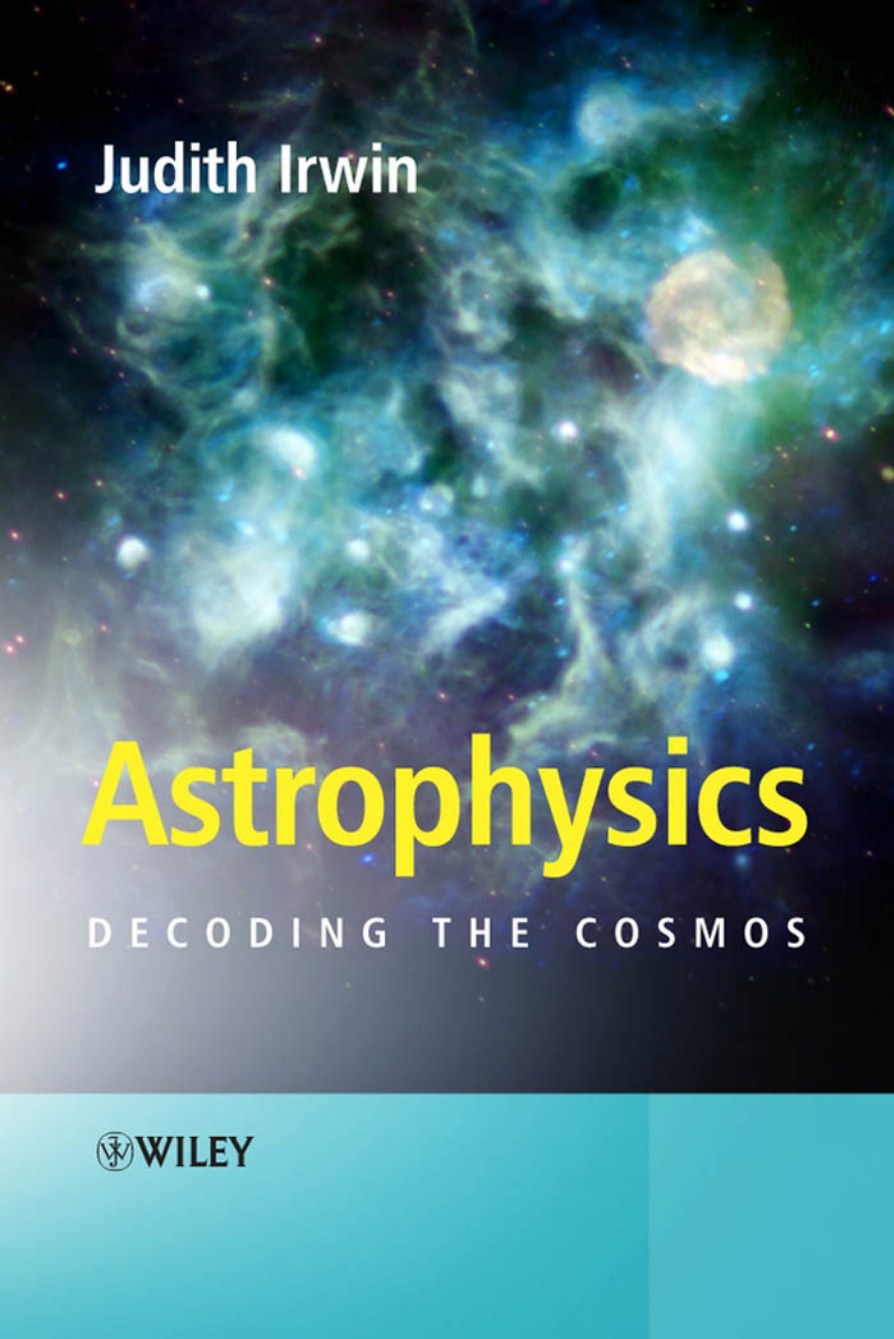

Judith Irwin
Astrophysics
DECODING THE COSMOS
WILEY

Astrophysics

Decoding the Cosmos

Judith A. Irwin

Queen's University, Kingston, Canada

John Wiley & Sons, Ltd

Astrophysics

Astrophysics

Decoding the Cosmos

Judith A. Irwin

Queen's University, Kingston, Canada

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data

Irwin, Judith A., 1954-
Astrophysics : decoding the cosmos / Judith A. Irwin.
p. cm.
ISBN 978-0-470-01305-2 (cloth)
1. Astrophysics. I. Title.
QB461.I79 2007
523.01–dc22 2007009980

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-470-01305-2 (HB)

ISBN 978-0-470-01306-9 (PB)

Typeset in 10.5/12.5 pt Times by Thomson Digital, India
Printed and bound in Great Britain by Antony Rowe Ltd., Chippenham, Wilts
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production

To Beth, Bob, Marg & Don

Contents

Preface

Like many textbooks, this one originated from lectures delivered over a number of years to undergraduate students at my home institution – Queen's University in Kingston, Canada. These students had already taken a first year (or two) of physics and one introductory astronomy course. Thus, this book is aimed at an intermediate level and is meant to be a stepping stone to more sophisticated and focussed courses, such as stellar structure, physics of the interstellar medium, cosmology, or others. The text may also be of some help to beginning graduate students with little background in astronomy or those who would like to see how physics is applied, in a practical way, to astronomical objects.

The astronomy prerequisite is helpful, but perhaps not required for students at a more senior level, since I make few assumptions as to prior knowledge of astronomy. I *do* assume that students have some familiarity with celestial coordinate systems (e.g. Right Ascension and Declination or others), although it is not necessary to know the details of such systems to understand the material in this text. I also do not provide any explanation as to how astronomical distances are obtained. Distances are simply assumed to be known or not known, as the case may be. I provide some figures that are meant to help with 'astronomical geography', but a basic knowledge of astronomical scales would also be an asset, such as understanding that the Solar System is tiny in comparison to the Galaxy and rotates about the Galactic centre.

As for approach, I had several goals in mind while organizing this material. First of all, I did not want to make the book too 'object-oriented'. That is, I did not want to write a great deal of descriptive material about specific astronomical objects. For one thing, astronomy is such a fast-paced field that these descriptions could easily and quickly become out of date. And for another, in the age of the internet, it is very easy for students to quickly download any number of descriptions of various astronomical objects at their leisure. What is more difficult is finding the thread of physics that links these objects, and it is this that I wanted to address.

Another goal was to keep the book practical, focussing on how we obtain information about our Universe from the signal that we actually detect. In the process, many equations are presented. While this might be a little intimidating to some students, the point should be made that the equations are our 'tools of the trade'. Without these tools, we would be quite helpless, but with them, we have access to the secrets that astronomical signals bring to us. With the increasing availability of computer algebra or other software, there is no longer any need to be encumbered by mathematics. Nevertheless, I have kept problems that require computer-based solutions to a minimum in this text, and have tried to include problems over a range of difficulty.

Astrophysics – Decoding the Cosmos will maintain a website at http://decoding.phy.queensu.ca. A Solutions manual to the problems is also available. I invite readers to visit the website and submit some problems of their own so that these can be shared with others. It is my sincere desire that this book will be a useful stepping-stone for students of astrophysics and, more importantly, that it may play a small part in illuminating this most remarkable and marvelous Universe that we live in.

Judith A. Irwin
Kingston, Ontario
October 2006

Acknowledgments

There are some important people that I feel honoured to thank for their help, patience, and critical assistance with this book. First of all, to the many people who generously allowed me to use their images and diagrams, I am very grateful. Astronomy is a visual science and the impact of these images cannot be overstated.

Thanks to the students, past and present, of Physics 315 for their questions and suggestions. I have more than once had to make corrections as a result of these queries and appreciate the keen and lively intelligence that these students have shown.

Thanks to Jeff Ross for his assistance with some of the 'nuts and bolts' of the references. And special thanks to Kris Marble and Aimee Burrows for their many contributions, working steadfastly through problems, and offering scientific expertise.

With much gratitude, I wish to acknowledge those individuals who have read sections or chapters of this book and offered constructive criticism – Terry Bridges, Diego Casadei, Mark Chen, Roland Diehl, Martin Duncan, David Gray, Jim Peebles, Ernie Seaquist, David Turner and Larry Widrow.

To my dear children, Alex and Irene, thank you for your understanding and good cheer when mom was working behind a closed door yet again, and also thanks to the encouragement of friends – Joanne, Wendy, and Carolyn.

My tenderest thanks to my husband, Richard Henriksen, who not only suggested the title to this book, but also read through and critiqued the *entire* manuscript. For patience, endurance, and gentle encouragement, I thank you. It would not have been accomplished without you.

J.A.I.

Introduction

Knowledge of our Universe has grown explosively over the past few decades, with discoveries of cometary objects in the far reaches of our Solar System, new-found planets around other stars, detections of powerful gamma ray bursts, galaxies in the process of formation in the infant Universe, and evidence of a mysterious force that appears to be accelerating the expansion of the Universe. From exotic black holes to the microwave background, the modern understanding of our larger cosmological home could barely be imagined just a generation ago. Headlines exclaim astonishing properties for astronomical objects – stars with densities equivalent to the mass of the sun compressed to the size of a city, million degree gas temperatures, energy sources of incredible power, and luminosities as great as an entire galaxy from a single dying star.

How could we possibly have reached these conclusions? How can we dare to describe objects so inconceivably distant that the only influence they have on our lives is through our very astonishment at their existence? With the exception of a few footprints on the Moon, no human being has ever ventured beyond the confines of the Earth. No spaceprobe has ever approached a star other than the Sun, let alone returned with material evidence. Yet we continue to amass information about our Universe and, indeed, to believe it. How?

In contrast to our attempts to reach into the Universe with space probes and radio beacons, the natural Universe itself has been continuously and quite effectively reaching us. The Earth has been bombarded with astronomical information in its various forms and in its own language. The forms that we think we understand are those of *matter* and *radiation*. Our challenge, in the absence of an ability to travel amongst the stars, is to find the best ways to detect and decipher such communications.

What astronomical matter is reaching us? The high energy subatomic particles and nuclei which make up *cosmic rays* that continuously bombard the earth (Sect. 3.6) as well as an influx of meteoritic dust, occasional meteorites and those rare objects that

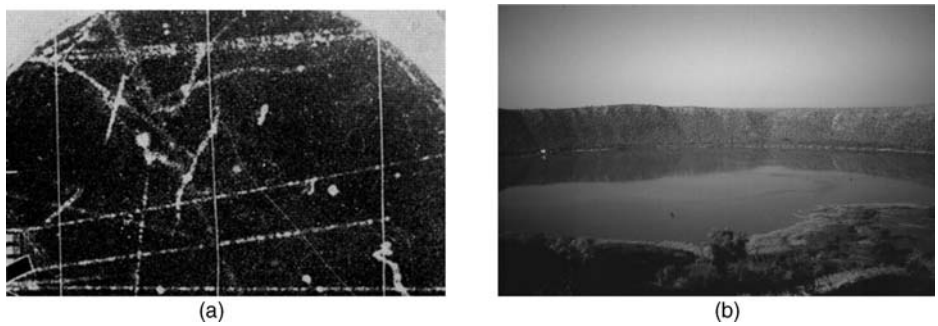


Figure I.1 (**a**) Early photograph of cosmic ray tracks in a cloud chamber. (Ref. [30]) (**b**) The 1.8 km diameter Lonar meteorite crater in India

are large enough to create impact craters. Such incoming matter provides us with information on a variety of astronomical sources, including our Sun, our Solar System, supernovae in the Galaxy[1], and other more mysterious sources of the highest energy cosmic rays. The striking contrast in size and effect between such particulate matter is illustrated in Figure I.1.

Radiation refers to *electromagnetic radiation* (or just 'light') over all wavebands from the radio to the gamma-ray part of the spectrum (Table G.6). Electromagnetic radiation can be described as a wave and identified by its wavelength, λ or its frequency, ν. However, it can also be thought of as a massless particle called a *photon* which has a particular energy, $E_{\rm ph}$. This energy can be expressed in terms of wavelength or frequency, $E_{\rm ph} = h\,c/\lambda = h\,\nu$, where h is Planck's constant. The wavelengths, frequencies and photon energies of various wavebands are given in Table G.6. The *wave–particle duality of light* is a deep issue in physics and related via the concept of probability. To quote Max Born, 'the wave and corpuscular descriptions are only to be regarded as complementary ways of viewing one and the same objective process, a process which only in definite limiting cases admits of complete pictorial interpretation' (Ref. [23]). Although, in principle, it may be possible to understand a physical process involving light from both points of view, there are some problems that are more easily addressed by one approach rather than another. For example, it is often more straightforward to consider waves when dealing with an interaction between light and an object that is small in comparison to the wavelength, and to consider photons when dealing with an interaction between light and an object that is large in comparison to the equivalent photon wavelength, $\lambda = h\,c/E_{\rm ph}$. Given this wave-particle duality, in this text we will apply whatever form is most useful for the task at hand. Some helpful expressions relating various properties of electromagnetic radiation are provided in Table I.1 and a diagram illustrating the wave nature of light is shown in Figure I.2.

[1]When 'Galaxy' is written with a capital G, it refers to our own Milky Way galaxy.

Table I.1. Useful expressions relevant to light, matter, and fields[a]

Meaning	Equation
Relation between wavelength and frequency	$c = \lambda\, \nu$
Lorentz factor	$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$
Energy of a photon	$E = h\,\nu = \frac{h\,c}{\lambda}$
Equivalent mass of a photon	$\mathrm{m} = E/c^2$
Equivalent wavelength of a mass (de Broglie wavelength)	$\lambda = h/(\mathrm{m}\, v)$
Momentum of a photon	$p = E/c$
Momentum of a particle	$p = \gamma\, \mathrm{m}\, v$
Snell's law of refraction[b]	$\mathrm{n}_1\, \sin\, \theta_1 = \mathrm{n}_2\, \sin\, \theta_2$
Index of refraction[c]	$\mathrm{n} = c/v$
Doppler shift[d]	$\frac{\Delta\lambda}{\lambda_0} = \frac{\lambda_{\mathrm{obs}} - \lambda_0}{\lambda_0} = \left(\frac{1+\frac{v_r}{c}}{1-\frac{v_r}{c}}\right)^{1/2} - 1$
	$\frac{\Delta\nu}{\nu_0} = \frac{\nu_{\mathrm{obs}} - \nu_0}{\nu_0} = \left(\frac{1-\frac{v_r}{c}}{1+\frac{v_r}{c}}\right)^{1/2} - 1$
	$\frac{\Delta\lambda}{\lambda_0} \approx \frac{v_r}{c}, \frac{\Delta\nu}{\nu_0} \approx \frac{-v_r}{c},\ (v \ll c)$
Electric field vector	$\vec{\mathrm{E}} = \vec{\mathrm{F}}/q$
Electric field vector of a wave[e]	$\vec{\mathrm{E}} = \vec{\mathrm{E}_0}\, \cos[2\,\pi(\frac{z}{\lambda} - \nu\, t) + \Delta\phi]$
Magnetic field vector of a wave[e]	$\vec{B} = \vec{B}_0\, \cos[2\,\pi(\frac{z}{\lambda} - \nu\, t) + \Delta\phi]$
Poynting flux[f]	$\vec{S} = (\frac{c}{4\,\pi})\, \vec{E} \times \vec{B}$
Time averaged Poynting flux[f]	$\langle S\rangle = \frac{c}{8\,\pi}\, E_0\, B_0$
Energy density of a magnetic field[g]	$u_B = \frac{B^2}{8\,\pi}$
Lorentz force[h]	$\vec{\mathrm{F}} = q\,(\vec{\mathrm{E}} + \frac{\vec{v}}{c} \times \vec{B})$
Electric field magnitude in a parallel-plate capacitor[i]	$\mathrm{E} = (4\,\pi\, N\, e)/A$
Electric dipole moment[j]	$\vec{p} = q\,\vec{r}$
Larmor's formula for power[k]	$P = \frac{dE}{dt} = \frac{2\,q^2}{3\,c^3}\, \lvert\ddot{\vec{r}}\rvert^2$
Heisenberg Uncertainty Principle[l]	$\Delta\, x\, \Delta\, p_x = \frac{h}{2\,\pi}$
	$\Delta\, E\, \Delta\, t = \frac{h}{2\,\pi}$
Universal law of gravitation[m]	$\mathrm{F}_G = G\, \frac{\mathrm{M}\,\mathrm{m}}{r^2}$
Centripetal force[n]	$\mathrm{F}_c = \frac{\mathrm{m}\, v^2}{r}$

a See Table A.3 for a list of symbols if they are not defined here.

[b] If an incoming ray is travelling from medium 1 with index of refraction, n_1, into medium 2 with index of refraction, n_2, then θ_1 is the angle between the incoming ray and the normal to the surface dividing the two media and θ_2 is the angle between the outgoing ray and the normal to the surface.

[c] v is the speed of light in the medium (the phase velocity) and c is the speed of light in a vacuum. Note that the index of refraction may also be expressed as a complex number whose real part is given by this equation and whose imaginery part corresponds to an absorbed component of light. See Appendix D.3 for an example.

[d] λ_0 is the wavelength of the light in the source's reference frame (the 'true' wavelength), λ_{obs} is the wavelength in the observer's reference frame (the measured wavelength), and v_r is the relative *radial* velocity between the source and the observer. v_r is taken to be positive if the source and observer are receding with respect to each other and negative if the source and observer are approaching each other.

[e] The wave is propagating in the z direction and $\Delta\,\phi$ is an arbitrary phase shift. The magnetic field strength, H, is given by $B = H\,\mu$ where μ is the permeability of the substance through which the wave is travelling (unitless in the cgs system). For em radiation in a vacuum (assumed here and throughout), this becomes $B = H$ since the permeability of free space takes the value, 1, in cgs units. Thus B is often stated as the magnetic field strength, rather than the magnetic flux density and is commonly expressed in units of Gauss. In cgs units, $\mathrm{E}\ (\mathrm{dyn\ esu}^{-1}) = B$ (Gauss).

[f] Energy flux carried by the wave in the direction of propagation. The cgs units are erg s^{-1} cm^{-2}. The time-averaged value is over one cycle.

[g] u_B has cgs units of erg cm^{-3} or dyn cm^{-2} (see Table A.2.).

[h] Force on a charge, q with velocity, $\vec{v}$, by an electric field, $\vec{\mathrm{E}}$ and magnetic field, $\vec{B}$.

[i] Here $N\,e$ is the charge on a plate and A is its area. In SI units, this equation would be $\mathrm{E} = \sigma/\epsilon_0$, where σ is the charge per unit area on a plate and ϵ_0 is the permittivity of free space (where we assume that free space is between the plates). In cgs units, $4\,\pi\,\epsilon_0 = 1$.

[j] $\vec{r}$ is the separation between the two charges of the dipole and q is the strength of one of the charges. The direction is negative to positive.

[k] Power emitted by a non-relativistic particle of charge, q, that is accelerating at a rate, $\ddot{\vec{r}}$.

[l] One cannot know the position and momentum ($x,\ p$) or the energy and time ($E,\ t$) of a particle or photon to arbitrary accuracy.

[m] Force between two masses, M and m, a distance, r, apart.

[n] Force on an object of mass, m, moving at speed, v, in a circular path of radius, r.

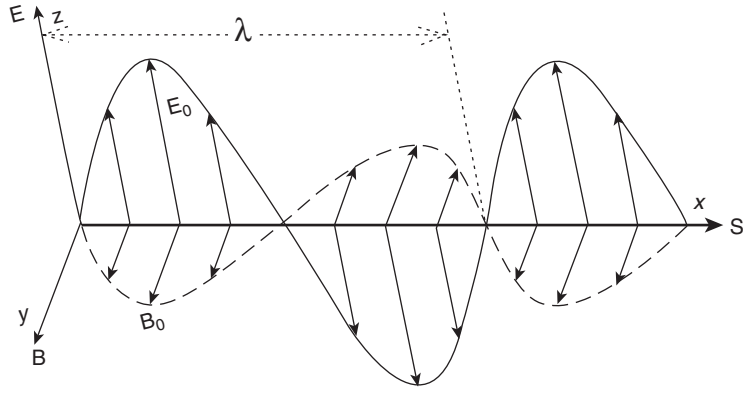
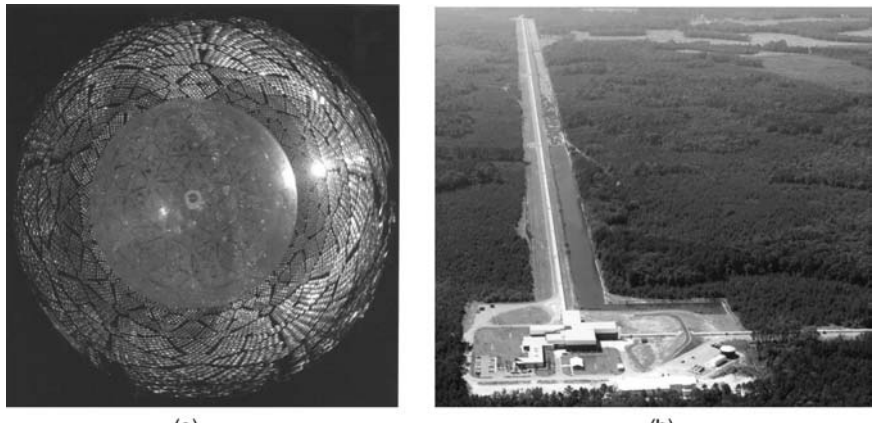


Figure I.2 Illustration of an electromagnetic wave showing the electric field and magnetic field perpendicular to each other and perpendicular to the direction of wave propagation which is in the x direction. The wavelength is denoted by λ.

If we now ask which of these information bearers provides us with most of our current knowledge of the universe, the answer is undoubtedly electromagnetic radiation, with cosmic rays a distant second. The world's astronomical volumes would be empty indeed were it not for an understanding of radiation and its interaction with matter. The radiation may come directly from the object of interest, as when sunlight travels straight to us, or it may be indirect, such as when we infer the presence of a black hole by the X-rays emitted from a surrounding accretion disk. Even when we send out exploratory astronomical probes, we still rely on man-made radiation to transfer the images and data back to earth.

This volume is thus largely devoted to understanding radiative processes and how such an understanding informs us about our Universe and the astronomical objects that inhabit it. It is interesting that, in order to understand the largest and grandest objects in the Universe, we must very often appeal to microscopic physics, for it is on such scales that the radiation is actually being generated and it is on such scales that matter interacts with it. We also focus on the 'how' of astronomy. How do we know the temperature of that asteroid? How do we find the speed of that star? How do we know the density of that interstellar cloud? How can we find the energy of that distant quasar? The answers are hidden in the radiation that they emit and, earthbound, we have at least a few keys to unlock their secrets. We can truly think of the detected signal as a coded message. To understand the message requires careful decoding.

