jets emit optical line radiation and the velocities of ejection of the beams can be found directly from the Doppler shifts of the lines. These velocities are found to be about $0.26c$. In addition, it has been shown convincingly from the velocity curves of the system and directly from radio maps made at different epochs that the beam precesses about the mean axis of the jets. These phenomena are beautifully illustrated by the movie made from 42 days of observation by the Very Long Baseline Array (VLBA) in the summer of 2004 (Mioduszewski *et al.*, 2004).¹ Many of the phenomena seen in SS 433 have their counterparts in extragalactic radio sources. In particular, the precession of the radio jet is reminiscent of the problem of accounting for multiple hot spots in the lobes of FR2 radio sources. Scheuer (1982) proposed that the multiple hot spots could be attributed to what he called the ‘dentist’s drill’ model in which the relativistic jet changes direction, in the process illuminating different regions of the interface between the cocoon

¹ see <http://www.nrao.edu/pr/2004/ss433/ss433.movie.gif>

**Fig. 21.13**

Radio images of the Galactic radio sources SS 433. A collimated outflow is observed, the direction of the jet precessing about the major axis of ejection as can be seen from the four radio maps which span a period of almost 200 days. The contour maps show the structure of the radio source at various phases in the cycle. The solid lines show a model of a precessing jet to account for the observed variations in structure. This phenomenon may also occur in extended extragalactic sources and account for the observation of multiple hot spots in the extended radio lobes (Hjellming and Johnston, 1981).

and the intergalactic medium. The precession of the jets seen in SS 433 may be the Galactic counterpart of such a process.

21.5 Jet physics

We have not answered the question of the physical origin of the relativistic jets in active galactic nuclei. Once a jet is created, the processes discussed in Sects 21.2.3 and 21.2.4 can lead to self-collimated structures, but the initial outflows must originate close to the active galactic nucleus itself. The model of a thick accretion disc with funnels along the axis of rotation illustrated in Fig. 14.10 may provide a collimation mechanism, but it has yet to be established whether or not these are stable structures.

One possibility is that the jet particles are associated with electromagnetic processes occurring in the magnetosphere of the rotating black hole, the rotation axis providing a natural axis along which the beams of particles are ejected. This model has similarities with the pulsar magnetospheres discussed in Sect. 13.9. The analysis of the electromagnetic structures of rotating black holes by Blandford and Znajek (1977) has been influential in

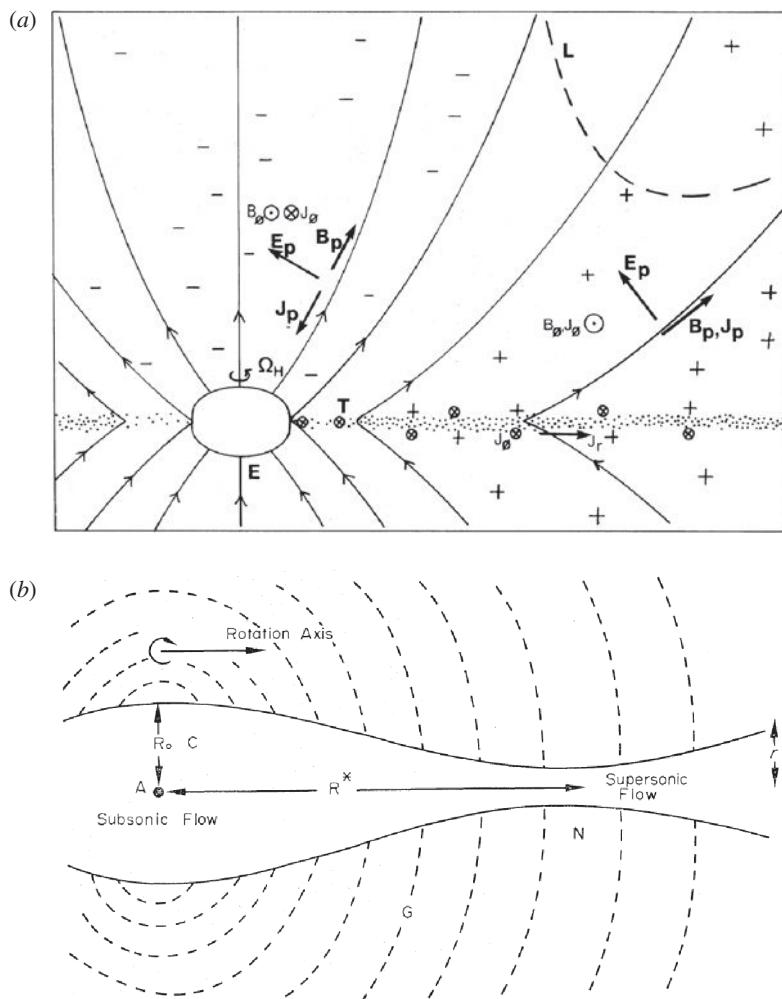


Fig. 21.14 (a) Illustrating the parabolic solution for the magnetic field structure about a rotating black hole (Blandford and Znajek, 1977). (b) The twin-exhaust' model of Blandford and Rees (1974), illustrating how gas dynamic processes could give rise to symmetric double outflows from the dense central regions of active galactic nuclei.

demonstrating not only how energy can be extracted electromagnetically from the rotational energy of the black hole, but also suggesting a means of collimating the beam of particles by magnetic fields lines which are tied to the rotation of the black hole. The type of parabolic solution which they proposed in their original paper is shown in Fig. 21.14.

The axis of the accretion disc provides a natural axis for the collimation of jets originating close to the black hole. Magnetic instabilities at the surface of the accretion disc could give rise to ‘flares’ which would result in a wind collimated perpendicular to the disc. Another possibility is that the jet could be created by gas dynamical processes occurring in the dense central regions surrounding the central black hole. The continuous generation of energy leads to the formation of a region of very high energy density of high energy particles. The

hot, possibly relativistic, gas will burst through the surrounding cloud in the direction of the poles of the cloud and form a collimated jet. The ‘twin-exhaust’ model of Blandford and Rees (1974) is an example of how the early formation of high energy jets could come about. It is not so clear that this type of structure can give rise to the extreme relativistic jets observed in superluminal sources (Sect. 22.1). This type of model provides a possible link between the properties of the radio jets and the regions of high density gas inferred to be present in the X-ray emitting and broad-line regions in the active galactic nucleus.

Compact extragalactic sources and superluminal motions

The evidence of Chap. 21 shows that the huge fluxes of relativistic material needed to power extended extragalactic radio sources originate close to the active galactic nuclei of the host galaxies. Direct evidence for extreme events in active galactic nuclei is provided by the superluminal motions observed in compact radio sources, by the properties of variable extragalactic γ -ray sources and by the γ -ray bursts. The extreme properties of these sources require them to be moving at highly relativistic velocities.

22.1 Compact radio sources

Direct evidence for the presence of ultra-relativistic electrons in the nuclei of active galaxies is provided by very long baseline interferometric (VLBI) studies of radio quasars and BL-Lac objects at centimetre wavelengths. Combining the angular sizes of these ultra-compact radio sources with their flux densities S_ν , the *brightness temperature* $T_b = (\lambda^2/2k)(S_\nu/\Omega)$ of the source region can be determined, where Ω is the solid angle subtended by the radio source. Observations of large samples of strong compact radio sources with structures on the scale of 1 milliarcsecond have shown that the maximum brightness temperatures are of the order of 10^{11} – 10^{12} K, none of them exceeding the limit of 10^{12} K at which catastrophic synchrotron self-Compton radiation would take place, as described in Sect. 9.6 (Kellermann *et al.*, 1998).

Brightness temperatures of the order of 10^{11} K can be compared with the thermodynamic temperature of electrons which have kinetic energies equal to their rest-mass energies, $kT \approx m_e c^2$, and so $T \approx 5 \times 10^9$ K. Assuming the emission is the incoherent radiation of high energy electrons, this argument demonstrates that the emitting electrons must have ultra-relativistic energies, $E \gg m_e c^2$, since particles cannot radiate at brightness temperatures greater than their thermodynamic temperatures. Generally, the angular sizes currently measurable for distant quasars by VLBI techniques correspond to about 1 pc compared with the Schwarzschild radius of a $10^8 M_\odot$ black hole of only 10^{-5} pc. Thus, the processes of jet formation probably occur on scales smaller than those currently accessible by centimetre VLBI.

A characteristic signature of compact radio sources is the observation of a synchrotron self-absorption cut-off at centimetre wavelengths, as indicated schematically in Fig. 8.12. The peak of the spectrum occurs roughly at the frequency at which the brightness temperature of the radio source is equal to the effective ‘thermal’ temperature of the electrons, as explained in Sect. 8.7.1. A useful reference calculation is how the spectrum of

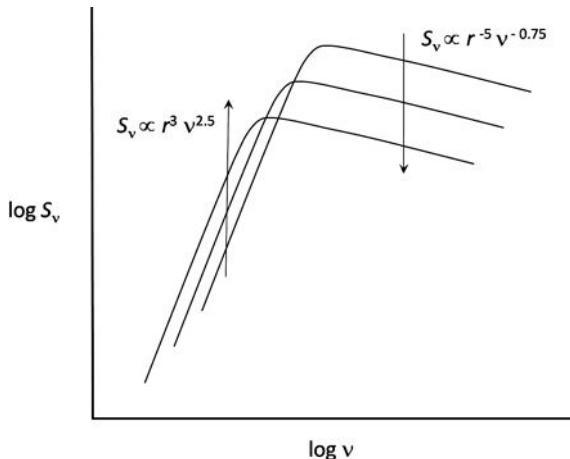


Fig. 22.1

Illustrating the evolution of the spectrum of a synchrotron self-absorbed spherical radio source in a uniform expansion in which the energies of the particles decrease adiabatically and the magnetic flux density decreases as r^{-2} . In this example, $\alpha = 0.75$, equivalent to $p = 2.5$.

such a synchrotron self-absorbed source evolves as it expands. In the idealised case of a uniformly expanding spherical source in which the relativistic electrons lose energy adiabatically as $E \propto r^{-1}$ and the magnetic flux density decreases as r^{-2} because of magnetic flux freezing, it was shown in Sect. 16.8 that the flux density of the source decreases as $S_v \propto v^{-(p-1)/2} r^{-2p} \propto v^{-\alpha} r^{-(4\alpha+2)}$, where the electron energy spectrum has the form $N(E) dE \propto E^{-p}$ and $\alpha = (p - 1)/2$. Supernova remnants such as Cassopaeia A decrease in flux density at roughly this rate.

A similar calculation can be carried out in the optically thick regime using the same approximate methods described in Sect. 8.7.1. The flux density of the source at frequency v is $S_v = (2kT_b/\lambda^2)\Omega$ where $\Omega \propto r^2$ is the solid angle subtended by the source. The Lorentz factor γ of the electrons radiation at this frequency is given approximately by $v \approx \gamma^2 v_g \propto \gamma^2 B$, where $v_g = eB/2\pi m_e$ is the non-relativistic gyrofrequency of the electron. As the source expands, $B \propto r^{-2}$ and so, at a fixed frequency, the electrons radiating at the frequency v must have energies proportional to r . Therefore, since $kT_b \propto \gamma m_e c^2$, it follows that the flux density in the optically thick regime increases as $S_v \propto v^{2.5} r^3$. This form of spectral evolution is illustrated schematically in Fig. 22.1 for a spectral index $\alpha = 0.75$ in the optically thin region of the spectrum. The maximum of the spectrum moves to lower frequencies as $v_{\max} \propto r^{-(4p+6)/(p+4)}$.

Different assumptions can be made about the variation of the magnetic flux density and particle energy as a function of radius. For example, the particles may be re-accelerated and the magnetic field may change as a different power of r in the expansion. These cases can be developed as straightforward generalisations of the above calculation, a number of useful examples for the expansion of a young radio supernova remnant being given by Chevalier (1998). Spectra of roughly the shape shown in Fig. 22.1 are observed in compact radio sources at high radio frequencies and are interpreted as synchrotron self-absorbed radio sources. The gigahertz peaked spectrum (GPS) objects are examples of such sources.

Often the observed source spectra are flat with $\alpha \sim 0$ in the centimetre waveband and these are interpreted as the superposition of synchrotron self-absorbed components within the source regions.

22.2 Superluminal motions

VLBI observations show that the extended radio jets observed in FR1 and FR2 radio sources have their origin in the associated active galactic nuclei, as discussed in Chap. 21. The large scale jets can be traced back more or less continuously to aligned structures on a scale of 10 pc or less, as has been observed in sources such as Cygnus A, Virgo A (M87) and 3C 31. One of the great surprises of the 1970s and 1980s was that many of the bright compact radio components observed by the VLBI technique were observed to move out from their nuclei at speeds in excess of the speed of light. One of the first examples of such *superluminal motions* was observed in the quasar 3C 273 in which a source component moved away from the nucleus at a speed of roughly eight times the speed of light (Fig. 22.2a, top panel) (Pearson *et al.*, 1981, 1982). With the construction of the Very Long Baseline Array and the European VLBI network, both dedicated to VLBI studies, systematic studies of compact radio sources were carried out. More recent examples of superluminal motions in the radio quasar 3C 279 and the nearby Seyfert 1 galaxy 3C 120 are shown in Fig. 22.2a (lower panel) and 22.2b, respectively. With the vastly improved imaging capabilities of these arrays, it has become clear that the typical superluminal jet consists of a series of components which travel outwards from the nucleus along more or less the same axis, as demonstrated by the example of 3C 120.

Characterising the nature of the population of superluminal sources has become possible with the imaging of large statistical samples of compact flat spectrum radio sources selected at high frequencies. This is a highly non-trivial problem because the sources are intrinsically variable. Cohen, Kellermann and their colleagues have carried out a major survey of these sources using the VLBA, the results of which are contained in a series of papers over a 10-year period (for detailed references, see the papers cited by Cohen (2007) (paper IV) and Homan (2009) (Paper VII) of the series). The sample of sources is based upon a complete survey of radio sources with flat radio spectra, $\alpha \leq 0.5$, at 5 GHz. This was then extrapolated to 15 GHz, at which frequency the major imaging survey was carried out and the observations then continued over the succeeding years to determine the kinematics of all the sources in the sample. The flux density limit of the sample was about 1.5 Jy at 15 GHz (2 cm wavelength). The first paper of the series presents images of the first 132 sources and gives a good impression of the typical structures observed (Kellermann *et al.*, 1998).

Cohen and his colleagues (2007) provide a summary of many of the key points of these observations, the sample having expanded to 225 compact radio sources observed at 15 GHz through the period 1994–2002. For 119 of these sources, transverse velocities were determined from multi-epoch observations, enabling velocities to be determined which satisfied their ‘excellent’ or ‘good’ criteria. Of the 199 sources, 10 are galaxies, 17 are

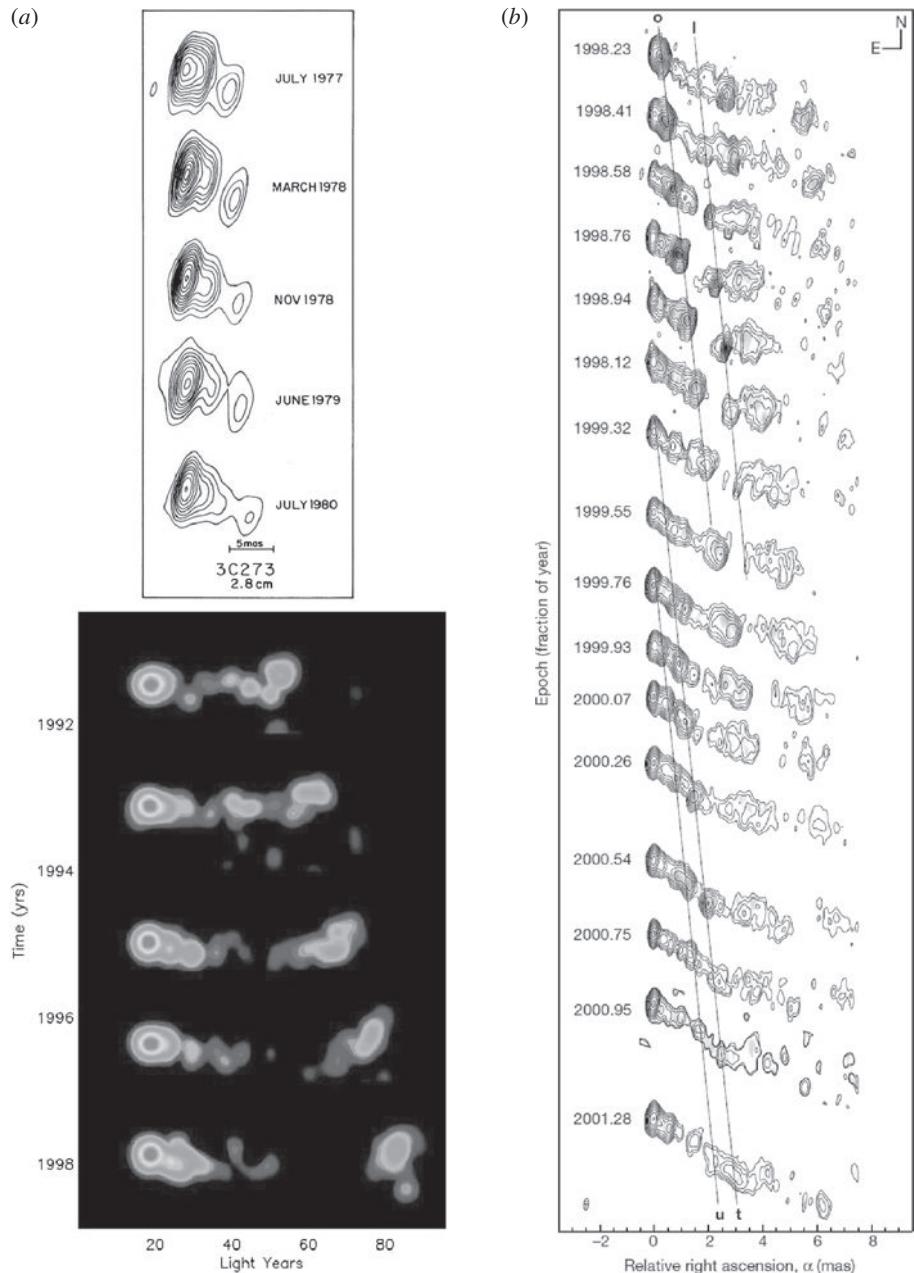


Fig. 22.2 (a) The superluminal radio source 3C 273 as observed by VLBI over the period July 1977 to July 1980 (top panel) (Pearson *et al.*, 1981, 1982). At a redshift $z = 0.158$, the hot spot in the radio jet moved a distance of about 25 light years over the three year period of the observation. In the lower panel, the superluminal expansion of 3C 273 is shown over a six year period. (Courtesy of NRAO and AUI.) (b) The superluminal radio source 3C 120 showing the stream of radio components moving out from the nucleus along the same axis at superluminal speeds of 4.5–5 times the speed of light (Marscher *et al.*, 2002).

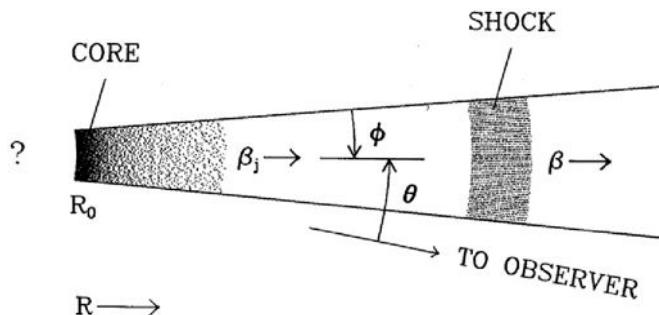


Fig. 22.3

A typical model for the observation of superluminal motion in compact radio sources (Marscher, 1993). The observer is located at an angle θ relative to the axis of ejection of the relativistic jet. The base of the jet is identified with the compact, flat spectrum source component and this part of the jet also moves out relativistically from the nucleus.

BL-Lac objects and 92 are quasars. Some of the key properties highlighted by Cohen and his colleagues are as follows:

- The transverse velocities range from mildly relativistic motion to transverse speeds up to 30 times the speed of light and monochromatic luminosities up to $3 \times 10^{28} \text{ W Hz}^{-1}$ at 15 GHz. Because of relativistic beaming, however, this is certainly an overestimate of the luminosity in the frame of the moving component which is likely to be about two orders of magnitude less, as described below.
- The sources possess a one-sided ‘core-jet’ structure in which the core is identified with the active galactic nucleus. The core component has a flatter spectrum than the component which moves superluminally away from the nucleus. The flatness of the spectrum of the core is naturally explained by high frequency synchrotron self-absorption.
- The jets consist of a series of components which move along the same trajectory from the nucleus. In the example of 3C 120 shown in Fig. 22.2b, the trajectories of four source components are shown by lines, their observed velocities lying in the range 4.5–5.0 c . Each of these ejections originated in an absorption event in the X-ray light curve as observed by the RXTE (Marscher *et al.*, 2002).
- Although many of the trajectories are straight, there are bends and kinks in some of the jets. As we will show, many of the jets must be travelling from the nucleus at angles close to the line of sight and so there will undoubtedly be considerable foreshortening of the actual trajectory. Right-angled kinks might in fact consist of quite small deviations from the axis of the ejection of the source components. For example, the precession observed in SS 433 seen end-on would result in dramatic kinks in the jet.
- The structures of the relativistically moving components often extend in the direction of the larger scale radio jets.

These properties can be naturally explained in terms of what has become the standard model for superluminal sources, many of the ideas being foreshadowed in the prescient papers of Rees of 1967 (Rees, 1967). A typical representation of the standard model is shown in Fig. 22.3 (Marscher, 1993). In this idealised picture, the jet material propagates along the axis of the beam at speed $\beta_j = v_j/c$. The opening angle of the beam is ϕ and

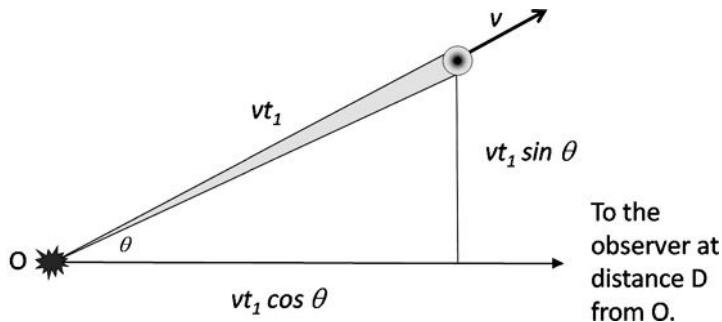


Fig. 22.4

Illustrating the motion of a relativistically moving source component as viewed from above. The ejection from the nucleus takes place at time t_0 from O. The diagram shows the configuration at time t_1 .

the angle between the axis of the beam and the line of sight to the observer is θ . The ejection events may give rise to shocks within the beam which move at speed $\beta = v/c$. These shocks may be identified with the radio source components which move down the beam axis. There will undoubtedly be a shock wave at the interface between the jet and the ambient interstellar medium. Just as in the case of supernova remnants, there is a difference between the speed of the jet material v_{jet} and the speed of the shock front v , which is often referred to as the *pattern speed*. The term pattern speed is used more widely than just in this context. The component may not be moving with the material which is the source of the emission. For example, a search-light beam illuminating high clouds can result in the appearance of a spot on the cloud moving faster than the speed of light if the beam is rotated fast enough. This apparent motion is the source's pattern speed. In this case, however, there is no movement of the clouds themselves.

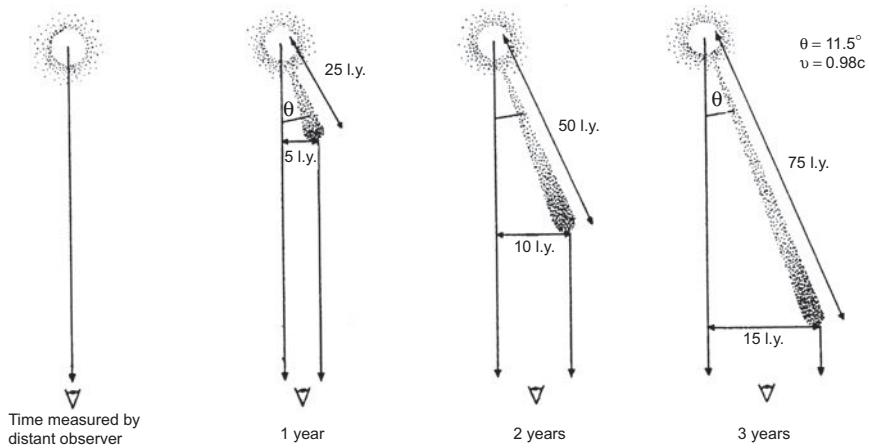
Before studying the observations in more detail, let us first work out some of the basic properties of relativistic beaming which will be used throughout the rest of this chapter.

22.3 Relativistic beaming

Let us begin with the simplest model for superluminal sources, what is often referred to as the *relativistic ballistic model*. The first part of the calculation is the determination of the *kinematics* of a relativistically moving source component, specifically the observed transverse speed of a component ejected at some angle θ to the line of sight at a relativistic velocity v (Fig. 22.4).

The observer is located at a distance D from the source. The source component is ejected from the origin O at some time t_0 and the signal from that event sets off towards the observer, where it arrives at time $t = D/c$ later. After time t_1 , the component is located at a distance vt_1 from the origin and so is observed at a projected distance $vt_1 \sin \theta$ according to the distant observer. The light signal bearing this information arrives at the observer at time

$$t_2 = t_1 + \frac{D - vt_1 \cos \theta}{c}, \quad (22.1)$$

**Fig. 22.5**

The origin of superluminal motion for a source component ejected at $0.98c$ from the nucleus of an active galaxy at an angle 11.5° to the line of sight as observed from above. Because the source of the radiation moves at a speed very close to that of light, it almost catches up with the radiation it emits.

since the signals have to travel a shorter distance $D - vt_1 \cos \theta$ to reach the observer. Therefore, according to the distant observer, the transverse speed of the component is

$$v_{\perp} = \frac{vt_1 \sin \theta}{t_2 - t} = \frac{vt_1 \sin \theta}{t_1 - \frac{vt_1 \cos \theta}{c}} = \frac{v \sin \theta}{1 - \frac{v \cos \theta}{c}}. \quad (22.2)$$

The maximum observed transverse speed is found by differentiating v_{\perp} with respect to θ and occurs at an angle $\cos \theta = v/c$; the maximum transverse velocity is $v_{\perp} = \gamma v$, where $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor. Thus, provided the source component moves at a speed close enough to the speed of light, apparent motions on the sky $v_{\perp} \gg c$ can be observed without violating causality and the postulates of special relativity. For example, if the source component were ejected at a speed $0.98c$, transverse velocities up to $\gamma c = 5c$ are perfectly feasible, the case illustrated in Fig. 22.5. The diagram shows that superluminal velocities are observed because the source component almost catches up with the radiation it emits.

Next we need to understand the effects of what is loosely referred to as ‘relativistic beaming’ upon the observed intensities of the source components. Let us consider first a classical undergraduate problem in relativity:

A rocket travels towards the Sun at speed $v = 0.8c$. Work out the luminosity, colour, angular size and brightness of the Sun as observed from the spaceship when it crosses the orbit of the Earth. It may be assumed that the Sun radiates like a uniform disc with a black-body spectrum at temperature T_0 .

This problem includes many of the effects found in relativistic beaming problems. Let us work out the separate effects involved in evaluating the intensity of radiation observed in the moving frame of reference. It is simplest to use four-vectors to work out the frequency

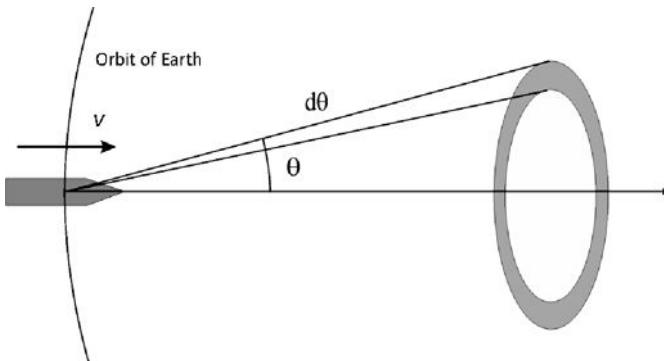


Fig. 22.6 Illustrating the geometry of the propagation of light from the Sun to the observer in a spaceship moving radially towards the Sun at speed v .

shifts and aberrations. Consider the radiation from an annulus of angular width $\Delta\theta$ at angle θ with respect to the centre of the Sun (Fig. 22.6).

- (i) *The frequency shift of the radiation* The frequency four-vector in the frame of the Solar System S is given in Rindler's notation as

$$\mathbf{K} = \left[\frac{\omega_0}{c}, \mathbf{k} \right] = \left[\frac{\omega_0}{c}, -k_0 \cos \theta, -k_0 \sin \theta, 0 \right], \quad (22.3)$$

where the light rays are assumed to propagate towards the observer at the orbit of the Earth, as illustrated in Fig. 22.6.¹ The frequency four-vector in the frame of reference of the spaceship S' is

$$\mathbf{K}' = \left[\frac{\omega'_0}{c}, \mathbf{k}' \right] = \left[\frac{\omega'_0}{c}, -k'_0 \cos \theta', -k'_0 \sin \theta', 0 \right]. \quad (22.4)$$

We use the time transform to relate the ‘time’ components of the four-vectors:

$$ct' = \gamma \left(ct - \frac{vx}{c} \right), \quad (22.5)$$

and so

$$\omega' = \gamma (\omega_0 + vk_0 \cos \theta). \quad (22.6)$$

Since $k_0 = \omega_0/c$,

$$v' = \gamma v_0 \left(1 + \frac{v}{c} \cos \theta \right) = \kappa v_0. \quad (22.7)$$

This is the expression for the ‘blueshift’ of the frequency of the radiation due to the motion of the spacecraft.

- (ii) The *waveband* $\Delta\nu$, in which the radiation is observed, is blueshifted by the same factor

$$\Delta\nu' = \kappa \Delta\nu_0. \quad (22.8)$$

¹ Notes on the relativistic conventions used are given in Appendix A.8.2.

- (iii) *Time intervals* are also different in the stationary and moving frames. The expression for the difference can be found by comparing the periods of the waves as observed in S and S'

$$v' = \frac{1}{T'} \quad ; \quad v_0 = \frac{1}{T_0} , \quad (22.9)$$

and so

$$\frac{T'}{T} = \frac{v_0}{v'} . \quad (22.10)$$

Since the periods T and T' can be considered to be the times measured on clocks, the radiation emitted in the time interval Δt is observed in the time interval $\Delta t'$ by the observer in S' such that $\Delta t' = \Delta t/\kappa$.

- (iv) *Solid angles* Finally, we work out how the solid angle subtended by the annulus shown in Fig. 22.6 changes between the two frames of reference. It is simplest to begin with the cosine transform, which is derived from the transformation of the x' component of the frequency four-vector:

$$\cos \theta' = \frac{\cos \theta + \frac{v}{c}}{1 + \frac{v}{c} \cos \theta} . \quad (22.11)$$

Differentiating with respect to θ on both sides of this relation,

$$\sin \theta' d\theta' = \frac{\sin \theta d\theta}{\gamma^2 \left(1 + \frac{v}{c} \cos \theta\right)^2} = \frac{\sin \theta d\theta}{\kappa^2} . \quad (22.12)$$

This result has been derived for an annular solid angle with respect to the x -axis, but we can readily generalise to any solid angle since $d\phi' = d\phi$ and so

$$\sin \theta' d\theta' d\phi' = \frac{\sin \theta d\theta d\phi}{\kappa^2} \quad d\Omega' = \frac{d\Omega}{\kappa^2} . \quad (22.13)$$

Thus, the solid angle in S' is smaller by a factor κ^2 as compared with that observed in S. This is a key aspect of the derivation of the aberration formulae. Exactly the same form of beaming occurred in the derivation of the formulae for synchrotron radiation.

We can now put these results together to work out how the intensity of radiation from the region of the Sun within solid angle $d\Omega$ changes between the two frames of reference. First of all, the intensity $I(v)$ is defined to be the power arriving at the observer per unit frequency interval per unit solid angle from the direction θ . The observer in the spacecraft observes the radiation arriving in the solid angle $d\Omega'$ about the angle θ' and we need to transform its other properties to those observed in S'. Let us enumerate how the factors change the observed intensity. The energy $h\nu N(v)$ received in S in the time interval Δt , in the frequency interval $\Delta\nu$ and in solid angle $\Delta\Omega$ is observed in S' as an energy $h\nu'N(v')$ in the time interval $\Delta t'$, in the frequency interval $\Delta\nu'$ and in solid angle $\Delta\Omega'$, where

$N(\nu) = N(\nu')$ is the invariant number of photons. Therefore, the intensity observed in S' is

$$I(\nu') = I(\nu) \times \frac{\kappa \times \kappa \times \kappa^2}{\kappa} = I(\nu)\kappa^3. \quad (22.14)$$

Let us now apply this result to the spectrum of black-body radiation, for which

$$I(\nu) = \frac{2h\nu^3}{c^2} (e^{h\nu/kT} - 1)^{-1}. \quad (22.15)$$

Then,

$$I(\nu') = \frac{2h\nu^3\kappa^3}{c^2} (e^{h\nu/kT} - 1)^{-1} = \frac{2h\nu'^3}{c^2} (e^{h\nu'/kT'} - 1)^{-1}, \quad (22.16)$$

where $T' = \kappa T$. In other words, the observer in S' observes a black-body radiation spectrum with temperature $T' = \kappa T$.

Another way of deriving the same result is to write the expression for black-body radiation as

$$I(\nu) = \frac{2h\nu^3}{c^2} n(\nu), \quad (22.17)$$

where $n(\nu)$ is the occupation number, which we demonstrated in Sect. 9.4.2 is a Lorentz invariant. Therefore, since $\nu' = \kappa \nu_0$,

$$I(\nu') = \frac{2h\nu'^3}{c^2} n(\nu') = \frac{2h\kappa^3\nu^3}{c^2} n(\nu), \quad (22.18)$$

as before. Furthermore, $n(\nu)$ need not be of black-body form, but applies to any occupation number and so the brightness temperature T_b also transforms as $T'_b(\nu') = \kappa T_b(\nu)$.

A number of useful results follow from this analysis. For example, (22.16) describes the temperature distribution of the Cosmic Microwave Background Radiation over the sky as observed from the Solar System, which is moving through the frame of reference in which the sky would be perfectly isotropic on the large scale at a velocity of about 600 km s^{-1} . Since $v/c \approx 2 \times 10^{-3}$ and $\gamma \approx 1$, the temperature distribution is rather precisely a dipole distribution, $T = T_0[1 + (v/c)\cos\theta]$, with respect to the direction of motion of the Solar System through the Cosmic Microwave Background Radiation.

In the example of the spacecraft travelling at $v = 0.8c$ towards the Sun, we can illustrate a number of the features of relativistic beaming. For example, if the motion of the spacecraft is significantly relativistic, the dipole pattern is distorted and the zeros of the polar diagram are shifted in the direction of motion. If $v = 0.8c$, $\gamma = 5/3$ and the angle at which there is no change of temperature, corresponding to $\gamma[1 + (v/c)\cos\theta] = 1$, is $\theta = 60^\circ$.

For the case of relativistically moving source components, all we need do is determine the value of κ for the source component moving at velocity v at an angle θ with respect to the line of sight from the observer to the distant quasar as illustrated in Fig. 22.4. In this case, a straightforward calculation shows that the value of κ is

$$\kappa = \frac{1}{\gamma \left(1 - \frac{v \cos\theta}{c}\right)}, \quad (22.19)$$

where the source is moving towards the observer as illustrated in the figure. As in the above example, the observed flux density of the source is therefore

$$S(v_{\text{obs}}) = \frac{L(v_0)}{4\pi D^2} \times \kappa^3, \quad (22.20)$$

where $v_{\text{obs}} = \kappa v_0$. In the case of superluminal sources, the spectra can often be described by a power law $L(v_0) \propto v_0^{-\alpha}$ and so

$$S(v_0) = \frac{L(v_0)}{4\pi D^2} \times \kappa^{3+\alpha}. \quad (22.21)$$

It is interesting to compare this result with the expressions used in the case in which the sources are at cosmological distances. In that case, we have to use the appropriate distance measure D to describe how the radiation spreads out over a sphere in the curved geometry of space (see Longair, 2008). In that case,

$$S(v_0) = \frac{L(v_0)}{4\pi D^2} \times \kappa^{1+\alpha}, \quad (22.22)$$

where $\kappa = (1+z)^{-1}$. Notice the difference of a factor of κ^2 between these formulae. The reason is that, in the cosmological case, the radiation is emitted isotropically over a sphere centred on the source, whereas in the case of the moving source component, the radiation is beamed towards the observer, the change of solid angle being given by κ^{-2} .

If the superluminal sources consisted of identical components ejected from the radio source at the same angle in opposite directions, the relative intensities of the two components would be in the ratio

$$\frac{S_1}{S_2} = \left(\frac{1 + \frac{v}{c} \cos \theta}{1 - \frac{v}{c} \cos \theta} \right)^{3+\alpha}. \quad (22.23)$$

It is therefore expected that there should be large differences in their observed intensities. For example, adopting the largest apparent velocities for a given value of γ , $\cos \theta = v/c$, in the limit $v \approx c$,

$$\frac{S_1}{S_2} = (2\gamma^2)^{3+\alpha}. \quad (22.24)$$

Since values of γ up to 30 are observed and $\alpha \sim 0-1$, it follows that the approaching component would be very much more luminous than the receding component. It is, therefore, not at all unexpected that the sources should be one-sided.

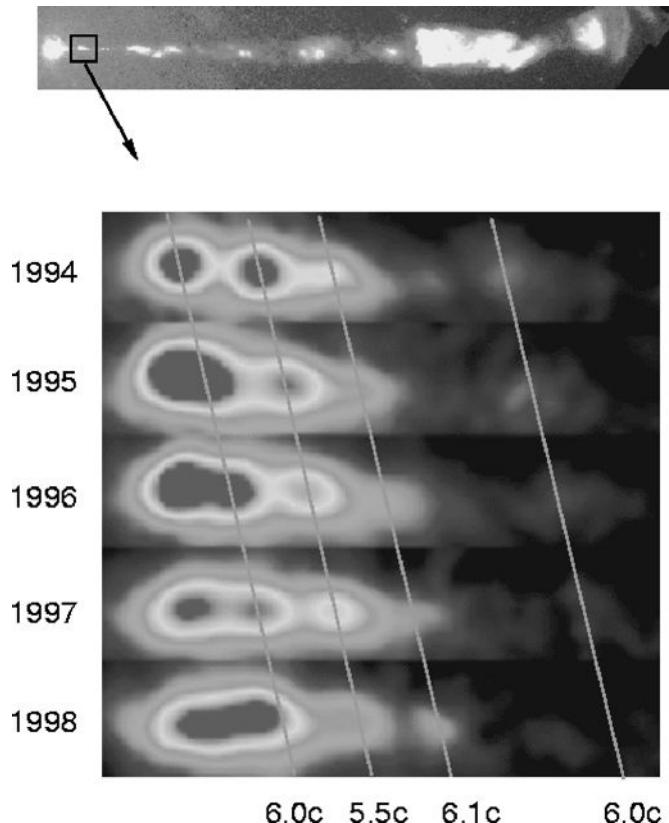
The above analysis has treated the source components as identical point sources ejected in opposite directions from the nucleus. In fact, the source components are better represented by relativistic jets, as illustrated in Fig. 22.3. This slightly modifies the above analysis. Suppose the superluminal source consists of two identical uniformly emitting jets moving in opposite directions from the nucleus. Let us consider the luminosity of a section of the jet of length $\Delta x'$ on either side of the nucleus. Then, the luminosity of these sections of the jet is $L' = \Delta x' \mu$ where μ is the luminosity per unit path length in the frame of the jet. The

increment of length $\Delta x'$ is observed in the observer's frame as an increment $\Delta x = \kappa \Delta x'$, as can be appreciated from the example shown in Fig. 22.5. Thus, the enhancement of the luminosity per unit path length is a factor of κ less than that given by (22.21), $S \propto \kappa^{2+\delta}$ rather than $S \propto \kappa^{3+\delta}$. In the same limit as in (22.24), $S_1/S_2 = (2\gamma^2)^{2+\alpha}$. Thus, the precise form of the relativistic beaming factor is model dependent. Often the relativistic beaming factor is written κ^δ where $\delta = \alpha + n$ with $n \sim 2-3$. It should also be noted that the external observer observes the source components at different proper times in their own rest frames. Therefore, if the jets were time variable and evolved in exactly the same way, the intrinsic luminosities of the components observed at a fixed time in the observer's frame would be different.

We can now revisit some aspects of the jet model shown in Fig. 22.3. The jet is assumed to be relativistic right down to the base of the jet which, on compact enough scales, becomes synchrotron self-absorbed. Thus, the compact core itself is also relativistically beamed and this accounts for the frequency with which compact cores as well as relativistic jets are observed. In a few extreme blazars, it is assumed that the jet is observed more or less precisely along the line of sight, but despite the brightness of the nucleus, high dynamic range synthesis imaging has enabled the underlying double radio sources to be observed. A good example is the blazar 3C 371, which looks to all appearances like an FR2 radio source observed end-on (Fig. 18.15).

Since the superluminal jets are pointing at a small angle to the line of sight, foreshortening effects are important in understanding the physical structure of the jets. A good example is the observed half-angle of the jet, ϕ in Fig. 22.3. If the jet is observed to have half-angle ϕ_{obs} , then the intrinsic opening half-angle is $\phi = \phi_{\text{obs}} \tan \theta$, where θ is the angle between the axis of the source and the line of sight. Marscher (1993) showed that, in the cases of the sources NRAO 140 and 3C 345, the intrinsic opening angles of the jets must be only $\phi \approx 1.5-2^\circ$. Thus, the relativistic beams must already be very well collimated indeed within a distance of only 1–10 pc from the nucleus. Bends and kinks in the jets are also strongly foreshortened and so, although the angles may look large on radio maps, when deprojected by the $\tan \theta$ factor, the bends in the jets are quite small. There are similar deviations from perfect symmetry in the large scale angular distribution of hot spots of about $7-10^\circ$ (Best *et al.*, 1995) and such deviations can probably encompass even the largest deviations observed in superluminal sources.

While VLBI provides the bulk of the imaging evidence for superluminal motions, the jet in M87 seen in Fig. 3.2a is an example of an FR1 radio source in which relativistic motions have also been measured by optical imaging. VLA and VLBI observations showed that the radio jet moves out from the nucleus at about $(0.28 \pm 0.08)c$, while knot A, the brightest feature in the jet, has a projected velocity of $(0.52 \pm 0.03)c$. Optical observations by the Wide Field Camera of the Hubble Space Telescope showed motions of the knots in the jet in the range $0.5-1c$, while the outermost knot has a velocity of only $0.1c$ (Biretta *et al.*, 1995). More recently, Biretta and his colleagues carried out optical imaging of features in the jet very close to the nucleus and found large superluminal velocities (Fig. 22.7). There are 10 superluminal features within the first 6 arcsec of the jet, eight of them having apparent speeds in the range $4-6c$ (Biretta *et al.*, 1999). These observations are consistent with the picture of FR1 radio galaxies discussed in Sect. 21.3, in which it was asserted that such

**Fig. 22.7**

The inner regions of the jet of M87 as observed by the Hubble Space Telescope through the period 1992–1998 (Biretta *et al.*, 1999). The optical knots are observed to move out from the nucleus at speeds up to $6c$.

sources have relativistic jets close to the nucleus which are decelerated to mildly relativistic velocities on their outward passage through the host galaxy .

22.4 The superluminal source population

Understanding the nature of the superluminal source phenomenon in compact extragalactic radio sources requires the systematic study of complete samples of such sources over a long time interval. As discussed in Sect. 22.2, this has now been achieved through the major surveys undertaken by Cohen, Kellermann and their colleagues (Cohen *et al.*, 2007). From the flux density limited survey described in that section, they assembled a statistical sample of 119 superluminal core-jet sources, most of which had been observed over a 10-year period and for which ‘excellent’ or ‘good’ structures had been determined. The maximum observed velocity $\beta_{\text{app}} = v_{\text{app}}/c$ for each source is plotted against observed luminosity L in Fig. 22.8. The symbols distinguish between quasars, BL-Lacs objects and galaxies, as described in the figure caption.

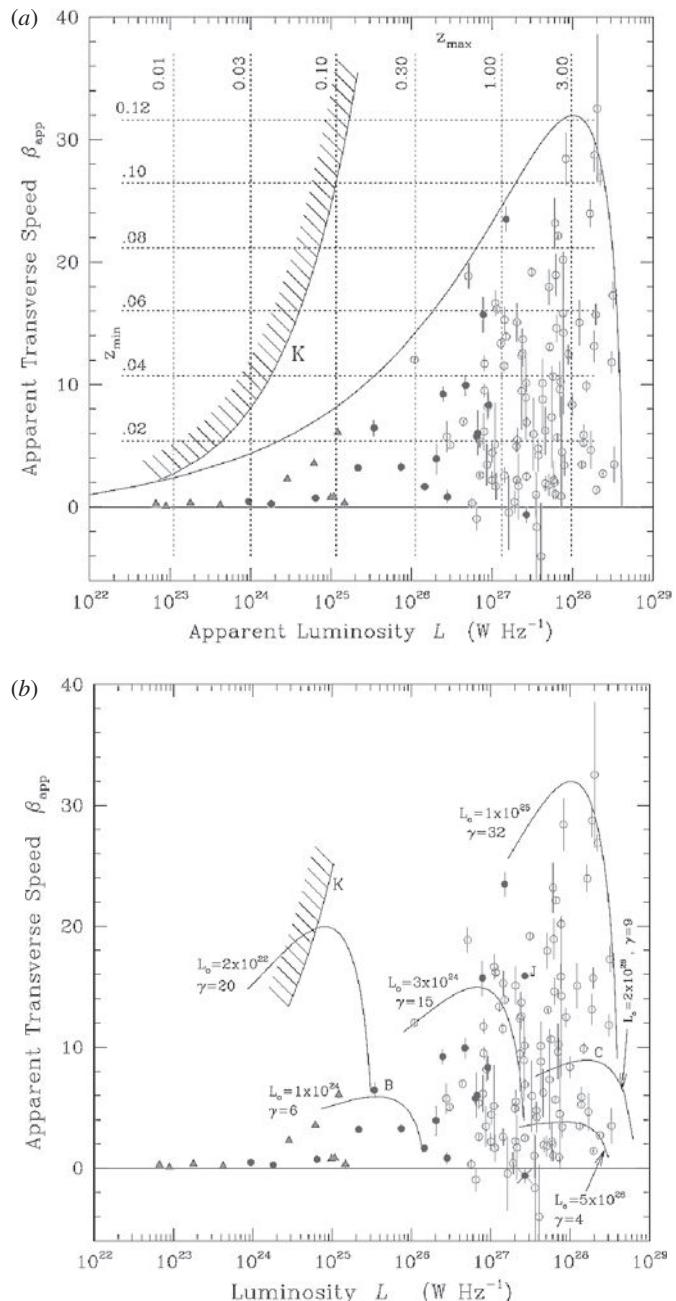


Fig. 22.8

(a) A plot of apparent transverse speed, $\beta_{\text{app}} = v_{\text{app}}/c$, against apparent luminosity, L , for the fastest component in each of 119 core-jet sources in the 2 cm VLBA survey (Cohen *et al.*, 2007). The aspect curve is the locus of (β_{app}, L) for sources with $\gamma = 32$ and $L_0 = 1.0 \times 10^{25} \text{ W Hz}^{-1}$, as θ varies. Curve K is an observational limit for $S_{\text{VLBA, med}} = 0.5 \text{ Jy}$ and $\mu = 4 \text{ mas yr}^{-1}$; sources in the hatched region are usually unobservable. The horizontal lines are the minimum values of redshift, $z_{\min}(\beta_{\text{app}})$, for which the angular velocity is below the limit $\mu = 4 \text{ mas yr}^{-1}$. The vertical lines are the maximum values of redshift, $z_{\max}(L)$, for which the flux density is greater than the limit, $S = 0.5 \text{ Jy}$. Open circles are quasars, full circles are BL-Lac objects and triangles are galaxies. (b) The same data as in Fig. 22.8, but with lines indicating the loci of sources with fixed values of β and L for values of θ within which there is a 92% chance of the sources being observed.

It is a non-trivial exercise to work out all the selection effects which determine which sources appear on Fig. 22.8 and even more of a challenge to interpret what the diagram is telling us. First of all, the sample is *flux-density limited* but the luminosities which appear on the diagram have been greatly enhanced from their intrinsic values by the effects of relativistic beaming, which depend upon both the Lorentz factors γ of the sources and the angles θ of their axes to the line of sight – both of these are unknown for any given source. Second, there are limits to the observable values of the proper motions of the source components. If they were too great, they would not be detected in the 10 years of the survey. The line K in Fig. 22.8a indicates an upper limit of 4 milliarcseconds year⁻¹. Sources in the hatched area to the left of K would not be observable in the flux density limited sample.

To interpret the diagram, we need the pair of equations (22.2), (22.17) and the expression for the observed luminosity of the superluminal source:

$$\beta_{\text{app}} c = v_{\perp} = \frac{v \sin \theta}{1 - \frac{v \cos \theta}{c}} ; \quad \kappa = \frac{1}{\gamma \left(1 - \frac{v \cos \theta}{c} \right)} ; \quad L_{\text{obs}} = \kappa^{\delta} L_0 , \quad (22.25)$$

where $\delta = \alpha + n$ and $n \sim 2-3$. The solid line on Fig. 22.8a shows the locus corresponding to an intrinsic luminosity $L_0 = 1.0 \times 10^{25}$ W Hz⁻¹ and a Lorentz factor $\gamma = 32$ as the angle θ is varied from 90° at the left-hand end of the line to 0° at the right-hand end. It can be seen that this locus provides a bounding curve within which all the superluminal sources are observed. Sources with these intrinsic properties are expected to lie along this locus, but the probability of the extreme high and low values being observed is very small, because they correspond to small solid angles on the sphere.

Cohen and his colleagues carried out a number of simulations to determine within what ranges of observed luminosities there would be a 92% chance that the source is observed, the truncation criteria being that the cumulative probabilities should lie between 4% at low luminosities and 96% at high luminosities. In Fig. 22.8b, the same curve with $L_0 = 1.0 \times 10^{25}$ W Hz⁻¹ and a Lorentz factor $\gamma = 32$ is shown, but with these truncation values included. It can be seen that the most likely values are quite closely clustered about the maximum observed luminosity and maximum apparent velocity – this rule holds good for the other truncation curves illustrated in Fig 22.8b. The reason for this is that the maximum enhancement takes place for angles close to the value $\cos \theta = v/c$ at which the greatest value of v_{\perp} is observed. These curves, referred to as *aspect curves* by Cohen and his colleagues, are useful tools in interpreting the distribution of sources in Fig. 22.8. For example, there is a real limit to the values of apparent luminosity at high values of apparent transverse speed, which is not the result of observational selection.

Cohen and his colleagues (2007) draw the following conclusions from their analysis:

- About half of the sources with $\beta_{\text{app}} > 4$ in a flux density limited sample have values of γ within 20% of the value of β_{app} .
- The aspect curve with $L_0 = 1.0 \times 10^{25}$ W Hz⁻¹ and a Lorentz factor $\gamma = 32$ forms an envelope within which all the sources in the flux density limited sample lie. From probability arguments, the maximum Lorentz factor is about 30 and the peak intrinsic luminosity $L_0 = 1.0 \times 10^{26}$ W Hz⁻¹.

- There is an observed correlation between β_{app} and L for the jets of quasars. High values of β_{app} are only found in radio jets with high values of L . This implies a corresponding correlation between γ and L_0 . In other words, high values of γ must preferentially exist in jets with high values of L_0 .
- The exponent n for a typical source must be less than 3 or else the highly luminous jets with the fastest superluminal speeds would have intrinsic luminosities comparable to those of slow nearby galaxies which are observed to have only mildly relativistic jets. If $n = 2$, the largest values of γ would be associated with intrinsically much more luminous jets. This value of n is consistent with the assumption that the superluminal sources are continuous jets with $n \sim 2$.
- According to probability arguments, there are too many low speed ($\beta_{\text{app}} < 3$) quasars in the sample. Cohen and his colleagues argue that some of these may have pattern speeds substantially lower than their beam speeds.
- The galaxies have a distribution of Lorentz factors up to $\gamma = 7$. Three show strong superluminal motion, but most are only mildly relativistic. They argue that these are not ‘off-axis’ versions of the powerful quasars.
- The pattern and beam speeds must be approximately equal for the fastest components in many sources.

It is a major theoretical challenge to account for the formation of jets moving at ultra-relativistic speeds from the nuclei of active galaxies. An attractive possibility is the development of the electromagnetic model of Blandford and Znajek illustrated in Fig. 21.14a. In a series of papers, Lovelace and his colleagues have developed the model taking account of the twisting up of the magnetic field lines which are tied to the accretion disc about the black hole – a brief summary of their results is given in the paper by Lovelace and Romanova (2003). As they show, the twisting of the magnetic field threading a differentially rotating accretion disc extracts angular momentum and energy from the disc magnetically. If there is significant mass flux in the outflow, the energy and angular momentum are carried outwards hydrodynamically by both the matter and the electromagnetic field. If the mass flux is negligible, energy and angular momentum are carried predominantly by the electromagnetic field, what are referred to as *Poynting outflows*. The structure of the Poynting outflow is illustrated schematically in Fig. 22.9. Lovelace and Romanova show that the speed of advance of the jet is given by the expression

$$\gamma \approx 8 \left(\frac{10}{\mathfrak{R}} \right)^{1/3} \left(\frac{B_0}{0.1 \text{ T}} \right)^{1/3} \left(\frac{10^6 \text{ m}^{-3}}{n_{\text{ext}}} \right)^{1/6}, \quad (22.26)$$

where $\mathfrak{R} = r_0/r_g \gg 1$, r_0 is the radius of the O-point of the magnetic field, $r_g = GM/c^2$, and B_0 is the magnetic flux density at the center of the disc. As expressed by Lovelace and Romanova,

As the ‘twist’, as measured by $\Omega_0 t$, increases, a high twist field configuration appears with a different topology. A *plasmoid* consisting of toroidal flux detaches from the disk and propagates outward. The plasmoid is bounded by a poloidal field line which has an X-point above the O-point on the disk. The occurrence of the X-point requires that there

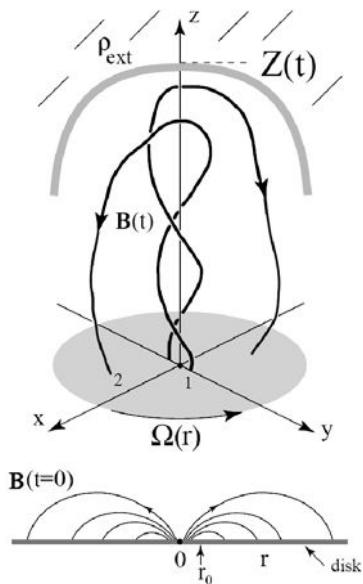


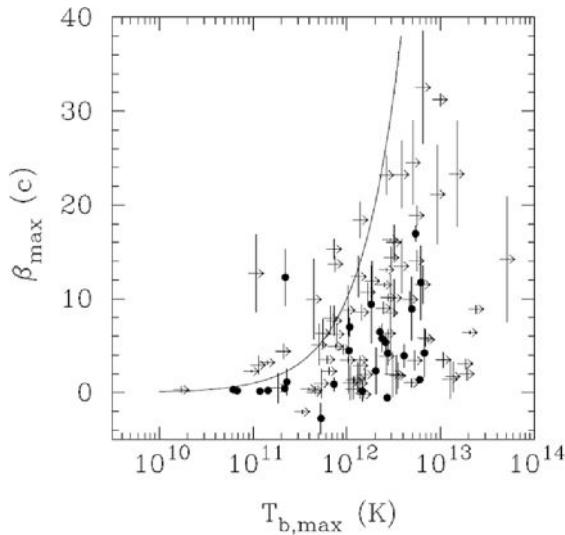
Fig. 22.9 A sketch of the magnetic field configuration of a Poynting jet from Lovelace and Romanova (2003). The lower part of the figure shows the initial dipole-like magnetic field threading the accretion disk which rotates at the angular rate $\Omega(r)$. The upper part of the figure shows the jet at some time later when the head of the jet is at a distance $Z(t)$. At the head of the jet there is pressure balance between the electromagnetic stress of the jet and the ram pressure of the ambient medium of density ρ_{ext} .

be at least a small amount of dissipation in the evolution from the poloidal dipole field and the Poynting jet configuration.

A key feature of the model is that the mass density within the jet should be low. In the picture described by Lovelace and Romanova, the mass in the jet consists of electron–positron pairs. The attraction of the model is that it results in highly relativistic collimated flows which originate in the inner regions of the accretion disc.

22.5 Synchro-Compton radiation and the inverse Compton catastrophe

Synchro-Compton radiation and the inverse Compton catastrophe were analysed in Sect. 9.6 where it was shown that, if the brightness temperature of the radio source exceeds 10^{12} K, the radio source loses energy catastrophically by synchrotron self-Compton radiation. The sample of core-jet sources discussed in Sects 22.2 and 22.4 have been analysed by Kovalev and his colleagues to determine the maximum brightness temperatures within their compact cores (Kovalev *et al.*, 2005). The analysis involved determining the visibility functions for the compact cores in the sources, both along and perpendicular to the axis

**Fig. 22.10**

Apparent velocity, β_{\max} , plotted against maximum core brightness temperature, $T_{b,\max}$, for the most compact components in core-jet superluminal sources for which ‘excellent’ or ‘good’ structures had been determined. Values of β_{\max} are taken for the fastest components of these sources. Lower limits to the brightness temperatures are plotted as arrows. The plotted curve is $T_{b,\max} = \beta_{\max} T_{\text{int}}$, where the intrinsic brightness temperature is taken to be $T_{\text{int}} = 10^{11}$ K.

of the jet directions. Not all the sources were resolved in both directions and so, in a number of cases, the measurements provide lower limits to the brightness temperatures. The median values of the observed brightness temperatures are about 10^{12} K with the values extending up to at least 5×10^{13} K. Figure 22.10 shows a plot of the apparent velocities of the superluminal cores plotted against their maximum brightness temperatures. The largest brightness temperatures tend to be associated with the largest values of apparent velocity. It is natural to associate the largest brightness temperatures with the effects of relativistic beaming. The solid line in Fig. 22.10 is the relation $T_{b,\max} = \kappa_{\max} T_{\text{int}} \approx \beta_{\max} T_{\text{int}}$ derived in Sect. 22.3, the intrinsic brightness temperature being taken to be $T_{\text{int}} = 10^{11}$ K. Thus, brightness temperatures greater than 10^{12} K can be accommodated assuming the cores are relativistically beamed.

The same argument reappears in a slightly different guise from studies of the time variability of the source components. The most compact blazars show variability on a time-scales Δt less than a day and so, according to the causality argument, the dimensions of the source would have to be less than $c \Delta t$. If, however, the source component is moving towards the observer at a relativistic velocity, the time-scale in the frame of reference of the source is $\Delta t' = \kappa \Delta t$ and so the causality constraint is relaxed. An additional factor of κ^2 therefore appears in the expression for the brightness temperature deduced from the flux density and time variability of the source,

$$T_{b,\text{obs}} = \kappa^3 T_{\text{int}} . \quad (22.27)$$

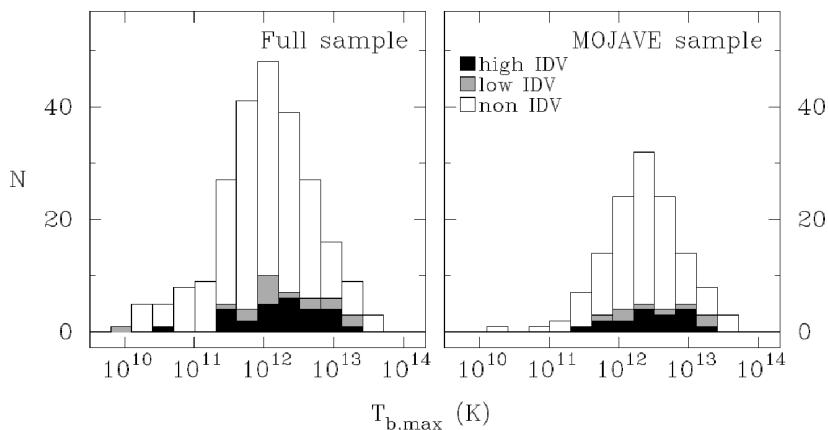


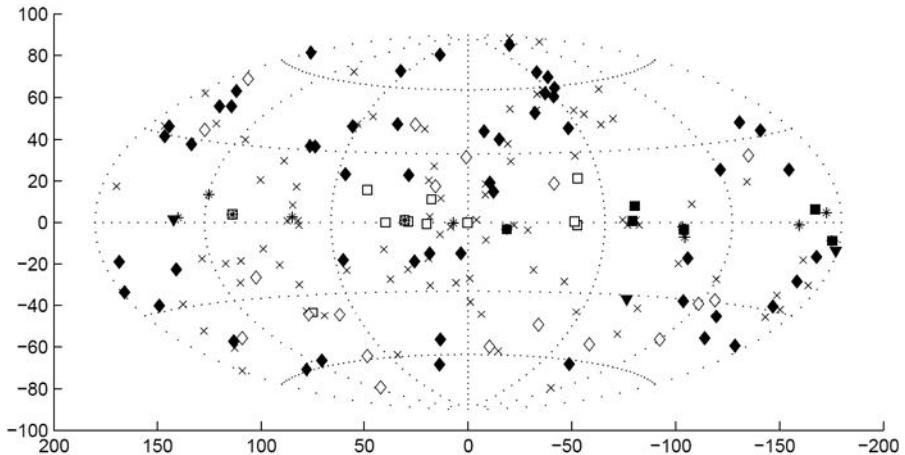
Fig. 22.11 Distributions of the maximum brightness temperature $T_{b,\max}$ for inter-day variable (IDV) selected and non-selected sources. The grey-scale coding shows the distributions for sources which display strong inter-day variability, weak variability and no variability (Kovalev *et al.*, 2005).

It should be emphasised that the exact form of the relativistic beaming ‘correction’ is somewhat model-dependent, but the sense of the changes can be appreciated from Fig. 22.5. The same type of reasoning reappears in studies of the highly variable γ -ray sources.

Figure 22.11 shows the histogram of the brightness temperature distributions of two large samples of core-jet sources, the grey-scale coding distinguishing between those sources which display strong inter-day variability, weak inter-day variability and no inter-day variability (Kovalev *et al.*, 2005). It is apparent that the strong variable sources are those which have high brightness temperatures as observed by VLBI, but their distribution is not so different from the majority of sources in the sample. On the other hand, Kovalev and his colleagues find that the inter-day variable sources are more compact and have more core dominant structure on sub-milliarcsecond scales than the non-variable sources.

22.6 γ -ray sources in active galactic nuclei

One of the key discoveries of the Compton Gamma-Ray Observatory (CGRO) was the observation of ultra-luminous, extragalactic γ -ray sources at energies $\epsilon \geq 100$ MeV. Figure 22.12 shows the distribution of point γ -ray sources from the revised catalogue of EGRET γ -ray sources of Casandjian and Grenier (2008) in a Hammer–Aitoff projection. There are 188 sources in the revised catalogue derived from the complete dataset obtained during the lifetime of the Compton Gamma-Ray Observatory. The figure caption shows that the bright sources comprise a variety of high energy astrophysical objects, including γ -ray pulsars, blazars, flat spectrum radio sources and supernova remnants; for 87 of the sources, identifications have not been made. A striking feature of the early observations made by the CGRO was that 24 of the brightest extragalactic sources were all associated with radio-loud, flat-spectrum radio sources, many of which exhibited superluminal motions (Michelson,

**Fig. 22.12**

The revised EGRET source catalogue in Galactic coordinates (Casandjian and Grenier, 2008). Identified pulsars are shown as black squares; other ATNF pulsars as open squares; blazar candidates as black diamonds; other flat-spectrum radio-sources as open diamonds; supernova remnants as stars; no counterpart as crosses.

1994). Figure 22.12 confirms that the majority of the extragalactic sources are associated with blazars, many of which exhibit superluminal motions.

These findings have been confirmed and greatly enhanced with the launch of the *Fermi Gamma-Ray Space Telescope*. Observations from the first three months of observation showed that the telescope has achieved much greater sensitivity than the CGRO with improved angular resolution. During this period, 132 bright sources were detected at $|b| > 10^\circ$ at better than the 10σ sensitivity level. 106 of these sources form the Large Area Telescope Bright Active Galactic Nuclei (AGN) Sample (LBAS). There are only two radio galaxies, Centaurus A and NGC 1275, but 104 blazars including 57 flat spectrum radio quasars, 42 BL Lac objects, and five blazars with uncertain classification (Abdo *et al.*, 2009b). All the blazars are variable γ -ray sources and are detected as radio sources. Kovalev and his colleagues (2009) monitored the radio properties of this sample with the VLBA almost contemporaneously and found that the γ -ray flux density is strongly correlated with the compact radio flux density. In particular, the γ -ray detected sources in the radio-selected complete sample generally have higher compact radio flux densities and their parsec-scale cores have higher brightness temperatures than the jets in non-detected objects. Kovalev and his colleagues inferred that the jets of bright γ -ray active galactic nuclei are associated with the more extreme superluminal sources. The radio jets were found to be in a more active radio state within several months of their detection as strong γ -ray sources. These results are consistent with a jet model in which the γ -ray and radio emission originate from the same collimated beam, as illustrated in Fig. 22.3.

The spectra of the powerful γ -ray sources show that most of the energy liberated by these objects is emitted in the γ -ray region of the spectrum. The γ -ray luminosities can be written in terms of a beaming factor f , which describes the fraction of the celestial sphere over which the source emits its observed intensity. Typically, the γ -ray luminosities amount to about $10^{41} f$ W, with a few examples up to 10 times greater than this value. These

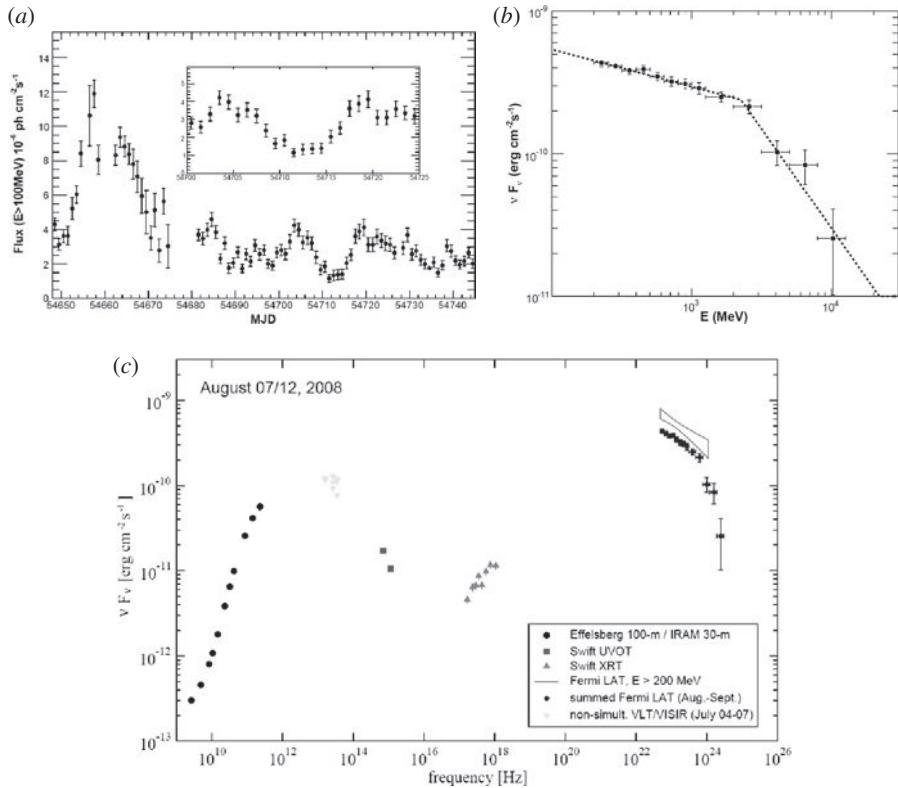


Fig. 22.13 The γ -ray properties of the blazar 3C 454.3 showing (a) the variability of the γ -ray source, (b) the γ -ray spectrum and (c) the broad-band energy spectrum from radio to γ -ray energies (Abdo *et al.*, 2009a).

are enormous luminosities, if the sources were to radiate isotropically, that is, if $f \sim 1$. Fermi Large Area Telescope observations of the quasar 3C 454.3 provide a good example of the properties of these strong γ -ray sources (Abdo *et al.*, 2009a). Over the period 7 July–6 October 2008, the γ -ray emission at energies greater than 100 MeV was highly variable (Figure 22.13a). Figure 22.10b shows the non-thermal γ -ray spectrum of the source, while Fig. 22.10c shows the energy spectrum of the source from radio to γ -ray energies. 3C 454.3 is a well known superluminal source with relativistic beaming factor $\kappa \approx 25$. It is apparent that the γ -ray emission originates in the same relativistic jets which give rise to the observed superluminal velocities.

A further argument concerns the enormous energy densities in γ -rays which must be present in these compact variable sources. The sources are also intense hard X-ray emitters and so the γ -rays lose energy in photon–photon collisions, leading to the creation of electron–positron pairs. The result of this process is the formation of what Blandford (1994) calls a ‘ γ -ray photosphere’ about the nucleus – the γ -rays we observe cannot originate from within this region.

A key role is played by the *compactness parameter* which determines whether or not the source is opaque for $\gamma\gamma$ collisions for electron–positron pair production. For simplicity,

consider the flux of γ -rays at the threshold for electron–positron pair production, $\varepsilon \sim m_e c^2$. The mean free path for $\gamma\gamma$ collisions is $\lambda = (N_\gamma \sigma)^{-1}$, where N_γ is the number density of photons with energies $\varepsilon = h\nu \sim m_e c^2$ and $\sigma \approx m_e r_e^2$, $r_e = 2.8 \times 10^{-15}$ m being the classical electron radius. If the source has luminosity L_γ and radius r , the number density of photons within the source region is

$$N_\gamma = \frac{L_\gamma}{4\pi r^2 c \varepsilon}. \quad (22.28)$$

The condition for the source to be opaque is $r \approx \lambda$, that is,

$$r \approx \frac{4\pi r^2 c m_e c^2}{L_\nu \sigma}, \quad \text{or} \quad \frac{L_\nu \sigma}{4\pi m_e c^3 r} \sim 1. \quad (22.29)$$

The compactness factor C is defined to be the quantity

$$C = \frac{L_\nu \sigma}{4\pi m_e c^3 r}. \quad (22.30)$$

Sometimes the compactness parameter is defined without the factor of 4π in the denominator. If C is much greater than unity, the γ -rays are all degraded by electron–positron pair production, resulting in a huge flux of electrons and positrons within the source region. Consequently, the source would no longer be a hard γ -ray source. The significance of the compactness parameter can be appreciated for a γ -ray source with isotropic luminosity $L_\gamma \sim 10^{41}$ W which is observed to vary on the time-scale of a day. Inserting these values into (22.30), $C \sim 10^3 \gg 1$ and so there is a problem in understanding why such sources should exist. Relativistic beaming again comes to the rescue. Because of relativistic beaming, the observed luminosity is enhanced by a factor $\kappa^{2+\alpha}$ and, in addition, because the time-scale of variability appears in the denominator of (22.27), the observed value is shorter by a factor κ and so the compactness parameter is increased by relativistic beaming by a factor of roughly $\kappa^{4+\alpha}$. Since $\alpha \approx 1$, $C \propto \kappa^5$. Notice again that the exact dependence upon κ is model dependent. Therefore, if $\gamma \geq 4$, in the frame of the source components themselves, the value of the compactness parameter can be reduced below the critical value $C = 1$, resolving the problem of the compactness parameter. From their more detailed analysis, Abdo and his colleagues (2009a) infer that, taking $C \propto \kappa^4$ the relativistic beaming factor has to be greater than $\kappa = 8$.

The favoured model is one in which the source of the γ -ray emission is the same relativistic jet which is responsible for the radio jet but, because of the compactness of the source, the emission probably originates much closer to the nucleus, in the regions in which the radio core is synchrotron self-absorbed, the region labelled ‘core’ in Fig. 22.3. Blandford (1994) argues that the most likely region in which the γ -rays are produced is at the point where the relativistic jet has just emerged from the γ -ray photosphere, which can be taken to be the region at which $C \approx 1$. The emission process is probably some version of inverse Compton scattering in which low energy photons are boosted to enormous energies by the high energy electrons present in the relativistic beam. The source of the photons which are scattered to high energies is a matter of speculation but various possibilities have been suggested. For example, the photons may be created in the inner regions of an accretion disc about a supermassive black hole. Alternatively, the photons from the accretion disc may be

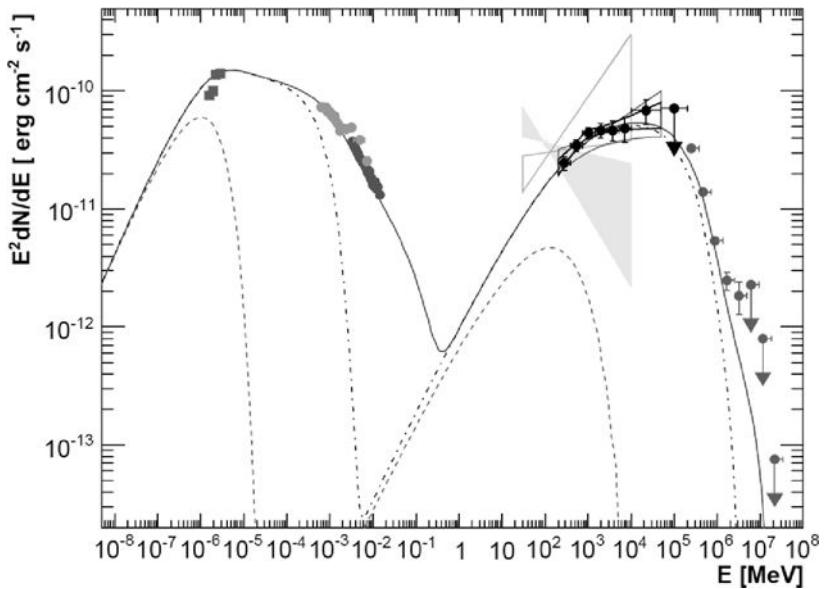


Fig. 22.14 The spectral energy distribution of the blazar PKS 2155–304 observed by the Fermi Gamma-Ray Space Telescope, the HESS atmospheric Cherenkov array, the ATOM telescope and the RXTE and Swift observatories derived from simultaneous observations through the period between 25 August and 6 September 2008 (Aharonian *et al.*, 2009). The solid line is a one-zone synchro-self-Compton model due to Band and Grindley (1985). The very high energy region of the spectral energy distribution is absorbed because of photon–photon interactions with extragalactic background radiation.

scattered into the beam by Thompson scattering by free electrons in the environment of the nucleus. A third possibility is that the soft photons are created within the relativistic jet itself.

The synchro-self-Compton scattering models can give a reasonable fit to the extremely broad-band observations made by combining observations made with the Fermi Gamma-Ray Space Telescope, the HESS atmospheric Cherenkov array and telescopes operating at X-ray and optical wavelengths. An example of this is shown in Fig. 22.14 of the blazar PKS 2155–304 (Aharonian *et al.*, 2009). The broad-band spectral energy distribution shows two broad peaks which are interpreted in terms of the models developed by Band and Grindley (1985), examples of which were discussed in Sect. 9.6 and illustrated in Fig. 9.15. To match the observations, it is assumed that both radiation peaks are relativistically beamed with a bulk beaming factor $\kappa = 32$. The radius of the emitting region is 0.05 pc and so the radiation originates from the inner regions of the jet. At the very highest TeV energies, absorption by $\gamma\gamma$ interactions of the γ -rays with the extragalactic optical and infrared background radiation result in an exponential cut-off to the observed spectrum.

The observations discussed above referred to the blazar PKS 2155–304 in its quiescent state but, during periods of activity, it is the most highly variable TeV γ -ray source yet observed. The outburst observed by the HESS telescope array in July 2006 was the most dramatic seen from any TeV γ -ray source (Aharonian *et al.*, 2007a). The light curve of the outburst of July 2006 reached two orders of magnitude greater flux densities than those observed in the ‘quiescent’ phase and was an order of magnitude greater than the flux

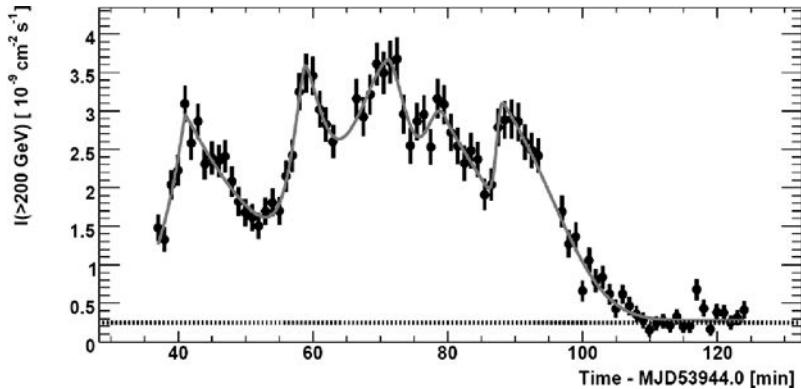


Fig. 22.15 The integral flux density for energies $\varepsilon \geq 200$ GeV observed from PKS 2155–304 on MJD 53944 versus time as observed by the HESS Cherenkov array telescope (Aharonian *et al.*, 2007a). The horizontal line represents the corresponding flux density observed from the Crab Nebula. Five separate bursts have been identified.

density of the Crab Nebula (Fig. 22.15). Variability was observed on the time-scale of a few minutes, the best measured individual flare having a rise-time of only 173 ± 23 s. Such rapid variability would require an extremely large Doppler factor. Aharonian and his colleagues argue that the host galaxy is a luminous galaxy and so is likely to host a black hole with mass $M \sim 10^9 M_\odot$. According to (14.8), the time-scale associated with the Schwarzschild radius of such a black hole is 10^4 s. Consequently, a relativistic beaming factor of $\kappa \geq 60$ would be required to account for the rapid changes in intensity. A similar limit would be required to account for the extreme values of the compactness parameter.

A survey of recent observations of extragalactic TeV γ -ray sources detected by ground-based γ -ray Cherenkov telescopes is given by Hinton (2009), who provides a useful list of currently active telescopes as well as a list of 19 active galactic nuclei from which TeV γ -rays have been detected. Most of the sources have redshifts $z \lesssim 0.2$ because of the cut-off due to $\gamma\gamma$ interactions with the extragalactic background radiation. Evidence that the cut-off is indeed due to this process is provided by the steepening of the TeV γ -ray spectra at redshifts $z \gtrsim 0.1$, although there is considerable dispersion in the spectral indices. The sensitivity of the detector arrays has, however, improving steadily. The recent detection of the flat spectrum radio quasar 3C 279 at a redshift $z = 0.536$ by the MAGIC collaboration represents a major advance (Berger *et al.*, 2009). These observations can be used to set significant limits to the extragalactic background intensity in the optical and infrared regions of the spectrum as illustrated by the analyses of Aharonian and his colleagues (Aharonian *et al.*, 2006).

22.7 γ -ray bursts

22.7.1 The discovery of γ -ray bursts

The discovery of γ -ray bursts was reported in a paper by Klebesadel, Strong and Olson of the Los Alamos National Laboratory in 1973, six years after the first of these events

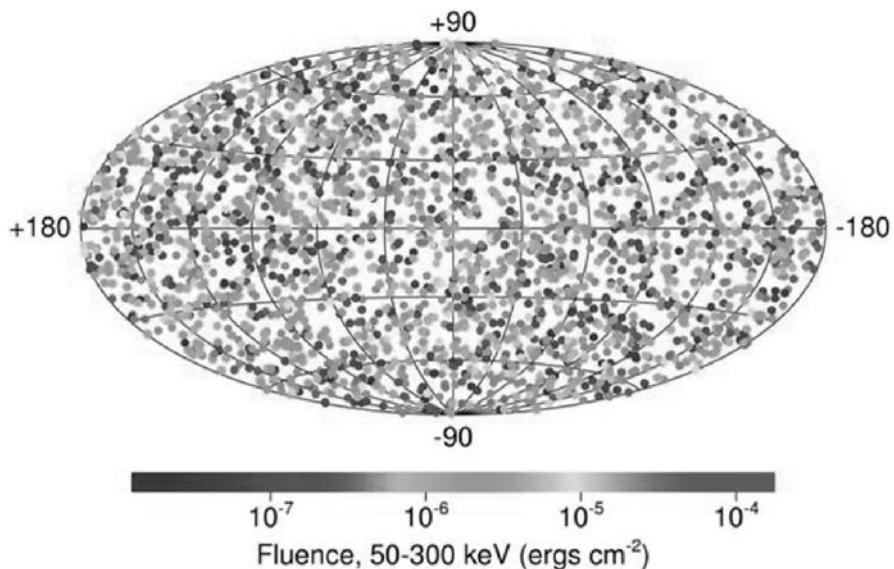
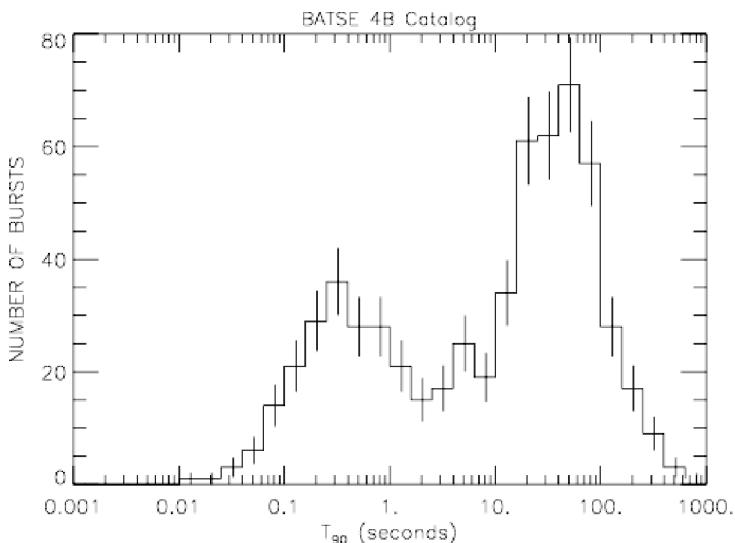


Fig. 22.16 The distribution of 2704 γ -ray burst over the sky as observed by the BATSE experiment of the CGRO. (Courtesy of NASA, G.J. Fishman and the CGRO Science Team.)

had been discovered by the Vela surveillance satellites (Klebesadel *et al.*, 1973). The delay in publication was not caused by security considerations, but rather because the authors wished to be certain that the γ -ray bursts were genuinely astronomical phenomena. The bursts lasted from a fraction of a second to several minutes (Fig. 22.15) and during which time the γ -ray bursts were the most luminous objects in the γ -ray sky. The Compton Gamma-Ray Observatory (CGRO) included a detector system specifically designed to detect and locate γ -ray bursts, the Burst and Transient Source Experiment (BATSE). During the 1990s, this instrument detected 2704 bursts and established a number of their key properties:

- The γ -ray bursts are uniformly distributed over the sky (Fig. 22.16).
- The number counts of the bursts show departures from those expected of a uniform Euclidean distribution of sources, suggesting a cosmological distribution of the sources.
- The distribution of γ -ray burst durations is bimodal: bursts which last less than 2 seconds being termed ‘short’ and those of greater duration ‘long’ (Fig. 22.17).

The γ -ray bursts were of such short duration and the angular resolution of the BATSE instrument only a few degrees that there was not time nor adequate positional accuracy to make secure identifications of the sources. The solution to the problem was stimulated by a theoretical paper by Mészáros and Rees (1993). They argued that, if the γ -ray bursts were extragalactic phenomena, the energy densities during the event would be so extreme that a relativistic shock wave would be created in which electrons would be accelerated to very high energies. These electrons would emit synchrotron radiation with a power-law spectrum, consistent with the observed spectra of the bursts. Although the intense

**Fig. 22.17**

The distribution of γ -ray burst durations from the bursts observed by the BATSE experiment of the CGRO showing the bimodality between the short and long bursts. (Courtesy of NASA, G.J. Fishman and the CGRO Science Team.)

γ -ray emission lasts only a short time at γ -ray energies, the emission would be expected to last much longer at lower energies – an *afterglow* was predicted which would appear successively in the X-ray, optical, infrared and radio wavebands.

Searches for the afterglows of γ -ray bursts were carried out by the Italian-Dutch BeppoSAX satellite. In 1997, the X-ray telescope was pointed at the position of the γ -ray burst GRB 970228 within 8 hours of the event having taken place and its X-ray afterglow discovered (Costa *et al.*, 1997). Subsequently, afterglows were observed throughout the electromagnetic spectrum. These observations enabled a precise position for the burst to be determined and the association of this burst with a very faint galaxy established by observations with the Hubble Space Telescope (Sahu *et al.*, 1997). This observation confirmed that the γ -ray bursts form a population of extragalactic objects. Within four years, afterglows from over 40 γ -ray bursts had been detected and 30 of these were associated with a distant host galaxy. A supernova- γ -ray burst association was suggested by the coincidence of the SN 1998bw with the γ -ray burst GBR 980425 by Galama and his colleagues (Galama *et al.*, 1998). This association was established with certainty for the γ -ray burst GRB 030329 observed on 29 March 2003 which was observed optically and showed that the characteristic broad lines of an extremely energetic supernova were observed within days of the event (Hjorth *et al.*, 2003).

The pace of discovery has increased greatly with observations from the HETE-2, SWIFT and FERMI γ -ray telescopes which have detector systems designed specifically to enable very rapid follow-up of γ -ray bursts to be made. SWIFT, in particular, has greatly increased the numbers of γ -ray bursts for which secure identifications have been made, both for the long and short bursts. The feature of SWIFT which made this possible is the automatic repointing of the telescope when a γ -ray burst occurs so that within two minutes of the

event occurring the onboard X-ray and optical-UV telescope can observe it. SWIFT has measured about 100 γ -ray bursts per year since its launch in 2005, 380 bursts had been detected by December 2008, for 126 of which redshifts had been measured. Most of the redshift measurements are for ‘long’ bursts, but 33 short bursts have been measured and for eight of these redshifts have been measured.

The γ -ray bursts are so luminous that they can be identified in extremely distant galaxies. The gamma-ray burst GRB 090423, discovered by the SWIFT γ -ray telescope, has the largest redshift of any known object, $z = 8.26^{+0.07}_{-0.08}$ (Tanvir *et al.*, 2009). This redshift can be compared with those of the most distant spectroscopically confirmed galaxy ($z = 6.96$) and quasar ($z = 6.43$). Cosmologically, these objects therefore provide probes of the reionisation era when the intergalactic gas was heated and reionised by ultraviolet radiation from the earliest generations of stars in galaxies. As expressed by Gehrels and his colleagues (2009) in their comprehensive review, ‘(γ -ray bursts) are now rapidly becoming powerful tools to study the detailed properties of the galaxies in which they are embedded and of the Universe in general’.

22.7.2 The properties of γ -ray bursts

With the greatly expanded sets of high quality observations from the present generation of γ -ray telescopes, a vast amount of detailed information is now available. The principal distinction remains between the long and short bursts, as is vividly demonstrated by the correlation diagrams shown in Fig. 22.17 from the review by Gehrels and his colleagues (2009). The main features of these diagrams are as follows:

- At cosmological distances, the energy releases of the γ -ray bursts are enormous. It is conventional to write the energy emitted as $E = E_{\text{iso}}(\Omega/4\pi)$, where Ω is the solid angle within which the energy is emitted. If the radiation were to be emitted isotropically, $\Omega = 4\pi$ and so $E = E_{\text{iso}}$. The diagrams make the point that the most energetic γ -ray bursts have energies $E = E_{\text{iso}}$ up to about 10^{47} J, corresponding to the rest-mass energy of $0.5 M_{\odot}$. However, the radiation is likely to be strongly beamed and so the energy requirement can be very significantly reduced.
- The overall time-scales of the bursts are parameterised by the quantity T_{90} which is defined to be the time within which 90% of the energy is liberated. The short bursts are significantly less energetic than the long bursts. The conventional picture is that the bimodality is associated with different types of progenitors for the long and short bursts.
- There is a difference in the γ -ray spectra of the short and long bursts, the short bursts having somewhat hard (flatter) spectra than the long bursts.
- The rise times of the long bursts can be as short as 100 ms and of the short bursts as short as 10 ms. The decay times are somewhat longer than the rise times.
- The variability or ‘spikiness’ of the burst is correlated with the energy or peak luminosity of the burst.
- For the long bursts, there is a time lag between the burst being observed at high and low γ -ray energies. For the SWIFT observations, the energy bands are defined to be 50–100

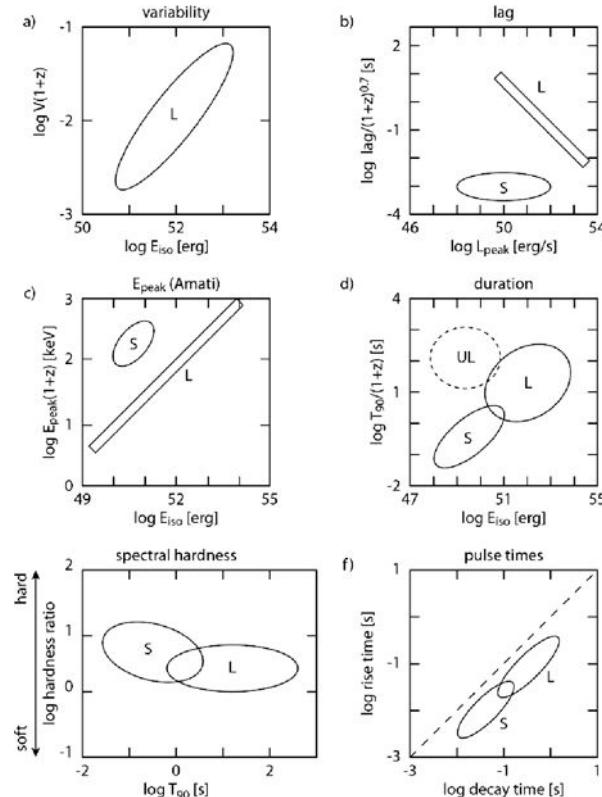


Fig. 22.18

Illustrating a number of the properties of γ -ray bursts, distinguishing between the long (L) and short (S) bursts (Gehrels *et al.*, 2009). UL stands for underluminous. The meanings of the various quantities are explained in the text.

keV (high energy) and 15–25 keV (low energy) channels. In contrast, no such lags are observed for the short bursts.

In addition to these differences, the host galaxies are different. The long bursts are associated with star-forming galaxies, with typical luminosities $L \approx 0.1 L^*$ and metallicities only about 10% of the solar value – these values are significantly smaller than galaxies from the Sloan Digital Sky Survey at the same redshift. The association of the long bursts with core-collapse supernovae has been convincingly demonstrated in many cases. The statistics of short bursts are much smaller than those of the long bursts but there is a clear correlation with regions of low star-formation, either in elliptical galaxies or in a region of low star-formation rate within galaxies. The luminosities of the elliptical host galaxies correspond to $L \sim L^*$ galaxies.

The afterglows have played a particularly important role in understanding the nature of the γ -ray bursts. Before the SWIFT observations began to appear in 2005, the afterglows were typically observed about 1000 seconds after the burst and the spectrum in the X-ray, optical and radio wavebands consisted of a power-law spectrum which steepened at later times. The SWIFT observations enabled the afterglows to be observed much sooner after

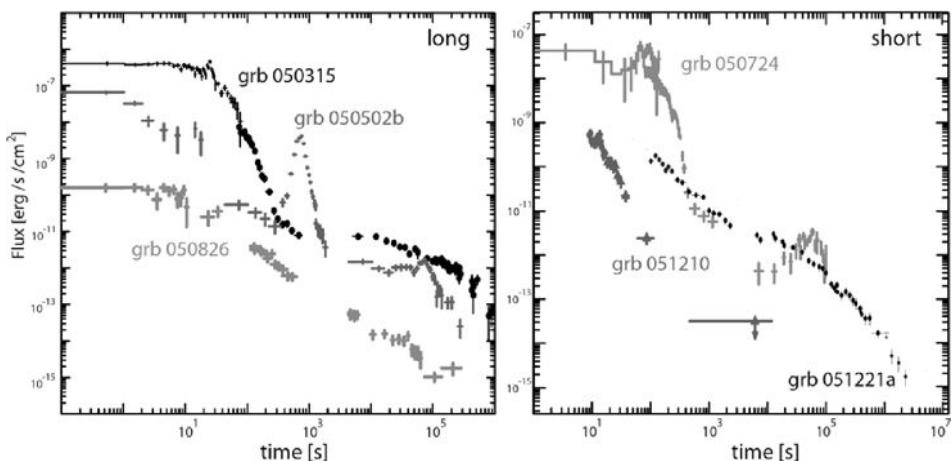


Fig. 22.19 Examples of the time evolution of the X-ray afterglows of representative examples of long and short γ -ray bursts (Gehrels *et al.*, 2009).

the burst and somewhat complicated this picture. Examples of the afterglows of long and short bursts are shown in Fig. 22.19. After about 10^3 seconds, the power-law decline of the X-ray afterglow is observed and in a number of cases an increased decline rate is observed after about 10^5 seconds, as inferred from the earlier observations. The new features are the very rapid decline in the first 100 seconds followed by a very slow decline, typically from about 100 to 1000 seconds. In addition, X-ray flares are observed in both the long and short bursts.

As might be expected, the wealth of data from γ -ray telescopes such as SWIFT have revolutionised our understanding of the nature of the γ -ray bursts, but also led to an appreciation of the diversity of phenomena associated with the bursts. The above broad-brush picture is adequate for our present purposes, but there are many interesting variants on these correlations which are thoroughly described by Gehrels and his colleagues (2009).

22.7.3 The physics of γ -ray bursts

We can use many of the tools developed in this and previous chapters to understand the physics relevant for the γ -ray bursts. A review of many aspects of the theory of γ -ray bursts has been provided by Mészáros (2002). According to the standard causality argument, a time-scale of the variability of the order of milliseconds indicates that the energy must originate from regions less than about 3×10^5 km in size. Hence the bursts must involve stellar-mass objects and huge energy releases, typically 10^{46} J if the emission were isotropic. The resulting energy density in γ -rays within the source region would then be so great that the same problems afflicting the extreme γ -ray luminosities of blazars and quasars, the degradation of the γ -rays by photon–photon interactions (Sect. 22.6), apply even more severely. From the energy and scale of the emitting region, it follows that the source region must expand relativistically and so, just as in the extreme extragalactic γ -ray sources, the

γ -ray bursts must involve relativistic bulk motions. The values of the relativistic beaming factors necessary to reduce the compactness factor below the critical value are $\kappa \sim 10^2\text{--}10^3$, greater even than those necessary for the extragalactic γ -ray sources and those observed by radio VLBI.

In the simplest picture, the source is taken to be a *relativistic fireball*, meaning a highly relativistically expanding sphere which heats the surrounding gas and drives a relativistic shock wave into it (Mészáros and Rees, 1993). If the sphere is optically thick, the radiation would be thermalised and a thermal spectrum observed. The observed spectra of the γ -ray bursts are however of non-thermal form, often with power-law spectra of the form $N(\varepsilon) \propto \varepsilon^{-\alpha}$, where $\alpha \sim 2\text{--}3$ at energies greater than about 0.1–1 MeV. Therefore the radiation must originate from optically thin regions later in the expansion. The relativistic shock wave then provides the means of accelerating electrons to very high energies by the first-order Fermi acceleration mechanism discussed in some detail in Sects 17.4 and 17.5. The structure of the expanding sphere resembles the structure shown in Fig. 16.10a. The relativistically expanding fireball drives an external shock into the ambient medium and there is a contact discontinuity between the material of the fireball and the shocked gas. As the sphere decelerates, an internal reverse shock is formed which is driven back towards the origin and reheats the expanding sphere. The difference as compared with Fig. 16.10a is that the external shock is relativistic and the reverse shock can be relativistic as well. All the arguments concerning the formation of a power-law spectrum of accelerated electrons and the growth of the magnetic flux density towards equipartition discussed in Sects 17.4 and 17.5 reappear, but now in relativistic guise.

Prior to the launch of the SWIFT satellite, the characteristic evolution of the synchrotron spectrum of such a sphere, the simplest version of which was discussed in Sect. 22.1, provided a reasonable match to the observed spectrum and time evolution of the afterglows in the X-ray, optical and radio wavebands. Examples of such spectra are shown in Fig. 22.20 (Granot and Sari, 2002). The origin of these spectra can be understood on the basis of the analysis of the properties of synchrotron radiation and the modifications to the spectra due to synchrotron losses discussed in Chaps 8 and 16. For example, in the case of spectrum 1, the standard optically thin synchrotron spectrum is observed in region G between the frequency ν_m , which corresponds to the lower limit of the electron energy spectrum, and ν_c at which the lifetimes of the electrons equals the synchrotron loss time. The power law at frequencies greater than ν_c in region H results from the assumption of the continuous acceleration under the influence of dominant synchrotron losses (Fig. 16.1a). The power law $L_\nu \propto \nu^{1/3}$ in region D corresponds to the low frequency tail of the spectrum of a synchrotron radiation of a single electron. The Rayleigh–Jeans spectrum in region B is the region of self-absorption for the radiation of the lowest energy electrons in the energy spectrum. In spectrum 2, there is a region A in which synchrotron self-absorption takes place with $\nu_m \ll \nu_{sa}$, resulting in the standard relation $I_\nu \propto \nu^{2.5}$. The other spectra correspond to different orderings of the various ‘critical’ frequencies, as discussed by Granot and Sari (2002). The reason for displaying these results is that it turned out that the simplest form of spectrum, example 1 in Fig. 22.20, could account for the spectral shape and time variability of the burst afterglows in the early observations and consequently could provide information about the physical properties of the relativistic fireball.

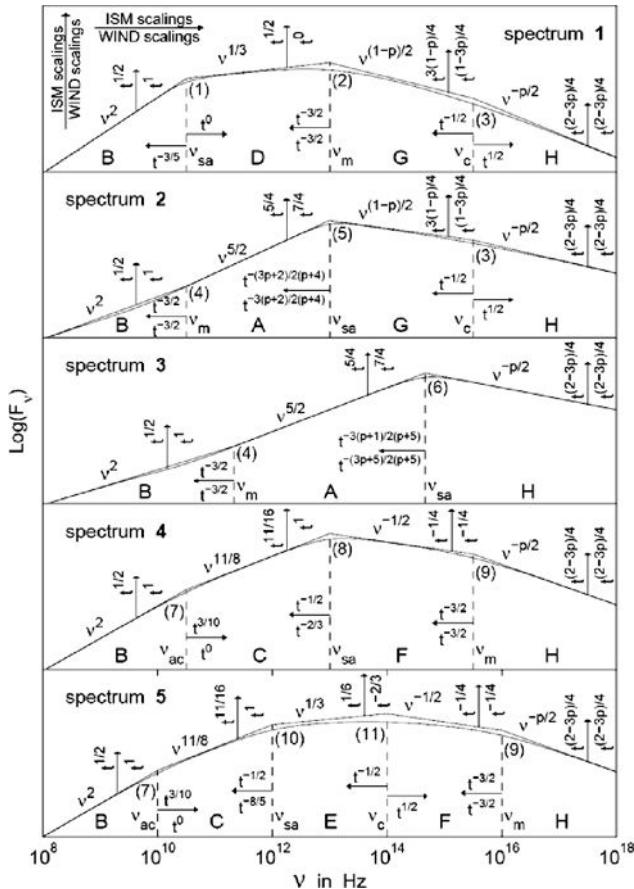


Fig. 22.20

Examples of different broad-band synchrotron spectra from a relativistic blast wave, in which electrons are accelerated with a power-law energy distribution $N(E) dE \propto E^{-p} dE$. The thin solid line shows the asymptotic power-law segments and their points of intersection, where various break frequencies are expected. The different orderings of the various break frequencies are discussed by Granot and Sari (2002).

Relativistic beaming alleviates the problems of $\gamma\gamma$ annihilation, but does not solve the energy problem. The fireball need not, however, be isotropic. In the case in which the energy of the γ -ray burst is emitted in a narrow collimated beam, the energy requirements can be very significantly reduced. In the case of a relativistic jet of opening angle θ , the same overall picture with external and internal shocks and a contact discontinuity illustrated in Fig. 21.7b applies, provided the opening angle θ of the jet is greater than γ^{-1} , where γ is the Lorentz factor associated with the bulk motion of the emitting region and the observer is located within the solid angle of the jet. As the jet sweeps up the surrounding gas and decelerates, however, its Lorentz factor decreases until $\theta < \gamma^{-1}$. Up till that point, the decrease in luminosity has been partly compensated by the fact that the relativistic beaming allows more of the jet to be observed. As γ continues to decrease when $\theta < \gamma^{-1}$, the emission from the whole jet is observed and the observed flux density decreases more rapidly with

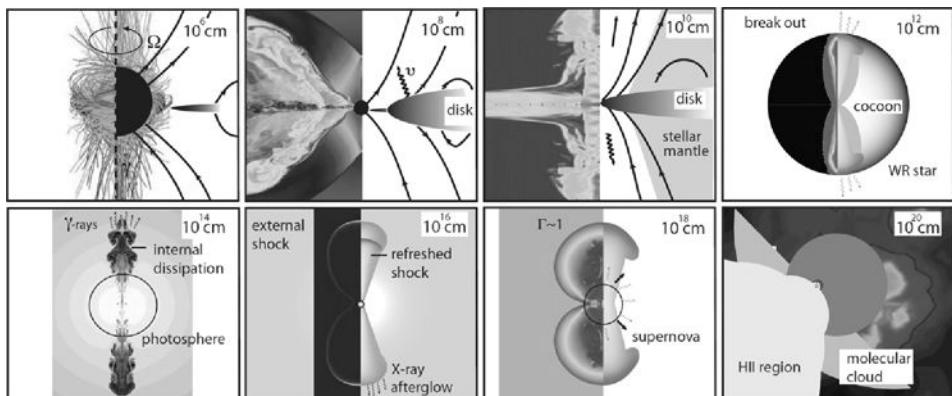


Fig. 22.21 Illustrating the evolution of a γ -ray burst on increasing physical scales (Gehrels *et al.*, 2009).

time. This change can explain the break in the afterglow light curve observed after about 10^5 seconds. In addition, the jet may expand sideways as it decelerates causing yet more rapid decline of the afterglow. If the jet model is adopted, the total energy requirements of the γ -ray bursts can be evaluated and it was found that, rather than a dispersion in intrinsic luminosities from about 10^{44} to 10^{47} J if the radiation is isotropic, there is only about an order of magnitude spread about the value 8×10^{43} J, a value typical of the energy released in the core collapse of a supernova explosion.

This picture of the evolution of the afterglows of γ -ray bursts needs some revision in the light of the SWIFT observations. The full implications of these observations are discussed by Granot (2008) who describes clearly the deficiencies of the model, in particular, the need to account for the initial steep decline of the X-ray light curve and its subsequent flattening. He also points out the problem that, although a break may be observed in the X-ray light curve, this is often not observed in the optical light curve. This observation is contrary to the expectations of the decelerating jet model in which the break in the light curves should be independent of wavelength. He suggests various solutions to these problems.

The association of the long bursts with core-collapse supernovae in which such energy releases are observed makes a compelling picture. A montage of images for the plausible evolution of γ -ray bursts is provided by the simulations cartoons presented by Gehrels and his colleagues (2009) (Fig. 22.21). There is only direct observational evidence for the later stages in the evolution, but the overall picture is consistent with the types of physical process which are inferred to take place in core-collapse supernovae. There are many similarities with the physical processes involved in the release of gravitational energy and the formation of relativistic jets in extragalactic radio sources discussed earlier in this chapter. The one big difference is that jet formation, collimation and the expansion into the surrounding medium take place very rapidly indeed within the collapsing envelope of the collapsing star. The diagrams make the important point that the γ -rays only become observable when the jets reach scales of order 10^7 km. If the relativistic beaming factor is of the order of $\kappa = 10^3$, time variations of the γ -ray burst would be expected on time-scales as short as 30 μ s. The appearance of the afterglows only occur significantly later in the evolution of the relativistic

jet as it is decelerated on passing through the stellar envelope and sweeping up mass. A popular picture is to relate the γ -ray bursts to the collapse of a very massive stars, such as the Wolf–Rayet stars, the cores of which would collapse to form a Kerr black hole and then to associate the relativistic jet with the presence of electric and magnetic fields in the vicinity of the hole, as illustrated in Fig. 22.19 and discussed in Sect. 22.4. Axisymmetric collapse of the outer envelope would continue onto the collapsed core. This picture is often referred to as a *collapsar* model for γ -ray bursts.

The short bursts are much less luminous than the long bursts and belong to different stellar populations. The very short time-scales of variability suggest that neutron stars and black holes are involved but, because they are not associated with star-forming regions in galaxies, they cannot be young stellar systems. An appealing possibility is that they are associated with the merger of binary neutron stars, or of a neutron star and a black hole, or of two black holes. Many of these possibilities are discussed in the review by Mészáros (2002). If this interpretation is correct, the statistics of the short γ -ray bursts provide estimates of the numbers of stellar binary mergers which might be detected by the gravitational wave detectors.

23.1 The cosmic evolution of galaxies and active galaxies

Evidence for strong evolutionary changes of the populations of extragalactic objects with cosmic epoch was first found in surveys of extragalactic radio sources and quasars in the 1950s and 1960s. An excess of faint sources was discovered in radio source and quasar surveys as compared with the expectations of uniform world models. The inference was that these classes of object were much more common at earlier cosmic epochs than they are at the present time. During the 1980s, the first deep counts of galaxies to very faint magnitudes became available thanks to the CCD revolution in optical detector technology. An excess of faint blue galaxies was discovered and these studies were extended to extremely faint apparent magnitudes by Hubble Space Telescope observations of the Hubble Deep Field and the Hubble Ultra-Deep Field.

This pattern of the discovery of excess numbers of faint objects at early cosmic epochs has been repeated in essentially all wavebands as deep surveys have become feasible. In the 1990s, surveys of the X-ray sky carried out by the ROSAT X-ray Observatory provided evidence for an excess of faint X-ray sources, similar to that found for the extragalactic radio sources and quasars. These studies were extended to much fainter X-ray sources by observations with the Chandra and XMM-Newton X-ray Observatories. The IRAS survey of the mid- and far-infrared sky, although not extending to as large redshifts as the radio and X-ray surveys, found evidence for an excess of faint sources. The deep mid- and far-infrared surveys carried out by the Spitzer Infrared Space Telescope have confirmed the large excess of faint infrared sources. Surveys became feasible in the submillimetre waveband in the late 1990s and a very large excess of faint submillimetre sources was discovered, the objects being associated with luminous dust-emitting galaxies.

The cosmological evolution of all classes of extragalactic object with cosmic epoch is a central issue for understanding the origin and evolution of galaxies. These studies have been enormously advanced by the availability of 8–10 metre ground-based optical-infrared telescopes, as well as the large scale surveys of galaxies and quasars resulting from the AAT 2dF Survey and the Sloan Digital Sky Survey.

The study of changes in source populations with cosmic epoch can be broadly separated into two overlapping areas of astrophysical study. The first is the evolution of the stellar and gaseous components of galaxies and clusters of galaxies with cosmic epoch. These studies include the evolution of the stellar populations and physical properties of galaxies, the changes in the numbers and chemical composition of the absorption line systems observed in the spectra of distant quasars, the evolution of the global star-formation rate with cosmic

epoch, and so on. We may think of these areas as comprising the evolution of the *thermal* components of the Universe.

The second concerns the evolution of the *non-thermal* Universe, meaning how the properties of active galactic nuclei – the quasars, the Seyfert galaxies, the extragalactic radio and X-ray sources and their close relatives – have evolved with cosmic epoch. These phenomena are associated with the presence of supermassive black holes in the nuclei of galaxies and their defining characteristic is that their intense emission is dominated by non-thermal radiation processes such as synchrotron radiation and inverse Compton scattering of high energy electrons.

These two aspects of cosmic evolution come together in studies of the co-evolution of the stellar and black hole properties of galaxies. It is now apparent that these are not independent, but rather there is a symbiosis between the evolutionary behaviour of the thermal and non-thermal Universes. For example, supermassive black holes can grow as a result of stellar mass-loss in galaxies, the material cooling and being channelled into the central regions of galaxies where it is consumed by the black hole. Once formed, the black holes influence the evolution of the stellar component of galaxies through various *feedback mechanisms*. A good example is the heating of the intracluster gas by radio source events which can prevent the infall of further baryonic matter into the galaxy and so limit the continued growth of the galaxies. In addition, the active systems provide important cosmological probes of the thermal Universe. For example, quasars and γ -ray bursts are among the most distant sources yet discovered and now extend into the end of the re-ionisation era, thus providing diagnostic tools for the study of physical conditions during the end of the ‘dark ages’.

The primary data involved in both aspects of these studies are large complete samples of objects with accurately known flux densities, or apparent magnitudes, and redshifts. Once redshifts and spectroscopic evidence become available for large samples of sources, the nature of these astrophysical changes with cosmic epoch can be determined in detail. The cosmological study of all classes of extragalactic object and the tools and techniques employed in their study were discussed in detail in my book *Galaxy Formation* (Longair, 2008) and the details of the background theory and procedures are not repeated here. The emphasis in this chapter is upon the evidence for strong cosmological evolution of active galaxies and its astrophysical implications.

23.2 The essential theoretical tools

We first summarise a number of the important theoretical results used in the comparison of the observations with the predictions of the standard cosmological models. These results are derived in some detail from first principles in my book *Galaxy Formation* (Longair, 2008).

23.2.1 Euclidean source counts

In a survey to a limiting flux density S for any class of extragalactic object, the *integral Euclidean source counts*, meaning the number of objects with flux densities greater than

flux density S , has the form $N(\geq S) \propto S^{-3/2}$, independent of the luminosity function $N(L)$ of the objects counted. In terms of apparent magnitudes, $m = \text{constant} - 2.5 \log_{10} S$, the Euclidean source counts become

$$N(\leq m) \propto 10^{0.6m} \quad \text{or} \quad \log N(\leq m) = 0.6m + \text{constant}. \quad (23.1)$$

The integral counts $N(\geq S)$ suffer from the disadvantage that the numbers of sources counted to different limiting flux densities are not independent since bright objects contribute to the counts at all lower flux densities. It is therefore statistically preferable to work in terms of *differential source counts* rather than integral counts, so that the numbers of sources counted in each flux density interval are independent. In this case,

$$dN(S) = N(S) dS \propto S^{-5/2} dS. \quad (23.2)$$

The corresponding expression in terms of apparent magnitudes is

$$dN(m) = N(m) dm \propto 10^{0.6m} dm. \quad (23.3)$$

These are useful reference relations and it is convenient to compare the observed counts and the expectations of various world models with them, as we illustrate below.

23.2.2 Number counts for the standard world models

For the standard world models, the Euclidean predictions are modified for the following reasons.

- The observed flux density now depends upon the spectrum of the source, because the radiation emitted at frequency ν_1 is observed at the redshifted frequency $\nu_0 = \nu_1/(1+z)$. For observations in the optical waveband, the differences between an inverse square law in luminosity distance and the predictions of the standard world models, when account is taken of the fact that a different region of the spectrum is observed, are often expressed in terms of K-corrections.
- In the standard world models, the distance measure D tends to a finite limit as $z \rightarrow \infty$. Unless the spectrum is strongly inverted, as is the case for dust spectra in the submillimetre waveband, the effects of observing the source spectrum at a redshifted frequency generally result in the sources becoming fainter with increasing redshift. To put it another way, the effects of the K-correction compensate for the fact that D tends to a finite value as $z \rightarrow \infty$.
- The volume element per unit redshift interval changes from $dN(z) \propto z^2 dz$ at small redshifts, $z \ll 1$, to $dN(z) \propto z^{-3/2} dz$ at redshifts $\Omega_0 z \gg 1$. The volume elements become smaller and smaller with increasing redshift, resulting in a ‘cut-off’ to the source distribution at redshifts $\Omega_0 z \gg 1$.

The differential number counts for different cosmological models for sources with power-law spectra $L(\nu) \propto \nu^{-1}$ are shown in Fig. 23.1. This choice of spectral index is a good approximation for the spectra of extragalactic radio sources, X-ray sources and quasars. An

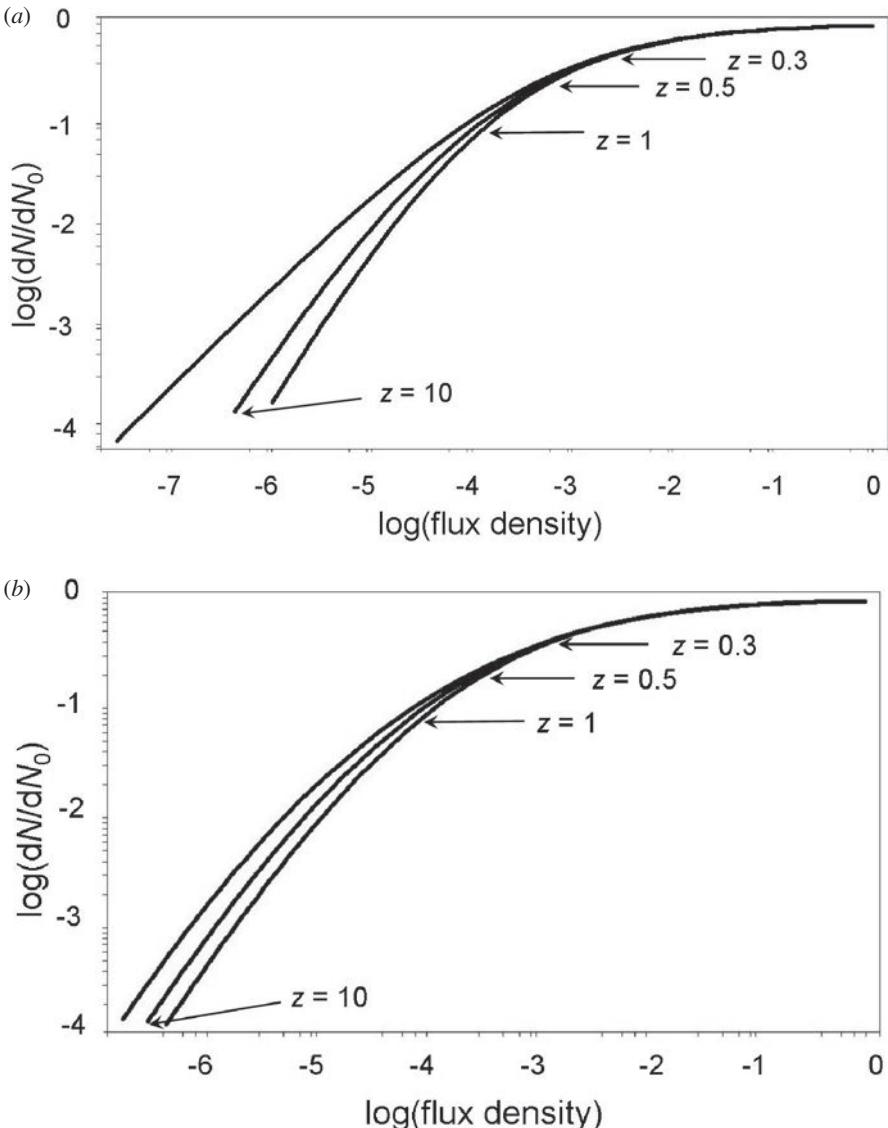


Fig. 23.1

(a) The predicted normalised differential source counts for a single luminosity class of source having spectral index $\alpha = 1$ for different values of the density parameter Ω_0 with $\Omega_\Lambda = 0$. From top to bottom, the differential counts are for models with $\Omega_0 = 0, 1$ and 2 (Longair, 1978). The arrows indicate the redshifts at which the sources are observed. The integration was terminated at $z = 10$. (b) The same as but for models with $\Omega_0 + \Omega_\Lambda = 1$ and spectral index $\alpha = 1$. From top to bottom, the differential counts are shown for models with $\Omega_0 = 0.3, 0.6$ and 1 .

advantage in using $\alpha = 1$ is that the results are also exact for *bolometric luminosities* and *flux densities*. These results are displayed as differential number counts relative to Euclidean prediction $dN(S) = N(S) dS \propto S^{-5/2} dS$. Figure 23.1a shows the normalised differential number counts for cosmological models with $\Omega_\Lambda = 0$ and Fig. 23.1b for flat cosmological models with $\Omega_0 + \Omega_\Lambda = 1$. The redshifts of objects at different flux densities are shown in

the figures. A number of important differences in comparison with the Euclidean number counts are immediately apparent.

The predicted differential counts depart rapidly from the Euclidean expectation even at relatively small redshifts. For example, for the case $\Omega_0 = 1$, the source counts at redshift $z = 0.5$ have differential slope -2.08 rather than -2.5 , corresponding to a slope of the integral source counts of -1.08 rather than -1.5 . The important point is that the effects of redshift set in at much smaller redshifts than might be expected. Similar results are found for models with finite values of Ω_Λ (Fig. 23.1b). The three models have $\Omega_0 + \Omega_\Lambda = 1$ and $\Omega_0 = 1, 0.6$ and 0.3 ; the world model with $\Omega_0 = 0.3$ corresponds closely to the concordance cosmological model.

In practice, the source populations cannot be represented by a single luminosity, but rather, the differential counts shown in Fig. 23.1 should be convolved with the luminosity function $N_0(L)$ of the sources. It is evident, however, that because all the relations shown in Fig. 23.1 are monotonically decreasing functions of decreasing flux density, convolution with *any* function must also produce a monotonically decreasing function of decreasing flux density. Thus, in all plausible world models, the slope of the differential and integral source counts must be smaller than the Euclidean predictions, that is, if $N(\geq S) \propto S^{-\beta}$, $\beta < 1.5$.

23.2.3 The V/V_{\max} or luminosity–volume test

A direct method of investigating the uniformity of a distribution of objects in space is to use the V/V_{\max} or *luminosity–volume test* (Schmidt, 1968; Rowan-Robinson, 1968). A sample of objects is selected which is known to be complete within well-defined flux density and apparent magnitude limits and for which complete redshift information is available. The test determines whether or not the distribution of objects is statistically uniform within the accessible volume of space defined by the observational selection criteria.

Suppose a sample of sources is *complete* in that all those having flux densities greater than some limiting value S_0 at a particular wavelength have been detected in some region of the sky. Assume also that redshifts, or distances, have been measured for all of them. Consider first sources of a single luminosity L . For each source, we can evaluate the quantity V/V_{\max} , where V is the volume of space enclosed by the redshift z of the source and V_{\max} is the volume of space within which the source could have been observed and still be included in the complete sample. The volume V_{\max} corresponds to the redshift z_{\max} at which a source of intrinsic luminosity L would have observed flux density S_0 . Thus,

$$\frac{V}{V_{\max}} = \frac{\int_0^r D^2 dr}{\int_0^{r_{\max}} D^2 dr}, \quad (23.4)$$

where D is the distance measure and r and r_{\max} are the comoving radial distance coordinates corresponding to z and z_{\max} , respectively. Notice that the volumes used in the test are comoving-coordinate volumes at the present epoch. Now suppose the distribution of sources

in space is uniform. The mean value of V/V_{\max} is then

$$\left\langle \frac{V}{V_{\max}} \right\rangle = \frac{\int_0^{r_0} \left(\frac{V}{V_{\max}} \right) D^2 dr}{\int_0^{r_0} D^2 dr}. \quad (23.5)$$

Setting

$$\int_0^r D^2 dr = X, \quad \frac{dX}{dr} = D^2, \quad (23.6)$$

and hence

$$\left\langle \frac{V}{V_{\max}} \right\rangle = \frac{\int_0^{r_0} X \left(\frac{dX}{dr} \right) dr}{X^2(r_0)} = \frac{1}{2}. \quad (23.7)$$

As is intuitively expected, for a uniform distribution of sources, they are observed on average half-way volume-wise to the limits of their observable volumes, independent of their luminosities.

Furthermore, if the source distribution is uniform, that is, if they have constant comoving number densities, the values of V/V_{\max} should be uniformly distributed between 0 and 1. The simplest statistical test for the significance of departures from the expected mean value of 0.5 can be developed as follows. For a uniform distribution of V/V_{\max} between 0 and 1, the standard deviation is $\sigma_0 = 1/\sqrt{12} = 0.288$. When the number of sources N is sufficiently large, the central limit theorem may be used and so the probability distribution of $\langle V/V_{\max} \rangle$ approaches a Gaussian distribution with standard deviation $N^{-1/2}$ of the original distribution, that is, $\sigma = \sigma_0/N^{1/2}$.

The V/V_{\max} procedure is a powerful technique for determining the uniformity and space density of any class of object. For example, in Felten's determination of the luminosity function of galaxies, he took account of the effects of extinction by interstellar dust in our Galaxy upon the luminosities and observable volumes within which galaxies of different intrinsic luminosities could be observed (Felten, 1977). In this case, the observable volume is not spherically symmetric about the observer – the key point is that the samples of sources studied should be statistically complete within well-defined observational selection criteria.

Generalising, the space density of any object can be written $N_i = 1/V_i$, where V_i is the volume of space within which the object could have been observed and still satisfy the selection criteria of the survey. By summing over all the objects in the complete sample, the luminosity function of the objects can be determined. The procedure can be further extended to consider the distribution of objects within a particular redshift interval, what is referred to as a *banded* V/V_{\max} test. This approach is useful in studying, for example, whether or not there is a cut-off in the distribution of quasars at large redshifts (Waddington *et al.*, 2001). These calculations can be carried out for all objects in the sample within given redshift limits and so luminosity functions as a function of cosmic epoch determined (Fig. 18.4).

23.3 The evolution of non-thermal sources with cosmic epoch

23.3.1 Number counts and V/V_{\max} tests for extragalactic radio sources

A compilation of counts of radio sources at a wide range of frequencies throughout the radio waveband is shown in Fig. 23.2 (Wall, 1996). The cause of the controversies in the 1950s and early 1960s was the steepness of the radio source counts at high flux densities. These early surveys were carried out at low radio frequencies, the counts at metre wavelengths (0.178 GHz) being the first to show the ‘inverted’ behaviour with $N(\geq S) \propto S^{-1.8}$ at high flux densities. The radio spectra of these sources have mean spectral index $\alpha \approx 0.8$ and a comparison of the high flux density behaviour of the counts with the corresponding theoretical predictions in Fig. 23.1 indicates that there must be many more faint sources than are predicted by all uniform world models. By the mid-1960s, it was known that the majority of these sources were at large redshifts and this led to the realisation that the source population had to evolve very strongly with cosmic epoch (Davidson and Davies, 1964; Longair, 1966).

When redshift data for complete samples of the radio quasars in the 3CR samples became available, Schmidt and Rowan-Robinson showed that these sources were concentrated towards the limits of their observable volumes (Rowan-Robinson, 1968; Schmidt, 1968) (Fig. 23.3). The values of $\langle V/V_{\max} \rangle$ for the quasars was found to be 0.686 ± 0.042 , significantly greater than the uniform expectation of 0.5. The radio galaxies in the 3CR sample were much more difficult to identify but by the 1990s they too were all identified and their redshifts measured. Their V/V_{\max} distribution and $\langle V/V_{\max} \rangle = 0.697 \pm 0.031$ were exactly the same as those of the radio quasars (Fig. 23.3). These analyses showed that the sources in the high flux density samples were much more common at large redshifts than nearby.

Radio source counts have now been determined over the entire radio waveband from low radio frequencies to short centimetre wavelengths (Fig. 23.2). They all show the same overall features – a steep source count at high flux densities, a plateau at intermediate flux densities and convergence at low flux densities. At the very lowest flux densities, $S \leq 10^{-3}$ Jy, the source counts flatten again.

In interpreting the counts of sources, the objective is to determine how the luminosity function of the sources $N(L)$ has changed with cosmic epoch. This is a time-consuming task since high resolution radio maps are needed in order to make optical identifications and obtain spectra of faint objects which typically have redshifts $z \geq 1$. It is convenient to consider separately the radio sources with steep ($\alpha \sim 0.8$) and flat ($\alpha \sim 0$) radio spectra. The former are the extended double radio sources which are the dominant population at low radio frequencies, $\nu < 1$ GHz; the latter are the much more common compact radio sources found in high frequency samples, $\nu > 1$ GHz. Two examples of the types of model which were consistent with all the data available to Dunlop and Peacock in the 1990s are shown in Fig. 23.4 (Dunlop and Peacock, 1990). Both the steep and flat spectrum radio sources show the same forms of evolutionary behaviour.

In the mid-1990s, the changes in the form of the radio luminosity function out to redshifts $z \approx 2$ could be conveniently described by what is termed *luminosity evolution*.

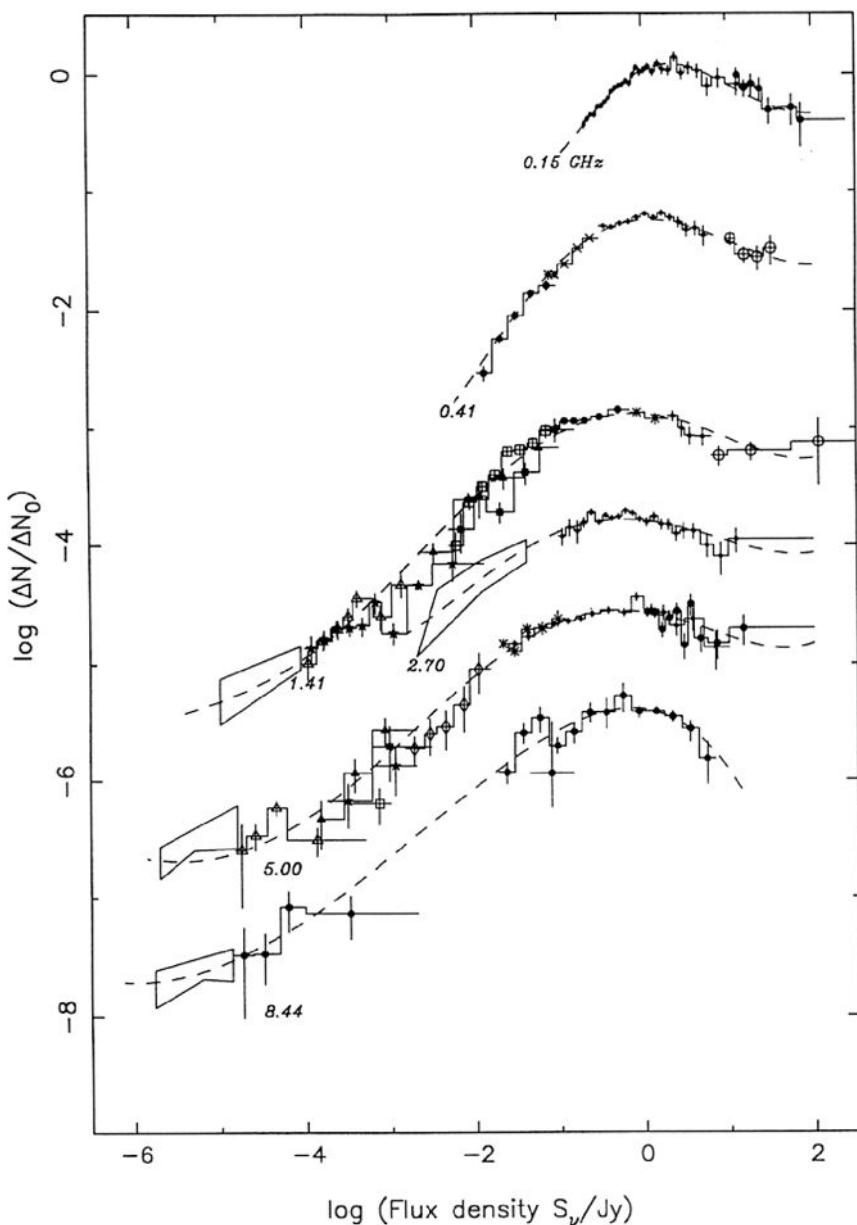


Fig. 23.2 The differential, normalised counts of extragalactic radio sources at a wide range of frequencies throughout the radio waveband. This compilation was kindly provided by Dr. Jasper Wall (Wall, 1996). The points show the number counts derived from surveys of complete samples of radio sources. The boxes indicate extrapolations of the source counts to very low flux densities using the $P(D)$ technique.

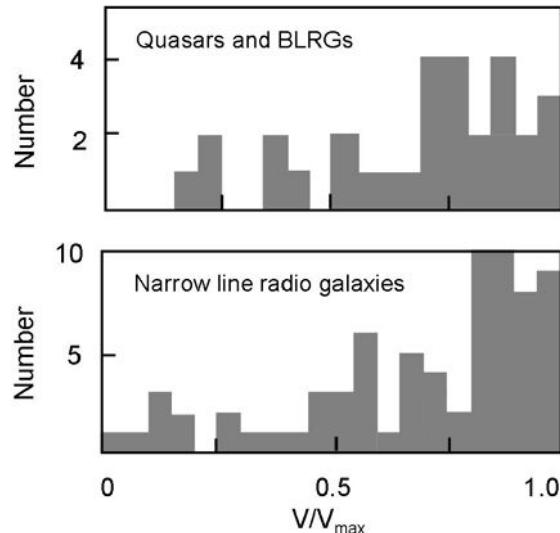


Fig. 23.3 The V/V_{\max} distributions for (a) quasars and broad-line radio galaxies and (b) narrow-line radio galaxies in a complete sample of extragalactic radio sources selected from the 3CR complete sample (Laing *et al.*, 1983). Redshifts have been measured for all the objects in the sample, which is complete to a limiting radio flux density of 9.7 Jy at 178 MHz (Longair, 1997a).

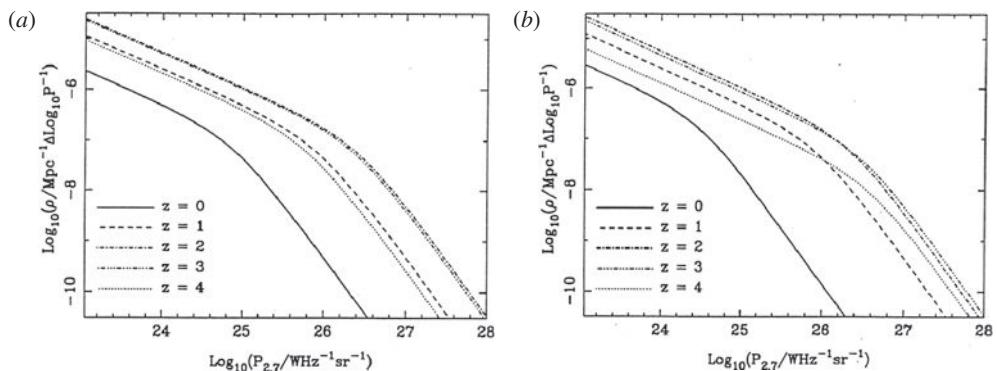


Fig. 23.4 Two examples of the forms of evolving radio luminosity function which can account for the radio source counts and redshift and identification data available in the mid-1990s (Dunlop, 1998). In model (a), the change in the radio luminosity function is described by luminosity evolution. In model (b), the changes in the radio luminosity function involve in addition negative density evolution at large redshifts. In both cases, the changes in the form of the radio luminosity function are described in terms of the number densities of sources per unit comoving volume.

In this form of evolution, the radio luminosity function of the powerful radio sources is shifted to greater radio luminosities with increasing redshift, whilst the normal galaxy radio luminosity function remains unchanged. These changes could be described by an increase in the radio luminosities of the strong radio sources with redshift as $L(z) = L_0(1+z)^3$ out to redshifts $z \approx 2$. Inspection of Fig. 23.4a shows that, although the shift in luminosity

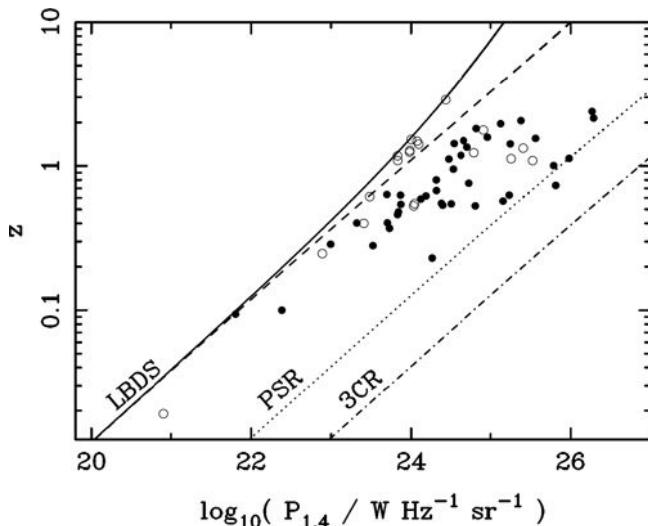


Fig. 23.5 The luminosity–redshift plane for radio sources with $S_{1.4} \geq 2 \text{ mJy}$ in the Leiden–Berkeley Deep Survey Hercules field. The flux density limits for the survey are shown for flat-spectrum (solid line, open circles) and steep-spectrum (dashed line, solid circles) sources, together with the limits for samples selected at higher flux densities, the PSR (dotted line) and 3CRR sample (dot-dash line) samples. Photometric redshifts have been adopted for those sources without spectroscopic redshifts (Waddington *et al.*, 2001).

amounts to a factor of 27 by a redshift of 2, the number density of sources of a given luminosity increases by a much greater factor. In the functions shown in Figure 23.4, the comoving number density of sources of luminosity $10^{27} \text{ W Hz}^{-1} \text{ sr}^{-1}$ increases by a factor of about 1000, indicating how strongly the luminosity function of extragalactic radio sources has evolved with cosmic epoch. Other parameterisations of the evolution of the luminosity function can account for the data equally well. Figure 23.4b shows what is referred to as luminosity-dependent density evolution which also provides a satisfactory fit to the data.

The dramatic changes in the luminosity function cannot continue to redshifts $z \gg 2$. For example, in Fig. 23.4, the luminosity functions at $z = 4$ have declined from the maximum at $z \approx 2\text{--}3$. If this decline at large redshifts did not take place, the observed convergence of the counts would not be reproduced and the radio background emission at low radio frequencies would be exceeded (Longair, 1995).

To address this issue, complete samples of radio sources are needed at low radio flux densities (Dunlop, 1998). The Leiden–Berkeley Deep Survey (LBDS) extended to flux densities about a factor of about 1000 fainter than the 3CRR sample. The radio luminosity–redshift diagram shown in Fig. 23.5 shows the flux density limits for the 3CRR and LBDS surveys, as well as the intermediate Parkes Selected Region (PSR) sample. The flux density limit for the LBDS survey is shown for flat-spectrum sources by a solid line and the flat spectrum sources in the survey by open circles; the steep-spectrum source limits are shown by a dashed line and the steep spectrum sources by solid circles (Waddington *et al.*, 2001). There is a rather abrupt cut-off to the source distribution for both the flat and steep spectrum sources at $z \sim 2$. These results show that the radio source and associated black hole activity

was at its greatest at redshifts $z \sim 2$ and that this activity decreased rather dramatically at earlier and later cosmological epochs. Waddington and his colleagues also showed that pure luminosity evolution was not a particularly good fit to the data when it was extended to low radio luminosities. Specifically, they found that there is a luminosity dependence of the high redshift cut-off – the lower luminosity sources with $P_{1.4\text{GHz}} \sim 10^{24} \text{ W Hz}^{-1} \text{ sr}^{-1}$ began to decline in comoving number density at redshifts $z \approx 1-1.5$, while higher luminosity sources with $P_{1.4\text{GHz}} \sim 10^{25}-10^{26} \text{ W Hz}^{-1} \text{ sr}^{-1}$ were in decline in comoving number density beyond $z \approx 2-2.5$.

23.3.2 Radio-quiet quasars

The radio-quiet quasars are about 50–100 times more common than the radio-loud variety (Sandage, 1965). The early surveys established that they also have a steep source count and a value of $\langle V/V_{\max} \rangle$ much greater than 0.5 (Bracegirdle *et al.*, 1970; Schmidt and Green, 1983). Subsequent studies have shown that the optically selected quasars exhibit evolutionary behaviour over cosmological time-scales similar to that established for the radio-loud quasars.

Number counts for radio-quiet quasars discovered in the Sloan Digital Sky Survey display a steep number count which converges at faint apparent magnitudes (Fig. 18.3). The changes of the optical luminosity function of optically selected quasars with cosmic epoch was derived by Boyle and his colleagues out to redshift $z \sim 2.3$ from a complete sample of 6000 quasars from the AAT 2dF Quasar Redshift survey (Fig. 18.4) (Boyle *et al.*, 2000). At zero redshift, the luminosity function of the radio-quiet quasar population joins smoothly on to that of the Seyfert galaxies, a natural continuity of the properties of these classes of active galactic nuclei. This is consistent with the procedure of Schmidt and Green who classified only 92 of their 114 large redshift ultraviolet excess objects as quasars since they had absolute magnitudes brighter than $M_B = -23$. The 22 lower luminosity objects were either Seyfert 1 nuclei or what they termed low-luminosity quasars.

With increasing redshift, many more quasars of a given luminosity are observed per unit comoving volume than expected in uniform world models, as is dramatically illustrated in Fig. 18.4. Over the redshift range $0 \leq z \leq 2.3$, the changes in the luminosity function can be described by ‘luminosity evolution’, of almost exactly the same form as that needed to account for the evolution of the radio luminosity function for luminous extragalactic radio sources, namely, $L(z) \propto (1+z)^\beta$ out to $z = 2$ with $\beta = 3.5$.

Systematic surveys have been undertaken to determine the large redshift evolution of the radio-quiet quasar population by a number of authors (Schmidt *et al.*, 1995; Fan *et al.*, 2001, 2004) and these have provided evidence for a decrease in the comoving space density of quasars at redshifts greater than 2–3. In the Sloan Digital Sky Survey observations, a five colour selection technique has enabled quasars up to $z = 5$ to be studied. There are strong redshift-dependent selection effects at $z \geq 2.3$ which need to be understood in quantitative detail before the evolving luminosity function can be derived (Richards *et al.*, 2006). Richards and his colleagues find that the number counts and luminosity functions agree well with the results of the Two Degree Field QSO Redshift Survey (2QZ) at redshifts where these datasets overlap in both redshift and luminosity. At redshifts greater than 2,

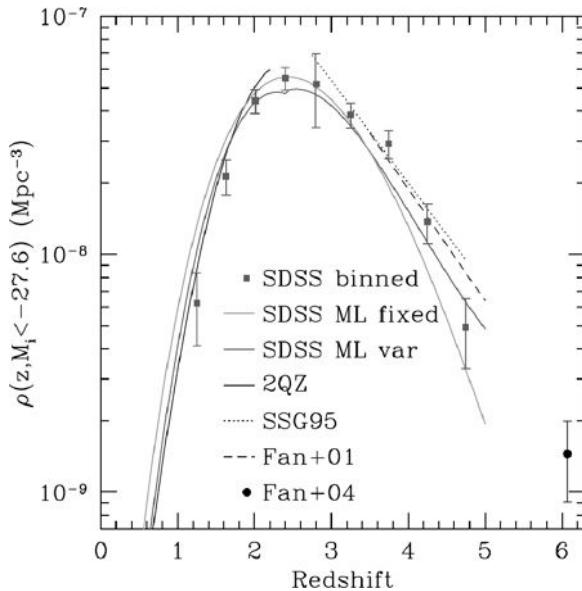
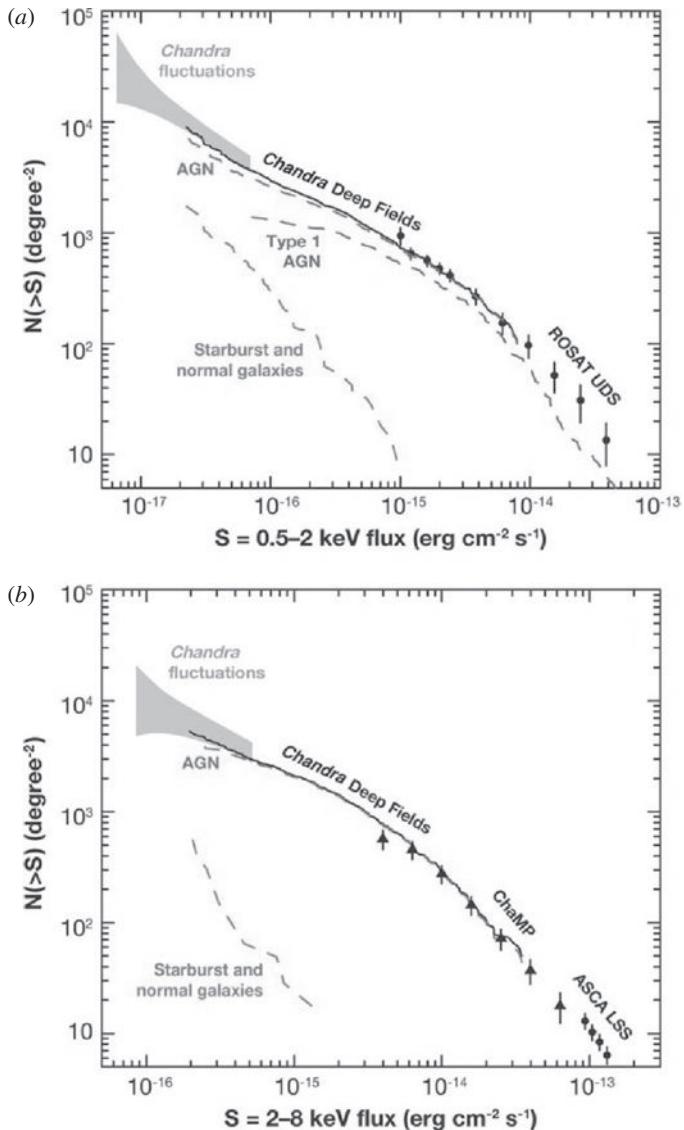


Fig. 23.6 The integrated i -band luminosity function for quasars more luminous than $M_i = -27.6$ (Richards *et al.*, 2006). The solid black line terminating at $z \approx 2$ is from the 2dF Quasar Redshift survey (Boyle *et al.*, 2000). The dashed and dotted lines are from Fan and Schmidt and their colleagues (Fan *et al.*, 2001; Schmidt *et al.*, 1995), respectively. The point at $z \sim 6$ is the comoving space density estimated by Fan and his colleagues (Fan *et al.*, 2004).

the comoving number density of luminous quasars peaks between redshifts 2 and 3 and then converges rapidly at larger redshifts, in agreement with earlier studies (Fig. 23.6). In addition, the slope of the luminosity function of quasars flattens at redshifts $z \geq 2.4$. They also included the results of Fan and his colleagues, who used the i -band drop-out technique to discover quasars with redshifts greater than $z = 5.7$. These observations showed that the decline in the comoving space density of luminous quasars continues out to redshift $z = 6$, as illustrated in Fig. 23.6.

23.3.3 X-ray source counts

The history of the number counts of X-ray sources has been surveyed by Brandt and Hasinger (2005). Following the bright sources surveys carried out by the UHURU, Ariel-V and HEAO-1 satellite observatories in the 1970s, the German-US-UK ROSAT mission surveyed the whole sky in the X-ray energy band 0.1–2.4 keV to a flux density limit about 100 times fainter than that of the HEAO-1 survey (Hasinger *et al.*, 1993). The next major advances were made by the Chandra and XMM-Newton X-ray Observatories. Deep and wide-field surveys with these telescopes enabled number counts of X-ray sources to be determined to faint X-ray flux densities throughout the 1–10 keV wavebands (Fig. 23.7). The counts have been extended to even lower flux densities using the statistical $P(D)$ technique. Most recently, more detailed number counts, particularly for the hard X-ray energies, $\varepsilon \sim 8$ keV, were described by Cappelluti and his colleagues (Cappelluti *et al.*, 2007).

**Fig. 23.7**

(a) Integral number counts of X-ray sources in the 0.52 keV band. The data points are derived from the ROSAT Lockman Hole field, the solid black curve from the Chandra Deep Fields and the grey areas from $P(D)$ analysis of the Chandra Deep Field-North. The dashed curves show number counts for all active galactic nuclei, Type 1 active galactic nuclei and starburst and normal galaxies. (b) Integral number counts of X-ray sources in the 2–8 keV band. The black circles are from the ASCA Large Sky Survey, the black triangles from the ChaMP study, the solid black curve from the Chandra Deep Fields study and the grey area from a $P(D)$ analysis of the Chandra Deep Field-North. Details of the sources of these data are given in the review by Brandt and Hasinger (2005).

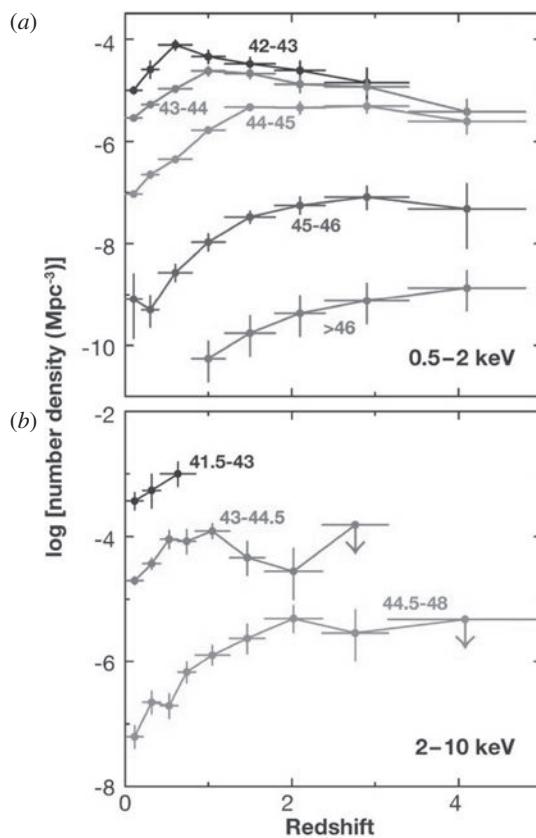
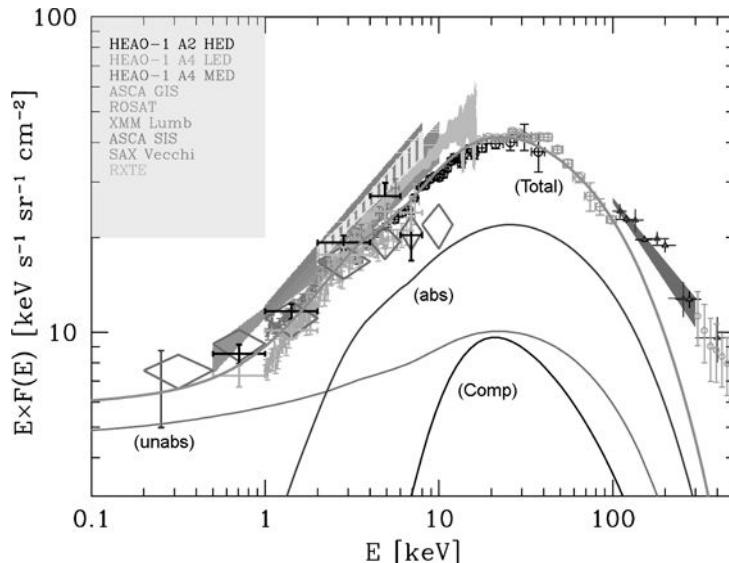


Fig. 23.8 (a) The comoving space density of X-ray selected active galactic nuclei in the $0.5\text{--}2 \text{ keV}$ energy band as a function of redshift. Results are shown for five luminosity ranges which are labelled by the logarithm of the X-ray luminosity in erg s^{-1} . (b) The same relation for X-ray active galactic nuclei selected in the $2\text{--}10 \text{ keV}$ energy band for three luminosity ranges (Brandt and Hasinger, 2005).