In the future, there may be new, exotic and perhaps unexpected ways in which we can gather information about our Universe. Already, we are seeing a trend towards such diversification. The decades-old 'Solar neutrino problem' has finally been solved from the vantage point of an underground mine in the Canadian shield (Figure I.3.a). Creative experiments are underway to detect the elusive *dark matter* that is believed to make up most of the mass of the Universe but whose nature is unknown (Sect. 3.2). Enormous international efforts are also underway to detect *gravitational waves*, weak perturbations in space-time predicted by Einstein's General Theory of Relativity



(a) (b)

Figure I.3 (**a**) The acrylic tank of the Sudbury Neutrino Observatory (SNO) looks like a coiled snakeskin in this fish-eye view from the bottom of the tank before the bottom-most panel of photomultiplier tubes was installed. The Canadian-led SNO project determined that neutrinos change 'flavour' en route from the Sun's interior, thus answering why previous experiments had detected fewer solar neutrinos than theory predicted. (Photo credit: Ernest Orlando, Lawrence Berkeley National Laboratory. Courtesy A. Mc Donald, SNO) (**b**) Aerial photo of the 'L' shaped Laser Interferometer Gravitational-Wave Observatory (LIGO) in Livingston, Louisiana, of length 4 km on each arm. Together with its sister observatory in Hanford, Washington, these interferometers may detect gravitational waves, tiny distortions in space–time produced by accelerating masses. (Reproduced courtesy of LIGO, Livingston, Lousiana)

(Figure I.3.b). Our astronomical observatories are no longer restricted to lonely mountain tops. They are deep underground, in space, and in the laboratory. Thus, 'decoding the cosmos' is an on-going and evolving process. Here are a few steps along the path.

Problems

0.1 Calculate the repulsive force of an electron on another electron a distance, 1 m, away in cgs units using Eq. 0.A.1 and in SI units using the equation,

$$\mathrm{F} = k\,\frac{q_1\,q_2}{r^2} \tag{I.1}$$

where the charge on an electron, in SI units, is 1.6×10^{-19} Coulombs (C) and the constant, $k = 8.988 \times 10^9$ N m^2 C^{-2}. Verify that the result is the same in the two systems.

0.2 Verify that the following equations have matching units on both sides of the equals sign: Eq. D.2 (the expression that includes the mass of the electron, m_e), Eq. 9.8 and Eq. F.19.

0.3 Verify that the equation for the *Ideal Gas Law* (Table 0.A.2) is equivalent to $\mathrm{P}V = \mathcal{N}\mathcal{R}\,\mathrm{T}$, where $\mathcal{N}$ is the number of *moles* and $\mathcal{R}$ is the *molar gas constant* (see Table G.2).

Appendix

Dimensions, Units and Equations

> When [chemists] had unscrambled the difficulties caused by the fact that chemists and engineers use different units ... they found that their strength predictions were not only frequently a thousandfold in error but bore no consistent relationship with experiment at all. After this they were inclined to give the whole thing up and to claim that the subject was of no interest or importance anyway.
>
> –*The New Science of Strong Materials*, by J. E. Gordon

The centimetre-gram-second (cgs) system of units is widely used by astronomers internationally and is the system adopted in this text. A summary of the units is given in Table 0.A.1 as well as corresponding conversions to Système Internationale (SI). If an equation is given without units, the cgs system is understood. The same symbols are generally used in both systems (Table 0.A.3) and SI prefixes (Table 0.A.4) are also equally applied to the cgs system (note that the base unit, cm, already has a prefix).

Table 0.A.1. Selected cgs – SI conversions[a]

Dimension	cgs unit (abbrev.)	Factor	SI unit[b] (abbrev.)
Length	centimetre (cm)	10^{-2}	metre (m)
Mass	gram (g)	10^{-3}	kilogram (kg)
Time	second (s)	1	second (s)
Energy	erg (erg)	10^{-7}	joule (J)
Power	erg second^{-1} (erg s^{-1})	10^{-7}	watt (W)
Temperature	kelvin (K)	1	kelvin (K)
Force	dyne (dyn)	10^{-5}	newton (N)
Pressure	dyne centimetre^{-2} (dyn cm^{-2})	0.1	newton metre^{-2} (N m^{-2})
	barye (ba)[c]	0.1	pascal (Pa)
Magnetic flux density (field)[d]	gauss (G)	10^{-4}	tesla (T)
Angle	radian (rad)	1	radian (rad)
Solid angle	steradian (sr)	1	steradian (sr)

[a] Value in cgs units times factor equals value in SI units.
[b] Système International d'Unités.
[c] This unit is rarely used in astronomy in favour of dyn cm^{-2}.
[d] See notes to Table I.1.

Table 0.A.2. Examples of equivalent units

Equation[a]	Name	Units
$\mathrm{P} = n\,k\,\mathrm{T}$	Ideal Gas Law	$\mathrm{dyn\,cm^{-2}} = \left(\frac{1}{\mathrm{cm^3}}\right)\left(\frac{\mathrm{erg}}{\mathrm{K}}\right)(\mathrm{K}) = \mathrm{erg\,cm^{-3}}$
$\mathrm{F} = \mathrm{m}\,a$	Newton's Second Law	$\mathrm{dyn} = \mathrm{g\,cm\,s^{-2}}$
$W = \mathrm{F}\,s$	Work Equation	$\mathrm{erg} = \mathrm{dyn\,cm} = \mathrm{g\,cm^2\,s^{-2}}$
$E_k = \frac{1}{2}\,\mathrm{m}\,v^2$	Kinetic Energy Equation	$\mathrm{erg} = \mathrm{g\,cm^2\,s^{-2}}$
$E_{\mathrm{PG}} = \mathrm{m}\,g\,\mathrm{h}$	Gravitational Potential Energy Equation	$\mathrm{erg} = \mathrm{g\,cm^2\,s^{-2}}$
$E_{\mathrm{Pe}} = \frac{q_1\,q_2}{r}$	Electrostatic Potential Energy Equation	$\mathrm{erg} = \mathrm{esu^2\,cm^{-1}}$

See Table 0.A.3 for the meaning of the symbols and Table G.2 for the meaning of the constants.

Table 0.A.3. List of symbols

Symbol	Meaning
a	radius, acceleration
A	atomic weight
A	area, albedo, total extinction
B	magnetic flux density, magnetic field strength[a]
$B(\mathrm{T})$	intensity of a black body (or specific intensity if subscripted with ν or λ)
D_{p}	degree of polarization
e	charge of the electron
E	energy, selective extinction
E	electric field strength
$\mathcal{EM}$	emission measure
$f_{i,j}$	oscillator strength between levels, i and j
f	correction factor
$F,\ S,\ f$	flux (or flux density, if subscripted with ν or λ)
F	force
g_n	statistical weight of level, n
g	Gaunt factor
I	intensity (or specific intensity, if subscripted with ν or λ)
j_ν	emission coefficient
J	mean intensity (or mean specific intensity, if subscripted with ν or λ)
$J(E)$	cosmic ray flux density per unit solid angle
$\bar{l}$	mean free path
L	luminosity (or spectral luminosity, if subscripted with ν or λ)
m, M	apparent magnitude & absolute magnitude, respectively
M, m	mass
n	index of refraction, principal quantum number
n, N	number density and number of (object), respectively
N	map noise
$\mathcal{N}$	number of moles, column density
p	momentum, electric dipole moment
P	power (or spectral power, if subscripted with ν or λ), probability
P	pressure
q	charge

(*Continued*)

Table 0.A.3. (*Continued*)

Symbol	Meaning
Q	efficiency factor
r, d, D, s, x, y, z, l, h, R	Position, separation, or distance
R	Rydberg constant
$\mathcal{R}$	collision rate
S	source function
t	time
T	kinetic energy
T	temperature
$\mathcal{T}$	period
u	energy density (or spectral energy density, if subscripted with ν or λ)
U, V, B, etc.	apparent magnitudes
$\mathcal{U}$	excitation parameter
v	velocity, speed
V	volume
X, Y, Z	mass fraction of hydrogen, helium and heavier elements, respectively
z	redshift, zenith angle
Z	atomic number
α	synchrotron spectral index, fine structure constant
α_ν	absorption coefficient
α_r	recombination coefficient
γ	Lorentz factor
$\gamma_{\rm coll}$	collision rate coefficient
Γ	spectral index of cosmic ray power law dist'n, damping constant
ϵ	permittivity, cosmological energy density
θ, ϕ	one-dimensional angle
κ_ν	mass absorption coefficient
λ	wavelength
μ	permeability, mean molecular weight
ν	frequency, collision rate per unit volume
ρ	mass density
σ	cross-sectional area, Stefan-Boltzmann constant, Gaussian dispersion
τ_ν, τ	optical depth & timescale, respectively
Φ	line shape function
χ	ionization potential
ω	angular frequency
Ω	solid angle, cosmological mass or energy density

[a] See notes to Table I.1.

Almost all equations used in this text look identical in the two systems and one need only ensure that the constants and input parameters are all consistently used in the adopted system. There are, however, a few cases in which the equation itself changes between cgs and SI. An example is the Coulomb (electrostatic) force,

$$\mathrm{F} = \frac{q_1\, q_2}{r^2} \tag{0.A.1}$$

Table 0.A.4. SI prefixes

Name	Prefix	Factor
yocto	y	10^{-24}
zepto	z	10^{-21}
atto	a	10^{-18}
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}
centi	c	10^{-2}
deci	d	10^{-1}
kilo	k	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}
exa	E	10^{18}
zeta	Z	10^{21}
yota	Y	10^{24}

With the two charges, q_1 and q_2 expressed in *electrostatic units* (esu, see Table G.2), and the separation, r, in cm, the answer will be in dynes. Note that there is no constant of proportionality in this equation, unlike the SI equivalent (Prob. 0.1). Equations in the cgs system show the most difference with their SI equivalents when electric and magnetic quantities are used. For example, in the cgs system, the permittivity and permeability of free space, ϵ_0 and μ_0, respectively, are both unitless and equal to 1 (see also notes to Table I.1).

Astronomers also have a number of units specific to the discipline. There are a variety of reasons for this, but the most common involves a process of normalisation. The value of some parameter is expressed in comparison to another known, or at least more familiar value. Some examples (see Table G.3) are the *astronomical unit* (AU) which is the distance between the Earth and the Sun. It is much easier to visualise the distance to Pluto as 40 AU than as 6×10^{14} cm. Expressing the masses of stars and galaxies in Solar masses ($M_\odot$) or an object's luminosity in Solar luminosities ($L_\odot$) is also very common. When such units are used, the parameter is often written as $M/M_\odot$, or $L/L_\odot$. Some examples can be seen in Eqs. 8.15 through 8.18 and in many other equations in this book.

A very valuable tool for checking the answer to a problem, or to help understand an equation, is that of *dimensional analysis*. The dimensions of an equation (e.g. time, velocity, distance) must agree and therefore their units (s, cm s^{-1}, cm, respectively) must also agree. Two quantities can be added or subtracted only if they have the same units, and logarithms and exponentials are unitless. In this process, it is helpful to recall some *equivalent units* which are revealed by writing

down some simple well-known equations in physics. A few examples are provided in Table 0.A.2. The example of the Ideal Gas Law in this table also shows the process of dimensional analysis, which involves writing down the units to every term and then cancelling where possible. A more complex example of dimensional analysis is given in Example A.1.

Example 0.A.1

For a gas in thermal equilibrium at some uniform temperature, T, and uniform density, n, the number density of particles with speeds[2] between v and $v + dv$ is given by the Maxwell–Boltzmann (or simply 'Maxwellian') velocity distribution,

$$n(v)\,dv = n\left(\frac{\mathrm{m}}{2\pi k\,\mathrm{T}}\right)^{3/2}\exp\left(-\frac{\mathrm{m}\,v^2}{2k\,\mathrm{T}}\right)4\pi\,v^2\,dv \tag{0.A.2}$$

where $n(v)$ is the gas density per unit velocity interval, m is the mass of a gas particle (taken here to be the same for all particles), and v is the particle speed. A check of the units gives,

$$\frac{1}{\mathrm{cm}^3\,\frac{\mathrm{cm}}{\mathrm{s}}}\frac{\mathrm{cm}}{\mathrm{s}} = \frac{1}{\mathrm{cm}^3}\left(\frac{\mathrm{g}}{\frac{\mathrm{erg}}{\mathrm{K}}\,\mathrm{K}}\right)^{3/2}\exp\left(-\frac{\mathrm{g}\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)^2}{\frac{\mathrm{erg}}{\mathrm{K}}\,\mathrm{K}}\right)\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)^2\frac{\mathrm{cm}}{\mathrm{s}} \tag{0.A.3}$$

Simplifying yields,

$$\frac{1}{\mathrm{cm}^3} = \frac{1}{\mathrm{s}^3}\left(\frac{\mathrm{g}}{\mathrm{erg}}\right)^{3/2}\exp\left(-\frac{\mathrm{g}\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)^2}{\mathrm{erg}}\right) \tag{0.A.4}$$

Using the equivalent units for energy (Table A.2), the exponential is unitless, as required, and we find,

$$\frac{1}{\mathrm{cm}^3} = \frac{1}{\mathrm{cm}^3} \tag{0.A.5}$$

Figure 3.12 shows a plot of this function. If Eq. (0.A.2) is integrated over all velocities, either on the left hand side (LHS) or the right hand side (RHS), the total density should result. Since an integration over all velocities is equivalent to a sum over the individual infinitesimal velocity intervals, the total density will also have the required units of $\frac{1}{\mathrm{cm}^3}$. This dimensional analysis helps to clarify the fact that, since the total density, n, appears on the RHS of Eq. (0.A.2) and is a constant, an integration over all terms, except n, on the RHS must be unitless and equal 1. (In fact these remaining terms represent a *probability distribution function*, see Sect. 3.4.1 for a description.) Also, since the number density is just the number of particles, N, divided by a constant (the volume), we could have substituted $N(v)\,dv$ on the LHS and N on the RHS for the density terms. Similarly, since

[2]Speed and velocity are taken to be equivalent in this text unless otherwise indicated.

the density and temperature are constant, we could have multiplied Eq. (0.A.2) by kT to turn it into an equation for the particle pressure in a velocity interval $\mathrm{P}(v)\,dv$ with the total pressure, P on the RHS (Table 0.A.2). Thus, while dimensional analysis says nothing about the origin or fundamentals of an equation, it can go a long way in revealing the meaning of one and how it might be manipulated.

An important comment is that one should be very careful of simplifying units without thinking about their meaning. A good example is the unit, erg s^{-1} Hz^{-1} which is a representation of a luminosity or power (see Sect. 1.1) per unit frequency (Hz) in some waveband. Since Hz can be represented as s^{-1}, the above could be written erg s^{-1} s^{1} = erg which is simply an energy and does not really express the intended meaning of the term. Similar difficulties can arise when a term is expressed as 'per cm of waveband'. Units of frequency or wavelength should not be combined with units of time or distance, respectively.

PART I
The Signal Observed

The radiation from an astronomical source can be thought of as a signal that provides us with information about it. In order to relate the signal received at the Earth to the physical conditions within an astronomical source, however, we first need ways to describe and measure light. This requires setting out the basic definitions for quantities involving light and the relationships between them. The definitions range from those associated with values measured at the Earth to those that are intrinsic to the source itself. We do this without regard (yet) for the processes that actually generate the light, an approach similar to what is often followed in Mechanics. For example, first one studies Kinematics which relates distances, velocities, and accelerations and the relations between them. Later, one considers Dynamics which deals with the forces that produce these motions. Measuring light involves a deep understanding of how the measurement process itself affects the signal and also how the Earth's atmosphere interferes. Our instrumentation imposes its own signature on an astronomical signal and it is important to account for this imposition. These steps are fundamental and lay the groundwork for turning the measurement of a weak glimmer of light into an understanding of what drives the most powerful objects in the Universe.

1

Defining the Signal

. . . the distance of the invisible background [is] so immense that no ray from it has yet been able to reach us at all.

–Edgar Allan Poe in *Eureka*, 1848

1.1 The power of light – luminosity and spectral power

The *luminosity*, L, of an object is the rate at which the object radiates away its energy (cgs units of erg s^{-1} or SI units of watts),

$$dE = L\,dt \tag{1.1}$$

This quantity has the same units as *power* and is simply the radiative power output from the object. It is an intrinsic quantity for a given object and does not depend on the observer's distance or viewing angle. If a star's luminosity is L_* at its surface, then at a distance r away, its luminosity is still L_*.

Any object that radiates, be it spherical or irregularly shaped, can be described by its luminosity. The Sun, for example, has a luminosity of $L_\odot = 3.85 \times 10^{33}$ erg s^{-1} (Table G.3), most of which is lost to space and not intercepted by the Earth (Example 1.1).

Example 1.1

Determine the fraction of the Sun's luminosity that is intercepted by the Earth. What luminosity does this correspond to?

At the distance of the Earth, the Sun's luminosity, $L_\odot$, is passing through the imaginary surface of a sphere of radius, $r_\oplus = 1$ AU. The Earth will be intercepting photons over only

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

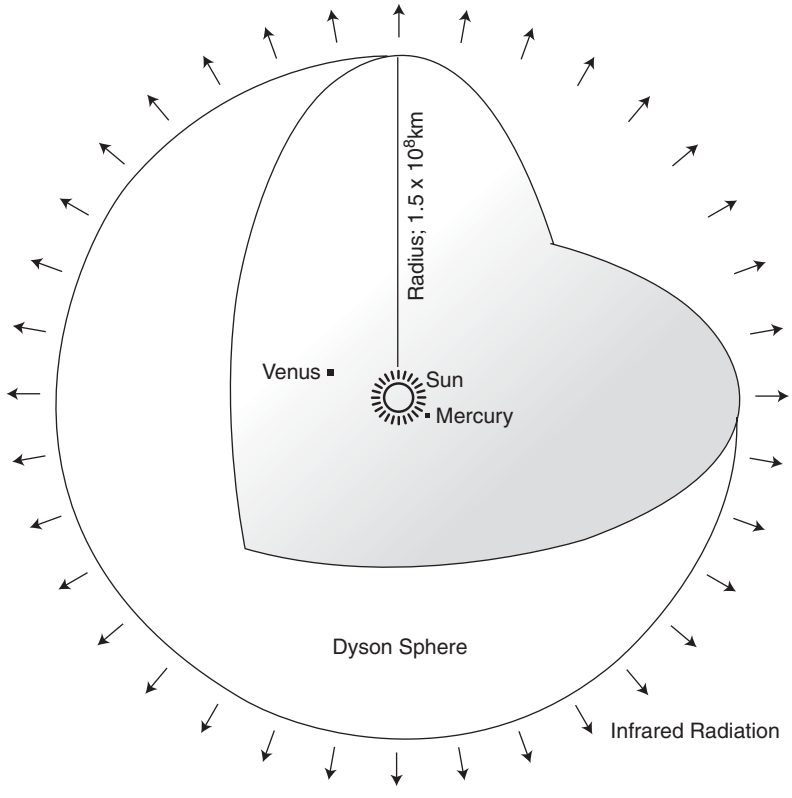


Figure 1.1. Illustration of a Dyson Sphere that could capture the entire luminous ouput from the Sun. Some have suggested that advanced civilizations, if they exist, would have discovered ways to build such spheres to harness all of the energy of their parent stars.

the cross-sectional area that is facing the Sun. This is because the Sun is so far away that incoming light rays are parallel. Thus, the fraction will be

$$\mathbf{f} = \frac{\pi {R_\oplus}^2}{4\pi r_\oplus^2} \tag{1.2}$$

where $R_\oplus$ is the radius of the Earth. Using the values of Table G.3, the fraction is $\mathbf{f} = 4.5 \times 10^{-10}$ and the intercepted luminosity is therefore $L_{\rm int} = \mathbf{f}\, L_\odot = 1.73 \times 10^{24}\,{\rm erg\,s^{-1}}$. A hypothetical shell around a star that would allow a civilization to intercept *all* of its luminosity is called a *Dyson Sphere* (Figure 1.1).

When one refers to the luminosity of an object, it is the *bolometric* luminosity that is understood, i.e. the luminosity over all wavebands. However, it is not possible to determine this quantity easily since observations at different wavelengths require different techniques, different kinds of telescopes and, in some wavebands, the

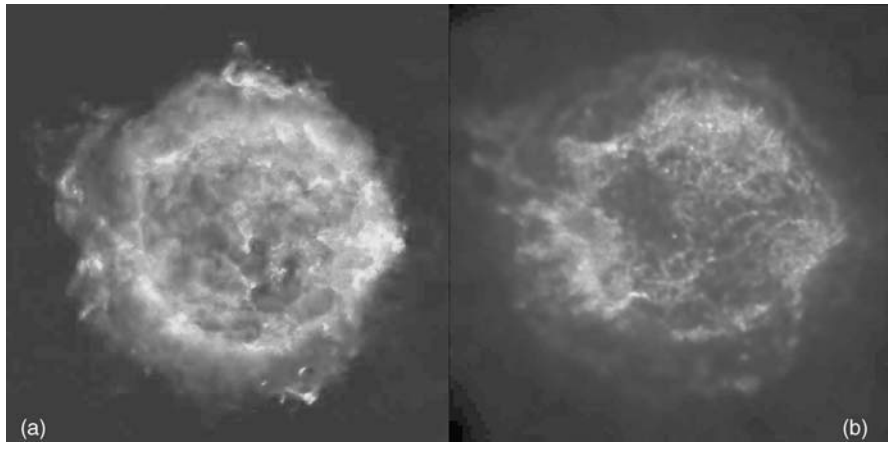


Figure 1.2. The supernova remnant, Cas A, at a distance of 3.4 kpc and with a linear diameter of $\approx$4 pc, was produced when a massive star exploded in the year AD 1680. It is currently expanding at a rate of 4000 km s^{-1} (Ref. [181]) and the proper motion (angular motion in the plane of the sky, see Sect. 7.2.1.1) of individual filaments have been observed. One side of the bipolar jet, emanating from the central object, can be seen at approximately 10 o'clock. **(a)** Radio image at λ20 cm shown in *false colour* (see Sect. 2.6) from Ref. [7]. Image courtesy of NRAO/AUI/NSF. **(b)** X-ray emission, with red, green and blue colours showing, respectively, the intensity of low, medium and high energy X-ray emission. (Reproduced courtesy of NASA/CXC/SAO) (see colour plate)

necessity of making measurements above the obscuring atmosphere of the Earth. Thus, it is common to specify the luminosity of an object for a given waveband (see Table G.6). For example, the *supernova remnant*, Cas A (Figure 1.2), has a radio luminosity (from $\nu_1 = 2 \times 10^7$ Hz to $\nu_2 = 2 \times 10^{10}$ Hz) of $L_{\text{radio}} = 3 \times 10^{35}$ erg s^{-1} (Ref. [6]) and an X-ray luminosity (from 0.3 to 10 keV) of $L_{\text{X-ray}} = 3 \times 10^{37}$ erg s^{-1} (Ref. [37]). Its bolometric luminosity is the sum of these values plus the luminosities from all other bands over which it emits. It can be seen that the radio luminosity might justifiably be neglected when computing the total power output of Cas A. Clearly, the source *spectrum* (the emission as a function of wavelength) is of some importance in understanding which wavebands, and which processes, are most important in terms of energy output. The spectrum may be represented mostly by *continuum emission* as implied here for Cas A (that is, emission that is continuous over some spectral region), or may include *spectral lines* (emission at discrete wavelengths, see Chapter 3, 5, or 9). Even very weak lines and weak continuum emission, however, can provide important clues about the processes that are occurring within an astronomical object, and must not be neglected if a full understanding of the source is to be achieved.