The interpretation of the X-ray source counts in the $0.5\text{--}10 \text{ keV}$ waveband is complicated by the fact that the X-ray spectra of sources selected at, say, 1 and 10 keV can be quite different. Figures 23.7a and b show that the principal contributors to the number counts at energies $\varepsilon \sim 1 \text{ keV}$ to the limits of the Chandra Deep Surveys are active galactic nuclei. Major efforts have been made to identify and obtain optical spectra for the X-ray sources in the Chandra surveys resulting in estimates of the evolution of the comoving space density of sources of different X-ray luminosities in the $0.5\text{--}2 \text{ keV}$ and $2\text{--}10 \text{ keV}$ wavebands (Fig. 23.8a and b). These both show the characteristic feature of the evolution of the population of active galactic nuclei that the most luminous sources show the strongest cosmological evolution out to redshifts $z \approx 2$ while the less luminous sources show much less dramatic evolution.

What the functions displayed in Fig. 23.8 disguise is the fact that the X-ray spectral types change quite dramatically with X-ray energy. This is most easily illustrated by considering the types of source which make up the total X-ray background intensity from soft ($\sim 1 \text{ keV}$)

**Fig. 23.9**

The spectrum of the cosmic X-ray background emission compared with the predicted contributions from populations of Compton-thin and Compton-thick active galactic nuclei. The estimates of the X-ray background intensity spectrum are discussed by Gilli and his colleagues, the sources of the data being listed in the top left of the diagram. Solid lines show the contributions of different types of active galactic nuclei according to the models of Gilli and his colleagues. These are: unobscured active galactic nuclei (unabs), obscured Compton-thin AGN (abs), Compton-thick active galactic nuclei (Comp) and the sum of these three components, including the contribution of clusters of galaxies (Total). The diamonds (and some of the other data points) show estimates of the background intensity by counting sources to the limits of the deep surveys (Gilli *et al.*, 2007).

and hard (~ 10 keV) X-ray energies. The spectrum of the X- and γ -ray background from 0.1 to 300 keV is shown in Fig. 23.9 from a multitude of experiments described by Gilli and his colleagues (Gilli *et al.*, 2007).

It is straightforward to determine how much of the X-ray background intensity in the 0.5–2 keV energy band can be attributed to discrete X-ray sources by integrating the number counts shown in Fig. 23.8a. There is general agreement that essentially all the background intensity in this waveband is associated with discrete sources and that this population is dominated by Type 1 active galactic nuclei in which the characteristic broad lines and continuum are observed. This means that the majority contributors to the X-ray background at these energies are unobscured sources in which the nuclear regions of the active galaxies are observed directly. If attention is restricted only to the main contributors in this soft waveband, the typical photon number spectral index is about 1.9 with a dispersion of about 0.2–0.3 about this mean value. This immediately poses a problem since the spectral index of the background spectrum is significantly smaller than this value in the 1–10 keV waveband. This problem can be understood by comparing the observed background spectrum with the predicted spectrum due to unabsorbed sources which are labelled (unabs) in Fig. 23.9. Therefore, the same population cannot explain the X-ray background intensity at harder X-ray energies.

Carrying out the same analysis at hard X-ray energies ($\varepsilon \sim 10$ keV), at least 50% and probably more of the background is associated with the discrete sources already detected, but their spectral mix is somewhat different. The most important difference is that there is a major population of sources which are strongly absorbed at soft X-ray energies. Sources have now been observed in which photoelectric absorption strongly suppresses the soft X-ray flux density and so ‘hardens’ the X-ray spectra. The inferred column densities can be up to 10^{24} cm $^{-2}$ and greater. For column depths greater than 10^{24} cm $^{-2}$, the effects of Comptonisation need to be taken into account and this process causes major distortions of the X-ray source spectra whatever the initial input photon spectrum. Gilli and his colleagues provide template spectra for this wide range of source which also include the effects of iron lines and reflected X-ray components from the inner regions of the active galactic nucleus (Gilli *et al.*, 2007).

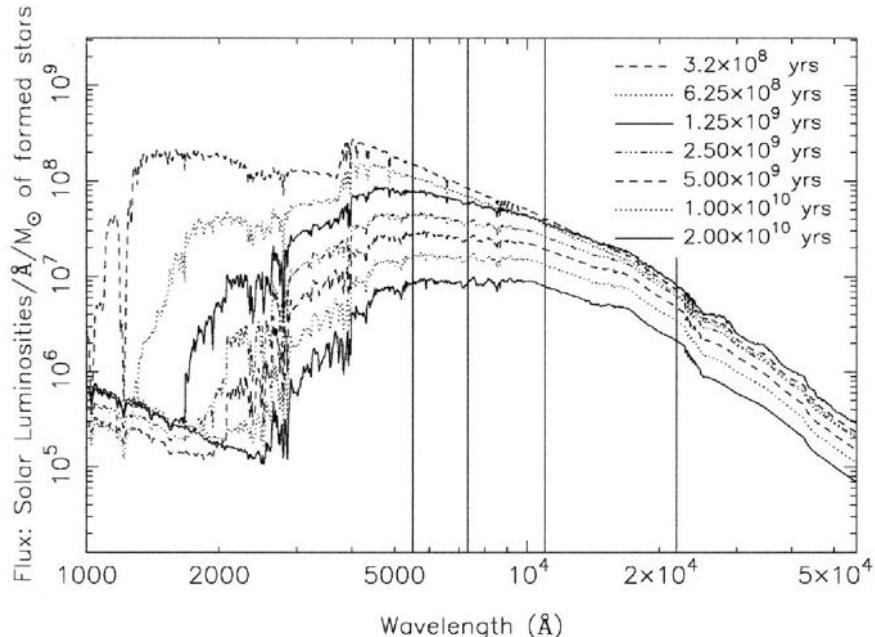
At energies greater than 10 keV, the number count data are much sparser, but the background spectrum continues to be remarkably flat up to about 30 keV, beyond which it changes slope rather dramatically and continues more or less as a power-law to γ -ray energies. It is assumed that Comptonised sources make a significant contribution to the background spectrum above 10 keV, as shown in Fig. 23.9, which also shows best estimates of the types of source which make up the observed X-ray background spectrum. Whilst the solid lines are models for the various contributions to the background, they are also consistent with the most recent data on the counts of X-ray sources in the 0.5–10 keV waveband reported by Cappelluti and his colleagues (2007). Notice also that, even taking account of the different types of known source, there must still be an additional component at γ -ray energies $\varepsilon \geq 100$ keV to account for the observed γ -ray background spectrum. These might be further Compton-thick hard X-ray sources.

23.4 The evolution of thermal sources with cosmic epoch

In this section, the term ‘thermal sources’ is taken to mean the emission of galaxies in the optical, infrared and submillimetre wavebands. More specifically, in the ultraviolet, optical and infrared wavebands, the radiation of galaxies is predominantly starlight, while in the mid- and far-infrared and submillimetre wavebands it is the emission of heated dust.

23.4.1 The predicted number counts for galaxies

For galaxies observed in the optical waveband, the predicted number counts depend upon knowledge of their spectra in the ultraviolet and optical wavebands and this is sensitive to their star-formation histories. As an example, Fig. 23.10 shows the results of computations of the evolution with cosmic epoch of the spectral energy distribution of a model giant elliptical galaxy in its rest frame. The model involves the *passive evolution* of the stellar energy distribution in the sense that the stellar population was formed in an initial starburst of duration 10^8 years and the subsequent evolution is simply the aging of that population. The stellar synthesis codes of Bruzual and Charlot (2003) were used in deriving Fig. 23.10.

**Fig. 23.10**

Illustrating the spectral evolution of a model giant elliptical galaxy in which the stellar population was formed in an initial starburst of duration 10^8 years. The computations used the stellar synthesis codes for the spectral energy distribution of galaxies developed by Bruzual and Charlot (2003). The predicted galaxy spectra are shown in equal logarithmic intervals of cosmic time up to the present day. The initial stellar population was chosen so that the observed spectrum of a giant elliptical galaxy at the present epoch was reproduced (Inskip *et al.*, 2002). The vertical lines show the rest wavelengths which would be observed at redshifts $z = 0, 1, 2$ and 3 from right to left, respectively, if the galaxy were observed at a wavelength of $2.2\text{ }\mu\text{m}$.

There is a one-to-one relation between the mass of a star, its luminosity and the waveband in which it emits most of this luminosity and so stars radiating in the ultraviolet waveband are massive and have short lifetimes. This is reflected in the rapid decline of the spectral energy distribution in the ultraviolet regions of the spectrum with cosmic epoch. Consequently, the K-corrections to be used in calculating the number counts of galaxies in the optical waveband are very sensitive to bursts of star-formation activity.

The situation is somewhat simpler if the observations are carried out in the infrared waveband (Inskip *et al.*, 2002). If observations are made at $2\text{ }\mu\text{m}$, for example, the corrections for the passive evolution of the stellar populations are much more stable and systematic as compared with the optical waveband since even at $z = 3$, the spectral energy distributions are much less influenced by star-formation events. The reasons for this can be appreciated from the following argument. Most of the infrared luminosity at wavelengths $\lambda > 1\text{ }\mu\text{m}$ originates from stars on the red giant branch and their luminosities and colours are essentially independent of the masses of their main sequence progenitors. The lifetimes of red giant stars are much less than the time their progenitors spend on the main sequence, and so the number of red giant stars is given by the product of the rate at which stars evolve off

the main sequence onto the giant branch and the time they spend there. In fact, we need only know the rate of evolution of stars off the main sequence for quite a narrow range of masses.

The mass-luminosity relation for stars on the main sequence is assumed to take the form $L/L_\odot = (M/M_\odot)^x$. Their main sequence lifetimes are determined by the time it takes the star to burn about 10% of its mass into helium, the *Schönberg-Chandrasekhar limit* (Sects 2.7 and 2.8). During evolution on the main sequence, the luminosities of stars of a given mass are remarkably constant (Kippenhahn and Weigert, 1990; Tayler, 1994). Since the available fuel is proportional to the mass of the star, the main sequence lifetime is $t = t_\odot(M/M_\odot)^{-(x-1)}$, where t_\odot is the main sequence lifetime of the Sun, which is about 10^{10} years. For stars with $M \sim M_\odot$, $x \sim 5$ and so the lifetimes of stars of mass $M = 2M_\odot$ is only 1/16 of the age of the Sun, that is, about 6×10^8 years. Thus, the range of masses of stars which contribute most of the light of the old stellar populations of galaxies lie in the range $1 < M_\odot < 2$ over the range of redshifts of interest.

Since the time stars spend on the giant branch, t_g , and their luminosities are relatively independent of their main sequence masses, we need only determine the rate at which stars evolve off the main sequence onto the giant branch as a function of main sequence mass (Gunn, 1978). It is assumed that all the stars were formed in an initial brief starburst and that the initial mass function of the stars was of Salpeter form, $dN = N(M) dM \propto M^{-y} dM$, where $y = 2.35$. Then, the number of stars on the giant branch N_g is

$$N_g = t_g \frac{dN}{dt} = t_g \left(\frac{dN}{dM} \right) \left(\frac{dM}{dt} \right). \quad (23.8)$$

Using the above relation between mass and main sequence lifetime,

$$L(t) = L(t_0) t^{-(x-y)/(x-1)}. \quad (23.9)$$

Inserting the values $x = 5$ and $y = 2.35$, we find $L \propto t^{-0.66}$. For the case of the critical world model, $t/t_0 = (1+z)^{-3/2}$ and so, to a good approximation, $L \propto (1+z)$. Thus, at a redshift of 1, the old stellar populations of galaxies should be about twice as luminous as they are at the present epoch, and at redshift $z = 3$, four times as luminous. This accounts for the rather systematic behaviour of the decrease in infrared luminosity of galaxies seen in Fig. 17.3. Typically, if the galaxies evolved passively, they would only be about a magnitude more luminous at redshifts $z \sim 1-2$ than they are at the present epoch.

23.4.2 Counts of galaxies

Ellis (1997) has described the various complications in determining reliable galaxy counts and of their interpretation. First of all, galaxies are extended objects, often with complex brightness distributions and care must be taken to ensure that the same types of object are compared at different magnitude limits and redshifts. Secondly, unlike the radio and X-ray sources, which are uniformly distributed over the sky, the distribution of galaxies is far from uniform on scales less than about $50 h^{-1}$ Mpc, as illustrated by the large voids and walls in the local distribution of galaxies (Fig. 4.2). Even at the faintest magnitudes, this ‘cellular’ structure in the distribution of galaxies results in fluctuations in the number counts

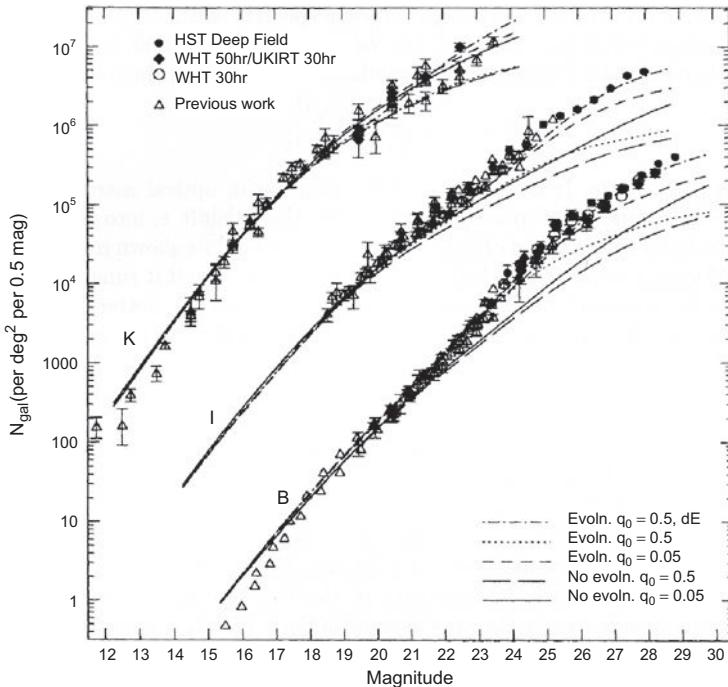


Fig. 23.11

The counts of faint galaxies observed in the B , I and K wavebands compared with the expectations of various uniform world models, as well as other models in which various forms of the evolution of the luminosity function of galaxies with redshift are assumed (Metcalfe *et al.*, 1996). The galaxy counts follow closely the expectations of uniform world models at magnitudes less than about 22, but there is an excess of galaxies in the B and I wavebands at fainter magnitudes.

of galaxies which exceed the statistical fluctuations expected in a random distribution. Thirdly, Fig. 3.8 shows that the probability of finding galaxies of different morphological types or sequences depends upon the environment in which the galaxy is located.

The counts of galaxies presented by Metcalfe and his colleagues (1996) provide a good impression of the overall counts of galaxies in the B (440 nm), I (800 nm) and K (2.2 μm) wavebands (Fig. 23.11). These were based upon a number of separate determinations by ground-based optical and infrared telescopes, as well as deep number counts in the Hubble Deep Field. The number counts of galaxies in the Hubble Deep Field join smoothly onto the ground-based counts. Furthermore, the Hubble Deep Field observations have enabled counts to be made in the I waveband, in which the background emission from the Earth's atmosphere becomes an increasingly important problem, and so to extend the counts by about a factor of 100 fainter than is possible from the ground. The lines labelled 'No evoln.' show the expectations of uniform world models and include appropriate K-corrections for the types of galaxy observed in bright galaxy samples.

These studies have been extended to even fainter magnitudes through Hubble Space Telescope observations of the Hubble Ultra-Deep Field (HUDF), which made use of the wider field capabilities of the Advanced Camera for Surveys (ACS) (Beckwith *et al.*, 2006).

This project resulted in the deepest image ever taken of the sky and involved 400 orbits of HST observing time, or about a million seconds of exposure time (Fig. 23.12a). The integral number counts of galaxies were complete to magnitudes $z_{850} = 28.7$ and $i_{775} = 29.2$. These counts are compared with those derived from the Great Observatories Origins Deep Survey (GOODS) which were complete to $z_{850} = 26.5$ in Fig. 23.12b. Interestingly, the number counts in the HUDF are 10% lower than those in the GOODS field, reflecting the influence of the large scale holes and walls of galaxies on the large scale.

A convenient recent survey of a large number of counts of galaxies has been presented by Ellis and Bland-Hawthorn (2006). Their Appendix A provides a comprehensive list of galaxy counts in the U , B , R , I and K wavebands. Interestingly, their objective was not in the use of galaxy counts for cosmological purposes, but rather to provide reliable number counts and their standard deviations for use in projects which required the reliable subtraction of the numbers of background galaxies. Their paper is particularly valuable in providing a standardised set of number counts for these wavebands and also for providing estimates of the standard deviations in these numbers, taking account of both the statistical uncertainties and the variations due to the fact that the large scale distribution of galaxies is highly correlated. As a result, the fluctuations in the number counts depend both upon the depth of the survey and angular scale over which the counts are made. For illustration, their number counts for observations made in a 15×15 arcmin 2 field are given in Table 23.1 and shown in Fig. 23.13. Because of the clustering tendencies of galaxies, the standard deviations are generally much greater than a $N^{1/2}$ estimate.

A number of general features of the galaxy counts are apparent from Figs 23.11 and 23.13. In the infrared K -waveband ($2.2 \mu\text{m}$), the counts follow reasonably closely the expectations of uniform world models with $q_0 \sim 0\text{--}0.5$. This is perhaps not too unexpected since the old stellar populations of galaxies are the principal contributors to the luminosities of galaxies in these wavebands. Even corrections for the passive evolution of the galaxy population are quite small, as discussed in Sect. 23.4.1. In contrast, in the R , B and U wavebands, there is a large excess of faint galaxies, particularly in the shortest wavelength bands. The departure from the expectations of the uniform models sets in at about $B = 23$ and, at fainter magnitudes, there is a large excess of faint blue galaxies. The lines in Fig. 2.12 illustrate the results of various modelling exercises to account for the observed counts.

The nature of the excess of blue galaxies was soon elucidated by studies with the Hubble Space Telescope. The high resolution images enabled the morphologies of galaxies to be classified into spheroidal/compact, spiral and irregular/peculiar/merger categories. The results of the pioneering analysis of Abraham and his colleagues (1996) are shown in Fig. 23.14. It is apparent that the spheroidal and spiral galaxies more or less follow the expectations of the uniform world models, while the objects classified as irregular/peculiar/merger systems show a distinct excess relative to their populations in bright galaxy samples, which amount to only about 1–2% of the galaxy population.

Part of the problem in understanding the nature of the faint blue galaxies is that, even with the present generation of 8–10 metre optical-infrared telescopes, spectroscopic observations are limited to $m_B \leq 24$. According to Cowie and his colleagues (1996), at this magnitude limit, the population of galaxies is composed of a mixture of normal galaxies at small redshifts plus galaxies undergoing rapid star formation from $z = 0.2$ to beyond $z = 1.7$.

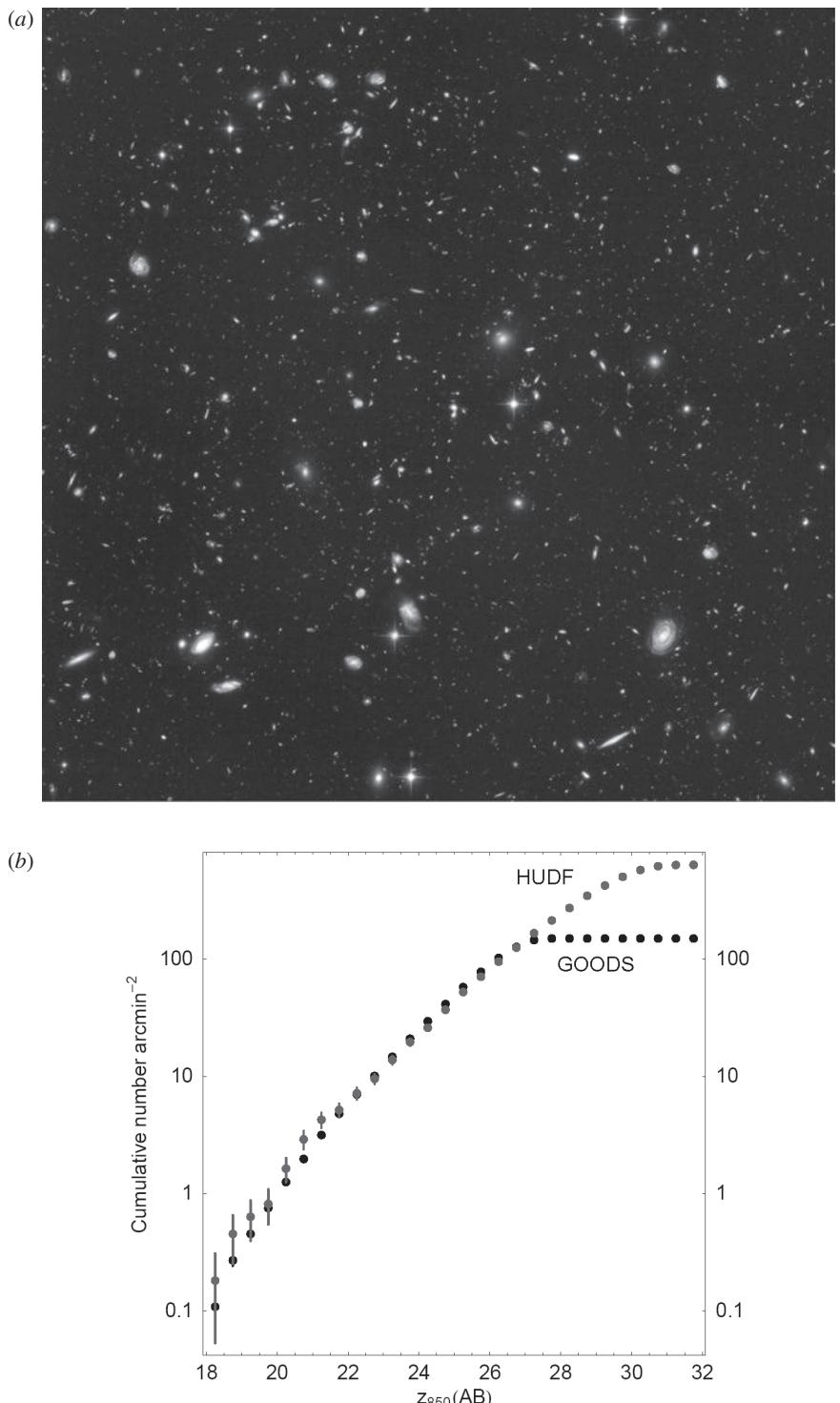


Fig. 23.12 (a) The Hubble Ultra-Deep Field (Beckwith *et al.*, 2006). (b) The integral counts of galaxies in the Hubble Ultra-Deep Field compared with the number counts in the GOODS survey. The completeness limit of the GOODS survey is $z_{850} = 26.5$, whereas that of the HUDF is about $z_{850} = 30$.

Table 23.1 Galaxy counts N and their standard deviations σ for an area of sky $15 \times 15 \text{ arcmin}^2$ in the U, B, R, I and K wavebands. The standard deviation depends upon the area of sky observed because of the strongly correlated distribution of galaxies (Ellis and Bland-Hawthorn, 2006). The numbers N quoted in the table are those between magnitudes m_{\min} and m_{\lim} .

m_{\lim}	m_{\min}	$N(\text{U})$ 19	σ_{U}	$N(\text{B})$ 15	σ_{B}	$N(\text{R})$ 15	σ_{R}	$N(\text{I})$ 14	σ_{I}	$N(\text{K})$ 13	σ_{K}
15										11	5
16								2	2		
17				0	1	4	2			144	28
18								32	11		
19				6	4	40	10			993	87
20								248	37		
21		66	19	50	14	295	36			4890	230
22								1410	100		
23		688	76	408	47	1650	110			18200	510
24								6410	270		
25		6020	310	4040	240	8930	360				
26								28710	790		
27		27210	780	20940	710	39390	1000				
28						74400	1600	101500	1900		
29				63400	1400						

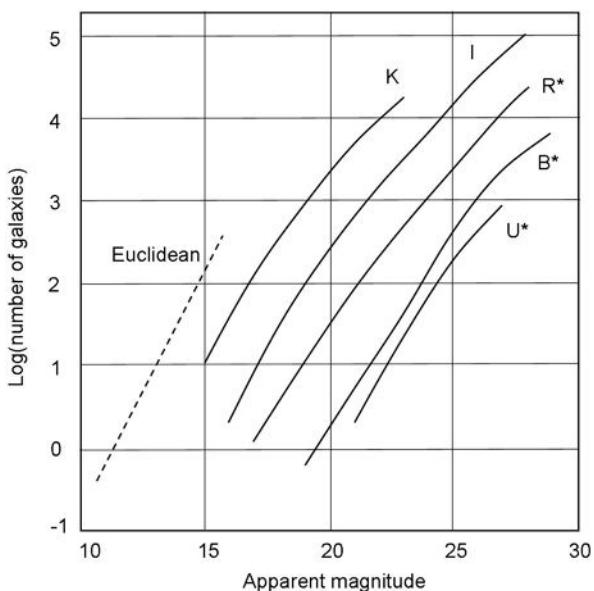


Fig. 23.13 The number counts of galaxies in the U, B, R, I and K wavebands from the survey of recent counts of galaxies for an area $15 \times 15 \text{ arcmin}^2$ on the sky (Ellis and Bland-Hawthorn, 2006). The number counts and their standard deviations are given in Table 17.1. Note that, for the sake of clarity, the number counts in the I -waveband have been displaced by -0.5 in $\log N$, the B -band number counts by -1 in $\log N$ and the R -band counts by -1.5 in $\log N$. For comparison, the Euclidean relation $N = 0.6m + \text{constant}$ is also shown.

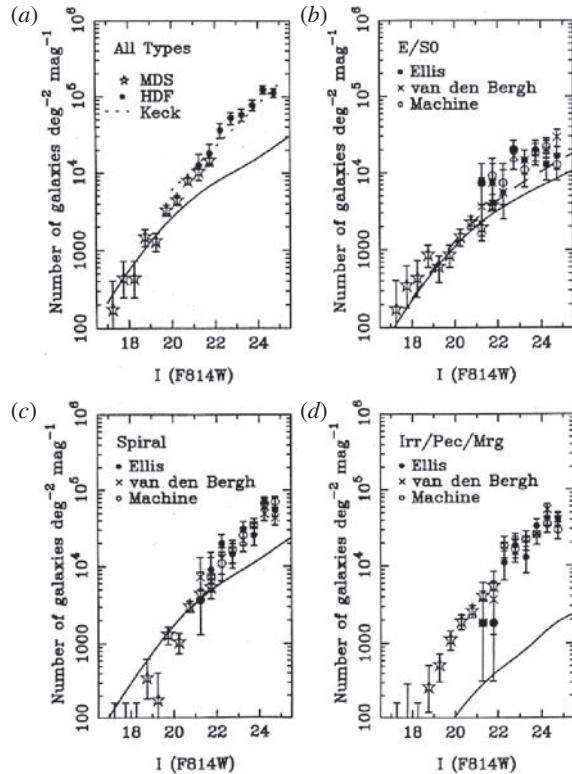


Fig. 23.14

The number–magnitude relation for morphologically segregated samples of galaxies from the Medium Deep Survey (MDS) and the Hubble Deep Field (Abraham *et al.*, 1996). The observations from the MDS survey are indicated by stars. The other symbols show the counts based upon morphological classifications carried out by Ellis, van den Bergh and by an automated machine-based classification algorithm. The dotted line in panel (a) shows the total counts of galaxies from a field observed by the Keck Telescope. The solid lines show the expected counts of the different morphological classes assuming their properties do not change with cosmic epoch (Glazebrook *et al.*, 1995).

These results are in accord with the conclusions of Ellis that there is unquestionably an increase in the numbers of star-forming galaxies with increasing redshift. Equally intriguing, Cowie and his colleagues find that there is little change in the K-band luminosity function out to redshifts $z \approx 1$, suggesting that most of their stellar populations were already in place by a redshift of 1. The existence of old, red galaxies at redshifts up to almost $z = 2$ has been confirmed by the studies by Cimatti and his colleagues (2004) from studies of the K20 sample of galaxies.

The imaging evidence from the Hubble Deep Field and the Hubble Ultra-Deep Field is even more compelling. There is no question about the fact that most of the faint objects appearing in Fig. 23.12a bear little resemblance to the classical forms of galaxy. This is particularly true if attention is restricted to the sample of galaxies which are certainly at large redshifts $z \geq 3.5$ through the observation of ‘drop-outs’ through the various filter bands. These are interpreted as star-forming galaxies in which the Lyman cut-off is

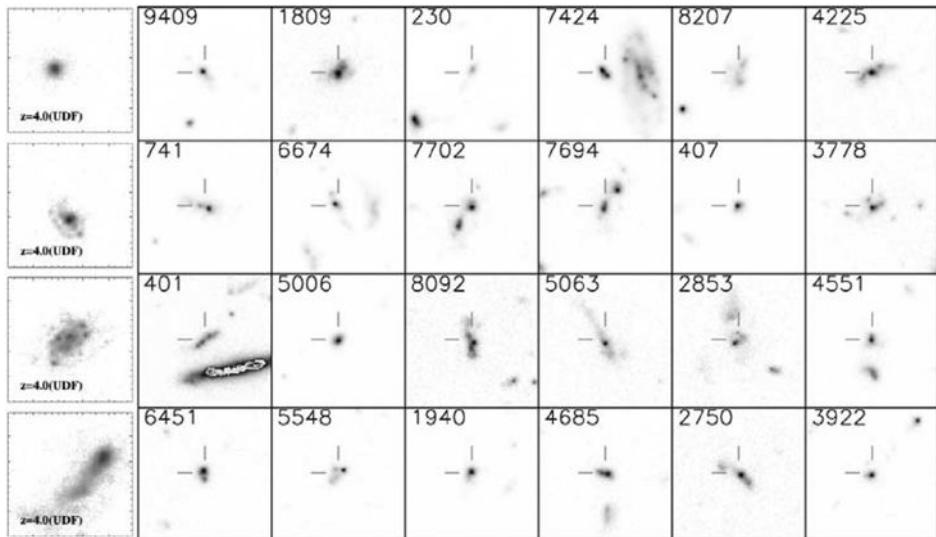


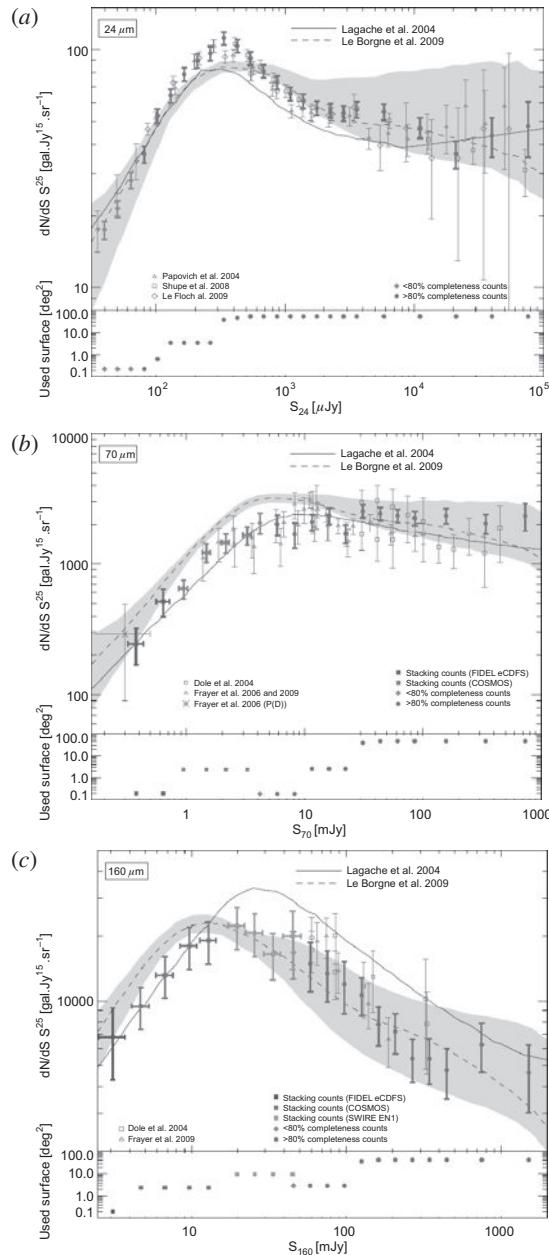
Fig. 23.15 Comparison of simulated images of nearby galaxies (four panels on the left) due to Lotz and her colleagues (2006) with images of the 24 brightest B435-dropout sources on the right. All figures correspond to 21 kpc at $z = 4.1$. The dropout sources are indicated by the small lines in the centre of each frame (Beckwith *et al.*, 2006).

redshifted into the optical waveband. The images in the paper by Beckwith and his colleagues (2006) illustrated many of the types of ‘drop-out’ galaxy observed in the HUDF (see Fig. 23.15). The galaxies are much smaller and much more irregular than those observed in the nearby Universe. Beckwith and his colleagues determined luminosity functions and size distributions for different types of drop-out galaxy.

23.5 Mid- and far-infrared number counts

The IRAS Satellite carried out the first essentially complete sky survey of the infrared sky in those wavebands which are inaccessible from the ground, namely the wavebands between 12.5 and 100 μm . The normalised differential counts showed that there were more faint IRAS galaxies than expected, in the same sense as the counts of radio sources, X-ray sources and quasars, although the range of redshifts sampled was very much smaller than those of the extragalactic radio sources and quasars (Oliver *et al.*, 1992).

The successor to the IRAS survey was the Infrared Space Observatory (ISO) of the European Space Agency which carried out deeper surveys in the mid- and far-infrared wavebands and confirmed the excesses of faint sources. Most recently, number counts in these wavebands have been made by the NASA Spitzer Space Telescope, an 85 cm cryogenic telescope with three cooled scientific instruments for imaging and spectroscopy in the 3.6–160 μm waveband. These instruments resulted in number counts at 24, 70 and 160 μm and all show a clear excess of faint sources, as well as a rather dramatic cut-off at low flux densities (Fig. 23.16).

**Fig. 23.16**

Differential number counts at (a) 24 μm , (b) 70 μm and (c) 160 μm (Béthermin *et al.*, 2010). The filled circles represent points obtained with $\geq 80\%$ completeness. The filled diamonds show points with 50–80% completeness. The continuous lines show the model calculation of Lagache and his colleagues (Lagache *et al.*, 2004). The dashed line and grey region show the models of Le Borgne and his colleagues (2009) and their 90% confidence region.

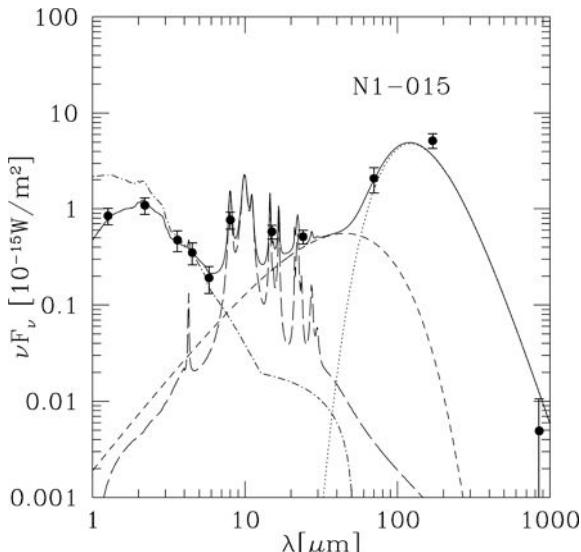


Fig. 23.17 An example of the spectral energy distribution of the luminous infrared galaxy N1-015 between 1 and 1000 μm (Sajina *et al.*, 2006). The components include a grey-body spectrum in the submillimetre waveband (dotted line), a warm power law (short-dashed line), PAH emission (long-dashed line), and stellar emission (dot-dashed line) which suffers extinction $e^{-\tau_\nu}$, where τ_ν describes the extinction as a function of wavelength. The total spectrum is shown by the continuous line and the observed broad-band flux densities by filled circles.

The complications of interpreting the number counts of mid- and far-infrared galaxies is illustrated by the example of the spectral energy distribution of one of the luminous far-infrared galaxies, N1-015 (Sajina *et al.*, 2006). Sajina and her collaborators showed that the spectrum can be decomposed into a number of components which are illustrated in Fig. 23.17. At the longest wavelengths, the submillimetre spectrum can be approximated by a grey-body spectrum, meaning a black-body spectrum with optically thin emissivity varying as $\nu^{1.5}$ (dotted line). In addition, to account for the gradual decline from the maximum of the black-body curve to shorter wavelengths, a warm power-law component is included, associated with dust grains of smaller size with decreasing wavelength (short-dashed line). In the 5–20 μm waveband the spectrum is dominated by the intense emission of polycyclic aromatic hydrocarbon molecules, PAHs, which can account for up to 20% of the total infrared luminosity (long-dashed line). At the shortest wavelengths, the stellar population of the galaxy is dominant (dot-dashed line), but it suffers extinction as $e^{-\tau_\nu}$, where τ_ν describes the attenuation of the spectral energy distribution as a function of wavelength. The total spectrum is shown by the continuous line and the observed broad-band flux densities by filled circles. Thus, the contributions of the various components in different classes of infrared galaxy need to be established. Various template-fitting procedures have been developed which enable estimates of the evolving luminosity function of galaxies in the infrared waveband to be established (Lagache *et al.*, 2003, 2004; Babbedge *et al.*, 2006).

There is reasonable agreement about the features needed to account for the number counts of galaxies in the mid- to far-infrared wavebands. First of all, the counts are dominated by

starburst activity. Babbedge and his colleagues use template fitting to study the evolution of active galactic nuclei in the SWIRE Legacy sample and find that they are far outnumbered by starburst galaxies (Babbedge *et al.*, 2006). They find the same form of evolution in the infrared waveband for the active galactic nuclei as described in Sect. 23.3. There are certainly some far-infrared sources in which the energy source is likely to be an active galactic nucleus, but X-ray studies of far-infrared and submillimetre sources have shown that there are relatively few X-ray active galactic nuclei as compared with those which are dominated by starbursts.

The phenomenological model of Lagache and his colleagues provides a simple picture for the evolutionary behaviour of the starburst population in the mid-infrared to submillimetre wavebands (Lagache *et al.*, 2003, 2004). The population is split into two components, a population of normal galaxies which evolves passively with cosmic epoch and a strongly evolving starburst component which consists of all the highest luminosity far-infrared sources, $L \geq 3 \times 10^{11} M_{\odot}$. The infrared spectra are not very different from that shown in Fig. 23.17, although the far-infrared maximum is somewhat colder for the normal galaxies as compared with the luminous starbursts. The strong evolution is associated with this luminous starbursting component. The form of evolution is significantly different from that found for the population of active galaxies in that all the strong evolution takes place between redshifts 0 and 1 and then remains at roughly the same level from redshifts $1 \leq z \leq 5$. Specifically, the evolution of the luminosity function for the luminous star-forming galaxies can be described by luminosity evolution of the form $L \propto (1+z)^3$ from $z = 0$ to $z = 1.5$, beyond which the luminosity function remains constant out to redshift $z = 5$. This result is similar to that found by Babbedge and his colleagues from their larger dataset. The various evolutionary predictions are compared with the observations in Fig. 23.16, where it can be seen that the number counts throughout the mid- and far-infrared wavebands are in reasonable agreement with this model, which can be tweaked to obtain better agreement with observation (Lagache *et al.*, 2004; Le Borgne *et al.*, 2009). Lagache and his colleagues show that the abrupt convergence of the counts at low flux densities seen in Fig. 23.16 is accounted for in their models. The reason for this is that the strong evolution takes place over the redshift range $0 \leq z \leq 1$ and then the comoving space density of sources is constant. As a result, the differential number counts abruptly change to the strong converging behaviour at $z \geq 1$ seen in Fig. 23.1.

23.6 Submillimetre number counts

The predictions shown in Fig. 23.1 are strongly modified for galaxies which are intense dust emitters in the submillimetre waveband, $0.1 \leq \lambda \leq 1$ mm. Examples of the stellar energy distributions from submillimetre to ultraviolet wavelengths for a sample of infrared galaxies selected at 60 μm are shown in Fig. 23.18a (Sanders and Mirabel, 1996). The maximum intensity at 50–100 μm corresponds to the temperatures to which the dust grains are heated, which are about 30–60 K. To the long-wavelength side of this maximum, the dust clouds are optically thin and the spectral index of their thermal emission is typically

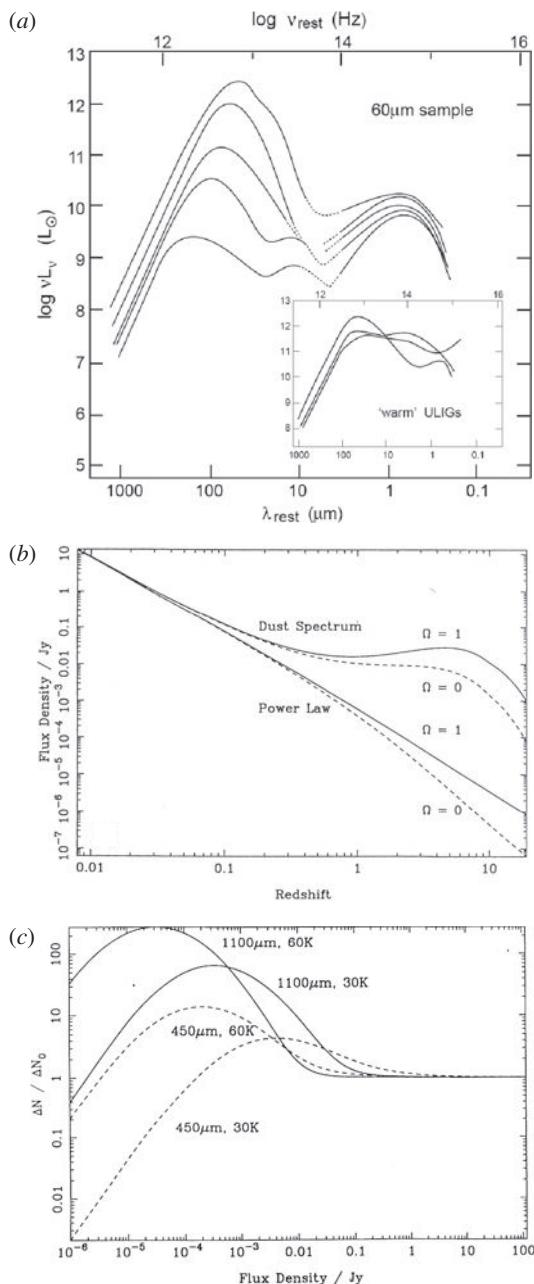


Fig. 23.18

(a) Examples of the spectral energy distributions from submillimetre to ultraviolet wavelengths for a sample of infrared galaxies selected at 60 μm. The insert shows the energy distributions for ultra-luminous infrared galaxies which have 'warm' infrared colours (Sanders and Mirabel, 1996). (b) The flux density–redshift relation in the submillimetre waveband for an infrared galaxy with the spectrum shown in (a), assuming the source has far-infrared luminosity $10^{13} L_\odot$ for world models with $\Omega_0 = 1$ and 0. For comparison, the flux density–redshift relations for a population of sources with power-law spectra $S_\nu \propto \nu^{-1}$ are also shown. (c) The predicted differential normalised counts for a uniform distribution of sources at 450 and 1100 μm assuming the galaxies have spectra similar to those of Fig. 23.18a and that the far-infrared luminosity function at 60 μm is that of IRAS galaxies (Saunders *et al.*, 1990). The predicted number counts at 450 and 1100 μm are shown for dust temperatures of 30 and 60 K (Blain and Longair, 1993).

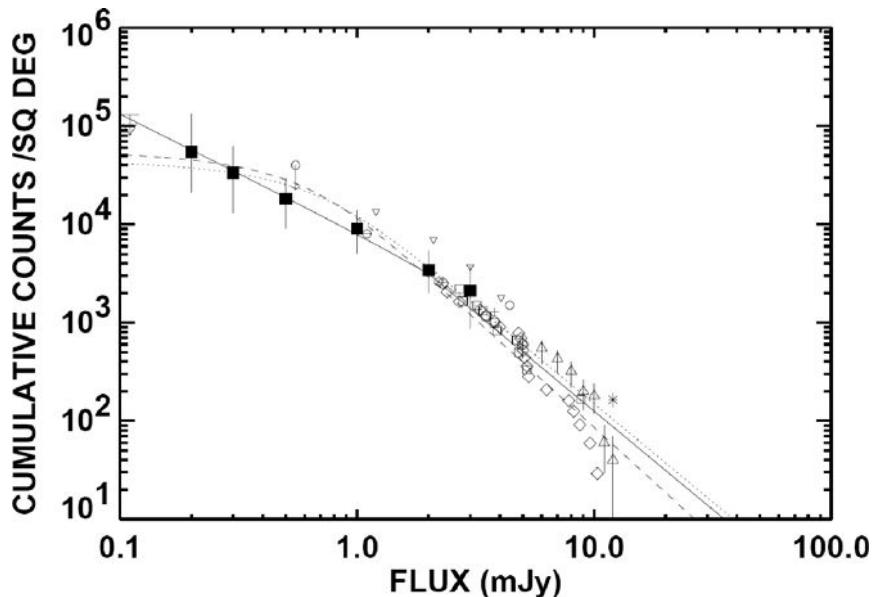


Fig. 23.19

The counts of submillimetre sources at $850\text{ }\mu\text{m}$ (Cowie *et al.*, 2002). The sources of these data are discussed by Cowie and his colleagues. The filled squares with error bars have been derived from an analysis of submillimetre sources observed in the vicinity of clusters of galaxies which enhances their flux densities because of gravitational lensing. The points with flux densities greater than about 2 Jy are derived from blank field surveys. The dotted and dashed lines show parametric fits to the counts which converge to the total extragalactic background emission observed in the submillimetre waveband.

$\alpha \approx -3$ to -4 . The luminosities in the far-infrared waveband can exceed those in the optical waveband and these types of galaxy are dominant contributors to the counts of sources in the submillimetre waveband.

The flux density–redshift relations at submillimetre wavelengths shown in Fig. 23.18*b* are quite different from those of active galaxies. The relations flatten, or even invert, at redshifts $z \sim 1$ because of the very large negative K-corrections in these wavebands (Blain and Longair, 1993, 1996). As a result, once galaxies of a given luminosity are observed at $z \sim 1$, the whole of the redshift range out to $z \sim 5$ – 10 becomes observable for these at roughly the same flux density. There is a corresponding dramatic increase in the number counts of sources at this flux density. In Fig. 23.16*c*, the predicted differential source counts are shown for a population of far-infrared sources with the far-infrared luminosity function of IRAS galaxies (Saunders *et al.*, 1990). When the source distribution extends to redshifts $z \geq 1$, the source counts are inverted with $\beta > 1.5$, even if there is no cosmological evolution of the source population.

Counts in the submillimetre waveband became possible thanks to the availability of the SCUBA array detector on the James Clerk Maxwell Telescope, which resulted in the discovery of a very large population of submillimetre sources (Smail *et al.*, 1997). Over the succeeding years a major effort was made to determine their number counts in detail and this is summarised in Fig. 23.19 (Cowie *et al.*, 2002). What Fig. 23.19 disguises is the fact

that these number counts greatly exceed the numbers of sources expected on the basis of the local 60 μm luminosity function of luminous far-infrared galaxies determined by the IRAS survey. Indeed the numbers exceeded the most extreme evolution models discussed by Blain and me in the early 1990s (Blain and Longair, 1993). The number counts of submillimetre galaxies must converge rapidly just below those flux densities shown in Fig. 23.19 or the submillimetre background intensity would be exceeded. The parametric fits to the counts shown by dashed and dotted lines in Fig. 17.20 converge to the submillimetre background intensity estimates by Puget, Hauser and their colleagues (Puget *et al.*, 1996; Hauser *et al.*, 1998; Hauser and Dwek, 2001).

23.7 The global star-formation rate

The observations discussed in Sects 23.4–23.6 are related the global rate of formation of stars as a function of cosmic epoch. This is an enormous subject and many of its different facets are described in Chap. 19 of my book *Galaxy Formation* (Longair, 2008), to which the reader is referred for many more details. A brief summary of the contents of that discussion is as follows:

- Lilly and Cowie first showed how the rate of formation of heavy elements could be inferred from the flat blue continuum spectra of star-forming galaxies and that these estimates are independent of the choice of cosmological model (Lilly and Cowie, 1987; Cowie, 1988). A prolonged burst of star formation has a remarkably flat intensity spectrum at wavelengths longer than the Lyman limit at 91.2 nm, as illustrated by the evolving model starbursts presented by White from computations using the spectral synthesis codes of Bruzual (White, 1989). Figure 23.20 shows the spectrum of a starburst galaxy as observed at different ages, assuming that the star-formation rate is constant with a Salpeter initial mass function. The distinctive spectrum of such a star-forming region, with a flat spectrum and a cut-off at the Lyman limit, has been used in many different ways to select star-forming galaxies and so estimate the rate of star formation out to redshifts $z \sim 6$.
- Cowie, Lilly and their colleagues undertook deep multi-colour surveys to discover flat spectrum star-forming galaxies at large redshifts and were successful in finding such objects with roughly equal intensities in the U , B and V wavebands. The background due to such objects amounted to about $10^{-24} \text{ W m}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1}$ which they interpreted as meaning that a significant fraction of the heavy elements, about $1.5 \times 10^{-31} \text{ kg m}^{-3}$, must have been synthesised at redshifts of about one (Cowie *et al.*, 1988). The same type of analysis was used to set limits to star formation rates in the Canada–France Redshift Survey by Lilly and his colleagues and in the Hubble Deep Field by Madau and his colleagues (Lilly *et al.*, 1995; Madau *et al.*, 1996)
- Steidel and his colleagues extended the multicolour technique for finding star-forming galaxies to redshifts $z > 3$ (Steidel, 1998; Steidel *et al.*, 1999). The idea was similar to that developed by Cowie and Lilly, but now the objective was to search for star-forming

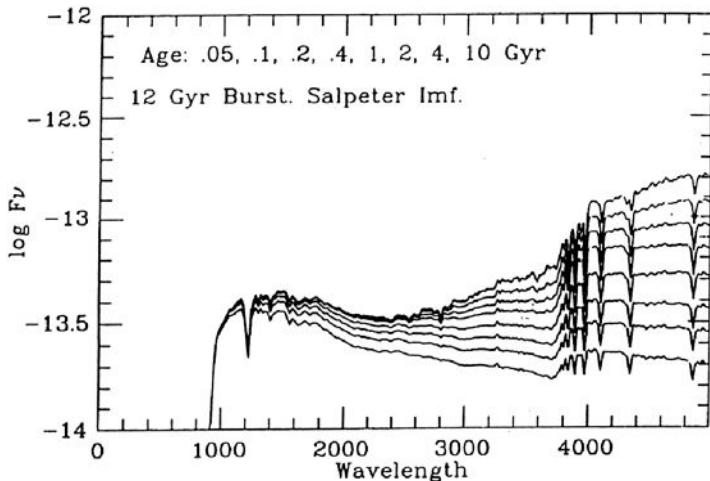


Fig. 23.20

Synthetic spectra for a starburst with constant star-formation rate as observed at the ages indicated. A Salpeter initial stellar mass function $N(M) dM \propto M^{-2.35} dM$ has been assumed with cut-offs at 75 and $0.08 M_{\odot}$. The spectra were generated by Bruzual using his evolutionary synthesis programs (White, 1989).

galaxies in which the Lyman limit is redshifted into the optical waveband. These objects are often referred to as *Lyman drop-out galaxies*.

- Observations of the Hubble Deep and Ultra-Deep Fields proved to be ideal for exploiting this approach because, in addition to very precise photometry in four wavebands spanning the wavelength range $300 < \lambda < 900$ nm, high resolution optical images enabled the morphologies of these galaxies to be studied (Beckwith *et al.*, 2006; Bouwens *et al.*, 2006). Of particular interest for the study of the star-formation rate is the analysis of the properties of the drop-out galaxies at progressively larger redshifts. Beckwith and his colleagues reported that, of the 10 040 objects appearing in the catalogue in the i_{775} waveband, there were 504 B_{435} dropouts, 204 V_{204} dropouts and 54 i_{775} dropouts, corresponding to galaxies with redshifts in the ranges $3.5 \leq z \leq 4.7$, $4.6 \leq z \leq 5.7$ and $5.7 \leq z \leq 7.4$, respectively.
- The Subaru Deep Surveys made use of the outstanding wide-field capability of the Japanese 8 metre Subaru telescope to show that, if attention is restricted to luminous far-ultraviolet galaxies, with $M_{\text{FUV}} \leq -21.3$, there is a factor of 10 decrease in the far-ultraviolet luminosity density between redshifts $z \sim 3$ and $z \sim 6$ (Shimasaku *et al.*, 2005). The less luminous sources, with $M_{\text{FUV}} \leq -19$, show very much less dramatic evolution but at the largest redshifts the less luminous sources produce about 30 times more ultraviolet luminosity than the luminous galaxies (Fig. 23.21).
- Similar results for the ultraviolet luminosity density were found from analyses of the data from the Hubble Ultra-Deep Field which suggest that it only decreased slightly from the values found at redshift $z \sim 2$ (Giavalisco *et al.*, 2004; Beckwith *et al.*, 2006). Beckwith and his colleagues agree with the conclusion that the flatness of the intensity distribution with increasing redshift is largely associated with the fainter galaxies in the samples at large redshifts. Thus, star formation was certainly well underway by the largest redshifts

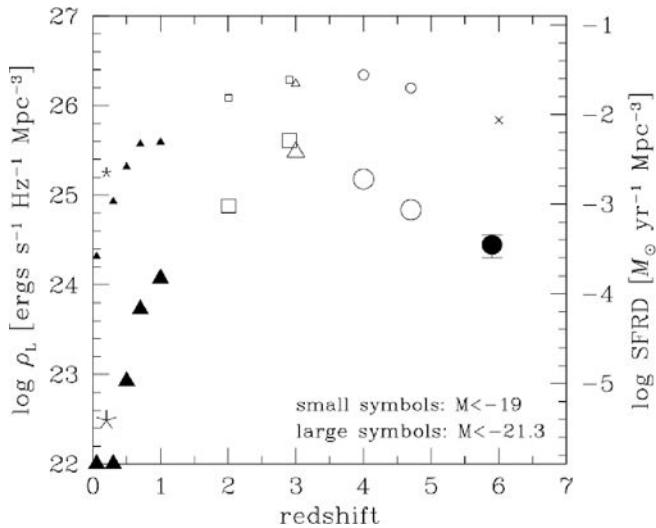


Fig. 23.21 Evolution of the far-ultraviolet (FUV) luminosity density and the corresponding star-formation rates. The large and small symbols indicate the luminosity density of galaxies brighter than far-ultraviolet absolute magnitudes $M = -21.3$ and $M = -19.0$, respectively (Shimasaku *et al.*, 2005).

accessible by the HUFD survey. There is general agreement, however, that there is an order of magnitude decrease in the ultraviolet ionising radiation between redshifts $z \sim 1$ and the present epoch, as described in the pioneering paper by Madau and his colleagues (Madau *et al.*, 1996).

- A concern about optical determinations of the cosmic star-formation rate is the extent to which the statistics are influenced by the effects of dust extinction. Galaxies undergoing bursts of star formation are not only sources of intense ultraviolet continuum radiation but are also strong emitters in the far-infrared waveband because of the presence of dust in the star-forming regions. The solution is to make observations in the submillimetre waveband in which star-forming regions are intense dust emitters and the total millimetre/submillimetre luminosity of a galaxy provides a measure of the star-formation rate. Wang, Cowie and Barger used a stacking technique to correlate the near-infrared and X-ray flux densities of galaxies in the GOODs survey areas with their submillimetre flux densities (Wang *et al.*, 2006). They showed that active galactic nuclei could at most contribute 15% of the total background, whereas the sources present in the field at 1.6 and 3.6 μm could account for at least 60% of the 850 μm background intensity. Most of the submillimetre background radiation was associated with Sb and Sc galaxies rather than with the Sd and irregular galaxies, many of the latter being the types of star-forming galaxy found in the deep optical-infrared surveys. The resulting plot of star-formation rate against redshift has a similar form to that derived from the optical-near infrared studies, namely, with increasing redshift the star-formation rate increases rapidly from zero to $z \approx 1$ and then remains roughly constant at redshifts up to four (Fig. 23.22). This result is similar to the estimates from the optical-UV continuum observations once appropriate extinction corrections are made, but the surprising result was that the blue star-forming

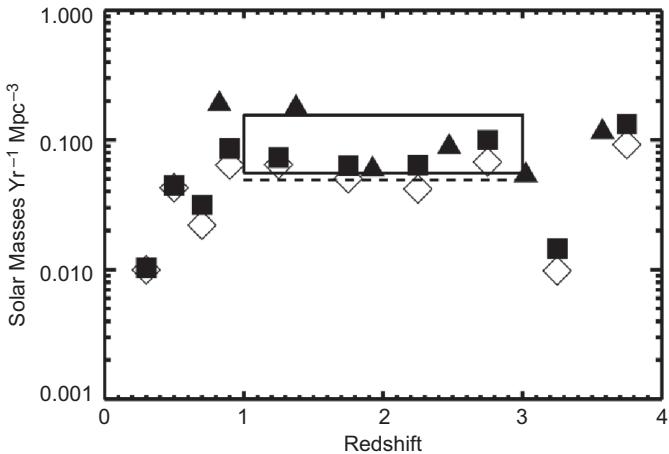


Fig. 23.22 The star-formation rate density as a function of redshift derived from various submillimetre surveys. The dashed horizontal line shows the rate assuming all the sources with $850 \mu\text{m}$ flux densities greater than 4 mJy lie at $1 \leq z \leq 3$. The rectangular region shows the density from the remaining submillimetre background that is not accounted for by the near-infrared observations, assuming that the sources lie in the redshift interval $1 \leq z \leq 3$ (Wang *et al.*, 2006).

galaxies did not contribute to the star-formation rates measured in the submillimetre waveband.

The reason for devoting attention to the rate of star formation in galaxies is that this process provides a number of essential ingredients for powering black holes in active galactic nuclei. First of all, it provides large amounts of fuel which can eventually be accreted onto the central black hole and second the accreted mass contributes to the growth of the mass of the black hole. Furthermore, there was much more star-formation activity at large redshifts as compared with the present epoch and this may well contribute to the observed strong evolution of the populations of active galaxies with cosmic epoch. The evidence suggests that the star-formation rate continues at a high rate at redshifts greater than about 3, whereas the active galaxies and quasars decrease in numbers with increasing redshift beyond $z \approx 3$. A further complication concerns the population of old red galaxies.

23.8 The old red galaxies

Various pieces of evidence indicate that some galaxies were already ‘old’ out to redshifts of at least $z = 2$ and so must have formed the bulk of their stellar populations at large redshifts. Furthermore, the most massive galaxies seem to have more or less the same maximum mass over the redshift interval $0 < z < 2$, contrary to what might have been expected if they had increased in mass and luminosity through the process of hierarchical clustering.

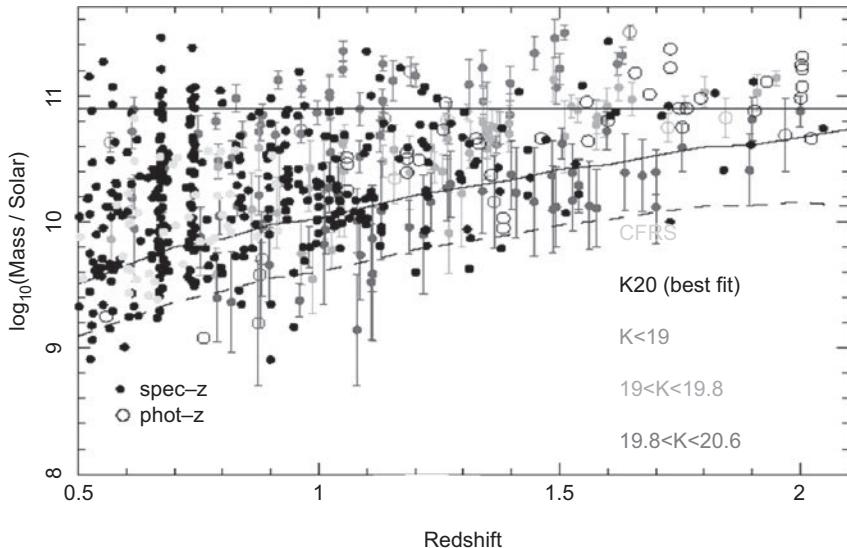


Fig. 23.23 A plot of stellar mass against redshift for galaxies selected from the Gemini Deep Deep Survey and K20 Surveys (McCarthy, 2006). The galaxies were selected in the K ($2.2 \mu\text{m}$) waveband. The stellar masses were estimated by fitting the extensive multicolour data to the stellar energy distribution codes of Bruzual and Charlot (2003).

- The plot of stellar mass against redshift for the Gemini Deep Deep and K20 surveys is shown in Fig. 23.23 (McCarthy, 2006). The galaxies were selected in the K ($2.2 \mu\text{m}$) waveband and so primarily sampled the old stellar populations of galaxies. The stellar masses of the galaxies were derived by fitting the extensive multicolour photometric data from the above surveys to the stellar energy distributions provided by the galaxy evolution codes of Bruzual and Charlot (2003). As emphasised by McCarthy, Fig. 23.23 represents the evolving stellar-mass density which is the complement of the star-formation history discussed in the last section (McCarthy, 2006). The horizontal solid line at $\log(M/M_\odot) = 10.9$ is the present day value of the stellar mass corresponding to the ‘turnover’ luminosity L^* of the galaxy luminosity function and the dashed line represents the detection limit of the survey at $K = 20.6$. There is little evidence for any increase in the baryonic masses of the galaxies with decreasing redshift. Glazebrook and his colleagues compared the data shown in Fig. 23.23 with the expected increase in mass according to the standard hierarchical clustering models of galaxy formation with decreasing redshift (Glazebrook *et al.*, 2004) and showed that the observations were inconsistent with these expectations. It is inferred that these galaxies must have completed the formation of their stellar populations by a redshift $z \sim 2$.
- In addition to photometry, 20 of the Gemini Deep Deep Survey galaxies with $z > 1.3$ were observed spectroscopically with very long exposures (McCarthy *et al.*, 2004). Among objects with spectroscopic redshifts, 51% of the galaxies with $I - K > 3.5$ showed unambiguous evidence for old stellar components while the redder galaxies with $I - K > 4$ showed an even greater tendency for old stellar spectra. McCarthy and his colleagues concluded that for 50–80% of the red galaxies in the sample at $z \geq 1.3$ the

optical-infrared emission is dominated by contributions from old stellar populations. Typically, the stellar populations must have formed by a redshift of 2.4 or greater for the reddest examples. As a result, the star-formation rates must have been very great during this initial ‘starburst’ phase, values of the order of $300\text{--}500 M_{\odot}$ year $^{-1}$ being deduced.

- Similar results have been reported by Cimatti and his colleagues from their analyses of the properties of four very red galaxies with redshifts in the range $1.6 \leq z \leq 1.9$ from the K20 sample (Cimatti *et al.*, 2004). The spectroscopic and morphological properties of these galaxies indicated that they were ‘old, fully assembled, spheroidal galaxies’ with masses $M > 10^{11} M_{\odot}$. These objects are similar to the old red galaxies discussed above. Cimatti and his colleagues showed that the build-up of massive early-type galaxies must have been much faster in the early Universe than expected.
- Evidence that the stellar populations of large redshift radio galaxies are likely to be old was presented by Lilly for the radio galaxy 0902+34 at a redshift $z = 3.4$ and by Dunlop and his colleagues for the radio galaxy LBDS 53W091 at redshift 1.55 (Lilly, 1988; Dunlop *et al.*, 1996). In the latter case, the age of its stellar population was at least 3.5×10^9 years so that these stars must have been formed at large redshifts. A similar result, namely an age of 4.5×10^9 years was found for the radio galaxy LBDS 53W069 at redshift $z = 1.43$ (Dey, 1997). Radio galaxies with qualitatively the same stellar properties have been observed to redshifts greater than 4 by Lacy, Spinrad and their colleagues (Lacy *et al.*, 1994; Spinrad *et al.*, 1995).
- Qualitatively similar conclusions can be derived from observations of the massive galaxies associated with the most powerful FR2 radio sources. The redshifts of these 3CR radio galaxies span the redshift interval $0.3 \leq z \leq 2$ and their mean stellar masses are remarkably constant over this redshift interval (Best *et al.*, 1998). These galaxies are among the most luminous known, having luminosities as great as those of the brightest galaxies in clusters at redshifts $z \sim 1$. Their stellar masses were independently estimated using the multicolour technique by McLure and his colleagues and were found to be $M \geq 3 \times 10^{11} M_{\odot}$ (Fig. 19.7b), that is, greater than those of the most massive galaxies plotted in Fig. 23.23 (McLure *et al.*, 2006). There is a strong correlation between the mass of the black hole in the quasar nucleus and the mass of the spheroid of the host galaxy (Magorrian *et al.*, 1998), the ratio of black hole mass M_{bh} to the mass of the spheroid M_{sph} being about 0.002 (Fig. 19.7a). Taking account of the fact that the spheroid mass must be about 10 times greater than the stellar mass to take account of the dark matter, these data would be consistent with a ratio $M_{\text{bh}}/M_{\text{sph}} \sim 0.002$. These data are naturally interpreted as evidence that the host galaxies of the quasars must have had masses $M \geq 10^{12} M_{\odot}$ and that their stellar populations must already have been assembled into galaxies, more or less as we know them today, by a redshift of 2–3.

It therefore seems that a significant fraction of all massive galaxies must have formed their stellar populations at large redshifts, $z \geq 2.5$, and that there was little subsequent increase in their masses through hierarchical clustering and subsequent star formation. This contrasts with the inference that the global star-formation rate was at its peak at redshifts $z \sim 1\text{--}2$. Most of the star formation must have taken place in less massive galaxies than those discussed in this section. This phenomenon that the massive galaxies formed their stellar

populations at early epochs and star formation continues for lower mass galaxies at much later epochs is referred to as ‘down-sizing’.

23.9 Putting it all together

How do all these observations fit together within the context of our understanding of the origin and evolution of galaxies and large scale structures in the Universe? There are no clear answers, but some clues are provided by the large scale simulations of structure and galaxy formation.¹

The concordance Λ CDM model has had considerable success in accounting for the observed features of the Universe on large scales, the presupposition being that, to create a fully-fledged theory of galaxy formation, many of these phenomena must be associated with the astrophysical processes by which the baryonic matter evolved within the dark matter haloes. This necessarily involves poorly understood processes such as the dependence of the star-formation rate upon the properties of the interstellar medium, the role of supernova explosions in sweeping matter away from star-forming regions and out of galaxies, the role of collisions between galaxies, the formation and growth of black holes in the nuclei of galaxies, and so on. These processes necessarily involve feedback mechanisms which can promote or suppress the formation of new stars and structures. Among these are the influence of the active galactic nucleus upon the astrophysics of the galaxy as a whole. The process of galaxy formation is therefore highly nonlinear.

Computational astrophysics has made enormous contributions to understanding in considerable detail the origin and evolution of cosmic structures and the observational fingerprints these leave in the Cosmic Microwave Background Radiation. For example, the *Millennium simulation* for the growth of dark matter perturbations under gravity involved following the evolution of over 10^{10} particles from a redshift $z = 127$ to the present epoch within a cubical region of comoving dimension $500h^{-1}$ Mpc (Springel *et al.*, 2005). The next step has been to build on the success of these programmes in order to address the astrophysical issues involved in the formation of stars and supermassive black holes in galactic nuclei. This is the domain of *semi-analytic models of galaxy formation*.

In the semi-analytic approach, rules need to be set up to describe how the baryonic matter cools, forms stars and completes the full life cycle of stellar evolution. In the words of Springel and his colleagues:

The term ‘semi-analytic’ conveys the notion that while in this approach the physics is parameterised in terms of simple analytic models, following the dark matter merger trees over time can only be carried out numerically. Semi-analytic models are hence best viewed as simplified simulations of the galaxy formation process.

In many ways, it is helpful to regard this approach as ‘experimental computational astrophysics’ in the sense that there are many possible realisations of these models and the

¹ For many more details, see my book *Galaxy Formation* (Longair, 2008).

aim is to constrain the necessary and essential physics on a trial and error basis. The list of necessary ingredients for the models include:

- Central to the whole procedure are the processes of *radiative cooling and star formation*. Cold condensed gas accumulates in the central regions of the dark matter haloes and is identified with the interstellar medium of the protogalaxy. The cool gas settles into a disc supported by rotation. Once the surface gas density exceeds the critical density found from observational studies such as those of Kennicutt (1989), star formation is assumed to take place with an efficiency of 10% on the dynamical time-scale of the disc. The parameters describing these processes are chosen to reproduce the phenomenological laws of star formation and the observed gas fractions in galaxies at small redshifts.
- The most massive stars formed have short lifetimes and explode as *supernovae*. Guided by observation, it is assumed that supernovae can blow gas out of star-forming discs and that the rate of mass ejection is proportional to the total mass of stars formed.
- The *morphologies of the galaxies* are parameterised by their bulge-to-disc ratios which are correlated with their Hubble types. Any surviving cold gas collapses to the nuclear regions of the galaxy where it gives rise to a nuclear starburst. The parameterisation of the latter process is derived from systematic studies of hydrodynamical simulations of galaxy collisions (Mihos and Hernquist, 1994, 1996).
- The models should also reproduce the observed spectra of galaxies at different stages of their evolution. The stellar population synthesis codes enable predictions to be made of the *spectrophotometric properties of galaxies* (Bruzual and Charlot, 2003). An important element of the spectral evolution of galaxies is the recycling of processed stellar material through the interstellar gas, both to the hot and cold phases – the next generation of stars is formed in the cold phases. Dust extinction needs to be included and this depends upon the enrichment of the interstellar gas with heavy elements and their condensation into dust grains.
- Most important for our present story is that account needs to be taken of the *formation of massive black holes in the nuclei of galaxies*. It is assumed that the growth of these black holes takes place as a result of galaxy mergers during which cold gas is dragged into the central regions under the influence of tidal forces. This process creates a nuclear starburst, as well as providing mass which is accreted into the black hole. The parameterisation of this process is chosen so that the observed relation between the mass of the central black hole and the mass of the bulge is obtained. The accretion process results in the huge luminosities of the quasars, so long as the supply of fuel lasts. The ultraviolet radiation of the quasars can contribute to the ionisation of the interstellar and intergalactic gas in their vicinity. In addition, the powerful relativistic jets of active galaxies have a profound influence upon the surrounding interstellar and intergalactic media, as illustrated in Fig. 4.9 for the case of the Perseus Cluster. The effect of this heating is to inhibit the further accretion of baryonic mass onto the galaxy. In the modelling, it is assumed that the heating rate is proportional to the mass of the black hole and to $T^{3/2}$, where T is the temperature of the gas. This process is adopted by the modellers to account for the fact that massive galaxies do not grow indefinitely and so provides a cut-off at the high mass end of the luminosity function.

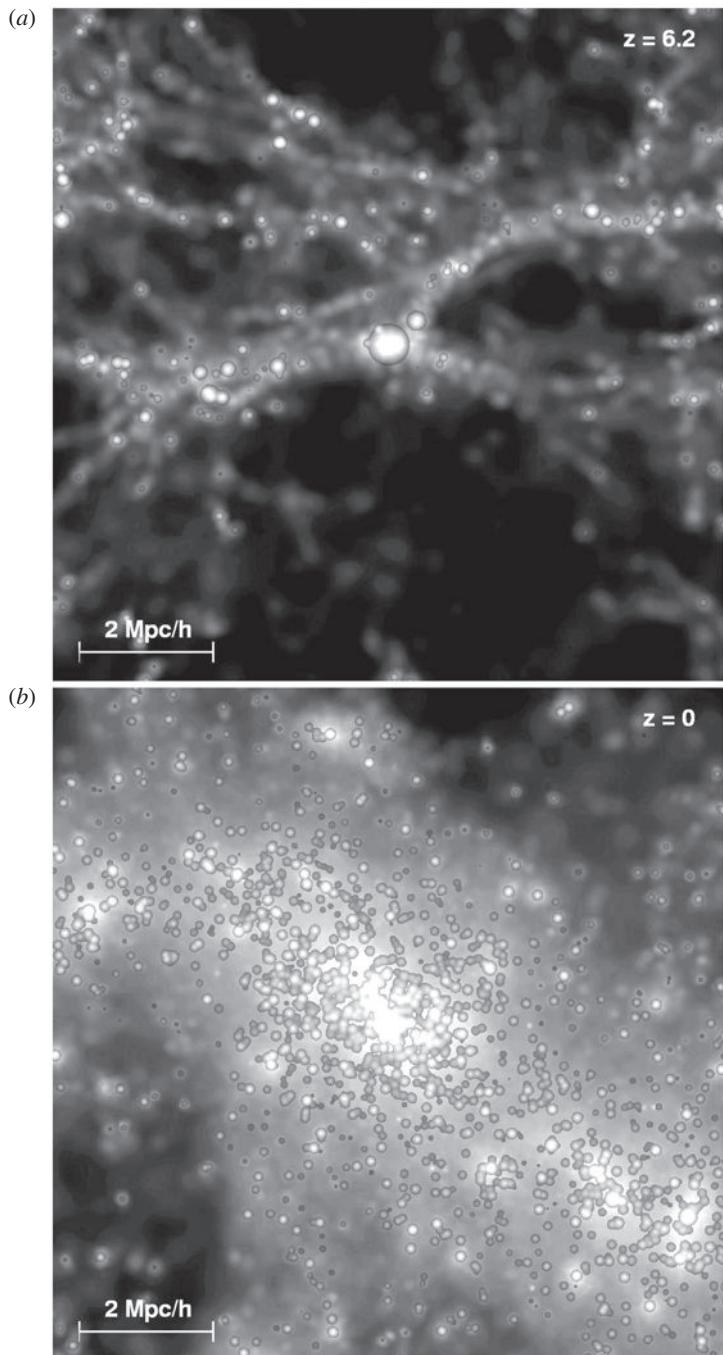


Fig. 23.24 (a) The environment of a ‘first quasar candidate’ in the Millennium simulation at redshift $z = 6.2$ and (b) at the present epoch in a cube of comoving dimension $10h^{-1}$ Mpc. The galaxies of the semi-analytic model are shown as circles overlayed on a grey-scale image of the dark matter density distribution. The volume of the sphere representing each galaxy is proportional to its stellar mass. At $z = 6.2$, all the galaxies are blue because of ongoing star formation, whereas many of the galaxies which have been accreted into the rich cluster at $z = 0$ have evolved into red galaxies (Springel *et al.*, 2005).