In the optical region of the spectrum, various *passbands* have been defined (Figure 1.3). The Sun's luminosity in V-band, for example, represents 93 per cent of its bolometric luminosity.

The *spectral luminosity* or *spectral power* is the luminosity per unit bandwidth and can be specified per unit wavelength, L_λ (cgs units of erg s^{-1} cm^{-1}) or per unit

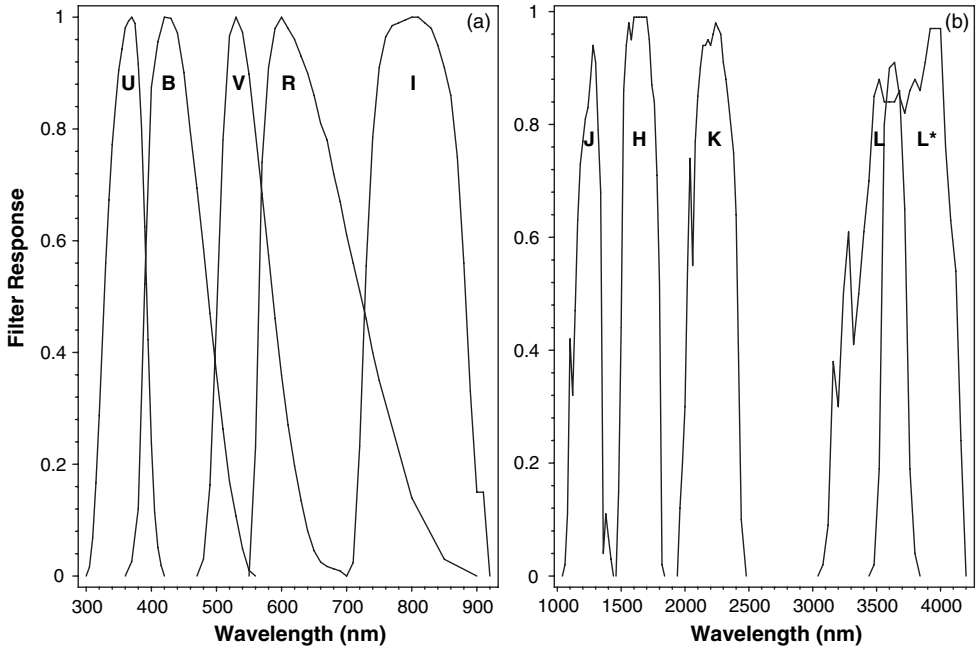


Figure 1.3. Filter bandpass responses for **(a)** the UBVRI bands (Ref. [17]) and **(b)** the JHKLL* bands (Ref. [19]). (The U and B bands correspond to UX and BX of Ref. [17].) Corresponding data can be found in Table 1.1

frequency, L_ν (erg s^{-1} Hz^{-1}),

$$dL = L_\lambda \, d\lambda = L_\nu \, d\nu \tag{1.3}$$

$$\text{so} \qquad L = \int L_\lambda \, d\lambda = \int L_\nu \, d\nu \tag{1.4}$$

Note that, since $\lambda = \frac{c}{\nu}$,

$$d\lambda = -\frac{c}{\nu^2} \, d\nu \tag{1.5}$$

so the magnitudes of L_λ and L_ν will not be equal (Prob. 1.1). The negative sign in Eq. (1.5) serves to indicate that, as wavelength increases, frequency decreases. In equations like Eq. (1.4) in which the wavelength and frequency versions of a function are related to each other, this negative is already taken into account by ensuring that the lower limit to the integral is always the lower wavelength or frequency. Note that the cgs units of L_λ (erg s^{-1} cm^{-1}) are rarely used since 1 cm of bandwidth is exceedingly large (Table G.6). Non-cgs units, such as erg s^{-1} Å^{-1} are sometimes used instead.

Luminosity is a very important quantity because it is a basic parameter of the source and is directly related to energetics. Integrated over time, it provides a measure of the energy required to make the object shine over that timescale. However, it is not a quantity that can be measured directly and must instead be derived from other measurable quantities that will shortly be described.

1.2 Light through a surface – flux and flux density

The *flux* of a source, f ($\mathrm{erg\,s^{-1}\,cm^{-2}}$), is the radiative energy per unit time passing through unit area,

$$dL = f\,dA \tag{1.6}$$

As with luminosity, we can define a flux in a given waveband or we can define it per unit spectral bandwidth. For example, the *spectral flux density*, or just *flux density* ($\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}}$ or $\mathrm{erg\,s^{-1}\,cm^{-2}\,cm^{-1}}$)[1] is the flux per unit spectral bandwidth, either frequency or wavelength, respectively,

$$\begin{aligned} dL_\nu &= f_\nu\,dA \qquad & dL_\lambda &= f_\lambda\,dA \\ df &= f_\nu\,d\nu \qquad & df &= f_\lambda\,d\lambda \end{aligned} \tag{1.7}$$

A special unit for flux density, called the *Jansky* (Jy) is utilized in astronomy, most often in the infrared and radio parts of the spectrum,

$$1\,\mathrm{Jy} = 10^{-26} \qquad \mathrm{W\,m^{-2}\,Hz^{-1}} = 10^{-23} \qquad \mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}} \tag{1.8}$$

Radio sources that are greater than 1 Jy are considered to be strong sources by astronomical standards (Prob. 1.3).

The spectral response is independent of other quantities such as area or time so Eq. (1.6) and the first line of Eq. (1.7) show the same relationships except for the subscripts. To avoid repitition, then, we will now give the relationships for the bolometric quantities and it will be understood that these relationships apply to the subscripted 'per unit bandwidth' quantities as well.

The luminosity, L, of a source can be found from its flux via,

$$L = \int f\,dA = 4\,\pi\,r^2 f \tag{1.9}$$

where r is the distance from the centre of the source to the position at which the flux has been determined. The $4\,\pi\,r^2$ on the right hand side (RHS) of Eq. (1.9) is strictly

[1]The two 'cm' designations should remain separate. See the Appendix at the end of the Introduction.

Figure 1.4. An image of the Centaurus A jet emanating from an *active galactic nucleus* (AGN) at the centre of this galaxy and at lower right of this image. Radio emission is shown in red and X-rays in blue. (Reproduced by permission of Hardcastle M.J., *et al.*, 2003 ApJ, **593**, 169.) Even though gaseous material may be moving along the jet in a highly directional fashion, the RHS of Eq. (1.9) may still be used, provided that photons generated within the jet (such as at the knot shown) escape in all directions. (see colour plate)

only true for sources in which the photons that are generated can escape in all directions, or *isotropically*. This is usually assumed to be true, even if the source itself is irregular in shape (Figure 1.4). These photons pass through the imaginary surfaces of spheres as they travel outwards. The $\frac{1}{r^2}$ fall-off of flux is just due to the geometry of a sphere (Figure 1.5.a). In principle, however, one could imagine other geometries. For example, the flux of a man-made laser beam would be constant with r if all emitted light rays are parallel and without losses (Figure 1.5.b). Light that is beamed into a narrow cone, such as may be occurring in pulsars[2] is an example of an intermediate case (Prob. 1.4).

For *stars*, we now define the *astrophysical flux*, F, to be the flux at the surface of the star,

$$L_* = 4\,\pi\,R_*^2\,F = 4\,\pi\,r^2 f \qquad \Rightarrow f = \left(\frac{R_*}{r}\right)^2 F \tag{1.10}$$

where L_* is the star's luminosity and R_* is its radius.

Using values from Table G.3, astrophysical flux of the Sun is $F_\odot = 6.33\times 10^{10}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ and the *Solar Constant*, which is the flux of the Sun at the distance

[2]Pulsars are rapidly spinning *neutron stars* with strong magnetic fields that emit their radiation in beamed cones. Neutron stars typically have about the mass of the Sun in a diameter only tens of km across.

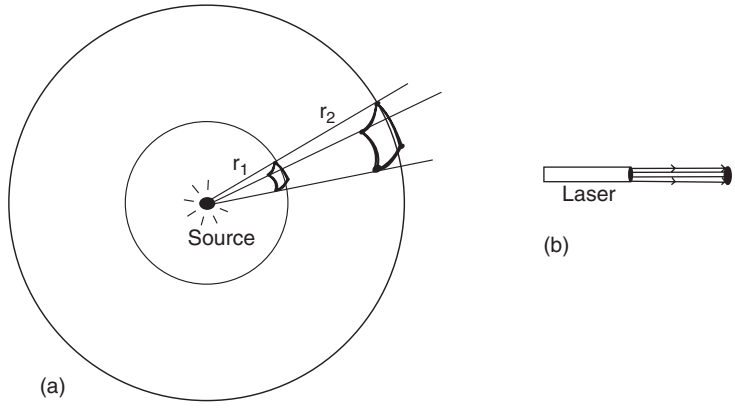


Figure 1.5. **(a)** Geometry illustrating the $\frac{1}{r^2}$ fall-off of flux with distance, r, from the source. The two spheres shown are imaginary surfaces. The same amount of energy per unit time is going through the two surface areas shown. Since the area at r_2 is greater than the one at r_1, the energy per unit time per unit cm^2 is smaller at r_2 than r_1. Since measurements are made over size scales so much smaller than astronomical distances, the detector need not be curved. **(b)** Geometry of an artificial laser. For a beam with no divergence, the flux does not change with distance.

of the Earth[3], denoted, $S_\odot$, is 1.367×10^6 erg s^{-1} cm^{-2}. The Solar Constant is of great importance since it is this flux that governs climate and life on Earth. Modern satellite data reveal that the solar 'constant' actually varies in magnitude, showing that our Sun is a variable star (Figure 1.6). Earth-bound measurements failed to detect this variation since it is quite small and corrections for the atmosphere and other effects are large in comparison (e.g. Prob. 1.5).

The flux of a source in a given waveband is a quantity that is measurable, provided corrections are made for atmospheric and telescopic responses, as required (see Sects. 2.2, 2.3). If the distance to the source is known, its luminosity can then be calculated from Eq. (1.9).

1.3 The brightness of light – intensity and specific intensity

The *intensity*, I (erg s^{-1} cm^{-2} sr^{-1}), is the radiative energy per unit time per unit solid angle passing through a unit area that is perpendicular to the direction of the emission. The *specific intensity* (erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} or erg s^{-1} cm^{-2} cm^{-1} sr^{-1}) is the radiative energy per unit time per unit solid angle per unit spectral bandwidth (either frequency or wavelength, respectively) passing through unit area perpendicular to the direction of the emission. The intensity is related to the flux via,

$$df = I \cos\theta \, d\Omega \tag{1.11}$$

[3]This is taken to be above the Earth's atmosphere.

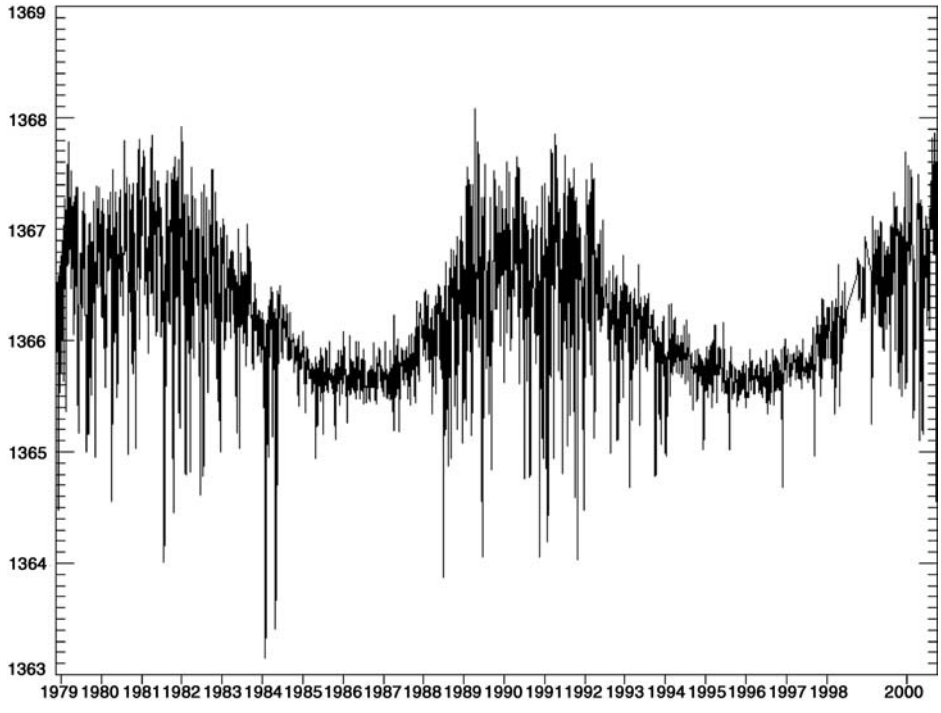


Figure 1.6. Plot of the Solar Constant (in W m^{-2}) as a function of time from satellite data. The variation follows the 11-year Sunspot cycle such that when there are more sunspots, the Sun, on average, is brighter. The peak to peak variation is less than 0.1 per cent. This plot provides definitive evidence that our Sun is a variable star. (Reproduced by permission of www.answers.com/topic/solar-variation)

As before, the same kind of relation could be written between the quantities per unit bandwidth, i.e. between the specific intensity and the flux density.

The specific intensity, I_ν, is the most basic of radiative quantities. Its formal definition is written,

$$dE = I_\nu \cos\theta \, d\nu \, d\Omega \, dA \, dt \tag{1.12}$$

Note that each elemental quantity is independent of the others so, when integrating, it doesn't matter in which order the integration is done.

The intensity isolates the emission that is within a given solid angle and at some angle from the perpendicular. The geometry is shown in Figure 1.7 for a situation in which a detector is receiving emission from a source in the sky and for a situation in which an imaginary detector is placed on the surface of a star. In the first case, the source subtends some solid angle in the sky in a direction, θ, from the zenith. The factor, cos θ accounts for the foreshortening of the detector area as emission falls on it.

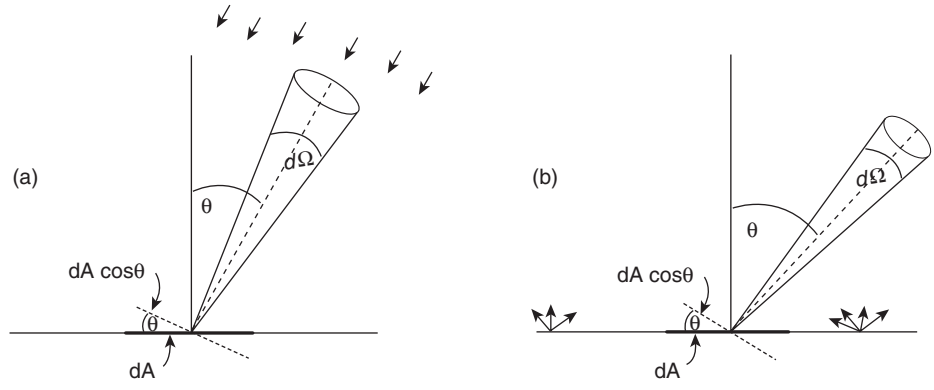


Figure 1.7. Diagrams showing intensity and its dependence on direction and solid angle, using a spherical coordinate system such as described in Appendix B. **(a)** Here *dA* would be an element of area of a detector on the Earth, the perpendicular upwards direction is towards the zenith, a source is in the sky in the direction, θ, and $d\Omega$ is an elemental solid angle on the source. The arrows show incoming rays from the *centre* of the source that flood the detector. **(b)** In this example, an imaginary detector is placed at the surface of a star. At each point on the surface, photons leave in all directions away from the surface. The intensity would be a measure of only those photons which pass through a given solid angle at a given angle, θ from the vertical

Usually, a detector would be pointed directly at the source of interest in which case $\cos\,\theta = 1$. In the second case, the coordinate system has been placed at the surface of a star. At any position on the star's surface, radiation is emitted over all directions away from the surface. The intensity refers to the emission in the direction, θ radiating *into* solid angle, $d\Omega$. Example 1.2 indicates how the intensity relates to the flux for these two examples. Figure 1.7 also helps to illustrate the generality of these quantities. One could place the coordinate system at the centre of a star, in interstellar space, or wherever we wish to determine these radiative properties of a source (Probs. 1.6, 1.7).

Example 1.2

(a) A detector pointed directly at a uniform intensity source in the sky of small solid angle, Ω, would measure a flux,

$$f = \int_{\Omega} I \cos\theta \, d\Omega \approx I\,\Omega \tag{1.13}$$

(b) The astrophysical flux at the surface of an object (e.g. a star) whose radiation is escaping freely at all angles outwards (i.e. over $2\,\pi$ sr), can be calculated by integrating

in spherical coordinates (see Appendix B),

$$F = \int I \cos\theta \, d\Omega = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} I \cos\theta \sin\theta d\theta \, d\phi = \pi I \tag{1.14}$$

Figure 1.8 shows a practical example as to how one might calculate the flux of a source for a case corresponding to Example 1.2a, but for which the intensity varies with position. The intensity in a given waveband is a measurable quantity, provided a solid angle can also be measured. If a source is so small or so far away that its angular size cannot be discerned (i.e. it is *unresolved*, see Sects. 2.2.3, 2.2.4, 2.3.2), then the intensity cannot be determined. In such cases, it is the flux that is measured, as shown in Figure 1.9. All stars other than the Sun would fall into this category[4].

Specific intensity is also referred to as *brightness* which has its intuitive meaning. A faint source has a lower value of specific intensity than a bright source. Note that it is possible for a source that is faint to have a larger flux density than a source that is bright if it subtends a larger solid angle in the sky (Prob. 1.9).

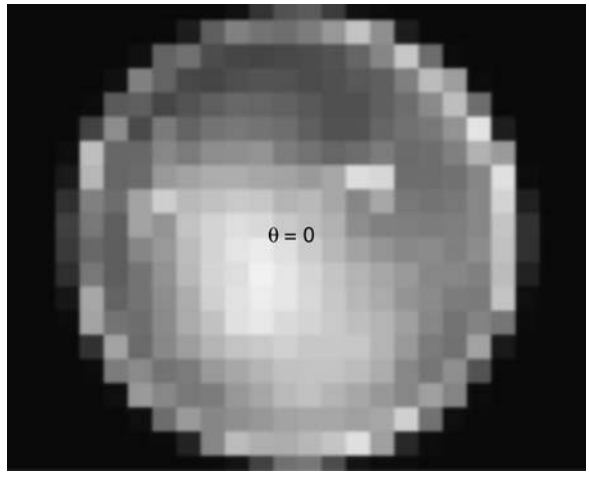


Figure 1.8. Looking directly at a hypothetical object in the sky corresponding to the situation shown in Figure 1.7.a but for $\theta = 0$ (i.e. the detector pointing directly at the source). The object subtends a total solid angle, Ω, which is small and therefore $\theta \approx 0$ at any location on the source. In this example, the object is of non-uniform brightness and Ω is split up into many small square solid angles, each of size, Ω_i and within which the intensity is I_i. Then we can approximate $f = \int I \cos\theta \, d\Omega$ using $f \approx \sum I_i \, \Omega_i$. Basically, to find the flux, we add up the individual fluxes of all elements.

[4]The exception is a few nearby stars for which special observing techniques are required.

Figure 1.9. In this case, a star has a very small angular size (left) and so, when detected in a square solid angle, Ω_p (right), which is determined by the properties of the detector, its light is 'smeared out' to fill that solid angle. In such a case, it is impossible to determine the intensity of the surface of the star. However, the flux of the star, f_*, is preserved, i.e. $f_* = I_* \Omega_* = \bar{I} \Omega_p$ (Eq. 1.13) where I_* is the true intensity of the star, Ω_* is the true solid angle subtended by the star, and $\bar{I}$ is the mean intensity in the square. Thus, for an object of angular size smaller than can be resolved by the available instruments (see Sects. 2.2.3, 2.2.4, and 2.3.2), we measure the flux (or flux density), but not intensity (or specific intensity) of the object

The intensity and specific intensity are *independent of distance* (constant with distance) in the absence of any intervening matter[5]. The easiest way to see this is via Eq. (1.13). Both f and Ω decline as $\frac{1}{r^2}$ (Eq. (1.11), Eq. (B.2), respectively) and therefore I is constant with distance. The Sun, for example, has $I_\odot = F_\odot/\pi = 2.01 \times 10^{10}$ erg s^{-1} cm^{-2} sr^{-1} as viewed from any source at which the Sun subtends a small, measurable solid angle. The constancy of I with distance is general, however, applying to large angles as well. This is a very important result, since a measurement of I allows the determination of some properties of the source without having to know its distance (e.g. Sect. 4.1).

1.4 Light from all angles – energy density and mean intensity

The *energy density*, u (erg cm^{-3}), is the radiative energy per unit volume. It describes the energy content of radiation in a unit volume of space,

$$du = \frac{dE}{dV} \tag{1.15}$$

The *specific energy density* is the energy density per unit bandwidth and, as usual, $u = \int u_\nu \, d\nu = \int u_\lambda \, d\lambda$. The energy density is related to the intensity (see Figure 1.10, Eq. 1.12) by,

$$u = \frac{1}{c} \int I \, d\Omega = \frac{4\pi}{c} J \tag{1.16}$$

[5]More accurately, I/n^2 is independent of distance along a ray path, where n is the index of refraction but the difference is negligible for our purposes.

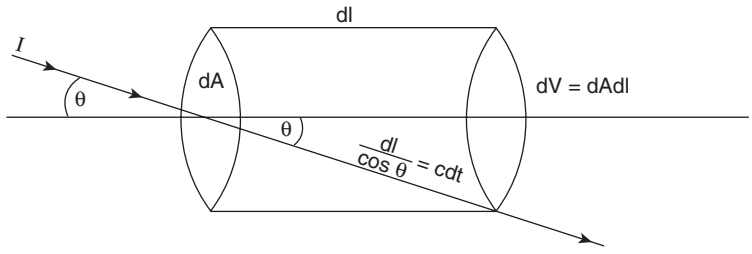


Figure 1.10. This diagram is helpful in relating the energy density (the radiative energy per unit volume) to the light intensity. An individual ray spends a time, $dt = dl/(c\,\cos\theta)$ in an infinitesimal cylindrical volume of size, $dV = dl\,dA$. Combined with Eq. (1.12), the result is Eq. (1.16).

where J is the *mean intensity*, defined by,

$$J \equiv \frac{1}{4\,\pi} \int I\,d\Omega \tag{1.17}$$

The mean intensity is therefore the intensity averaged over all directions. In an isotropic radiation field, $J = I$. In reality, radiation fields are generally not isotropic, but some are close to it or can be approximated as isotropic, for example, in the centres of stars or when considering the 2.7 K cosmic microwave background radiation (Sect. 3.1). In a non-isotropic radiation field, J is not constant with distance, even though I is. Example 1.3 provides a sample computation.

Example 1.3

Compute the mean intensity and the energy density at the distance of Mars. Assume that the only important source is the Sun.

$$\begin{aligned} J &= \frac{1}{4\,\pi} \int_0^{4\pi} I\,d\Omega \\ &= \frac{1}{4\,\pi} \int_{\Omega_\odot} I_\odot\,d\Omega \;\approx\; \frac{I_\odot\,\Omega_\odot}{4\,\pi} = \frac{I_\odot}{4\,\pi}\,\frac{\pi\,{\theta_\odot}^2}{4} = \frac{I_\odot}{16}\left(\frac{2\,R_\odot}{r_{\rm Mars}}\right)^2 \end{aligned} \tag{1.18}$$

where we have used Eq. (B.3) to express the solid angle in terms of the linear angle, and Eq. (B.1) to express the linear angle in terms of the size of the Sun and the distance of Mars. Inserting $I_\odot = 2.01 \times 10^{10}\,{\rm erg\,s^{-1}\,cm^{-2}\ sr^{-1}}$ (Sect. 1.3), $R_\odot = 6.96 \times 10^{10}$ cm, and $r_{\rm Mars} = 2.28 \times 10^{13}$ cm (Tables G.3, G.4), we find, $J = 4.7 \times 10^4\,{\rm erg\,s^{-1}\,cm^{-2}\ sr^{-1}}$. Then $u = \frac{4\,\pi}{c}\,(4.7 \times 10^4) = 2.0 \times 10^{-5}\ {\rm erg\ cm^{-3}}$.

The radiation field (u or J) in interstellar space due to randomly distributed stars (Prob. 1.10) must be computed over a solid angle of 4π steradians, given that

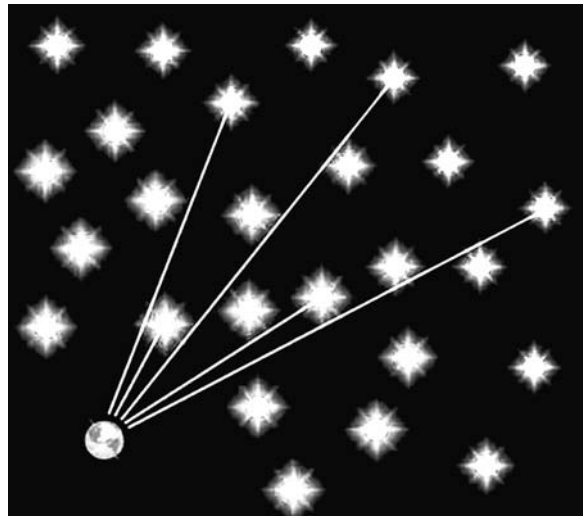


Figure 1.11. Why is the night sky dark? If the Universe is infinite and populated in all directions by stars, then eventually every sight line should intersect the surface of a star. Since I is constant with distance, the night sky should be as bright as the surface of a typical star. This is known as Olbers' Paradox, though Olbers was not the first to note this discrepancy. See Sect. 1.4.

starlight contributes from many directions in the sky. However, in this case, $J \neq I$ because there is no emission from directions between the stars. If there were so many stars that every line of sight eventually intersected the surface of a star of brightness, I_*, then $J = I_*$ and the entire sky would appear as bright as I_*. This would be true even if the stars were at great distances since I_*, being an intensity, is independent of distance. If this is the case, we would say that the stellar *covering factor* is unity.

A variant of this concept is called *Olbers' Paradox* after the German astronomer, Heinrich Wilhelm Olbers who popularized it in the 19th century. It was discussed as early as 1610, though, by the German astronomer, Johannes Kepler, and was based on the idea of an infinite starry Universe which had been propounded by the English astronomer and mathematician, Thomas Digges, around 1576. If the Universe is infinite and populated throughout with stars, then every line of sight should eventually intersect a star and the night sky as seen from Earth should be as bright as a typical stellar surface. Why, then, is the night sky dark?[6] Kepler took the simple observation of a dark night sky as an argument for the finite extent of the Universe, or at least of its stars. The modern explanation, however, lies with the

[6]The earlier form of the question was posed somewhat differently, referring to increasing numbers of stars on increasingly larger shells with distance from the Earth.

intimate relation between time and space on cosmological scales (Sect. 7.1). Since the speed of light is constant, as we look farther into space, we also look farther back in time. The Universe, though, is not infinitely old but rather had a beginning (Sect. 3.1) and the formation of stars occurred afterwards. The required number of stars for a bright night sky is $\approx 10^{60}$ and the volume needed to contain this quantity of stars implies a distance of 10^{23} light years (Ref. [74]). This means that we need to see stars at an epoch corresponding to 10^{23} years ago for the night sky to be bright. The Universe, however, is younger than this by 13 orders of magnitude (Sect. 3.1)! Thus, as we look out into space and back in time, our sight lines eventually reach an epoch prior to the formation of the first stars when the covering factor is still much less than unity. (Today, we refer to this epoch as *the dark ages*.) Remarkably, this solution was hinted at by Edgar Allan Poe in his prose-poem, *Eureka* in 1848 (see the prologue to this chapter).

1.5. How light pushes – radiation pressure

Radiation pressure is the momentum flux of radiation (the rate of momentum transfer due to photons, per unit area). It can also be thought of as the force per unit area exerted by radiation and, since force is a vector, we will treat radiation pressure in this way as well[7]. Thus, the pressure can be separated into its normal, $\mathrm{P}_\perp$, and tangential, $\mathrm{P}_\parallel$, components with respect to the surface of a wall. The normal radiation pressure will be,

$$d\mathrm{P}_\perp = \frac{d\mathrm{F}_\perp}{dA} = \frac{dp}{dt\,dA}\cos\theta = \frac{dE}{c\,dt\,dA}\cos\theta \tag{1.19}$$

where we have expressed the momentum of a photon in terms of its energy (Table I.1). Using Eq. (1.12) we obtain,

$$d\mathrm{P}_\perp = \left(\frac{1}{c}\right) I\,\cos^2\theta\,d\Omega \tag{1.20}$$

For the tangential pressure, we use the same development but take the sine of the incident angle, yielding,

$$d\mathrm{P}_\parallel = \left(\frac{1}{c}\right) I\,\cos\theta\sin\theta d\Omega \tag{1.21}$$

[7]Pressure is actually a *tensor* which is a mathematical quantity described by a matrix (a vector is a specific kind of tensor). We do not need a full mathematical treatment of pressure as a tensor, however, to appreciate the meaning of radiation pressure.

Then for isotropic radiation,

$$\mathrm{P}_{\perp} = \left(\frac{1}{c}\right) \int_{4\pi} I \cos^2\theta \, d\Omega = \frac{4\,\pi}{3\,c}\, I$$

$$\mathrm{P}_{\parallel} = \left(\frac{1}{c}\right) \int_{4\pi} I \cos\theta \sin\theta d\Omega = 0$$

$$\text{therefore} \qquad \mathrm{P} = \sqrt{\mathrm{P}_{\perp}{}^2 + \mathrm{P}_{\parallel}{}^2} = \frac{4\,\pi}{3\,c}\, I = \frac{1}{3}\, u \tag{1.22}$$

where we have used a spherical coordinate system for the integration (Appendix B), Eq. (1.16), and the fact that $J = I$ in an isotropic radiation field. Note that the units of pressure are equivalent to the units of energy density, as indicated in Table 0.A.2. Since photons carry momentum, the pressure is not zero in an isotropic radiation field. A surface placed within an isotropic radiation field will not experience a *net* force, however. This is similar to the pressure of particles in a thermal gas. There is no net force in one direction or another, but there is still a pressure associated with such a gas (Sect. 3.4.2).

We can also consider a case in which the incoming radiation is from a fixed angle, θ and the solid angle subtended by the radiation source, Ω, is small. This would result in an acceleration of the wall, but the result depends on the kind of surface the photons are hitting. We consider two cases, illustrated in Figure 1.12: that in which the photon loses all of its energy to the wall (perfect absorption) and that in which the photon loses none of its energy to the wall (perfect reflection).

For perfect absorption, integrating Eqs. (1.20), (1.21) with θ, Ω, constant, yields,

$$\mathrm{P}_{\perp} = \left(\frac{1}{c}\right) I\,\Omega\, \cos^2\theta = \frac{f}{c} \cos^2\theta$$

$$\mathrm{P}_{\parallel} = \left(\frac{1}{c}\right) I\,\Omega\, \cos\theta\, \sin\theta = \frac{f}{c} \cos\theta\, \sin\theta$$

$$\mathrm{P} = \sqrt{\mathrm{P}_{\perp}{}^2 + \mathrm{P}_{\parallel}{}^2} = \frac{f}{c} \cos\theta \tag{1.23}$$

where we have used Eq. (1.13) with f the flux along the directed beam[8].

For perfect reflection, only the normal component will have any effect against the wall (as if the surface were hit by a ball that bounces off). Also, because the momentum of the photon reverses direction upon reflection, the change in momentum is twice the value of the absorption case. Thus, the situation can be described by Eq. (1.20) except

[8]For a narrow beam, this is equivalent to the Poynting flux (Table I.1).

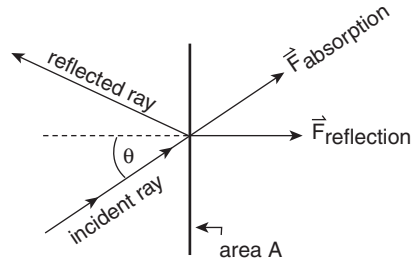


Figure 1.12. An incoming photon exerts a pressure on a surface area. For perfect absorption, the area will experience a force in the direction, $\vec{F}_{\text{absorption}}$. For perfect reflection, only the normal component of the force is effective and the resulting force will be in the direction, $\vec{F}_{\text{reflection}}$

for a factor of 2.

$$\mathrm{P} = \mathrm{P}_{\perp} = \left(\frac{2}{c}\right) I\,\Omega\,\cos^2\theta = \frac{2f}{c}\cos^2\theta \tag{1.24}$$

A comparison of Eqs. (1.23) and (1.24) shows that, provided the incident angle is not very large, a reflecting surface will experience a considerably greater radiation force than an absorbing surface. Moreover, as Figure 1.12 illustrates, the direction of the surface is *not* directly away from the source of radiation as it must be for the absorbing case. These principles are fundamental to the concept of a *Solar sail* (Figure 1.13).

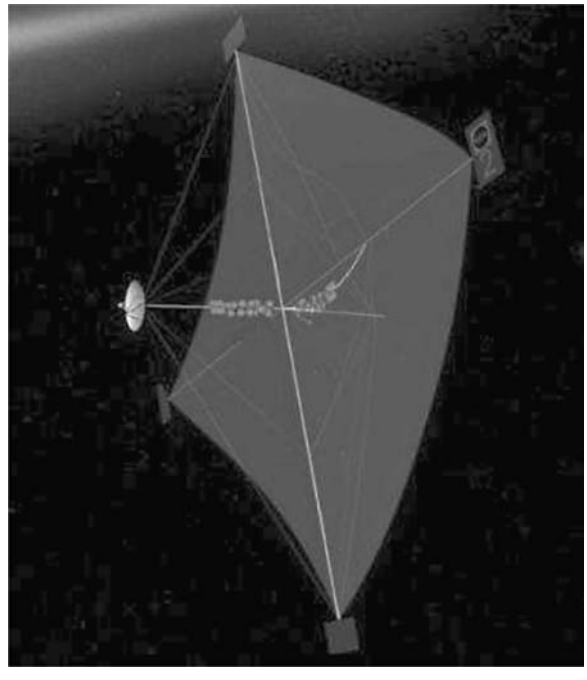


Figure 1.13. Artist's conception of a thin, square, reflective Solar sail, half a kilometre across. Reproduced by permission of NASA/MSFC

Since the direction of motion depends on the angle between the radiation source and the normal to the surface, it would be possible to 'tack' a Solar sail by altering the angle of the sail, in a fashion similar to the way in which a sailboat tacks in the wind. Moreover, even if the acceleration is initially very small, it is *continuous* and thus very high velocities could eventually be achieved for spacecraft designed with Solar sails (Prob. 1.11).

1.6 The human perception of light – magnitudes

Magnitudes are used to characterize light in the optical part of the spectrum, including the near IR and near UV. This is a logarithmic system for light, similar to decibels for sound, and is based on the fact that the response of the eye is logarithmic. It was first introduced in a rudimentary form by Hipparchus of Nicaea in about 150 B.C. who labelled the brightest stars he could see by eye as 'first magnitude', the second brightest as 'second magnitude', and so on. Thus began a system in which brighter stars have *lower* numerical magnitudes, a sometimes confusing fact. As the human eye has been the dominant astronomical detector throughout most of history, a logarithmic system has been quite appropriate. Today, the need for such a system is less obvious since the detector of choice is the CCD (charge coupled device, Sect. 2.2.2) whose response is linear. However, since magnitudes are entrenched in the astronomical literature, still widely used today, and well-characterized and calibrated, it is very important to understand this system.

1.6.1 Apparent magnitude

The *apparent magnitude* and its corresponding flux density are values as measured above the Earth's atmosphere or, equivalently, as measured from the Earth's surface, corrected for the effects of the atmosphere,

$$\begin{aligned} m_\lambda - m_{\lambda_0} &= -2.5 \log\left(\frac{f_\lambda}{f_{\lambda_0}}\right) \\ m_\nu - m_{\nu_0} &= -2.5 \log\left(\frac{f_\nu}{f_{\nu_0}}\right) \end{aligned} \tag{1.25}$$

where the subscript, 0, refers to a standard calibrator used as a reference, m_λ and m_ν are apparent magnitudes in some waveband and f_λ, f_ν are flux densities in the same band. Note that this equation could also be written as a ratio of fluxes, since this would only require multiplying the flux density (numerator and denominator) by an effective bandwidth to make this conversion. This system is a *relative* one, such that the magnitude of the star of interest can be related to that of *any* other star in the same waveband via Eq. (1.25). For example, if a star has a flux density that is 100 times greater

Table 1.1. Standard filters and magnitude calibration[a]

	U	B	V	R	I	J	H	K	L	L*
$\lambda_{\rm eff}^{b}$	0.366	0.438	0.545	0.641	0.798	1.22	1.63	2.19	3.45	3.80
$\Delta\lambda^{c}$	0.065	0.098	0.085	0.156	0.154	0.206	0.298	0.396	0.495	0.588
$f_{\nu_0}^{d}$	1.790	4.063	3.636	3.064	2.416	1.589	1.021	0.640	0.285	0.238
$f_{\lambda_0}^{e}$	417.5	632	363.1	217.7	112.6	31.47	11.38	3.961	0.708	0.489
ZP_ν	0.770	−0.120	0.000	0.186	0.444	0.899	1.379	1.886	2.765	2.961
ZP_λ	−0.152	−0.601	0.000	0.555	1.271	2.655	3.760	4.906	6.775	7.177

[a]UBVRIJHKL Cousins–Glass–Johnson system (Ref. [18]). The table values are for a fictitious A0 star which has 0 magnitude in all bands. A star of flux density, f_ν in units of $10^{-20}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}}$ or f_λ in units of $10^{-11}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$, will have a magnitude, $m_\nu = -2.5\log(f_\nu) - 48.598 - \mathrm{ZP}_\nu$ or $m_\lambda = -2.5\log(f_\lambda) - 21.100 - \mathrm{ZP}_\lambda$, respectively. [b]The effective wavelength, in μm, is defined by $\lambda_{\rm eff} = [\int \lambda f(\lambda) R_W(\lambda)\, d\lambda]/[\int f(\lambda) R_W(\lambda)\, d\lambda]$, where $f(\lambda)$ is the flux of the star at wavelength, λ, and $R_W(\lambda)$ is the response function of the filter in band W (see Figure 1.3). Thus, the effective wavelength varies with the spectrum of the star considered. [c]Full width at half-maximum (FWHM) of the filters in μm. [d]Units of $10^{-20}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}}$. [e]Units of $10^{-11}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$.

than a second star, then its magnitude will be 5 less than the second star. However, in order to assign a specific magnitude to a specific star, it is necessary to identify certain standard stars with known flux densities to which all others can be compared.

Several slightly different calibration systems have evolved over the years so, for careful and precise work, it is necessary to specify which system is being used when measuring or stating a magnitude. An example of such a system is the UBVRIJHKL Cousins–Glass–Johnson system for which parameters are provided in Table 1.1. The corresponding wavebands, U, B, V, etc., are illustrated in Figure 1.3. The apparent magnitude is commonly written in such a way as to specify these wavebands directly, e.g.

$$\begin{aligned} \mathrm{V} - \mathrm{V}_0 &= -2.5\log\left(\frac{f_\mathrm{V}}{f_{\mathrm{V}_0}}\right) \\ \mathrm{B} - \mathrm{B}_0 &= -2.5\log\left(\frac{f_\mathrm{B}}{f_{\mathrm{B}_0}}\right)\text{etc.} \end{aligned} \tag{1.26}$$

where the flux densities can be expressed in either their λ-dependent or ν-dependent forms. The V-band, especially, since it corresponds to the waveband in which the eye is most sensitive (cf. Table G.5), has been widely and extensively used. Some examples of apparent magnitudes are provided in Table 1.2.

The standard calibrator in most systems has historically been the star, Vega. Thus, Vega would have a magnitude of 0 in all wavebands (i.e. $\mathrm{U}_0 = 0$, $\mathrm{B}_0 = 0$, etc) and its flux density in these bands would be tabulated. However, concerns over possible variability of this star, its possible IR excess, and the fact that it is not observable from the Southern hemisphere, has led to modified approaches in which the star Sirius is also taken as a calibrator and/or in which a model star is used instead. The latter approach has been taken in Table 1.1 which lists the reference flux densities for

Table 1.2. Examples of apparent visual magnitudes[a]

Object or item	Visual magnitude	Comments
Sun	−26.8	
Approx. maximum of a supernova	−15	assuming V = 0 precursor
Full Moon	−12.7	
Venus	−3.8 to −4.5[b]	brightest planet
Jupiter	−1.7 to −2.5[b]	
Sirius	−1.44	brightest nighttime star
Vega	0.03	star in constellation Lyra
Betelgeuse	0.45	star in Orion
Spica	0.98	star in Virgo
Deneb	1.23	star in Cygnus
Aldebaran	1.54	star in Taurus
Polaris	1.97	the North Star[c]
Limiting magnitude[d]	3.0	major city
Ganymede	4.6	brightest moon of Jupiter
Uranus	5.7[e]	
Limiting magnitude[d]	6.5	dark clear sky
Ceres	6.8[e]	brightest asteroid
Pluto	13.8[e]	
Jupiter-like planet	26.5	at a distance of 10 pc
Limiting magnitude of HST[f]	28.8	1 h on A0V star
Limiting magnitude of OWL[g]	38	future 100 m telescope

[a]From Ref. [71] (probable error at most 0.03 mag) and on-line sources. [b]Typical range over a year. [c]A variable star. [d]This is the faintest star that could be observed by eye without a telescope. It will vary with the individual and conditions. [e]At or close to 'opposition' (180° from the Sun as seen from the Earth). [f]'Hubble Space Telescope', from Space Telescope Science Institute on-line documentation. The limiting magnitude varies with instrument used. The quoted value is a best case. [g]OWL (the 'Overwhelmingly Large Telescope') refers to the European Southern Observatory's concept for a 100 m diameter telescope with possible completion in 2020.

reference magnitudes of zero in all filters. The flux density and reference flux density must be in the same units. The above equations can be rewritten as,

$$\begin{aligned} m_\lambda &= -2.5\log(f_\lambda) - 21.100 - \mathrm{ZP}_\lambda \\ m_\nu &= -2.5\log(f_\nu) - 48.598 - \mathrm{ZP}_\nu \end{aligned} \tag{1.27}$$

where f_λ is in units of erg cm^{-2} s^{-1} Å^{-1}, f_ν is in units of erg cm^{-2} s^{-1} Hz^{-1}, and ZP_λ, ZP_ν are called *zero point* values (Ref. [18]). Example 1.4 provides a sample calculation.

Example 1.4

An apparent magnitude of $B = 1.95$ *is measured for the star, Betelgeuse. Determine its flux density in units of* erg cm^{-2} s^{-1} Å^{-1}.

From Eq. (1.26) and the values from Table 1.1, we have,

$$\mathrm{B} - \mathrm{B}_0 = 1.95 - 0 = -2.5 \log\left(\frac{f_\mathrm{B}}{632 \times 10^{-11}}\right) \tag{1.28}$$

Solving, this gives $f_\mathrm{B} = 1.0 \times 10^{-9}$ erg cm^{-2} s^{-1} Å^{-1} for Betelgeuse.

Eq. (1.27) can also be used,

$$\mathrm{B} = 1.95 = -2.5 \log(f_\mathrm{B}) - 21.100 + 0.601 \tag{1.29}$$

which, on solving, gives the same result.

1.6.2 Absolute magnitude

Since flux densities fall off as $\frac{1}{r^2}$, measurements of apparent magnitude between stars do not provide a useful comparison of the intrinsic properties of stars without taking into account their various distances. Thus the *absolute magnitude* has been introduced, either as a bolometric quantity, M_bol, or in some waveband (e.g. M_V, M_B, etc). The absolute magnitude of a star is the magnitude that would be measured if the star were placed at a distance of 10 pc. Since the magnitude scale is relative, we can let the reference star in Eq. (1.25) be the *same* star as is being measured but placed at a distance of 10 pc,

$$m - M = -2.5 \log\left(\frac{f}{f_{10\,\mathrm{pc}}}\right) = -5 + 5 \log\left(\frac{d}{\mathrm{pc}}\right) \tag{1.30}$$

where we have dropped the subscripts for simplicity and used Eq. (1.9). Here d is the distance to the star in pc. Eq. (1.30) provides the relationship between the apparent and absolute magnitudes for any given star. The quantity, $m - M$, is called the *distance modulus*. Since this quantity is directly related to the distance, it is sometimes quoted as a proxy for distance. Writing a similar equation for a reference star and combining with Eq. (1.30) (e.g. Prob. 1.12) we find,

$$M - M_{b\odot} = -2.5 \log\left(\frac{L}{L_\odot}\right) \tag{1.31}$$

where we have used the Sun for the reference star. Eq. (1.31) has been explicitly written with bolometric quantities (Table G.3) but one could also isolate specific bands, as before, provided the correct reference values are used.