This list indicates the types of input physics which the semi-analytic modellers have to adopt in order to account for the observed properties of galaxies. It is an ambitious list, but it is hard to argue the proposition that these processes are not important in the evolution of galaxies. The models provide practical realisations of the many different physical processes needed to create a self-consistent picture of galaxy formation.

Let us give a few examples of the successes of the semi-analytic modelling procedure. Having incorporated all the baryonic physics into the Millennium dark matter simulations, the two-point correlation function for galaxies provides a much better fit to the observations than that of the dark matter haloes. In addition, the models can account for the galaxy two-point correlation function for galaxies of different luminosities and colours (Springel *et al.*, 2005). The models also provide clear predictions about the evolution of the mass function of clusters of galaxies as a function of cosmic epoch. There remain some significant problems, including accounting for the zero-point of the Tully–Fisher relation and details of the fundamental plane for elliptical galaxies.

From our present perspective, the models can account for the formation of quasars at redshifts as large as $z = 6$. Figure 23.24a shows the region of the simulation containing a massive galaxy at redshift $z = 6.2$, as well as the cellular large scale structures about it. Although these high density peaks are very rare, they do occur statistically and lead to the early formation of a galaxy massive enough to host a supermassive black hole. The same region can be tracked to the present epoch when it turns out that the overdense region has evolved into a rich cluster of galaxies with a massive galaxy at its centre (Fig. 23.24b) (Springel *et al.*, 2005). The simulations also indicate that the epoch of maximum quasar activity occurred at redshifts $z \sim 2\text{--}3$.

From my personal perspective, the importance of the semi-analytic approach is that it provides a powerful new tool for testing hypotheses about the physical processes involved in galaxy and supermassive black hole formation. The simulations also provided strong motivations for future programmes of observation and theoretical analysis. One of the challenges is to keep track of the role of the non-linear astrophysics involved in these remarkable simulations and make them an integral part of the toolkit of extragalactic high energy astrophysics.

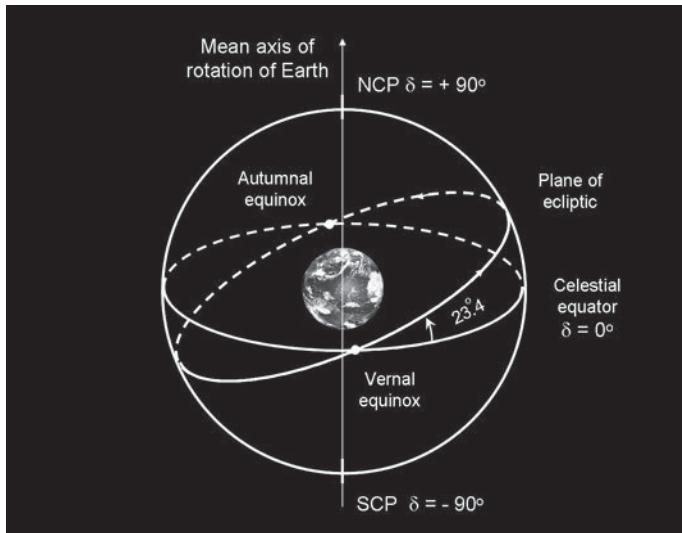
A.1 Galactic coordinates and projections of the celestial sphere onto a plane

The complexities of defining the celestial system of coordinates go far beyond what is needed in this text. These arise because the Earth does not move in a perfectly elliptical orbit about the Sun but is subject to wobbles and precessions because of the perturbing influence of the Moon and planets. These issues are dealt with in the textbooks by Smart and Murray (Smart, 1977; Murray, 1983).

The positions of celestial objects are described by a fixed set of spherical polar coordinates on the sky known as *right ascension* (RA or α) and *declination* (Dec or δ). The north celestial pole (NCP) is defined to be the mean direction of the rotation axis of the Earth and declination is the polar angle measured from the equator ($\delta = 0^\circ$) towards the north celestial pole ($\delta = 90^\circ$) (Fig. A.1). The south celestial pole (SCP) has declination $\delta = -90^\circ$. In the year 2000.0, the Earth's rotation axis was tilted at an angle of $23^\circ 26' 21.448''$ with respect to the direction perpendicular to the plane of the ecliptic, which is the plane of the Earth's orbit about the Sun (Fig. A.1). The coordinates of right ascension and declination are referred to the reference epoch 2000.0 which is known as the 2000.0 coordinate system. The Earth's rotation axis points more or less in the same direction as the Earth moves round the Sun and this gives rise to the seasonal changes of climate. There are, however, important long term precessional changes of the rotation axis of the Earth due to tidal torques acting on the aspherical Earth and these need to be taken into account in transforming from epoch 2000.0 coordinates to the epoch of observation.

The zero point of right ascension is taken to be the point at which the celestial equator intersects the plane of the ecliptic. This point is known as the *First Point in Aries*, or the *Vernal equinox*, and is defined to be 0 hours of right ascension. Lines of constant longitude around the celestial equator are defined in terms of hours, minutes and seconds of time such that 24 hours corresponds to completing a great circle of 360° . Thus, the units of position on the celestial sphere are different in right ascension and declination. The angles in declination can be directly related to arcseconds but, in right ascension, a difference of $\Delta\alpha$ seconds of time corresponds to an angle of $15 \cos \delta \Delta\alpha$ arcseconds at declination δ .

We live in a spiral galaxy in which a disc of stars and gas rotates about the Galactic Centre. Because of the absence of dust extinction at radio wavelengths, the zero of the galactic coordinate system was defined in 1959 by radio observations of the distribution of neutral hydrogen in the inner galaxy and by the location of the compact non-thermal radio source Sagittarius A in the Galactic Centre (Blaauw *et al.*, 1959). Since that time,

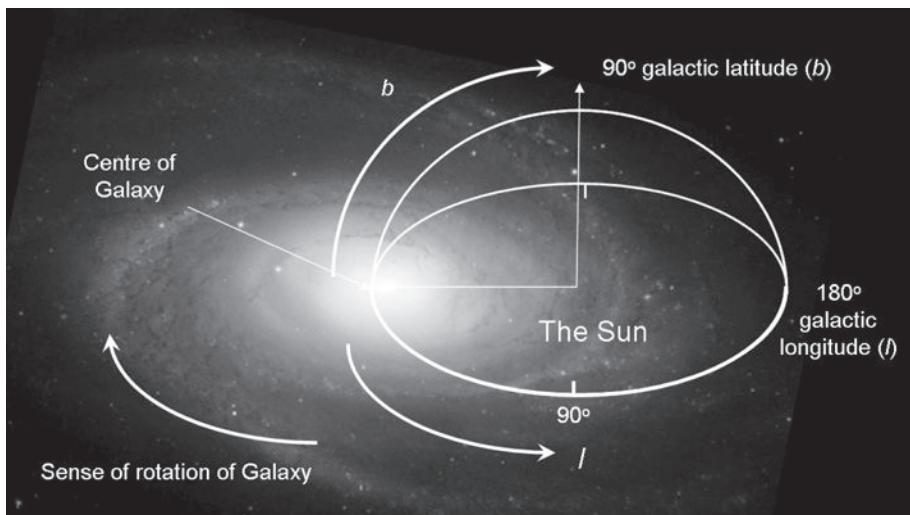
**Fig. A.1**

Illustrating the definition of the system of celestial coordinates used in astronomy. NCP and SCP stand for north and south celestial poles, respectively. The plane of the ecliptic is the plane in which the Earth orbits the Sun.

higher angular resolution observations of Sagittarius A have provided evidence for a very compact radio source Sagittarius A* (Sgr A*) which is now known to be coincident with a supermassive black hole. Sgr A* have provided evidence for a very compact radio source, Sagittarius A* (Sgr A*) is very close to the zero point of *galactic latitude* (b) and *longitude* (l) defined by Blaauw and his colleagues.

In the 2000.0 coordinate system, the Galactic Centre is defined as the direction $\alpha = 17^{\text{h}} 45^{\text{m}} 37.224^{\text{s}}$, $\delta = -28^{\circ} 56' 10.23''$. The Galactic equator is taken to be the great circle which passes through the centre of the Galactic plane as illustrated in Fig. A.2. *Galactic longitude* is defined in the sense that it increases in a counter-clockwise direction starting at $l = 0^{\circ}$ at the Galactic Centre, reaches 180° in the anti-Centre direction and returns to zero after completing a circuit of 360° . The *Galactic north pole* is defined to be the pole of the Galactic coordinate system which lies in the northern celestial hemisphere at celestial coordinates $\alpha = 12^{\text{h}} 51^{\text{m}} 26.282^{\text{s}}$; $\delta = +27^{\circ} 07' 42.01''$ in the 2000.0 coordinate system. The Galactic equator is at $b = 0^{\circ}$ and the Galactic north pole at $b = +90^{\circ}$; the Galactic south pole lies at $b = -90^{\circ}$. The inclination of the Galactic equator to the Earth's equator is 62.9 degrees. The intersection, or node line, of the two equators is at $\alpha = 18^{\text{h}} 51.4^{\text{m}}$, $\delta = 0^{\circ} 00'$ or, in Galactic coordinates, at $l = 33^{\circ}$ degrees, $b = 0^{\circ}$. The massive black hole in Sgr A* probably provides the best definition of the dynamical centre of the Galaxy and has galactic coordinates $l = 359^{\circ} 56' 39.4''$, $b = -0^{\circ} 2' 46.2''$.

In order to represent the images of the sky in different wavebands on a two-dimensional sheet of paper, the coordinate frame has to be distorted and two examples of how this can be achieved without excessive distortion of features present on the celestial sphere are considered here. Many more details of these and other types of projection of the celestial sphere onto different surfaces are described by Calabretta and Greisen (2002).

**Fig. A.2**

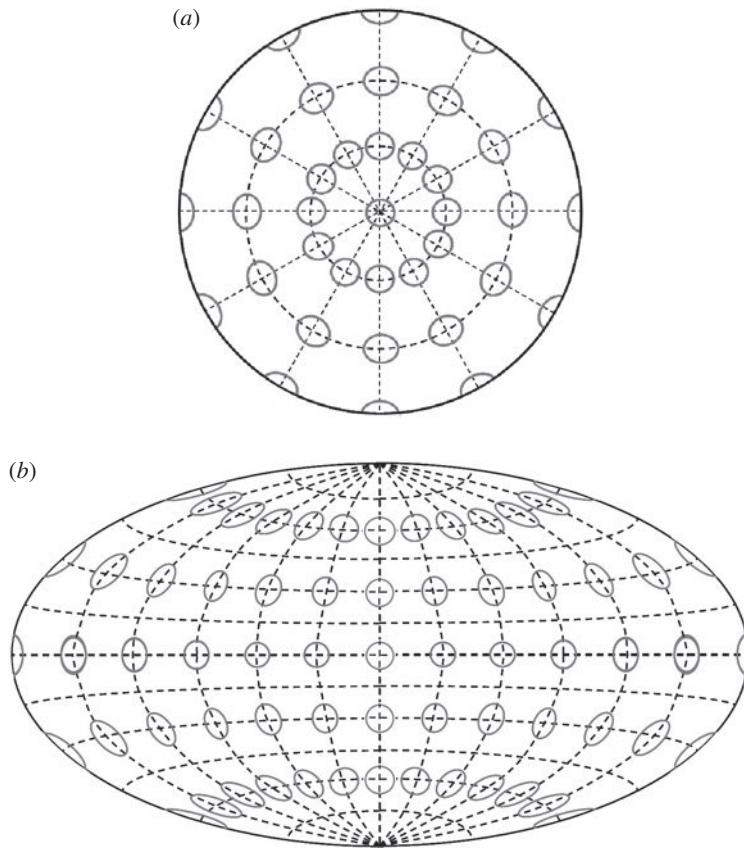
Illustrating the definition of the system of Galactic coordinates (l, b). The galaxy M81 has been rotated and inverted to simulate what our Galaxy might look like as seen by an external observer. (Image of M81 courtesy of NASA and the Space Telescope Science Institute.)

Figure 1.2a of Chapter 1 shows images of the northern and southern celestial hemisphere plotted in *equidistant azimuthal polar projections* or *zenith equidistant projections* photographed and created by Axel Mellinger from 51 wide-angle images. The northern celestial hemisphere on the left has the north celestial pole, $\delta = 90^\circ$, in the centre while the celestial equator, $\delta = 0^\circ$, is the bounding circle around the edge of the image. The right-hand image shows the southern celestial hemisphere, centred on the southern celestial pole at $\delta = -90^\circ$. Because two images have been used to span the whole sky, the distortions are not too great as shown by the diagram in Fig. A.3a. In both diagrams, the Milky Way is clearly seen as a broad band of mission cutting across both diagrams.

Another popular projection is the *Hammer–Aitoff projection* or *Aitoff projection* which shows the complete celestial sphere in a single diagram. This projection involves a reasonable compromise between the shape and scale distortions. The northern and southern Galactic poles ($b = \pm 90^\circ$) are at the top and bottom of the image (Fig. A.3b). The scale of Galactic longitude runs from 0° at the centre, which is the direction of the Galactic Centre, through $+180^\circ$ at the left of the image, the anti-Centre direction and then from $+180^\circ$ at the right of the image to 360° (or 0°) at the Centre. The distortions of the image caused by the projection from the surface of a sphere to a two-dimensional plane is shown in Fig. A.3b. These become large towards the northern and southern Galactic poles.

A.2 Distances in astronomy

The unit of distance used in astronomy is the *parallax-second* or *parsecs*. It is defined to be the distance at which the mean radius of the Earth's orbit about the Sun subtends an

**Fig. A.3**

(a) Illustrating the distortions of equal areas on the celestial sphere in an equidistant azimuthal polar or zenith equidistant projection. (b) Illustrating the distortions of the Hammer–Aitoff projection.

angle of one second of arc. In metres, the parsec, abbreviated to pc, is 3.0856×10^{16} m. For many purposes, it is sufficiently accurate to adopt $1\text{ pc} = 3 \times 10^{16}$ m. The parsec is a recognised SI unit and it is often convenient to work in kiloparsecs ($1\text{ kpc} = 1000\text{ pc} = 3 \times 10^{19}$ m), megaparsecs ($1\text{ Mpc} = 10^6\text{ pc} = 3 \times 10^{22}$ m) or even gigaparsecs ($1\text{ Gpc} = 10^9\text{ pc} = 3 \times 10^{25}$ m). Sometimes, it is convenient to measure distances in *light-years* which is the distance light travels in one year: $1\text{ light-year} = 9.4605 \times 10^{15}$ m. Thus, $1\text{ pc} = 3.26$ light-years. Another commonly used distance unit in astronomy is the *astronomical unit*, abbreviated to AU, which is the mean radius of the Earth's orbit about the Sun. $1\text{ AU} = 1.49578 \times 10^{11}$ m. The very nearest stars to the Earth are at a distance of about 1 pc and so they are about 2×10^5 times as far away as the Earth is from the Sun. These astronomical units are summarised in Table A.1, as well as approximate values for order of magnitude estimates.

Accurate distances are among the most difficult measurements to make in astronomy and there must, therefore, exist corresponding uncertainties in all the derived physical properties of astronomical objects. The most accurate direct distance measurements are

Table A.1 Astronomical units.

1 astronomical unit (AU)	$= 1.496 \times 10^{11}$ m	$\approx 1.5 \times 10^{11}$ m
1 parallax-second (parsec, pc)	$= 3.0856 \times 10^{16}$ m	$\approx 3 \times 10^{16}$ m
1 light year (ly)	$= 9.4605 \times 10^{15}$ m	$\approx 10^{16}$ m
1 solar mass (M_{\odot})	$= 1.989 \times 10^{30}$ kg	$\approx 2 \times 10^{30}$ kg
1 solar radius (R_{\odot})	$= 6.9598 \times 10^8$ m	$\approx 7 \times 10^8$ m
Luminosity of Sun (L_{\odot})	$= 3.90 \times 10^{26}$ W	$\approx 4 \times 10^{26}$ W

derived from measurements of the *parallaxes* of nearby stars, that is, the apparent motion of nearby stars against the background of very distant stars due to the motion of the Earth in its orbit about the Sun. A star at a distance of 1 pc is observed to move an angular distance of one arcsecond relative to the background stars when the Earth moves a distance equal to the Earth's mean radius about the Sun, perpendicular to the line of sight to the star. Even for the stars closest to the Sun, few distances are known with precision much better than about 5%. The accuracy with which this local distance scale is known has been greatly improved thanks to observations by the Hipparcos astrometric satellite of the European Space Agency.

To extend the distance scale from nearby stars to Galactic and extragalactic distances, *distance indicators* are used to provide *relative distances* for distant objects as compared to similar objects nearby. The complexities of undertaking these observations to extend the distance scale to cosmological distances have been reviewed by Rowan-Robinson (1985; 1988). The range of distances over which the different distance indicators can be used is shown in Fig. A.4. For example, *Cepheid variables* are used to extend the local distance scale from our own Galaxy to nearby galaxies. The Cepheid variable stars are pulsating stars which have characteristic light curves, that is, the variation of their luminosities with the phase of the pulsation. In 1912, Henrietta Leavitt showed that there exists a tight correlation between the intrinsic luminosity of the variable star and its pulsation period (Leavitt, 1912). Therefore, if the pulsation periods of Cepheid variables in nearby galaxies can be measured, their intrinsic luminosities L can be measured and then, by measuring their flux densities S , their distances r can be found from the inverse square law $r = (L/4\pi S)^{1/2}$.

The use of distance indicators can be extended to extragalactic distances using other 'standard' properties of galaxies or the brightest star systems in them, as indicated in Fig. A.4. On the very largest scale, relative distances can be measured using the *velocity-distance relation* for galaxies. In 1929, Hubble showed that the system of galaxies is not stationary but is expanding uniformly such that, nearby, the distance of a galaxy r from our own Galaxy is proportional to its velocity of recession v away from it, $v = H_0 r$, where H_0 is Hubble's constant (Hubble, 1929). Hubble's constant is often written $H_0 = 100h$ km s⁻¹ Mpc⁻¹ because of uncertainties in the absolute value of Hubble's constant – the present best value of h is about 0.73 with a 1- σ uncertainty of about 10%.

The problem of the accumulating errors using distance indicators can be avoided if astrophysical methods are used to estimate physical sizes and distances directly at large

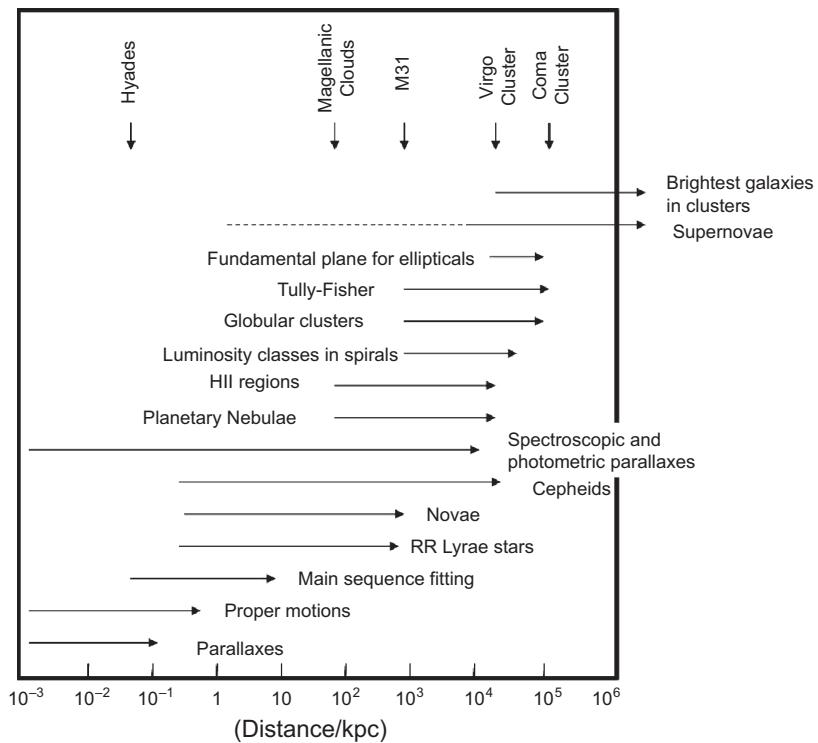


Fig. A.4 Illustrating the 'cosmological distance ladder' (Rowan-Robinson, 1985, 1988). The diagram shows roughly the range of distances over which different classes of object can be used to estimate astronomical distances. The diagram has been redrawn and updated from Rowan-Robinson's presentation.

distances. If the angular size $\Delta\theta$ of some object is measured for which the physical size d can be estimated by physical arguments, the distance r is found immediately from $r = d/\Delta\theta$. An example of this approach is the *Baade–Wesselink method* to measure distances to supernova explosions in distant galaxies. In a simple approximation, the expanding sphere may be considered a black-body at temperature T_1 . As the sphere expands from diameter d_1 to d_2 , its luminosity changes because the sphere has a larger surface area at some new temperature T_2 . By measuring the change in temperature and flux density of the supernova, the change in angular size $\Delta\theta$ of the supernova can be measured using the Stefan–Boltzmann law since at t_1 the flux density is $S_1 = (\sigma T_1^4/4)(d_1/r)^2 = (\sigma T_1^4/4)\theta_1^2$ and at t_2 $S_1 = (\sigma T_2^4/4)\theta_2^2$. The physical size of the expansion ($d_2 - d_1$) can be measured from the velocity of expansion of the shell v and the times t_1 and t_2 between which the temperatures T_1 and T_2 were measured, $(d_2 - d_1) = 2v(t_2 - t_1)$. Hence the distance r can be found. This method can potentially be used to measure very large cosmological distances directly.

Another method of measuring directly physical sizes for distant objects is to use the Sunyaev–Zeldovich effect as applied to the very hot gas observed in clusters of galaxies (see Sect. 4.6). In this case, the pressure of the gas can be measured from the magnitude of the Sunyaev–Zeldovich decrement in the Rayleigh–Jeans region of the spectrum of the

Cosmic Microwave Background radiation,

$$\frac{\Delta T}{T} = -2\sigma_T \int \left(\frac{kT}{m_e c^2} \right) N_e dl \quad (\text{A.1})$$

and the bremsstrahlung spectrum and intensity of the X-ray emission provide estimates of the temperature and the quantity

$$L_X \propto \int n_e^2 T^{-1/2} dl . \quad (\text{A.2})$$

The properties of the hot gas cloud are overdetermined and so its physical size can be found. Then, by measuring its angular size, the distance of the cluster can be found.

Another direct physical method of measuring extragalactic distances is to use the properties of *gravitational lenses*. If the background source is variable, then the different images of the source display the same variability but with time delays depending upon the geometry of the lensing galaxy. From detailed modelling of the distorted high resolution images observed by long baseline interferometry, the gravitational potential of the lensing galaxy can be found and physical dimensions of the lensed image can be deduced.

The great advantage of the last three methods is that they eliminate the necessity of measuring distances through the hierarchy of the cosmological distance ladder. The best estimates of Hubble's constant from the last two methods are in good agreement with the values $h = 0.73$ found from the Hubble Space Telescope Key project and from modelling the power spectrum of intensity fluctuations in the Cosmic Microwave Background Radiation.

A.3 Masses in astronomy

It is convenient to describe the masses of celestial objects in terms of the mass of the Sun. 1 solar mass, written $1 M_\odot$, is $1.989 \times 10^{30} \approx 2 \times 10^{30}$ kg. Jupiter has mass $M_J \approx 0.001 M_\odot$, our own Galaxy has mass about $10^{11} M_\odot$ and the Coma Cluster of galaxies has mass about $2 \times 10^{15} M_\odot$.

Most direct methods of measuring masses in astronomy involve the combination of Newton's laws of motion with Newton's law of gravity and the assumption that the system is in statistic equilibrium under gravity. For the Sun, for example, its mass is found by equating the centripetal accelerations of the planets to their gravitational accelerations due to the Sun $v^2/r = GM_\odot/r^2$, $M_\odot = v^2 r / G$. For binary star systems, similar methods can be used. For a cluster of stars or galaxies, the total mass is found using the *viral theorem* which states that, for a system in dynamical equilibrium, the internal kinetic energy is equal to half its gravitational potential energy, $\frac{1}{2}M\langle v^2 \rangle = \frac{1}{2}GM^2/r_s$, $M \approx r_s \langle v^2 \rangle / G$ where $\langle v^2 \rangle$ is the mean square velocity and r_s is some suitably chosen radius.¹ Variants on this approach use the fact that different types of observation provide evidence on the *mass distribution* in the system, for example, by determining the mass distribution of hot gas in the deep

¹ This general result is proved in Sect. 3.3 of my book *Galaxy Formation* (Longair, 2008).

gravitational potential wells in clusters of galaxies, or by studying the distortions of the images of background galaxies by the mass within systems along the line of sight.

A.4 Flux densities, luminosities, magnitudes and colours

The standard measure of the power received from a celestial object is its *flux density* S_ν , which is the energy incident per second per unit area per unit frequency band at the telescope. Thus, if an energy ΔE is detected in a bandwidth ν to $\nu + \Delta\nu$ in time Δt and A is the collecting area of the telescope, which is assumed to be 100% efficient at detecting the radiation, the flux density of the source at frequency ν is defined to be $S_\nu = \Delta E / (A\Delta t \Delta\nu)$. The SI unit of flux density is $\text{W m}^{-2} \text{ Hz}^{-1}$. If this flux density originates within solid angle $d\Omega$ on the sky, the *intensity* of the source is $I_\nu = S_\nu / \Delta\Omega$; the SI unit of intensity is $\text{W m}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1}$. The flux densities of astronomical objects are very small and so radio astronomers introduced the unit $10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$ which is known as 1 jansky, abbreviated to Jy. Fainter sources can be described in millijanskys ($10^{-3} \text{ Jy} = 10^{-29} \text{ W m}^{-2} \text{ Hz}^{-1}$) and microjanskys ($10^{-6} \text{ Jy} = 10^{-32} \text{ W m}^{-2} \text{ Hz}^{-1}$). In X-ray and γ -ray astronomy, flux densities are often quoted in terms of the flux of photons rather than of energy, so that the units photons per second per unit area per unit energy range, for example, photons $\text{s}^{-1} \text{ m}^{-2} \text{ keV}^{-1}$, are often used. In optical and infrared astronomy, flux densities and intensities are often quoted per unit wavelength I_λ rather than per unit frequency. Since $I_\nu d\nu = I_\lambda d\lambda$, it follows that $I_\nu = cI_\lambda/\nu^2$.

In optical and infrared astronomy, although flux densities are becoming commoner, the traditional system of *optical magnitudes* still dominates the astronomical literature. The underlying reason for this is that flux densities and intensities are generally measured through rather broad filter bands which can contain a number of spectral features. The astronomical magnitudes thus measure the total power received within a given filter band.

The concept of magnitudes was introduced by the Greek astronomers who placed stars into five ‘magnitude’ classes, the magnitudes being assigned on the basis of the brightnesses of the stars to the naked eye. Fifth magnitude stars were the faintest objects visible. The system was put on a systematic basis by Pogson in 1854, who showed that the magnitude scale was logarithmic in flux density, like many physiological sensations, and that the magnitude scale of the ancients could be approximated by a rule in which five logarithmic magnitudes correspond to a factor of 100 in flux density.

Flux densities are converted to *apparent magnitudes* m which are negative logarithmic measures of flux density defined by

$$m = \text{constant} - 2.5 \log_{10} S \quad (\text{A.3})$$

where S is the flux density of the source. Thus, a difference of five magnitudes corresponds to a factor of 100 in flux density. The magnitude system is normalised so that a standard star, chosen to be the bright star Vega, or α -Lyrae, in the constellation of Lyra, has zero magnitude at all wavelengths. In this way, magnitudes can be defined at all wavelengths (see Table A.2). The very brightest stars in the sky have magnitude about 0. The

Table A.2 The characteristics of some of the optical and infrared wavebands used in ground-based astronomy.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>U</i>	0.365	8.2×10^{14}	0.068	4.2×10^{-8}	1.88×10^{-23}	22	150	10
<i>B</i>	0.44	6.8×10^{14}	0.098	7.2×10^{-8}	4.64×10^{-23}	23	100	10
<i>V</i>	0.55	5.5×10^{14}	0.089	4.0×10^{-8}	3.95×10^{-23}	22	170	XX
<i>R</i>	0.70	4.3×10^{14}	0.22	1.8×10^{-8}	2.87×10^{-23}	21	250	55
<i>I</i>	0.90	3.3×10^{14}	0.24	8.3×10^{-9}	2.24×10^{-23}	18.5	1.5×10^{3b}	370
<i>J</i>	1.25	2.4×10^{14}	0.28	3.07×10^{-9}	1.60×10^{-23}	16	1.0×10^{4b}	2.8×10^3
<i>H</i>	1.65	1.8×10^{14}	0.30	1.12×10^{-9}	1.02×10^{-23}	13	5.6×10^{4b}	1.7×10^4
<i>K</i>	2.2	1.4×10^{14}	0.42	4.07×10^{-10}	6.57×10^{-24}	12.5	4.4×10^4	1.8×10^4
<i>L</i>	3.45	8.7×10^{13}	0.60	7.30×10^{-11}	2.90×10^{-24}	5.5	8.0×10^6	5.0×10^6
<i>M</i>	4.7	6.4×10^{13}	0.67	2.12×10^{-11}	1.63×10^{-24}	2	1.0×10^8	7.0×10^7
<i>N</i>	10.2	2.9×10^{13}	5.2 ^a	1.10×10^{-12}	3.90×10^{-25}	-3	1.0×10^9	5.0×10^9
<i>Q</i>	20.0	1.5×10^{13}	5.2 ^a	7.80×10^{-14}	1.04×10^{-25}	-5	6.0×10^8	3.0×10^9

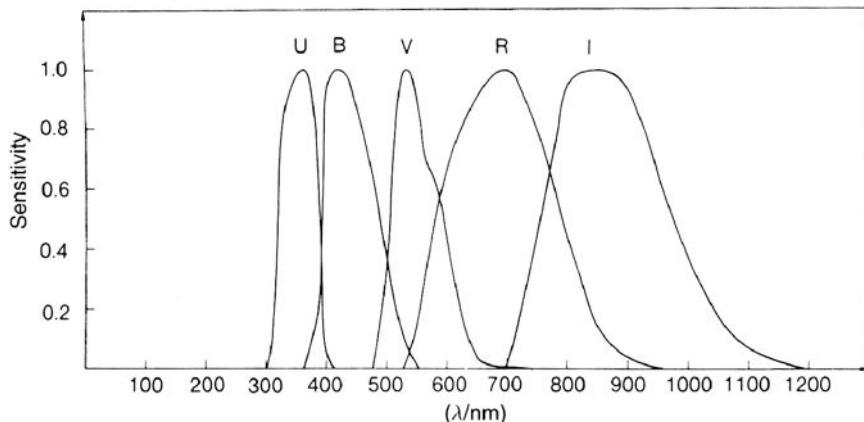
^a These bands may be observed with narrower filters to reduce the background intensity.

^b In these wavebands the background is dominated by hydroxyl (OH) airglow emission lines.

Column Labels

- (1) Name of waveband
- (2) Effective wavelength of the waveband ($\lambda_{\text{eff}}/\mu\text{m}$).
- (3) Effective frequency of the waveband ($\nu_{\text{eff}}/\text{Hz}$).
- (4) Effective width of the band ($\Delta\lambda_{\text{eff}}/\mu\text{m}$).
- (5) Flux density of a zero magnitude star per unit wavelength ($S_{\lambda}(0)/\text{W m}^{-2} \mu\text{m}^{-1}$).
- (6) Flux density of a zero magnitude star per unit frequency range ($S_{\nu}(0)/\text{W m}^{-2} \text{Hz}^{-1}$).
- (7) Background intensity in magnitudes arcsec⁻².
- (8) Background photon intensity per unit waveband
($I(\lambda)/\text{photons m}^{-2} \text{arcsec}^{-2} \text{s}^{-1} \mu\text{m}^{-1}$).
- (9) Background photon intensity in standard waveband given in column (4) ($I/\text{photons m}^{-2} \text{arcsec}^{-2} \text{s}^{-1}$).

The background intensities should be used as a rough guide in making estimates of the observing time necessary to reach a given limiting magnitude. In the above table, it is assumed that the observations are made from a good dark site in the absence of scattered moonlight. In the infrared waveband at wavelengths $\lambda \geq 2 \mu\text{m}$, it is assumed that the telescope and the atmosphere are at a temperature of 288 K and that the emissivity of the telescope is 8%. Lower background fluxes are obtained if narrow filters are used. In addition, at those wavelengths at which air glow is the dominant source of noise, lower background fluxes can be obtained by observing between the strong OH emission lines. This is now possible with the latest generation of infrared spectrometers.

**Fig. A.5**

The transmission functions of some of the standard filters for the 0.3–1 μm waveband (see also Table A.2) (Karttunen *et al.*, 2007).

faintest stars which can be seen with the naked eye have $m \approx 5$. The faintest stars which can now be observed with a 4 metre telescope in a 5 minute observation using a CCD camera have magnitudes about 25. In the longest exposures made by the Hubble Space Telescope in the Hubble Ultra-Deep Field, galaxies as faint as 32 were observed in the I waveband.

In practice, the flux density S is measured within some range of frequencies, ν_1 to ν_2 , determined by the transmission of the atmosphere, the properties of the telescope mirror and the instrumentation of the telescope as a function of frequency. In general, observations are made through a filter with a certain transmission function T_ν which describes the fraction of the energy incident upon the filter which is transmitted to the detector as a function of frequency ν . Thus, for a particular filter i , $S_i = \int_0^\infty T_i(\nu) S_\nu \, d\nu$. If $T_i(\nu) = 1$ for all frequencies, the flux density S_i is called a *bolometric flux density* or, converting to magnitudes, a *bolometric apparent magnitude*. This magnitude corresponds to the total radiation emitted by the source at all wavelengths.

At optical and infrared wavelengths, a number of standard magnitudes systems have been established by adopting particular filters as standards. In the days of photographic astronomy, the filters were determined largely by the sensitivity response of the photographic emulsions and the available filter materials. Nowadays, a wide range of filter transmission functions are in use, examples of those which span the optical and near infrared wavebands being illustrated in Fig. A.5. For each filter, an effective central wavelength (or frequency) and effective band pass (λ_{eff} , $\Delta\lambda_{\text{eff}}$) can be defined. Values for these quantities as well as typical background intensities and zero points in each of the wavebands, corresponding to the flux density of Vega in these wavebands, are listed in Table A.2. This table is useful for rough calculations but it must be stressed that these figures assume that the spectrum of the object within the filter passband is the same as α -Lyrae. In precise work, it is essential to use the transmission function $T_i(\nu)$ for each filter and the spectrum of the object $S(\nu)$ to derive precise magnitudes for comparison with theory.

The intrinsic *luminosities* of cosmic sources are related to their flux densities by

$$L = 4\pi r^2 S, \quad (\text{A.4})$$

where r is the distance of the source. This is the total emitted power of the source, assuming that the radiation is emitted isotropically by the source, and is measured in units of W Hz^{-1} . If S is the *bolometric flux density* of the source, L is the *bolometric luminosity* of the source and is measured in W .

In optical and infrared astronomy, astronomers often use *absolute magnitudes* M . These are defined to be the magnitudes which the objects would have if they were placed at a distance of 10 pc. An object of intrinsic luminosity L has flux density S and therefore, for any object,

$$M = m - 5 \log_{10}(r/10) \quad (\text{A.5})$$

where the distance r is measured in pc. For stars of different luminosities,

$$M = M_\odot - 2.5 \log_{10}(L/L_\odot). \quad (\text{A.6})$$

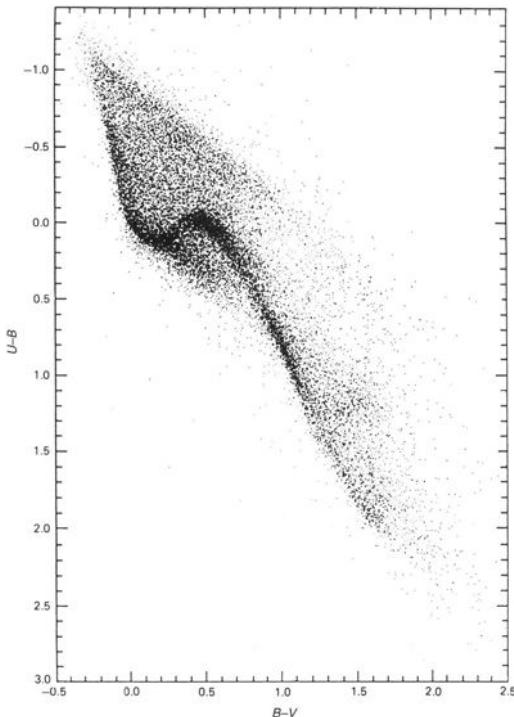
Just as in the case of apparent magnitudes, the absolute magnitudes can either refer to a specific wavelength or waveband or else to a bolometric absolute magnitude which corresponds to the total luminosity of the object integrated over all wavelengths. The absolute bolometric magnitude of the Sun is $M_\odot = 4.75$ and hence in general we can write

$$M = 4.75 - 2.5 \log_{10}(L/L_\odot), \quad (\text{A.7})$$

where L is the bolometric luminosity of the object.

The luminosities of celestial objects are conveniently described in terms of the *bolometric luminosity* of the Sun. The bolometric luminosity of the Sun, written as L_\odot , is the total luminosity of the Sun integrated over all wavelengths. Its value is $L_\odot = 3.90 \times 10^{26}$ W and is almost entirely emitted in the ultraviolet, optical and near-infrared wavebands. Astronomers use units of L_\odot to describe the luminosities of objects far outside those wavebands in which the Sun's luminosity is emitted. For example, the far-infrared luminosities of regions of star formation and the X-ray luminosities of binary X-ray sources are often quoted in units of L_\odot and this provides a useful measure of the energetics of these sources relative to the optical/infrared luminosity of the Sun.

Astronomers define the *colours* of stars in terms of the differences in their magnitudes at different wavelengths. For example, a commonly used colour is the difference between the magnitudes of a star in the blue (B) waveband (centred on 440 nm) and the visual (V) waveband (centred at 550 nm). This *colour index* or *colour* ($B - V$) is a measure of how blue or red the star is. Since stars may be thought of as black-body emitters in the simplest approximation, blue stars are hot and red stars are cool. This statement is quantified by working in terms of colours such as $B - V$. The use of colour provides a means of discriminating the general spectral shapes of different classes of object. A common way of displaying this information is in terms of a two-colour diagram in which, for example, $(U - B)$ is plotted against $(B - V)$. Figure A.6 is an example of a two-colour diagram for 46 000 stars. Colours such as $(U - B)$ or $(B - V)$ are strong functions of the surface temperature of the star and hence, in plotting a luminosity–temperature diagram, it

**Fig. A.6**

A two-colour plot of $(U - B)$ against $(B - V)$ for 46 000 stars. Most stars lie along the S-shaped locus which is defined by stars lying on the main sequence. Hot blue stars are located at the top left of the diagram and cool red stars to the bottom right. If stars radiated like black-bodies, they would lie on a locus stretching smoothly from the bottom right to top left of the diagram. The 'kink' in the relation is due to the effect of the Balmer continuum absorption which causes the spectra to deviate from those of black-bodies. The effect of reddening by interstellar dust is to move the colours of each star towards the bottom right corner of the diagram. The rather sharp upper envelope to the distribution of hot stars above the main sequence follows precisely this reddening line (Nicolet, 1980).

is more convenient to plot one of these colours rather than the surface temperature of the star which requires a detailed knowledge of radiative transport through the stellar atmosphere.

A.5 Diffraction-limited telescopes

The angular resolving power of a telescope is defined by the angular distribution of the intensity of radiation received from a point source at infinity. In the case of telescopes for the hard X-ray and γ -ray wavebands, the angular resolving power is determined by the geometric optics of the collimator. In the ultraviolet, optical, infrared and radio wavebands, the angular resolving power is determined by diffraction optics.²

² The classic reference for understanding the details of diffraction-limited optics is the text *Principles of Optics* by Born and Wolf (1999). See, in particular, Chap. 8, *Elements of the Theory of Diffraction*

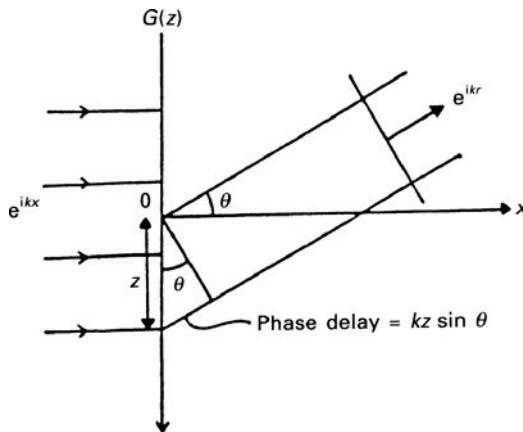


Fig. A.7 Illustrating the phase delay associated with the constructive interference of waves at an angle θ with respect to the direction perpendicular to the aperture.

Consider a one-dimensional aperture with transmission function $G(z)$, that is, the fraction of the electromagnetic radiation incident in the distance interval dz at the position z on the aperture which is transmitted by the aperture is $G(z)$ (Fig. A.7). Using Huygen's principle, the wavefront of the radiation can be replaced by a superposition of point sources of spherical waves. We can therefore represent the light transmitted at z by a point source of spherical waves of amplitude, or field strength, $G(z) dz$. The phase delay of the radiation in the direction θ with respect to the normal vector through the centre of the aperture is $kz \sin \theta$ where $k = |\mathbf{k}|$ is the wavevector of the radiation transmitted by the aperture. The field strength of the radiation in the direction θ from the increment dz of the aperture is therefore proportional to $G(z) \exp(-ikz \sin \theta) dz$ and the total field strength is

$$U(\theta) = \int_{-\infty}^{\infty} G(z) \exp(-ikz \sin \theta) dz . \quad (\text{A.8})$$

In the limit of small angles, $\sin \theta \approx \theta$ and hence

$$U(\theta) = \int_{-\infty}^{\infty} G(z) \exp(-ik\theta z) dz . \quad (\text{A.9})$$

In this limit, $U(\theta)$ is the Fourier transform of the aperture distribution $G(z)$. Extending this analysis to two dimensions for a rectangular aperture in the $y-z$ plane with corresponding angles ϕ and θ , we find

$$U(\theta, \phi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(y, z) \exp[-ik(\phi y + \theta z)] dy dz . \quad (\text{A.10})$$

For the case of a uniformly illuminated one-dimensional aperture of length $2a$, $G(z) = 1$ for $-a \leq z \leq a$ and is zero otherwise. Therefore

$$U(\theta) = \int_{-a}^{a} \exp(-ik\theta z) dz = 2 \frac{\sin(k\theta a)}{k\theta} .$$

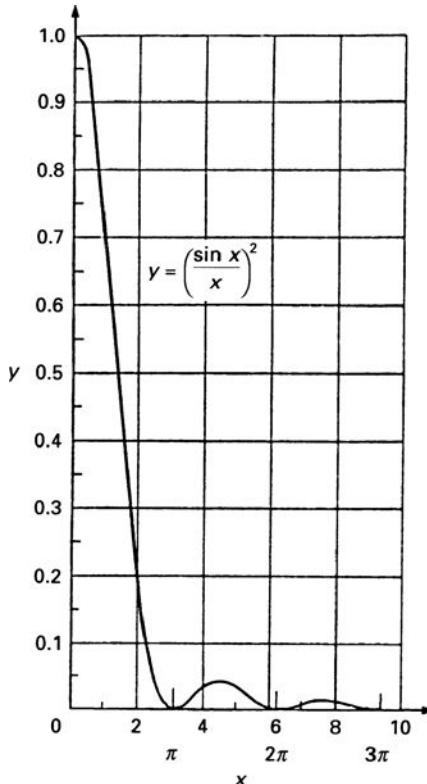


Fig. A.8 The diffraction pattern of a rectangular one-dimensional aperture, where $x = k\theta a$ and $I_0 = 1$ (Born and Wolf, 1999).

To find the intensity of radiation we take the square of the field strength

$$I(\theta) \propto U^2(\theta) \propto \frac{\sin^2(k\theta a)}{(k\theta)^2}.$$

Normalising this intensity distribution to the central intensity I_0 at $\theta = 0$,

$$I(\theta) = I_0 \left[\frac{\sin(k\theta a)}{k\theta a} \right]^2. \quad (\text{A.11})$$

By exactly the same procedure, a rectangular aperture of dimensions $2a \times 2b$ has diffraction pattern

$$I(\theta, \phi) = I_0 \left[\frac{\sin(k\theta a)}{k\theta a} \right]^2 \left[\frac{\sin(k\phi b)}{k\phi b} \right]^2. \quad (\text{A.12})$$

The expressions (A.11) and (A.12) are the well-known *Fraunhofer diffraction patterns* for a single slit and a rectangular slit, respectively, the function $[\sin(k\theta a)/(k\theta a)]^2$ being displayed in Fig. A.8. The term Fraunhofer diffraction is used when the approximation $\sin \theta \approx \theta$ can be applied and higher off-axis terms in θ^2 are neglected. More generally, the condition for Fraunhofer diffraction is that the input and output beams from the aperture

have the same curvature, the simplest case being that in which both beams are parallel at infinity, as in our treatment.

As can be seen from Fig. A.8, the first zero of the diffraction pattern occurs at $k\theta a = \pi$ and the intensity of radiation falls to half its central intensity at $k\theta a = 1.39$. In terms of the wavelength λ , this occurs when

$$\theta_{1/2} = 0.44\lambda/2a . \quad (\text{A.13})$$

This result means that the phases of the waves only add constructively within an angle $\theta_{1/2}$ to produce a strong signal. At greater angles, the phases of the waves become progressively more and more out of step and do not interfere constructively.

The calculation can be repeated for a circular aperture, the analogue of the circular mirror or lens of an optical telescope. First, the expression (A.10) is transformed into polar coordinates (r, ψ) such that the point (y, z) on the aperture is given by $y = r \cos \psi$, $z = r \sin \psi$ and the element of area is $r dr d\psi$. Correspondingly, the angles on the sky are transformed to the corresponding angular coordinates (ρ, σ) such that $\phi = \rho \cos \sigma$, $\theta = \rho \sin \sigma$. ρ measures the angular direction from the axis of the circular aperture and σ the azimuthal direction. Integrating over the complete (y, z) or (r, ψ) plane, (A.10) becomes

$$\begin{aligned} U(\rho, \sigma) &= \int_0^\infty \int_0^{2\pi} G(r, \psi) \exp[-ik\rho r(\cos \psi \cos \sigma + \sin \psi \sin \sigma)] d\psi r dr \\ &= \int_0^\infty \int_0^{2\pi} G(r, \psi) \exp[-ik\rho r \cos(\psi - \sigma)] d\psi r dr , \end{aligned} \quad (\text{A.14})$$

where the radius of the aperture is a . For a uniformly illuminated aperture, $G(r, \psi) = 1$ for $r \leq a$ and $G(r, \psi) = 0$ at $r > a$. In this case, the diffraction pattern is symmetrical about the axis $\rho = 0$ and so σ can be set to any fixed value, which can be taken to be zero for convenience,

$$U(\rho) = \int_0^a \int_0^{2\pi} \exp(-ik\rho r \cos \psi) d\psi r dr , \quad (\text{A.15})$$

The ψ integral is then evaluated using the integral representation of the Bessel function $J_0(z)$,

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} \exp(-iz \cos \alpha) d\alpha .$$

Then

$$U(\rho) = \frac{2\pi}{(k\rho)^2} \int_0^a J_0(x)x dx ,$$

where $x = k\rho r$.

Using the Bessel function relation

$$\frac{d}{dx}[x^{n+1} J_{n+1}(x)] = x^{n+1} J_n(x) ,$$

for the case $n = 0$, we find

$$U(\rho) = \frac{2\pi a}{k\rho} J_1(k\rho a) = 2A \left[\frac{J_1(k\rho a)}{k\rho a} \right], \quad (\text{A.16})$$

where $A = \pi a^2$ is the area of the aperture. As before, the intensity is found by taking the square of the field strength and, normalising to the central intensity I_0 ,

$$I(\rho) = I_0 \left[\frac{2J_1(k\rho a)}{k\rho a} \right]^2. \quad (\text{A.17})$$

Note that in the limit of small values of x , $J_1(x)/x \rightarrow \frac{1}{2}$.

The diffraction pattern (A.16) is shown in Fig. A.9a and is the well-known *Airy pattern* of a circular aperture. Outside the central bright region, which is known as the *Airy disc*, there are concentric rings which contain a small but significant fraction of the energy transmitted by the aperture. This is illustrated in Fig. A.9b in which the total encircled energy is displayed as a function of angular distance from the axis of the aperture,

$$f(\rho) = \frac{\int_0^\rho I(\rho)\rho \, d\rho}{\int_0^\infty I(\rho)\rho \, d\rho}.$$

It can be seen that 84% of the light lies within the first dark ring and more than 90% within the second. This function, the *encircled energy fraction*, is an important measure of the performance of a telescope system, particularly when high precision photometry is required.

The first dark ring of the Airy diffraction pattern occurs at $k\rho a = 1.220\pi = 3.883$. Expressing the angle ρ in terms of the wavelength λ and the radius of the aperture a , we find

$$\rho = 1.22\lambda/D,$$

where $D = 2a$ is the diameter of the aperture. This corresponds to the *Rayleigh criterion* for resolving two nearby point sources. The criterion corresponds to one object being located on the first dark ring of the Airy diffraction pattern of the other.

The point spread function of the aperture depends upon the function $G(z)$ and so can be modified by adopting different forms of the transmission function $G(z)$. For example, if $G(z)$ is taken to be a Gaussian function $G(z) \propto \exp(-\rho^2/2\rho_0^2)$, the point spread function $I(\rho)$ is also of Gaussian form because the Fourier transform of a Gaussian function is just another Gaussian. A Gaussian transmission function broadens the point spread function but it eliminates the dark rings around the Airy disc. Thus, there is a trade-off between angular resolution and eliminating the dark rings in the Airy pattern. In radio astronomy, the rings external to the Airy disc are called *side-lobes* of the main beam and can be eliminated or strongly attenuated by ‘grading the aperture’. In optical parlance, the modification of the point spread function by changing the function $G(z)$ is called *apodisation*.

The success of an imaging system is crucially dependent upon the quality of the optical system and this is often quoted in terms of its *wavefront error*. If there are random phase errors imposed on the wavefront, the amplitudes of the waves add together slightly out of phase. This has two consequences – the central intensity is reduced and the point spread

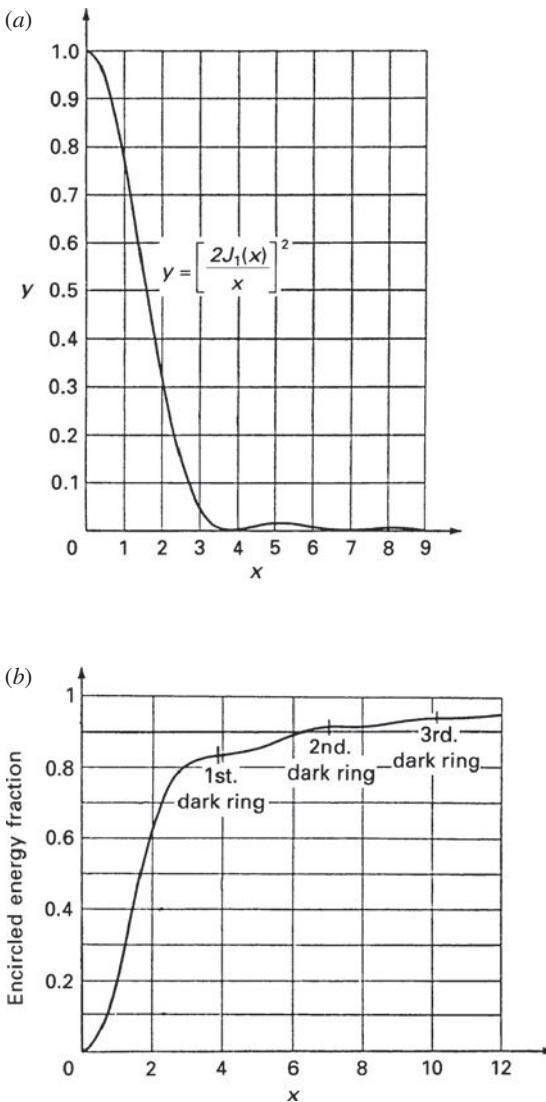
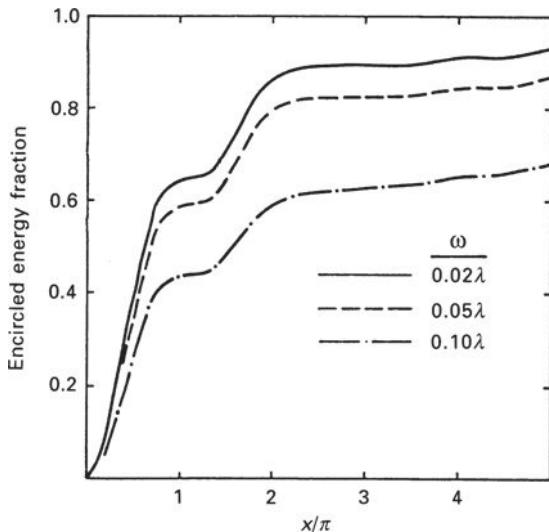


Fig. A.9 (a) The radial variation of intensity of the diffraction pattern of a uniform circular aperture. (b) The encircled energy fraction as a function of angular distance from the axis of a uniform circular aperture (Born and Wolf, 1999).

function is broadened – power is removed from the central maximum and scattered into the wings of the point spread function. Figure A.10 shows that there is a significant decrease in the central intensity if the random phase errors are even as small as $\lambda/10$. The usual criteria for a diffraction-limited system is that the random phase errors should amount to less than $\lambda/20$.

In optical astronomy, diffraction-limited performance is not achieved without taking special measures. The atmosphere above the telescope is not smooth – there is small-scale turbulence which gives rise to fluctuations in the refractive index of the atmosphere and these turbulent cells cause the image of a point source to be scattered in angle. This

**Fig. A.10**

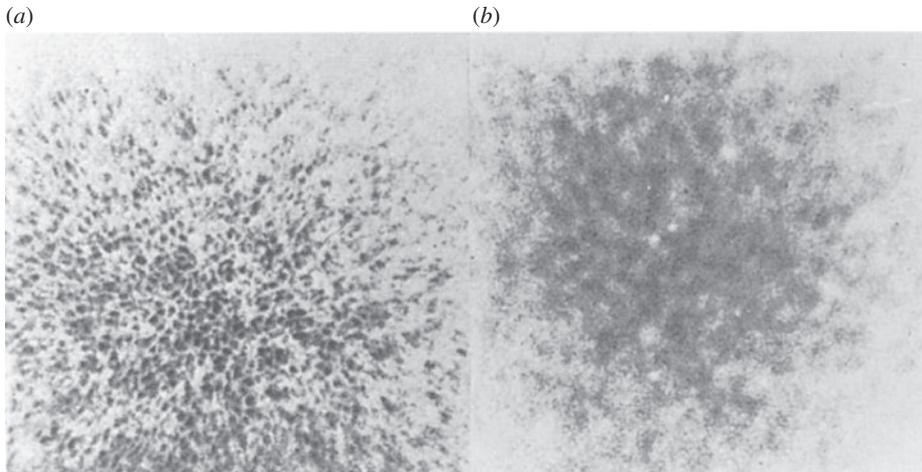
The encircled energy fraction as a function of angle from the axis of a Cassegrain telescope and its dependence upon the rms surface errors of the surface of the primary mirror. The radius of the central hole is 33% of that of the primary mirror. The solid line shows the encircled energy fraction for a perfect mirror. The dashed line shows the response when the rms surface errors on small scales is $\lambda/20$. The dot-dashed line shows the response if the rms error is $\lambda/10$. Small scale irregularities in the surface of the mirror result in energy being removed from the main beam and scattered into the wings of the diffraction pattern. These calculations have been carried out for the mirror of the Hubble Space Telescope by D. J. Schroeder (2000).

phenomenon is illustrated by very short exposure images of bright stars (Fig. A.11). If a star is photographed with an exposure of about 10^{-2} s, its image appears as a random superposition of speckles, each speckle corresponding to the image formed by the light passing coherently through one of the turbulent cells. This phenomenon of *astronomical seeing* is the prime cause of degradation of the diffraction-limited performance of all large optical-infrared telescopes. On a good site, the size of the ‘seeing disc’ can be as small as 0.4 arcsec. Thus, for a 5 metre class telescope, the degradation of the image quality corresponds to about a factor of 20 as compared with its theoretical angular resolution.

If the seeing disc is characterised by an angular scale θ_s , this characteristic angular size decreases slowly with increasing wavelength as $\theta_s \propto \lambda^{-0.2}$ so that if the seeing disc is typically about 1 arcsec at 500 nm, it is expected to be only about 0.6 arcsec at 5 μm and 0.5 arcsec at 10 μm .

To improve the imaging properties of optical-infrared telescopes, *adaptive optics* can be employed to eliminate in real time the phase fluctuations induced by the atmosphere. Flexible mirrors are used to compensate in real time for the distortions imposed on the wavefronts of the incoming waves and thus provide diffraction-limit performance.

One way of avoiding the problems of astronomical seeing is to place the telescope above the Earth’s atmosphere, as has been spectacularly achieved by the NASA–ESA Hubble Space Telescope. Being located above the Earth’s atmosphere, the problem of astronomical seeing is eliminated and the whole of the ultraviolet waveband is opened up for observation.

**Fig. A.11**

Illustrating the phenomenon of astronomical seeing. (a) A very short exposure (0.02 s) of an unresolved star showing the splitting up of the image into a large number of ‘speckles’. Roughly speaking, the angular size of each of the speckles corresponds to the diffraction limit of the aperture. (b) A similar very short exposure of the giant star Betelgeuse illustrating the difference in speckle pattern as compared with (a) when the star has an angular size greater than the diffraction limit of the telescope. In both images (a) and (b), the overall angular size of the image is about 2 arcsec. When observations are made with longer integration times, the speckles overlap and the image is blurred with a typical full-width half-maximum of about 1 arcsec (Labeyrie, 1978).

A.6 Interferometry and synthesis imaging

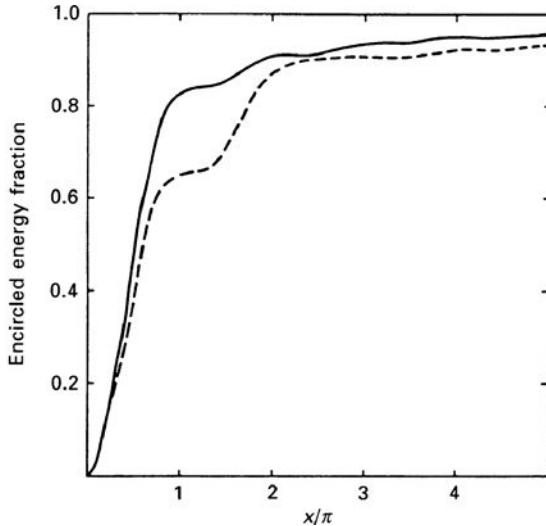
The simplest way of extending the above concepts to interferometers, aperture synthesis, speckle interferometry, and so on, is to use the theorem derived in expression (A.9) that the amplitude of the response to a point source is the two-dimensional Fourier transform of the aperture distribution. We can determine the response to a distant source using various theorems in Fourier analysis (Bracewell, 1986).

Let us first work out the point spread function for a mirror with a circular hole in the centre, such as is adopted in the Cassegrain configuration adopted in many modern telescopes, including the Hubble Space Telescope. The amplitude of the response to a point source can be found using the *addition theorem* for Fourier transforms according to which

If $f(x)$ and $g(x)$ have Fourier transforms $F(s)$ and $G(s)$, then the Fourier transform of $f(x) + g(x)$ is $F(s) + G(s)$.

Thus, by subtracting the Fourier transform of the central hole from that of the full mirror and then squaring the result we can find the polar diagram of the mirror and its encircled energy fraction (Fig. A.12).

Let us consider next the case of a two-element interferometer in one dimension which can be represented by two uniform apertures separated by a distance D (Fig. A.13a). To describe this aperture distribution, we use the convolution theorem according to which

**Fig. A.12**

The encircled energy fraction as a function of angle from the axis of a circular aperture (solid line) compared with that of the same aperture but with a circular hole cut in the middle of the primary with radius 33% of that of the primary mirror (dashed line). This encircled energy distribution is similar to that expected for the Hubble Space Telescope (Schroeder, 2000).

The Fourier transform of the convolution of two functions $f(x) * g(x)$ is the product of their Fourier transforms $F(s) \times G(s)$.

We can therefore convolve the aperture distribution of a single aperture with two delta functions separated by distance D , that is

$$T(z) = f(z) * g(z),$$

where $f(z) = 1$ if $-a \leq z \leq a$ and $f(z) = 0$ otherwise; $g(z)$ is represented by two delta-functions $\delta(z)$ separated by distance D

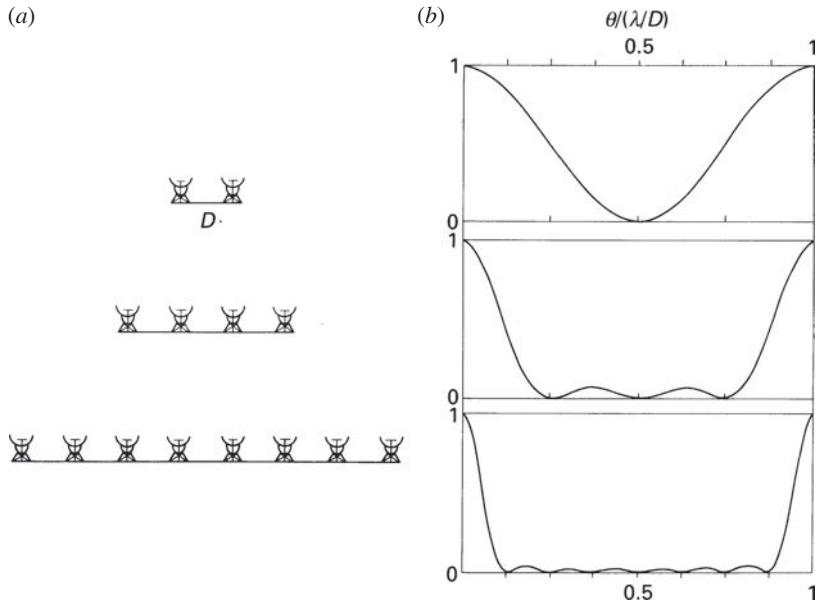
$$g(z) = \delta(-D/2) + \delta(D/2).$$

The amplitude of the diffracted image is given by the Fourier transform of $T(x)$ which, according to the convolution theorem, is just proportional to the product of the Fourier transforms of $f(z)$ and $g(z)$. For the case of two uniform apertures the amplitude is therefore

$$\begin{aligned} F(\theta) &= I_0 \sin(k\theta a) / k\theta a \\ G(\theta) &= (2\pi)^{-1/2} \int g(z) \exp(-ik\theta z) dz \\ &= [\exp(ik\theta D/2) + \exp(-ik\theta D/2)] / (2\pi)^{-1/2}. \end{aligned}$$

and so

$$U(\theta) = F(\theta)G(\theta) = \frac{2I_0}{(2\pi)^{1/2}} \frac{\sin(k\theta a)}{k\theta a} \cos\left(\frac{k\theta D}{2}\right).$$

**Fig. A.13**

(a) A schematic diagram showing grating arrays with increasing numbers of antennae. (b) The angular response of the grating arrays shown in (a). These diagrams show only the part of the beam associated with the interferometer pattern and should be multiplied by the response of an individual antenna to find the overall response of the grating array, i.e. in the case of the rectangular apertures used in the example in the text, the response should be multiplied by the function $[\sin(k\theta a)/k\theta a]^2$ (see A.11). It can be seen that the more antennae added to the array, the narrower the beam becomes relative to the separation between the grating lobes. This is the principle of grating array telescopes.

The power polar diagram of the interferometer is proportional to the square of the amplitude $I(\theta) \propto U^2(\theta)$ and so

$$I(\theta) \propto \left[\frac{\sin(k\theta a)}{k\theta a} \right]^2 \cos^2 \left(\frac{k\theta D}{2} \right). \quad (\text{A.18})$$

This response is shown in Fig. 8.9b. It is straightforward to extend this analysis to an array of telescopes, each separated by distance D to produce what is known as a *grating telescope*. For example, for a four-element interferometer, the function $g(z)$ is

$$g(z) = \delta\left(-\frac{3D}{2}\right) + \delta\left(-\frac{D}{2}\right) + \delta\left(\frac{D}{2}\right) + \delta\left(\frac{3D}{2}\right),$$

and so on for more antennae. Figure A.13a show the distribution of antennae and the resulting power polar diagrams. It can be seen that, as the number of antennae increases, the polar diagram becomes sharper and the first grating side-lobe occurs several beamwidths away from the central maximum.

The extension to two dimensions is straightforward and this technique is used to map the distribution of radio emission from extended sources. The strategy is to arrange for the first grating side-lobe to lie outside the angular distribution of radiation on the sky. In the case

of *aperture synthesis*, the above procedure is used to fill in all the baselines which would be needed to reconstruct a fully filled aperture of diameter equal to the longest baseline used in the interferometric array.

Radio astronomers realised that, in general, the sky does not change with time and so there is no need to obtain all the information at one time. Provided the amplitude and relative phases of the signals arriving at two separate antennae can be measured, the information appropriate to that separation of elements of a large antenna is obtained. In the most elegant implementation of this scheme, the rotation of the Earth is used to move one antenna with respect to another and this technique is often referred to as *Earth-rotation aperture synthesis*. Numerous examples of radio astronomical images obtained by the technique of aperture synthesis are presented in the main text, for example, the remarkable image of the brightest extragalactic radio source in the northern sky Cygnus A obtained with the Very Large Array (VLA) (Fig. 21.1a).

A.7 The sensitivities of astronomical detectors

All signals in astronomy can only be measured with a certain statistical precision. It is convenient to distinguish between those wavebands in which the intensity of radiation is described as a flux of photons, $h\nu \gg kT$, and those in which it is better to think of the radiation as a superposition of electromagnetic waves, $h\nu \ll kT$. In both cases, there are limitations because of the finite number of photons or waves counted per second and because of the background against which the signal has to be detected.

A.7.1 Optical and infrared detectors

If $h\nu \gg kT$, the flux of radiation consists of photons of energy $h\nu$. The simplest limit to the precision with which this flux of radiation can be measured is the statistics of the number of photons counted. In the absence of noise, Poisson statistics tells us that the uncertainty in this number is roughly $n \pm n^{1/2}$. Normally, however, this signal is observed against a noise background, for example, thermal fluctuations in the receiver, fluctuations in the intensity of the sky background and so on. In long exposures, the background signal can far exceed the strength of the signal we are trying to detect. We can however use the central limit theorem to show how the background intensity can be reduced statistically. The theorem may be expressed in the following way:

If N estimates of a quantity x_i are made which are randomly selected from an arbitrary probability density function $p(x)$, the best estimate of the mean value of x is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

and the probability distribution of this value \bar{x} about the true mean is a Gaussian distribution with standard deviation $\sigma_0/N^{1/2}$ where σ_0 is the standard deviation of the probability density function $p(x)$.

This theorem is proved in all the standard text books on statistics and is very important. Notice that the probability density function can be of any form, in particular of non-Gaussian form, and we need only know σ_0 for this non-Gaussian distribution. The other theorem we need concerns the variance of the sum of two quantities x and y , each of which have separately variances σ_x^2 and σ_y^2 , σ_x and σ_y being their standard deviations. The variance of $x + y$ is $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$.

The key point is that, every time we make an astronomical observation, we average over a large number of independent estimates of the random noise signal, x_i . Consequently, whatever the characteristics of the noise described by $p(x)$, it is safe to assume that the probability distribution of \bar{x} is of Gaussian form. Furthermore, as N increases, the mean value of the noise is determined with greater precision and, specifically, the standard deviation about the mean decreases as $N^{-1/2}$ so that, provided the backgrounds do not change, the noise can be determined with very high accuracy by making a sufficiently large number of observations.

Suppose the flux density of the source we wish to detect is $N(\nu)$ photons $\text{m}^{-2} \text{s}^{-1} \text{Hz}^{-1}$ and the source is observed for a time t in a waveband $\Delta\nu$. The effective aperture of the telescope A_{eff} is the area of the telescope once account is taken of all the losses between the flux of radiation entering the aperture, and this flux being detected as registered photons at the detector. A_{eff} includes, for example, the quantum efficiency of the detector, the losses in the optical system, and so on. Therefore, the number of photons detected from the source is

$$S = N(\nu) A_{\text{eff}} t \Delta\nu \text{ photons}.$$

The statistical uncertainty in this number of photons is $\pm n^{1/2}$ so that the signal-to-noise ratio, in the absence of other sources of noise, is

$$\frac{S}{N} = (N(\nu) A_{\text{eff}} t \Delta\nu)^{1/2}. \quad (\text{A.19})$$

To work out the signal-to-noise ratio in the presence of other sources of noise we add the variances of all the sources of noise in the detector system and those due to the environment. Let us list some of these.

- (i) First there is the variance associated with the total number of photons detected given by (A.8).
- (ii) There is background radiation from the sky and the telescope. In the optical and near-infrared wavebands, $\lambda \lesssim 2 \mu\text{m}$, the background from the sky is the primary source of unwanted background radiation. If the intensity of the night sky is $B(\nu)$ photons $\text{m}^{-2} \text{s}^{-1} \text{Hz}^{-1} \text{sr}^{-1}$, we need to know the solid angle Ω subtended by the detector on the sky. Therefore, the signal from the night sky is $A_{\text{eff}} \Omega B(\nu) t \Delta\nu$, which by the same reasoning as in (i) is also the variance σ_2^2 about the mean value of the background. In the case of the thermal infrared wavebands, $\lambda \geq 2 \mu\text{m}$, the telescope itself is a strong source of background radiation, particularly at wavelengths $\lambda \sim 10 \mu\text{m}$ at which a Planck function at 300 K has maximum intensity. The aim of the design of infrared telescopes is to minimise the emissivity of the telescope and, in particular, to ensure that the detector sees as little of the telescope structure as possible. We include the

thermal background of the telescope in the same formula for the background given above.

- (iii) There will be a certain ‘dark current’ in the detector, by which is meant noise electrons produced within the detector volume due to a variety of causes, for example, the thermal excitation of electrons into the potential wells of a CCD detector. If C is the number of electrons generated per second, the total variance during the observation is $\sigma_3^2 = Ct$, following the same reasoning as above.
- (iv) There may also be a noise contribution when the signal is read out through the output amplifier. If this amounts to R electrons rms, the variance is $\sigma_4^2 = R^2$ for a single read-out. If the detector is read out many times in the course of an observation, this source of noise can amount to an important contribution to the noise signal against which faint objects have to be detected.

We now find the total noise signal by summing the variances of these individual noise contributions

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2. \quad (\text{A.20})$$

The signal-to-noise ratio is therefore

$$\frac{S}{N} = \frac{N(\nu) A_{\text{eff}} t \Delta\nu}{\sigma}. \quad (\text{A.21})$$

Let us look at some simple applications of these formulae. We can recover our previous result when the primary source of noise is simply photon statistics. In this case, $\sigma = \sigma_1 = (N(\nu) A_{\text{eff}} \Delta\nu t)^{1/2}$ and

$$S/N = [N(\nu) A_{\text{eff}} t \Delta\nu]^{1/2}. \quad (\text{A.22})$$

If the observations are *detector-noise-limited*, $\sigma = \sigma_3 = (Ct)^{1/2}$ and

$$\frac{S}{N} = \frac{N(\nu) A_{\text{eff}} \Delta\nu}{C^{1/2}} t^{1/2}. \quad (\text{A.23})$$

For very faint objects, particularly in the infrared waveband, the observations are often *background-limited* in which case $\sigma = \sigma_2$ and

$$\frac{S}{N} = N(\nu) \left(\frac{A_{\text{eff}} t}{\Omega B(\nu)} \right)^{1/2}. \quad (\text{A.24})$$

These results can be used to show how the signal-to-noise ratio depends upon the size of the telescope, A_{eff} , and the integration time necessary to achieve this signal-to-noise ratio. An interesting way of writing these relations is in terms of the time needed to achieve a given signal-to-noise ratio. If the diameter of the primary mirror of the telescope is D , $A_{\text{eff}} \propto D^2$ and then the relations (A.11), (A.12) and (A.13) become

$$\text{Photon-noise-limited} \quad t \propto A_{\text{eff}}^{-1} \propto D^{-2} \quad (\text{A.25})$$

$$\text{Detector-noise-limited} \quad t \propto A_{\text{eff}}^{-2} \propto D^{-4} \quad (\text{A.26})$$

$$\text{Background-limited} \quad t \propto \Omega A_{\text{eff}}^{-1} \propto \Omega D^{-2}. \quad (\text{A.27})$$

Another way of expressing these results is in terms of the limiting flux density which can be observed with a given signal-to-noise ratio in a given time. In this case we find

$$\text{Photon-noise-limited} \quad S \propto A_{\text{eff}}^{-1} \propto D^{-2} \quad (\text{A.28})$$

$$\text{Detector-noise-limited} \quad S \propto A_{\text{eff}}^{-1} \propto D^{-2} \quad (\text{A.29})$$

$$\text{Background-limited} \quad S \propto \Omega^{\frac{1}{2}} A_{\text{eff}}^{-1} \propto \Omega^{\frac{1}{2}} D^{-1}. \quad (\text{A.30})$$

In many programmes, the observations are made at the very limit of what is possible technologically and generally speaking these observations are background-limited. It can be seen from the relations (A.16) and (A.19) that the limiting flux density is then strongly dependent upon Ω , the solid angle subtended by the detector on the sky. For example, the fact that the Hubble Space Telescope is diffraction limited with $\theta \approx 0.1$ arcsec in the optical waveband results in a large gain in sensitivity for point sources.