1.6.3 The colour index, bolometric correction, and HR diagram

The *colour index* is the difference between two magnitudes in different bandpasses for the same star, for example,

$$\mathrm{B}-\mathrm{V}=-2.5\log\left(\frac{f_{\mathrm{B}}}{f_{\mathrm{V}}}\frac{f_{\mathrm{V}_0}}{f_{\mathrm{B}_0}}\right)=-2.5\log\left(\frac{f_{\mathrm{B}}}{f_{\mathrm{V}}}\right)-(\mathrm{ZP}_{\mathrm{B}}-\mathrm{ZP}_{\mathrm{V}}) \tag{1.32}$$

or between any other two bands. Eq. (1.32) is derivable from Eqs. (1.26) or (1.27). Various colour indices are provided for different kinds of stars[9] in Table G.7. Since this quantity is basically a measure of the ratio of flux densities at two different wavelengths (with a correction for zero point), it is an indication of the *colour* of the star. A positive value for B − V, for example, means that the flux density in the V band is *higher* than that in the B band and hence the star will appear more 'yellow' than 'blue' (see Table G.5).

The colour index, since it applies to a single star, is *independent of distance*. (To see this, note that converting the flux density to a distance-corrected luminosity would require the same factors in the numerator and denominator of Eq. (1.32)). Consequently, the colour index can be compared directly, star to star, without concern for the star's distance. We will see in Sect. 4.1.3 that colour indices are a measure of the surface temperature of a star. This means that stellar temperatures can be determined without having to know their distances.

Since a colour index could be written between any two bands, one can also define an index between one band and *all* bands. This is called the *bolometric correction*, usually defined for the V band,

$$\mathrm{BC}=m_{\mathrm{bol}}-\mathrm{V}=M_{\mathrm{bol}}-M_{\mathrm{V}} \tag{1.33}$$

For any given star, this quantity is a correction factor that allows one to convert from a V band measurement to the bolometric magnitude (Prob. 1.16). Values of BC are provided for various stellar types in Table G.7.

Figure 1.14 shows a plot of absolute magnitude as a function of colour index for over 5000 stars in the Galaxy near the Sun. Such a plot is called a *Hertzsprung–Russell (HR) diagram* or a *colour–magnitude diagram (CMD)*. The absolute magnitude can be converted into luminosity (see Eq. 1.31) and the colour index can be converted to a temperature (see the calibration of Figure G.1) which are more physically meaningful parameters for stars. The HR Diagram is an essential tool for the study of stars and stellar evolution and shows that stars do *not* have arbitrary temperatures and luminosities but rather fall along well-defined regions in $L-\mathrm{T}$ parameter space. The positions of stars in this diagram provide important information about stellar parameters and stellar evolution, as described in Sect. 3.3.2. It is important to note that the distance to each star must be measured in order to obtain absolute magnitudes, a feat that has been accomplished to unprecedented accuracy by Hipparcos, a European satellite launched

[9] Stellar spectral types will be discussed in Sect. 3.4.7.

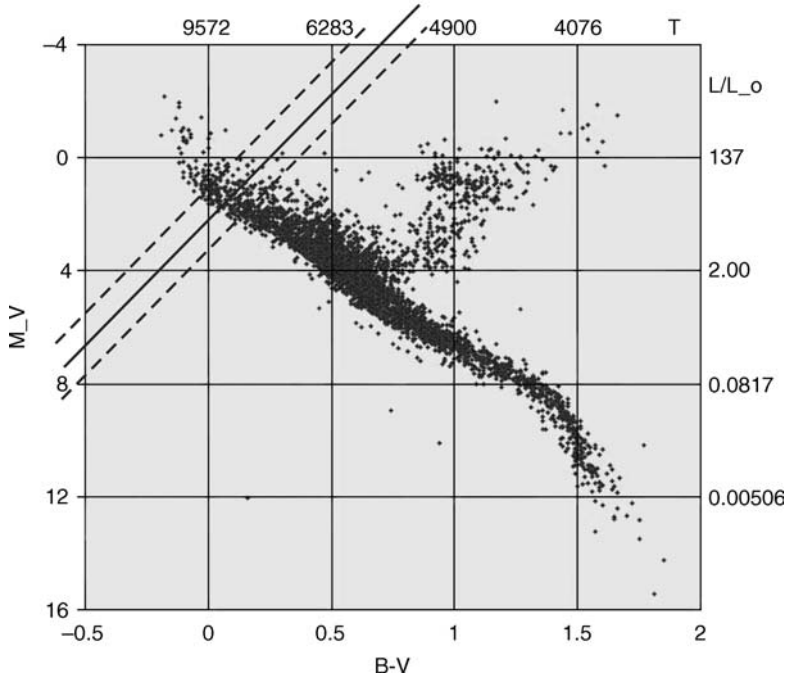


Figure 1.14. Hertzsprung–Russell (HR) diagram for approximately 5000 stars taken from the Hipparcos catalogue. Hipparcos determined the distances to stars, allowing absolute magnitude to be determined. The M_V data have errors of about ± 0.1 magnitudes and the $B - V$ data have errors of < 0.025 magnitudes. Most stars fall on a region passing from upper left (hot, luminous stars) to lower right (dim, cool stars), called the *main sequence*. The main sequence is defined by stars that are burning hydrogen into helium in their cores (see Sect. 3.3.2). The temperature, shown at the top and applicable to the main sequence, was determined from the Temperature $-B - V$ calibration of Table G.7 and Figure G.1. The luminosity, shown at the right, was determined from Eqs. (1.31) and (1.33), using the bolometric correction from Table G.7. The straight line, defined by the equation, $M_V = -8.58(B - V) + 2.27$, locates the central ridge of the instability strip for Galactic Cepheid variable stars and the dashed lines show its boundaries (see Ref. [166a] and Sect. 5.4.2).

in 1989. The future GAIA (Global Astrometric Interferometer for Astrophysics) satellite, also a European project with an estimated launch date of 2011, promises to make spectacular improvements. This satellite will provide a census of 10^9 stars and it is estimated that it will discover 100 new asteroids and 30 new stars that have planetary systems *per day*.

1.6.4 Magnitudes beyond stars

Magnitudes are widely used in optical astronomy and, though the system developed to describe stars (for which specific intensities cannot be measured, Sect. 1.3), it can be applied to any object, extended or point-like, as an alternate description of the flux

density. One could quote an apparent magnitude for other point-like sources such as distant QSOs ('quasi-stellar objects')[10], or of extended sources like galaxies. Since the specific intensity of an extended object is a flux density per unit solid angle, it is also common to express this quantity in terms of magnitudes per unit solid angle (Prob. 1.17).

1.7 Light aligned – polarization

The magnetic and electric field vectors of a wave are perpendicular to each other and to the direction of propagation (Figure I.2). A signal consists of many such waves travelling in the same direction in which case the electric field vectors are usually randomly oriented around the plane perpendicular to the propagation direction. However, if all of the electric field vectors are aligned (say all along the z axis of Figure I.2) then the signal is said to be polarized. Partial polarization occurs if some of the waves are aligned but others randomly oriented. The *degree of polarization*, $D_{\rm p}$ is defined as the fraction of total intensity that is polarized (often expressed as a percentage),

$$D_{\rm p} \equiv \frac{I_{\rm pol}}{I_{\rm tot}} = \frac{I_{\rm pol}}{I_{\rm pol} + I_{\rm unpol}} \tag{1.34}$$

Polarization can be generated internally by processes intrinsic to the energy generation mechanism (see Sect. 8.5, for example), or polarization can result from the scattering of light from particles, be they electrons, atoms, or dust particles (see Sect. 5.1 or Appendix D). When polarization is observed, $D_{\rm p}$ is usually of order only a few percent and 'strong' polarization, such as seen in radio jets (Figure 1.4), typically is $D_{\rm p} \lesssim 15$ per cent (Ref. [175]). Values of $D_{\rm p}$ over 90 per cent, however, have been detected in the jets of some pulsars (Ref. [113]) and in the low frequency radio emission from planets (Sect. 8.5.1). For the Milky Way, $D_{\rm p} = 2$ per cent over distances 2 to 6 kpc from us due to scattering by dust (Ref. [175]). In practical terms, this means that polarized emission is usually much fainter than unpolarized emission and requires greater effort to detect.

Problems

1.1 Assuming Cas A (see Sect. 1.1 for more data) has a spectral luminosity between $\nu_1 = 2 \times 10^7$ Hz and $\nu_2 = 2 \times 10^{10}$ Hz, of the form, $L_\nu = K\,\nu^{-0.7}$, where K is a constant. Determine the value of K, and of L_ν and L_λ at $\nu = 10^9$ Hz. What are the units of K?

1.2 Find, by comparison with exact trigonometry, the angle, θ (provide a numerical value in degrees), above which the small angle approximation, Eq. (B.1), departs from the exact

[10] A QSO is the bright active nucleus of a very distant galaxy that looks star-like at optical wavelengths. QSOs that also emit strongly at radio wavelengths are called *quasars*.

result by more than 1 per cent. How does this compare with the relatively large angle subtended by the Sun?

1.3 Determine the flux density (in Jy) of a cell phone that emits 2 mW cm^{-2} at a frequency of 1900 MHz over a bandwidth of 30 kHz, and of the Sun, as measured at the Earth, at the same frequency. (Eq. 4.6 provides an expression for calculating the specific intensity of the Sun.) Compare these to the flux density of the supernova remnant, Cas A ($\sim$ 1900 Jy as measured at the Earth at 1900 MHz) which is the strongest radio source in the sky after the Sun. Comment on the potential of cell phones to interfere with the detection of astronomical signals.

1.4 (a) Consider a pulsar with radiation that is beamed uniformly into a circular cone of solid angle, Ω. Rewrite the right hand side (RHS) of Eq. (1.9) for this case.

(b) If $\Omega = 0.02$ sr, determine the error that would result in L if the RHS of Eq. (1.9) were used rather than the correct result from part (a).

1.5 Determine the percentage variation in the solar flux incident on the Earth due to its elliptical orbit. Compare this to the variation shown in Figure 1.6.

1.6 Determine the flux in a perfectly isotropic radiation field (i.e. I constant in all directions).

1.7 (a) Determine the flux and intensity of the Sun (i) at its surface, (ii) at the mean distance of Mars, and (iii) at the mean distance of Pluto.

(b) How large (in arcmin) would the Sun appear in the sky at the distances of the two planets? Would it appear resolved or as a point source to the naked eye at these locations? That is, would the angular diameter of the Sun be larger than the resolution of the human eye (Table G.5) or smaller? (Sects. 2.2, 2.2.3, and 2.3.2 provide more information on the meaning of 'resolution'.)

1.8 The radio spectrum of Cas A, whose image is shown in Figure 1.2, is given in Figure 8.14 in a log–log plot. The plotted specific intensity can be represented by, $I_\nu = I_{\nu 0}\,(\nu/\nu_0)^\alpha$ in the part of the graph that is declining with frequency, where ν_0 is any reference frequency in this part of the plot, $I_{\nu 0}$ is the specific intensity measured at ν_0 and α is the slope. In Prob. 1.1, we assumed that $\alpha = -0.7$. Now, instead, measure this value from the graph and determine, for the radio band from $\nu_1 = 2 \times 10^7$ to $\nu_2 = 2 \times 10^{10}$ Hz,

(a) the intensity of Cas A, I,

(b) the solid angle that it subtends in the sky, Ω,

(c) its flux, f,

(d) its radio luminosity, $L_{\rm rad}$. Confirm that this value is approximately equal to the value given in Sect. 1.1.

1.9 On average, the brightness of the Whirlpool Galaxy, M 51 (see Figure 3.8 or 9.8), which subtends an ellipse of major × minor axis, $11.2' \times 6.9'$ in the sky, is 2.1 times that of the Andromeda Galaxy, M 31 (subtending $190' \times 60'$). What is the ratio of their flux densities?

1.10 Where does the Solar System end? To answer this, find the distance (in AU) at which the radiation energy density from the Sun is equivalent to the ambient mean energy density of interstellar space, the latter about 10^{-12} erg cm^{-3}. After 30 years or more of space travel, how far away are the Pioneer 10 and Voyager 1 spacecraft? See `http://spaceprojects.arc.nasa.gov/Space_Projects/pioneer/PNhome.html` and `http://voyager.jpl.nasa.gov`. Are they out of the Solar System?

1.11 Consider a circular, perfectly reflecting Solar sail that is initially at rest at a distance of 1 AU from the Sun and pointing directly at it. The sail is carrying a payload of 1000 kg (which dominates its mass) and its radius is $R_s = 500$ m.

(a) Derive an expression for the acceleration as a function of distance, $a(r)$, for this Solar sail. Include the Sun's gravity as well as its radiation pressure. (The constants may be evaluated to simplify the expression.)

(b) Manipulate and integrate this equation to find an expression for the velocity of the Solar sail as a function of distance, $v(\mathrm{r})$. Evaluate the expression to find the velocity of the Solar sail by the time it reaches the orbit of Mars.

(c) Finally, derive an expression for the time it would take for the sail to reach the orbit of Mars. Evaluate it to find the time. Express the time as seconds, months, or years, whatever is most appropriate.

1.12 Derive Eq. (1.31) (see text).

1.13 Repeat *Example 1.4* but expressing the flux density in its frequency-dependent form.

1.14 For the U band and L* band filters, verify that f_ν corresponds to f_λ in Table 1.1. Why might there be minor differences?

1.15 Refer to Table 1.2 for the following.

(a) Determine the range (ratio of maximum to minimum flux density) over which the unaided human eye can detect light from astronomical objects. Research the range of human hearing from 'barely audible' to the 'pain threshold' and compare the resulting range to the eye.

(b) Use the 'Multiparameter Search Tool' of the *Research Tools* at the web site of the Hipparcos satellite (`http://www.rssd.esa.int/Hipparcos/`) to determine what percentage of stars in the night sky one would lose by moving from a very dark country site into a nearby light polluted city.

(c) The star, Betelgeuse, is at a distance of 130 pc. Determine how far away it would have to be before it would be invisible to the unaided eye, if it were to undergo a supernova explosion.

1.16 A star at a distance of 25 pc is measured to have an apparent magnitude of $V = 7.5$. This particular type of star is known to have a bolometric correction of $BC = -0.18$. Determine the following quantities: (a) the flux density, f_V in units of erg cm^{-2} s^{-1} Å^{-1}, (b) the absolute V magnitude, M_V, (c) the distance modulus, (d) the bolometric apparent and absolute magnitudes, m_{bol}, and M_{bol}, respectively, and (e) the luminosity, L in units of $L_\odot$.

1.17 A galaxy of uniform brightness at a distance of 16 Mpc appears elliptical on the sky with major and minor axis dimensions, $7.9' \times 1.4'$. It is observed first in the radio band centred at 1.4 GHz (bandwidth = 600 MHz) to have a specific intensity of 4.8 mJy beam^{-1}, where the 'beam' is a circular solid angle of diameter, $15''$ (see also Example 2.3d). A measurement is then made in the optical B band of 22.8 magnitudes per pixel, where the pixel corresponds to a square on the sky which is one arcsecond on a side. Determine (all in cgs units) I_ν, f_ν, f, and L of the galaxy in each band. In which band is the source brighter? In which band is it more luminous?

1.18 The limiting magnitude of some instruments can be pushed fainter by taking extremely long exposures. Estimate the limiting magnitude of the Hubble Ultra-Deep Field from the information given in Figure 3.1.

2

Measuring the Signal

All these facts were discovered . . . with the aid of a spyglass which I devised, after first being illuminated by divine grace.

–Galileo Galilei in *The Starry Messenger*

2.1 Spectral filters and the panchromatic universe

The first astronomical instrument was the human eye. From prehistoric times until today, human beings have surveyed the heavens with this most elegant and effective 'telescope'. The eye, like any other instrument, acts like a filter, accepting light at some wavelengths and filtering out others. The *spectral response function* of the human eye is shown in Figure 2.1 (see also Table G.5) and illustrates the relative ability of the eye to detect light at different wavelengths in the visual part of the electromagnetic spectrum[1]. If an incoming signal were uniformly bright across all wavelengths, the eye would not perceive it that way, but rather would see light in the range 500–550 nm (depending on conditions) as the brightest. Outside of this relatively narrow visual band, the eye is unable to detect a signal at all.

Attempts have been made to improve upon the human eye for at least 700 years. Early versions of eyeglasses, for example, date to 1284 or 1285 AD and possibly earlier in a more primitive fashion. However, it was not until Galileo first turned a telescope on the Moon, the Sun and planets in the early 17th Century, that combinations of lenses were put towards the purpose of astronomical observation. Since Galileo, we have enjoyed 400 years of increasingly larger and more sophisticated optical telescopes. By contrast, other wavebands were not even known to exist until

[1] 'Visual' will be used to indicate the part of the spectrum to which the human eye is sensitive, whereas 'optical' will refer to the part of the spectrum to which ground-based optical telescopes are sensitive; the latter includes the near-IR and the near-UV.

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

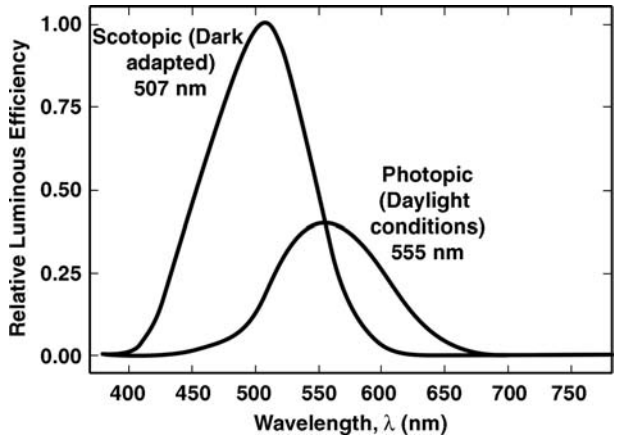


Figure 2.1. Response function of the human eye for photopic (daylight) and scotopic (dark-adapted) conditions. The peak wavelengths are indicated. The dark-adapted response is dominated by rods in the eye's retina whereas daylight vision is dominated by the cones. Sensitivity to light is much higher under dark-adapted conditions, as shown, but the ability to distinguish colour is markedly diminished. Related information can be found in Table G.5

the discovery of the infrared by William Herschel in 1800. Once it was understood that there could be emission invisible to the human eye (see Table G.6), it was only a matter of time before a variety of developing technologies provided the means to detect and form images in non-visual wavebands. The technology to detect radio signals, for example, matured rapidly after World War II. In a sense, such technologies provide us with eyeglasses to enhance the response of the human eye 'sideways' into other wavebands.

One other important constraint inhibited the discovery of far-IR, UV, X-ray and γ-ray emission from astronomical objects until the latter-half of the 20th Century – that of the Earth's atmosphere. The spectral response function of the atmosphere is shown in Figure 2.2 and reveals that only in the optical and radio wavebands is the atmosphere sufficiently transparent for ground-based observations. The 1957 launch of Sputnik by the former Soviet Union marked the dawn of the space-age and, with it, a rich era of astronomical discovery. This era continues today, with space-based probes realising increasingly sensitive and highly detailed images of astronomical objects at the previously inaccessible 'invisible' wavelengths.

Thus, the history of astronomy has largely been driven by two spectral response functions: that of the human eye and that of the Earth's atmosphere. Our knowledge of the Universe is still based on information obtained through a strong optical and ground-based bias. In 2005, for example, there were 45 optical research telescopes greater than 2.4 m in diameter and approximately 17 more in either the construction or planning stages. Only one of these (the Hubble Space Telescope) was not ground based. In the same year, about 17 space-based astronomical telescopes were in operation with another 17 planned or proposed. The amount and quality of non-optical and

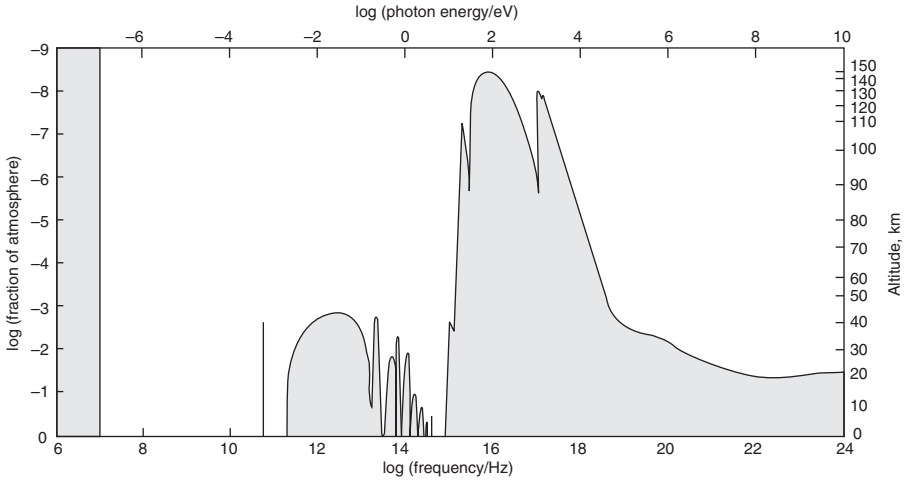


Figure 2.2. Atmospheric transparency curve for the Earth. The altitude (right hand side) indicates how high one would have to put a telescope in order for the atmosphere to be transparent at that frequency. This is expressed as a fraction of the total atmosphere on the left hand side. In this plot, wavelength decreases to the right in contrast to Figure 2.1. The two atmospheric windows are in the radio and optical parts of the spectrum. Most atmospheric absorption is due to water vapour (H_2O), then carbon dioxide (CO_2), and then ozone (O_3). The single spike around 60 GHz is due to oxygen (O_2). Adapted from Ref. [101]

space-based data are rapidly increasing, however[2], and are revealing aspects of our Universe never before seen, or even contemplated.

Why is such a concerted effort required over so many wavebands? Figure 2.3, which shows the spectrum of the galaxy, NGC 2903, provides a good illustration. This rather normal spiral galaxy (see Figure 2.11 for an optical image) emits over a wavelength range spanning ten orders of magnitude, and probably more if the observations were available. A wealth of information is hidden in this plot with each band revealing new and different insights about the source. If an image were taken at each waveband, the images would not be identical to each other. Further clues about the system are uncovered by this changing appearance with waveband. It is clearly impossible to truly understand the galaxy with observations in the optical band alone – an approach is required that is 'panchromatic', or spanning all wavebands. If we now consider the wide range of objects one could observe in the Universe besides normal galaxies, the imperative to widen our waveband horizons becomes even greater.

Astronomy is now in a discovery era, driven by technological improvements in the telescope and its concomitant instruments. It is therefore essential to have some understanding of how this technology, whether space-based or under the blanket of

[2]For a good sampling of the number of ground-based and space-based facilities available, see the website of the Royal Astronomical Society http://www.ras.org.uk under *astrolinks*.