A.7.2 Radio and millimetre-wave receivers

The basic theorem which describes the amplitude of thermal fluctuations in an electrical circuit is *Nyquist's theorem*.³ In thermal equilibrium, the average energy per mode is given by the relation

$$\bar{E} = \frac{h\nu}{\exp(h\nu/kT) - 1}, \quad (\text{A.31})$$

where ν is the frequency of that mode. By considering the modes of a transmission line of wave impedance Z_0 terminated by matched resistors R , it can be readily shown that the resistors each deliver a noise power

$$P = \frac{h\nu}{\exp(h\nu/kT) - 1} \text{ W Hz}^{-1} \quad (\text{A.32})$$

if they are maintained at temperature T . At low frequencies, $h\nu \ll kT$, which is normally the case for radio and microwave receivers, this expression reduces to

$$P = kT, \quad (\text{A.33})$$

which is the familiar form of Nyquist's theorem. Thus, at radio wavelengths, the noise power delivered by the resistor in the frequency range ν to $\nu + \Delta\nu$ is $P d\nu = kT d\nu$. In the opposite limit, $h\nu \gg kT$, the noise power decreases exponentially as

$$P = h\nu \exp(h\nu/kT).$$

For radio receivers, the expression (A.21) provides a convenient way of describing the performance of a receiver which delivers a certain noise power P_n . We can define an *equivalent noise temperature* T_n for the performance of the receiver at frequency ν by the relation

$$T_n = P_n/k. \quad (\text{A.34})$$

³ I have shown in the Appendix to Chapter 15 of *Theoretical Concepts in Physics* how this theorem can be derived from Einstein's prescription for the fluctuations in black-body radiation (Longair, 2003).

There is another important feature of the fluctuations in this noise signal which we have already noted. We note that, per unit frequency range per second, the electrical noise power is kT which corresponds exactly to the energy of a single wave mode in thermodynamic equilibrium. Since this energy is in the form of the superposition of electromagnetic waves, the fluctuations in this wave mode correspond to $\Delta E/E = 1$. Therefore, we expect the amplitude of the noise fluctuations per unit frequency interval to be kT .

A wavepacket consisting of waves with frequencies in the interval ν to $\nu + \Delta\nu$ remains coherent for a time $\tau \approx \Delta\nu^{-1}$ which is called the *coherence time*.⁴ To make independent estimates of the field strength, the samples should not be taken more than about once per coherence time. Thus, whereas for photons we obtain an independent piece of information every time a photon arrives, in the case of waves, we obtain independent estimates only once per coherence time $\tau \approx \Delta\nu^{-1}$. If we observe a source for time t , we obtain $t/\tau = t\Delta\nu$ independent estimates of the intensity of the source.

Normally, we are interested in detecting very weak signals in the presence of a much greater noise signal generated by the receiver. As discussed above, this noise power fluctuates with amplitude $\Delta E/E = 1$ per mode per second. We can therefore reduce the amplitude of the fluctuations by increasing the length of the integration and by increasing the bandwidth of the observations. In both cases, we increase the number of independent estimates of the strength of the signal by a factor $\Delta\nu t$ and hence the amplitude of the power fluctuations is reduced to

$$\Delta P = kT/(\Delta\nu t)^{1/2}. \quad (\text{A.35})$$

Thus, if we are able to integrate long enough and use large enough bandwidths, very weak sources can be observed.

Radio astronomers extend the use of temperatures to define other aspects of the performance of the telescope, its receivers and the objects to be observed. If the intensity of radiation from a region of sky is I_ν , an equivalent *brightness temperature* T_b can be defined using the expression for the intensity of black-body radiation

$$I_\nu = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT_b) - 1}. \quad (\text{A.36})$$

This brightness temperature T_b is a lower limit to the temperature of the region itself because thermodynamically no region can emit radiation with intensity greater than that of a black-body at its thermodynamic temperature, unless the population of the levels is inverted as in the case of masers and lasers. The concept of brightness temperature is thus a very useful way of setting a lower limit to the temperature of the region for sources in which the radiation mechanism is incoherent. Very often in radio astronomy the Rayleigh–Jeans approximation for the brightness temperature can be used because $h\nu \ll kT$ and consequently the expression for brightness temperature reduces to

$$I_\nu = 2kT_b/\lambda^2. \quad (\text{A.37})$$

⁴ This is also demonstrated in the Appendix to Chapter 15 of *Theoretical Concepts in Physics*.

We have already shown that there is a certain amount of noise power available at the terminals of a resistance R at temperature T . We can replace the resistance R by a matched antenna of radiation resistance R and, if the antenna is then placed in a black-body cavity at temperature T , the same noise power will be available at the output of the antenna because the system is in thermodynamic equilibrium. If, instead, the antenna is looking at the sky and the brightness temperature distribution is $T_b(\theta, \phi)$, the mean power available at the output of the antenna is

$$W = \frac{1}{2} A_e \int T_b(\theta, \phi) P(\theta, \phi) d\Omega = k T_A , \quad (\text{A.38})$$

where A_e is the effective area of the antenna and is related to the beam area of the antenna Ω_A by $A_e \Omega_A = \lambda^2$, $P(\theta, \phi)$ is the polar diagram of the antenna and T_A is defined to be the *antenna temperature* due to the radiation incident upon it. If the source is a point object or is small compared with the half-power beamwidth, the flux density of the source is

$$S = (2k T_A / \lambda^2) \Omega_A . \quad (\text{A.39})$$

This antenna temperature has to be detected in the presence of noise and the main contributions are from the radio background emission from our Galaxy, T_{gal} , the emission from the Earth's atmosphere, T_{atm} , and the noise temperature of the complete receiving system, T_{sys} . The last term contains all components which contribute to the noise power present in the receiver at the point where the signal is detected. This includes the contributions from the antenna itself, the antenna noise temperature, T_{na} , from the transmission lines, T_m , and the noise temperature of the receiver, T_n . The sum of these sources of noise can be expressed

$$T_{\text{tot}} = \sum_i T_i .$$

In an observation of integration time t and with bandwidth $\Delta\nu$, the minimum detectable antenna temperature is therefore

$$\Delta T_{\min} = T_{\text{tot}} / (\Delta\nu t)^{\frac{1}{2}} ,$$

as shown above.

A.8 Units and relativistic notation

A.8.1 SI units

We use SI units throughout this text, although this is not yet common practice in much of the astronomical literature. Historically, astronomers and high energy particle physicists have used convenient non-SI units and, indeed, many professionals may not recognise some of the units I employ – teslas for magnetic flux density is a good example. The literature is full of ergs, gauss, centimetres, solar neutrino units, and so on, as well as specifically astronomical terms such as astronomical units, parsecs, magnitudes, colours, etc. SI units

Table A.3 Physical constants.

Velocity of light	$c = 2.997925 \times 10^8 \text{ m s}^{-1}$
Gravitational constant	$G = 6.6726 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$
Planck's constant	$2\pi\hbar = h = 6.6261 \times 10^{-34} \text{ J s}$
Electron charge	$e = 1.6022 \times 10^{-19} \text{ C}$
Mass of electron	$m_e = 9.109 \times 10^{-31} \text{ kg}$
Mass of proton	$m_p = 1.6726 \times 10^{-27} \text{ kg}$
Boltzmann's constant	$k = 1.3807 \times 10^{-23} \text{ J K}^{-1}$
Gas constant	$R = 8.315 \text{ J K}^{-1} \text{ mol}^{-1}$
Avogadro's number	$N_0 = 6.0221 \times 10^{23} \text{ mol}^{-1}$
Permittivity of free space	$\epsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \text{ m}^{-2} \text{ N}^{-1}$
Permeability of free space	$\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1}$
Fine structure constant ($\alpha = e^2/4\pi\epsilon_0\hbar c$)	$\alpha = 7.297 \times 10^{-3}$ $\alpha^{-1} = 137.04$
Classical electron radius ($r_e = e^2/4\pi\epsilon_0 m_e c^2$)	$r_e = 2.818 \times 10^{-15} \text{ m}$
Magnetic moment of 1 Bohr magneton ($\mu_B = eh/4\pi m_e$)	$\mu_B = 9.274 \times 10^{-24} \text{ J T}^{-1}$
Radiation density constant ($a = 8\pi^5 k^4 / 15c^3 h^3$)	$a = 7.5660 \times 10^{-16} \text{ J m}^{-3} \text{ K}^{-4}$
Stefan–Boltzmann constant ($\sigma = ac/4$)	$\sigma = 5.6705 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$

are not always the most convenient units to use since they may result in enormously large or small numbers. Unfortunately, there is little likelihood of standardisation in the near future and one just has to learn to live with it. My approach is as follows. All formulae and derivations are given using SI units and I quote numerical values in SI units. I relax my procedures when observational results are quoted but relate these non-SI units to SI units where appropriate. I introduce non-standard units at the appropriate points in the text. It is useful to gather together all the fundamental constants we need in one place in SI units and these are given in Table A.3.

Energies plays an important part in our story. Normally, particle energies are quoted in *electron volts* (eV) with the usual SI modifications for greater orders of magnitude:

$$\begin{aligned} 1 \text{ kiloelectron volt} &= 1 \text{ keV} = 10^3 \text{ eV}, \\ 1 \text{ megaelectron volt} &= 1 \text{ MeV} = 10^6 \text{ eV}, \\ 1 \text{ gigaelectron volt} &= 1 \text{ GeV} = 10^9 \text{ eV}, \\ 1 \text{ teraelectron volt} &= 1 \text{ TeV} = 10^{12} \text{ eV}, \\ 1 \text{ petaelectron volt} &= 1 \text{ PeV} = 10^{15} \text{ eV}, \\ 1 \text{ exaelectron volt} &= 1 \text{ EeV} = 10^{18} \text{ eV, etc.} \end{aligned}$$

The following conversion factors are useful:

$$\begin{aligned} (\text{proton rest mass}) \times c^2 &= m_p c^2 = 938.3 \text{ MeV} \approx 10^9 \text{ eV} = 1 \text{ GeV}, \\ (\text{electron rest mass}) \times c^2 &= m_e c^2 = 0.511 \text{ MeV} \approx 5 \times 10^5 \text{ eV} = 0.5 \text{ MeV}, \\ 1 \text{ eV} &= 1.602 \times 10^{-19} \text{ J} = 1.602 \times 10^{-12} \text{ erg}. \end{aligned}$$

Other useful energy conversion factors for photons are as follows:

$$E = h\nu = hc/\lambda = 1.2399/\lambda = 4.136 \times 10^{-15} \nu \text{ eV},$$

where the frequency ν is measured in hertz (or cycles s^{-1}) and the wavelengths λ in microns (i.e. 10^{-6} m or 10^3 nm). Although it is usual to use angstroms (\AA) as wavelength units in astronomy, I prefer to use microns or micrometres (μm) and nanometres (nm), recalling that $1 \text{ \AA} = 0.1 \text{ nm} = 10^{-4} \mu\text{m}$.

Very often, we will be interested in the radiation from hot bodies and then a useful conversion factor is:

$$E = kT = 1.380 \times 10^{-23} T \text{ J} = 8.617 \times 10^{-5} T \text{ eV},$$

where the temperature T is measured in kelvins.

A.8.2 Special relativity, four-vectors and basic energy relations

Everyone has their own favourite conventions for calculations in special relativity. I use the notation and conventions for special relativity as presented by Rindler in his book *Relativity: Special, General and Cosmological* (Rindler, 2001). According to his conventions, the Lorentz transformations are written:

$$\left. \begin{array}{l} ct' = \gamma(ct - Vx/c), \\ x' = \gamma(x - Vt), \\ y' = y, \\ z' = z, \end{array} \right\} \quad (\text{A.40})$$

where the Lorentz factor γ is defined to be

$$\gamma = (1 - V^2/c^2)^{-1/2}, \quad (\text{A.41})$$

where V is the relative velocity of the inertial frames of reference S and S' . The velocity of light is always written explicitly in all calculations. The components of four-vectors are defined to correspond to the components of the primitive displacement four-vector \mathbf{R} which is taken to be the quantity

$$\mathbf{R} \equiv [ct, x, y, z]. \quad (\text{A.42})$$

The norm of the four-vector is then the quantity

$$\text{norm}(\mathbf{R}) = c^2 t^2 - x^2 - y^2 - z^2 \quad (\text{A.43})$$

and is a Lorentz invariant in any inertial frame of reference. Note that there is a plus sign in front of the $c^2 t^2$ and minus signs in front of x^2 , y^2 and z^2 ; in the language of relativity, the *signature* of the metric is $[1, -1, -1, -1]$. The time component has a different status from the spatial components, and so I use the convention that the time component has a subscript 0 and refer to it as the zeroth component of the four-vector.

The components of other four-vectors are defined to be the quantities which transform like $[ct, x, y, z, t]$. Thus, if a four-vector is written as $\mathbf{A} = [A_0, A_1, A_2, A_3]$, its norm is

$$|\mathbf{A}|^2 = A_0^2 - A_1^2 - A_2^2 - A_3^2, \quad (\text{A.44})$$

exactly the equivalent to $R^2 = x^2 + y^2 + z^2$ for three-vectors. The transformation rules for the components of A can be found from (A.40) with the identifications

$$\begin{aligned} ct' &\rightarrow A'_0, \quad ct \rightarrow A_0; \\ x' &\rightarrow A'_1, \quad x \rightarrow A_1; \\ y' &\rightarrow A'_2, \quad y \rightarrow A_2; \\ z' &\rightarrow A'_3, \quad z \rightarrow A_3. \end{aligned}$$

In this way, the following four-vectors are defined:

$$\left. \begin{aligned} \text{four-velocity } \mathbf{U} &\equiv [\gamma c, \gamma u_x, \gamma u_y, \gamma u_z], \\ \text{four-momentum } \mathbf{P} &\equiv m_0 \mathbf{U} \equiv [\gamma m_0 c, \gamma m_0 u_x, \gamma m_0 u_y, \gamma m_0 u_z], \\ \text{four-frequency } \mathbf{K} &\equiv [\hbar\omega/c, \hbar\mathbf{k}], \end{aligned} \right\} \quad (\text{A.45})$$

and so on. For more details of the derivations of the forms of these and other four-vectors and their uses, the reader may consult Rindler's text or my own version in *Theoretical Concepts in Physics* (Longair, 2003).

It is useful to recall the various types of energy which we meet in dealing with relativistic particles. We deal with the *total energies, kinetic energies and momenta* of particles. We recall the standard results. If m_0 is the rest mass of the particle, v its velocity and c the velocity of light,

$$\begin{aligned} \text{total energy} &= (\text{rest energy} + \text{kinetic energy}) = \gamma m_0 c^2, \\ \text{kinetic energy} &= \text{total energy} - \text{rest mass energy} = (\gamma - 1)m_0 c^2, \\ \text{relativistic three-momentum} &= \gamma m v. \end{aligned}$$

In the *non-relativistic case*, $v/c \ll 1$, $\gamma \rightarrow 1$

$$\begin{aligned} \text{kinetic energy} &= \frac{1}{2}m_0 v^2, \\ \text{non-relativistic three-momentum} &= m_0 v. \end{aligned}$$

In the *ultra-relativistic case*, $\gamma \gg 1$, $v \approx c$, we find

$$\begin{aligned} \text{total energy} &\approx \text{kinetic energy}, \\ \text{kinetic energy} &\approx \gamma m_0 c^2 \\ \text{relativistic three-momentum} &\approx \gamma m_0 c, \\ \text{total energy} &\approx \text{Kinetic energy} \approx (\text{relativistic three-momentum})c. \end{aligned}$$

Bibliography

- Aaronson, M. & Mould, J. (1983). A distance scale from the infrared magnitude/H I velocity-width relation. IV – The morphological type dependence and scatter in the relation; the distances to nearby groups, *Astrophysical Journal*, **265**, 1–17.
- Abdo, A. A., Ackermann, M., Ajello, M., *et al.* (2009a). Early Fermi Gamma-Ray Space Telescope observations of the quasar 3C 454.3, *Astrophysical Journal*, **699**, 817–823.
- Abdo, A. A., Ackermann, M., Ajello, M., *et al.* (2009b). Bright active galactic nuclei source list from the first three months of the Fermi Large Area Telescope All-Sky Survey, *Astrophysical Journal*, **700**, 597–622.
- Abdurashitov, J. N., Bowles, T. J., Cleveland, B. T., *et al.* (2003). Measurement of the solar neutrino capture rate in SAGE, *Nuclear Physics B Proceedings Supplements*, **118**, 39–46.
- Abdurashitov, J. N., Veretenkin, E. P., Vermul, V. M., *et al.* (2002). Solar neutrino flux measurements by the Soviet–American Gallium Experiment (SAGE) for half the 22-year solar cycle, *Soviet Journal of Experimental and Theoretical Physics*, **95**, 181–193.
- Abell, G. O. (1958). The distribution of rich clusters of galaxies, *Astrophysical Journal Supplement*, **3**, 221–288.
- Abell, G. O., Corwin Jr, H. G., & Olowin, R. P. (1989). A catalogue of rich clusters of galaxies, *Astrophysical Journal Supplement*, **70**, 1–138.
- Abraham, J., Abreu, P., Aglietta, M., & Pierre Auger Collaboration (2010a). Measurement of the depth of maximum of extensive air showers above 10^{18} eV, *Physical Review Letters*, **104**, 091101 1–7.
- Abraham, J., Abreu, P., Aglietta, M., & Pierre Auger Collaboration (2010b). Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Array, *Physics Letters B*, **685**, 239–246.
- Abraham, R. G., Tanvir, N. R., Santiago, B., *et al.* (1996). Galaxy morphology to $I = 25$ mag in the Hubble Deep Field, *Monthly Notices of the Royal Astronomical Society*, **279**, L47–L52.
- Abramovitz, M. & Stegun, I. A. (1965). *Handbook of Mathematical Functions*. New York: Dover.
- Abramowicz, M. A., Jaroszyński, M., & Sikora, M. (1978). Relativistic, accreting disks, *Astronomy and Astrophysics*, **63**, 221–224.
- Adams, F. C., Lada, C. J., & Shu, F. H. (1987). Spectral evolution of young stellar objects, *Astrophysical Journal*, **312**, 788–806.
- Adams, F. C. & Shu, F. H. (1985). Infrared emission from protostars, *Astrophysical Journal*, **296**, 655–669.

- Afonso, C., Albert, J. N., Andersen, J., *et al.* (2003). Limits on Galactic dark matter with 5 years of EROS SMC data, *Astronomy and Astrophysics*, **400**, 951–956.
- Aharonian, F., Akhperjanian, A. G., Anton, G., *et al.* (2009). Simultaneous observations of PKS 2155–304 with HESS, Fermi, RXTE, and Atom: Spectral energy distributions and variability in a low state, *Astrophysical Journal Letters*, **696**, L150–L155.
- Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., *et al.* (2007a). An Exceptional very high energy gamma-ray flare of PKS 2155–304, *Astrophysical Journal Letters*, **664**, L71–L74.
- Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., *et al.* (2006). A low level of extragalactic background light as revealed by γ -rays from blazars, *Nature*, **440**, 1018–1021.
- Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., *et al.* (2007b). H.E.S.S. Observations of the supernova remnant RX J0852.0–4622: Shell-type morphology and spectrum of a widely extended very high energy gamma-ray source, *Astrophysical Journal*, **661**, 236–249.
- Aharonian, F. A., Akhperjanian, A. G., Aye, K.-M., *et al.* (2004). High-energy particle acceleration in the shell of a supernova remnant, *Nature*, **432**, 75–77.
- Ahmad, Q. R., Allen, R. C., Andersen, T. C., *et al.* (2002). Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory, *Physical Review Letters*, **89**, 011301.
- Aitken, D. K., Smith, C. H., James, S. D., *et al.* (1988). 10 micron spectral observations of SN 1987A – The first year, *Monthly Notices of the Royal Astronomical Society*, **235**, 19P–31P.
- Akerib, D. S., Attisha, M. J., Bailey, C. N., *et al.* (2006). Limits on spin-independent interactions of weakly interacting massive particles with nucleons from the Two-Tower run of the Cryogenic Dark Matter Search, *Physical Review Letters*, **96**, 011302.
- Alcock, C., Akerlof, C. W., Allsman, R. A., *et al.* (1993a). Possible gravitational microlensing of a star in the Large Magellanic Cloud, *Nature*, **365**, 621–623.
- Alcock, C., Allsman, R. A., Alves, D. R., *et al.* (2000). The MACHO project: Microlensing results from 5.7 years of Large Magellanic Cloud observations, *Astrophysical Journal*, **542**, 281–307.
- Alcock, C., Allsman, R. A., Axelrod, T. S., *et al.* (1993b). The MACHO project – A search for the dark matter in the Milky Way, in *Sky Surveys: Protostars to Protogalaxies*, ed. Soifer, T., pp. 291–296. San Francisco: Astronomical Society of the Pacific Conference Series.
- Alexander, P. (2006). Models of young powerful radio sources, *Monthly Notices of the Royal Astronomical Society*, **368**, 1404–1410.
- Alexander, P., Brown, M. T., & Scott, P. F. (1984). A multi-frequency radio study of Cygnus A, *Monthly Notices of the Royal Astronomical Society*, **209**, 851–868.
- Alfvén, H. & Herlofson, N. (1950). Cosmic radiation and radio stars, *Physical Review*, **78**, 616.
- Aliu, E., Andringa, S., Aoki, S., *et al.* (2005). Evidence for muon neutrino oscillation in an accelerator-based experiment, *Physical Review Letters*, **94**(8), 081802.

- Aloisio, R., Berezinsky, V., & Gazizov, A. (2009). Ultra high energy cosmic rays: The disappointing model, in *ArXiv e-print 0907.5194*.
- Amsler, C., Doser, M., Antonelli, M., et al. (2008). Review of particle physics, *Physics Letters B*, **667**, 1–5. These data can be found at <http://pdg.lbl.gov>.
- Anderson, C. (1932). The apparent existence of easily deflected positives, *Science*, **76**, 238–239.
- Anderson, C. & Neddermeyer, S. (1936). Cloud chamber observations of cosmic rays at 4300 metres elevation and near sea-level, *Physical Review*, **50**, 263–271.
- Antoni, T., Apel, W. D., Badea, A. F., et al. (2005). KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems, *Astroparticle Physics*, **24**, 1–25.
- Antonucci, R. R. (1993). Unified models for active galactic nuclei and quasars, *Annual Review of Astronomy and Astrophysics*, **31**, 473–521.
- Antonucci, R. R. & Miller, J. S. (1985). Spectropolarimetry and the nature of NGC 1068, *Astrophysical Journal*, **297**, 621–632.
- Arnett, D. (2004). Stellar nucleosynthesis: A status report 2003, in *Origin and Evolution of the Elements*, eds McWilliam, A. & Rauch, M., pp. 12–26. Cambridge: Cambridge University Press.
- Arnett, W. D. & Clayton, D. D. (1970). Explosive nucleosynthesis in stars, *Nature*, **227**, 780–784.
- Arzoumanian, Z., Chernoff, D. F., & Cordes, J. M. (2002). The velocity distribution of isolated radio pulsars, *Astrophysical Journal*, **568**, 289–301.
- Ashie, Y., Hosaka, J., Ishihara, K., et al. (2005). Measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I, *Physical Review D*, **71**, 112005.
- Aublin, J. D. (2009). Discriminating potential astrophysical sources of the highest energy cosmic rays with the Pierre Auger Observatory, in *Proceedings of the 31st International Cosmic Ray Conference (Lodz, Poland)*, ArXiv: 0906.2347v2 [astro-ph.E].
- Auger, P., Ehrenfest Jr., P., Maze, R., et al. (1939). Extensive air showers, *Reviews of Modern Physics*, **11**, 288–291.
- Axford, W. I., Leer, E., & Skadron, G. (1977). The acceleration of cosmic rays by shock waves, *Proceedings of the 15th International Cosmic Ray Conference*, **11**, 132–135.
- Baade, W. & Minkowski, R. (1954). Identification of the radio sources in Cassiopeia, Cygnus A, and Puppis A, *Astrophysical Journal*, **119**, 206–214.
- Babbedge, T. S. R., Rowan-Robinson, M., Vaccari, M., et al. (2006). Luminosity functions for galaxies and quasars in the Spitzer Wide-area Infrared Extragalactic Legacy Survey, *Monthly Notices of the Royal Astronomical Society*, **370**, 1159–1180.
- Backer, D. C., Kulkarni, S. R., Heiles, C., et al. (1982). A millisecond pulsar, *Nature*, **300**, 615–618.
- Bahcall, J. N. (1989). *Neutrino Astrophysics*. Cambridge: Cambridge University Press.
- Bahcall, J. N. & Bethe, H. (1990). A solution of the solar neutrino problem, *Physical Review Letters*, **65**, 2233–2235.
- Bahcall, J. N., Kirhakos, S., Saxe, D. H., et al. (1997a). Hubble Space Telescope images of a sample of 20 nearby luminous quasars, *Astrophysical Journal*, **479**, 642–658.

- Bahcall, J. N., Pinsonneault, M. H., Basu, S., *et al.* (1997b). Are standard solar models reliable?, *Physical Review Letters*, **78**, 171–174.
- Bahcall, N. A. (1977). Clusters of galaxies, *Annual Review of Astronomy and Astrophysics*, **15**, 505–540.
- Bahcall, N. A., Dong, F., Hao, L., *et al.* (2003a). The richness-dependent cluster correlation function: Early Sloan Digital Sky Survey data, *Astrophysical Journal*, **599**, 814–819.
- Bahcall, N. A., McKay, T. A., Annis, J., *et al.* (2003b). A merged catalog of clusters of galaxies from early Sloan Digital Sky Survey data, *Astrophysical Journal Supplement*, **148**, 243–274.
- Balbus, S. A. & Hawley, J. F. (1991). A powerful local shear instability in weakly magnetized disks. I – Linear analysis. II – Nonlinear evolution, *Astrophysical Journal*, **376**, 214–233.
- Baldry, I. K., Glazebrook, K., Brinkmann, J., *et al.* (2004). Quantifying the bimodal color-magnitude distribution of galaxies, *Astrophysical Journal*, **600**, 681–694.
- Ballard, K. R. & Heavens, A. F. (1992). Shock acceleration and steep-spectrum synchrotron sources, *Monthly Notices of the Royal Astronomical Society*, **259**, 89–94.
- Band, D. L. & Grindlay, J. E. (1985). The synchrotron–self-Compton process in spherical geometries. I. Theoretical framework, *Astrophysical Journal*, **298**, 128–146.
- Barthel, P. D. (1989). Is every quasar beamed?, *Astrophysical Journal*, **336**, 606–611.
- Barthel, P. D. (1994). Unified schemes of FR2 radio galaxies and quasars, in *First Stromlo Symposium: Physics of Active Galactic Nuclei*, eds Bicknell, G. V., Dopita, M. A., & Quinn, P. J., pp. 175–186. San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 54.
- Batchelor, G. (1970). *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.
- Bearden, J. A. & Burr, A. F. (1967). Reevaluation of X-ray atomic energy levels, *Reviews of Modern Physics*, **39**, 125–142.
- Beck, R., Brandenburg, A., Moss, D., Shukurov, A., & Sokoloff, D. (1996). Galactic magnetism: Recent developments and perspectives, *Annual Review of Astronomy and Astrophysics*, **34**, 155–206.
- Beck, R., Carilli, C. L., Holdaway, M. A., *et al.* (1994). Multifrequency observations of the radio continuum emission from NGC 253. 1: Magnetic fields and rotation measures in the bar and halo, *Astronomy and Astrophysics*, **292**, 409–424.
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., *et al.* (2006). The Hubble Ultra Deep Field, *Astronomical Journal*, **132**, 1729–1755.
- Begelman, M. C. (1996). Baby Cygnus A's, in *Cygnus A – Study of a Radio Galaxy*, eds Carilli, C. L. and Harris, D. E., pp. 209–214. Cambridge: Cambridge University Press.
- Bekefi, G. (1966). *Radiation Processes in Plasmas*. New York: John Wiley.
- Bell, A. R. (1978). The acceleration of cosmic rays in shock fronts. I, *Monthly Notices of the Royal Astronomical Society*, **182**, 147–156.
- Bell, A. R. (2004). Turbulent amplification of magnetic field and diffusive shock acceleration of cosmic rays, *Monthly Notices of the Royal Astronomical Society*, **353**, 550–558.
- Bell, A. R. (2005). The interaction of cosmic rays and magnetized plasma, *Monthly Notices of the Royal Astronomical Society*, **358**, 181–187.

- Bell, A. R. & Lucek, S. G. (2001). Cosmic ray acceleration to very high energy through the non-linear amplification by cosmic rays of the seed magnetic field, *Monthly Notices of the Royal Astronomical Society*, **321**, 433–438.
- Bell, E. F., McIntosh, D. H., Katz, N., et al. (2003). The optical and near-infrared properties of galaxies: I. Luminosity and stellar mass functions, *Astrophysical Journal Supplement Series*, **149**, 289–312.
- Bell-Burnell, J. (1983). The discovery of pulsars, in *Serendipitous Discoveries in Radio Astronomy*, eds Kellermann, K. & Sheets, B., pp. 160–170. Green Bank, West Virginia: National Radio Astronomy Publications.
- Bender, R., Burstein, D., & Faber, S. M. (1993). Dynamically hot galaxies ii. global stellar populations, *Astrophysical Journal*, **411**, 153–169.
- Bennett, C., Halpern, M., Hinshaw, G., et al. (2003). First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results, *Astrophysical Journal Supplement Series*, **148**, 1–27.
- Bennett, C. L., Banday, A. J., Gorski, K. M., et al. (1996). Four-year COBE DMR Cosmic Microwave Background observations: Maps and basic results, *Astrophysical Journal*, **464**, L1–L4.
- Benson, B. A., Church, S. E., Ade, P. A. R., et al. (2004). Measurements of Sunyaev–Zel'dovich effect scaling relations for clusters of galaxies, *Astrophysical Journal*, **617**, 829–846.
- Beresnyak, A., Jones, T. W., & Lazarian, A. (2009). Turbulence-induced magnetic fields and the structure of cosmic ray modified shocks, *Astrophysical Journal* **707**, 1541–1549.
- Berezhko, E. G. & Völk, H. J. (2007). Spectrum of cosmic rays produced in supernova remnants, *Astrophysical Journal Letters*, **661**, L175–L178.
- Berezinsky, V. (2007). On origin of ultra high energy cosmic rays, *Astrophysics and Space Science*, **309**, 453–463.
- Berezinsky, V., Gazizov, A. & Grigorjeva, S. (2006). On astrophysical solution to ultrahigh energy cosmic rays, *Physical Review*, **D74**, 043005, pp. 1–35.
- Berger, K., Majumdar, P., Lindfors, E., et al. (2009). MAGIC observations of the distant quasar 3C 279 during an optical outburst in 2007, *ArXiv e-prints* 0907.1046.
- Bertola, F., Bettoni, D., Danziger, J., et al. (1991). Testing the gravitational field in elliptical galaxies: NGC 5077, *Astrophysical Journal*, **373**, 369–390.
- Bertola, F. & Galletta, G. (1979). Ellipticity and twisting of isophotes in elliptical galaxies, *Astronomy and Astrophysics*, **77**, 363–365.
- Best, P. N., Bailer, D. M., Longair, M. S., et al. (1995). Radio source asymmetries and unified schemes, *Monthly Notices of the Royal Astronomical Society*, **275**, 1171–1184.
- Best, P. N., Longair, M. S., & Röttrung, H. J. A. (1996). Evolution of the aligned structures in $z \sim 1$ radio galaxies, *Monthly Notices of the Royal Astronomical Society*, **280**, L9–L12.
- Best, P. N., Longair, M. S., & Röttrung, H. J. A. (1997). HST, radio and infrared observations of 28 3CR radio galaxies at redshift $z \sim 1$. I - The observations, *Monthly Notices of the Royal Astronomical Society*, **292**, 758–794.
- Best, P. N., Longair, M. S., & Röttrung, H. J. A. (1998). HST, radio and infrared observations of 28 3CR radio galaxies at redshift z approximately equal to 1. II – Old stellar populations

- in central cluster galaxies, *Monthly Notices of the Royal Astronomical Society*, **295**, 549–567.
- Best, P. N., Longair, M. S., & Röttrig, H. J. A. (2000). Ionization, shocks and evolution of the emission-line gas of distant 3CR radio galaxies, *Monthly Notices of the Royal Astronomical Society*, **311**, 23–36.
- Bethe, H. & Heitler, W. (1934). On the stopping of fast particles and on the creation of positive electrons, *Proceedings of the Royal Society of London*, **A146**, 83–112.
- Béthermin, M., Dole, H., Beelen, A., *et al.* (2010). Spitzer deep and wide legacy mid- and far-infrared number counts and lower limits of cosmic infrared background, *Astronomy and Astrophysics*, **512**, A78, 1–14.
- Bignami, G. F., Caraveo, P. A., Luca, A. D., *et al.* (2003). The magnetic field of an isolated neutron star from X-ray cyclotron absorption lines, *Nature*, **423**, 725–727.
- Bildsten, L., Chakrabarty, D., Chiu, J., *et al.* (1997). Observations of accreting pulsars, *Astrophysical Journal Supplement Series*, **113**, 367–408.
- Binney, J. (1978). On the rotation of elliptical galaxies, *Monthly Notices of the Royal Astronomical Society*, **183**, 501–514.
- Binney, J. & Merrifield, M. (1998). *Galactic Astronomy*. Princeton: Princeton University Press.
- Binney, J. & Tremaine, S. (2008). *Galactic Dynamics*. Princeton: Princeton University Press.
- Biretta, J. A., Sparks, W. B., & Macchetto, F. (1999). Hubble Space Telescope observations of superluminal motion in the M87 jet, *Astrophysical Journal*, **520**, 621–626.
- Biretta, J. A., Zhou, F., & Owen, F. N. (1995). Detection of proper motions in the M87 jet, *Astrophysical Journal*, **447**, 582–596.
- Blaauw, A., Gum, C. S., Pawsey, J. L., *et al.* (1959). Note: Definition of the new I.A.U. system of galactic co-ordinates, *Astrophysical Journal*, **130**, 702–703.
- Blackett, P. & Occhialini, G. (1933). Some photographs of the tracks of penetrating radiation, *Proceedings of the Royal Society of London*, **A139**, 699–722.
- Blain, A. W. & Longair, M. S. (1993). Sub-millimetre cosmology, *Monthly Notices of the Royal Astronomical Society*, **264**, 509–521.
- Blain, A. W. & Longair, M. S. (1996). Observing strategies for blank-field surveys in the sub-millimetre waveband, *Monthly Notices of the Royal Astronomical Society*, **279**, 847–858.
- Blandford, R. & Eichler, D. (1987). Particle acceleration at astrophysical shocks – A theory of cosmic-ray origin, *Physics Reports*, **154**, 1–75.
- Blandford, R. D. (1990). Physical processes in active galactic nuclei, in *Active Galactic Nuclei*, eds Blandford, R. D., Netzer, H., Woltjer, L., *et al.* pp. 161–275, Saas-Fee Advanced Course 20. Berlin: Springer-Verlag.
- Blandford, R. D. (1994). Holes, disks, stars and jets in active galactic nuclei, in *The Physics of Active Galaxies*, ed. Bicknell, G. V., Dopita, M. A., pp. 23–32. San Francisco: Astronomical Society of the Pacific Conference Series. Vol. 54 .
- Blandford, R. D. & Begelman, M. C. (1999). On the fate of gas accreting at a low rate onto a black hole, *Monthly Notices of the Royal Astronomical Society*, **303**, L1–L5.

- Blandford, R. D. & McKee, C. F. (1982). Reverberation mapping of the emission line regions of Seyfert galaxies and quasars, *Astrophysical Journal*, **255**, 419–439.

Blandford, R. D. & Narayan, R. (1992). Cosmological applications of gravitational lensing, *Annual Review of Astronomy and Astrophysics*, **30**, 311–358.

Blandford, R. D. & Ostriker, J. P. (1978). Particle acceleration by astrophysical shocks, *Astrophysical Journal*, **221**, L29–L32.

Blandford, R. D. & Rees, M. J. (1974). A 'twin-exhaust' model for double radio sources, *Monthly Notices of the Royal Astronomical Society*, **169**, 395–415.

Blandford, R. D. & Znajek, R. L. (1977). Electromagnetic extraction of energy from Kerr black holes, *Monthly Notices of the Royal Astronomical Society*, **179**, 433–456.

Blanton, M., Hogg, D., Bahcall, N., et al. (2003). The galaxy luminosity function and luminosity density at redshift $z = 0.1$, *Astrophysical Journal*, **592**, 819–838.

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. (2003). The broadband optical properties of galaxies with redshifts $0.02 \leq z \leq 0.22$, *Astrophysical Journal*, **594**, 186–207.

Blumenthal, G. R. & Gould, R. J. (1970). Bremsstrahlung, synchrotron radiation, and Compton scattering of high-energy electrons traversing dilute gases, *Reviews of Modern Physics*, **42**, 237–271.

Bolton, C. T. (1972). Identifications of CYG X-1 with HDE 226868, *Nature*, **235**, 271–273.

Bondi, H. (1952). On spherically symmetrical accretion, *Monthly Notices of the Royal Astronomical Society*, **112**, 195–204.

Bondi, H. & Hoyle, F. (1944). On the mechanism of accretion by stars, *Monthly Notices of the Royal Astronomical Society*, **104**, 273–282.

Born, M. & Wolf, E. (1999). *Principles of Optics*, 7th edition. Cambridge: Cambridge University Press.

Bosma, A. (1981). 21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types, *Astronomical Journal*, **86**, 1825–1846.

Bothe, W. & Kolhörster, W. (1929). The nature of the high-altitude radiation, *Zeitschrift für Physik*, **56**, 751–777.

Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P., et al. (2006). Galaxies at $z \sim 6$: The UV luminosity function and luminosity density from 506 HUDF, HUDF parallel ACS field, and GOODS i-dropouts, *Astrophysical Journal*, **653**, 53–85.

Boyle, B. J., Griffiths, R. E., Shanks, T., et al. (1993). A deep ROSAT survey. I – The QSO X-ray luminosity function, *Monthly Notices of the Royal Astronomical Society*, **260**, 49–58.

Boyle, B. J., Shanks, T., Croom, S. M., et al. (2000). The 2dF QSO Redshift Survey – I. The optical luminosity function of quasi-stellar objects, *Monthly Notices of the Royal Astronomical Society*, **317**, 1014–1022.

Bracassi, A., Formiggini, L., & Gandolfi, E. (1970). Magnitudes, colours and coordinates of 175 ultraviolet excess objects in the field $13^h, +36^\circ$, *Astronomy and Astrophysics*, **5**, 264?–279. Erratum: *Astronomy and Astrophysics*, **23**, 159.

Bracewell, R. (1986). *The Fourier Transform and its Applications*. New York: McGraw-Hill.

- Brandt, W. N. & Hasinger, G. (2005). Deep extragalactic X-Ray surveys, *Annual Review of Astronomy and Astrophysics*, **43**, 827–859.
- Browne, I. W. A. & Murphy, D. W. (1987). Beaming and the X-ray, optical and radio properties of quasars, *Monthly Notices of the Royal Astronomical Society*, **226**, 601–627.
- Bruzual, G. & Charlot, S. (2003). Stellar population synthesis at the resolution of 2003, *Monthly Notices of the Royal Astronomical Society*, **344**, 1000–1028.
- Burbidge, E. M., Burbidge, G. R., & Sandage, A. R. (1963). Evidence for the occurrence of violent events in the nuclei of galaxies, *Reviews of Modern Physics*, **35**, 947–972.
- Burbidge, G. R. (1956). On synchrotron radiation from Messier 87, *Astrophysical Journal*, **124**, 416–429.
- Burbidge, G. R. (1959). Estimates of the total energy in particles and magnetic field in the non-thermal radio sources, *Astrophysical Journal*, **129**, 849–851.
- Caffee, M. W., Reedy, R. C., Goswami, J. N., et al. (1988). Irradiation records in meteorites, in *Meteorites and the Early Solar System*, eds Kerridge, J. & Matthews, M., pp. 205–245. Tucson: University of Arizona Press.
- Calabretta, M. R. & Greisen, E. W. (2002). Representations of celestial coordinates in FITS, *Astronomy and Astrophysics*, **395**, 1077–1122.
- Camenzind, M. (2007). *Compact Objects in Astrophysics – White Dwarfs, Neutron Stars and Black Holes*. Berlin: Springer-Verlag.
- Cameron, A. G. W. (1973). Abundances of the elements in the Solar System, *Space Science Reviews*, **15**, 121–146.
- Cappelluti, N., Hasinger, G., Brusa, M., et al. (2007). The XMM-Newton wide-field survey in the COSMOS field II: X-ray data and the log N–log S relations, *Astrophysical Journal Supplement*, **172**, 341–352.
- Carilli, C. L. & Barthel, P. D. (1996). Cygnus A, *Astronomy and Astrophysics Reviews*, **7**, 1–54.
- Carilli, C. L., Perley, R. A., Dreher, J. W., et al. (1991). Multifrequency radio observations of Cygnus A – Spectral aging in powerful radio galaxies, *Astrophysical Journal*, **383**, 554–573.
- Carlstrom, J. E., Joy, M. K., Grego, L., et al. (2000). Imaging the Sunyaev–Zel'dovich effect, in *Particle Physics and the Universe: Proceedings of Nobel Symposium 198*, eds Bergström, L., Carlson, P., & Fransson, C., pp. 148–155. Stockholm: Physica Scripta.
- Carron, N. (2007). *An Introduction to the Passage of Energetic Particles Through Matter*. London: Taylor and Francis.
- Carter, B. (1971). Axisymmetric black hole has only two degrees of freedom, *Physical Review Letters*, **26**, 331–333.
- Casandjian, J. & Grenier, I. A. (2008). A revised catalogue of EGRET γ -ray sources, *Astronomy and Astrophysics*, **489**, 849–883.
- Caswell, J. L. (1976). A map of the northern sky at 10 MHz, *Monthly Notices of the Royal Astronomical Society*, **177**, 601–616.
- Cavaliere, A. (1980). Models of X-ray emission from clusters of galaxies, in *X-ray Astronomy*, eds Giacconi, R. & Setti, G., pp. 217–237. Dordrecht: Reidel.
- Cesarsky, C. J. (1980). Cosmic-ray confinement in the galaxy, *Annual Review of Astronomy and Astrophysics*, **18**, 289–319.

- Challinor, A. & Lasenby, A. (1998). Relativistic corrections to the Sunyaev–Zeldovich effect, *Astrophysical Journal*, **499**, 1–6.
- Chambers, K. C., Miley, G. K., & van Breugel, W. J. M. (1987). Alignment of radio and optical orientations in high-redshift radio galaxies, *Nature*, **329**, 604–606.
- Chandrasekhar, S. (1981). *Hydrodynamic and Hydromagnetic Stability*. New York: Dover.
- Chandrasekhar, S. (1983). *The Mathematical Theory of Black Holes*. Oxford: Clarendon Press.
- Charbonneau, D., Brown, T. M., Latham, D. W., et al. (2000). Detection of planetary transits across a Sun-like star, *Astrophysical Journal*, **529**, L45–L48.
- Charlot, S. & Longhetti, M. (2001). Nebular emission from star-forming galaxies, *Monthly Notices of the Royal Astronomical Society*, **323**, 887–903.
- Chevalier, R. A. (1998). Synchrotron self-absorption in radio supernovae, *Astrophysical Journal*, **499**, 810–819.
- Christensen-Dalsgaard, J. (2002). Helioseismology, *Reviews of Modern Physics*, **74**, 1073–1129.
- Christian, D. J. (2002). The third Extreme Ultraviolet Explorer right angle program catalog: The last years, *Astronomical Journal*, **124**, 3478–3484.
- Chupp, E. L. (1976). *Gamma-Ray Astronomy: Nuclear Transition Region*. Dordrecht, Reidel. (Geophysics and Astrophysics Monographs. Volume 14.)
- Chwolson, O. (1924). Über eine mögliche Form fiktiver Doppelsterne, *Astronomische Nachrichten*, **221**, 329.
- Cimatti, A., Daddi, E., Renzini, A., et al. (2004). Old galaxies in the young Universe, *Nature*, **430**, 184–187.
- Clark, J. S., Goodwin, S. P., Crowther, P. A., et al. (2002). Physical parameters of the high-mass X-ray binary 4U1700–37, *Astronomy and Astrophysics*, **392**, 909–920.
- Clavel, J., Reichert, G. A., & 56 authors (1991). Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. I – An 8 month campaign of monitoring NGC 5548 with IUE, *Astrophysical Journal*, **366**, 64–81.
- Clemmow, P. C. W. & Dougherty, J. P. (1969). *Electrodynamics of Particles and Plasmas*. Reading, Massachusetts: Addison-Wesley.
- Coburn, W., Kretschmar, P., Kreykenbohm, I., et al. (2006). Cyclotron features in X-ray spectra of accreting pulsars, *Advances in Space Research*, **38**, 2747–2751.
- Cohen, M. H., Lister, M. L., Homan, D. C., et al. (2007). Relativistic beaming and the intrinsic properties of extragalactic radio jets, *Astrophysical Journal*, **658**, 232–244.
- Cole, S., Percival, W. J., Peacock, J. A., et al. (2005). The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final data set and cosmological implications, *Monthly Notices of the Royal Astronomical Society*, **362**, 505–534.
- Colless, M., Dalton, G., Maddox, S., et al. (2001). The 2dF Galaxy Redshift Survey: Spectra and redshifts, *Monthly Notices of the Royal Astronomical Society*, **328**, 1039–1063.
- Colless, M. & Dunn, A. M. (1996). Structure and dynamics of the Coma Cluster, *Astrophysical Journal*, **458**, 435–454.
- Compton, A. H. (1923). The spectrum of scattered X-rays, *Physical Review*, **22**, 409–413.

- Condon, J. J. (1989). The 1.4 gigahertz luminosity function and its evolution, *Astrophysical Journal*, **338**, 13–23.
- Condon, J. J. (1992). Radio emission from normal galaxies, *Annual Review of Astronomy and Astrophysics*, **30**, 575–611.
- Cordes, J. M. & Lazio, T. J. W. (2002). NE2001. I. A new model for the Galactic distribution of free electrons and its fluctuations, *ArXiv Astrophysics e-prints* astro-ph/0207156.
- Cordes, J. M. & Lazio, T. J. W. (2003). NE2001. II. Using radio propagation data to construct a model for the Galactic distribution of free electrons, *ArXiv Astrophysics e-prints* astro-ph/0301598.
- Costa, E., Frontera, F., Heise, J., et al. (1997). Discovery of an X-ray afterglow associated with the gamma-ray burst of 28 February 1997, *Nature*, **387**, 783–785.
- Côté, P., McLaughlin, D. E., Cohen, J. G., et al. (2003). Dynamics of the globular cluster system associated with M49 (NGC 4472): Cluster orbital properties and the distribution of dark matter, *Astrophysical Journal*, **591**, 850–877.
- Côté, P., McLaughlin, D. E., Hanes, D. A., et al. (2001). Dynamics of the globular cluster system associated with M87 (NGC 4486). II. Analysis, *Astrophysical Journal*, **559**, 828–850.
- Cowie, L. (1988). Protogalaxies, in *The Post-Recombination Universe*, eds Kaiser, N. & Lasenby, A. N., pp. 1–18. Dordrecht: Kluwer.
- Cowie, L., Lilly, S., Gardner, J., et al. (1988). A cosmologically significant population of galaxies dominated by very young star formation, *Astrophysical Journal*, **332**, L29–L32.
- Cowie, L. L., Barger, A. J., & Kneib, J.-P. (2002). Faint submillimeter counts from deep 850 micron observations of the lensing clusters A370, A851, and A2390, *Astronomical Journal*, **123**, 2197–2205.
- Cowie, L. L., Songaila, A., Hu, E. M., et al. (1996). New insight on galaxy formation and evolution from Keck spectroscopy of the Hawaii Deep Fields, *Astronomical Journal*, **112**, 839–864.
- Cox, D. P. & Smith, B. W. (1974). Large-scale effects of supernova remnants on the Galaxy: Generation and maintenance of a hot network of tunnels, *Astrophysical Journal Letters*, **189**, L105–L108.
- Cristiani, S. (1986). Optical variability in quasars, in *Structure and Evolution of Active Galactic Nuclei*, eds Giuricin, G., Mardirossian, F., Mezzetti, M., et al., pp. 83–91. Dordrecht: Reidel.
- Cruddace, R., Paresce, F., Bowyer, S., et al. (1974). On the opacity of the interstellar medium to ultrasoft X-rays and extreme-ultraviolet radiation, *Astrophysical Journal*, **187**, 497–504.
- Dabrowski, Y., Fabian, A. C., Iwasawa, K., et al. (1997). The profile and equivalent width of the X-ray iron emission line from a disc around a Kerr black hole, *Monthly Notices of the Royal Astronomical Society*, **288**, L11–L15.
- Damon, P. E., Kaime, D., Kocharov, G. E., et al. (1995). Radiocarbon production by the gamma-ray component of supernova explosions, *Radiocarbon*, **37**, 599–604.
- Damon, P. E., Lerman, J. C., & Long, A. (1978). Temporal fluctuations of atmospheric ^{14}C : Causal factors and implications, *Annual Review of Earth and Planetary Science*, **6**, 457–494.

- Davidson, W. & Davies, M. (1964). Interpretation of the counts of radio sources in terms of a 4-parameter family of evolutionary universes, *Monthly Notices of the Royal Astronomical Society*, **127**, 241–255.
- Davies, R. D. (2006). An anomalous dust emission component? – The observations, in *CMB and Physics of the Early Universe*, pp. 1–8. Proceedings of Science – online journal. http://pos.sissa.it//archive/conferences/027/018/CMB2006_018.pdf
- Davies, R. L., Efstathiou, G., Fall, S. M., et al. (1983). The kinematic properties of faint elliptical galaxies, *Astrophysical Journal*, **266**, 41–57.
- Davis, L. & Greenstein, J. L. (1951). The polarization of starlight by aligned dust grains, *Astrophysical Journal*, **114**, 206–240.
- de Plaa, J., Kaastra, J. S., Méndez, M., et al. (2005). The temperature structure in the core of Sérsic 159–03, *Advances in Space Research*, **36**, 601–604.
- de Vaucouleurs, G., de Vaucouleurs, A., Corwin Jr., H. G., et al. (1991). *Third Reference Catalogue of Bright Galaxies: Containing Information on 23,024 Galaxies with Reference to Papers Published Between 1913 and 1988*. Berlin: Springer-Verlag.
- Dermer, C. D. (1986). Secondary production of neutral pi-mesons and the diffuse galactic gamma radiation, *Astronomy and Astrophysics*, **157**, 223–229.
- Dermer, C. D. & Menon, G. (2009). *High Energy Radiation from Black Holes: Gamma Rays, Cosmic Rays, and Neutrinos*. Princeton: Princeton University Press.
- Deubner, F.-L. & Gough, D. (1984). Helioseismology: Oscillations as a diagnostic of the Solar interior, *Annual Review of Astronomy and Astrophysics*, **22**, 593–619.
- Dey, A. (1997). The host galaxies of distant radio sources, in *The Hubble Space Telescope and the High Redshift Universe*, eds Tanvir, N. R., Aragón-Salamanca, A., & Wall, J. V., pp. 373–376. Singapore: World Scientific.
- Diehl, R., Halloin, H., Kretschmer, K., et al. (2006a). Radioactive ^{26}Al from massive stars in the Galaxy, *Nature*, **439**, 45–47.
- Diehl, R., Prantzos, N., & von Ballmoos, P. (2006b). Astrophysical constraints from gamma-ray spectroscopy, *Nuclear Physics A*, **777**, 70–97.
- Dirac, P. (1928a). The quantum theory of the electron, *Proceedings of the Royal Society of London*, **A117**, 610–624.
- Dirac, P. (1928b). The quantum theory of the electron II, *Proceedings of the Royal Society of London*, **A118**, 351–361.
- Djorgovski, S. G. & Davis, M. (1987). Fundamental properties of elliptical galaxies, *Astrophysical Journal*, **313**, 59–68.
- Doeleman, S. S., Weintraub, J., Rogers, A. E. E., et al. (2008). Event-horizon-scale structure in the supermassive black hole candidate at the Galactic Centre, *Nature*, **455**, 78–80.
- Dombrovski, V. A. (1954). On the nature of the radiation from the Crab Nebula, *Dokladi Akademii Nauk SSSR*, **94**, 1021–1024.
- Dopita, M. A. & Sutherland, R. S. (1996). Spectral signatures of fast shocks. I. Low-density model grid, *Astrophysical Journal Supplement*, **102**, 161–188.
- Draine, B. T. (2003). Interstellar dust grains, *Annual Reviews of Astronomy and Astrophysics*, **41**, 241–289.

- Draine, B. T. (2004). Astrophysics of dust in cold clouds, in *The Cold Universe, Saas-Fee Advanced Course 32*, eds Blain, A. W., Combes, F., Draine, B. T., Pfenniger, D., & Revaz, Y., pp. 213–304. Berlin: Springer-Verlag.
- Draine, B. T. & Lazarian, A. (1998). Electric dipole radiation from spinning dust grains, *Astrophysical Journal*, **508**, 157–179.
- Dreher, J. W., Carilli, C. L., & Perley, R. A. (1987). The Faraday rotation of Cygnus A – Magnetic fields in cluster gas, *Astrophysical Journal*, **316**, 611–625.
- Dressler, A. (1980). Galaxy morphology in rich clusters – Implications for the formation and evolution of galaxies, *Astrophysical Journal*, **236**, 351–365.
- Dressler, A., Lynden-Bell, D., Burstein, D., et al. (1987). Spectroscopy and photometry of elliptical galaxies. I – A new distance estimator, *Astrophysical Journal*, **313**, 42–58.
- Driver, S. P., Allen, P. D., Graham, A. W., et al. (2006). The Millennium Galaxy Catalogue: Morphological classification and bimodality in the colour–concentration plane, *Monthly Notices of the Royal Astronomical Society*, **368**, 414–434.
- Drury, L. O., Duffy, P., Eichler, D., et al. (1999). On ‘box’ models of shock acceleration and electron synchrotron spectra, *Astronomy and Astrophysics*, **347**, 370–374.
- Drury, L. O. & Falle, S. A. E. G. (1986). On the stability of shocks modified by particle acceleration, *Monthly Notices of the Royal Astronomical Society*, **223**, 353–376.
- Dunlop, J. S. (1998). Cosmic star-formation and radio source evolution, in *Astrophysics and Space Science Library, Dordrecht: Kluwer Vol. 226: Observational Cosmology with the New Radio Surveys*, eds Bremer, M. N., Jackson, N., & Perez-Fournon, I., pp. 157–164.
- Dunlop, J. S. & Peacock, J. A. (1990). The redshift cut-off in the luminosity function of radio galaxies and quasars, *Monthly Notices of the Royal Astronomical Society*, **247**, 19–42.
- Dunlop, J. S., Peacock, J. A., Spinrad, H., et al. (1996). A 3.5-Gyr-old galaxy at redshift 1.55, *Nature*, **381**, 581–584.
- Edelson, R., Vaughan, S., Warwick, R., et al. (1999). The ROSAT Wide Field Camera Extragalactic Survey, *Monthly Notices of the Royal Astronomical Society*, **307**, 91–98.
- Efstathiou, G. (1990). Cosmological perturbations, in *Physics of the Early Universe*, eds Peacock, J. A., Heavens, A. F., & Davies, A. T., pp. 361–463. Edinburgh: SUSSP Publications.
- Eguchi, K., Enomoto, S., & 97 authors (2003). First results from Kamland: Evidence for reactor anti-neutrino disappearance, *Physical Review Letters*, **90**, 021802.
- Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt, *Annalen der Physik*, **322**, 132–148.
- Einstein, A. (1915). Die Feldgleichung der Gravitation (The Field Equations of Gravitation), *Sitzungsberichte, Königlich Preussische Akademie der Wissenschaften (Berlin)*, **II**, 844–847.
- Einstein, A. (1936). Lens-like action of a star by the deviation of light in the gravitational field, *Science*, **84**, 506–507.
- Ellis, G. R. A. (1982). Galactic radio emission below 16.5 MHz and the galactic emission measure, *Australian Journal of Physics*, **35**, 91–104.
- Ellis, R. G. (1997). Faint blue galaxies, *Annual Review of Astronomy and Astrophysics*, **35**, 389–443.

- Ellis, S. C. & Bland-Hawthorn, J. (2006). GalaxyCount: A JAVA calculator of galaxy counts and variances in multiband wide-field surveys to 28 AB mag, *Monthly Notices of the Royal Astronomical Society*, **377**, 815–828.
- Enge, H. A. (1966). *Introduction to Nuclear Physics*. London: Addison-Wesley.
- Erber, T. (1966). High-energy electromagnetic conversion processes in intense magnetic fields, *Reviews of modern physics*, **38**, 626–659.
- Eugster, O., Herzog, G. F., Marti, K., et al. (2006). Irradiation records, cosmic-ray exposure ages, and transfer times of meteorites, in *Meteorites and the Early Solar System II*, eds Lauretta, D. S. and McSween, H. Y., Jr., pp. 829–851. Tucson: University of Arizona Press.
- Fabbiano, G., Trinchieri, G., Elvis, M., et al. (1984). An X-ray survey of a complete sample of 3CR radio galaxies, *Astrophysical Journal*, **277**, 115–131.
- Faber, S. M. (1973). Variations in spectral-energy distributions and absorption-line strengths among elliptical galaxies, *Astrophysical Journal*, **179**, 731–754.
- Faber, S. M. (1999). Black holes in galaxy centers, in *Formation of Structure in the Universe*, eds Dekel, A. & Ostriker, J. P., pp. 337–359. Cambridge: Cambridge University Press.
- Faber, S. M. & Jackson, R. E. (1976). Velocity dispersions and mass-to-light ratios for elliptical galaxies, *Astrophysical Journal*, **204**, 668–683.
- Fabian, A. C. (1994). Cooling flows in clusters of galaxies, *Annual Review of Astronomy and Astrophysics*, **32**, 277–318.
- Fabian, A. C. (1998). Emission lines: signatures of relativistic rotation, in *Theory of Black Hole Accretion Disks*, eds Abramowicz, M. A., Bjornsson, G., & Pringle, J. E., pp. 123–133. Cambridge: Cambridge University Press.
- Fabian, A. C. (2009). Black holes at work, *Astronomy and Geophysics*, **30**, 3.18–3.24.
- Fabian, A. C. & Rees, M. J. (1995). The accretion luminosity of a massive black hole in an elliptical galaxy, *Monthly Notices of the Royal Astronomical Society*, **277**, L55–L58.
- Fabian, A. C., Sanders, J. S., Ettori, S., et al. (2000). Chandra imaging of the complex X-ray core of the Perseus Cluster, *Monthly Notices of the Royal Astronomical Society*, **318**, L65–L68.
- Fabian, A. C., Sanders, J. S., Taylor, G. B., et al. (2006). A very deep Chandra observation of the Perseus Cluster: Shocks, ripples and conduction, *Monthly Notices of the Royal Astronomical Society*, **366**, 417–428.
- Fabian, A. C., Vaughan, S., Nandra, K., et al. (2002). A long hard look at MCG-6-30-15 with XMM-Newton, *Monthly Notices of the Royal Astronomical Society*, **335**, L1–L5.
- Fabricant, D. G., Lecar, M., & Gorenstein, P. (1980). X-ray measurements of the mass of M87, *Astrophysical Journal*, **241**, 552–560.
- Fan, X., Hennawi, J. F., Richards, G. T., et al. (2004). A survey of $z \geq 5.7$ quasars in the Sloan Digital Sky Survey. III. Discovery of five additional quasars, *Astronomical Journal*, **128**, 515–522.
- Fan, X., Narayanan, V. K., Lupton, R. H., et al. (2001). A survey of $z \geq 5.8$ quasars in the Sloan Digital Sky Survey. I. Discovery of three new quasars and the spatial density of luminous quasars at $z \sim 6$, *Astronomical Journal*, **122**, 2833–2849.

- Fanaroff, B. L. & Riley, J. M. (1974). The morphology of extragalactic radio sources of high and low luminosity, *Monthly Notices of the Royal Astronomical Society*, **167**, 31P–36P.
- Felten, J. (1977). Study of the luminosity function for field galaxies, *Astronomical Journal*, **82**, 861–878.
- Felten, J. (1985). Galaxy luminosity functions, M/L ratios, and closure of the Universe – Numbers and problems, *Comments on Astrophysics*, **11**, 53–67.
- Fermi, E. (1949). On the origin of the cosmic radiation, *Physical Review*, **75**, 1169–1174.
- Fernini, I., Burns, J. O., Leahy, J. P., et al. (1991). Depolarization asymmetry in the quasar 3C 47, *Astrophysical Journal*, **381**, 63–71.
- Ferrario, D. T., Wickramsinghe, D. T., Bailey, I. R., et al. (1989). EXO 033319-2554.2: An eclipsing AM Herculis system showing cyclotron emission features, *Astrophysical Journal*, **337**, 832–842.
- Feynman, R., Leighton, R. B., & Sands, M. L. (1965). *Feynman Lectures on Physics*. Redwood City, California: Addison-Wesley.
- Feynman, R. P. (1972). *Statistical Mechanics: A Set of Lectures*. Reading, Massachusetts: W. A. Benjamin.
- Fich, M. & Tremaine, S. (1991). The mass of the Galaxy, *Annual Review of Astronomy and Astrophysics*, **29**, 409–445.
- Field, G. B. (1965). Thermal instability, *Astrophysical Journal*, **142**, 531–567.
- Field, G. B., Goldsmith, D. W., & Habing, H. J. (1969). Cosmic-ray heating of the interstellar gas, *Astrophysical Journal Letters*, **55**, L149–L154.
- Fitch, W. S., Pacholczyk, A. G., & Weymann, R. J. (1967). Light variations of the Seyfert galaxy NGC 4151, *Astrophysical Journal*, **150**, L67–L70.
- Fitzpatrick, R. (2008). *The Physics of Plasmas*. Lulu publishers. available at: <http://farside.ph.utexas.edu/teaching/plasma/380.pdf>.
- Ford, H. C., Harms, R. J., Tsvetanov, Z. I., et al. (1994). Narrowband HST Images of M87: Evidence for a disk of ionized gas around a massive black hole, *Astrophysical Journal Letters*, **435**, L27–L30.
- Forman, W., Jones, C., Cominsky, L., et al. (1978). The fourth UHURU catalog of X-ray sources, *Astrophysical Journal Supplement Series*, **38**, 357–412.
- Fort, B. & Mellier, Y. (1994). Arc(let)s in clusters of galaxies., *Astronomy and Astrophysics Reviews*, **5**, 239–292.
- Francis, P. J., Hewett, P. C., Foltz, C. B., et al. (1991). A high signal-to-noise ratio composite quasar spectrum, *Astrophysical Journal*, **373**, 465–470.
- Frank, J., King, A., & Raine, D. J. (2002). *Accretion Power in Astrophysics*, 3rd edition. Cambridge: Cambridge University Press.
- Frank, J., King, A. R., & Lasota, J.-P. (1987). The light curves of low-mass X-ray binaries, *Astronomy and Astrophysics*, **178**, 137–142.
- Frolov, V. P. & Novikov, I. D. (1998). *Black Hole Physics: Basic Concepts and New Developments*. Dordrecht: Kluwer.
- Fukuda, S., Fukuda, Y., & 117 authors (2001). Solar ${}^8\text{B}$ and hep neutrino measurements from 1258 days of Super-Kamiokande data, *Physical Review Letters*, **86**, 5651–5655.
- Galama, T. J., Vreeswijk, P. M., van Paradijs, J., et al. (1998). An unusual supernova in the error box of the γ -ray burst of 25 April 1998, *Nature*, **395**, 670–672.

- Garcia-Munoz, M., Simpson, J. A., Guzik, T. G., *et al.* (1987). Cosmic-ray propagation in the Galaxy and in the heliosphere – The path-length distribution at low energy, *Astrophysical Journal Supplement*, **64**, 269–304.
- Garrington, S. T., Leahy, J. P., Conway, R. G., *et al.* (1988). A systematic asymmetry in the polarization properties of double radio sources with one jet, *Nature*, **331**, 147–149.
- Gavazzi, R., Treu, T., Rhodes, J. D., *et al.* (2007). The Sloan Lens ACS Survey. IV: The mass density profile of early-type galaxies out to 100 effective radii, *Astrophysical Journal*, **667**, 176–190.
- Gehrels, N., Ramirez-Ruiz, E., & Fox, D. B. (2009). Gamma-ray bursts in the Swift era, *Annual Review of Astronomy and Astrophysics*, **47**, 567–617.
- Geiger, H. & Müller, W. (1928). Das Electronenzählrohr (The electron-counting tube), *Physicalische Zeitschrift*, **29**, 839–841.
- Geiger, H. & Müller, W. (1929). Technische Bemerkungen zum Electronenzählrohr (Technical remarks on the electron counting tube), *Physicalische Zeitschrift*, **30**, 489–493.
- Genzel, R., Schödel, R., Ott, T., *et al.* (2003). Near-infrared flares from accreting gas around the supermassive black hole at the galactic centre, *Nature*, **425**, 934–937.
- Ghez, A. M., Morris, M., Becklin, E. E., *et al.* (2000). The accelerations of stars orbiting the milky way's central black hole, *Nature*, **407**, 349–351.
- Giacalone, J. & Jokipii, J. R. (2007). Magnetic field amplification by shocks in turbulent fluids, *Astrophysical Journal Letters*, **663**, L41–L44.
- Giacconi, R., Gursky, H., Kellogg, E., *et al.* (1971). Discovery of periodic X-ray pulsations in Centaurus X-3 from UHURU, *Astrophysical Journal*, **167**, L67–L73.
- Giacconi, R., Gursky, H., & van Speybroeck, L. P. (1968). Observational techniques in X-ray astronomy, *Annual Review of Astronomy and Astrophysics*, **6**, 373–416.
- Giavalisco, M., Dickinson, M., Ferguson, H. C., *et al.* (2004). The rest-frame ultraviolet luminosity density of star-forming galaxies at redshifts $z \geq 3.5$, *Astrophysical Journal Letters*, **600**, L103–L106.
- Gilli, R., Comastri, A., & Hasinger, G. (2007). The synthesis of the cosmic X-ray background in the Chandra and XMM-Newton era, *Astronomy and Astrophysics*, **463**, 79–96.
- Ginzburg, V. L. (1951). Cosmic rays as a source of Galactic radio-radiation, *Doklady Akademii Nauk SSSR*, **76**, 377–380.
- Ginzburg, V. L., Sazonov, V. N., & Syrovatskii, S. I. (1968). Synchrotron radiation and its reabsorption, *Soviet Physics Uspekhi*, **11**, 34.
- Ginzburg, V. L. & Syrovatskii, S. I. (1964). *The Origin of Cosmic Rays*. Oxford: Pergamon Press.
- Ginzburg, V. L. & Syrovatskii, S. I. (1965). Cosmic magnetobremssstrahlung (synchrotron radiation), *Annual Review of Astronomy and Astrophysics*, **3**, 297–350.
- Ginzburg, V. L. & Syrovatskii, S. I. (1969). Developments in the theory of synchrotron radiation and its reabsorption, *Annual Review of Astronomy and Astrophysics*, **7**, 375–420.
- Glazebrook, K., Abraham, R. G., McCarthy, P. J., *et al.* (2004). High abundance of massive galaxies 3–6 billion years after the Big Bang, *Nature*, **430**, 181–184.

- Glazebrook, K., Ellis, R. S., Colless, M., *et al.* (1995). The morphological identification of the rapidly evolving population of faint galaxies, *Monthly Notices of the Royal Astronomical Society*, **275**, L19–L22.
- Gold, T. (1968). Rotating neutron stars as the origin of pulsating radio sources, *Nature*, **218**, 731–732.
- Goldreich, P. & Julian, W. H. (1969). Pulsar electrodynamics, *Astrophysical Journal*, **157**, 869–880.
- Goodrich, R. W. & Osterbrock, D. E. (1983). MRK 744 and MRK 1066 – Two Seyfert galaxies with strong absorption-line spectra, *Astrophysical Journal*, **269**, 416–422.
- Gould, R. J. (2005). *Electromagnetic Processes*. Princeton: Princeton University Press.
- Gradshteyn, I. S. & Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products*. New York: Dover.
- Granot, J. (2008). Critical review of basic afterglow concepts, *ArXiv e-prints* 0811.1657.
- Granot, J. & Sari, R. (2002). The shape of spectral breaks in gamma-ray burst afterglows, *Astrophysical Journal*, **568**, 820–829.
- Green, D. A., Reynolds, S. P., Borkowski, K. J., *et al.* (2008). The radio expansion and brightening of the very young supernova remnant G1.9+0.3, *Monthly Notices of the Royal Astronomical Society*, **387**, L54–L58.
- Greenhill, L. J., Henkel, C., Becker, R., *et al.* (1995a). Centripetal acceleration within the subparsec nuclear maser disk of NGC4258., *Astronomy and Astrophysics*, **304**, 21–33.
- Greenhill, L. J., Jiang, D. R., Moran, J. M., *et al.* (1995b). Detection of a subparsec diameter disk in the nucleus of NGC 4258, *Astrophysical Journal*, **440**, 619–627.
- Greiner, J., Cuby, J. G., & McCaughrean, M. J. (2001). An unusually massive stellar black hole in the Galaxy, *Nature*, **414**, 522–525.
- Greisen, K. (1966). End to the cosmic-ray spectrum?, *Physical Review Letters*, **16**, 748–750.
- Griffin, R. F. (1985). The distributions of periods and amplitudes of late-type spectroscopic binaries, in *Interacting Binaries*, eds Eggleton, P. P. & Pringle, J. E., pp. 1–12, Dordrecht: Reidel.
- Gueth, F. & Guilloteau, S. (1999). The jet-driven molecular outflow of HH 211, *Astronomy and Astrophysics*, **343**, 571–584.
- Gugliucci, N. E., Taylor, G. B., Peck, A. B., *et al.* (2005). Dating COINS: kinematic ages for compact symmetric objects, *Astrophysical Journal*, **622**, 136–148.
- Gull, S. F. (1975). The X-ray, optical and radio properties of young supernova remnants, *Monthly Notices of the Royal Astronomical Society*, **171**, 263–278.
- Gull, S. F. & Northover, K. J. E. (1973). Bubble model of extragalactic radio sources, *Nature*, **244**, 80–83.
- Gunn, J. E. (1978). The Friedmann models and optical observations in cosmology, in *Observational Cosmology: 8th Advanced Course, Swiss Society of Astronomy and Astrophysics, Saas-Fee 1978*, eds Maeder, A., Martinet, L., & Tamman, G., pp. 1–121. Geneva: Geneva Observatory Publications.
- Gunn, J. E. & Ostriker, J. P. (1970). On the nature of pulsars. III. Analysis of observations, *Astrophysical Journal*, **160**, 979–1002.
- Haberl, F. (2007). The magnificent seven: magnetic fields and surface temperature distributions, *Astrophysics and Space Science*, **308**, 181–190.

- Hague, J. D. (2009). "Astrophysical sources of cosmic rays and related measurements with the Pierre Auger Observatory", ArXiv:090b. 2347v2 [astro-ph. HE].
- Hampel, W., Handt, J., Heusser, G., et al. (1999). GALLEX solar neutrino observations: Results for GALLEX IV, *Physics Letters B*, **447**, 127–133.
- Häring, N. & Rix, H. (2004). On the black hole mass–bulge mass relation, *Astrophysical Journal Letters*, **604**, L89–L92.
- Harms, R. J., Ford, H. C., Tsvetanov, Z. I., et al. (1994). HST FOS spectroscopy of M87: Evidence for a disk of ionized gas around a massive black hole, *Astrophysical Journal Letters*, **435**, L35–L38.
- Hasinger, G., Burg, R., Giacconi, R., et al. (1993). A deep X-ray survey in the Lockman Hole and the soft X-ray Log N –Log S , *Astronomy and Astrophysics*, **275**, 1–15.
- Hasinger, G. & van der Klis, M. (1989). Two patterns of correlated X-ray timing and spectral behaviour in low-mass X-ray binaries, *Astronomy and Astrophysics*, **225**, 79–96.
- Hauser, M. G., Arendt, R. G., Kelsall, T., et al. (1998). The COBE diffuse infrared background experiment search for the cosmic infrared background. I. Limits and detections, *Astrophysical Journal*, **508**, 25–43.
- Hauser, M. G. & Dwek, E. (2001). The cosmic infrared background: Measurements and implications, *Annual Review of Astronomy and Astrophysics*, **39**, 249–307.
- Hawking, S. W. (1972). Black holes in general relativity, *Communications in Mathematical Physics*, **25**, 152–166.
- Hawking, S. W. (1975). Particle creation by black holes, in *Quantum Gravity: Proceedings of the Oxford Symposium*, eds Isham, C. J., Penrose, R., & Sciama, D. W., pp. 219–267. Oxford: Clarendon Press.
- Hawking, S. W. & Ellis, G. R. (1973). *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- Hawkins, M. R. S. (1986). On the nature of objects detected as faint long-term variables, *Monthly Notices of the Royal Astronomical Society*, **219**, 417–426.
- Hayashi, C. (1961). Stellar evolution in early phases of gravitational contraction, *Publications of the Astronomical Society of Japan*, **13**, 450–452.
- Hazard, C., Mackey, M. B., & Shimmims, A. J. (1963). Investigation of the radio source 3C 273 by the method of lunar occultations, *Nature*, **197**, 1037–1039.
- Heckman, T. M. (1980). An optical and radio survey of the nuclei of bright galaxies – Activity in normal galactic nuclei, *Astronomy and Astrophysics*, **87**, 152–164.
- Heiles, C. (1976). The interstellar magnetic field, *Annual Review of Astronomy and Astrophysics*, **14**, 1–22.
- Heitler, W. (1954). *The Quantum Theory of Radiation*. Oxford: Oxford University Press.
- Hess, V. (1912). Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten, (Concerning observations of penetrating radiation on seven free balloon flights), *Physikalische Zeitschrift*, **13**, 1084–1091.
- Hesser, J. E., Harris, W. E., Vandenberg, D. A., et al. (1987). A CCD color–magnitude study of 47 Tucanae, *Publications of the Astronomical Society of the Pacific*, **99**, 739–808.
- Hewish, A. (1986). The pulsar era, *Quarterly Journal of the Royal Astronomical Society*, **27**, 548–558.