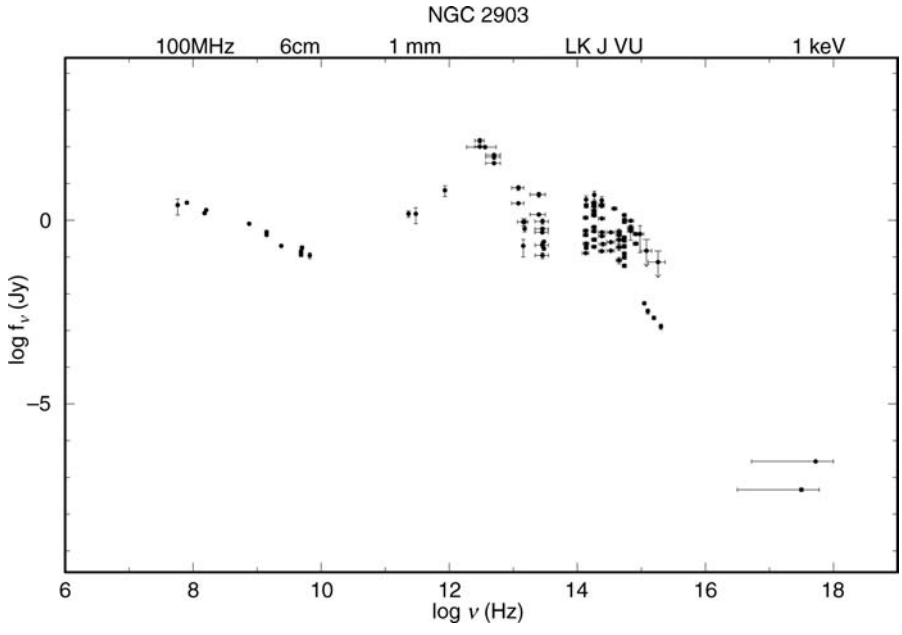


Figure 2.3. Spectrum of the galaxy, NGC 2903 (Ref. [105]). The frequency is shown at the bottom and some specific wavebands are labelled at the top. See Sect. 10.1.1 for a discussion of this spectrum. (Adapted from the NASA/IPAC Extra galactic Database)

our atmosphere, collects, filters, distorts, and ultimately reveals the secrets of the astronomical sources that so intrigue us.

2.2 Catching the signal – the telescope

Although the technology required to collect light and the quality of the result vary enormously across wavebands, the basic principles are the same: the job of a telescope is to collect as much light as possible and to focus it on a detector[3], forming an image, if possible. For scientific purposes, there are two main ingredients in this process: a surface that collects and focusses the light called the *objective* (also called the primary lens, or mirror or antenna, depending on what is used), and a device (the *detector*) that detects and converts the received signal into some form, usually digital, for storage and analysis via computer. This is analogous to what the human eye does. The pupil is the opening through which light can pass, acting as the collector, focussing is achieved by the lens and the interior fluid, and the detector is the retina with its plethora of rods and cones. The eye's detector then converts the light into electrical signals which are sent to the brain to be analysed. A telescope, therefore, acts rather like a giant eye.

[3]Focussing is not possible, however, at the highest γ-ray energies (see Figure 2.7.b).

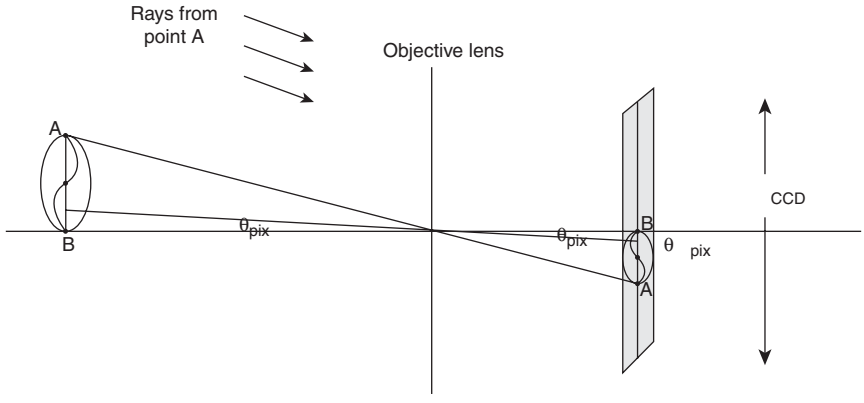


Figure 2.4. Diagram of a simple optical telescope, showing the relationships between the focal length, the linear scale on the detector, and the angular scale in the sky. Note that the angles have been exaggerated. The rays from a very distant point, A, all impinge upon the objective parallel to each other and at the same angle. The same is true of the rays from point B. The rays shown are those passing through the centre of the lens which experience no net bending due to lens symmetry. If visual observing is desired, it would be necessary to place a lens (called the *ocular* or *eyepiece*) in the focal plane of the telescope instead of the detector. The ocular would then bend the light again so that a human being can view the image.

Figure 2.4 provides a diagram of a simple telescope that uses a lens as the objective. However, the results, below, are the same for a reflecting instrument. Since astronomical signals are at a great distance, the incoming wavefronts from a source at any angle in the sky will be plane parallel. This means that light rays that trace the direction of the wave front are also parallel to each other, and are intercepted by the entire aperture, as shown. It also means that the position of the focussed source behind the lens is the actual focal length of the lens, f, and this is where the detector must therefore be placed. The geometry of the diagram shows how the angular size of any object on the sky, θ, is related to a linear size on the detector, l,

$$\theta \approx \tan\ \theta\ = \frac{l}{f} \tag{2.1}$$

where θ (radians) is presumed small. Thus telescopes with longer focal lengths result in larger images on the detector. Since f is a constant for any telescope, so is the quantity, θ/l, sometimes called the *plate scale* for optical instruments.

Telescopes are designated by their diameters and *focal ratios*, the latter being the ratio of the focal length to the diameter of the objective. For example, a 3 m f/5 telescope would be 3 m in diameter and have a focal length of $f = 5 \times 3 = 15$ m. Bigger telescopes can detect fainter signals (Sect. 2.2.1) and also result in images with better clarity or finer detail, also called higher *resolution* (Sect. 2.2.4). However, the latter is subject to other constraints which will be outlined in Sects. 2.2.3, and 2.3.2.

2.2.1 Collecting and focussing the signal

At optical wavelengths, collecting and focussing light is done via either a *lens* (refracting the light) or *curved mirror* (reflecting the light). For everything else equal, larger collecting areas result in brighter images on the detector because a larger area will intercept more of the source luminosity. Since astronomical sources are typically very faint, telescopes are built as large as costs and mechanical structures allow. Large optical telescopes used for research always use mirrors, since they require finishing only on one side and can be supported more easily with the mirror at the bottom of the telescope tube. Moreover, since reflection affects all wavelengths of optical light the same way whereas refraction bends light differently for different wavelenghts (e.g. see the Appendix to this chapter) the use of mirrors avoids certain *aberrations* associated with lenses[4]. Since the light then reflects upwards, either a detector must be placed at the focal point high up in the tube, called the *prime focus*, or else another *secondary mirror* must be used to reflect the signal back down again to a detector near the bottom (Figure 2.5). If the incoming

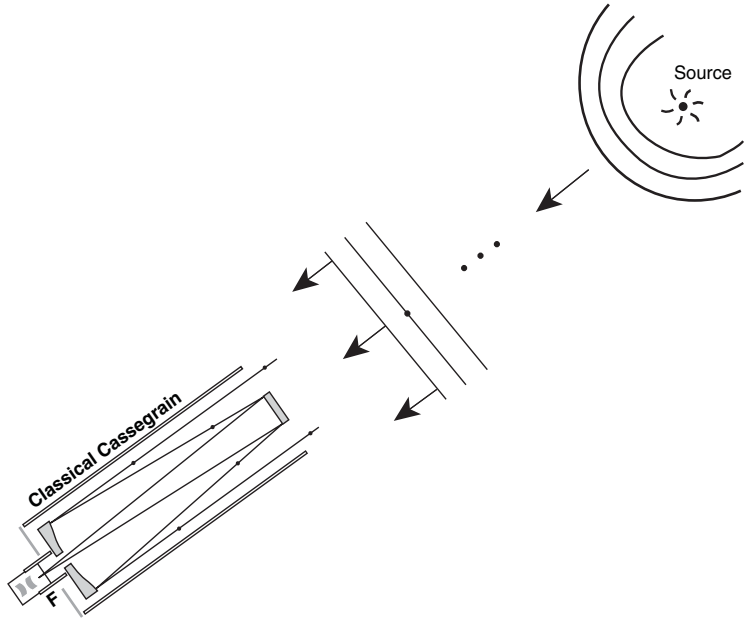


Figure 2.5. A distant point source emits its light in all directions, but the wavefront is plane parallel by the time it reaches the telescope. Rays denote the direction of the incoming plane parallel wavefront, in this case from a point source 'on-axis', i.e. at the centre of the field of view. This telescope shows a typical design in which curved mirrors collect and focus the light onto a detector behind the primary mirror. Such a design may have different names depending on the curvature of the mirrors and whether additional corrective lenses are used

[4]In particular, lenses are subject to *chromatic aberration* because of wavelength-dependent refraction, which puts the focal point of blue light closer to the lens than the focal point of red light.

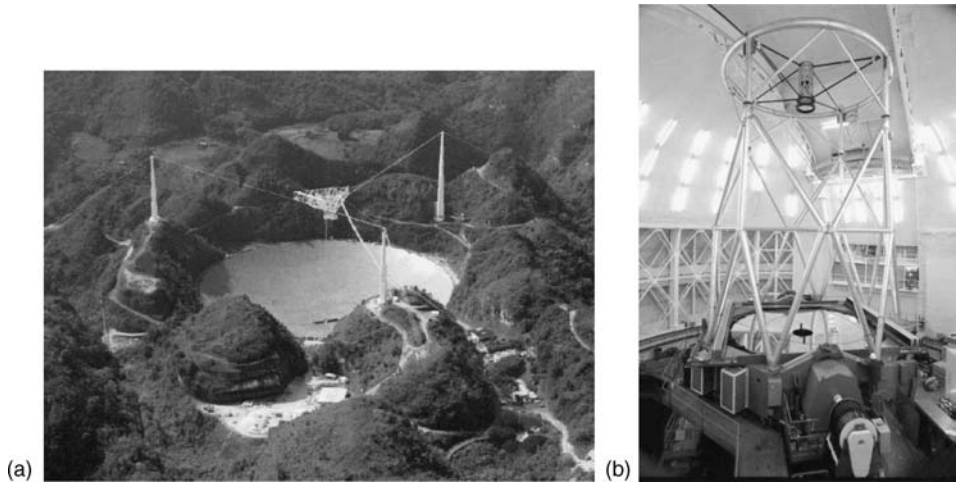


Figure 2.6. **(a)** The Arecibo Telescope, which operates at radio wavelengths, has a primary reflector 305 m in diameter. Suspended approximately 137 m above the centre of the dish at its prime focus is a platform containing several more small reflectors and the detectors. (Photo courtesy of the NAIC – Arecibo Observatory, a facility of the NSF.) **(b)** A view of Gemini North, an 8.1 m diameter optical telescope situated on Mauna Kea in Hawaii. Light coming down the open tube reflects back from the primary mirror to a secondary mirror which reflects the light again through a hole in the primary to the detectors below. The effective focal length is 128.12 m. (Reproduced by permission of the Gemini observatory/AURA)

light changes direction via mirrors, the light path is said to be *folded*. Modern detection instruments can be rather massive and so are usually placed at or near the bottom of the telescope. A variety of light paths are possible, however, including those with secondary mirrors at positions that do not block the aperture and those with detectors on platforms adjacent to the telescope tube.

Single-dish radio telescopes, infrared, optical, and ultraviolet telescopes all operate similarly, with collection and focussing achieved via a primary reflector. Figure 2.6 shows a radio and optical telescope example.

To maintain the integrity of the signal, irregularities on the surface of the primary reflector should be a small fraction (e.g. 1/10th if possible) of a wavelength. Thus optical telescopes require finely engineered smooth mirrors whereas a wire mesh may suffice at radio wavelengths. At X-ray wavelengths, however, even very smooth mirrors have rather limited focussing capabilities since X-rays tend to be absorbed, rather than reflected, when hitting a mirror perpendicularly. Instead X-rays are focussed via a series of glancing reflections (Figure 2.7.a)[5]. At even higher energies, γ-rays cannot be focussed at all and other techniques, such as tracing the signal direction through multiple detectors, must be used to determine the direction in the sky from which the light has come. (Figure 2.7.b).

[5]For X-ray telescopes, the ability of the telescope to focus onto a detector is what limits the spatial resolution.

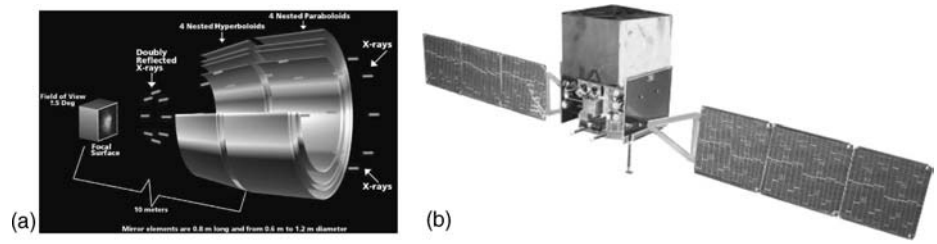


Figure 2.7. **(a)** A schematic view of the space-based Chandra X-ray telescope, showing how successive grazing reflections are used to focus the light onto a detector. (Reproduced by permission of NASA/CXC/SAO) **(b)** Artist's drawing of the Gamma-ray Large Area Space Telescope (GLAST). This telescope is basically a large cube in space with no mirrors or tubes. The detector itself has built-in capabilities to measure the directions and energies of incident gamma rays. GLAST is a joint European–American–Japanese project with a proposed launch date in 2007. (Reproduced by permission of NASA E/PO, Sonoma State University)

2.2.2 Detecting the signal

The detector is a device for turning the collected light into another form that can be analysed. A wide variety of detectors are used depending on waveband and desired output: photographic films, radio receivers, microchannel plates[6], CCDs (see below) and many others.

The simplest detector is one that accepts a signal from a single position on the sky. The telescope is pointed at a position on the sky and the detector records values (usually a spectrum, that is, the emission as a function of wavelength or frequency) that correspond to only this one position. If a map of an extended source is desired, the telescope must re-point to another position on the source, the detector makes a recording again, and so on until a picture of the source is built up consecutively. Many radio and mm-wave telescopes, which have complex receiver systems, still work on this principle. However, this process is slow and, wherever possible, has been replaced by imaging detectors which consist of many 'picture elements' or *pixels*. In general, imaging detectors are called *focal plane arrays*.

For optical work, an example of a focal plane array is the photographic plate which has been widely used in the past and is still sometimes used today, especially when very large fields of view are desired or if images, rather than numerical scientific results, are desired. However, the photograph has been virtually universally replaced for scientific work by the *Charge Coupled Device*, or *CCD* (Figure 2.8.a), which is a semi-conductor detector (a 'chip') like those used in digital cameras. The CCD's high sensitivity to low light levels, its linear response to increasing light levels[7], and its direct interface to computers make it the ideal detector for most scientific purposes. Larger format CCDs are also now becoming available, making earlier limitations of small fields of view less problematic. In each pixel of the CCD, electrons are released when exposed to light in

[6]These are detection and amplifying devices used to convert a single high energy photon, such as an X-ray, into many electrons at the output.
[7]Any given pixel, however, will become *saturated* if exposed to too high a level.

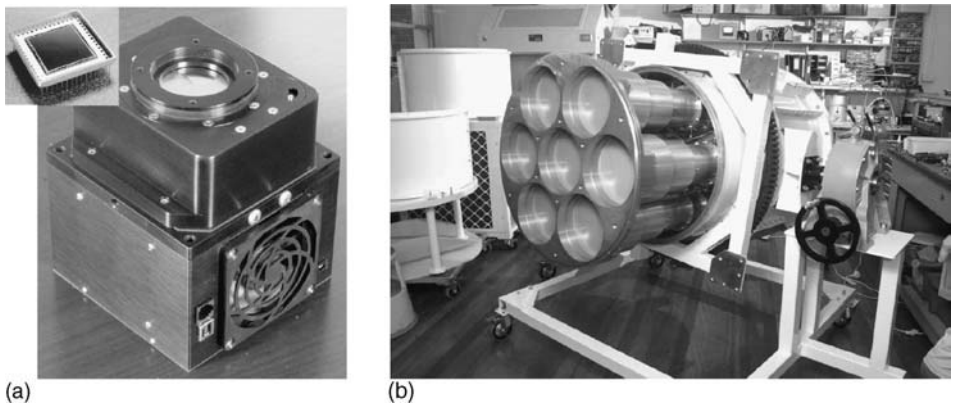


Figure 2.8 **(a)** A CCD camera like this one can be affixed to the back of an optical telescope to obtain images. Most of the weight of the camera is in supportive systems such as the cooling system. The CCD itself (upper left inset) is quite small. This particular CCD is 2.46 cm on a side with 1024×1024 pixels in the array, each of which is square and $24\,\mu$m on a side. (Reproduced by permission of Finger Lakes Instrumentation) **(b)** By contrast, the Arecibo L-Band Feed Array (ALFA) weighs 600 kg and fills a small laboratory. When placed at the focal plane of the telescope, its seven circular pixels of diameter 25 cm sample seven nearby regions on the sky with small gaps in between. (Reproduced by permission of CSIRO Australia. Credit: David Smyth)

some energy range for which the CCD has been designed. This is essentially the *photoelectric effect*. The number of electrons released is proportional to the number of photons incident. The electrons remain at the location of the pixel during the exposure and are read out at the end, line by line, via an applied voltage. The information is then stored on the computer in a file that represents it as a two-dimensional numerical array with array locations corresponding to the location on the chip, higher numbers representing brighter light.

Multi-pixel detectors in the focal plane of the telescope are used across all wavebands from radio to X-rays, with the CCD replaced by whatever detector is required for the given band. At radio wavelengths, for example, the pixels consist of feedhorns which guide the signal into other electronic equipment. Figure 2.8.b, illustrates this, showing the 7 pixel focal plane array used at the Arecibo Radio Telescope.

2.2.3 Field of view and pixel resolution

Since any detector has a finite pixel size, $l_{\rm pix}$, and a finite diameter, $l_{\rm det}$, Eq. (2.1) indicates that these limits will impose corresponding limits on the angular resolution, $\theta_{\rm pix}$, also called the *pixel field of view*, and the angular field of view, $\theta_{\rm FOV}$, respectively. Detectors with many small pixels have the potential to produce higher resolution images, provided there are no other limitations on the resolution (see next sections). For example, a photographic emulsion with tiny grains will show more detail than one with large grains. An illustration of an image on a CCD in which the resolution is

limited by a large pixel size is shown in Figure 2.11.c. A previous example, showing a single star smaller than the pixel field of view was given in Figure 1.9. As for the angular field of view, physically larger detectors are able to accept light from a larger range of angles, provided the telescope tube itself does not interfere.

2.2.4 Diffraction and diffraction-limited resolution

Since the telescope objective, however large it might be, collects only a portion of the light from any source that illuminates the Earth, it is like a circular aperture which accepts light that falls within it and rejects light that does not. This is analogous to a laboratory situation in which light passes through a small hole and is subsequently *diffracted*. Diffraction is the net result of interference *within* the aperture. Each point on any plane wavefront can be thought of as a new source of circular waves. This is known as *Huygen's Principle*. If there were no barrier (an infinite aperture), then the net result of these circular waves interfering with each other would still be a plane wave. However, with the barrier in place, only the points within the aperture interfere with each other and the net result is a wave that fans out in a circular fashion at the edges. The resulting diffraction pattern falling on a screen (or on a detector in the case of a telescope) is shown in Figure 2.9 for a narrow slit. For small angles and a circular aperture, the angular distance of the first null from the central point, θ_N, and the full width of the central peak at half maximum intensity (FWHM), θ_{FWHM}, are, respectively,

$$\theta_N = 1.22\lambda/D \quad \theta_{FWHM} = 1.02\lambda/D \tag{2.2}$$

where λ is the wavelength of the light and D is the diameter of the aperture. D could be somewhat less than the physical size of the aperture if the light path is partially blocked

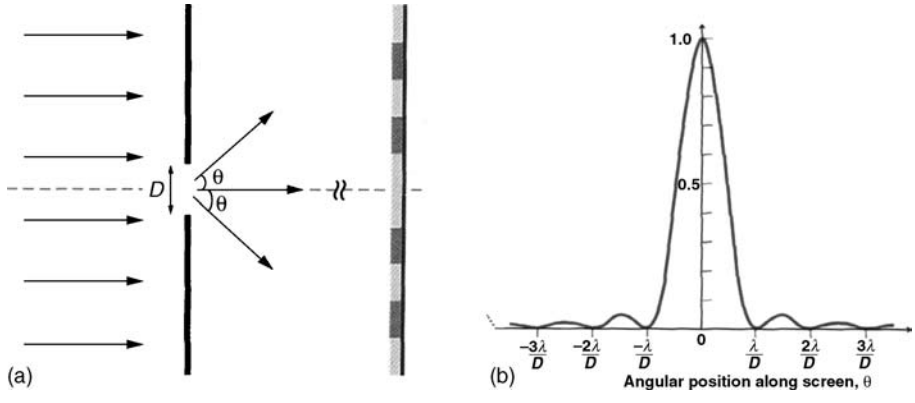


Figure 2.9. A laboratory aperture + screen is analogous to a telescope + detector. **(a)** Light is accepted only through a narrow slit and projected onto a screen some distance away. A diffraction pattern results from interference within the opening. **(b)** A plot of intensity as a function of angle, θ, across the screen yields this pattern.

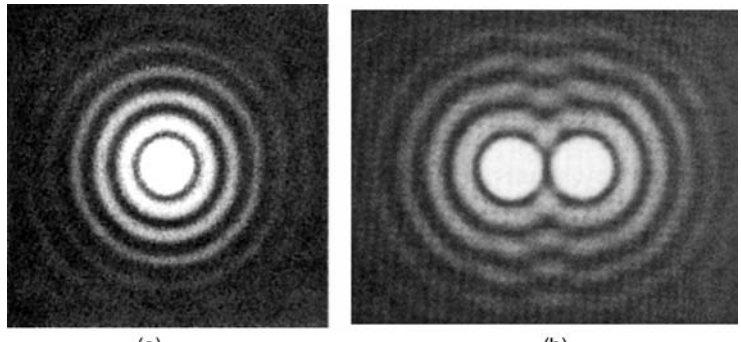


Figure 2.10. **(a)** The two-dimensional diffraction pattern from a circular aperture illuminated by a single point source at infinity. The bright central spot is called the *Airy disk*. **(b)** Two closely spaced point sources in the sky would create a pattern like this on the detector.

or otherwise imperfect. In such a case, D must be replaced by the *effective aperture*, $D_{\rm eff}$. It is clear from Eq. (2.2) that if we want a crisp small image of a point source, then a larger telescope (larger D) is required. It is also worth remembering that *higher* or *larger* θ corresponds to *lower* resolution.

A face-on view of the diffraction pattern in two dimensions for a circular aperture is shown in Figure 2.10.a. This is the *spatial response function* of the telescope to a point source that is evenly illuminating the aperture. Most of the emission falls within the central peak called the *Airy disk* in optical astronomy and called the *main lobe* or *main beam* in radio astronomy. However, a small amount of emission falls within the other surrounding peaks or diffraction rings (*sidelobes* in radio astronomy). For uniform aperture illumination, the Airy disk contains 84 per cent of the power falling on the detector. The peak of the first diffraction ring occurs at a level 1.7 per cent of the central peak and contains 8.6 per cent of the power, while the values for the tenth ring are 10^{-3} and 0.2, respectively (Ref. [178]). Thus, a point source in the sky will not form a point source on the detector, but rather a diffraction pattern in which the light is spread out as illustrated.

It is sometimes possible to modify the response function of the telescope. For example, the power going into the diffraction rings could be reduced by weighting the aperture so that it accepts more light towards the centre in comparison to the edges. However, such weighting has the effect of reducing the effective size, D, of the aperture thereby worsening its resolution. The opposite is also true. 'Super-resolution' can be achieved by weighting the outer parts of the dish or mirror but only by increasing the power going into the diffraction rings[8].

[8]In the limit, one could split up the aperture into many small elements and spread these elements out over a much larger area, effectively increasing D to a very large value. The signal can be reconstructed by analysing the cross-correlation between the various elements. This is the principle of interferometry commonly used in radio astronomy. The technique is increasingly used in optical astronomy as well, though it is more technically demanding at the smaller wavelengths. Another way to weight an aperture is to apply a mask which blocks light from certain regions.

If a second star is now present, displaced from the first by some angle, θ, then two sets of parallel light rays evenly illuminate the aperture, separated by this angle (see, e.g. points A and B in Figure 2.4). Two diffraction patterns will now be on the detector with their centres separated by θ. If these stars were separated from each other by smaller and smaller angles in the sky, the diffraction patterns on the detector would become progressively closer together, overlapping (as shown in Figure 2.10.b) and finally merging. There is a minimum angle on the sky, $\theta_{\rm d}$, at which it is still possible to *just* distinguish that there are two sources rather than one. This limit is called the *diffraction-limited resolution* of the telescope, or just the *diffraction limit*, and provides a more technical definition of the concept of resolution. For a circular aperture, the resolution is often specified by the *Rayleigh criterion* which states that this will occur when the maximum of one image is placed at the first minimum (first null) of the other. In practise, separation by the FWHM is an adequate criterion and, as a rule of thumb, the diffraction-limited resolution of the telescope may be found from,

$$\theta_{\rm d} \approx \lambda / D \tag{2.3}$$

The brightness distribution of an astronomical source can be thought of as a set of many individual closely spaced point sources of different brightness. On the detector, each point in the image will have superimposed on it the diffraction pattern of a point source. Mathematically, this can be represented as a *convolution* (see Appendix A.4) of the brightness distribution of the source with the diffraction pattern of the telescope. The broader the central lobe of the diffraction pattern, the poorer, or lower is the resolution and the source will appear as if out of focus. Such a comparison is shown between Figure 2.11.a and Figure 2.11.b. Since the

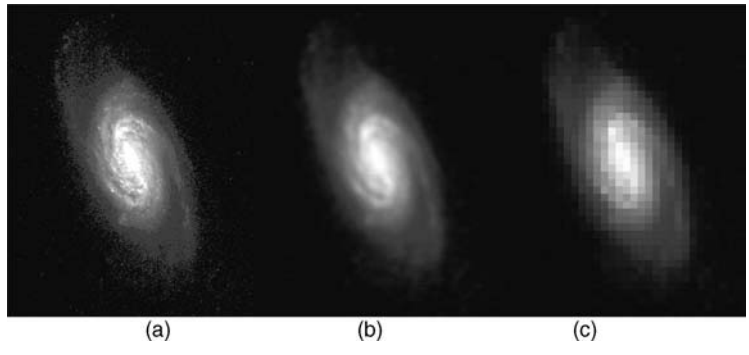


Figure 2.11. Illustration of resolution for the spiral galaxy, NGC 2903 (distance, $D = 8.6$ Mpc). The disk of this galaxy is circular, appearing as an ellipse in projection, with the minor axis of the ellipse subtending the angle, 6.0′, and the major axis, 12.6′, on the sky. **(a)** A high resolution image such as might result from using a large aperture telescope, **(b)** a low resolution image such as from a telescope with a small aperture, or from a large aperture telescope with poor seeing and **(c)** a low resolution image due to large pixels on the detector.

light in the outer diffraction rings is much lower in intensity than the main lobe, these rings do not show up in Figure 2.11.b.

Lower resolution, whether due to telescope diffraction or other limitations (Sect. 2.3.2), has the effect of spreading out the signal and lowering its peak (note that the flux is preserved, cf. Figure 1.9). In the limit, if the resolution is very low, the angular size of the source, θ_S, becomes much less than θ_d and the source is said to be *unresolved* as if it were a point source (Prob. 2.8). Its apparent angular size will then equal the angular size of the central lobe of the beam or Airy disk.

2.3 The corrupted signal – the atmosphere

Any observation from the Earth's surface must contend with the atmosphere. Not only does the signal pass through this scattering, absorbing and emitting layer, but the effect that the atmosphere has on a signal is strongly wavelength-dependent and is also variable with time and with local conditions. Details of scattering and absorption will be discussed more fully in Chapter 5 but here we consider the atmosphere in terms of its undesirable effects on observations.

The spectral response curve of the atmosphere (Figure 2.2) indicates that there are only two atmospheric windows for which ground-based observations are possible, the radio and the optical. Of these, the most transparent band is the radio. Radio wavelengths are large in comparison to the typical size of atmospheric particles (atoms, molecules, and dust particles) so the probability of interaction with these particles is small and atmospheric effects are mostly negligible. This means that observations in the radio band can be made day or night, during cloud cover or clear weather and from sea level, if desired. The sharp cutoff at very low radio frequency (≈ 10 MHz, Figure 2.2) is due to the *ionosphere* which reflects, rather than transmits incoming waves (see Appendix E). The ionosphere is an ionised layer ranging from a height of 50 km to 500 km with a maximum electron density (the number density of free electrons) of between 10^4 and $10^6\,\mathrm{cm}^{-3}$, depending on conditions. At the high frequency end of the radio band ($\nu \gtrsim 300$ GHz, $\lambda \lesssim 0.1$ cm, Figure 2.2) the atmosphere also becomes more important and radio telescopes operating in this range must be placed on high mountains and are put in domes just like optical telescopes, as Figure 2.12 illustrates. Techniques such as *beam switching* or *chopping* are also employed at these as well as IR wavelengths, which involve rapidly switching the light path back and forth from the source to a nearby region of blank sky. The image of the source is then corrected for time variable atmospheric emission by subtracting off the nearby sky brightness distribution which closely corresponds in time.

The optical window (which includes the near IR and near-UV) is strongly affected by the atmosphere and therefore optical research telescopes are placed at high altitudes above as much of the 'weather' as possible. Some specific effects of the atmosphere on an optical signal are discussed below.

Figure 2.12. The James Clerk Maxwell Telescope, operating at sub-mm wavelengths, is located on Mauna Kea in Hawaii at an elevation of 4200 m. Although this is a radio telescope, it is housed in a dome and, under normal operating conditions, the dome opening is covered by a protective cloth membrane that is transparent at the operating wavelengths. Contrast this image to the Arecibo radio telescope that operates at longer wavelengths, shown in Figure 2.6.a, which is at an altitude of only 497 m. (Reproduced by permission of Robin Phillips)

2.3.1 Atmospheric refraction

The bending of light as it passes from one type of medium to another is described by *Snell's Law* (Table I.1). Such bending occurs as light passes from space into denser and denser sections of the atmosphere of the Earth. The bending increases systematically with *zenith angle* which is the angle between the zenith and the source, making the source appear higher in the sky than it actually is. Extended objects that span a range of zenith angle will therefore also appear distorted, as does the Sun when it is close to the horizon at sunrise or sunset (Figure 2.13, Prob. 2.2). Thus, models accounting for known systematic refraction must be built into telescope control software in order to accurately point the telescope to a source in the sky. The details of atmospheric

Figure 2.13. The Sun appears slightly flattened at sunrise or sunset.

refraction at optical wavelengths are provided in the Appendix at the end of this chapter[9]. A more general discussion of refraction is given in Sect. 5.3.

2.3.2 Seeing

The atmosphere is turbulent and this turbulence can be represented by 'cells' of gas that are constantly in motion. Many such cells with different sizes and slightly different temperatures may be present above the telescope aperture. A good approximation to atmospheric turbulence is one in which the pressure remains constant but the temperature fluctuates, resulting in fluctuations in the index of refraction (see Eq. 2.A.4). This introduces a small random element of refraction superimposed on the systematic refraction discussed above. The net result is that an incoming wave that is originally plane parallel will become distorted, and the tilt of the wavefront results in a focal point on the detector that is shifted slightly from the non-tilted position. If there are multiple cells above the aperture, then there will be multiple distorted images (each convolved with the telescope's diffraction pattern) on the detector (Figure 2.14). Each individual image is called a *speckle*. The larger the aperture, the more cells can be in front of it and the more speckles there will be on the detector. The largest effects are due to cells near the ground that are blown across the aperture by wind. The sizes of these atmospheric cells vary, but a typical value is $r_0 \approx 10\,\mathrm{cm}$ in the visual, though it can be up to $\approx 20\,\mathrm{cm}$ at a high altitude where there is more atmospheric stability.

Since atmospheric cells are constantly in motion, the speckle pattern changes rapidly with time. The timescale over which significant changes in the pattern occur is of order the time it takes for a cell to move across the aperture,

$$t = a\,\frac{r_0}{v} \tag{2.4}$$

where v is the wind speed and the constant, a, has a value of ≈ 0.3. For a wind speed of $v = 300\,\mathrm{cm\ s^{-1}}$ ($11\,\mathrm{km\ h^{-1}}$) and $r_0 = 10\,\mathrm{cm}$, therefore, this timescale is 10 ms. Thus, to see diffraction limited images using a large aperture telescope, it is necessary to take very rapid exposures shorter than 10 ms in duration. This poses a problem, however, because exposures shorter than 10 ms are generally not long enough to produce a significant response on the detector except for the very brightest stars.

In order for the detector to register a signal from a faint source, it is necessary to take exposures that are much longer than 10 ms. During the exposure, the shifting speckles build up a signal on the detector, filling in the gaps and forming a smeared out image over a small region (Figure 2.14). The characteristic angular size of this region, θ_s, taken to be the FWHM, is called the *seeing disk* or just the *seeing*. Seeing, therefore, refers to the change in angular position of the signal, or the change in the phase of the wavefront, with time due to the atmosphere. The apparent position of the star in the sky is at the centre of

[9]Refraction exists at radio wavelengths as well and is similar in magnitude to the value for the near-IR.

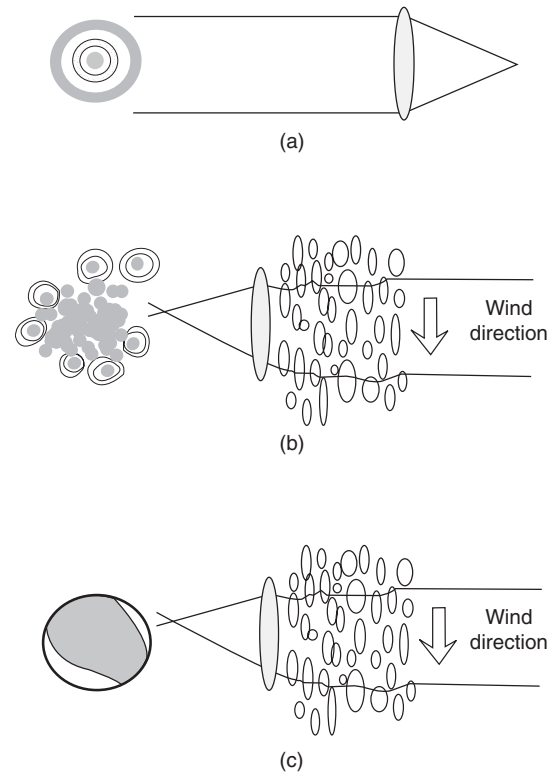


Figure 2.14. Effect of the atmosphere on a stellar image. **(a)** If the telescope were in space, the image of a point source would be the diffraction pattern of the telescope with the width of the central peak given by θ_d. **(b)** Individual atmospheric cells of typical size, 10 cm, move across the telescope aperture effectively breaking up the aperture into many small apertures about this size. Since the refraction is slightly different for the different cells, each one creates its own image, called a speckle, on the detector at a slightly shifted location. Each individual image has its own diffraction pattern and the image changes rapidly with time as the atmospheric cells move across the aperture. **(c)** If a time exposure of the star is taken, the speckle pattern is seen as a blended spot, called the seeing disk.

its seeing disk. Seeing varies with wavelength, and with altitude and geographic location. It also varies from night to night and can change over the course of the night.

Thus, the resolution of a large optical ground-based telescope is limited, not by diffraction, θ_d, but by the seeing, θ_s, since over the timescale of a typical exposure, the stellar image will smear out over a size, θ_s. Large research telescopes have historically been designed to collect more light, allowing fainter sources to be detected. However, their resolution is limited by the size of an atmospheric cell. The resolution, $\theta_s = \lambda/r_0$, so the effective aperture of the telescope is equivalent to the size of the *atmospheric cell*, not the telescope mirror. If $r_0 = 10$ cm, then a 10 cm diameter telescope will have the same resolution as a 10 m telescope ($\approx 1''$, Eq. 2.3). For the 10 cm diameter or smaller telescope, the resolution is *diffraction-limited* and for any larger telescope, it

will be *seeing-limited*. Depending on the linear size of the pixels on the CCD, however, the resolution could also be *pixel-limited* (Example 2.1). The image of a point source on the detector, as actually measured, is called the *point spread function* (PSF).

Example 2.1

An 8 inch f/10 telescope is outfitted with a 1024×1024 *pixel*^{2} *CCD, each pixel* 9 μm square. Determine the field of view, and also the resolution when this telescope is used, (a) in a back yard with seeing of $\theta_s = 2''$ and (b) on a high mountain with $\theta_s = 0.4''$. (c) Could this telescope achieve higher resolution at either of these locations?

Using *cgs* units, the telescope diameter, $D = 20.32\,\text{cm}$ and pixel size $l_{\text{pix}} = 9 \times 10^{-4}\,\text{cm}$. The CCD size is $l_{\text{det}} = 1024 \times l_{\text{pix}} = 0.92\,\text{cm}$ and the focal length, $f = 10 \times D = 203.2\,\text{cm}$. From Eq. (2.1), the field of view is $\theta_{\text{FOV}} = l_{\text{det}}/f = 4.5 \times 10^{-3}$ rad, or $15.6'$ square, independent of location. By the same equation, the pixel field of view (converting to arcseconds) is $\theta_{\text{pix}} = 0.91''$. The diffraction limit of the telescope, from Eq. (2.3), adopting an observing wavelength of $\lambda = 507 \times 10^{-7}\,\text{cm}$ (Table G.5), is $\theta_{\text{d}} = \lambda/D = 2.5 \times 10^{-6}$ rad $= 0.5''$.

(a) Since θ_s is larger than both the pixel resolution, θ_{pix}, and the diffraction-limited resolution, θ_{d}, the observations are seeing-limited and the resolution is $2''$[10].

(b) In this case, $\theta_{\text{pix}} > \theta_{\text{d}} > \theta_s$ so the observations are pixel-limited and the resolution is $0.91''$.

(c) No improvement can be achieved in the backyard case since the seeing cannot be changed. However, on the mountain, by changing the CCD to one which has pixels about half the size or smaller, the telescope could become diffraction-limited with a resolution of $0.5''$.

2.3.3 Adaptive optics

One solution to the problem of seeing is to launch telescopes, even those working at optical wavelengths, into orbit as was done, for example, for the Hubble Space Telescope. This solution, however, is very expensive. Another is to improve the design of observatories to minimise local temperature gradients that can produce turbulence (Figure 2.15).

However, there is now a very effective way of dealing with the problem of seeing – that of *adaptive optics*. Now being routinely employed at major facilities in the near-IR where the atmosphere is more stable, adaptive optics is a method of correcting for the

[10] A more accurate treatment would consider how each effect contributes to the PSF.

Figure 2.15. The Gemini North Telescope on Mauna Kea, in Hawaii, showing the open sides which help to minimise local turbulence. (Reproduced by permission of Neelan Crawford, courtesy of Gemini Observatory/AURA)

seeing in real-time by making rapid adjustments to a deformable mirror. An example of an adaptive optics system is shown in Figure 2.16.

After passing through the telescope aperture, the signal, along with a reference signal from a point source, is reflected from a deformable mirror. The reference signal is then split off to a 'wavefront sensor' which must sample the wavefront on ms timescales. The wavefront sensor breaks up the reference signal into many images via

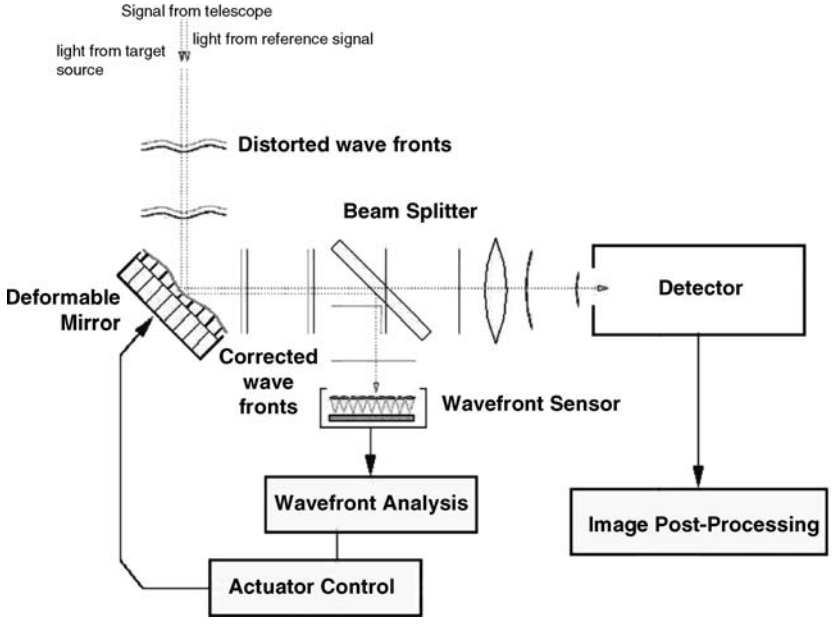


Figure 2.16. Simplified diagram of an adaptive optics system. Adapted from http://cfao.ucolick.org/ao/how.php)

lenses. If the wavefront is planar, then the reference signal images are evenly spaced on the detector, but if the wavefront is distorted, then the images are unevenly spaced. After analysis of the spacing, a signal is sent to actuators at the rear of the deformable mirror which adjust the shape of the mirror to compensate for the wavefront distortion. Since the signal of astronomical interest is also reflecting from the same deformable mirror, its image is therefore corrected as well.

In order for this system to work, the reference signal must be bright and must be close enough in the sky ($< 1'$) to the astronomical source of interest that the two signals pass through the same atmosphere and same optical path. However, there are very few bright stars in the sky that are fortuitously placed close to any arbitrary astronomical source. The solution has been to generate a *laser guide star*. A laser beam is emitted into the sky and excites or scatters off of particles in the atmosphere at an altitude of about 100 km, forming a point-like source as the reference beacon. A natural guide star must still be used for absolute positional information, but it need not be so bright or close to the target source.

The image quality that has been achieved by adaptive optics is remarkable (Figure 2.17) and this technique promises to provide more improvements as the technology becomes more refined. At present, adaptive optics systems have been

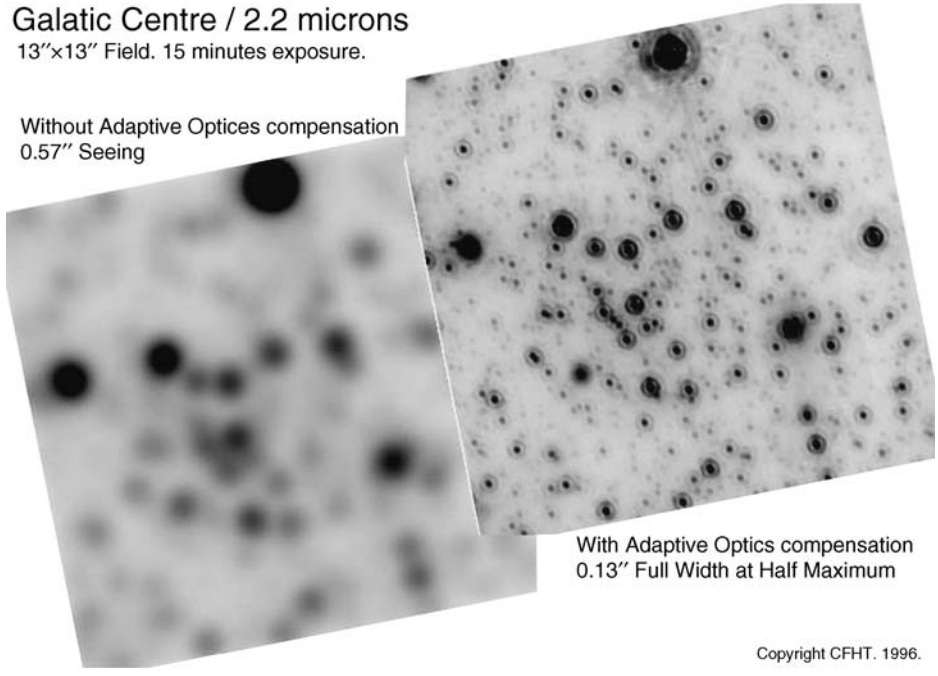


Figure 2.17. Example of the improvement that can result using adaptive optics. The image at left shows many stars crowded together in the region of the Galactic Centre. The image appears out of focus because of the seeing. At right is the same field after using adaptive optics. The star images now show the diffraction pattern of the telescope and the resolution approaches the diffraction limit. (Reproduced by permission of the Canada-France-Hawaii Telescope/1996)

added on to existing telescopes, but they will be an integral part of the design of new and larger telescopes in the future (Ref. [65]).

2.3.4 Scintillation

Scintillation, commonly called 'twinkling', is the rapid change in amplitude (brightness) of a signal with time due to the atmosphere. It occurs for similar reasons as seeing, as atmospheric cells act like lenses focussing, defocussing, and shifting the signal. Scintillation, however, is primarily due to changes in the curvature of the wavefront rather than its tilt and there is a greater effect from atmospheric cells that are higher up, rather than near the ground. The most characteristic timescale for scintillation is of order a ms, but variations are seen on a range of timescales from microseconds to seconds or longer (Ref. [53]). The fluctuations in amplitude will be about a mean value which is the desired measurement and, since most exposure times are longer than the scintillation timescale, scintillation is not a serious problem for most astronomical measurements.