- Hewish, A., Bell, S. J., Pilkington, J. D. H., *et al.* (1968). Observations of a rapidly pulsating radio source, *Nature*, **217**, 709–713.
- Hewitt, J. N., Turner, E. L., Burke, B. F., *et al.* (1987). A VLA gravitational lens survey, in *Observational Cosmology: IAU Symposium No. 124*, eds Hewitt, A., Burbidge, G., & Fang, L. Z., pp. 747–750. Dordrecht: Reidel.
- Heyvaerts, J. (1981). Particle acceleration in solar flares, in *Solar Flare Magnetohydrodynamics*, ed. Priest, E. R., pp. 429–555. London: Gordon and Breach.
- Hildebrand, R. H. (1983). The determination of cloud masses and dust characteristics from submillimetre thermal emission, *Quarterly Journal of the Royal Astronomical Society*, **24**, 267–282.
- Hillas, A. M. (1984). The origin of ultra-high-energy cosmic rays, *Annual Review of Astronomy and Astrophysics*, **22**, 425–444.
- Hillebrandt, W. & Niemeyer, J. C. (2000). Type IA supernova explosion models, *Annual Review of Astronomy and Astrophysics*, **38**, 191–230.
- Hillenbrand, L. A. (1997). On the stellar population and star-forming history of the Orion Nebula Cluster, *Astronomical Journal*, **113**, 1733–1768.
- Hinton, J. (2009). Ground-based gamma-ray astronomy with Cherenkov telescopes, *New Journal of Physics*, **11**(5), 055005.
- Hirata, K. S., Inoue, K., Kajita, T., *et al.* (1990). Results from one thousand days of real-time, directional solar-neutrino data, *Physical Review Letters*, **65**, 1297–1300.
- Hjellming, R. M. & Johnston, K. J. (1981). An analysis of the proper motions of SS 433 radio jets, *Astrophysical Journal Letters*, **246**, L141–L145.
- Hjorth, J., Sollerman, J., Møller, P., *et al.* (2003). A very energetic supernova associated with the γ -ray burst of 29 March 2003, *Nature*, **423**, 847–850.
- Hoekstra, H., Yee, H. K. C., & Gladders, M. D. (2004). Properties of galaxy dark matter halos from weak lensing, *Astrophysical Journal*, **606**, 67–77.
- Hogg, D. W., Blanton, M. R., Brinchmann, J., *et al.* (2004). The dependence on environment of the color–magnitude relation of galaxies, *Astrophysical Journal*, **601**, L29–L32.
- Holloway, N. J. & Pryce, M. H. L. (1981). Properties of gaps in pulsar magnetospheres, *Monthly Notices of the Royal Astronomical Society*, **194**, 95–110.
- Homan, D. C., Kadler, M., Kellermann, K. I., *et al.* (2009). MOJAVE: Monitoring of Jets in Active Galactic Nuclei with VLBA Experiments. VII. Blazar jet acceleration, *Astrophysical Journal*, **706**, 1253–1268.
- Hook, I. M., McMahon, R. G., Boyle, B. J., *et al.* (1991). The variability of a large sample of quasars, in *The Space Distribution of Quasars*, ed. Crampton, D., pp. 67–75. San Francisco: Astronomical Society of the Pacific Conference Series Vol. 21.
- Horne, K. & Marsh, T. R. (1986). Indirect imaging of accretion disks in binaries, in *The Physics of Accretion onto Compact Objects*, eds Mason, K. O., Watson, M. G., & White, N. E., pp. 1–13. Berlin: Springer Verlag.
- Hosaka, J., Ishihara, K., Kameda, J., *et al.* (2006). Solar neutrino measurements in Super-Kamiokande-I, *Physical Review D*, **73**, 112001.
- Hoyle, F. & Fowler, W. A. (1963). On the nature of strong radio sources, *Monthly Notices of the Royal Astronomical Society*, **125**, 169–176. Also, Nature of strong radio sources, (1963), *Nature*, **197**, 533–535.

- Hoyle, F. & Lyttleton, R. A. (1939). The effect of interstellar matter on climatic variation, *Proceedings of the Cambridge Philosophical Society*, **35**, 405–415.
- Huarte-Espinosa, M., Krause, M. & Alexander, P. (2010). Interaction of Fanaroff–Riley II radio jets with turbulent intracluster magnetic fields, *Monthly Notices of the Royal Astronomical Society* (in press).
- Hubble, E. P. (1929). A relation between distance and radial velocity among extra-galactic nebulae, *Proceedings of the National Academy of Sciences*, **15**, 168–173.
- Hubble, E. P. (1936). *The Realm of the Nebulae*. New Haven: Yale University Press.
- Huchra, J., Jarrett, T., Skrutskie, M., et al. (2005). The 2MASS Redshift Survey and low galactic latitude large-scale structure, in *Nearby Large-Scale Structures and the Zone of Avoidance*, ed. Fairall, K. P. & Woudt, P. A., pp. 135–146. San Francisco: *Astronomical Society of the Pacific Conference Series*, vol. 329.
- Hughes, P. A. (ed.) (1991). *Beams and Jets in Astrophysics*. Cambridge: Cambridge University Press.
- Hulse, R. A. & Taylor, J. H. (1975). Discovery of a pulsar in a binary system, *Astrophysical Journal Letters*, **195**, L51–L53.
- Illingworth, G. (1977). Rotation (?) in 13 elliptical galaxies, *Astrophysical Journal Letters*, **218**, L43–L47.
- Inskip, K. J., Best, P. N., Longair, M. S., et al. (2002). Infrared magnitude–redshift relations for luminous radio galaxies, *Monthly Notices of the Royal Astronomical Society*, **329**, 277–289.
- Irwin, M., McMahon, R. G., & Hazard, C. (1991). APM optical surveys for high redshift quasars, in *The Space Distribution of Quasars*, ed. Crampton, D., pp. 117–126, San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 21.
- Iyudin, A. F., Diehl, R., Bloemen, H., et al. (1994). COMPTEL observations of Ti-44 gamma-ray line emission from CAS A, *Astronomy and Astrophysics*, **284**, L1–L4.
- Jackson, J. D. (1999). *Classical Electrodynamics*. New York: John Wiley.
- Jenkins, E. B. (1987). Observations of absorption lines from highly ionized atoms, in *Exploring the Universe with the IUE Satellite*, ed. Kondo, Y., volume 129 of *Astrophysics and Space Science Library*, pp. 531–548, Dordrecht: Reidel.
- Jennison, R. C. & Das Gupta, M. K. (1953). Fine structure of the extra-terrestrial radio source Cygnus 1, *Nature*, **172**, 996–997.
- Jones, T. W. (2008). The role of MHD in the ICM and its interactions with AGN outflows, in *Extragalactic Jets: Theory and Observation from Radio to Gamma Ray*, ed. Rector, T. A. and De Young, D. S. pp. 398–409. San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 386.
- Jokipii, J. R. (1973). Turbulence and scintillations in the interplanetary plasma, *Annual Review of Astronomy and Astrophysics*, **11**, 1–28.
- Kaastra, J. S., Tamura, T., Peterson, J. R., et al. (2004). Spatially resolved X-ray spectroscopy of cooling clusters of galaxies, *Astronomy and Astrophysics*, **413**, 415–439.
- Kaiser, C. R. & Alexander, P. (1997). A self-similar model for extragalactic radio sources, *Monthly Notices of the Royal Astronomical Society*, **286**, 215–222.
- Kaler, J. (2001). Planetary nebulae, *Encyclopedia of Astronomy and Astrophysics*, **3**, 2066–2074.

- Kang, H. & Jones, T. W. (2006). Numerical studies of diffusive shock acceleration at spherical shocks, *Astroparticle Physics*, **25**, 246–258.
- Kang, H., Ryu, D., & Jones, T. W. (2009). Self-similar evolution of cosmic-ray modified shocks: The cosmic-ray spectrum, *Astrophysical Journal*, **695**, 1273–1288.
- Kapahi, V. K. & Saikia, D. J. (1982). Relativistic beaming in the central components of double radio quasars, *Journal of Astrophysics and Astronomy*, **3**, 465–483.
- Karttunen, H., Kroger, P., Oja, H., et al. (2007). *Fundamental Astronomy*. Heidelberg: Springer-Verlag.
- Karzas, W. J. & Latter, R. (1961). Electron radiative transitions in a Coulomb field, *Astrophysical Journal Supplement*, **6**, 167–212.
- Katz, D. M., Kassim, N. E., Lazio, T. J. W. et al. (2000). Spatial variations of the synchrotron spectrum within Tycho's supernova remnant (3C 10): A spectral tomography analysis of radio observations at 20 and 90 centimeter wavelengths, *Astrophysical Journal*, **529**, 453–462.
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. (2003). The dependence of star formation history and internal structure on stellar mass for 10^5 low-redshift Galaxies, *Monthly Notices of the Royal Astronomical Society*, **341**, 54–69.
- Kellermann, K. I., Vermeulen, R. C., Zensus, J. A., et al. (1998). Sub-milliarcsecond imaging of quasars and active galactic nuclei, *Astronomical Journal*, **115**, 1295–1318.
- Kembhavi, A., Feigelson, E. D., & Singh, K. P. (1986). X-ray and radio core emission in radio quasars, *Monthly Notices of the Royal Astronomical Society*, **220**, 51–67.
- Kembhavi, A. K. & Narlikar, J. V. (1999). *Quasars and Active Galactic Nuclei – An Introduction*. Cambridge: Cambridge University Press.
- Kennicutt, R. (1989). The star formation law in galactic discs, *Astrophysical Journal*, **344**, 685–703.
- Kennicutt, R. (2006). Young spirals get older, *Nature*, **442**, 753–754.
- Kennicutt, R. C., Edgar, B. K., & Hodge, P. W. (1989). Properties of H II region populations in galaxies. II – The H II region luminosity Function, *Astrophysical Journal*, **337**, 761–781.
- Kennicutt, Jr., R. C. (1998). The global Schmidt law in star-forming galaxies, *Astrophysical Journal*, **498**, 541–552.
- Kent, S. M. & Gunn, J. E. (1982). The dynamics of rich clusters of galaxies. I - The Coma Cluster, *Astronomical Journal*, **87**, 945–971.
- Kerr, R. P. (1963). Gravitational field of a spinning mass as an example of algebraically special metrics, *Physical Review Letters*, **11**, 237–238.
- Khachikian, E. Y. & Weedman, D. W. (1971). A spectroscopic study of luminous galactic nuclei, *Astrofizika*, **7**, 389–406.
- Khachikian, E. Y. & Weedman, D. W. (1974). An atlas of Seyfert galaxies, *Astrophysical Journal*, **192**, 581–589.
- Kidger, M. R. (1989). The optical variability of 3C 345, *Astronomy and Astrophysics*, **226**, 6–22.
- Kiepenheuer, K. O. (1950). Cosmic rays as the source of general galactic radio emission, *Physical Review*, **79**, 738–739.

- King, I. R. (1966). The structure of star clusters. III. Some simple dynamical models, *Astronomical Journal*, **71**, 64–75.
- King, I. R. (1981). The dynamics of globular clusters, *Quarterly Journal of the Royal Astronomical Society*, **22**, 227–243.
- Kippenhahn, R. & Weigert, A. (1990). *Stellar Structure and Evolution*. Berlin: Springer-Verlag.
- Klebesadel, R. W., Strong, I. B., & Olson, R. A. (1973). Observations of gamma-ray bursts of cosmic origin, *Astrophysical Journal Letters*, **182**, L85–L88.
- Klochkov, D., Staubert, R., Postnov, K., et al. (2008). INTEGRAL observations of Hercules X-1, *Astronomy and Astrophysics*, **482**, 907–915.
- Kneib, J. P. (1993). *PhD Dissertation*. Universit Paul Sabatier, Toulouse.
- Koch, H. W. & Motz, J. W. (1959). Bremsstrahlung cross-section formulas and related data, *Reviews of Modern Physics*, **31**, 920–955.
- Kolb, E. W. & Turner, M. S. (1990). *The Early Universe*. Redwood City, California: Addison-Wesley.
- Kolhörster, W. (1913). Messungen der Durchdringenden Strahlung im Freiballon in Grösseren Höhen, *Physikalische Zeitschrift*, **14**, 1153–1156.
- Kompaneets, A. (1956). The establishment of thermal equilibrium between quanta and electrons, *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, **31**, 876–885. (English translation: 1957, *Soviet Physics*, **4**, 730–737).
- Koo, D. C. & Kron, R. (1982). QSO counts – A complete survey of stellar objects to $B = 23$, *Astronomy and Astrophysics*, **105**, 107–119.
- Kormendy, J. & Bender, R. (1999). The double nucleus and central black hole of M31, *Astrophysical Journal*, **522**, 772–792.
- Kormendy, J. & Richstone, D. O. (1995). Inward bound – The search for supermassive black holes in galactic nuclei, *Annual Review of Astronomy and Astrophysics*, **33**, 581–624.
- Kovalev, Y. Y., Aller, H. D., Aller, M. F., et al. (2009). The relation between AGN gamma-ray emission and parsec-scale radio jets, *Astrophysical Journal Letters*, **696**, L17–L21.
- Kovalev, Y. Y., Kellermann, K. I., Lister, M. L., et al. (2005). Sub-milliarcsecond imaging of quasars and active galactic nuclei. IV. Fine-scale structure, *Astronomical Journal*, **130**, 2473–2505.
- Kowal, G., Lazarian, A., Vishniac, E. T., et al. (2009). Numerical tests of fast reconnection in weakly stochastic magnetic fields, *Astrophysical Journal*, **700**, 63–85.
- Kramer, M., Stairs, I. H., Manchester, R. N., et al. (2006). Tests of general relativity from timing the double pulsar, *Science*, **314**, 97–102.
- Krause, M. (2005). Very light jets II: Bipolar large scale simulations in King atmospheres, *Astronomy and Astrophysics*, **431**, 45–64.
- Krause, O., Birkmann, S. M., Usuda, T., et al. (2008a). The Cassiopeia A supernova was of Type IIb, *Science*, **320**, 1195–1197.
- Krause, O., Tanaka, M., Usuda, T., et al. (2008b). Tycho Brahe's 1572 supernova as a standard type Ia as revealed by its light-echo spectrum, *Nature*, **456**, 617–619.
- Krolik, J. H. (1999). *Active Galactic Nuclei – From the Central Black Hole to the Galactic Environment*. Princeton, NJ: Princeton University Press.

- Krymsky, G. F. (1977). A regular mechanism for the acceleration of charged particles on the front of a shock wave, *Doklady Akademii Nauk SSSR*, **234**, 1306–08.
- Krzeminski, W. (1974). The identification and UBV photometry of the visible component of the Centaurus X-3 binary system, *Astrophysical Journal Letters*, **192**, L135–L138.
- Ku, W., Helfand, D. J., & Lucy, L. B. (1980). X-ray properties of quasars, *Nature*, **288**, 323–328.
- Kubota, A. & Makishima, K. (2005). Observational studies of stellar black hole binaries and ULXs, *ArXiv Astrophysics e-prints*, astro-ph/0507271. See also: Advances in Space Research, Special Issue Proceedings of 35th COSPAR Conference, Paris, France, 18–25 July 2004.
- Kulsrud, R. & Pearce, W. P. (1969). The effect of wave-particle interactions on the propagation of cosmic rays, *Astrophysical Journal*, **156**, 445–469.
- Kulsrud, R. M. (2005). *Plasma Physics for Astrophysics*. Princeton, NJ: Princeton University Press.
- Labeyrie, A. (1978). Stellar interferometry methods, *Annual Review of Astronomy and Astrophysics*, **16**, 77–102.
- Lacy, M., Miley, G., Rawlings, S., *et al.* (1994). 8C 1435+635: A radio galaxy at $z = 4.25$, *Monthly Notices of the Royal Astronomical Society*, **271**, 504–512.
- Lagache, G., Dole, H., & Puget, J.-L. (2003). Modelling infrared galaxy evolution using a phenomenological approach, *Monthly Notices of the Royal Astronomical Society*, **338**, 555–571.
- Lagache, G., Dole, H., Puget, J.-L., *et al.* (2004). Polycyclic aromatic hydrocarbon contribution to the infrared output energy of the Universe at $z \simeq 2$, *Astrophysical Journal Supplement*, **154**, 112–117.
- Lagage, P. O. & Cesarsky, C. J. (1983). The maximum energy of cosmic rays accelerated by supernova shocks, *Astronomy and Astrophysics*, **125**, 249–257.
- Laing, R. A. (1988). The sidedness of jets and depolarization in powerful extragalactic radio sources, *Nature*, **331**, 149–151.
- Laing, R. A. (1993). Radio observations of jets: large scales, in *Astrophysical Jets*, eds Burgarella, D., Livio, M., & O'Dea, C., pp. 95–119. Cambridge: Cambridge University Press.
- Laing, R. A. & Bridle, A. H. (2002). Relativistic models and the jet velocity field in the radio galaxy 3C 31, *Monthly Notices of the Royal Astronomical Society*, **336**, 328–352.
- Laing, R. A., Riley, J. M., & Longair, M. S. (1983). Bright radio sources at 178 MHz – Flux densities, optical identifications and the cosmological evolution of powerful radio galaxies, *Monthly Notices of the Royal Astronomical Society*, **204**, 151–187.
- Lal, D. (1972). Hard rock cosmic ray archaeology, *Space Science Reviews*, **14**, 3–102.
- Lamb, H. (1932). *Hydrodynamics, 6th edition*. Cambridge: Cambridge University Press.
- Landau, L. D. & Lifshitz, E. M. (1987). *Fluid Mechanics, 2nd edition*. Oxford: Butterworth-Heinemann.
- Larmor, J. (1884). Electromagnetic induction in conducting sheets and solid bodies, *Philosophical Magazine, Series 5*, **17**, 1–23.

- Lattes, C., Occhialini, G., & Powell, C. (1947). Observations on the tracks of slow mesons in photographic emulsions, *Nature*, **160**, 453–456.
- Lawson, K. D., Mayer, C. J., Osborne, J. L., *et al.* (1987). Variations in the spectral index of the Galactic radio continuum emission in the northern hemisphere, *Monthly Notices of the Royal Astronomical Society*, **225**, 307–327.
- Lazarian, A. & Vishniac, E. T. (1999). Reconnection in a weakly stochastic field, *Astrophysical Journal*, **517**, 700–718.
- Lazarian, A., Vishniac, E. T. & Cho, J. (2004). Magnetic field structure and stochastic reconnection in a partially ionized gas, *Astrophysical Journal*, **603**, 180–197.
- Le Borgne, D., Elbaz, D., Ocvirk, P., *et al.* (2009). Cosmic star-formation history from a non-parametric inversion of infrared galaxy counts, *Astronomy and Astrophysics*, **504**, 727–740.
- Le Roux, E. (1961). Étude théorique du rayonnement synchrotron des radiosources, *Annales d'Astrophysique*, **24**, 71–85.
- Leavitt, H. S. (1912). Periods of 25 variable stars in the Small Magellanic Cloud, *Harvard College Observatory Circular*, No. **173**, 1–2.
- Lee, J., Koo, B., Raymond, J., *et al.* (2007). Subaru HDS observations of a Balmer-dominated shock in Tycho's supernova remnant, *Astrophysical Journal Letters*, **659**, L133–L136.
- Lee, J.-J., Raymond, J.C., Park, S., *et al.* (2010). Resolved shock structure of the Balmer-dominated filaments in Tycho's supernova remnant: Cosmic-ray precursor?, *Astrophysical Journal Letters*, **715**, L146–L149.
- Leger, A. & Puget, J. L. (1984). Identification of the ‘Unidentified’ IR emission features of interstellar dust?, *Astronomy and Astrophysics*, **137**, L5–L8.
- Legg, M. P. C. & Westfold, K. C. (1968). Elliptic polarization of synchrotron radiation, *Astrophysical Journal*, **154**, 499–514.
- Leibundgut, B. (2000). Type Ia supernovae, *Astronomy and Astrophysics Reviews*, **10**, 179–209.
- Leighton, R. (1959). *Introduction to Modern Physics*. San Francisco: Addison-Wesley.
- Lequeux, J., Peimbert, M., Rayo, J. F., *et al.* (1979). Chemical composition and evolution of irregular and blue compact galaxies, *Astronomy and Astrophysics*, **80**, 155–166.
- Lewin, W. H. G. and van der Klis, M. (eds) (2006). *Compact Stellar X-ray Sources*. Cambridge Astrophysics Series, No. 39. Cambridge: Cambridge University Press.
- Liedahl, D. A. (1999). The X-ray spectral properties of photoionized plasma and transient plasmas, in *X-Ray Spectroscopy in Astrophysics*, eds van Paradijs, J. & Bleeker, J. A. M., volume 520 of *Lecture Notes in Physics*. Berlin: Springer Verlag, pp. 189–268.
- Lightman, A. P. & Eardley, D. M. (1974). Black holes in binary systems: Instability of disk accretion, *Astrophysical Journal*, **187**, L1–L3.
- Lilly, S. & Cowie, L. (1987). Deep infrared surveys, in *Infrared Astronomy with Arrays*, eds Wynn-Williams, C. & Becklin, E., pp. 473–482. Honolulu: Institute for Astronomy, University of Hawaii Publications.
- Lilly, S. J. (1988). Discovery of a radio galaxy at a redshift of 3.395, *Astrophysical Journal*, **333**, L161–L167.

- Lilly, S. J., Tresse, L., Hammer, F., *et al.* (1995). The Canada–France Redshift Survey. VI. Evolution of the galaxy luminosity function to $z \sim 1$, *Astrophysical Journal*, **455**, 108–124.
- Lin, R. P., Krucker, S., Hurford, G. J., *et al.* (2003). RHESSI observations of particle acceleration and energy release in an intense solar gamma-ray line flare, *Astrophysical Journal Letters*, **595**, L69–L76.
- Liu, Q. Z., van Paradijs, J., & van den Heuvel, E. P. J. (2006). Catalogue of high-mass X-ray binaries in the Galaxy (4th edition), *Astronomy and Astrophysics*, **455**, 1165–1168.
- Longair, M. S. (1966). On the interpretation of radio source counts, *Monthly Notices of the Royal Astronomical Society*, **133**, 421–436.
- Longair, M. S. (1978). Radio astronomy and cosmology, in *Observational Cosmology: 8th Advanced Course, Swiss Society of Astronomy and Astrophysics, Saas-Fee 1978*, eds Maeder, A., Martinet, L., & Tammann, G., pp. 125–257. Geneva: Geneva Observatory Publications.
- Longair, M. S. (1981). *High Energy Astrophysics, 1st edition*. Cambridge: Cambridge University Press.
- Longair, M. S. (1988). The new astrophysics, in *The New Physics*, ed. Davies, P., pp. 94–208. Cambridge: Cambridge University Press.
- Longair, M. S. (1995). The physics of background radiation, in *The Deep Universe*, by Sandage, A.R., Kron, R.G. and Longair, M.S., eds Binggeli, B. & Buser, R., pp. 317–514, Saas - Fee Advanced Course 23. Berlin: Springer-Verlag.
- Longair, M. S. (1997a). Active galactic nuclei – The redshift one 3CR galaxies, *Astronomy and Geophysics*, **38**, 10–15.
- Longair, M. S. (1997b). *High Energy Astrophysics, Volume 1 (revised 2nd edition)*. Cambridge: Cambridge University Press.
- Longair, M. S. (1997c). *High Energy Astrophysics, Volume 2 (revised 2nd edition)*. Cambridge: Cambridge University Press.
- Longair, M. S. (2003). *Theoretical Concepts in Physics: An Alternative View of Theoretical Reasoning in Physics*. Cambridge: Cambridge University Press.
- Longair, M. S. (2006). *The Cosmic Century: A History of Astrophysics and Cosmology*. Cambridge: Cambridge University Press.
- Longair, M. S. (2008). *Galaxy Formation*, 2nd edition. Berlin: Springer-Verlag.
- Longair, M. S. & Riley, J. M. (1979). Statistical evidence on the dynamical evolution of extended radio sources, *Monthly Notices of the Royal Astronomical Society*, **188**, 625–635.
- Longair, M. S., Ryle, M., & Scheuer, P. A. G. (1973). Models of extended radiosources, *Monthly Notices of the Royal Astronomical Society*, **164**, 243–270.
- Lorimer, D. & Kramer, M. (2005). *Handbook of Pulsar Astronomy*. Cambridge: Cambridge University Press.
- Lotz, J. M., Madau, P., Giavalisco, M., *et al.* (2006). The rest-frame far-ultraviolet morphologies of star-forming galaxies at $z \sim 1.5$ and 4, *Astrophysical Journal*, **636**, 592–609.
- Lovelace, R. V. E. & Romanova, M. M. (2003). Relativistic Poynting jets from accretion disks, *Astrophysical Journal*, **596**, L159–L162.

- Lucek, S. G. & Bell, A. R. (2000). Non-linear amplification of a magnetic field driven by cosmic ray streaming, *Monthly Notices of the Royal Astronomical Society*, **314**, 65–74.
- Lund, N. (1984). Cosmic ray abundances, elemental and isotopic, in *Cosmic Radiation in Contemporary Astrophysics*, ed. Shapiro, M. M., pp. 1–26. Dordrecht: Reidel.
- Luo, D., McCray, D., & Slavin, J. (1994). The impact of SN 1987A with its interstellar ring, *Astrophysical Journal*, **430**, 264–276.
- Lyne, A. G., Burgay, M., Kramer, M., et al. (2004). A double-pulsar system: A rare laboratory for relativistic gravity and plasma physics, *Science*, **303**, 1153–1157.
- Lyne, A. G. & Graham-Smith, F. (2006). *Pulsar Astronomy*, 3rd edition. Cambridge: Cambridge University Press.
- Madau, P., Ferguson, H., Dickinson, M., et al. (1996). High-redshift galaxies in the *Hubble Deep Field*: Colour selection and star formation history to $z \sim 4$, *Monthly Notices of the Royal Astronomical Society*, **283**, 1388–1404.
- Maeder, A. & Meynet, G. (1989). Grid of evolutionary models from 0.85 to 120 solar masses—observational tests and the mass limits, *Astronomy and Astrophysics*, **210**, 155–173.
- Magorrian, J., Tremaine, S., Richstone, D., et al. (1998). The demography of massive dark objects in galaxy centers, *Astronomical Journal*, **115**, 2285–2305.
- Mahoney, W. A., Varnell, L. S., Jacobson, A. S., et al. (1988). Gamma-ray observations of Co-56 in SN 1987A, *Astrophysical Journal Letters*, **334**, L81–L85.
- Majewski, S. R., Munn, J. A., Kron, R. G., et al. (1991). A proper motion and variability QSO survey to $B = 22.5$, in *The Space Distribution of Quasars*, ed. Crampton, D., pp. 55–65. San Francisco: Astronomical Society of the Pacific Conference Series, vol. 21.
- Malkan, M. & Sargent, W. L. (1982). The ultraviolet excess of Seyfert 1 galaxies and quasars, *Astrophysical Journal*, **254**, 22–37.
- Manchester, R. (2005). CSIRO Research Highlights, 2005. See <http://www.atnf.csiro.au/research/highlights/2005/manchester/manchester.html>
- Manchester, R. N., Hobbs, G. B., Teoh, A. et al. (2005). The Australia National Facility Pulsar Catalogue, *Astronomical Journal*, **129**, 1993–2006.
- Manchester, R. N. & Taylor, J. H. (1977). *Pulsars*. San Francisco: W. H. Freeman.
- Margon, B. & Ostriker, J. P. (1973). The luminosity function of Galactic X-ray sources – A cutoff and a “standard candle”? *Astrophysical Journal*, **186**, 91–96.
- Markarian, B. E. (1967). Galaxies with an ultraviolet continuum, *Astrofizika*, **3**, 24–38.
- Markarian, B. E., Lipovetsky, V. A., & Stepanian, D. A. (1981). Galaxies with ultraviolet continuum XV, *Astrofizika*, **17**, 619–627. Translation: (1982), *Astrophysics*, **17**, 321–332.
- Marscher, A. P. (1993). Compact extragalactic radio jets, in *Astrophysical Jets*, ed. Bur-garella, D. Livio, M. and O’Dea, C. P., *Astrophysics and Space Science Library*, Vol. 103, pp. 73–94. Cambridge: Cambridge University Press.
- Marscher, A. P., Jorstad, S. G., Gómez, J., et al. (2002). Observational evidence for the accretion-disk origin for a radio jet in an active galaxy, *Nature*, **417**, 625–627.
- Marsh, T. R., Horne, K., Schlegel, E. M., et al. (1990). Doppler imaging of the dwarf nova U Geminorum, *Astrophysical Journal*, **364**, 637–646.

- Matt, G., Fabian, A. C., & Reynolds, C. S. (1997). Geometrical and chemical dependence of K-shell X-ray features, *Monthly Notices of the Royal Astronomical Society*, **289**, 175–184.
- Matthews, T. A., Morgan, W. W., & Schmidt, M. (1964). A discussion of galaxies identified with radio sources, *Astrophysical Journal*, **140**, 35–49.
- Matthews, T. A. & Sandage, A. R. (1963). Optical identification of 3C 48, 3C 196 and 3C 286 with stellar objects, *Astrophysical Journal*, **138**, 30–56.
- Matthewson, D. S. & Ford, V. L. (1970). Polarization observations of 1800 stars, *Memoirs of the Royal Astronomical Society*, **74**, 139–182.
- Matthiae, G. (2010). Observations of ultra high energy cosmic rays, *Journal of Physics Conference Series*, **203**, 012016, pp. 1–6.
- Matz, S. M., Share, G. H., Leising, M. D., et al. (1988). Gamma-ray line emission from SN 1987A, *Nature*, **331**, 416–418.
- Mayor, M. & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star, *Nature*, **378**, 355–359.
- McCarthy, P. J. (2006). Galaxy formation and cosmology in the ELT era, in *Scientific Requirements for Extremely Large Telescopes: IAU Symposium No. 232*, eds Whitelock, P., Dennefeld, M., & Leibundgut, B., pp. 119–129. Cambridge: Cambridge University Press.
- McCarthy, P. J., Le Borgne, D., Crampton, D., et al. (2004). Evolved galaxies at $z \geq 1.5$ from the Gemini Deep Deep Survey: The formation epoch of massive stellar systems, *Astrophysical Journal Letters*, **614**, L9–L12.
- McCarthy, P. J., van Breugel, W. J. M., Spinrad, H., et al. (1987). A correlation between the radio and optical morphologies of distant 3CR radio galaxies, *Astrophysical Journal*, **321**, L29–L33.
- McClintock, J. E. and Remillard, R. A. (2006). Black hole binaries, in *Compact Stellar X-ray Sources*, ed Lewin, W. H. G. and van der Klis, M., 157–213. Cambridge Astrophysics Series, No. 39. Cambridge: Cambridge University Press.
- McLeod, J. M. & Andrew, B. H. (1968). The radio source VRO 42.22.01, *Astrophysical Letters*, **1**, 243.
- McLure, R. J., Jarvis, M. J., Targett, T. A., et al. (2006). On the evolution of the black hole:spheroid mass ratio, *Monthly Notices of the Royal Astronomical Society*, **368**, 1395–1403.
- Melia, F. & Falcke, H. (2001). The supermassive black hole at the Galactic Center, *Annual Review of Astronomy and Astrophysics*, **39**, 309–352.
- Mellinger, A. (2007). Web-address: <http://home.arcor-online.de/axel.mellinger/>.
- Menjo, H., Miyahara, H., Kuwana, K., et al. (2005). Possibility of the detection of past supernova explosions by radiocarbon measurement, in *International Cosmic Ray Conference, Pune 2005*, volume 2, eds Sripathi Acharya, B., Gupta, S., Jagadeesan, P., Jain, A., Karthikeyan, S., Morris, S., and Tonwar, S. pp. 357–360. Mumbai: Tata Institute of Fundamental Research.
- Merritt, D. (1987). The Distribution of Dark Matter in the Coma Cluster, *Astrophysical Journal*, **313**, 121–135.
- Mestel, L. (1999). *Stellar Magnetism*. Oxford: Clarendon Press.

- Mészáros, P. (2002). Theories of gamma-ray bursts, *Annual Review of Astronomy and Astrophysics*, **40**, 137–169.
- Mészáros, P. & Rees, M. J. (1993). Gamma-ray bursts: Multiwaveband spectral predictions for blast wave models, *Astrophysical Journal*, **418**, L59–L62.
- Metcalfe, N., Shanks, T., Campos, A., *et al.* (1996). Galaxy formation at high redshifts, *Nature*, **383**, 236–237.
- Mewaldt, A. R. & Webber, R. W. (1990). Cosmic ray source abundances derived from high energy measurements of Fe-group nuclei, in *International Cosmic Ray Conference*, Vol. 3, ed. Protheroe, R. J., Adelaide: University of Adelaide. pp. 432–435.
- Meyer, P. (1979). Cosmic rays, in *Proceedings of the 16th International Conference on Cosmic Rays*, volume 2 of *International Cosmic Ray Conference*.
- Michell, J. (1784). On the means of discovering the distance, magnitude, etc. of the fixed stars, in consequence of the diminution of the velocity of their light, in case such a diminution should be found to take place in any of them, and such other data should be procured from observations, as would be farther necessary for that purpose, *Philosophical Transactions of the Royal Society*, **74**, 35–57.
- Michelson, P. (1994). High energy gamma ray emission from active galaxies: EGRET observations and implications, in *The Physics of Active Galaxies*, ed. Bicknell, G. V., Dopita, M. A., Quinn, P. J., pp. 13–21 of San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 54.
- Mihos, J. C. & Hernquist, L. (1994). Triggering of starbursts in galaxies by minor mergers, *Astrophysical Journal*, **425**, L13–L16.
- Mihos, J. C. & Hernquist, L. (1996). Gasdynamics and starbursts in major mergers, *Astrophysical Journal*, **464**, 641–663.
- Mikheyev, S. P. & Smirnov, A. Y. (1985). Resonance enhancement of oscillations in matter and solar neutrino spectroscopy, *Soviet Journal Nuclear Physics*, **42**, 913–917.
- Miller, G. E. & Scalo, J. M. (1979). The initial mass function and stellar birthrate in the solar neighborhood, *Astrophysical Journal Supplement Series*, **41**, 513–547.
- Miller, J. M., Fabian, A. C., Wijnands, R., *et al.* (2002). Evidence of spin and energy extraction in a galactic black hole candidate: The XMM-Newton/EPIC-pn spectrum of XTE J1650–500, *Astrophysical Journal Letters*, **570**, L69–L73.
- Miller, J. S. (1994). The unification of active galaxies: Seyferts and beyond, in *The Physics of Active Galaxies*, eds Bicknell, G. V., Dopita, M. A., & Quinn, P. J., pp. 149–157 of San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 54.
- Minkowski, R. (1960). A new distant cluster of galaxies, *Astrophysical Journal*, **132**, 908–908.
- Mioduszewski, A. J., Rupen, M. P., Walker, R. C., *et al.* (2004). A summer of SS433: Forty days of VLBA imaging, in *Bulletin of the American Astronomical Society*, **36**, 967.
- Mirabel, I. F. & Rodrigues, L. F. (1994). A superluminal source in the galaxy, *Nature*, **371**, 46–48.
- Mirabel, I. F. & Rodrigues, L. F. (1998). Microquasars in our galaxy, *Nature*, **392**, 673–676.
- Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation*. San Francisco: W.H. Freeman.

- Mitchell, R. J., Culhane, J. L., Davison, P. J. N., *et al.* (1976). Ariel 5 observations of the X-ray spectrum of the Perseus Cluster, *Monthly Notices of the Royal Astronomical Society*, **175**, 29P–34P.
- Miyoshi, M., Moran, J., Herrnstein, J., *et al.* (1995). Evidence for a black hole from high rotation velocities in a sub-parsec region of NGC4258, *Nature*, **373**, 127–129.
- Moore, C. E. & Merrill, P. W. (1968). *Partial Grotian Diagrams of Astrophysical Interest*. Washington: US Department of Commerce, National Bureau of Standards.
- Morgan, W. W. (1958). A preliminary classification of the forms of galaxies according to their stellar population, *Publications of the Astronomical Society of the Pacific*, **70**, 364–391.
- Mukai, K., Wood, J. H., Naylor, T., *et al.* (1997). The X-ray eclipse of the dwarf nova HT Cassiopeiae: Results from ASCA and ROSAT HRI observations, *Astrophysical Journal*, **475**, 812–822.
- Murray, C. A. (1983). *Vectorial Astrometry*. Bristol: Adam Hilger.
- Mushotzky, R. (1980). The X-ray spectra of clusters of galaxies, in *X-ray Astronomy*, eds Giacconi, R. & Setti, G., pp. 171–179. Dordrecht: Reidel.
- Nagano, M. & Watson, A. A. (2000). Observations and implications of the ultrahigh-energy cosmic rays, *Reviews of Modern Physics*, **72**, 689–732.
- Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., *et al.* (1995). Discovery of a cool brown dwarf, *Nature*, **378**, 463–465.
- Narayan, R. (1991). Instabilities in thick disks, in *Structure and Emission Properties of Accretion Disks*, Proceedings of IAU Colloquium 129, the 6th Institute d’Astrophysique de Paris (IAP) Meeting, ed. Bertout, C., Collin-Souffrin, S., and Lasota, J. P., 153–160. Gif-sur-Yvette: Editions Frontières.
- Narayan, R. and Goodman, J. (1989). Non-axisymmetric shear instabilities in thick accretion disks, in *Theory of Accretion Disk*, ed F. Meyer, 231–247. Proceedings of a NATO Advanced Research Workshop, volume 290. Dordrecht: Kluwer.
- Narayan, R., Igumenshev, I. V., & Abramowicz, M. A. (2000). Self-similar accretion flows with accretion, *Astrophysical Journal*, **539**, 798–808.
- Narayan, R. & Yi, I. (1994). Advection-dominated accretion: a self-similar solution, *Astrophysical Journal Letters*, **428**, L13–L16.
- Neininger, N. (1992). The magnetic field structure of M51, *Astronomy and Astrophysics*, **263**, 30–36.
- Newman, E. T., Couch, K., Chinnapared, K., *et al.* (1965). Metric of a rotating charged mass, *Journal of Mathematical Physics*, **6**, 918–919.
- Nicolet, B. (1980). A plot of UBV diagram, *Astronomy and Astrophysics Supplement*, **42**, 283–284.
- Northrop, T. G. (1963). *The Adiabatic Motion of Charged Particles*. New York: Interscience.
- Novikov, I. D. & Thorne, K. S. (1973). Astrophysics of black Holes., in *Black Holes*, eds DeWitt, C. & DeWitt, B. S., pp. 343–450. New York: Gordon and Breach.
- Oemler, A. J. (1974). The systematic properties of clusters of galaxies. Photometry of 15 clusters, *Astrophysical Journal*, **194**, 1–20.
- Ohira, Y., Terasawa, T., & Takahara, F. (2009). Plasma instabilities as a result of charge exchange in the downstream region of supernova remnant shocks, *Astrophysical Journal Letters*, **703**, L59–L62.

- Oliver, S. J., Rowan-Robinson, M., & Saunders, W. (1992). Infrared background constraints on the evolution of IRAS galaxies, *Monthly Notices of the Royal Astronomical Society*, **256**, 15P–22P.
- Oort, J. H. & Walraven, T. (1956). Polarization and composition of the Crab Nebula, *Bulletin of the Astronomical Institutes of the Netherlands*, **12**, 285–311.
- Orosz, J. A. (2007). Home-page of Jerome A. Orosz. <http://mintaka.sdsu.edu/faculty/orosz/web/>.
- Orosz, J. A., McClintock, J. E., Narayan, R., et al. (2007). A 15.65-solar-mass black hole in an eclipsing binary in the nearby spiral galaxy M33, *Nature*, **449**, 872–875.
- Orr, M. J. L. & Browne, I. W. A. (1982). Relativistic beaming and quasar statistics, *Monthly Notices of the Royal Astronomical Society*, **200**, 1067–1080.
- Osmer, P. S. (1982). Evidence for a decrease in the space density of quasars at z more than about 3.5, *Astrophysical Journal*, **253**, 28–37.
- Osterbrock, D. E. (1978). Optical emission-line spectra of Seyfert galaxies and radio galaxies, *Physica Scripta*, **17**, 137–143.
- Osterbrock, D. E. & Ferland, G. J. (2005). *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. Mill Valley, California: University Science Books.
- Ostriker, J. P. & Peebles, P. J. E. (1973). A numerical study of the stability of flattened galaxies: or, can cold galaxies survive?, *Astrophysical Journal*, **186**, 467–480.
- Owen, F. N. & Ledlow, M. J. (1994). The FR I/II break and the bivariate luminosity function in Abell Clusters of galaxies, in *First Stromlo Symposium: Physics of Active Galactic Nuclei*, eds Bicknell, G. V., Dopita, M. A., & Quinn, P. J., pp. 319–323. San Francisco: Astronomical Society of the Pacific Conference Series, Vol. 34.
- Owen, F. N., Ledlow, M. J., Morrison, G. E. et al. (1997). The cluster of galaxies surrounding Cygnus A, *Astrophysical Journal Letters*, **488**, L15–L17.
- Pachoczyk, A. G. (1970). *Radio Astrophysics*. San Francisco: W. H. Freeman.
- Pacini, F. (1967). Energy emission from a neutron star, *Nature*, **216**, 567–568.
- Pacini, F. (1968). Rotating neutron stars, pulsars and supernova remnants, *Nature*, **219**, 145–146.
- Page, L. (1997). Review of observations of the cosmic microwave background, in *Critical Dialogues in Cosmology*, ed. Turok, N., pp. 343–362. Singapore: World Scientific.
- Pagel, B. (1997). *Nucleosynthesis and Chemical Evolution of Galaxies*. Cambridge: Cambridge University Press.
- Panagia, N., Gilmozzi, R., Macchietto, F., et al. (1991). Properties of the SN 1987A circumstellar ring and the distance to the Large Magellanic Cloud, *Astrophysical Journal*, **380**, L23–L26.
- Papaloizou, J. C. B. & Pringle, J. E. (1984). The dynamical stability of differentially rotating discs with constant specific angular momentum, *Monthly Notices of the Royal Astronomical Society*, **208**, 721–750.
- Parker, E. N. (1957). Sweet's mechanism for merging magnetic fields in conducting fluids, *Journal of Geophysical Research*, **62**, 509–520.
- Parker, E. N. (1965). Cosmic rays and their formation of a Galactic halo, *Astrophysical Journal*, **142**, 584–590.
- Parker, E. N. (1979). *Cosmical Magnetic Fields*. Oxford: Clarendon Press.

- Pearson, T. J., Unwin, S. C., Cohen, M. H., *et al.* (1981). Superluminal expansion of quasar 3C 273, *Nature*, **290**, 365–368.
- Pearson, T. J., Unwin, S. C., Cohen, M. H., *et al.* (1982). Superluminal expansion of 3C 273, in *Extragalactic Radio Sources*, eds Heeschen, D. S. & Wade, C. M., pp. 355–356. Dordrecht: Reidel.
- Pengelly, R. M. (1964). Recombination spectra, I, *Monthly Notices of the Royal Astronomical Society*, **127**, 145–163.
- Penrose, R. (1969). Gravitational collapse: The role of general relativity, *Rivista Nuovo Cimento*, **1**, 252–276.
- Penzias, A. A. & Wilson, R. W. (1965). A measurement of excess antenna temperature at 4080 MHz, *Astrophysical Journal*, **142**, 419–421.
- Perley, R. A., Dreher, J. W., & Cowan, J. J. (1984). The jet and filaments in Cygnus A, *Astrophysical Journal*, **285**, L35–L38.
- Perlmutter, S., Gabi, S., Goldhaber, G., *et al.* (1997). Measurements of the cosmological parameters omega and lambda from the first seven supernovae at $z > 0.35$, *Astrophysical Journal*, **483**, 565–581.
- Perlmutter, S. (2003). Supernovae, dark energy, and the accelerating universe, *Physics Today*, **56**, 53–62.
- Peterson, B. M. (1997). *An Introduction to Active Galactic Nuclei*. Cambridge: Cambridge University Press.
- Peterson, B. M., Balonek, T. J., & 63 authors (1991). Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. II – An intensive study of NGC 5548 at optical wavelengths, *Astrophysical Journal*, **368**, 119–137.
- Petschek, H. E. (1964). Magnetic field annihilation, in *The Physics of Solar Flares*, Proceedings of AAS-NASA Symposium, ed. Hess, W. N. 425–439. NASA SP-50.
- Phillips, M. M. (1993). The absolute magnitudes of Type IA supernovae, *Astrophysical Journal*, **413**, L105–L108.
- Pierre Auger Collaboration (2007). Correlation of the highest-energy cosmic rays with nearby extragalactic objects, *Science*, **318**, 938–943.
- Plüschke, S., Diehl, R., Schönfelder, V., *et al.* (2001). The COMPTEL 1.809 MeV survey, in *Exploring the Gamma-Ray Universe*, eds Gimenez, A., Reglero, V., & Winkler, C., pp. 55–58. *ESA Special Publication*, Vol. 459.
- Powell, C. F., Fowler, P., & Perkins, D. (1959). *The Study of Elementary Particles by the Photographic Method*. Oxford: Pergamon Press.
- Pozdnyakov, L. A., Sobol, I. M., & Sunyaev, R. A. (1983). Comptonization and the shaping of X-ray source spectra – Monte Carlo calculations, *Soviet Scientific Reviews, Section E: Astrophysics and Space Physics Reviews*, **2**, 189–31.
- Pratt, G. W. & Arnaud, M. (2002). The mass profile of A1413 observed with XMM-Newton: Implications for the $M-T$ relation, *Astronomy and Astrophysics*, **394**, 375–393.
- Price, P. & Fleischer, R. (1971). Identification of energetic heavy nuclei with solid dielectric track detectors: Applications to astrophysical and planetary studies, *Annual Review of Nuclear Science*, **21**, 295–334.
- Priest, E. & Forbes, T. (2000). *Magnetic Reconnection*. Cambridge: Cambridge University Press.

- Priest, E. R. (1982). *Solar Magneto-Hydrodynamics*. Dordrecht: Reidel, Geophysics and Astrophysics Monographs, Volume 21.
- Priest, E. R. & Forbes, T. G. (1986). New models for fast steady state magnetic reconnection, *Journal of Geophysical Research*, **91**, 5579–5588.
- Pringle, J. E. (1981). Accretion discs in astrophysics, *Annual Review of Astronomy and Astrophysics*, **19**, 137–162.
- Pringle, J. E. & King, A. R. (2007). *Astrophysical Flows*. Cambridge: Cambridge University Press.
- Puget, J.-L., Abergel, A., Bernard, J.-P., *et al.* (1996). Tentative detection of a cosmic far-infrared background with COBE, *Astronomy and Astrophysics*, **308**, L5–L8.
- Pye, J. P., McGale, P. A., Allan, D. J., *et al.* (1995). The ROSAT Wide Field Camera all-sky survey of extreme-ultraviolet sources – II. The 2RE Source Catalogue, *Monthly Notices of the Royal Astronomical Society*, **274**, 1165–1193.
- Quest, K. B. & Shapiro, V. D. (1996). Evolution of the fire-hose instability: Linear theory and wave-wave coupling, *Journal of Geophysical Research*, **101**, 24457–24470.
- Ramana Murthy, P. V. & Wolfendale, A. W. (1993). *Gamma-Ray Astronomy*, 2nd edition. Cambridge: University Press.
- Ramaty, R. & Lingenfelter, R. E. (1979). Gamma-ray line astronomy, *Nature*, **278**, 127–132.
- Ratcliffe, J. A. (1972). *An Introduction to the Ionosphere and Magnetosphere*. Cambridge: Cambridge University Press.
- Reedy, R., Arnold, J., & Lal, D. (1983). Cosmic-ray record in Solar System matter, *Annual Review of Nuclear Science*, **33**, 505–537.
- Rees, M. J. (1967). Studies in radio source structure – I. A relativistically expanding model for variable quasi-stellar radio sources, *Monthly Notices of the Royal Astronomical Society*, **135**, 345–360.
- Rees, M. J. (1984). Black hole models for active galactic nuclei, *Annual Review of Astronomy and Astrophysics*, **22**, 471–506.
- Reimer, P. J., Baillie, M. G. L., Bard, E., *et al.* (2004). IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP, *Radiocarbon*, **46**, 1029–1058.
- Remillard, R. A. & McClintock, J. E. (2006). X-ray properties of black-hole binaries, *Annual Review of Astronomy and Astrophysics*, **44**, 49–92.
- Rest, A., Welch, D. L., Suntzeff, N. B., *et al.* (2008). Scattered-light echoes from the historical galactic supernovae Cassiopeia A and Tycho (SN 1572), *Astrophysical Journal Letters*, **681**, L81–L84.
- Reynolds, R. J. (1990). The low density ionized component of the interstellar medium and free–free absorption at high galactic latitudes, in *Low Frequency Astrophysics from Space*, eds Kassim, N. E. & Weiler, K. W., volume 362 of *Lecture Notes in Physics*, Berlin: Springer Verlag, pp. 121–129.
- Richards, G. T., Strauss, M. A., Fan, X., *et al.* (2006). The Sloan Digital Sky Survey Quasar Survey: Quasar luminosity function from data release 3, *Astronomical Journal*, **131**, 2766–2787.
- Rindler, W. (2001). *Relativity: Special, General and Cosmological*. Oxford: Oxford University Press.

- Roberts, M. S. & Haynes, M. P. (1994). Physical parameters along the Hubble sequence, *Annual Review of Astronomy and Astrophysics*, **26**, 115–152.
- Robinson, I., Schild, A. and Schucking, E. L. (eds) (1965). *Quasi-Stellar Sources and Gravitational Collapse*. Chicago: University of Chicago Press.
- Robson, I. E. (1999). *Active Galactic Nuclei*. Chichester: John Wiley, in association with Praxis Publishing.
- Rochester, G. & Butler, C. (1947). Evidence for the existence of new unstable elementary particles, *Nature*, **160**, 855–857.
- Rossi, B. & Greisen, K. (1941). Cosmic-ray theory, *Reviews of Modern Physics*, **13**, 240–309.
- Rowan-Robinson, M. (1968). The determination of the evolutionary properties of quasars by means of the luminosity–volume test, *Monthly Notices of the Royal Astronomical Society*, **141**, 445–458.
- Rowan-Robinson, M. (1985). *The Cosmological Distance Ladder*. New York: W. H. Freeman.
- Rowan-Robinson, M. (1988). The extragalactic distance scale, *Space Science Reviews*, **48**, 1–71.
- Ruderman, M. A. & Sutherland, P. G. (1975). Theory of pulsars – Polar caps, sparks, and coherent microwave radiation, *Astrophysical Journal*, **196**, 51–72.
- Rybicki, G. B. & Lightman, A. P. (1979). *Radiative Processes in Astrophysics*. New York: John Wiley.
- Sahu, K. C., Livio, M., Petro, L., et al. (1997). The optical counterpart to gamma-ray burst GRB 970228 observed using the Hubble Space Telescope, *Nature*, **387**, 476–478.
- Sajina, A., Scott, D., Dennefeld, M., et al. (2006). The 1–1000 μm spectral energy distributions of far-infrared galaxies, *Monthly Notices of the Royal Astronomical Society*, **369**, 939–957.
- Salpeter, E. E. (1955). The luminosity function and stellar evolution, *Astrophysical Journal*, **121**, 161–167.
- Salpeter, E. E. (1964). Accretion of interstellar matter by massive objects, *Astrophysical Journal*, **140**, 796–800.
- Sandage, A. (1957). Observational approach to evolution. II. A computed luminosity function for K0-K2 stars from $M_{\{v\}} = +5$ to $M_{\{v\}} = -4.5$., *Astrophysical Journal*, **125**, 435–444.
- Sandage, A. R. (1965). The existence of a major new constituent of the Universe: The quasistellar galaxies, *Astrophysical Journal*, **141**, 1560–1578.
- Sanders, D. B. & Mirabel, I. F. (1996). Luminous infrared galaxies, *Annual Review of Astronomy and Astrophysics*, **34**, 749–792.
- Sanders, D. B., Soifer, B. T., Elias, J. H., et al. (1988). Ultraluminous infrared galaxies and the origin of quasars, *Astrophysical Journal*, **325**, 74–91.
- Sargent, W. L. W. (1970). A spectroscopic survey of compact and peculiar galaxies, *Astrophysical Journal*, **160**, 405–427.
- Sargent, W. L. W., Young, P. J., Lynds, C. R., et al. (1978). Dynamical evidence for a central mass concentration in the galaxy M87, *Astrophysical Journal*, **221**, 731–744.

- Saunders, W., Rowan-Robinson, M., Lawrence, A., *et al.* (1990). The 60-micron and far-infrared luminosity functions of IRAS galaxies, *Monthly Notices of the Royal Astronomical Society*, **242**, 318–337.
- Savage, B. D. & de Boer, K. S. (1979). Observational evidence for a hot gaseous Galactic corona, *Astrophysical Journal Letters*, **230**, L77–L82.
- Scheuer, P. A. G. (1966). Radiation processes in radio astronomy, in *Plasma Astrophysics: Proceedings of the International School of Physics ‘Enrico Fermi’*, ed. Sturrock, P. A., volume 39, pp. 289–306. New York: Academic Press.
- Scheuer, P. A. G. (1974). Models of extragalactic radio sources with a continuous energy supply from a central object, *Monthly Notices of the Royal Astronomical Society*, **166**, 513–528.
- Scheuer, P. A. G. (1982). Morphology and power of radio sources, in *Extragalactic Radio Sources*, ed. Heeschen, D. S. and Wade, C. M., volume 97 of *IAU Symposium*, pp. 163–165. Dordrecht: Reidel.
- Scheuer, P. A. G. & Readhead, A. C. S. (1979). Superluminally expanding radio sources and the radio-quiet QSOs, *Nature*, **277**, 182–185.
- Schmidt, B. P., Kirshner, R. P., & Eastman, R. G. (1992). Expanding photospheres of Type II supernovae and the extragalactic distance scale, *Astrophysical Journal*, **395**, 366–386.
- Schmidt, M. (1959). The rate of star formation, *Astrophysical Journal*, **129**, 243–258.
- Schmidt, M. (1963). 3C 273: A star-like object with large red-shift, *Nature*, **197**, 1040–1040.
- Schmidt, M. (1965). Large redshifts of five quasi-stellar sources, *Astrophysical Journal*, **141**, 1295–1300.
- Schmidt, M. (1968). Space distribution and luminosity functions of quasi-stellar sources, *Astrophysical Journal*, **151**, 393–409.
- Schmidt, M. & Green, R. F. (1983). Quasar evolution derived from the Palomar Bright Quasar Survey and other complete quasar surveys, *Astrophysical Journal*, **269**, 352–374.
- Schmidt, M., Schneider, D. P., & Gunn, J. E. (1995). Spectroscopic CCD surveys for quasars at large redshift. IV. Evolution of the luminosity function from quasars detected by their Lyman-alpha emission, *Astronomical Journal*, **110**, 68–77.
- Schneider, D., Schmidt, M., & Gunn, J. E. (1991). PC 1247+3406 – An optically selected quasar with a redshift of 4.897, *Astronomical Journal*, **102**, 837–840.
- Schneider, P. (2006). *Extragalactic Astronomy and Cosmology*. Berlin: Springer–Verlag.
- Schneider, P., Kochanek, C. S., & Wambsganss, J. (2006). *Gravitational Lensing: Strong, Weak and Micro*. Saas-Fee Advanced Course 33: eds. Meylan, G. and Jetzer, P. and North, P. Berlin: Springer–Verlag.
- Schödel, R., Ott, T., Genzel, R., *et al.* (2002). A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way, *Nature*, **419**, 694–696.
- Schönberg, M. & Chandrasekhar, S. (1942). On the evolution of the main-sequence stars, *Astrophysical Journal*, **96**, 161–171.
- Schreier, E., Levinson, R., Gursky, H., *et al.* (1972). Evidence for the binary nature of Centaurus X-3 from UHURU X-Ray Observations, *Astrophysical Journal*, **172**, L79–L89.

- Schroeder, D. (2000). *Astronomical Optic*. 2nd edn. San Diego: Academic Press.
- Schwarzschild, K. (1916). Über das Gravitationsfeld einis Massenpunktes nach der Einsteinschen Theorie (On the gravitational field of a point mass according to Einsteinian theory), *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, **1**, 189–196.
- Schwarzschild, M. (1979). A numerical model for a triaxial stellar system in dynamical equilibrium, *Astrophysical Journal*, **232**, 236–228.
- Sedov, L. I. (1959). *Similarity and Dimensional Methods in Mechanics*. New York: Academic Press.
- Sekido, Y. & Elliot, H. (1985). *Early History of Cosmic Ray Studies*. Dordrecht: Reidel.
- Sellgren, K. (1984). The near-infrared continuum emission of visual reflection nebulae, *Astrophysical Journal*, **277**, 623–633.
- Serkowski, K. (1973). Interstellar polarization, in *Interstellar Dust and Related Topics, IAU Symposium No. 52*, eds Greenberg, J. M. & van der Hulst, H. C., pp. 145–152. Dordrecht: Reidel.
- Serkowski, K., Mathewson, D. S., & Ford, V. L. (1975). Wavelength dependence of interstellar polarization and ratio of total to selective extinction, *Astrophysical Journal*, **196**, 261–290.
- Sérsic, J. L. (1968). *Atlas de Galaxias Australes*. Cordoba, Argentina: Observatorio Astronomico.
- Seward, F. D. & Charles, P. A. (2010). *Exploring the X-Ray Universe*. 2nd edn. Cambridge: Cambridge University Press.
- Shakura, N. & Sunyaev, R. A. (1973). Black holes in binary systems. Observational appearance, *Astronomy and Astrophysics*, **24**, 337–355.
- Shapiro, M. M. (1991). A brief introduction to the cosmic radiation, in *Cosmic Rays, Supernovae and the Interstellar Medium*, eds Shapiro, M. M., Silberberg, R., & Wefel, J. P., pp. 1–28. Dordrecht: Kluwer.
- Shapiro, P. R. & Field, G. B. (1976). Consequences of a new hot component of the interstellar medium, *Astrophysical Journal*, **205**, 762–765.
- Shapiro, S. L. & Teukolsky, S. A. (1983). *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects*. New York: Wiley-Interscience.
- Shimasaku, K., Ouchi, M., Furusawa, H., et al. (2005). Number density of bright Lyman-break galaxies at $z \sim 6$ in the Subaru Deep Field, *Publications of the Astronomical Society of Japan*, **57**, 447–458.
- Shklovsky, I. S. (1953). On the nature of the radiation from the Crab Nebula, *Dokladi Akademii Nauk SSSR*, **90**, 983–986.
- Shu, F. H. (1992). *Physics of Astrophysics*, Vol. II. Mill Valley, California: University Science Books.
- Shu, F. H., Adams, F. C., & Lizano, S. (1987). Star formation in molecular clouds – Observation and theory, *Annual Review of Astronomy and Astrophysics*, **25**, 23–81.
- Silberberg, R., Tsao, C. H., & Letaw, J. R. (1988). Recent improvement of spallation cross section calculations, applicable to cosmic ray physics, in *NATO ASIC Proc. 220: Genesis and Propagation of Cosmic Rays*, eds Shapiro, M. M. & Wefel, J. P., pp. 357–374.

- Silva, D. R. & Cornell, M. E. (1992). A new library of stellar optical spectra, *Astrophysical Journal Supplement Series*, **81**, 865–881.
- Simpson, J. (1983). Elemental and isotopic composition of Galactic cosmic rays, *Annual Reviews of Nuclear and Particle Science*, **33**, 323–381.
- Skilling, J. (1971). Cosmic rays in the Galaxy: Convection or diffusion?, *Astrophysical Journal*, **170**, 265–273.
- Skobelzyn, D. (1929). Über eine neue Art sehr schneller β -strahlen (On a new type of very fast β -rays), *Zeitschrift für Physik*, **54**, 686–702.
- Smail, I., Ivison, R. J., & Blain, A. W. (1997). A deep sub-millimeter survey of lensing clusters: A new window on galaxy formation and evolution, *Astrophysical Journal Letters*, **490**, L5–L8.
- Smart, W. (1977). *Textbook on Spherical Astronomy*, 6th edition, with revisions by R.M. Green. Cambridge: Cambridge University Press.
- Smith, H. J. & Hoffleit, D. (1963). Light variations in the superluminous radio galaxy 3C 273, *Nature*, **198**, 650–651.
- Snellen, I. A. G., Mack, K., Schilizzi, R. T., et al. (2004). The CORALZ sample – I. Young radio-loud active galactic nuclei at low redshift, *Monthly Notices of the Royal Astronomical Society*, **348**, 227–234.
- Soifer, B. T., Sanders, D. B., Madore, B. F., et al. (1987). The IRAS bright galaxy sample. II – The sample and luminosity function, *Astrophysical Journal*, **320**, 238–257.
- Soltan, A. (1982). Masses of quasars, *Monthly Notices of the Royal Astronomical Society*, **200**, 115–122.
- Sparke, L. & Gallagher, J. (2000). *Galaxies in the Universe: An Introduction*. Cambridge: Cambridge University Press.
- Spinrad, H., Dey, A., & Graham, J. R. (1995). Keck observations of the most distant galaxy: 8C 1435+63 at $z = 4.25$, *Astrophysical Journal*, **438**, L51–L54.
- Spitzer, L. (1962). *Physics of Fully Ionized Gases*, 2nd edition. New York: Interscience.
- Spitzer, L. (1968). *Diffuse Matter in Space*. New York: Interscience.
- Spitzer, L. & Härm, R. (1953). Transport phenomena in a completely ionized gas, *Physical Review*, **89**, 977–981.
- Spitzer, L. J. & Hart, M. H. (1971). Random gravitational encounters and the evolution of spherical systems. I. Method, *Astrophysical Journal*, **164**, 399–409.
- Springel, V., White, S. D. M., Jenkins, A., et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature*, **435**, 629–636.
- Stahler, S. W. & Palla, F. (2005). *The Formation of Stars*. New York: Interscience.
- Stahler, S. W., Shu, F. H., & Taam, R. E. (1980). The evolution of protostars. I – Global formulation and results, *Astrophysical Journal*, **241**, 637–654.
- Stairs, I. H. (2004). Pulsars in binary systems: Probing binary stellar evolution and general relativity, *Science*, **304**, 547–552.
- Starrfield, S. (1988). The classical nova outburst, in *Multiwavelength Astrophysics*, ed. Cordova, F. A., pp. 159–188, Cambridge: Cambridge University Press.
- Stecker, F. W. (1975). Gamma ray astrophysics, in *Origin of Cosmic Rays*, eds Osborne, J. L. & Wolfendale, A. W., pp. 267–334, Dordrecht: Reidel.

- Stecker, F. W. & Salamon, M. H. (1999). Photodisintegration of ultra-high-energy cosmic rays: A new determination, *Astrophysical Journal*, **512**, 521–526.
- Steidel, C. (1998). Galaxy evolution: Has the ‘epoch of galaxy formation’ been found?, in *Eighteenth Texas Symposium on Relativistic Astrophysics and Cosmology*, eds Olinto, A., Frieman, J., & Schramm, D., pp. 124–135. River Edge, N.J. : World Scientific.
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., et al. (1999). Lyman-break galaxies at $z \geq 4$ and the evolution of the ultraviolet luminosity density at high redshift, *Astrophysical Journal*, **519**, 1–17.
- Steigman, G. (2004). Big Bang nucleosynthesis: Probing the first 20 minutes, in *Measuring and Modeling the Universe*, ed. Freedman, W. L., pp. 169–195. Cambridge: Cambridge University Press.
- Stephenson, F. R. & Green, D. A. (2002). *Historical Supernovae and their Remnants*. Oxford: Clarendon Press.
- Stockton, A. & Ridgway, S. (1996). Optical and near IR observations of Cygnus A, in *Cygnus A – Study of a Radio Galaxy*, ed. Carilli, C. L. & Harris, D. E., pp. 1–4, Cambridge: Cambridge University Press.
- Strong, A. W., Moskalenko, I. V., & Reimer, O. (2000). Diffuse continuum gamma rays from the galaxy, *Astrophysical Journal*, **537**, 763–784.
- Strong, A. W., Moskalenko, I. V., & Reimer, O. (2004). Diffuse Galactic continuum gamma rays: A model compatible with EGRET data and cosmic-ray measurements, *Astrophysical Journal*, **613**, 962–976.
- Stuiver, M., Reimer, P. J., & Braziunas, T. F. (1998). Radiocarbon age calibration for terrestrial and marine samples, *Radiocarbon*, **40**, 1127–1151.
- Suganuma, M., Yoshii, Y., Kobayashi, Y., et al. (2006). Reverberation measurements of the inner radius of the dust torus in nearby Seyfert 1 galaxies, *Astrophysical Journal*, **639**, 46–63.
- Sunyaev, R. A. (1980). The microwave background radiation in the direction toward clusters of galaxies, *Soviet Astronomy Letters*, **6**, 213–216.
- Sunyaev, R. A. & Titarchuk, L. G. (1980). Comptonization of X-rays in plasma clouds – Typical radiation spectra, *Astronomy and Astrophysics*, **86**, 121–138.
- Sunyaev, R. A. & Zeldovich, Y. B. (1980). Microwave background radiation as a probe of the contemporary structure and history of the Universe, *Annual Review of Astronomy and Astrophysics*, **18**, 537–560.
- Sutherland, R. S. (1998). Accurate free-free Gaunt factors for astrophysical plasmas, *Monthly Notices of the Royal Astronomical Society*, **300**, 321–330.
- Sweet, P. A. (1958). The neutral point theory of solar flares, in *Electromagnetic Phenomena in Cosmical Physics*, ed. Lehnert, B., volume 6 of *IAU Symposium*, pp. 123–134.
- Tananbaum, H., Gursky, H., Kellogg, E. M., et al. (1972). Discovery of a periodic binary X-ray source in Hercules from UHURU, *Astrophysical Journal*, **174**, L144–L149.
- Tandberg-Hanssen, E. & Emslie, A. G. (1988). *The Physics of Solar Flares*. Cambridge: Cambridge University Press.
- Tanvir, N. R., Fox, D. B., Levan, A. J., et al. (2009). A γ -ray burst at a redshift of $z \sim 8.2$, *Nature*, **461**, 1254–1257.

- Taylor, R. J. (1972). *The Origin of the Chemical Elements*. The Wykeham Science Series, London: Wykeham Publications.
- Taylor, R. J. (1994). *The Stars: Their Structure and Evolution*. Cambridge: Cambridge University Press.
- Taylor, G. I. (1950a). The formation of a blast wave by a very intense explosion. I. Theoretical discussion, *Proceedings of the Royal Society of London, A* **201**, 159–174.
- Taylor, G. I. (1950b). The formation of a blast wave by a very intense explosion. II. The atomic explosion of 1945, *Proceedings of the Royal Society of London, A* **201**, 175–186.
- Taylor, J. H. & Cordes, J. M. (1993). Pulsar distances and the Galactic distribution of free Electrons, *Astrophysical Journal*, **411**, 674–684.
- Thompson, C. & Duncan, R. C. (1995). The soft gamma repeaters as very strongly magnetized neutron stars – I. Radiative mechanism for outbursts, *Monthly Notices of the Royal Astronomical Society*, **275**, 255–300.
- Thompson, C. & Duncan, R. C. (1996). The soft gamma repeaters as very strongly magnetized neutron stars – II. Quiescent neutrino, X-Ray, and Alfvén wave emission, *Astrophysical Journal*, **473**, 322–342.
- Thomson, J. J. (1906). *Conduction of Electricity Through Gases*. Cambridge: Cambridge University Press.
- Thorne, K., Price, R., & Macdonald, D. (1986). *Black Holes: The Membrane Paradigm*. New Haven: Yale University Press.
- Toller, G. N. (1990). Optical observations of Galactic and extragalactic light – Implications for Galactic structure, in *The Galactic and Extragalactic Background Radiation*, eds Bowyer, S. & Leinert, C., *IAU Symposium No. 139*, pp. 21–34. Dordrecht: Kluwer.
- Toomre, A. & Toomre, J. (1972). Galactic bridges and tails, *Astrophysical Journal*, **178**, 623–666.
- Tremaine, S. & Gunn, J. (1979). Dynamical role of light neutral leptons in cosmology, *Physical Review Letters*, **42**, 407–410.
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. (2004). The origin of the mass–metallicity relation: Insights from 53,000 star-forming galaxies in the Sloan Digital Sky Survey, *Astrophysical Journal*, **613**, 898–913.
- Trodden, M. (2006). Physics of the very early Universe: What can we learn from particle collider experiments?, *Proceedings of Science, CMB2006*, 1–9. This electronic publication can be found at <http://pos.sissa.it/archive/conferences/027/003/CMB2006-003.pdf>.
- Tsao, C. H. & Silberberg, R. (1979). Improved semiempirical estimates of cross sections, in *International Cosmic Ray Conference*, Vol. 2, ed. Kamata, K., pp. 202–205. Tokyo: Institute of Cosmic Ray Research.
- Tully, R. B. & Fisher, J. R. (1977). A new method of determining distances to galaxies, *Astronomy and Astrophysics*, **54**, 661–673.
- Turland, B. D. & Scheuer, P. A. G. (1976). Instabilities of Kelvin–Helmholtz type for relativistic streaming, *Monthly Notices of the Royal Astronomical Society*, **176**, 421–441.

- Ulrich, M. H., Boksenberg, A., Bromage, G. E., *et al.* (1984a). Detailed observations of NGC 4151 with IUE – III. Variability of the strong emission lines from 1978 February to 1980 May, *Monthly Notices of the Royal Astronomical Society*, **206**, 221–238.
- Ulrich, M. H., Boksenberg, A., Bromage, G. E., *et al.* (1984b). Detailed observations of NGC 4151 with IUE – III. Variability of the strong emission lines from 1978 February to 1980 May, *Monthly Notices of the Royal Astronomical Society*, **209**, 479.
- van den Heuvel, E. P. J. (1987). Millisecond pulsar formation and evolution, in *The Origin and Evolution of Neutron Stars, IAU Symposium No. 125*, eds Helfand, D. J. & Huang, J.-H., pp. 393–404, Dordrecht: Reidel.
- van der Klis, M. (2000). Millisecond oscillations in X-ray binaries, *Annual Review of Astronomy and Astrophysics*, **38**, 717–760.
- Vashakidze, M. A. (1954). On the degree of polarization of the light near extragalactic nebulae and the Crab Nebula, *Astronomicheskikh Tsirkular*, No. **147**, 11–13.
- Veilleux, S. (1999). Spectroscopy of luminous infrared galaxies, in *Galaxy Interactions at Low and High Redshift*, eds Barnes, J. E. & Sanders, D. B., volume 186 of *IAU Symposium*, pp. 295–301, Dordrecht: Reidel.
- Velikhov, E. P. (1959). Stability of an ideally conducting liquid flowing between cylinders rotating in a magnetic field, *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, **36**, 1398–1404. Translation: (1959), *Soviet Physics – JETP*, **9**, 995–998.
- Venturi, T., Cotton, W. D., Feretti, L., *et al.* (1996). VLBI observations of FRI radio galaxies, in *Extragalactic Radio Sources*, ed. R. D. Ekers, C. Fanti, & L. Padrielli, volume 175 of *IAU Symposium*, pp. 124–126. Dordrecht: Reidel.
- Véron-Cetty, M.-P. & Véron, P. (2006). A catalogue of quasars and active nuclei: 12th edition, *Astronomy and Astrophysics*, **455**, 773–777.
- Vink, J. & Laming, J. M. (2003). On the magnetic fields and particle acceleration in Cassiopeia A, *Astrophysical Journal*, **584**, 758–769.
- Visvanathan, N. & Sandage, A. R. (1977). The color–absolute magnitude relation for E and S0 Galaxies. I - Calibration and tests for universality using Virgo and eight other nearby clusters, *Astrophysical Journal*, **216**, 214–226.
- Völk, H. J., Berezhko, E. G., & Ksenofontov, L. T. (2005). Magnetic field amplification in Tycho and other shell-type supernova remnants, *Astronomy and Astrophysics*, **433**, 229–240.
- Waddington, I., Dunlop, J. S., Peacock, J. A., *et al.* (2001). The LBDS Hercules sample of mJy radio sources at 1.4 GHz - II. Redshift distribution, radio luminosity function, and the high-redshift cut-off, *Monthly Notices of the Royal Astronomical Society*, **328**, 882–896.
- Wall, J. V. (1996). Space distribution of radio source populations, in *Extragalactic Radio Sources, IAU Symposium No. 175*, eds Ekers, R., Fanti, C., & Padrielli, L., pp. 547–552. Dordrecht: Kluwer.
- Wall, J. V. & Peacock, J. A. (1985). Bright extragalactic radio sources at 2.7 GHz. III – The all-sky catalogue, *Monthly Notices of the Royal Astronomical Society*, **216**, 173–192.
- Wambsganss, J. (1998). Gravitational lensing in astronomy, *Living Review in Relativity*, **1**. Online article: accepted 28 August 1998; last amended 31 August 2001 <http://www.livingreviews.org/lrr-1998-12>.

- Wandel, A. & Mushotzky, R. F. (1986). Observational determination of the masses of active galactic nuclei, *Astrophysical Journal*, **306**, L61–L66.
- Wang, W., Harris, M. J., Diehl, R., et al. (2007). SPI observations of the diffuse ^{60}Fe emission in the Galaxy, *Astronomy and Astrophysics*, **469**, 1005–1012.
- Wang, W.-H., Cowie, L. L., & Barger, A. J. (2006). A near-infrared analysis of the submillimeter background and the cosmic star-formation history, *Astrophysical Journal*, **647**, 74–85.
- Warner, B. (1995). *Cataclysmic Variable Stars*. Cambridge: Cambridge University Press.
- Warren, S. J., Hewett, P. C., Irwin, M. J., et al. (1987). First observation of a quasar with a redshift of 4, *Nature*, **325**, 131–133.
- Wasson, J. (1985). *Meteorites: Their Record of Early Solar System History*. San Francisco: W.H. Freeman.
- Watson, M. G. & King, A. R. (1991). Accretion discs in low-mass X-ray binaries., in *IAU Colloquium 129: The 6th Institute d'Astrophysique de Paris (IAP) Meeting: Structure and Emission Properties of Accretion Disks*, Bertout, C., Collin - Souffrin, S. and Lasota, J. P., pp. 19–32, Gif-sur-Yvette: Editions Frontieres.
- Wdowczyk, J. & Wolfendale, A. W. (1984). Galactic cosmic rays above 10^{18} eV, *Journal of Physics G Nuclear Physics*, **10**, 1453–1463.
- Wdowczyk, J. & Wolfendale, A. W. (1989). Highest energy cosmic rays, *Annual Review of Nuclear and Particle Science*, **39**, 43–71.
- Webber, W. R. (1983). Cosmic ray electrons and positrons – A review of current measurements and some implications, in *NATO ASIC Proc. 107: Composition and Origin of Cosmic Rays*, ed. Shapiro, M. M., pp. 83–100.
- Webber, W. R., Kish, J. C., & Schrier, D. A. (1985). Cosmic ray isotope measurements with a new Cerenkov X total energy telescope, *19th International Cosmic Ray Conference*, **2**, 88–91.
- Webber, W. R., Kish, J. C., & Schrier, D. A. (1990a). Formula for calculating partial cross sections for nuclear reactions of nuclei with $E \gtrsim 200$ MeV/nucleon in hydrogen targets, *Physical Review C*, **41**, 566–571.
- Webber, W. R., Kish, J. C., & Schrier, D. A. (1990b). Individual charge changing fragmentation cross sections of relativistic nuclei in hydrogen, helium, and carbon targets, *Physical Review C*, **41**, 533–546.
- Webber, W. R., Kish, J. C., & Schrier, D. A. (1990c). Individual isotopic fragmentation cross sections of relativistic nuclei in hydrogen, helium, and carbon targets, *Physical Review C*, **41**, 547–565.
- Webber, W. R., Kish, J. C., & Schrier, D. A. (1990d). Total charge and mass changing cross sections of relativistic nuclei in hydrogen, helium, and carbon targets, *Physical Review C*, **41**, 520–532.
- Weber, J. (1969). Evidence for discovery of gravitational radiation, *Physical Review Letters*, **22**, 1320–1324.
- Weber, J. (1970). Anisotropy and polarization in the gravitational-radiation experiments, *Physical Review Letters*, **25**, 180–184.
- Webster, A. S. (1970). On the diffusion-loss model of cosmic ray electron propagation in the Galaxy, *Astrophysical Letters*, **5**, 189–192.

- Webster, A. S. (1971). *Cosmic Ray Electrons and Galactic Radio Emission*. PhD Dissertation. Cambridge University.
- Webster, A. S. (1974). The spectrum of the galactic non-thermal background radiational observations at 408, 610 and 1407 MHz, *Monthly Notices of the Royal Astronomical Society*, **166**, 355–372.
- Webster, B. L. & Murdin, P. (1972). Cygnus X-1: A spectroscopic binary with a heavy companion?, *Nature*, **235**, 37–38.
- Wefel, J. P. (1988). An overview of cosmic ray research – Composition, acceleration and propagation, in *Genesis and Propagation of Cosmic Rays*, eds Shapiro, M. M. & Wefel, J. P., pp. 1–9, Dordrecht: Reidel.
- Wefel, J. P. (1991). The composition of the cosmic rays: An update, in *NATO ASIC Proc. 337: Cosmic Rays, Supernovae and the Interstellar Medium*, eds Shapiro, M. M., Silberberg, R., & Wefel, J. P., pp. 29–56, Dordrecht: Kluwer.
- Weinheimer, C. (2001). Neutrino mass from tritium β -decay, in *Dark Matter in Astro- and Particle Physics, Proceedings of the International Conference DARK 2000*, ed. Klapdor-Kleingrothaus, H. V., pp. 513–519. Berlin: Springer-Verlag.
- Weisskopf, V. F. (1981). The formation of Cooper pairs and the nature of superconducting currents, *Contemporary Physics*, **22**, 375–395.
- Wentzel, D. G. (1974). Cosmic-ray propagation in the galaxy – Collective effects, *Annual Review of Astronomy and Astrophysics*, **12**, 71–96.
- Westfold, K. C. (1959). The polarisation of synchrotron radiation, *Astrophysical Journal*, **130**, 241–258.
- Wheeler, J. A. (1968). Our Universe: The known and the unknown, *American Scientist*, **56**, 1–20.
- White, D. A., Fabian, A. C., Allen, S. W., et al. (1994). A ROSAT HRI observation of the ABELL:478 cluster of galaxies, *Monthly Notices of the Royal Astronomical Society*, **269**, 589–606.
- White, S. D. (1989). Observable signatures of young galaxies, in *The Epoch of Galaxy Formation*, eds Frenk, C. S., Ellis, R. S., Shanks, T., et al. pp. 15–30. Dordrecht: Kluwer.
- Whiteoak, J. B. (1974). The observed characteristics of the local magnetic field, in *Galactic Radio Astronomy*, eds Kerr, F. J. & Simonson, S. C., volume 60 of *IAU Symposium*, pp. 137–150, Dordrecht: Reidel.
- Wielebinski, R. (1993). Radio astronomy techniques of observing magnetic fields: The Galaxy, in *The Cosmic Dynamo*, eds Krause, F., Radler, K. H., & Rudiger, G., volume 157 of *IAU Symposium*, pp. 271–277, Dordrecht: Kluwer.
- Wilkes, B. (1999). The spectral energy distributions of active galactic nuclei, in *Quasars and Cosmology*, eds Ferland, G. & Baldwin, J., San Francisco: pp. 15–42. Astronomical Society of the Pacific Conference Series, Vol. 162.
- Wilkes, B. J., Tananbaum, H., Worrall, D. M., et al. (1994). The Einstein database of IPC X-ray observations of optically selected and radio-selected quasars, 1, *Astrophysical Journal Supplement*, **92**, 53–109.
- Wilkinson, P. N., Henstock, D. R., Browne, I. W., et al. (2001). Limits on the cosmological abundance of supermassive compact objects from a search for multiple imaging in compact radio sources, *Physical Review Letters*, **86**, 584–587.

- Wilkinson, P. N., Polatidis, A. G., Readhead, A. C. S., *et al.* (1994). Two-sided ejection in powerful radio sources: The compact symmetric objects, *Astrophysical Journal Letters*, **432**, L87–L90.
- Willingale, R., Bleeker, J. A. M., van der Heyden, K. J., *et al.* (2003). The mass and energy budget of Cassiopeia A, *Astronomy and Astrophysics*, **398**, 1021–1028.
- Willis, A. J., van der Hucht, K. A., Conti, P. S., *et al.* (1986). An atlas of high resolution IUE ultraviolet spectra of 14 Wolf–Rayet stars, *Astronomy and Astrophysics Supplement Series*, **63**, 417–599.
- Wilson, A. S., Arnaud, K. A., Smith, D. A., *et al.* (2002). Cygnus A, in *New Visions of the Universe in the XMM-Newton and Chandra Era*, ed. Jansen, F., European Space Agency ESA SP-488, ArXiv: astro-ph/0202319v1, 1–9.
- Wilson, A. S., Young, A. J., & Shopbell, P. L. (2000). Chandra observations of Cygnus A: Magnetic field strengths in the hot spots of a radio galaxy, *Astrophysical Journal Letters*, **544**, L27–L30.
- Wolfenstein, L. (1978). Neutrino oscillations in matter, *Physical Review D*, **17**, 2369–2374.
- Woltjer, L. (1990). Phenomenology of active galactic nuclei, in *Saas-Fee Advanced Course 20. Active Galactic Nuclei*, eds Courvoisier, T. J.-L. & Mayor, M., pp. 1–55. Berlin: Springer-Verlag.
- Woosley, S. & Janka, T. (2005). The physics of core-collapse supernovae, *Nature Physics*, **1**, 147–154.
- Woosley, S. E. (1986). Nucleosynthesis and stellar evolution, in *Saas-Fee Advanced Course 16: Nucleosynthesis and Chemical Evolution*, ed Hauck, B., Maeder A., and Meynet, G. Geneva: Geneva Observatory Publications. pp. 1–195.
- Woosley, S. E. & Weaver, T. A. (1986). The physics of supernova explosions, *Annual Review of Astronomy and Astrophysics*, **24**, 205–253.
- Wrobel, J. M. & Lind, K. R. (1990). The double-lobed blazar 3C 371, *Astrophysical Journal*, **348**, 135–140.
- Yanasak, N. E., Wiedenbeck, M. E., Mewaldt, R. A., *et al.* (2001). Measurement of the secondary radionuclides ^{10}Be , ^{26}Al , ^{36}Cl , ^{54}Mn , and ^{14}C and implications for the Galactic cosmic-ray age, *Astrophysical Journal*, **563**, 768–792.
- Young, P. J., Westphal, J. A., Kristian, J., *et al.* (1978). Evidence for a supermassive object in the nucleus of the galaxy M87 from SIT and CCD area photometry, *Astrophysical Journal*, **221**, 721–730.
- Yukawa, H. (1935). On the interaction of elementary particles. I, *Proceedings of the Physical-Mathematical Society of Japan*, **17**, 48–57.
- Zamorani, G., Henry, J. P., Maccacaro, T., *et al.* (1981). X-ray studies of quasars with the Einstein Observatory. II, *Astrophysical Journal*, **245**, 357–374.
- Zaritsky, D., Kennicutt, R. C., & Huchra, J. P. (1994). H II regions and the abundance properties of spiral galaxies, *Astrophysical Journal*, **420**, 87–109.
- Zatsepin, G. T. & Kuz'min, V. A. (1966). Upper limit of the spectrum of cosmic rays, *Soviet Journal of Experimental and Theoretical Physics Letters*, **4**, 78–80.
- Zavlin, V. E. (2009). Thermal emission from isolated neutron stars: Theoretical and observational aspects, in *Neutron Stars and Pulsars*, ed. Becker, W., volume 357 of *Astronomy and Space Science Library*. Berlin: Springer-Verlag.

- Zeldovich, Y. & Sunyaev, R. (1969). The interaction of matter and radiation in a hot-model universe, *Astrophysics and Space Science*, **4**, 301–316.
- Zeldovich, Y. B. & Raizer, Y. P. (2002). *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*. Mineola, New York: Dover. Originally published in English by Academic Press, New York in two volumes, 1966, 1967.
- Zombeck, M. V. (2006). *Handbook of Space Astronomy and Astrophysics*, 3rd edition. Cambridge: Cambridge University Press.
- Zwicky, F. & Zwicky, M. A. (1971). *Catalogue of Selected Compact Galaxies and of Post-Eruptive Galaxies*. Guemligen: Zwicky.

Name index

- Abell, George, 99, 101
Abraham, Robert, 733
Abramovitz, Milton, 206, 209
Adams, Fred, 353, 369
Amsler, Claude, 275
Anderson, Carl, 29, 30, 163
Arnaud, Monique, 110
Arnett, David, 386
Arzoumanian, Zaven, 420
Auger, Pierre, 29
- Babbedge, T., 740
Backer, Donald, 417, 418
Bahcall, John, 55, 57, 58
Bahcall, Neta, 105
Balbus, Steven, 455
Band, David, 264
Barger, Amy, 745
Beckwith, Steven, 737
Beccquerel, Henri, 146
Bekesi, George, 193
Bell(-Burnell), Jocelyn, 19, 406
Bennett, Charles, 16
Bethe, Hans, 57, 163, 166, 175
Bignami, Giovanni, 197
Binney, James, 106, 153
Blaauw, Adriaan, 754
Blackett, Patrick, 29
Blain, Andrew, 743
Bland-Hawthorn, Jonathan, 733
Blandford, Roger, 251
Blumenthal, George, 163, 175, 242
Bondi, Hermann, 443
Bothe, Walter, 28
Boyle, Brian, 591, 724
Bracessi, Alessandro, 588
Brandt, W., 725, 726
Bruzual, Gustavo, 729, 730, 743, 744, 747
Butler, Clifford, 30
- Côté, Patrick, 94, 95
Calabretta, Mark, 754
Camenzind, Max, 403, 404
Cameron, Alistair, 60, 498
Cannon, Annie, 37
- Cappelluti, Nico, 729
Carter, Brandon, 434
Caswell, James, 226
Cavaliere, Alfonso, 110
Cesarsky, Catherine, 187, 189
Challinor, Anthony, 115, 259
Chandrasekhar, Subrahmanyam, 302, 429, 434, 455
Charlot, Stéphane, 729, 730, 747
Chwolson, O., 117
Cimatti, Andrea, 736, 748
Clayton, Donald, 386
Clemmow, Phillip, 267
Colless, Matthew, 108, 109
Compton, Arthur, 231
Cordes, James, 420
Cowie, Lennox, 733, 736, 743, 745
Cox, Donald, 357
- Damon, Paul, 297
Davies, Rodney, 376
Davis, Leverett, 373
Davis, Raymond, 32, 54, 55
de Vaucouleurs, Gérard, 77, 78
Dermer, Charles, 505
Deubner, Franz-Ludwig, 51
Diehl, Roland, 287
Dirac, Adrian, 29
Djorgovski, George, 88
Dougherty, John, 267
Draine, Bruce, 351, 372, 373, 375, 376
Dressler, Alan, 84, 88
Dunlop, James, 720, 748
Dunn, Andrew, 108, 109
Dyakov, Sergei, 250
- Eastman, Ronald, 393
Efstatiou, George, 125
Einstein, Albert, 117, 171, 231, 246
Ellis, George, 433
Ellis, Richard, 118, 731, 736
Ellis, Simon, 733
Emslie, Gordon, 322, 324
Eugene, 323
Ewen, Harold, 20
Ezer, D., 60

- Faber, Sandra, 87, 88
 Fabian, Andrew, 113, 114
 Fan, Xiaohui, 725
 Felten, James, 96, 97, 719
 Ferland, Gary, 341, 343
 Fermi, Enrico, 141
 Feynman, Richard, 247, 249, 424
 Field, George, 360, 361
 Fisher, Richard, 87
 Fitzpatrick, Richard, 298, 299
 Forbes, Terry, 324, 327, 329
 Ford, Vincent, 374
 Fort, Bernard, 120
 Frank, Juhani, 451, 455, 456, 459, 460, 461, 462, 463,
 470, 476, 478, 481, 656, 657, 659
 Frolov, Valery, 429
- Garcia-Munoz, M., 514
 Gavazzi, Raphaël, 122
 Gel'fand, Israel, 250
 Genzel, Reinhard, 622
 Ghez, Andrea, 621
 Gilli, R., 602, 728, 729
 Ginzburg, Vitali L., 193, 220
 Glazebrook, Karl, 747
 Gold, Thomas, 406
 Goldreich, Peter, 425, 426
 Goldsmith, Donald, 360, 361
 Gough, Douglas, 51
 Gould, Robert, 154, 163, 175, 242
 Graham Smith, Francis, 406, 416, 418, 420, 422, 424,
 429
 Green, Richard, 589, 591, 724
 Greenstein, Jesse, 373
 Greisen, Eric, 754
 Greisen, Kenneth, 273
 Grindlay, Jonathan, 264
 Gunn, James, 107, 126, 420
- Habing, Harm, 360, 361
 Hall, John, 372
 Hasinger, Günther, 484, 725, 726
 Hauser, Michael, 743
 Hawking, Stephen, 433, 438
 Hawkins, Michael, 591
 Hawley, John, 455
 Haynes, Martha, 86, 87
 Heitler, Wilhelm, 163, 166, 175
 Hess, Victor, 27, 28
 Hewish, Antony, 19, 406
 Hewitt, Jacqueline, 123
 Hildebrand, Roger, 352
 Hillas, Michael, 502, 529, 530, 580
 Hillebrandt, Wolfgang, 380, 383
 Hiltner, William, 372
 Holloway, Nigel, 429
- Hoyle, Fred, 443
 Hubble, Edwin, 77, 78, 757
 Hulse, Russell, 415
 Härm, Richard, 304
- Jackson, John, 267
 Jackson, John D., 137, 141
 Jackson, Robert, 87
 Janka, Hans-Thomas, 384, 386, 388, 420
 Jansky, Karl, 17
 Jokipii, Randy, 186, 187
 Julian, William, 425, 426
- Karttunen, Hannu, 35
 Karzas, William, 169, 229
 Kauffmann, Guinevere, 85
 Kennicutt, Robert, 87, 362
 Kent, Stephen, 107
 Kerr, Roy, 434
 King, Andrew, 451, 456, 459, 460, 461, 462, 463, 470,
 476, 478, 481, 656, 659
 King, Ivan, 105
 Kippenhahn, Rudolf, 395, 398, 399
 Kippenhahn, Rudolph, 35, 48, 60, 61, 63
 Kirshner, Robert, 393
 Kneib, Jean-Pierre, 118
 Koch, H. William, 163, 175
 Kochanek, Christopher, 118
 Kolb, Rocky, 125
 Kolhörster, Werner, 27, 28, 29
 Kompaneets, Aleksander, 250
 Koo, David, 590
 Kormendy, John, 77
 Kramer, Michael, 406
 Krause, Oliver, 382
 Kron, Richard, 590
 Kulsrud, Russell, 189, 298, 308, 310, 324, 329
- Lacy, Mark, 748
 Lagache, Guilaine, 740
 Landau, Lev, 250, 314, 316, 317, 452, 456
 Larmor, Joseph, 425
 Lasenby, Anthony, 115, 259
 Lasota, Jean-Pierre, 481
 Latter, Richard, 169, 229
 Lazarian, Alexander, 375
 Le Roux, Edouard, 206
 Leavitt, Henrietta, 757
 Leger, Alain, 350, 351
 Leibundgut, Bruno, 380
 Lequeux, James, 88
 Liedahl, Duane, 243, 245, 246, 250, 251, 253
 Lifshitz, Evgenii, 314, 316, 317
 Lightman, Alan, 193, 202, 209, 210, 243, 250
 Lilly, Simon, 743, 748
 Lingenfelter, Richard, 290

- Liu, Q.Z., 485
 Lizano, Susana, 369
 Lochner, James, 449
 Longair, Malcolm, 743
 Lorimer, Duncan, 406
 Lotz, Jennifer, 737
 Lyne, Andrew, 406, 416, 418, 420, 422, 424, 429
 Lyttleton, Raymond, 443
- Mészáros, Peter, 705, 709, 713
 MacDonald, Douglas, 438
 Madau, Piero, 743, 745
 Majewski, Steven, 591
 Malin, David, 389
 Manchester, Richard, 410
 Marsh, Thomas, 475
 Mathewson, Donald, 374
 Mayor, Michel, 66
 McCarthy, Patrick, 747
 McCaughean, Mark, 365
 McClintock, Jeffrey, 487, 488, 490
 McCray, Richard, 393
 Mellier, Yannick, 120
 Mellinger, Axel, 7, 8, 755
 Menjo, Hiroaki, 297
 Merritt, David, 107, 108
 Mestel, Leon, 424, 427
 Metcalfe, Nigel, 732
 Mewaldt, Richard, 515
 Meyer, Peter, 497
 Michell, John, 429, 431
 Millikan, Robert, 29
 Mirabel, Felix, 440
 Misner, Charles, 434
 Motz, J., 163, 175
 Murray, Andrew, 753
 Mushotzky, Richard, 448, 449
- Neddermeyer, Seth, 29, 30
 Neininger, N., 370
 Newman, Ted, 434
 Niemeyer, Jens, 380, 383
 Nikolic, Bojan, 349
 Northrop, Theodore, 183
 Novikov, Igor, 429
- Occhialini, Giuseppe, 29
 Oort, Jan, 20
 Orosz, Jerome, 441, 442
 Osmer, Patrick, 591
 Osterbrock, Donald, 341, 343
 Ostriker, Jeremiah, 93, 420
- Pacholczyk, Andrej, 193
 Pacini, Franco, 406, 424
 Pagel, Bernard, 497
- Palla, Francesco, 361
 Panagia, Nino, 392
 Parker, Eugene, 324
 Peacock, John, 267, 720
 Pearce, W. P., 189
 Peebles, James, 93
 Pengelly, R.M., 341
 Penrose, Roger, 433, 437
 Penzias, Arno, 14
 Petschek, Harry, 327
 petschek64, 327
 Pogson, Norman, 760
 Pozdnyakov, L., 243, 253, 255, 256
 Pratt, Gabriel, 110
 Price, Richard, 437
 Priest, Eric, 322, 323, 324, 327, 329
 Pringle, James, 451
 Pryce, M.H.L, 429
 Puget, Jean-Loup, 350, 351, 743
 Purcell, Edward, 20
- Queloz, Didier, 66
- Röntgen, Wilhelm, 146
 Raine, Derek, 451, 456, 459, 460, 461, 462, 463, 470, 476, 478, 656, 659
 Raizer, Yuri, 314
 Ramaty, Reuven, 290
 Rees, Martin, 705
 Reimer, Paula, 296
 Remillard, Ronald, 487, 488, 490
 Rest, Armin, 382
 Reynolds, R.J., 341
 Richards, Gordon, 724
 Richer, John, 360
 Roberts, Morton, 86, 87
 Rochester, George, 30
 Rodrigues, Juan, 440
 Rossi, Bruno, 273
 Rowan-Robinson, Michael, 720, 757, 758
 Rubin, Vera, 92
 Ruderman, Malvin, 428
 Rybicki, George, 193, 202, 209, 210, 243, 250
- Sérsic, José Luis, 82
 Sajina, Anna, 739
 Sandage, Allan, 77, 88
 Scheuer, Peter, 210
 Schmidt, Brian, 393
 Schmidt, Maarten, 362, 589, 591, 720, 724, 725
 Schneider, Peter, 118, 122
 Schroeder, Daniel, 770
 Schwarzschild, Karl, 430
 Schwarzschild, Martin, 94
 Shakura, Nicolai, 454, 459
 Shapiro, Stuart, 399, 400, 403, 422, 433, 434

- Shu, Frank, 352, 353, 360, 369
Silberberg, Rein, 283, 285, 286
Simpson, John, 497, 498, 511
Skobeltsyn, Dmitri, 28, 29
Smart, William, 753
Smith, Barham, 357
Sobol, I., 243
Spinrad, Hyron, 748
Spitzer, Lyman, 298, 302, 304
Springel, Volker, 749
Stahler, Steven, 361
Starrfield, Sumner, 478, 479
Stecker, Floyd, 504, 505, 534, 535
Stegun, Irene, 209
Steidel, Charles, 743
Steigman, Gary, 123
Strong, Andrew, 175, 504, 505
Sunyaev, Rashid, 151, 243, 257, 454, 459
Sutherland, Ralph, 169
Sweet, Peter, 324
Syrovatskii, Sergei I., 193, 220
- Tandberg-Hanssen, Einar, 322
Tayler, Roger, 35, 47, 48, 496
Taylor, Joseph, 415
Tesla, Nikola, 163
Teukolsky, Saul, 399, 400, 403, 422, 433, 434
Thomson, J. J., 155, 156
Thomson, John Joseph (J.J.), 232
Thorne, Kip, 434, 435, 437
Tremaine, Scott, 106, 126, 153
Trodden, Mark, 126
Tsao, Chen-Hsiang, 283, 285, 286
Tully, Brent, 87
Turner, Michael, 125
- Ulrich, Roger, 55
- van de Hulst, Henk, 20
van den Bergh, Sidney, 77, 736
- van den Heuvel, Edward, 473
van der Kris, Michiel, 484
VandenBerg, Donald, 39
Velikhov, E.P., 455
Visvanathan, N., 88
- Waddington, Ian, 724
Wall, Jasper, 721
Wambsganss, Joachim, 118
Wandel, Amri, 448, 449
Wang, Wei-Hao, 745
Warren, Stephen, 590
Wasson, John, 150
Wdowczyk, Jerzy, 503
Weaver, Thomas, 383
Webber, William, 495, 515
Weber, Joseph, 33
Weber, William, 283
Wefel, John, 515
Weigert, Alfred, 35, 48, 61, 395, 398,
 399
Weigert, Alfred., 60
Weisskopf, Victor, 405
Wentzel, Donat, 187
Westfold, Kevin, 206
Wheeler, John, 434
White, Simon, 743
Wilkinson, Peter, 123
Willingale, Richard, 385
Wilson, Robert, 14
Wolfendale, Arnold, 503
Woon, David, 335
Woosley, Stan, 383, 384, 386, 387, 388, 420,
 478
- Yukawa, Hideki, 29, 30
- Zavlin, Vyacheslav, 421
Zeldovich, Yakov, 250, 257, 314
Zwickly, Fritz, 101

Object index

- 14016+2610, 354
1E1207.4–5209 (isolated neutron star), 197
3C 273, 450, 587, 588, 638, 684
3C 58, 379
3C273, 19, 24, 26
47 Tucanae (47 Tuc), 39, 66, 69, 73, 418
4C 21.53, 417, 418
4U 1700–37, 413
51 Peg, 66, 67
- A 0620–00, 487
A0535+262, 471, 472
Abell 478, 112, 113
Abell 1413, 110
Abell 2218, 99, 100, 118
Antennae, 79, 80
- B1937+21 (millisecond pulsar), 417
B2334+61, 421
Becklin–Neugabauer (B–N) object, 364, 365
Betelgeuse, 771
- Cartwheel Galaxy, 80
Cassiopæia A (Cas A), 289, 335, 379, 385, 386, 553, 557, 559
as a type IIb supernova, 385
Centaurus X–3 (Cen X–3), 411
Coma Cluster of galaxies (Abell 1656), 24, 106–109
core radius of, 105
mass of, 107, 759
mass-luminosity ratio of, 108
X-ray image of, 108, 109
Crab Nebula, 23, 24, 26, 193
Crab Nebula (M1, NGC1952), 379
continuous injection of energy into, 410
energy requirements of, 409
pulsar in, 406, 409, 417, 418, 422, 423, 424, 425, 428
Cyg X–1, 24
Cyg X–3, 24
Cygnus A, 24, 261, 489, 774
Cygnus X–1, 413
hard X-ray spectrum of, 256
Cygnus X–1 (Cyg X–1), 438, 442, 448, 449, 487
Cygnus X–2 (Cyg X–2), 483, 484
- η -Carinae, 71, 72
mass loss rate of, 71
EXO 033319–2554.2, 197
- G1.9+0.3 (supernova remnant), 379
Galactic Centre, 7, 621
supermassive black hole in, 11
‘Geminga’, 421
Gliese 229B
brown dwarf companion of Gliese 229, 66
spectrum similar to that of Jupiter, 66
GRB 030329, 706
GRO 1744–28, 483
GRO J1655–40, 487
GRO J1744–28, 472
GRS 1915+105, 440
mass function for, 440
GRS1915+105, 441
Guitar Nebula, 420
GX5–1, 483, 484
- HD 209458, 67
HD 93131 (Wolf–Rayet star), 71
Her X–1 (pulsating X-ray binary), 197, 405
Hercules X–1 (Her X–1), 412
discovery records of, 412
modulation of light curve, 483
precession of rotation axis of neutron star, 483
unpinning of crust and magnetic field from neutron superfluid, 483
HH211, 365, 366
HH30, 365, 366
HH34, 365, 366
Homunculus Nebula (η -Carinae), 72
HT Cas, 478
total eclipse observed in, 478
- Kepler’s supernova, 379
Kleinmann–Low Nebula, 364, 365
- Large Magellanic Cloud (LMC), 8, 11, 13, 26, 32, 379, 388, 389
LMC X–1, 442
LMC X–3, 442, 487, 489

- M1, *see* Crab Nebula (M1, NGC 1952)
 M3, 39
 M31, Andromeda Nebula, 92, 97
 M33, 439, 440
 M33 X–7 (X-ray binary), 440
 M49 (NGC 4472), 94, 95
 M51, 370, 371
 M51 (NGC 5194), 79
 M67, 39
 M81, 755
 M87 (NGC 4486), 23, 78, 79, 94
 non-thermal optical jet of, 78
 Magellanic Clouds, 339
 MCG–6–30–15, 448, 449, 611, 626, 627, 628
 Milky Way, 7, 8
 MXB 1730–335 (rapid burster), 483

 NGC 1300, 79
 NGC 2362, 39
 NGC 2787, 79
 NGC 4258, 621
 NGC 4839, 108
 NGC 5195, 79
 NGC 5506, 448, 449
 NGC 7023, 351
 North Pole star, 7

 Ophiuchus molecular cloud, 354
 Orion giant molecular clouds, 363, 364
 Orion Molecular Cloud, 13
 Orion Molecular Cloud A, 363, 364
 Orion Nebula, 13, 17, 363, 364, 365
 luminous far-infrared sources in, 364
 Orion star cluster, 66, 69, 70
 Orion, constellation of, 363, 364

 Perseus Cluster of galaxies, 24, 114
 Ly α and Ly β emission lines of highly ionised iron,
 Fe $^{+25}$ in, 169
 X-ray bremsstrahlung of hot intracluster gas in, 169
 X-ray spectrum of, 169
 Plough or Great Bear, 7
 PSR 1919+21 (CP 1919), 406, 407
 discovery records of, 407
 PSR B0540–69, 409
 PSR B0656+14, 421
 PSR B1055–52, 421

 PSR B1509–58, 409
 PSR B1913+16 (binary pulsar), 415, 416
 change of period due to emission of gravitational
 waves, 416
 tests of general relativity and, 415
 PSR J0538+2817, 421
 PSR J0737–3039, 416
 PSR J0737–3039, 416, 417
 change of period due to emission of gravitational
 waves, 416
 PSR J1119–6127, 409

 Sérsic 159–03, 113
 Sagittarius A, 753, 754
 Sagittarius A* (Sgr A*), 621, 622, 623
 Sanduleak –69 202, 32, 389
 Sco X–1, 24
 Scorpius X–1 (Sco X–1), 483
 Small Magellanic Cloud (SMC), 8, 11, 13
 SN 1006, 297, 379
 SN 1054, *see* Crab Nebula (M1, NGC 1952)
 SN 1181, 379
 SN 1572, *see* Tycho's supernova
 SN 1604, *see* Kepler's supernova
 SN 1987A, 32, 73, 279, 288, 289, 379, 388–393
 SU Aur, 354

 Taurus molecular cloud, 354
 Trapezium stars in Orion Nebula, 364, 365
 Tycho's supernova, 379, 382
 remnant of, 382
 typing of by light echos, 382

 U Gem, 475, 476

 Vega (α -Lyrae), 760, 762
 Vela supernova remnant, 26
 pulsar in, 406, 421, 422, 423, 424
 Vela X–1, 413
 Virgo Cluster of galaxies, 24, 99, 100
 VSSG 23, 354

 X1822–371, 480
 XBT0748–676, 480
 XTE J1650–500, 492

 Z Cha, 475

Index

- Abell Catalogues of rich clusters of galaxies
selection criteria of clusters in, 100
space density of clusters in, 100
- Abell Clusters of galaxies, 99–102
spatial correlations with clusters and galaxies, 101
- aberration formula, relativistic, 238
- absolute magnitude, 763
- absorption coefficient χ_v , 170
for thermal bremsstrahlung, 170
corrected for stimulated emission, 172
uncorrected for stimulated emission, 172
- absorption edges at X-ray energies, 228–230
K-edges, 229
- accelerated charged particles
polar diagram of radiation of, 156
- acceleration four-vector \mathbf{A} , 161
- acceleration of an electron in the electrostatic field of
a proton or nucleus, 163
- acceleration of high energy particles, 561–582
beyond the standard model, 574–580
diffusive shock acceleration in strong shock waves,
568–574
energy spectrum of cosmic rays at and above the
'knee', 579–580
Fermi acceleration – original version, 564–568
general principles of acceleration, 561–562
highest energy cosmic rays, 580–582
in solar flares, 562–563
magnetic fields in supernova shock fronts, 574–576
nonlinear diffuse shock acceleration, 576–579
- accretion
as an energy source for X-ray sources, 24
in X-ray binary systems, 412
maximum energy release for Schwarzschild black
hole, 413
- accretion columns in magnetic cataclysmic variables,
477–478
emission of far ultraviolet and soft X-ray emission
from, 477
shock fronts in, 477
temperature of shocked gas in, 477
- accretion disc, thick, 437
- accretion discs, *see* thin accretion discs
boundary layer
emission from, in cataclysmic variables, 477
luminosity of, 457
- boundary layer at inner edge of, 456
energy flow in, 457
formation of, 444
luminosities of, 457
outward transfer of angular momentum, 444, 445
viscous dissipation of energy in, 445
- accretion luminosity, 444
- Accretion Power in Astrophysics* (Frank, King and
Raine), 451
- accretion power in astrophysics, 443–492
accreting binary systems, 473–486
accretion in binary systems, 464–473
black holes in X-ray binaries, 486–491
Eddington limiting luminosity, 445–447
efficiency of the accretion process, 443–445
efficiency of the accretion process, ξ , 444
for neutron stars, 444
for white dwarfs, 444
onto black holes, 444
- general considerations, 443–450
thick discs and advective flows, 461–463
thin accretion discs, 451–461
- accretion radius, *see* capture, or accretion, radius
- action integral in Lagrangian mechanics, 182, 183
- active galactic nuclei
'blue-bump' component in spectra of, 450
ratio of black hole to spheroid masses, 748, 750
Type 1, 602, 728
unobscured, 602, 728
- active galaxies, 585–609
- adaptive optics, 7, 770
- adiabatic changes in classical mechanics, 182
- adiabatic expansion
of a magnetic field, 307
- adiabatic invariance, 184
principle of, 182
- adiabatic loss problem and the acceleration of high
energy particles, 556–560
- adiabatic motion of charged particle, 134
- Adiabatic Motion of Charged Particles* (Northrop),
183
- Advanced Camera for Surveys (ACS), 122, 732
- advection, 656, 657, 659
- advective transport of mass and energy, 462–463
supermassive black holes and, 463
through black hole horizon, 463

- Airy diffraction pattern, 768
 Airy disc, 768
 Aitoff projection, 755
 Alfvén and hydromagnetic waves in interstellar medium, 187–189
 damping by neutral particles, 189
 damping rate of, 189
 energy density of, 188
 growth rate of, 188
 momentum density of, 188
 Alfvén radius, 469, 470
 for white dwarfs, 470
 Alfvén speed, 188, 189, 321, 323, 325, 326, 327
 alignment effect in radio galaxies, 654, 655
 alignment of interstellar grains, 372–374
 magnetic field parallel to optical polarisation, 374
 physical mechanisms for, 373
 Barnett effect, 373
 paramagnetic dissipation, 373
 Rowland effect, 373
 suprathermal processes, 373
 α -discs, 455, 459, 461
 AM Herculis binaries, 197
 ambipolar diffusion, 311
 Ampère's theorem, 324
 Anglo Australian Telescope 2dF galaxy survey, 97, 101
 Anglo-Australian telescope, 589
 Anglo-Australian Telescope 2dF survey (AAT 2dF), 4, 77, 80, 95, 96
 angular cyclotron frequency, 179
 angular diameter distance, 118
 in general relativity, 430, 431, 432
 angular frequency ω_0 of electron in atom, 135
 angular gyrofrequency, 179, 183, 189
 non-relativistic, 201
 relativistic, 201, 211
 angular momentum transport by viscosity, 455, 456
 angular plasma frequency, 446
 angular resolving power, 764
 anomalous resistivity of a plasma, 328
 anomalous X-ray pulsars, 418–419
 location on $P - \dot{P}$ diagram, 419
 antenna temperature, 779
 minimum detectable, 779
 aperture grating, 768
 aperture synthesis, 771–774
 Earth-rotation, 774
 principles of, 21
 apodisation, 768
 Apollo 12 and the Surveyor satellite, 147
 Apollo 14, 149
 Apollo 17, 147
 apparent magnitude
 bolometric, 762
 definition of, 760
 Archimedean spiral, 313
 Ariel-V satellite, 725
 ASCA Large Sky Survey, 725
 associated Legendre functions, 51
 asteroids
 asteroid belt and, 148
 parent bodies of meteorites, 148
 astronomical seeing, 770, 771
 astronomical unit (AU), 756
 astronomical units, 757
 astroparticle physics, 127
Astrophysical Flows (Pringle and King), 451
Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Osterbrock and Ferland), 341
 asymptotic giant branch stars, 288
 Atacama Large Millimetre Array (ALMA), 14
 atmosphere as convertor for cosmic rays, 502
 atmospheric turbulence and astronomical seeing, 769, 771
 atomic binding energy, 135
 Auger air-shower array, 31
 Auger ultra-high energy cosmic ray observatory, 275
 aurorae, 320
 green and red lines of oxygen in, 320
 auroral zone, 320
 Australia Telescope Compact Array, 393
 Australia Telescope National Facility (ATNF), 21
 Avogadro's number, 175, 271
 Baade–Wesselink method, 392, 758
 background intensities in ground-based observations, 761, 775
 background radiation
 submillimetre and far-infrared
 contribution of active galactic nuclei, 745
 main contributors to, 745, 746
 ultraviolet
 decrease at large redshifts, 744, 745
 decrease in intensity from $z = 1$ to present epoch, 745
 Baksan Neutrino Observatory in the northern Caucasus mountains, 57
 Balmer absorption line index $H\delta_A$ as age indicator, 85, 86
 Balmer break, or discontinuity, $D_n(4000)$ as age indicator, 85, 86
 Balmer continuum absorption, 764
 Balmer decrement, 341
 Balmer series of hydrogen, 450, 587, 594, 642, 646, 647
 BATSE instrument of the Compton Gamma-ray Observatory, 472
 Beckwith, Steven, 744
 BeppoSAX gamma-ray telescope, 289
 Bernoulli's equation, 315
 Bessel functions $J_0(z)$, $J_1(z)$, 767, 768
 betatron, 193

- Bethe–Bloch formula, 140–141, 163, 167
 bias parameter, 122
 BIMA Millimetre Array, 115, 116
 binary pulsars, 415–417
 binary star systems, 464–467
 - close, 464
 - contact binaries, 464
 - common envelope of, 464
 - mass–luminosity relation for, 465
 - equipotential surfaces in the rotating system, 464, 465
 - evolution of stars in, 465, 466
 - feeding the accretion disc, 467
 - Roche lobe overflow, 467, 468
 - stellar mass loss and, 468
 - stellar mass loss and winds, 467
 - mass transfer in, 465, 466
 - non-contact close binaries, 465
 - periods of, 464
 - role of magnetic fields, 469–473
 - accretion columns and, 469
 - magnetic pressure of, 469
 - mass flow onto magnetic poles, 469, 470
 - ram pressure and, 469
 - spectroscopic binaries, 464
 - statistics of, 464
 - symbiotic stars and, 467, 474
 - visual binaries, 464
 - binary X-ray sources, 438
 - bipolar outflows, 353, 355, 364–367
 - magnetic fields in, 367
 - model for, 367
 - molecular beams in, 364
 - origin of, 369
 - magnetic fields and, 369
 - polarisation observations of, 367
 - similarity to extragalactic radio sources, 367, 369
 - BL-Lacertae objects (BL-Lac objects), 19
 - Black Hole Physics: Basic Concepts and New Developments* (Frolov and Novikov), 429
 - Black hole X-ray binaries, 488
 - black hole X-ray binaries
 - quasi-periodic oscillations (QPOs), 488, 490, 491
 - increase in frequency with luminosity, 490
 - black holes, 76, 123, 378, 429–442
 - accretion luminosity of, 458
 - angular momentum on last stable orbit, 458
 - condition of matter to fall into, 458
 - dragging of inertial frames, 435
 - electrodynamics of, 437, 438
 - ergospheres of rotating, 437
 - evaporation of, 438
 - extraction of rotational energy from rotating, 437
 - ‘flickering’ of the X-ray intensity and, 448, 449
 - general case in general relativity, 434–438
 - Hawking radiation from very low mass, 438
 - horizon of rotating black hole, 435
 - in binary X-ray sources, 25
 - in X-ray binaries, 486–491
 - different active states of, 487, 488
 - disc fraction f , 487
 - hard state, 487, 489–490
 - hard state and presence of radio jets, 489
 - hot corona in, 490, 491
 - iron fluorescence lines, 487, 491, 492
 - luminosity of accretion disc, 488, 489
 - power density spectrum of the variability, 488
 - quasi-periodic oscillations (QPOs), 490
 - steep power-law, 487
 - steep power-law state, 490
 - temperature distribution in accretion disc about, 488
 - thermal state, 487, 488–489
 - last stable orbit, 448, 449, 458, 463, 490, 491, 611, 629, 630, 631, 633, 656
 - last stable orbit about, 445
 - evidence for, 489
 - light-travel time across the last stable orbit, 448
 - magnetic fields threading, 437
 - mass estimates for, in binary X-ray systems, 438, 439
 - masses from time variability, 448, 450
 - maximally rotating, 436
 - maximum angular momenta of, 435
 - maximum energy release of Schwarzschild, 436, 445
 - observational evidence for, 438–442
 - primordial, 438
 - resistivity of, 437, 438
 - Schwarzschild, 436
 - sketches of, in binary X-ray systems, 441, 442
 - specific angular momentum of, 458
 - static radius about rotating, 435
 - supermassive, 123, 715
 - epoch of maximum quasar activity, 723
 - non-thermal radiation processes and, 715
 - temperature of, 438
 - temperature of gas at last stable orbit, 449
 - Black Holes – The Membrane Paradigm*, 437
 - black holes in the nuclei of galaxies, 610–636
 - black holes in X-ray binaries and active galactic nuclei, 448–450
 - black-body radiation
 - intensity of, 218, 778
 - Rayleigh–Jeans limit of, 218
 - spectrum of, 5
 - blue sequence or blue cloud, 80
 - avoid regions of rich clustering, 85
 - properties of galaxies of, 81
 - Bohr model of the atom, 135
 - Bohr radius of the hydrogen atom, 174
 - bolometric absolute magnitude, 763
 - of Sun, 763
 - bolometric flux density, 763

- bolometric luminosity of Sun, 763
 Boltzmann distribution, 171, 299, 333
 Boltzmann equation, 250, 259
 evolution of photon occupation number and, 250
 Boltzmann relation, 247
 Bose–Einstein distribution
 as a solution of Kompaneets equation, 253
 Bose–Einstein spectrum, 257
 Bose–Einstein statistics, 404
 bound–free absorption, *see* photoelectric absorption
 bound–free absorption in stars, 47
 bow shock, 320
 density enhancement across, 320
 Boyer–Lindquist coordinates, 434
 braking index, 407, 408, 409
 energy loss from pulsars and, 407
 for magnetic dipole radiation, 408
 magnetic breaking and, 407
 ‘braking radiation’, *see* bremsstrahlung
 bremsstrahlung, 99, 109, 115, 133, 154, 163–177,
 301, 759
 collision parameters b_{\max} and b_{\min} for, 163, 166
 cooling rate, 111
 emissivity of, 115
 low frequency spectrum of, 165, 166
 non-relativistic energy loss rate, 166–167
 non-thermal, 290
 relativistic, 173–177
 average energy of photons emitted in, 177
 catastrophic losses in the atmosphere, 176
 collision parameters b_{\max} and b_{\min} for, 173–175
 correction for electron–electron interactions,
 175
 critical energy, 175
 in terms of the photon number flux density, 175,
 176
 radiation length X_0 for, 175
 the low energy γ -ray emission of the interstellar
 medium and, 177
 total energy loss rate, 174
 total stopping power for different materials, 175
 total stopping power in different materials, 175
 spectral emissivity of, 109
 spectrum of, 110, 115
 thermal, 167–173
 absorption, 170–173
 Gaunt factors for, 168
 Gaunt factors for radio wavelengths, 168
 Gaunt factors for X-ray energies, 168
 mass of gas and dark matter in clusters of
 galaxies and, 169
 spectral emissivity of, 167–169
 spectral emissivity of, in terms of photon number
 per unit energy interval, 169
 total energy loss rate of, 168
 bremsstrahlung absorption in stars, 47
 bremsstrahlung, X-ray, 394
 brightness temperature T_b , 217, 219, 778, 779
 Rayleigh–Jeans approximation for, 778
 brown dwarfs, 66, 123, 124
 discovered in 2MASS survey, 66
 discovered in Pleiades, Orion and ρ Ophiuchus
 clusters, 66
 discovered in Sloan Digital Sky Survey, 66
 BUGS experiment of Ariel-VI mission, 499, 500

 CalTech Submillimetre Observatory, 259
 CalTech Submillimetre Observatory (CSO), 14, 115
 Canada–France Redshift Survey, 743
 canonical coordinates, 182
 canonical momentum, 182
 of particle in a magnetic field, 182
 capture, or accretion, radius, 485, 486
 carbon burning, 64
 carbon monoxide (CO) map of Galaxy, 17
 carbon recombination lines in gaseous nebulae
 very high order transitions, 342
 carbon–nitrogen–oxygen (CNO) cycle, 44, 45, 48, 50,
 62, 72, 478, 479
 Cassegrain telescope, 770, 771
 cataclysmic variables, 383, 443, 447, 467, 473–478,
 479
 accretion columns in magnetic, 477–478
 AM Herculis stars, 474
 classical novae, 473, 478, 479
 event rate in our Galaxy and M31, 478
 thermonuclear runaway and, 473, 478, 479
 dwarf novae, 473, 474, 475, 476, 478
 eclipse mapping of, 475
 temperature mapping, 475
 intermediate polars, 474
 magnetic, 474
 novae-like stars, 474
 polars, 474
 recurrent novae, 474
 strong emitters in the 1–10 keV X-ray wavebands,
 477
 strong emitters in the EUV and soft X-ray
 wavebands, 477
 strong winds from, 478
 symbiotic stars, 474
 celestial equator, 753, 755
 celestial hemisphere
 north, 755
 south, 755
 celestial sphere, 755
 central limit theorem, 719, 774
 Gaussian statistics and, 775
 Cepheid variables, 757
 luminosity–period relation for, 757
 Cerro Tololo InterAmerican Observatory, 11
 ChaMP study, 725
 Chandra Deep Fields, 725, 727

- Chandra X-ray Observatory, 25, 114, 379, 385, 393, 546, 663, 669, 714, 725, 727
 Chandrasekhar limit, 383, 397–406
 Chandrasekhar mass, 382, 383, 384, 400, 466
 for neutron stars, 399, 400
 for white dwarfs, 399
 charge-coupled devices (CCD), 7, 776
 charged particles in magnetic fields
 dynamics in time-varying field, 180–184
 adiabatic invariant approach to, 182–184
 physical approach to, 181–182
 dynamics of, 178–189
 in static uniform field, 178–180, 193
 scattering by Alfvén and hydromagnetic waves, 187–189
 scattering by irregularities in the field, 184–187
 spiral motion in uniform field, 179
 chemical potential μ , 246, 247, 257
 Cherenkov radiation, 264–270
 ‘shock wave’ interpretation of, 266
 Cherenkov cone, 269
 condition for, 266
 energy loss rate per unit bandwidth, 270
 circumstellar disc, 353
Classical Electrodynamics (Jackson), 129
 classical electron radius, 233
 classical novae, *see* cataclysmic variables, classical novae
 clusters of galaxies, 99–127
 Bautz–Morgan classification, 102
 central concentration of galaxies and central profile, 102
 central mass density of, 105
 core radius, 102, 104
 crossing time, 107
 dark matter in, 123–127
 galaxy content, 102
 isothermal gas spheres and, 102–106
 mass segregation, 102
 morphologies of, 99–102
 most massive galaxies in, 153
 regular, intermediate and irregular, 102
 symmetry, 102
 X-ray emission of, 25
 clusters of stars, 39
 coincidence counting, 28
 collision of high energy particle with stationary
 electron
 duration of, 134
 limits to collision parameters, 133–136
 relativistic case, 139–140
 non-relativistic treatment, 131–133
 maximum energy loss, 132
 relativistic treatment, 139–140
 collision parameter, 132
 collision parameters
 maximum and minimum for electrostatic scattering, 302
 collision time of particles in a plasma, 301
 collisionless plasma, 305
 colour index
 definition of, 763
 colour–colour diagram for stars, 763, 764
 colour–magnitude diagram for stars, *see* Hertzsprung–Russell diagram
 colours
 definition of, 763
 comoving radial distance coordinate, 718
 comoving volume
 variation with redshift, 716
 compact extragalactic sources and superluminal motions, 681–713
 compact H II region
 spectrum at radio wavelengths of, 172, 173
 compact radio sources
 relativistic bulk motion in, 264
 variability of, 263
 Compton effect, 236
 Compton Gamma-ray Observatory (CGRO), 288, 289, 292
 EGRET instrument of, 26, 504
 Compton optical depth, 245, 246, 251, 255, 257, 258
 Compton scattering, 114, 235–237
 average energy change of photon in, 244
 average energy gain of photons by, 244
 derivation of formulae for, 235–237
 exchange of energy between electrons and radiation field, 236
 inverse, 715
 Klein–Nishina cross-section, 237
 probability distribution for a single scattering, 260
 recoil effect, 236
 Compton scattering optical depth, 115
 Comptonisation, 243–257, 490
 basic physics of, 243–246
 Bose–Einstein distribution
 formation of, 246, 247
 with finite chemical potential μ , 246
 Compton optical depth, 245, 246
 condition for significant distortions of photon spectrum, 245
 examples of astrophysical applications, 243
 interchange of energy between matter and radiation and, 243–246
 differential equation for, 245
 Kompaneets equation, 250–257
 number of scatterings to approach saturation, 245, 246
 occupation number, 246–250
 recoil effect in, 244

- computer simulations
 hydrodynamical simulations of galaxy collisions, 750
 origin and evolution of cosmic structures and, 749
- concentration index $C = (R90/R50)$, 85, 86
- conductivity of a plasma, 133, 425
- cone diagram of the distribution of galaxies, 101
- conservation of angular momentum, 431, 444
- conservation of energy for Newtonian gravity, 431, 432, 444
- conservation of energy in general relativity, 432, 444
- conservation of mass, 47
- convective transport of energy in stars, 45, 48
 condition for, 45, 60
 in main sequence stars, 61
 in pre-main sequence stars, 61
 in red giant stars, 61
 of different masses on the main sequence, 61
- convective transport of mass and energy, 463
- cooling flows in clusters of galaxies, 110–114
- Cooper pairs, 404
- coordinate systems and projections used in astronomy, 7
- coordinate systems in astronomy, 753–755
- coordinate time in general relativity, 431
- coordinate-space diagram for diffusion-loss equation, 191
- Copernicus satellite, 337, 338
- coronal loops, 308, 309, 312, 322
- coronal mass ejection events, 314
- coronene, 351
- COS-B satellite, 26
- Cosmic Background Explorer, 376
- Cosmic Background Explorer (COBE), 15, 16
 DIRBE instrument of, 11, 12, 13
- The Cosmic Century: A History of Astrophysics and Cosmology* (Longair), xiii, xiv, 3, 5, 48, 193
- Cosmic Microwave Background Radiation, 14–16, 17, 99, 114, 115, 376, 759
 cosmological fluctuations in, 16
 dipole component of, 15, 16
 distortions from a perfect black-body spectrum, 257
 due to energy injection after the recombination era, 257
 due to energy injection prior to the recombination era, 257
 limits to, 259
- motion of the Solar System relative to, 15, 16
 radiation temperature of, 15
- Sunyaev–Zeldovich distortions of, 114, 115, 259
- cosmic ray astrophysics, 27–32
 extensive air-showers, 29
 from space and from the ground, 30–32
 history of, 27–30
 ionisation of the atmosphere with increasing altitude, 28, 29
- cosmic ray clocks, 501
- cosmic ray electrons
 spectrum of, 212, 225, 226
- cosmic ray protons and nuclei
 solar modulation of, 185, 187
- cosmic rays, 493–535
 abundances of the elements in the cosmic rays, 496–502
 air-shower technique at energies $E \gtrsim 10^{15}$ eV, 493, 494
 antiparticles in the flux of, 496
 chemical abundances of, 497–500
 features of, 497–499
 chemical composition of, 30
 confinement time in the Galaxy and cosmic ray clocks, 515–517
 confinement volume for, 517–520
 differential energy spectra for different species, 493–495
 differential energy spectra for electrons, 495–496
 differential energy spectrum of
 ‘ankle’ in, 493, 494
 ‘knee’ in, 493, 494
 overall, 493, 494
 energy density of, 358
 energy spectra of cosmic ray protons and nuclei, 493–496
 solar modulation and, 493
 energy spectrum of, 30, 31
 Galactic halo and, 520–522
 Greisen–Kuzmin–Zatsepin (GKZ) cut-off, 531–535
 highest energy cosmic rays and extensive air-showers, 522–524
 isotopic abundance anomalies in
 origin of, 515
 isotopic abundances of, 500–502
 differences from cosmic abundances, 501
 isotropy and energy density of, 502–503
 isotropy of
 as a function of energy, 502
 net stream of cosmic rays and, 503
 underground muons and, 502
 isotropy of ultra-high energy cosmic rays, 529–531
 local energy density of, 503
 compared with other local energy densities, 503
 observation of the highest energy, 524–529
 origin of the light elements in, 507–515
 abundance differences compared with cosmic abundances, 507
 overall statistics for, 493
 particle detectors in space observatories, 493
 Solar System abundances of the elements, 496–497, 499
 source abundances compared with the local Galactic abundances, 511, 512

- correlation with first ionisation potential, 513
 first ionisation potential and, 511
 spectrum of, 212
 transfer equation for light nuclei, 507–512
 ultra-high energy, 31
 anisotropies in distribution of, 31
 cut-off at very high energies, 32
 variations in the chemical composition of cosmic rays with energy, 513–515
 boron-to-carbon ratio, 513, 514
 chromium-to-iron ratio, 513, 514
- cosmic rays and the discovery of new particles, 29–30
 charged and neutral kaons, 30
 mesotron, 29
 muon, 30
 pion, 30
 positron, 29
 strange particles, 30
- cosmic rays in the atmosphere, 292–297
 electromagnetic cascades, 292–294
 formation rate and half-life of ^{14}C , 295
 formation rate and half-life of ^3H , 295
 nucleonic cascades, 292–294
 path-length for nuclear interaction, 293
 pion production, 292–293
- radioactive nuclei produced by cosmic rays in the atmosphere, 294–297
 formation of carbon-14 ^{14}C , 294
 formation of tritium ^3H , 294
 residence time in atmosphere, 295
- secondary fluxes of relativistic electrons, 294
 vertical fluxes of, 295
 vertical fluxes of at different heights, 293
- cosmic rays, solar
 underabundances correlated with first ionisation potential, 511
- cosmic star formation rate
 maximum at redshifts $z \sim 1–2$, 748
 problem of dust extinction, 745
 submillimetre determinations of
 as a function of redshift, 745
- cosmic web, 101
- cosmological aspects of high energy astrophysics, 714–752
- cosmological distance ladder, 758, 759
- Coulomb's law of electrostatics, 157
 relativistic transformation of, 136–139
- counts of active galaxies
 evolution in infrared waveband, 740
- counts of extragalactic radio sources, 720
 excess of faint radio sources, 720
 evidence for strong cosmological evolution, 720
- counts of galaxies, 731–737
 excess of faint blue galaxies, 732, 733
 and starburst galaxies, 733, 737
 nature of, 733–737
- fluctuations in, due to large-scale clustering, 732, 733, 736
- for irregular/peculiar/merger systems, 733, 736
- for spheroidal and spiral galaxies, 733, 736
- in infrared K -band, 733
- in infrared K -waveband, 736
- in U, B, R, I and K wavebands, 736
- problems of determining, 731
- counts of galaxies and active galaxies, predicted, 715
 at submillimetre wavelengths, 740–742
 normalised differential counts, 741, 742
- Euclidean, 715–716
 differential, 716
 integral, 715
- for standard world models, 716–731
 for sources with power-law spectra, 718
 infrared counts for galaxies, 730
 normalised, differential, 718
 optical counts for galaxies, 729–731
 slopes of integral and differential, 718
- counts of infrared and submillimetre sources
 convergence of, at mid-infrared wavelengths, 740
- IRAS galaxies
 excess of faint sources, 737
- Spitzer Survey
 excess of faint sources, 737
- submillimetre wavelengths, 742
 excess of faint sources, 743
- counts of infrared sources, 737
- counts of mid and far-infrared sources, 740
- counts of radio quiet quasars, 591, 592, 724
- counts of submillimetre sources, 740–743
- counts of X-ray sources, 725–729
 evidence for evolution of the source populations, 727
- hard X-ray energies, 2–10 keV, 602, 727, 729
 history of, 725
 problems of interpretation, 727
 soft X-ray energies, 0.5–2 keV, 727
 and the integrated X-ray background emission, 728
- soft X-ray energies, 0.5–2 keV, 727
- Crab Nebula (M1, NGC1952), 409, 549, 586, 644, 704
 pulsar in, 379
- critical density for degenerate gas, 396
- critical angular frequency for synchrotron radiation, 207, 208
- critical brightness temperature, 263, 264
- critical cosmological density, 127
 in neutrinos with finite mass, 125
- critical density
 for common ions, 342
- critical density for stabilisation in degenerate neutron gas, 402
- critical density for star formation, 750
- critical Fermi momentum p_F , 402

- critical frequency for synchrotron radiation, 201, 211, 212, 218, 223
 crossing time, 107
 crossing time τ_{cr} of star or galaxy in a cluster, 153
 Cryogenic Dark Matter Search (CDMS II), 127
 current sheets, 323
 curvature of space-time in general relativity, 430
 curvature radiation, 201, 428
 cusp catastrophe, 120
 cyclotron absorption features, 405
 cyclotron absorption features in accreting X-ray sources, 197, 198
 magentic field estimates from, 197
 cyclotron absorption features in isolated neutron star, 197
 magentic field estimates from, 198
 cyclotron radiation, 195–198
 broadening of spectral lines of, 196
 circular polarisation of, 197
 geometry of magnetic field configuration and, 197
 magentic field estimates from, 197
 harmonics of gyrofrequency, 196, 197
 mildly relativistic, 196
 cyclotron radius of charged particle in magnetic field, 179

 dark energy, 381
 dark matter, 123–127
 astrophysical and experimental limits, 126–127
 axions and, 125
 baryonic, 123–125
 black holes and, 123
 limits to mass density from gravitational lensing, 123
 brown dwarfs and baryonic, 123
 clusters of galaxies and, 99, 108, 109
 forms of, 123–127
 in clusters of galaxies, 108, 110
 in early type galaxies, 122
 in galaxies, 77, 93
 and stability of disc galaxies, 93
 MACHOs and, 125
 gravitational microlensing and, 124
 masses of dark matter particles, 126, 127
 neutrinos with finite rest mass and, 125
 non-baryonic, 125–126
 searches for dark matter particles, 127
 structure of, 122
 WIMPs and, 125
 dark matter haloes of galaxies and clusters, 122
 Davis–Greenstein alignment mechanism, 374
 de Broglie wavelength, 279
 de Vaucouleurs $r^{1/4}$ law for surface brightness, 82, 105
 dead stars, 378–442

 Debye length, 298–300, 302, 446
 as shielding distance in plasma, 299, 300
 decay of binary orbits and gravitational waves, 32
 declination (Dec or δ), 753
 deflagrations, 383
 degeneracy pressure, 394
 condition for use of relativistic equation of state, 396
 conditions under which important, 395, 396
 degenerate gas
 non-relativistic equation of state for, 395, 396
 relativistic equation of state for, 397
 of electrons, 397
 of neutrons, 397
 degenerate stars
 internal structure of, 394–397
 dendrochronology, *see* tree-ring-dating
 density parameter
 for stars in the Universe, 98
 in baryons, 98
 density wave theory of spiral structure, 357, 462
 forcing mechanisms and, 357
 depolarisation of polarised radio signals, *see* Faraday depolarisation
 detection of ultrahigh energy γ -rays, 265
 detonations, 383
 deuterium formation in stars, 44, 54
 neutrinos from, 44
 differential rotation of Galactic disc, 353
 diffraction optics, 764
 diffraction-limited telescopes, 764–770
 diffusion coefficient D , scalar, 189
 diffusion coefficients for high energy particles in
 fluctuations in magnetic field, 186
 physical model for, 186–187
 diffusion loss equation for high energy electrons, 540–544
 distortion of the injection energy spectrum, 540–543
 diffusion of charged particles, 301–303
 mean free path and, 301
 diffusion time for magnetic field in a plasma, 310, 311, 324
 diffusion-loss equation, 253
 escape time, 253
 source term, 253
 statistical acceleration and, 252
 diffusion-loss equation for high energy particles, 189–192
 coordinate-space approach, 191–192
 elementary approach, 190
 including spallation gains and losses, catastrophic loss of particles, radioactive decay, 192
 diffusive velocity, 326
 dipole radiation, 158, 198, 199
 Dirac δ -function, 300

- Dirac delta function, 160, 165
 Fourier transform of, 165
- dispersion measure, 419
- dispersion measure of pulsars, 344–345, 346
- displacement four-vector, 136, 154
- displacement four-vector \mathbf{R} , 781
- dissipation time-scale, 324
- distance indicators, 757
 problems of using, 757
- distance measure D , 716
- distances in astronomy, 9, 755–759
 astrophysical methods, 757, 759
 Baade–Wesselink method, 758
 gravitational lenses, 759
 Sunyaev–Zeldovich effect, 758
 using distance indicators, 757
- Doppler shift formula, relativistic, 238
- ‘drop-out’ galaxies, 736, 737, 744
- Drude model, *see* electrical conductivity of a fully ionised plasma, Drude model for
- dust, *see* interstellar dust
- dust emission in millimetre and submillimetre waveband, 16
- dust extinction of galaxy spectral energy distribution, 739
- dust shells about giant stars, 74
- dwarf novae, *see* cataclysmic variables, dwarf novae
- dynamical friction, 131, 151–153
 clusters of galaxies and, 153
 galaxies and, 153
 globular clusters and, 153
 limits to collision parameters and, 152
 regular clusters of galaxies and, 153
- dynamical time-scale of a star, 41
- dynamics of a charged particle in a magnetic field with small scale fluctuations, 186
- early-type galaxies, 78, 80
 mass distribution in, 122
- Earth’s magnetosphere, *see* magnetosphere of Earth
- eclipsing X-ray binary with stellar mass black hole, 439
- eddies, turbulent, 454
- Eddington limiting luminosity, 71, 445–447, 449, 450, 454, 461, 462, 469, 470, 473, 478, 483, 490, 612, 634, 636, 656, 657, 658, 671
- Eddington luminosity, 413
- effective aperture of an antenna, 775
- effective area of the antenna A_e , 779
- effective temperature T_{eff} , 35, 48
- effective temperature of relativistic gas, 218
- Effelsberg 100 m radio telescope, 371
- 8–10 metre optical-infrared telescopes, 7
- Einstein angle, 118, 120
- Einstein coefficients
 relations between, 249
- Einstein coefficients for absorption and spontaneous and induced emission, 220
- Einstein coefficients for spontaneous and stimulated emission and stimulated absorption, 171
- Einstein radius, 117, 120
- Einstein ring, 117, 118
- Einstein X-ray Observatory, 25, 70, 113
- Einstein–de Sitter world model, 731
- electric dipole moment, 157
- electrical conductivity of a fully ionised plasma, 303–304, 308, 323, 324
 dissipation of energy and, 323
- Drude model for, 300, 303, 304
 including effect of electron–electron collisions, 304
 Lorentz approximation for, 303
- electrical conductivity of metals
 typical values for, 304
- electromagnetic field
 energy density of, in a medium, 267
- Electromagnetic Processes* (Gould), 129
- electromagnetic showers, 265, 272–275
 critical energy E_c for, 273, 274
 degradation of energy by ionisation losses at low energies, 273
 degradation of energy through the atmosphere, 273
 properties of, 275
 simple model for, 272, 273
 total number of particles with depth through a medium, 274
- electromotive force, 305, 306
- electromotive force \mathcal{E} , 181
- electron cyclotron radiation features in the X-ray spectra, 469
- electron degeneracy pressure, 68, 76, 378
- electron scattering, *see* Thomson scattering, 460, 461, 462, 659
- electron–photon cascades, *see* electromagnetic showers
- electron–positron annihilation line from the direction of the Galactic Centre, 278
- electron–positron pair production, 270–272
 cross-section at intermediate photon energies, 271
 cross-section in the ultrarelativistic limit, 271
 impossibility in free space, 271
 radiation length for, 271
 similarity of radiation length for bremsstrahlung and for, 271, 272
- electron–positron annihilation, 275–278
 annihilation at rest or in flight, 277
 broadening of γ -ray lines, 277
 cross-section for
 in extreme relativistic limit, 277
 thermal electrons and positrons, 278
 maximum and minimum energies of photons, 277
 of positronium atoms, 277
- electron–positron annihilation line, 290, 291

- elliptical galaxies
 as triaxial systems, 93, 94
 stability of, 94
 constancy of velocity dispersion with radius, 95
 dark matter in, 94, 95
 flattening not due to rotation, 93, 94
 mass-to-luminosity ratios for, 93
 rotation along minor axis, 93
 rotation of, 94
 variation of ellipticity with radius, 93
 emission from the Earth's atmosphere T_{atm} , 779
 emissivity κ_v , 170
 in terms of elementary atomic processes, 171
 emissivity of telescope in infrared waveband, 775
 encircled energy fraction, 768, 769
 for circular mirror with central hole, 772
 with varying wavefront errors, 770
 end-points of stellar evolution, 76
 energies
 useful conversion formulae, 780
 useful conversion formulae for hot bodies, 781
 useful conversion formulae for photons, 781
 energy dissipation rate by viscous forces, 456, 457
 energy generation, 47
 energy loss processes for high energy electrons, 536–540
 adiabatic losses, 538–539
 bremsstrahlung, 537
 inverse Compton scattering, 539–540
 ionisation losses, 537
 synchrotron radiation, 496, 539
 energy transport in stars, 45–47
 enthalpy per unit mass, 315
 for perfect gas, 316
 entropy of plasma
 conservation of, 308
 equation of hydrostatic support, 103, 451
 equation of mass conservation, 103
 equation of state of stellar material, 40, 47, 52
 equations of stellar structure, 47–50, 397
 equidistant azimuthal polar projection, 7, 755, 756
 equivalent current loop for particle gyrating in magnetic field, 181
 equivalent noise temperature T_n , 777
 EROS project, 125
 ESO-SERC Southern Sky Survey, 99
 Euler's constant, 168
 evolution of active galaxies with cosmic epoch, 720–729
 extragalactic radio sources, 720–724
 ‘luminosity evolution’, 720
 cut-off of strong evolution beyond redshift $z \sim 2–3$, 723
 for sources with steep and flat radio spectra, 720, 722
 luminosity-dependent density evolution, 723, 724
 radio-quiet quasars, 724–725
 evolution of galaxies and active galaxies with cosmic epoch, 714–743
 mid and far-infrared number counts, 737–740
 active galaxies, 715, 720–729
 extragalactic radio sources, 720–724
 radio-quiet quasars, 724–725
 X-ray clusters of galaxies, 729
 X-ray sources, 725–729
 brief history of evidence for, 714–715
 co-evolution of stellar and black hole properties of galaxies, 715
 counts of galaxies, 731–737
 counts of galaxies and active galaxies, 715
 Euclidean source counts, 715–716
 for standard world models, 731
 submillimetre counts of dusty galaxies, 740–742
 counts of galaxies and active galaxies, predicted
 for standard world models, 716
 submillimetre number counts, 740
 stellar and gaseous components of galaxies, 714
 submillimetre number counts, 743
 V/V_{\max} or luminosity–volume test, 718–719
 excess of faint blue galaxies
 irregular nature of, 737
 exoplanets, 66, 68
 Doppler technique for discovering, 66
 highly elliptical orbits of, 68
 Jupiter mass planets very close to parent stars, 68
 occultation technique for discovering, 67
 problems of accounting for orbits of, 68
 EXOSAT X-ray telescope, 449, 480, 483
 explosive nucleosynthesis, 65
 extensive air-showers, 293
 extinction
 role of absorption and scattering, 347–350
 extinction by interstellar dust grains, 9, 11
 extinction coefficient, 9
 extinction curves, 347
 extinction efficiency Q , 348, 349
 extragalactic radio sources, 661–680
 luminosity function of
 evolution with cosmic epoch, 720–724
 Extreme Ultraviolet Explorer (EUVE), 22, 23
- Faber–Jackson relation and fundamental plane, 87–88
 and the distances of galaxies, 88
 Faraday depolarisation, 346–347
 Faraday rotation of linearly polarised radio signals, 345–347, 369–371
 direction of magnetic field and, 346
 estimates of the Galactic magnetic field and, 346
 refractive indices of elliptically polarised waves, 345

- right- and left-handed elliptically polarised waves, 345
 rotation measure, 346
 Fermi energy, 402
 Fermi momentum of a degenerate Fermi gas, 397
 Fermi–Dirac distribution, 397
 Fermi–Thomas model of the atom, 174
 filling factor, 341
 fine structure constant, 135, 167, 229
 fine structure constant in electrodynamics, 400
 Finkelstein coordinates, 433
 FIRAS instrument of the Cosmic Background Explorer, 259
 first adiabatic invariant, 182
 first point in Aries, 753
 fluctuations in black-body radiation, 777
 fluctuations in the Cosmic Microwave Background Radiation, 749
 fluctuations per mode in black body radiation, 778
Fluid Mechanics (Landau and Lifshitz), 314
 flux densities, luminosity, magnitudes and colours in astronomy, 760–764
 flux density, 762
 bolometric, 762, 763
 definition of, 760
 Fokker–Planck equation for stars in clusters, 105
 fold catastrophe, 120
 forbidden transitions in spectra of gaseous nebulae, 342–343
 critical density for, 342
 metastable energy levels and, 342
 radiative de-excitation and, 342
Formation of Stars (Stahler and Palla), 361
 48-inch Schmidt Telescope Palomar Sky Survey (POSS), 99
 four-dimensional momentum space
 volume element of as a Lorentz invariant, 250
 four-frequency **K**, 782
 four-momentum **P**, 782
 four-velocity **U**, 782
 Fourier analysis, 771–774
 addition theorem, 771
 convolution theorem, 771, 772
 Fourier transform, 211
 Fourier transformation, 162, 163, 184, 200, 201, 202
 Fraunhofer diffraction, 766
 for circular aperture, 767, 768, 769
 encircled energy fraction, 768
 for circular mirror or lens, 767
 for rectangular slit, 766
 for single slit, 766
 free-fall velocity in general relativity, 431
 free-free emission, *see* bremsstrahlung
 frequencies of galaxies of different types, 736
 in different galactic environments, 732
 fully convective stars
 stability criterion for, 59
 fundamental plane, 87–88
Galactic Astronomy (Binney and Merrifield), 77
 Galactic centre, 753, 754, 755
 Galactic coordinates, 753, 754, 755
Galactic Dynamics (Binney and Tremaine), 77
 Galactic equator, 754
 Galactic latitude (*b*), 754
 Galactic longitude (*l*), 754, 755
 Galactic magnetic field, 225, 369–377
 aligned with local spiral arm, 374
 Faraday rotation in the interstellar medium, 369–371
 large fluctuations in, 369, 370
 large-scale order in, 369, 370, 374
 magnetic flux density from pulsar rotation measures, 226
 mean magnetic flux density, 377
 North Polar Spur and, 374
 optical polarisation of starlight, 371–374
 radio emission of spinning dust grains and, 374–376
 radio emission of the Galaxy and, 377
 summary of the information on, 377
 Zeeman splitting of 21-cm line radiation and, 377
 Galactic north pole, 754, 755
 Galactic south pole, 754, 755
 galaxies, 77–98
 barred spiral SB, 78
 collisions between, 79
 colour–luminosity relation for elliptical galaxies, 88
 elliptical E, 78
 ellipticities of, 78
 further correlations among the properties of galaxies, 86–89
 Faber–Jackson relation and fundamental plane, 87–88
 mass–metallicity relation for galaxies, 88–89
 Tully–Fisher relation for spiral galaxies, 87
 gas phase metallicity–luminosity relation for late-type galaxies, 88
 Hubble sequence, 78–80
 interactions between, 99
 irregular, 78
 large-scale distribution of
 holes and voids in, 101
 sheets and filaments in, 101
 lenticular or S0, 79
 luminosity density of the Universe due to, 97
 luminosity function of, 95–98
 mass–metallicity relation for galaxies, 88, 89
 mass-to-luminosity ratios for spiral galaxies, 92, 93
 masses of, 89–95
 mean number density in the Universe, 97
 mean separation in the Universe, 97
 normal spiral S, 78

- galaxies, (*cont.*)
 passive evolution of, 730
 importance of red giant branch, 730
 peculiar, 79
 red and blue sequences, 80–85
 ring, 80
 statistics of galaxies belonging to red and blue
 sequences, 85
 surface brightness distributions of, 81, 85
 discs, 82, 84
 spheroids or bulges, 82, 84
 ‘tuning fork’ diagram for, 78
- Galaxies in the Universe: An Introduction* (Sparke and Gallagher), 77
- galaxies, old red, 736
- Galaxy
 dark matter in, 356
 radio emission of, 224–227
 radio disc, 224
 thickness of radio disc, 225
 rotation curve of, 353, 355
- Galaxy Formation* (Longair), 118
- Galaxy Formation* (Longair), xiii, xiv, 78, 100, 127, 759
- galaxy formation
 collisions between galaxies and, 749
 ‘down-sizing’, 749
 feedback mechanisms and, 749
 growth of supermassive black holes and, 749
- Lyman-break galaxies
 multicolour technique for discovering, 744
- observed global star formation rate, 743
 star-forming galaxies and, 743
 starburst galaxies at large redshift, 743
 star and element formation and, 746
 star formation and, 749
- Galaxy, mass of, 759
- GALLEX solar neutrino experiment, 32, 57
- γ -ray background emission
 Comptonisation and, 729
 spectrum of, 728
- γ -ray bursts, 27, 384
- γ -ray luminosity of the Galaxy, 504
- γ -ray observations of the Galaxy, 493, 503–506
 distribution of cosmic rays in the Galaxy and, 503–506
- pion production and, 503
 cross-section for, 504
- relativistic bremsstrahlung of ultrarelativistic electrons, 504, 506
- spectrum of γ -rays produced by inverse Compton scattering of starlight, 504, 506
- spectrum of γ -rays produced from neutral pion decay, 504, 505
- γ -ray waveband, 25–27, 504
- γ -ray line emission, 27
- Cherenkov detection technique, 27
- detectors for, 26
- gauge selection in electrodynamics, 160
- Gaunt factor, 110, 111, 168
 frequency averaged, 168
- Gaunt factors, 302, 304
 for diffusion of particles in a plasma, 302
 for electrical conductivity of plasma, 302
- Gauss’s theorem for Newtonian gravity, 92, 119
- Gaussian point-spread function, 768
- Gemini Deep Deep Survey, 747
- Gemini North telescope, 440
- general relativity, 430–433
 dynamics of particles about point mass, 431
- GEO600 gravitational wave experiment, 33
- geodesic distance, 430
- giant branch, 37, 39, 49, 61, 69, 73, 74
- giant molecular clouds, 311, 362–364
 formation of, 356
 in spiral density waves, 357
 percolation processes and, 357
 supernova explosions and, 357
- globular clusters, 39, 69
 in M49, 94, 95
 in M49 and M87, 94
 oldest, 70
 typical parameters for, 153
- Goldstake solar neutrino experiment, 32
- GOLF experiment of the ESA SOHO mission, 52, 53
- Gran Sasso Laboratory, in Central Italy, 57
- grating telescope, 773
- gravitational deflection of light rays, 120
 by the Sun, 117
 collision parameter for, 117
- gravitational fine structure constant, 400
- gravitational lensing, 99, 759
 gravitational lensing by galaxies and clusters of galaxies, 116–122
 caustics and cusps in, 120
 cluster masses from, 120
 critical surface density for, 119
 distortion of background images by, 120
 galaxy–galaxy imaging and, 122
 necessary conditions for, 119
- gravitational lensing, strong, 122
- gravitational lensing, weak, 122
- Gravitational Lensing: Strong, Weak and Micro* (Schneider, Kochanek and Wambsganss), 118
- gravitational relaxation time τ_f , 152–153
- grazing incidence optics, 22, 23
- Great Observatories Origins Deep Survey (GOODS), 733, 734
- Great Observatories Origins Deep Surveys (GOODS), 745
- grey-body spectrum, 739

- Grotrian diagram, 342
group velocity, 344
guiding centre, 193, 208
guiding centre motion, 179, 182, 184, 186
gyration radius of cosmic ray protons in magnetic field, 502
gyrofrequency, 344, 345
non-relativistic, 195, 201, 202, 208, 209
relativistic, 193, 196, 201, 211, 214
gyrofrequency v_g , 179
non-relativistic, 179
gyroradius, 184
and rigidity R , 185
gyroradius of charged particle in magnetic field, 179
gyroscopic precession about rotating black holes, 435
- Hammer–Aitoff projection, 8, 9, 12, 13, 15, 18, 23, 26, 755, 756
Handbook of Space Astronomy and Astrophysics (Zombeck), 129
- hard low mass X-ray binary stars, 278
Harvard spectral classification system, 37
spectra of main sequence stars in, 37, 38
Hawking radiation, 438
Hawking, Stephen, 434
Hayashi tracks, 59–61, 64
for fully convective stars, 60
HEAO-1 X-ray Observatory, 725
HEAO-A2 experiment, 169
HEAO-C satellite, 27
HEAO-C2 cosmic ray experiment, 497
HEAO-C3 cosmic ray experiment, 499, 500
heat diffusion equation, 45
Heisenberg Uncertainty Principle, 135, 139
Heisenberg's uncertainty principle, 279, 339, 394, 395, 396
helioseismology, 51–54
acoustic or p -modes, 51
gravity or g -modes, 51
probing the structure of the Sun and, 51–54
helium
primordial nucleosynthesis of, 54
cosmic abundance of, 40
helium burning, 64, 72
helium flash, 73, 401
helium shell burning, 64
Herbig–Haro (HH) objects, 364, 366
Hertzsprung–Russell diagram, 35
Hertzsprung–Russell diagram, 35, 36, 37, 38, 39, 49, 59, 63, 64, 66, 68, 69, 70, 73, 74
for clusters of stars, 39
for white dwarfs, 402
theorist's, 68, 69
HESS Cherenkov γ -ray array, 27
Hess ultra γ -radiation, 28
- hierarchical models of galaxy formation
old red galaxies and, 747
High Energy Astrophysics (Longair), xiii
high energy astrophysics
definition of, 3
high energy astrophysics and modern physics and astronomy, 3–4
high energy electron energy spectrum in the local interstellar medium, 543–544
high mass X-ray binaries, 485–486
capture, or accretion, radius for stellar wind, 485, 486
O and B stars in, 485
properties of, 485
X-ray luminosity due to accretion, 486
Hipparcos astrometric satellite, 36, 757
Homestake gold-mine, South Dakota, 55
homologous stellar models, 47–48, 68
energy generation rates, 47
inadequacies of, 48
horizontal branch, 39, 40, 69, 73
evolution of stars on, 73
mass loss and, 39
hot gas in clusters of galaxies
absence of cool gas in, 113
associated with heating by radio lobes, 114
models to explain the, 114
abundance of iron, 110
characteristic cooling time for, 111
cooling flows and, 110–114
cooling flows in, 112
mass inflow rates of, 113
cooling time of, 112
distribution of, 109–114
iron line Fe xxvi from, 109
sound waves in, 114
HR diagram for stars, *see* Hertzsprung–Russell diagram
Hubble Deep Field, 714, 732, 736, 743, 744
Hubble sequence of galaxies, 77, 78–80, 86
correlations along, 86–89
luminosity function of H II regions, 87
neutral hydrogen, 86
star formation rates and, 87
total surface density and surface density of neutral hydrogen, 87
Hubble Space Telescope, 19, 71, 74, 78, 118, 122, 365, 379, 385, 392, 393, 578, 600, 603, 615, 618, 654, 663, 692, 693, 706, 714, 732, 733, 770, 771, 772, 777
Hubble Ultra Deep Field, 762
Hubble Ultra Deep Field (HUDF), 714, 732, 733, 734, 736, 737, 744, 745
Hubble's constant, 759
from Sunyaev–Zeldovich effect, 115
Hubble Space Telescope Key project, 759

- Hubble's constant, (*cont.*)
 measured from power spectrum of fluctuations in
 Cosmic Microwave Background Radiation,
 759
- Huygen's principle, 765
- Huygens' construction, 266, 269
- Hydrodynamics* (Lamb), 51
- hydrogen
 ionisation potential of, 42
 hydrogen recombination lines from warm component
 of the interstellar gas, 341
 hydrogen recombination lines in gaseous nebulae,
 341–342
 H β line of the Balmer series, 341
 very high order transitions, 342
- hydrogen shell burning, 64
- hydromagnetic waves, circularly polarised, 188
- hydrostatic equilibrium, 39, 47, 109
- hydrostatic support, equation of, 40
- IMB solar neutrino experiment, 32
- impedance of free space Z_0 , 157
- Infrared Astronomical Satellite (IRAS), 10, 13, 74
- infrared cirrus, 13
- infrared luminosity function of galaxies
 evolution of, 739, 740
- Infrared Space Observatory (ISO), 737
- infrared waveband, 9–14
 all-sky images in, 11–14
 emission by dust grains at thermal infrared
 wavelengths, 12, 13
 near-infrared wavelengths in, 10
 observing it, 9–11
 thermal infrared wavelengths in, 10
 wavelength windows in, 10
- initial mass function for stars, 361–362
- Miller and Scalo, 361
- Salpeter, 361
- inner Lagrangian point, 465, 467, 468, 475, 481,
 486
- instability strip, 73
- instantaneous rest frame, 158
- INTEGRAL γ -ray observatory, 27, 197, 257, 278
- INTEGRAL Gamma-ray Observatory, 288, 289, 292
- intensity
 definition of, 760
- intensity of radiation I_ν , 170
- interactions of high energy photons, 228–278
- intercloud medium, 342
- intercombination lines, 342
- interferometer
 eight-element
 power polar diagram of, 773
 four-element
 power polar diagram of, 773
- grating telescope, 773
- two-element, 771
 power polar diagram of, 773
- interferometry and synthesis imaging, 771–774
- internal energy per unit mass, 315
- International Linear Collider (ILC), 125
- International Ultraviolet Explorer (IUE), 22, 71, 392,
 638, 647, 651
- International Ultraviolet Observatory (IUE), 338
- interplanetary magnetic field, 303
- interstellar chemistry, 21, 145, 338
- interstellar dust, 347–353
 217.5 nm absorption feature, 350
 associated with $\pi \rightarrow \pi^*$ transitions, 350
 condensed matter physics of, 373
 cross-section for scattering and absorption, 348
 diffuse interstellar bands and, 348, 350
 electric charging of, 372–373
 coupling of grains and neutral particles through,
 373
 electric dipole moments of, 375
 emissivity of heated, 351
 extinction law for, 347–348
 formation of interstellar molecules and, 350
 graphite grains and, 350
 heavy elements in the interstellar medium and, 347
 Mie theory of scattering and absorption, 348, 349
 obscuration in our Galaxy and, 347
 optical depth for, 347
 ratio of total to selective absorption R_V , 347
 reddening, 347, 348
 reradiation of heated, 351
 star formation and, 351
 rotation frequency–grain radius relation
 for different phases of the interstellar medium,
 375
 rotation speed of dust grains, 372–373, 375
 rotation speed of PAH molecules, 375
 selective absorption, 347
 shielding of molecules by, 359, 363
 silicate absorption feature of, 348
 silicate absorption features at 9.7 and 18 μm , 350
 size parameter, 348, 349
 small grains in, 350
 transient heating of, 350
 sublimation temperature of, 347, 352
 water ice feature at 3.1 μm , 350
 wide range of grain sizes present in, 350
- interstellar gas
 average properties of, 333
 cooling mechanisms, 358–359
 bound–bound or bound–free emission, 359
 bremsstrahlung, 359
 in gaseous nebulae, 359
 interstellar dust emission, 359
 low lying energy levels of common elements, 359
 molecular line emission, 359

- diagnostic tools
 21-cm line emission and absorption, 333–335
 column depth from X-ray absorption, 340
 dispersion measure of pulsars, 344–345
 Faraday rotation of linearly polarised radio signals, 345–347
 ionised interstellar gas, 340–347
 molecular radio lines, 335–338
 neutral interstellar gas, 333–340
 optical and ultraviolet absorption lines, 338–339
 permitted and forbidden transitions in gaseous nebula, 340–343
 thermal bremsstrahlung, 340
 X-ray absorption, 339–340
 heating mechanisms, 357–358
 collisions of old supernova shells, 357
 cosmic rays and, 358
 ionisation rate by ionisation losses, 358
 other types of, 358
 supernova explosions, 357
 ultraviolet radiation of hot stars, 358
 hot component of, 357, 358
 overall picture of, 353–361
 large scale dynamics, 353–357
 overall state of, 359–361
 thermal instabilities, 361
 condition for, 360–361
 interstellar gas and magnetic fields, 331–377
 interstellar medium
 phases of, 359, 360
 cold neutral medium, 360
 coronal gas, 360, 361
 diffuse clouds, 360
 giant molecular clouds, 360
 H I clouds, 360
 intercloud medium, 360
 two-phase model for, 361
 interstellar medium and the life cycle of stars, 333
 interstellar molecules, 20
 hydroxyl molecule OH, 20
 line emission of, 16
An Introduction to the Ionosphere and Magnetosphere
 (Ratcliffe), 305
 invariant four-volume, 240
 invariant four-volume in four-momentum space, 240, 250
 inverse β -decay, 385, 404
 inverse β -decay process, 401
 inverse Compton radiation, 490
 inverse Compton scattering, 237–243, 244, 260
 average energy of scattered photons, 242
 derivation of formulae for, 237–243
 energy density of radiation in moving frame, 238–240
 geometry of, 238
 maximum energy of scattered photon, 242
 of radio, infrared and optical photons, 243
 similarity of loss rate to synchrotron loss-rate, 241, 261
 spectral index of radiation of a power-law distribution of electron energies, 241
 spectrum of scattered radiation by a single electron, 241, 242
 total energy loss rate of, 241
 inverse square laws of electrostatics and gravitation, 151
 ion-acoustic instability, 328
 ionisation loss formula
 adapted for radiation damage density, 147
 ionisation losses, 292, 301, 358
 ionisation losses of electrons, 145, 146
 maximum energy transfer per interaction, 146
 ionisation losses of protons and nuclei, 131–153
 average energy loss per unit path length, 133
 Bethe–Bloch formula, 140–141
 cancer therapy and, 145
 density effect, 141
 effects of polarisation of medium, 141
 energy spectrum of ejected electrons, 135
 heating of giant molecular clouds and, 131, 145
 lower limit b_{\min} to collision parameters
 classical limit, 134
 quantum limit, 134
 mean energy loss rates in different materials, 142, 143
 minimum loss rate, 141, 142
 non-relativistic treatment, 131–136
 nuclear emulsions and, 146
 number of ion–electron pairs produced, 145
 particle detectors and, 131
 practical forms of the ionisation loss formulae, 141–145
 range R and, 143, 144
 relativistic treatment, 136–141
 stopping power, 142
 straggling, 144
 upper limit b_{\max} to collision parameters, 133
 ionisation potential I , 135, 142
 IRAM millimetre interferometer, Plateau de Bure, 365
 IRAS galaxies, 741, 742
 IRAS infrared observatory, 714, 737
 IRAS space mission, 353
 Irvine–Michigan–Brookhaven (IMB) neutrino experiment, 389
 ISAAC infrared spectrograph of the ESO VLT, 441
 isothermal carbon core, 64
 isothermal gas spheres, 102–106, 119
 core radius, 104
 projected, 104
 singular, 119, 120
 structural index, or structural length, 104

- isothermal gas spheres, (*cont.*)
 tidal radius of, 105
 truncated, 105
 isothermal helium core, 63
- Jacobian for transformation between inertial frames of reference, 240
- James Clerk Maxwell Telescope, 742
- James Clerk Maxwell Telescope (JCMT), 14
- James Webb Space Telescope (JWST), 11
- Jansky (Jy), 760
- Jeans length, 368
- Jeans' instability, 368
 fragmentation within giant molecular clouds, 368
 in molecular clouds, 368
 time-scale for collapse, 368
- Jeans' length, 368
- Jeans' mass, 368
- Jet Propulsion Laboratory, 391
- Jupiter, mass of, 66, 759
- K-correction, 716
- K*-corrections, 732
- K20 sample of galaxies, 736
- K20 Survey, 747
 very red galaxies in, 748
- Kamiokande neutrino experiment, 389
- Kamiokande solar neutrino experiment, 32
- Keck 10 metre telescope, 736
- Kelvin–Helmholtz instability, 320
- Kelvin–Helmholz time-scale, 43
- Kennicutt, Robert, 750
- Kepler's third law of planetary motion, 92, 356
- Keplerian orbits, 451, 452, 456, 457, 630
- Keplerian velocities, 456, 457
- Keplerian velocity, 454, 463, 471, 472
- Kerr black holes, 434
 maximum energy release, 436, 445
- Kerr metric, 434, 435
- kinetic energy per nucleon, 180
 ionisation losses and, 142
- King profiles for mass distribution in clusters, 105, 107
- Kirchhoff's law, 340
- Klein–Nishina cross-section, 237, 264
- Kompaneets equation, 246, 250–257, 259, 260
 ‘current’ of photons in phase space and, 251
 diffusion coefficient for photons in phase space, 252
 diffusion of photons in phase space and, 252
 expansion to higher orders in $\partial n/\partial x$, 259
 formation of power-law spectrum by thermal processes, 255
- formation of Wien peak, 253–255
 induced scattering and, 250
 Monte Carlo solutions of, 254
 photon conservation and, 251
- power-law solutions of, 255–257
 recoil effect and, 251, 255
 recoil effect and induced scattering, 252
 solutions for $\hbar\omega \gg kT$, 253
 spectra of X-ray sources and solutions of, 253–257
- Kramers opacity, 460, 461
- Kruskal coordinates, 433
- Kuiper Airborne Observatory, 14, 359
- lack holes
 mass from kinematics of nuclear gas clouds, 448, 450
- Lagrangian formulation of classical dynamics, 182
- Lagrangian mechanics
 action integral in, 182
- Λ CDM model of galaxy and structure formation, 749–752
 problems with
 accounting for Faber–Jackson relation and Tully–Fisher relation, 752
- Lane–Emden equation, 103
- Lane–Emden equation, 398
- Laplace's equation, 425
- Large Hadron Collider (LHC), 125
- Large Magellanic Cloud, 125
 distance from observations of SN 1987A, 392
- Larmor formula, 374, 376
- Larmor's formula, 158
- Larmor's formula for radiation of accelerated electron, 232
- laser emission, 778
- Laser Interferometer Gravitational-Wave Observatory (LIGO), 33
- last stable circular orbit, 432, 433, 436
- late-type galaxies, 78, 80
- Lectures on Physics, Vol. III* (Feynman), 247
- Leiden–Berkeley Deep Survey (LBDS), 723
- Lexan polycarbonate, 147, 148
- Liénard–Wiechert potentials, 200, 202
 doppler shift factor in, 202, 203
- Liénard–Wiechert potentials, 161, 266
- lifetimes of stars of different mass, 361
- Lifshitz, Evgenii, 452, 456
- light echo technique, 385
- light, L, group of elements, 508
- light-year, definition, 756
- line blanketing, 88
- Lockman Hole survey field, 725
- long period variables, 73
- Lorentz factor, 139, 141, 146, 178, 180, 293, 431, 502
- Lorentz factor γ , 193, 211, 222, 781
- Lorentz force, 178, 183, 425
- Lorentz gauge, 160
- Lorentz transformation
 inverse, 140

- Lorentz transformations, 155, 781
 inverse, 239
- Lorentz transforms for electric and magnetic fields, 137–139, 194
- low mass X-ray binaries, 479–484, 490
 accretion disc corona, 480, 481
 dependence of observed properties on angle of inclination, 479
 eclipses and ‘dips’ in X-ray light curves, 479
 quasi-periodic oscillations (QPOs), 483, 484
 models for, 484
- thick absorbing screen about accretion disc, 481
- X-ray burst sources (bursters), 481
 comparison of accretion and thermonuclear runaway luminosities, 482
 properties of, 481
 Type I, 481–483
 Type II, 483
 X-ray colour–colour diagram for, 484
- luminosity
 definition of, 763
- luminosity classes for stars, 37, 38
- luminosity density of starlight in the Universe, 97
- luminosity function of galaxies, 95–98, 719
 for red and blue galaxies, 96
 luminosity density of starlight and, 97
 mass-to-luminosity ratio
 for the Universe as a whole, 97
- Schechter, 96
 break absolute magnitude M^* , 96
 break luminosity L^* , 96
 written in absolute magnitudes, 96
 useful statistics about galaxies, 97–98
- luminosity function of stars, 361
- luminosity indicators for stars, 37, 38
- luminosity–temperature diagram for stars, *see* Hertzsprung–Russell diagram
- luminous infrared galaxies
 evolution of, 740
 N1–015
 spectral energy distribution of, 739
- luminous infrared galaxy N1-015
 spectral energy distribution of, 739
- Lundquist number, 326
- Lyman continuum absorption, 22
- Lyman limit for hydrogen, 22
- Mach number, 316
- Mach number, M , 452, 454, 554, 570, 577, 579, 671
 Alfvén, 575
- MACHO project, 124, 125
- MACHOs, 124
 mean mass of, 125
- Magellanic Clouds, 446, 447
- magnetars, 418–419
 magnetic fields in, 419
- magnetic buoyancy, 321–323
 formation of magnetic loops and, 323
 in a plane-parallel stratified atmosphere, 322
- magnetic dipole radiation, 422, 425, 427
 pulsars and, 408
- magnetic field fluctuations in interplanetary medium
 power spectrum of, 184, 185, 187
- magnetic field, energy density of, 307
- magnetic field, Galactic, 225
- magnetic fields
 in extragalactic radio sources, 18
 slowly varying, 180, 182
- magnetic fields in our Galaxy, *see* Galactic magnetic field
- magnetic fields in supernova shock fronts, 574–576
- magnetic fluctuations generated by Alfvén and hydromagnetic waves in interstellar medium, 187–189
 physical model for, 187
- magnetic flux freezing, 304–314, 405, 425, 463
 change of magnetic flux density with density of plasma, 310
 diffusion time for magnetic field in a plasma and, 310
- magnetohydrodynamic approach, 307–311
- physical approach, 305–307
 similarity to adiabatic motion of charged particles, 183
- magnetic lines of force
 Faraday’s concept of, 321
 magnetic flux freezing and, 305
- magnetic lines of force, reconnection of, *see* reconnection of magnetic lines of force
- magnetic mirroring, 181
 in the Earth’s radiation belts, 181
- magnetic moment μ of the current loop, 181
- Magnetic Reconnection* (Priest and Forbes), 329
- magnetic reconnection, *see* reconnection of magnetic lines of force
 magnetic Reynolds’ number, 311
 ‘longitudinal’, 326
 Lundquist number and, 326
- magnetic rigidity, *see* rigidity R
- magnetic tubes of force, 322
- magnetoactive medium, 345
- magnetohydrodynamics
 aspects of, 298–329
 derivation from microscopic description of particle dynamics in a magnetic field, 184
- equations of, 307–308
 entropy equation, 308
 equation of continuity, 307
 force equation, 307
 Maxwell’s equation, 307
 Ohm’s law, 308
- magnetopause, 320

- magnetosheath, 321
 magnetosphere of Earth, 319–321, 323
 collisionless plasmas in, 321
 collisionless shocks in, 321
 plasma wave interactions in, 321
 magnetopause, 320
 magnetotail, 320
 neutral sheet, 320
 plasma sheet, 320
 shock wave discontinuity in, 321
 magnetotail, 320, 323
 magnitude
 absolute, definition of, 763
 apparent, definition of, 760
 bolometric absolute, definition of, 763
 bolometric apparent, definition of, 762
 magnitudes in optical and infrared astronomy, 760
 main sequence, 36, 37, 38, 39, 40, 44, 48, 49, 50, 61,
 62, 63, 68, 69, 70, 71, 764
 main sequence termination point, 39, 68, 73
 main sequence lifetimes of sun and stars, 50
 Mariner 4, 184, 185
 maser action of interstellar molecules, 21
 maser emission, 778
 mass absorption coefficient for high energy photons,
 271, 272
 mass conservation, equation of, 41
 mass function, 439
 mass function of a binary system, 414, 439
 mass loss
 formation of white dwarfs and, 378
 quiescent, 70
 rates of for massive stars, 71
 mass–luminosity ratio
 early-type galaxies, 122
 mass–luminosity relation for main sequence stars, 36,
 48, 50, 68
 mass–luminosity relation for stars, 361, 731
 mass–metallicity relation for galaxies, 88–89
 mass-to-luminosity ratio
 for the Universe as a whole, 97
 for Universe as a whole, 97
 for visible parts of galaxies, 98
 of elliptical galaxies, 97
 of spiral galaxies, 97
 masses in astronomy, 9, 759–760
 masses of galaxies, 89–95
 masses of elliptical galaxies, 93–95
 from widths of absorption lines, 93, 95
 rotation curves of spiral galaxies, 92–93
 virial theorem for galaxies and clusters, 89–92
 massive galaxies in clusters, 750
Mathematical Theory of Black Holes
 (Chandrasekhar), 429
 Mauna Kea Observatory, Hawaii, 14
 Maxwell's equations in free space, 158
 Maxwellian velocity distribution, 168
 McGill pulsar group, 419
 mean free path
 for pitch angle scattering, 186
 of particle in a plasma, 301
 of a proton in the interplanetary medium, 303
 mean free path λ , 454
 mean free time between particle collisions in a
 plasma, 303
 Medium Deep Survey of galaxies (MDS), 736
 medium, M, group of elements, 508
 metallicity Z, 68, 69, 88
 metals, 40
 meteorites, 148–151
 constancy of the flux of cosmic rays and, 151
 exposure ages of, 151
 fossil tracks of cosmic rays and, 148–150
 Galactic cosmic rays and, 149, 150
 history of Solar System and, 148–151
 Solar cosmic rays and, 149, 150
 Mg₂ index, 88
 correlation with ($B-V$) colour, 88
 microquasar, 440
 Mie theory of scattering and absorption, 348, 349
 Mikheyev–Smirnov–Wolfenstein (MSW) effect, 57,
 58
 Milky Way, 755
 Millennium galaxy catalogue, 83, 97
 Millennium Simulation, 749, 751, 752
 first quasar candidate, 751
 Miller and Scalo initial mass function, 361, 362
 millimetre and submillimetre waveband, 14–17
 all-sky images in, 14–17
 Cosmic Microwave Background Radiation in, 14
 molecular lines in, 14
 observing in, 14
 millisecond pulsars, 405, 417–418
 as members of old galactic populations, 420
 discovery of, 418
 distribution in the Galaxy, 420
 in globular clusters, 418
 limit to spin-up of, 418
 location on $P-\dot{P}$ diagram, 418
 magnetic fields in, 418
 space velocities of, 420
 spin-up of, 418
 minimum energy requirements for synchrotron
 radiation, 549–552
 Minkowski metric, 430
 MKK spectral classification system, 37
 modified Bessel functions of order 2/3 and 1/3, 206
 asymptotic expressions for, 210
 modified Bessel functions of order zero and one, 164
 asymptotic values of, 164
 molecular hydrogen
 distribution in the Galaxy, 356

- molecular radio lines, 335–338
 list of molecules identified by their, 335
 maser action and, 335
 molecular doubling processes and, 337
 rotational ladder of, 337
- molecules in the interstellar gas, 335–338
 acetylenic, 338
 CO as a tracer of molecular hydrogen, 337
 discovery of, 335
 electronic transitions of, 335
 glycine, 338
 in large redshift quasars, 338
 list of known species, 335
 molecular hydrogen, 337
 rotational transitions of, 335
 self-shielding, 337
 shielding by dust, 337
 types of, 338
 unsaturated, 338
 vibrational transitions of, 335
- momentum four-vector, 139, 140, 154
 for electrons, 235
 for photons, 235, 275
- momentum impulse, 132, 139, 165, 301
- morphological classification of galaxies, 77–80
- morphologies of clusters of galaxies, 99–102
- Mount Hopkins Observatory, Arizona, 11
- muons
 decay of, 293
 mean lifetime of, 293
 test of relativistic time dilation and length contraction, 293
- underground
 measurements of intensity and isotropy of cosmic rays from, 293
- negative pressure equation of state, 433
- neutral hydrogen
 21-cm line emission of, 20
 21-cm map of the Galaxy, 20
 distribution in the Galaxy, 356
- neutral sheets, 320, 324, 325, 327
 thickness of, 326
- Neutrino Astrophysics* (Bahcall), 54
- neutrino interactions
 charged current interaction (CC), 57, 58, 59
 elastic scattering (ES), 57, 58, 59
 neutral current interaction (NC), 57, 58, 59
- neutrino oscillations, 57, 125
 and solar neutrinos, 57, 58, 59
 in atmospheric μ neutrinos, 58
 in terrestrial neutrino experiments, 58
- neutrinos
 laboratory limits to the masses of, 125
 laboratory limits to the number of species of, 126
- neutrinos, solar, *see* solar neutrinos, 279
- neutron capture γ -ray line at 2.223 MeV, 290, 291
- neutron degeneracy pressure, 76, 378
- neutron drip, 403
- neutron production by (γ, n) interaction, 297
- neutron star, 311
 binding energy of, 379
 diffusion time for magnetic field from, 311
 magnetic flux density of, 311
- neutron stars, 19, 68, 76, 378, 401–406
 cooling by neutrino emission, 421
 discovery of, 406–419
 ellipticity ϵ due to rotation, 424
 galactic population of, 419–420
 in X-ray binary systems, 410–415
 internal structure of, 401–406
 core region, 404
 inner crust, 404
 inverse β -decay process and, 401
 neutron liquid phase, 404
 outer crust, 404
 superconductivity and superfluidity and, 404–405
 surface layers, 403
 zones within, with increasing density, 403–404
- magnetic dipole moment of, 408
- mass estimates in binary star systems, 416, 417, 439
- predicted surface temperatures, 421
- rotating neutron superfluid and, 424
 pinning of quantised vortices in, 424
 quantisation of angular momentum, 424
- rotation of, 405
- rotational break-up speed, 405
- scale height of atmospheres of, 421
- thermal emission of, 421–422
 ‘truly isolated’, 422
- X-ray emission from surfaces of, 421
- neutronisation, 402
- Newton’s law of gravity, 759
- Newton’s laws of motion, 759
- Nobeyama Radio Observatory, 45 m millimetre telescope, 364
- noise power, 779
 fluctuations of, 778
 in thermal equilibrium, 777
- noise temperature
 of the receiving system T_{sys} , 779
- non-electromagnetic astrophysics, 32–34
 astroparticle physics, 33–34
 gravitational waves, 32–33
 neutrino astrophysics, 32
- nonlinear diffuse shock acceleration, 579
- non-relativistic gyroradiation, 195–198
 energy loss rate by, 195
 linear and circular polarisation of, 195, 214
- non-thermal emission, 193
- non-thermal radiation, 5

- non-thermal sources
 spectra of, 212
- nonlinear diffuse shock acceleration, 576
- norm of four-vector, 161
- norm of the four-vector, 781
- north celestial pole (NCP), 753, 754, 755
- North Polar Spur, 224, 374
- northern celestial hemisphere, 754
- novae, 383, 478–479
- nuclear γ -ray lines
 important, 292
- nuclear cascades, 275
- number of particle produce in, 281
- nuclear deflagration, 479
- nuclear emission lines, 287–292
 ^{44}Ti from supernova remnant Cas A, 289
 asymptotic giant branch stars, 288
 collisional excitation of nuclei, 289–292
 cross-sections for, 289, 290
 in the interstellar medium, 289, 290, 291
 decay of radioactive isotopes, 287–289
 astrophysically important examples of, 287
 conditions for observability of, 287–288
 diffuse ^{26}Al and ^{60}Fe emission from the Galaxy, 288
 diffuse ^{26}Al emission from the Galaxy, 288
 explosive nucleosynthesis and, 288
 supernovae and, 288
 Wolf–Rayet stars and, 288
- nuclear emulsions, 146
 development of, 146
 discovery of elementary particle and, 146
 silver bromide crystals, AgBr, 146
- nuclear energy generation rates in stars, 43–45
- nuclear interaction cross-section, 279, 281
- nuclear interactions, 279–297
 cosmic rays in the atmosphere, 292–297
 high energy astrophysics and, 279–282
 multiple scattering within nuclei, 279, 280
 neutron production in, 280
 nuclear emission lines, 287–292
 nucleus–nucleus collision, 281–282
 pion and strange particle production in, 280, 281
 spallation cross-sections, 282–283
 spallation fragments, 280
- nuclear interactions in the atmosphere
 mean free path for, 281
- nuclear interactions of high energy particles with
 nuclei of atoms and molecules, 131
- nucleonic cascades
 particle detectors and, 293
 total ionisation as a measure of energy of primary
 cosmic ray, 293
- nucleosynthesis
 explosive, 386–388
 late stages of, in massive stars, 385
- shell burning, 384
- steady-state, 386–388
- steady-state hydrostatic, 386
- synthesis of isotopes and, 386, 387
- Nucleosynthesis and Chemical Evolution of Galaxies*
 (Pagel), 497
- nucleosynthesis in massive stars, 378
- nucleosynthesis, primordial, 123, 500
- nucleus, radius of, 279
- Nyquist's theorem, 777
- OB supergiant mass-loss rates, 485, 486
- observability of the sky in different astronomical
 wavebands, 5, 6
- observations in cosmology for Friedman world models
 flux density–redshift relations
 for starburst galaxies in the submillimetre
 waveband, 741, 742
- occupation number, 240, 246–250
 as a Lorentz invariant, 250
 for Bose–Einstein distribution, 249
 for Planck distribution, 249
 in Rayleigh–Jeans region of Planck spectrum, 249
- mean, 248
 spontaneous and induced processes and, 246
- OH/IR stars, 73
- old red galaxies, 746–749
 constancy of masses with redshift, 747
 early formation of stellar populations of, 748
 evidence for old stellar populations in, 747
 evolving stellar mass density and, 747
 massive, 746–749
 need for early starbursts, 748
 selected in the K waveband, 747
 star-forming galaxies at large redshift and, 748
 stellar masses as a function of redshift, 747
- 'onion-skin' chemical structure of massive stars, 65
- opacity κ of stellar material, 45–47, 48
- optical and infrared wavebands used in ground-based
 astronomy, 761
- optical and ultraviolet absorption lines, 338–339
 curve of growth of, 339
 D lines of sodium, 338
 deficit of heavy element in the interstellar gas,
 339
 deuterium and, 339
 Doppler broadening of, 339
 equivalent width of, 339
 H and K lines of calcium, 338
 highly ionised OVI and, 339
 molecular hydrogen and, 339
 radiation or natural broadening of, 339
- optical depth for radiation τ , 172
- optical polarisation of starlight, 371–374
 alignment of interstellar grains and, 372–374
 as a function of Galactic coordinates, 374

- degree of polarisation correlated with extinction, 372
maximum degree of, 371
parallel to minor axis of grains, 372
optical waveband, 5–9
all-sky images in, 7–9
observing in, 5–7
Orbital Astronomical Observatories II (OAO-II), 338
orbital migration, 68
origin of cosmic rays in our Galaxy, 536–560
The Origin of the Chemical Elements (Tayler), 496
OSO III satellite, 26
overdensity of galaxies about any galaxy, 84
OVRO Millimetre Array, 115, 116
oxygen burning, 64
- P(D)* distribution, 720, 725
P-Cygni profiles, 70–73, 478
mass outflows and, 71
p-p chain, *see* proton–proton chain (p-p chain)
paleogeomagnetic studies of Earth's magnetic field, 296
Palomar 200-inch telescope, 7, 591
parallax-second (or parsec, pc), definition, 755
parallaxes, 757
Parkes Radio Telescope, 410
Parkes Selected Region (PSR) sample, 723
Parseval's theorem, 162–163, 269
applied to magnetic field fluctuations in interplanetary medium, 184
spectral distribution of the radiation of an accelerated electron and, 162–163
partial spallation cross-sections, *see* spallation cross-sections, partial
particle–antiparticle annihilation, 275
Pauli exclusion principle, 126
 \dot{P} versus P diagram for pulsars, 410
‘graveyard’ region of, 410
as an evolutionary sequence for pulsars, 410
death line for pulsars on, 410, 418
pulsar equivalent of the Hertzsprung–Russell diagram, 410
peculiar galaxies, 79
strong gravitational encounters or collisions, 79
tails associated with prograde encounter between galaxies, 80
Penrose process, 437
perchloroethylene C₂Cl₄, 54
perfect gas law, 109
permitted and forbidden transitions in gaseous nebula, 340–343
excitation by ultraviolet emission of hot stars, 340
photoexcitation and photoionisation, 340
temperature of emitting regions, $T_{\text{gas}} \approx T_{\star}$, 341, 358, 359
- permittivity
relation to refractive index n , 267
Perseus Cluster of galaxies, 750
Petrosian r-band luminosity, 85
Petschek mechanism of magnetic reconnection, 327
phase space
elementary volume of, 248
photodisintegration of iron nuclei, 385
photoelectric absorption, 228–230
cross-sections for, 228–230
spectra of X-ray sources at energies $\varepsilon \sim 1$ keV, 228–230
photoelectric effect, 228
photoionisation
cross-section for hydrogen atoms, 341
photon loss processes
summary of absorption coefficient for, 271
photon–photon collisions, 275
as a source of opacity for high-energy γ -rays, 276
cross-section for
in the regime $\bar{\varepsilon} \approx m_e c^2$, 277
in ultrarelativistic limit, 276
threshold energy for electron–positron pair production, 275–276
examples of, 276
physical constants in SI units, 780
The Physics of Fully Ionised Gases (Spitzer), 298, 302
The Physics of Plasmas (Fitzpatrick), 298
Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena (Zeldovich and Raizer), 314
Pioneer missions of NASA, 312
pionisation, *see* cosmic rays in the atmosphere, pion production
pions
decay of, 293
mean lifetime of, 293
pitch angle, 193, 195, 196, 198, 201, 203, 206, 207, 208, 213, 215, 222
isotropic distribution of, 195, 213
pitch angle θ , 178, 179, 184
pitch angle scattering, 187
physical model for, 184
Planck distribution, 352, 459
Planck spectrum, 5, 246, 247, 248
as a solution of Kompaneets equation, 253
Planck spectrum of black-body radiation, 170, 171
plane of the ecliptic, 753, 754
planetary nebulae, 73–74
central stars of, 74
evolutionary tracks for, 75
images of, 74
mass loss and, 74
mass loss events and the structures of, 74
plasma frequency, 298–300, 344, 345
angular, 298

- plasma physics
 aspects of, 298–329
 elementary concepts in, 298–304
Plasma Physics for Astrophysics (Kulsrud), 298, 329
 plasma sheet, 320
 Plummer model for elliptical galaxies, 106
 point-spread function, 768
 Poisson statistics, 774
 Poisson's equation
 in electrostatics, 299
 Poisson's equation for gravity, 106, 368
 polars, 197
 polycyclic aromatic hydrocarbon (PAH) molecules, 350–351
 planes of hexagonal benzene rings and, 351
 unidentified infrared lines and, 351
 polycyclic aromatic hydrocarbon molecules (PAHs), 739
 polytropes, 398
 polytropic index, 398
 positron production mechanisms, 275–278
 positronium atoms, 277
 positrons
 sources of, 275
 creation of electron–positron pairs, 275
 decay of π^+ pions, 275
 decay of long-lived radioactive isotopes, 275
 Poynting vector, 157
 Poynting vector flux, 267, 268
 Poynting's theorem, 233
 primordial nucleosynthesis, 98
 principle of detailed balance, 170, 217
 Principle of Equivalence, 430
 principle of jump rate symmetry, 248
Principles of Optics (Born and Wolf, 764)
 probability density function, 775
 projected surface brightness, 110
 projections of celestial sphere onto a plane, 753–755
 proper acceleration, 158, 161
 proper time
 in general relativity, 431, 432
 in special relativity, 430
 proton decay, 390
 proton–proton chain (p-p chain), 44, 48, 50, 54, 56
 pp1, pp2 and pp3 branches, 54
 protostars, 311, 352–353
 accretion shock in, 353
 as intense far infrared sources, 353
 diffusion time for magnetic field from, 311
 dust photosphere of, 352, 353
 hydrostatic core of, 352
 release of binding energy by far infrared emission, 353
 spectra of, 353, 354
 structure of, 352–353
 pulsar glitches, 409, 422–424
 changes in moment of inertia of neutron star and, 422
 glitch function, 423
 healing parameter, 423
 migration of quantised vortices and, 424
 spin-up following, 422
 time constant of τ_c , 422, 423
 starquakes and, 423
 two-component model of interior of neutron star and, 422, 423
 pulsar magnetosphere, 424–427
 beaming of radio emission and, 427
 closed field lines in, 426
 corotation radius, 426
 critical field line, 426
 dominated by electromagnetic forces, 425
 inevitability of fully conducting plasma surrounding, 425
 inner and outer acceleration gaps, 428, 429
 radio emission from, 429
 light cylinder, 426
 magnetic field distribution in, 426
 open field lines in, 426
 polar cap regions of, 427
 potential differences in, 427
 sparks in, 428
 space charge distribution in, 425
 strong electric fields in, 425
 zero charge cones, 426, 429
 pulsars, 345
 ages from braking index, 409
 ages of, 409
 anomalous X-ray, 418–419
 as magnetised rotating neutron stars, 424
 as rotating, magnetised neutron stars, 406, 408
 as spiral arm populations, 419
 associated with supernova remnants, 419
 binary, 415–417
 dispersion measures of, 419
 high velocities of
 asymmetric supernova explosions and, 467
 disruption of binary systems and, 467
 incoherent infrared, optical and X-ray emission
 from outer gap regions, 429
 kinematic ages of, 420
 large velocities due asymmetric collapse of
 core-collapse supernovae, 420
 large velocities due to disruption of close binaries, 420
 luminosity function of, 420
 magnetic flux density from slow-down rate, 410
 millisecond, 405, 417–418
 normal radio, 406–410
 \dot{P} versus P diagram, 410, 411
 radio emission of, 405
 rate of loss of rotational energy, 409

- rates of formation of, 420
 scale-height in the Galaxy, 420
 space densities of, 420
 space velocities of, 420
 spin-up, 411, 471, 472
 by accretion, 473
 maximum rotation rate and, 472
 spin-up rate–luminosity relation, 471, 472
 timing noise, 422
- quantisation of rotational angular momentum, 335
 quantum theory of gravity, 433
 quark matter, 404
 quasar spectra
 Lyman- α emission line, 589, 590
 prominent emission lines in, 590
- quasars
 3CR sample of radio, 722
 early formation of supermassive black holes in, 748
 low-luminosity, 592, 724
 radio quiet, *see* radio quiet quasars
 ratio of black hole to spheroid masses, 748, 750
- quasi-Stellar radio sources, 19
- radiation damage, 131, 146–151
 in plastics and meteorites, 146–151
 polymers and, 147
- radiation damage density J , 147, 148
- radiation losses of accelerated charged particles
 non-relativistic, 155–161
 relativistic, 161–162
- radiation of accelerated charged particles, 154–163
 from Maxwell's equations, 158–161
 J.J. Thomson's treatment, 155–158
 loss rate of, 193
 polar diagram of, 158
 polarisation of, 158
 properties of, 158
 relativistic invariants, 154–155
 total energy loss rate, 154, 155, 174
 total energy loss rate of, 158
- radiation pressure in stars, 45, 48
- Radiation Processes in Astrophysics* (Rybicki and Lightman), 129
- radiation resistance, 779
- radiation-driven pulsational instability for massive stars, 71
- radiative transport of energy in stars, 45–47
- radio background emission from our Galaxy T_{gal} , 779
- radio emission of spinning dust grains, 374–376
 Galactic background radiation and, 376
- radio emission of the Galaxy, 224–227
 comparison with predicted emission using local energy spectrum of cosmic ray electrons, 226
- determination of radio spectrum and radio emissivity, 224–226
- radio emissivity of, 225
 in direction of opaque regions of ionised hydrogen, 226
- radio spectrum of, 224, 225
- radio galaxies
 3CR sample of, 722
 discovery of, 18
 energy requirements of, 18
 evidence for old stellar populations in, 748
- radio galaxy 0902+34, 748
- radio galaxy LBDS 53W069, 748
- radio galaxy LBDS 53W091, 748
- radio map of the sky, 224
- radio pulsars, 19, 76
 brightness temperatures of, 427, 428
 coherent radiation of, 427
 maser emission of, 428
 radio emission of, 427–429
 curvature radiation and, 428
 electron/death line and, 428
 electron/positron–photon cascades and, 428
- radio quiet quasars
 counts of, 591, 592, 724
 cut-off at large redshifts, 590, 724, 725
 definition of complete samples of, 591
 completeness of, 591
 dispersion prism-grating techniques, 591
 multi-colour photometric technique, 590–591
 searches for ‘*i*-band drop-outs’, 591, 725
 searches for Lyman- α and CIV emission lines, 591
 searches for variability of, 591
 ultraviolet excess technique, 588–589
- discovery of, 724
 evolution of, with cosmic epoch, 591, 724
 luminosity function of
 ‘luminosity evolution’ of, 724
 evolution of, 591, 593, 724, 725
- radio waveband, 17–21
 neutral hydrogen and molecular line astronomy, 20–21
 observing the sky in, 21
 origin of high energy astrophysics and, 17–20
- Radiocarbon*, 295
- radiocarbon dating, 294–297
 calibration of, 295, 296
 impact of nuclear test explosions, 297
 influence of Earth's magnetic field, 296
 using coral samples, 296
 using tree ring dating, 296
 detection of supernovae in tree-ring data, 297
- radionuclides, 279
- radius of curvature
 instantaneous, 201, 203, 204, 207
- range R of high energy particles, 143
 in different materials, 144

- rapid or r-process, 65, 388
 formation of elements beyond the iron peak and, 388
 proto-neutron stars and, 388
 ratio of specific heat capacities
 for non-relativistic degenerate gas, 395, 396
 Rayleigh criterion for resolving two point sources, 768
 Rayleigh scattering, 349, 350
 Rayleigh–Taylor instability, 383
 Rayleigh–Jeans limit, 218
 Rayleigh–Jeans region of Planck spectrum, 173
 Rayleigh–Jeans spectrum, 459, 460, 710
 recoil effect, 236, 244, 255
 reconnection of magnetic lines of force, 320,
 323–329
 diffusive time-scale, 326
 formation of current filaments by tearing mode
 instabilities, 328
 ohmic losses, 324
 Petschek mechanism, 327
 reconnection velocity, 326, 327
 self-consistency of models for, 329
 Sweet–Parker mechanism for, 324–326
 rectangular one-dimensional aperture
 diffraction pattern of, 766
 red and blue sequences of galaxies, 80–85
 colour and absolute magnitude, 81
 effect of galaxy environment, 84–85
 mean stellar age and concentration index C , 85
 Sérsic index and colour, 81–84
 statistics of galaxies belonging to, 85
 red sequence, 80
 preferentially found in rich cluster environments, 84
 properties of galaxies of, 80
 Red-Sequence Cluster Survey, 122
 redshift
 gravitational, 431
 reduced mass of molecule, 337
 reference frames in standard configuration, 136
 Reissner–Nordstrøm metric, 434
 relativistic aberration, 198
 formulae, 199, 200
 relativistic beaming
 synchrotron radiation and, 198–201, 205, 210
 relativistic invariants, 154–155
 total energy loss rate, 154, 155, 161, 174, 195
 relativistic jets
 heating of the interstellar gas by, 750
 relativistic kinetic energy, 782
 relativistic length contraction, 166
 relativistic Maxwellian distribution, 218
 relativistic plasma, 18
 relativistic three-momentum, 180, 183, 782
 relativistic total energy, 782
 relativistic transformation of an inverse square law
 Coulomb field, 136–139
- Relativity: Special, General and Cosmological*
 (Rindler), 781
 relaxation time of particles in a plasma, 302–303
 retarded potentials, 267
 retarded time, 160, 203
 Reuven Ramaty High Energy Solar Spectroscopic
 Imager (RHESSI), 289, 291
 Reynolds number \mathcal{R} , 453, 454
 right ascension (RA or α), 753
 rigidity R , 180, 185, 186
 Rindler, Wolfgang, 433
 ring galaxies, 80
 Roche lobe, 461, 464, 465, 466, 467, 468, 480, 481
 overflow, 467, 473, 485, 486
 Roche lobe overflow, 72
 ROSAT satellite
 ultraviolet Wide Field Camera of, 357
 ROSAT Wide Field Camera, 22
 ROSAT X-Ray Observatory, 24, 25, 112, 422, 714,
 725
 Rosseland Mean Opacity, 47
 Rosseland mean opacity, 460
 Rossi X-ray Timing Explorer (RXTE), 418, 449, 483,
 486, 489, 653
 rotation curves of spiral galaxies, 92–93
 constancy at large radial distances, 92
 rotation measure, 346, 369–371
 as a function of galactic latitude, 369, 370
 pulsar, 370, 371
 RR-Lyrae variable stars, 73
 masses of, 73
 Sérsic index n , 82, 83, 84, 85
 SAGE solar neutrino experiment, 32
 Sagittarius A* (Sgr A*), 754
 supermassive black hole in, 754
 Salpeter initial mass function, 361, 362, 731, 743, 744
 SAS-2 γ -ray satellite, 504
 SAS-2 satellite, 26
 scalar potential ϕ , 158
 scale height of an atmosphere, 322
 scattering of charged particles by irregularities in
 magnetic field, 184–187
 scattering of high energy particles by Alfvén and
 hydromagnetic waves, 187–189
 Schönberg–Chandrasekhar limit, 49, 59
 Schönberg–Chandrasekhar limit, 731
 Schechter luminosity function for galaxies
 turnover luminosity L^* , 747
 Schmidt law of star formation, *see* Schmidt-Kennicutt
 law
 Schwarzschild black hole, 490
 Schwarzschild metric, 430–433, 434, 435
 Schwarzschild radius, 19, 413, 429, 431, 433, 435,
 444, 446, 448, 610, 611, 612, 621, 623, 629,
 681, 704

- coordinate singularity in, 433
 physical singularity at $r = 0$, 433
 X-ray emission from the vicinity of, 25
 SCUBA submillimetre bolometer array, 14, 742
 seeing disc, 770
 variation in size with wavelength, 770
 self-collision time of particles in a plasma, 302
 semi-analytic models of galaxy formation, 749–752
 as ‘experimental computational astrophysics’, 749
 dust extinction and, 750
 formation of massive black holes in the nuclei of galaxies and, 750
 formation of supermassive black holes and, 752
 in centres of rich clusters, 752
 heating of the intergalactic gas in clusters and, 750
 nuclear starbursts and, 750
 objectives of, 749
 quasars at large redshifts and, 751
 radiative cooling and star formation, 750
 spectrophotometric properties of galaxies and, 750
 supernova explosions and, 750
 the epoch of maximum quasar activity and, 752
 two-point correlation function for galaxies and
 for different luminosities and colours, 752
 semi-forbidden transitions in spectra of gaseous nebulae, 342–343
 example of C III], 342
 intercombination lines, 342
 sensitivities of astronomical detectors, 774–779
 in the photon limit, 774–777
 in the wave limit, 777–779
 increasing signal-to-noise by increasing bandwidth and integration time, 778
 optical and infrared detectors, 774–777
 radio and millimetre-wave receivers, 777–779
 Seyfert galaxies, 592, 724
 counts of, 591, 724
 Shapiro time delay, 416
 shear stress, 452, 453
 shock waves, 314–319, 357
 basic properties of plane, 314–317
 collisionless, 321
 conservation relations for, 314
 energy flux conservation, 315
 mass conservation, 314, 315
 momentum flux conservation, 315
 explosions and, 314
 heating of gas to high temperatures and, 317
 Mach number, 316
 oblique, 319
 passage of streamlines through, 319
 role of atomic or molecular viscosities, 317
 shock conditions, 316
 speed of sound and, 314, 315
 stand-off distance, 319
 strong, 317
 supersonic piston, 317–319
 SI (Système International) system of units, 129
 sidelobes, 768
 signal-to-noise ratio
 background-limited, 776, 777
 dependence on size of telescope for different limiting cases, 776
 detector-noise-limited, 776, 777
 in the presence of other sources of noise, 775, 776
 background signal from sky and telescope, 775
 dark current in the detector, 776
 read-out noise, 776
 photon statistics and, 775
 photon statistics limit, 776, 777
 signature of metric, 781
 silicate dust grains
 complex dielectric constant for, 348
 silicon burning, 64
 single line spectroscopic binaries, 439
 single-line spectroscopic binaries, 414
 singularity theorems of Penrose and Hawking, 433
 sky in different astronomical wavebands, 4–5
 Sloan Digital Sky Survey (SDSS), 4, 7, 77, 80, 81, 82,
 84, 85, 88, 97, 98, 100, 590, 591, 592, 714,
 724, 725
 slow or s-process, 65
 Smithsonian Sub-Millimetre Array (SMA), 14
 Smoluchowski’s envelope, 103
 soft γ -ray repeaters, 418–419
 soft X-ray emission from the Galactic plane, 359
 Solar and Heliospheric Observatory (SOHO) of ESA,
 52, 53
 solar flares, 322
 γ -ray spectrum of, 289, 291
 solar luminosity, 36
Solar Magnetohydrodynamics (Priest), 322
 solar mass, 36, 759
 Solar Maximum Mission, 391
 solar modulation, 185, 187, 225, 358, 502, 505
 cosmic ray energy spectrum and, 493
 diffusive-convective model for, 493
 electron energy spectrum and, 495
 solar neutrino experiment, 54
 solar neutrino problem, 55
 solar neutrinos, 54–58
 and neutrino oscillations, 57, 58
 chlorine detector for, 54, 55
 GALLEX solar neutrino experiment, 56
 gallium experiments, 56
 Kamiokande II experiment, 55
 SAGE experiment, 56
 Sudbury Neutrino Observatory (SNO), Ontario,
 Canada, 57, 58, 59
 SuperKamiokande experiment, 56
 solar radius, 36
 solar seismology, *see* helioseismology

- Solar System abundances of the elements, 496–497, 499
 local interstellar abundances, 497, 498
 meteoritic abundances, 497
- Solar Wind, 19, 70, 303, 312–314, 319, 426, 493
 high-speed winds observed in, 312, 313
 magnetic field in, 312
 magnetosphere and the, 319–321
 drag forces associated with, 320
 rotating garden sprinkler and trajectories of particles in, 312
 spiral configuration of magnetic field in, 313
 out to 20–25 AU, 312
 typical properties of, 312
- Soudan Cryogenic Dark Matter Search (CDMS), 34
- source function, 222
- south celestial pole (SCP), 753, 754
- South Pole Observatory, 14
- Soviet atomic and hydrogen bomb programme, 250
- spallation, 150–151, 279, 495, 499, 500, 507
 formation of L from M elements, 508
 formation of radioactive isotopes in meteorites and, 150, 151
 formation of rare isotopes in meteorites and, 150
 path length distribution for, 509–511
 characteristic escape time τ_e , 510
 dependence upon cosmic ray energy, 513, 514
 exponential, 510
 Gaussian, 510
 isotopic anomalies, 515
 leaky box model for, 510
 truncated exponential, 510
 path length through interstellar gas for, 507
 production of ^3He from ^4He , 509
 products of, of iron nuclei, 509
 radioactive products, 500
 slab model for, 508, 510
- spallation cross-sections, 282–283, 507
 dependence upon particle energy, 283
 for iron nuclei, 283
 features of, 283
 methods of determining
 experimental, 282
 Monte Carlo techniques, 282
 semi-empirical relations, 282
- partial, 282–283, 507
- total inelastic, 283
 weighted average total inelastic, 508, 509
- spallation fragments, 280
 relativistic, 282
- special relativity
 notation, 781–782
- specific angular momentum, 431, 433
- specific angular momentum in general relativity, 434
- specific heat capacities γ_{SH}
 ratio of, for perfect gas, 218
 ratio of, for relativistic gas, 218
- specific heat capacity per unit mass, 308
- specific volume, 315
- speckles, 771
- spectral energy distribution of galaxies
 evolution of, 729–731
 in far-infrared and submillimetre wavebands, 740, 741
- spectral index, 212, 217
- spectrum of radiation of an arbitrarily moving electron, 202–203
- spin temperature for 21-cm absorption, 335
- spinning dust grains
 radio emission of, 374–376
- spiral arm tracers in the Galaxy, 356
- spiral galaxies
 dark matter haloes, 93
- Spitzer infrared space observatory, 714
- Spitzer infrared space telescope, 737
- Spitzer Space Telescope, 10, 11
- spontaneous and induced processes, 247, 249
 rules for, 247, 249
- stability of disc galaxies and dark matter haloes, 93
- stand-off distance, 319, 320
- stand-off shock waves, 314, 319
- standard wavebands used in ground-based astronomy, 761, 762
- star clusters
 ages of, 68
 open, 69
- star formation, 75, 361–369
 angular momentum problem, 367
 energy problem, 367
 initial mass function, 361–362
 issues in the theory of, 367–369
 Jeans' instability, 368
 magnetic field problem, 368
 model for, 353, 355
 pre-main sequence evolutionary tracks, 70
 pre-main sequence stars and, 353, 354
 regions of, 362–367
 Schmidt–Kennicutt law, 361–362
- star formation in molecular clouds
 radiation of heated dust grains at infrared wavelengths as a signature of, 352
 role of interstellar dust grains, 351
- star-forming galaxies, 352
 spectrum of, 737
- starburst galaxies, 740
 ultraviolet spectra of, 743, 744
- starquakes, 423
- stars and stellar evolution, 35–76
 basic observations of, 35–39

- equations of energy generation and energy transport, 43–47
 equations of hydrostatic support and mass conservation, 40–41
 equations of stellar structure, 47–50
 evolution of high and low mass stars, 59–68
 high mass stars, 62–65
 low mass stars, 62, 66–68
 Hayashi tracks, 59–61
 helioseismology and the internal structure of the Sun, 51–54
 importance of mass loss, 71, 72
 mass loss, 70–75
 horizontal branch, 73
 overall mass loss rate, 75
 P-Cygni profiles and Wolf–Rayet stars, 70–73
 planetary nebulae, 73–74
 observations of solar neutrinos, 54–58
 stellar evolution on the colour–magnitude diagram, 68–70
 stellar structure, 39–43
 sun as a star, 50–58
 virial theorem for stars, 42–43
 statistical equation of motion for mean drift velocity of electrons in electric field, 303
 statistical equilibrium, 91, 759
 statistical weights, 248
 statistical weights g_1, g_2 , 171
 Stefan–Boltzmann law, 37
 Stefan–Boltzmann law, 35, 46, 459, 758
 stellar coronae, 70
 stellar evolution
 formation of isothermal core, 59
 growth of central helium core, 62
 hydrogen-shell burning, 59
 lower mass limit of $0.08 M_{\odot}$ for hydrogen burning stars, 66
 of a $1.3 M_{\odot}$ star, 49
 of a $15 M_{\odot}$ star, 384
 core collapse of, 385
 formation of iron core, 385
 neutrino luminosity of, 385
 of a $5 M_{\odot}$ star, 62, 63
Stellar Magnetism (Mestel), 424
 stellar structure
 equations of, 39–50
 stellar winds, 70
 stellar X-ray sources in Magellanic Clouds
 X-ray luminosities of, 446, 447
 Stokes' theorem, 324
 stopping power, 142
 straggling, 144
 structural index, or structural length, 104
 sub-dwarf (sdB) stars, 589
 sub-dwarf stars, 38
 sub-giant stars, 38
 Subaru 8-metre telescope, 744
 Subaru 8.2 metre telescope, 382
 Subaru Deep Surveys, 744
 submillimetre background radiation, 743
 Sudbury Neutrino Observatory (SNO), 32
 Sun
 central pressure of, 41
 corona of, 312
 main sequence lifetime of, 50
 minimum central temperature of, 42
 origin of magnetic field of, 54
 properties of, 36
 rotation of, 312
 speed of sound as a function of radius, 52–54
 sunspots, 322
 Sunyaev–Zeldovich effect, 257–259
 hot gas in clusters of galaxies and, 259
 Hubble's constant and, 259
 in Rayleigh–Jeans region of the spectrum, 258
 increase in energy density of the background radiation, 258
 physical nature of result in Rayleigh–Jeans spectral region, 259
 Sunyaev–Zeldovich effect in hot intracluster gas, 99, 114–115, 758
 distance estimates using, 115
 in the Rayleigh–Jeans limit, 115
 shape of spectral distortions, 259
 spectral shape of distortions, 115
 spectral signature of, 115
 SuZIE experiment and measurement of shape of spectral distortions, 259
 SuZIE experiment and measurement of spectral shape of distortions, 115
 super-Alfvénic motions, 321
 superCDMS experiment, 127
 superclustering of galaxies, 101
 superconductivity and superfluidity in neutron stars, 404–405
 supergiant stars, 38
 superluminal motions
 in microquasars, 440
 supernova light curve
 radioactive origin of, 288
 supernova remnants, 379, 394
 as X-ray sources, 394
 evolution of, 554–556
 supernova remnants as sources of high energy electrons, 553
 supernova remnants as sources of high energy particles, 545–548
 γ -ray observations of shell-like, 547–548
 radio observations of, 545
 radio observations of shell-like, 547
 supernova SN 1987A, 388–393
 ^{44}Ti energy source for light curve, 289, 392

- supernova SN 1987A, (*cont.*)
 ^{56}Co γ -ray emission lines from, 288, 391
 ^{57}Co energy source for light curve, 392
 $[\text{Co II}]$ and $[\text{Ni II}]$ infrared emission lines from, 391, 392
collision of envelope with mass-loss ring, 392, 393
observed at optical, X-ray and radio wavelengths, 393
distance estimate from $[\text{O III}]$ ring, 392
dust formation in, 392
light curve of, 388, 389, 390, 391, 392
limits to neutrino mass, 390
neutrino luminosity of, 390
neutrinos from, 279
progenitor of, 389
rings of $[\text{O III}]$ about, 392, 393
stellar mass-loss prior to, 389, 392, 393
Type IIP, 388
supernova units (SNu), 381
supernovae, 75, 378–394
as sources of high energy particles, 394
classification of, 379
core-collapse, 384–386
bounce mechanism for, 386
formation of neutron stars and black holes and, 384–386
inhomogeneous expulsion of outer layers, 386
kinetic energy release, 385
neutrinos from, 390
historical, 378–380
kinetic energy of ejected material, 394
light curves, 380
rates of occurrence, 382
Type Ia, 380–384, 466
Type II, 467
Types I and II, 379
differences between, 380
typology, 378–380
supersonic motion, *see* shock waves
supersonic piston, 317–319
location of shock front relative to, 318
supernovae and the, 319
Sutherland, Peter, 428
Sweet–Parker mechanism of magnetic reconnection, 324–326
dissipation rate of, 324, 326
effects of gas pressure, 325
energy released in, 326
SWIFT satellite, 418, 706, 707, 708, 709, 710, 712
SWIRE Legacy sample, 740
symbiotic star, 383
symbiotic stars, 461
synchro–Compton catastrophe, 261–264
synchrotron radiation, 193–227, 377, 410, 489, 715
as the relativistic limit of cyclotron radiation, 197
asymptotic expression for emissivity at high and low frequencies, 210
critical angular frequency for, 207, 208
critical frequency for, 201, 211, 212, 218, 223
high frequency limit, 210–211
map of Galactic, 18
minimum energy requirements for, 549–552
polarisation of, 214–217
circular, 217
fractional, for power-law electron energy spectrum, 216–217
spectra of orthogonal polarisations I_{\perp} and I_{\parallel} , 214
system of coordinates for, 204
total energy in polarisations I_{\perp} and I_{\parallel} , 215
radio emission of the Galaxy and, 224–227
relation between spectral index α and spectral index of electron spectrum p , 212
self-absorption of, 217–222
absorption coefficient for, 219–222
physical arguments, 217–219
spectral emissivity of
in orthogonal polarisations, 208
maximum of, 209
total, 208
spectrum of a power law distribution of electron energies, 212–214
emissivity per unit volume, 213
full analysis, 213
physical arguments, 212
spectrum of, detailed analysis, 202–211
algebra of, 204–207
of radiation of an arbitrarily moving electron, 202–203
results of, 207–211
system of coordinates for, 203–204
spectrum of, physical arguments, 198–202
spontaneous transition probability for, 219
total energy loss rate, 193–195, 210
useful numerical results, 222–224
absorption coefficient χ_v for a random magnetic field, 224
critical frequency v_c , 223
emission spectrum of a single electron, 222
radiation spectrum of a power-law electron energy distribution, 223
total energy loss rate, 222
synchrotron radiation as the scattering of virtual photons, 241
synchrotron radiation facilities, 209
synchrotron self-absorption, 217–222, 262, 263
absorption coefficient for, 219–222
for a randomly oriented magnetic field, 221
two-level system for, 219
evidence for presence of relativistic electrons, 219
evidence for relativistic electron in compact radio sources, 263

- physical arguments, 217–219
 polarisation of synchrotron self-absorbed source, 222
 spectrum of radio source exhibiting, 219
 synchrotron–self-Compton radiation, 260–264
 models of sources of, 264, 265
 homogeneous, 264
 inhomogeneous, 264
 synchro–Compton catastrophe, 261–264
Système International d'Unités (SI units), 779–781
- TAMA gravitational wave experiment, 33
 tearing mode instabilities, 327
 technetium Tc, 65
 telluric absorption, 10
 temperature-frequency relation for black-body radiation, 6
 term diagram for doubly ionised oxygen O III, 342, 343
The Physics of Solar Flares (Tandberg-Hanssen and Emslie), 322
Theoretical Concepts in Physics (Longair), xiii, 160, 429, 777, 778, 782
 thermal bremsstrahlung, *see* bremsstrahlung, thermal, 340, 477
 absorption in plane of Galaxy, 340
 absorption, at radio wavelengths, 340
 at soft X-ray energies from hot component of interstellar gas, 340
 at X-ray energies from intracluster gas, 340
 and FeXXV emission lines, 340
 at X-ray energies from supernova remnants and FeXXV emission lines, 340
 thermal energy per unit mass, 42
 thermal instabilities, 360–361
 thermal paradox for stars, 43
 thermal time-scale for stars, 43
 thermalisation time for stellar systems, 93
 thermalisation time of particles in a plasma, 302–303
 thermonuclear runaway, 473, 478, 482
 stabilisation at about 10^8 K, 479
 thick accretion discs, 461–463
 funnels along polar directions, 462, 463
 Newtonian rotationally supported vorticity-free torus, 462, 463
 stability of, 462
 thin accretion discs, 451–461
 about black holes, 458
 inner boundary condition, 458
 conditions for, 451–452
 continuum spectra of, 476
 detailed models of, 459–461
 Doppler tomography of, 475, 476
 emission spectra of, 458–459
 in black-body approximation, 458
 emission spectrum of, 459, 460, 462
 emission spectrum of, in optically thin region, 461
 formation of hot spot, 467, 474, 475
 instabilities in, 478
 opacity of material of, 460
 role of viscosity – the α parameter, 452–455
 structure of thin discs, 455–458
 temperature distribution in, 458–459
 transition from dominance of gas pressure to radiation pressure, 459, 460
- Third Reference Catalogue of Bright Galaxies* (de Vaucouleurs *et al.*), 86
- Thomson and Compton scattering, 231–237
 Thomson cross-section, 46, 229, 446
 Thomson scattering, 46, 48, 232–445
 degree of polarisation of, 234
 derivation of formulae for, 232–235
 differential cross-section for, 233, 241, 250
 Eddington limiting luminosity and, 413
 for 100% polarised emission, 234
 geometry of, 232
 mean free path for, 235
 optical depth for, 235, 244, 253, 254, 256
 polarisation of, 234
 properties of, 233–235
 total scattering cross-section for, 233
- Thomson scattering cross-section σ_T , 155, 195
- Thorne, Kip, 434
- 3CR radio galaxies, 654, 655, 720, 748
 early formation of bulk of stellar population, 748
 stellar masses of large redshift, 748
- 3CR radio sources, 720, 723
- threshold detectors, 265
- tidal radius of cluster of galaxies, 104
- tidal radius of galaxies, 122
- time dilation formula
 in general relativity, 431
- time to reach a given signal-to-noise ratio, 777
- tokamaks, 323
- torque, magnetic, 470, 471, 473
- torque, viscous, 453, 455, 470
- total depth of the atmosphere in kg m^{-2} , 176
- transfer equation for radiation, 170, 171, 221
- in terms of Einstein coefficients for spontaneous and induced emission, 249
- in terms of occupation numbers, 249
- Transition Region and Coronal Explorer (TRACE)
 spacecraft of NASA, 309
- transmission function $G(z)$ of one-dimensional aperture, 765
 Fourier transform of, 765
 Gaussian, 768
- transmission function T_ν of astronomical filters, 762
- transparency of the atmosphere as a function of wavelength, 5, 6
- in the infrared and submillimetre wavebands, 10, 14
- windows in the submillimetre waveband, 14

- tree-ring dating, 295
 triple- α process, 64
 Tully–Fisher relation for spiral galaxies, 87
 and the distances of galaxies, 87
 infrared, 87
 turbulence
 magnetohydrodynamic, 455
 21-cm line emission and absorption, 333–335
 absorption coefficient for, 334–335
 column density of neutral hydrogen from, 334
 emissivity of, 334
 kinematics of gas in galaxies from, 334
 spin temperature for, 335
 spontaneous transition probability of, 333
 2dF galaxy redshift survey, 589, 714
 2dF quasar redshift survey, 589, 590, 591, 592, 593,
 724, 725
 Two Micron All Sky Survey (2MASS), 11, 98
 two-colour diagram for stars, 763, 764
 Type Ia supernovae, 380–384, 479
 as standard candles, 381
 evolution of colours of, 382
 formation of ^{56}Ni , 382
 light curves
 radioactive decay of ^{56}Co , 382, 383
 light curves of, 380
 luminosity–width relation, 380
 most luminous supernovae, 381
 rates of occurrence, 381
 redshift–distance relation and, 381
 thermonuclear explosions and, 380, 382
 type Ia supernovae, 478–479
- UHURU X-ray Observatory, 23, 24, 725
 UHURU X-ray observatory, 410, 485, 600
 UK Schmidt Telescope, 589
 ultra-high energy γ -ray telescopes, 275
 ultrahigh energy γ -rays
 detection of, 272–275
 ultraluminous infrared galaxies (ULIRGs), 741
 ultraviolet waveband, 21–22
 extreme ultraviolet (EUV), 22
 observations of resonance lines of common
 elements, 22
 Ulysses mission of the European Space Agency and
 NASA, 312, 313
 unstable stars, 73
 useful statistics about galaxies, 97–98
- V/V_{\max} or luminosity–volume test, 718–719
 banded, 719
 for radio quiet quasars, 724
 space distribution of galaxies, quasars active
 galaxies, 718–719
 van den Heuvel, Edward, 411
 variance, 775
 of sum of two quantities, 775
 total, as the sum of variances, 776
 vector potential A , 158, 160, 183, 210
 Vela satellites, 27
 velocity cone, 214, 215, 217
 velocity of Solar System through the Cosmic
 Microwave Background Radiation, 16
 velocity–distance relation for galaxies, 757
 vernal equinox, 753
 Very Large Array (VLA), 21, 123, 774
 Very Long Baseline Interferometry (VLBI), 21, 219,
 263
 vicinity of the black hole, 637–660
 violent interstellar medium, 359
 supernova explosions and, 359
 VIRGO gravitational wave experiment, 33
 virial theorem, 103, 400
 virial theorem for galaxies and clusters, 89–92, 93, 94,
 759
 problems of application to observed systems, 91, 92
 virial theorem for stars, 42–43, 89
 viscosity, 453, 454, 459, 656
 effective, in shock fronts, 574
 kinematic, 445, 657
 turbulent, 462, 469
 viscosity parameter α , 459, 461, 478, 660
 viscosity, dynamic, 455, 457
 viscosity, dynamic or shear η , 452, 454
 viscosity, kinematic v , 453
 viscosity, turbulent, 454, 455, 459
 visible matter in the Universe
 integrated luminosity of, 97
 VLT Antu telescope, ISAAC infrared camera of, 365
 VLT Kueyen Telescope, FORS2 instrument of, 365,
 366
 voids in the distribution of galaxies, 731
 Voyager missions of NASA, 312
- W UMa-type binaries, 464
 wave impedance Z_0 , 777
 wavebands used in ground-based astronomy, 761
 waveform errors, 768, 770
 Wheeler, John, 434
 white dwarf, 310
 diffusion time for magnetic field from, 310, 311
 magnetic flux density of, 310
 white dwarf stars, 37, 38, 68, 74, 76
 white dwarfs, 378, 401, 443, 589
 carbon-oxygen, 382
 cooling curves for, 402
 cooling times for, 401
 Hertzsprung–Russell diagram for, 402
 ultimate fate of, 401
 with carbon-oxygen cores, 401
 white dwarfs, neutron stars and the Chandrasekhar
 limit, 394–401

- Wien distribution, 246, 247
 average energy of photons of, 246
 Wien's displacement law, 5
 Wien's law, 255
 Wilkinson Microwave Anisotropy Probe (WMAP), 16
 WIMPs
 astrophysical limits to the masses of, 126
 candidates for, 125
 laboratory detection of, 127
 laboratory limits to the masses of, 126
 suppression mechanisms for, 127
 Wolf–Rayet star, 466
 Wolf–Rayet stars
 mass loss rates of, 72
 Wolf–Rayet star, 466, 467, 515, 713
 Wolf–Rayet stars, 70–73, 288
 WC and WN types, 72
 W^\pm and Z^0 bosons, 126

 X mass fraction of hydrogen, 69
 X-ray absorption, 339–340
 X-ray absorption coefficient of interstellar matter,
 229, 231
 column depth for, 230
 optical depth and, 230
 X-ray atomic energy levels, 230
 X-ray background at soft X-ray energies, 24, 25
 X-ray background emission
 at hard X-ray energies
 nature of sources contributing to, 602, 729
 at soft X-ray energies
 nature of sources contributing to, 728
 contributions of different classes of source to,
 728
 spectrum of, 728
 X-ray binaries, 467
 X-ray binary systems, 410–415, 461, 479–491
 accretion in, 413
 estimate of mass of neutron stars in, 414–415
 estimate of masses of neutron stars and black holes
 in, 415
 high mass X-ray binaries, 485–486
 high-mass X-ray binaries, 413
 late O or early B type companion stars, 413, 414

 low mass X-ray binaries, 479–484
 Galactic bulge sources, 479
 in globular clusters, 479
 low-mass X-ray binaries, 413
 in globular clusters, 414
 luminosities of, 413
 soft X-ray transients, 414
 symbiotic X-ray binaries, 414
 typical temperature of, 413
 X-ray bursters, 414
 X-ray pulsars, 414
 X-ray emission of intracluster gas, 99
 X-ray nova, 487
 X-ray pulsars, 24, 438
 X-ray source spectra and solutions of the Kompaneets
 equation, 253–257
 X-ray sources
 binary pulsating
 as rotating magnetised neutron stars, 446
 Comptonisation and, 602, 729
 evolution of luminosity function of, with cosmic
 epoch, 727
 populations at soft and hard X-ray energies, 602,
 728
 strongly absorbed, 602, 729
 template spectra for strongly absorbed sources, 602,
 729
 X-ray waveband, 22–25
 detectors for, 23
 observing in, 22–23
 sky in, 23–25
 XMM-Newton X-Ray Observatory, 25, 108, 110, 113,
 197, 257, 386, 421, 628, 629, 633, 714, 725

 Y mass fraction of helium, 69
 year 2000.0 coordinate system, 753, 754
 Yerkes spectral classification system, 37

 Z mass fraction of ‘metals’, 69
 Zeeman splitting
 of 21-cm line radiation, 377
 circularly polarised components of, 377
 of OH absorption lines, 377
 zenith equidistant projection, 7, 755, 756