A star, or other point-like source, viewed through the atmosphere can be seen by eye to fluctuate in intensity. An extended object has a brightness distribution that is convolved with the seeing disk. That is, the object's brightness distribution can be thought of as a series of point sources of different brightness, each one 'blurred' to the size of the seeing disk and each seeing disk is scintillating. If the eye could spatially resolve each of these points, it would see brightness fluctuations across a source. However, the resolution of the human eye ($\approx 1'$) is much poorer than the seeing ($\approx 1''$). Thus, each of the fluctuations across an extended source (some positive, some negative about the mean) are not seen individually by the eye but are instead integrated together. The result is that the eye perceives an extended source as steadily shining. The old adage 'stars twinkle and planets shine' is true for the planets visible by eye, but exceptions can occur (Prob. 2.9).

2.3.5 Atmospheric reddening

Reddening is the result of wavelength dependent absorption and scattering. Shorter wavelength (blue) light is more readily scattered out of the line of sight than is light of longer (redder) wavelengths. Therefore longer, redder wavelengths will have greater intensity when viewed through an absorbing and scattering medium like the atmosphere. This is true at both the zenith and the horizon, but since viewing towards the horizon corresponds to viewing over a longer path length, the effect is stronger, as is readily seen when viewing a red sunset. Reddening of starlight can also be noticed by eye by observing a given star when it is high in the sky and then again when the star is about to set. Molecules as well as dust and aerosols contribute to the reddening in the Earth's atmosphere. Absorption and scattering will be treated in greater detail in Chapter 5.

2.4 Processing the signal

It is not possible to detect a signal without affecting it. Both the atmosphere, as well as the telescope and detector, impose their spatial and spectral response patterns on the signal, as already seen. As light passes through other optical devices within the telescope (for example, through lenses, wavelength-dispersive media such as spectrographs, through filters or reflected from smaller mirrors), each 'interaction' also affects the signal. Moreover, other sources of emission which can pollute the image (such as detector noise or atmospheric emission) must be removed. The image that is detected, called the *raw image*, requires corrections for these effects, or at least a characterisation of them, to know their limitations on the data. In addition, the corrected data, which consist of an array or arrays of numbers, must be converted to commonly-understood units (e.g. magnitudes or Jy per unit solid angle) that are independent of telescope or location, i.e. the signal must be *calibrated*. The end result will be a *processed image* such as would be measured with 'perfect' instrumentation above the Earth's atmosphere.

2.4.1 Correcting the signal

The steps involved in correcting the signal vary depending on waveband, the specifics of the telescope and its instrumentation. For example, if the telescope is in orbit, corrections are required for variations in the signal due to variations in the orbit. If images are taken rapidly, a new image may retain a 'memory', due to incomplete CCD readout from the previous image. Data from radio telescopes require radio interference to be excised and beam sidelobes to be carefully characterised. Optical telescopes must have unwanted cosmic ray hits removed and atmospheric emission lines either corrected for or avoided. Non-uniform responses across the field of view as well as in frequency must be 'flattened' and higher order distortions corrected if present.

For work with an optical telescope and CCD, the minimum corrections would include the following steps. The CCD image would first have a constant positive offset subtracted from the image to account for a known DC voltage level that exists in CCDs (the bias correction). The signal that results from thermal electrons in the detector, without any exposure to light would then be subtracted (the dark or thermal correction). The non-uniform response of each pixel to light across the CCD must be corrected; this is done by dividing the image by another image that has been exposed to a source that is known to be evenly illuminated (the flat field correction). Multiple frames of the same source, if they are present, must be medianed (rather than averaged) together, a process which removes random cosmic rays (Sect. 3.6). Finally, a background sky emission level, determined from a region on the CCD in which there is no source, would be subtracted from the image. Depending on the complexity of the instrumentation, it is usually more time-consuming to correct and characterise the data than to acquire the data in the first place.

2.4.2 Calibrating the signal

The most common method of calibrating the signal is to compare it to a reference signal (the *calibrator*) of known flux density or brightness. There are usually only a few *primary flux calibrators* at any wavelength and all other observations are brought in line with them. The stars, Vega and Sirius, are good examples in the optical band, as are the quasars (see Sect. 1.6.4), 3C 286 and 3C 48 at radio wavelengths. Asteroids and planets are often used at sub-mm and/or IR wavelengths. Secondary calibrators may also be used. These would be sources whose fluxes are tied to those of the primary calibrators and are more widely dispersed in the sky so that one or more are accessible during any observing session.

Atmospheric attenuation is important at optical and sub-mm wavelengths and increases with zenith angle. At optical wavelengths, a calibrator must be monitored periodically at a variety of zenith angles over which the source of interest is also observed. The source data are then corrected according to the calibrator signal at the corresponding zenith angle. At sub-mm wavelengths, at least once per night a measurement of sky brightness as a function of zenith angle is made (called *sky dips*) and this information is combined with a model of the atmosphere to compute the attenuation.

Once the calibrators have been observed, it is then a straightforward matter of multiplying the source counts, in arbitrary units, by the factor determined from the calibrator (e.g. magnitudes/counts, Jy/counts, etc.). Since the calibrator and source are observed with the same equipment, the attenuation of the signal due to absorptions and scatterings in the instrumentation are also taken into account.

Clearly, the primary flux calibrators must be measured on an absolute scale for this system to work. This process is non-trivial and considerable effort has been expended to make accurate measurements and bring them to a common system (e.g. Refs. [11], [136]). It consists of observing the source along with a man-made source of known brightness. At radio or sub-mm wavelengths, this might be a black absorbing vane (a black body, see Sect. 4.1 and Example 4.1) at fixed known temperature and, at optical wavelengths, a calibrated lamp. The flux density of Sirius and Vega are also alternatively determined by calculating the expected value, using other known stellar parameters such as temperature, distance, and the shape of the spectrum. Absolute calibrations are also sometimes used during normal observations, rather than reverting to the relative scale discussed above, particularly at radio and sub-mm wavelengths.

2.5 Analysing the signal

The final result is a reduced, or processed image, in units that are common to astronomy, ready for analysis, measurement, and other scientific scrutiny. Measurements of specific intensity and flux density can be made as described in Sect. 1.3 and illustrated in Figure 1.8, and other parameters can then be determined, if possible, as described in the previous chapter. It is therefore important to understand the characteristics of the images in order to place limits on the source parameters which are derived from them.

Once an image has been obtained (assuming no artifacts), it will be characterised by eight properties, three providing spatial information, three providing spectral information, and two related to the amplitude of the signal:

(1) Position of map centre: The map centre is usually placed at the location of the target source in the sky and specified by some coordinates, Right Ascension and Declination (RA, DEC) being most common.

(2) Field of view: The FOV is set by the detector size in the case of a simple system (Sect. 2.2.3) though it can be smaller for more complex, partially blocked light paths.

(3) Spatial resolution: This is affected by telescope size, pixel size, or seeing, as described in Sect. 2.3.2.

(4) Central wavelength or frequency: Different instrumentation is required for different wavebands but, within any given band, more accurate centering in wavelength or frequency can be achieved by tuning or using filters.

(5) Bandwidth: Implicit in the setting of central wavelength or frequency is the fact that a range of wavelength (λ_1 to λ_2) or frequency (ν_1 to ν_2) will be detected.

(6) Spectral resolution: If detailed spectral information is required, the waveband may be broken up into wavelength or frequency *channels* and each examined separately.

(7) Noise level: A spatial region of the image that is devoid of emission (after all corrections and sky subtraction) will contain positive and negative values about a mean of zero due to noise in the system. The value is measured as a *root-mean-square*, i.e. $N = \sqrt{< x^2 >}$, where x is the calibrated pixel value and the angle brackets represent an average.

(8) Signal-to-noise ratio (S/N): This is the ratio of the signal strength at any image location to the rms noise. In general, longer integration times (that is, longer exposure times) result in higher S/N images because N decreases with increasing integration time. In order for a signal to be considered a real detection, the S/N ratio must be high enough to be sure that it is not just a random peak in the noise level. A minimum value of ≈ 3 to 10 will often be used, although other factors can also play a part in this decision, such as the angular size of the signal on the detector or whether it is seen in the same location at other wavebands. Another way of characterising the image as a whole is by the *dynamic range* which is the ratio of the peak signal to the minimum detectable signal on the map. This is therefore a measure of the 'stretch' in brightness that has been detected. A practical measurement is to take the peak signal on the map over the rms noise. The S/N ratio may be different at every point in the map but there is only one value of dynamic range for a map.

These properties must be kept in mind when determining and interpreting the scientific results that derive from the data. Example 2.2 provides an illustration of the kinds of questions one should ask prior to embarking on an observing run.

Example 2.2

A galaxy at a distance of $D = 25$ Mpc will be observed with a spatial resolution of 1″ at a frequency of 4.57×10^{14} Hz and bandwidth of $\Delta\nu = 2 \times 10^{12}$ Hz in order to detect the HII regions (these are ionised hydrogen regions around hot stars, see Figures 3.13 and 8.4) within it. The map noise expected for the integration time proposed is $N = 2 \times 10^{-17}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}. Could a spherical HII region of typical diameter, $d = 500$ pc, and luminosity, $L = 10^{39}$ erg s^{-1}, in this frequency band be detected from these observations? Would it be resolved?

At $D = 25\,\mathrm{Mpc} = 7.71 \times 10^{25}$ cm, the flux (Eq. 1.9) is $f = 1.34 \times 10^{-14}$ erg s^{-1} cm^{-2} and the mean flux density in a band of width, 2×10^{12} Hz, is (Eq. 1.7) $f_\nu = 6.69 \times 10^{-27}$ erg s^{-1} cm^{-2} Hz^{-1}. A diameter of 500 pc corresponds (Eqs. B.1, B.3) to a solid angle of $\Omega = 3.14 \times 10^{-10}$ sr so the specific intensity is (Eq. 1.13) $I_\nu = 2.13 \times 10^{-17}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}. This is just at the rms noise level, so the HII region would not be at the minimum level for detection. A longer integration time should be chosen to increase the S/N. An angular resolution of 1″ corresponds (Eq. B.1) to a linear scale of 121 pc at the distance of this galaxy, so the HII region would indeed be resolved, if it were detected.

2.6 Visualising the signal

The final image can be displayed in a variety of ways. The galaxy, NGC 2903, has already been shown for different resolutions (Figure 2.11). Figure 2.18 shows the same galaxy, but using three different visualisation schemes. In all three cases, the data (i.e. the numerical values in the 2-D array) are exactly the same. The first image shows the data as a linear greyscale (Figure 2.18.a) similar to the linear response of the CCD with which it was obtained. This represents the data well, except that the range of brightness is very large and it is difficult to show the entire range on paper. Details near the centre of the galaxy are 'burned out' and the faint outer regions of the galaxy do not show up well. An alternative is Figure 2.18.b which applies a logarithmic weighting to the data prior to displaying it. This has the effect of compressing the range of brightnesses so that both the bright and dim parts of the image are easily seen. The drawback is that there is less contrast in any part of the image so the true range of brightness is not obvious. Figure 2.18.c is a third possibility. Apart from being visually appealing, the use of colour, depending on how it is applied, allows details in both the bright and dim parts of the image to be highlighted. This is called *false colour* because the adopted colour scheme is chosen only for illustrative purposes and has no scientific meaning otherwise.

Figure 2.19 illustrates a more sophisticated representation of data when comparisons are desired between wavebands. Contours (Figure 2.19.a) are most often used alone or

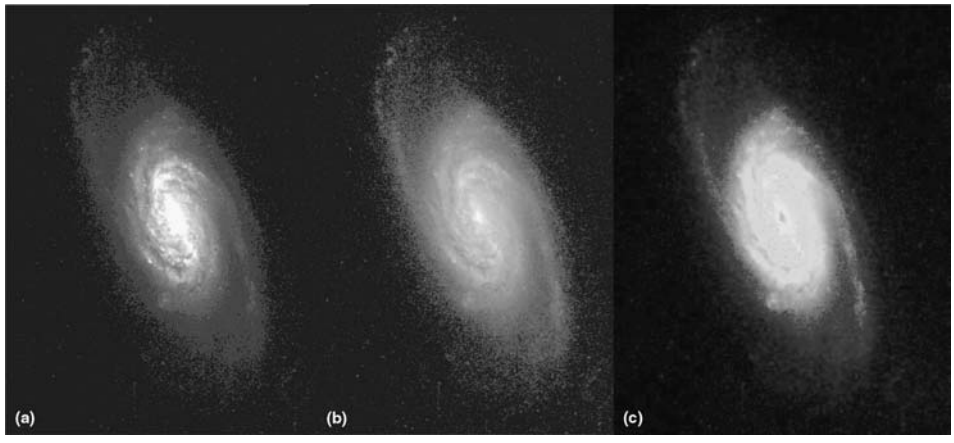


Figure 2.18. An optical image of the galaxy, NGC 2903 (distance, $D = 8.6$ Mpc), taken in a band centred at 500 nm, is represented in three different ways: **(a)** linear greyscale, **(b)** logarithmic greyscale, and **(c)** logarithmic false colour. (see colour plate)

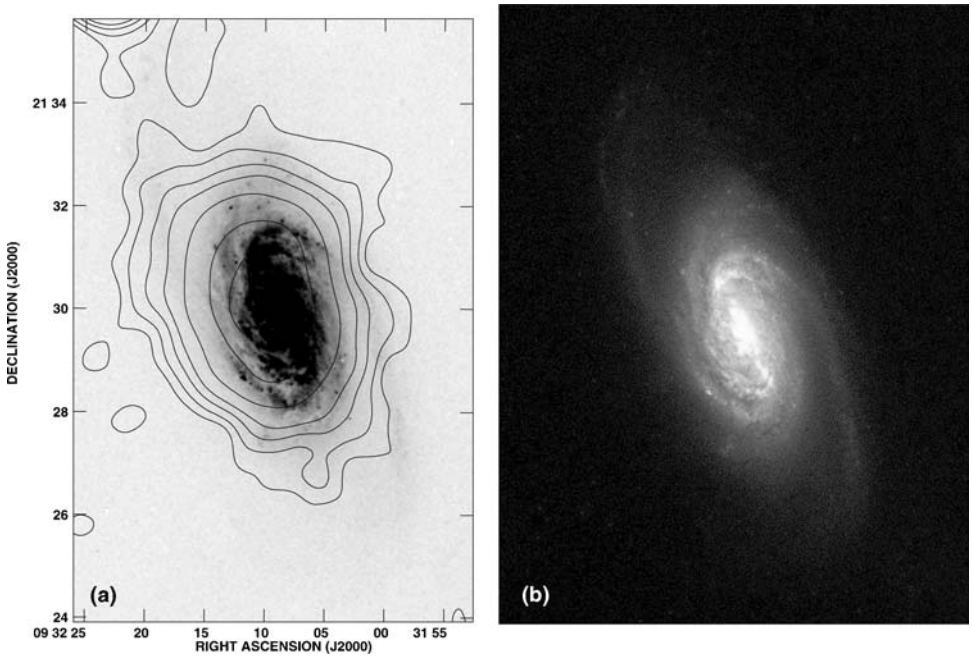


Figure 2.19. **(a)** Radio continuum contours at λ20 cm overlaid (single-channel bandwidth = 0.66 MHz) on the same image of NGC 2903 as in Figure 2.18. The radio contours are at 0.6, 1.5, 3, 5, 10, 25, and 50 mJy beam^{-1}, the peak map level is 117 mJy beam^{-1}, the beam FWHM is 54.4″ (see Sect. 2.2.4), there are 1.7″ per pixel, and the rms noise is 0.5 mJy beam^{-1}. **(b)** Combination of three different images in three wavebands centred at λ500 nm (shown as blue), λ650 nm (green) and λ820 nm (red). (see colour plate)

in overlays like this when the spatial resolution is low or the image structure is not very complex. Otherwise the contours become crowded. In Figure 2.19.b, another optical image is shown. At first glance, this might be interpreted as being a single image for which colours have been applied (i.e. false colour). In fact, it is a combination of three different images taken in three different wavebands for which different colours have been adopted. Again, the colours corresponding to these wavebands were not intended to describe what the eye would see. They are simply chosen to distinguish between the three bands in order to convey more information. For scientific analysis, it is best to consider the image in each band separately.

Any of these approaches is valid and astronomers generally use whatever weighting, contour, or colour scheme allows them to convey the information about the source that is desired. Since the 'true' colour of an astronomical object applies only to the narrow visual range of the electromagnetic spectrum to which the eye is sensitive, from a scientific point of view there is little motivation to reproduce true colour images of astronomical objects. However, the careful adoption of a visualisation scheme can sometimes provide useful information 'at a glance', as Example 2.3 illustrates.

Example 2.3

Consider the strong background source just to the north of the HII region, IC 5146*, that is visible in the diffraction-limited radio image, shown by contours, in Figure* 8.4*.*

(a) What quantity is represented by the contours?

(b) Is this source resolved or unresolved?

(c) Estimate the flux density of this source (Jy).

(d) Express the peak specific intensity of this source in cgs units.

(a) The contour units are in milli-Janskys (see Eq. 1.8) per beam-solid-angle, (mJy beam^{-1}) so these are contours of specific intensity.

(b) If the source is unresolved (source angular diameter, $\theta_S, \ll$ the beam angular diameter, θ_b), then its observed size will be the same as the beam size since its emission is spread out over the shape of the beam (Sect. 2.2.4). If θ_S approaches the size of the beam or exceeds it, then the observed size of the source will be broader than the beam. The source in the figure is the brightest source in the map (see contours) so its peak is (see caption) $I_{\nu\max} = 318$ mJy beam^{-1}. The half-maximum contour is then $I_\nu = 159$ mJy beam^{-1}. Measuring the full width of the half-maximum contour using the tick marks on the Declination (ordinate, or north-south) axis[11] shows that the apparent source size,

[11]Do not use the Right Ascension axis to measure angles since it is expressed in units of time.

though somewhat oval in shape, is not significantly greater than the stated beam FWHM of $\approx 80''$ in any direction. Therefore this source is unresolved, a conclusion that is reached by a quick comparison of its angular size at half-maximum with the beam size.

(c) The flux density, f_S, of an unresolved source is preserved (Figure 1.9, Sect. 2.2.4). Since the main beam has a Gaussian shape and the telescope is pointing directly at the source, the integral under the beam (of solid angle, Ω_b) can be approximated as (Eq. 1.13), $f_\nu = \int I_\nu \cos(\theta)\, d\Omega \approx I_{\nu\max}\, \Omega_b = 318\,\mathrm{mJy}$. Basically, the flux density of an unresolved source is confined to a single beam, and therefore read simply from the peak at its position for contours in 'per beam' units.

(d) This part requires more than a 'quick glance' at the map. From the definition of a milli-Jansky (Eq. 1.8), $I_\nu = 318 \times 10^{-26}$ erg s^{-1} cm^{-2} Hz^{-1} beam^{-1}. Converting the beam FWHM to radians, we find, $\theta_b = (80/60/60/180) * \pi = 3.88 \times 10^{-4}$ rad. Then the beam solid angle is (Eq. B.3 for a beam that is approximately circular on the sky), $\Omega_b = \pi {\theta_b}^2/4 = 1.18 \times 10^{-7}$ sr. We finally find, $I_\nu = (318 \times 10^{-26}$ erg s^{-1} cm^{-2} Hz^{-1} beam$^{-1}) \div (1.18 \times 10^{-7}$ sr beam$^{-1}) = 2.7 \times 10^{-17}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} in cgs units.

There is one possible element of an image that is still missing, however – that of the spectral dependence of the emission. The above images apply only to data obtained in a single spectral channel. However, if the frequency band has been split into many frequency channels (or equivalent in the wavelength regime), then an image is obtained in each of these channels. Rather than a two-dimensional image, we now have a three-dimensional *data cube*, the third axis being either frequency, or wavelength. If emission is observed in a spectral line (see Chapter 9), then the Doppler relation (Table I.1, Sect. 7.2.1) can be used to convert the third axis into a velocity. Figure 2.20 illustrates this for neutral hydrogen (HI, see Sect. 3.4 for nomenclature) data obtained for NGC 2903. Data cubes like this contain a rich wealth of information about the source, its contents and dynamics.

Problems

2.1 Using Figure 2.4 as a starting point, (a) draw a diagram showing the rays that indicate the field of view on the sky, and (b) draw a diagram which illustrates why larger apertures should result in brighter images.

2.2 For the conditions given in the Appendix at the end of this chapter, determine the ellipticity (ratio of minor to major axis) of the sun when its lower limb appears to be at the horizon.

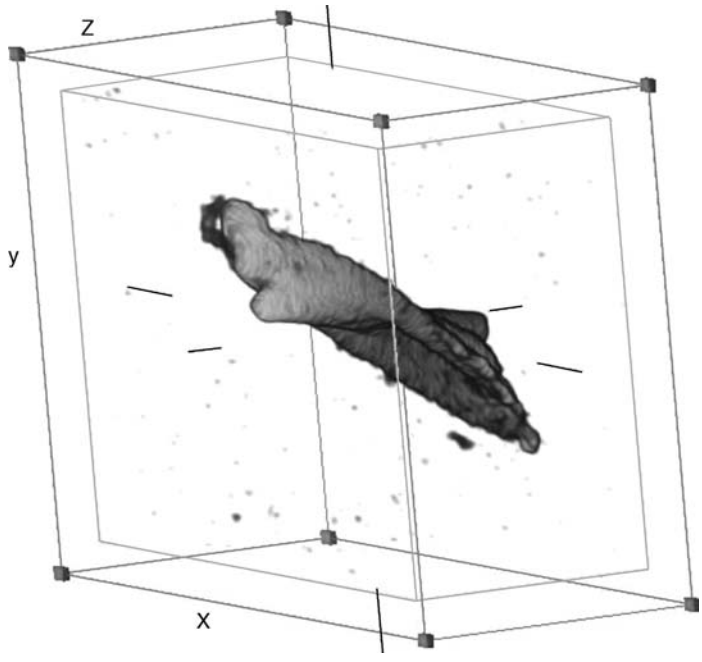


Figure 2.20. A *data cube* showing λ 21 cm HI emission from the galaxy, NGC 2903 in a three-dimensional 'space' with the x axis being Right Ascension, the y axis being Declination, and the z axis being either wavelength, frequency, or velocity. For this galaxy, the emission on the z axis extends from an observed wavelength of $\lambda_1 = 21.13215$ cm to $\lambda_2 = 21.15837$ cm, with the centre at $\lambda_{sys} = 21.14526$ cm. An analysis of the emission in this cube can determine how the galaxy is rotating and how much mass is present (Sect. 7.2.1).

2.3 (a) An observer, using the V-band filter (see Table 1.1), centres a star on his 256×256 pixel CCD detector. The star's observed zenith angle is $60°$, the CCD pixel separation corresponds to $1''$ on the sky, and the seeing is sub-arcsecond. He then changes to a U-band filter followed by an I-band filter. Determine and sketch where the star will appear on the CCD in each of the filters, as well as its true position for the reference pressure and temperature given in the Appendix at the end of this chapter.

(b) Repeat part (a) but for a temperature of $0°$C.

(c) If the filters are removed and if the CCD could detect all wavelengths from the U band to the I band with uniform sensitivity, what would the image of the star look like on the CCD?

2.4 For observations at $\lambda 2.2 \mu$m through atmospheric turbulence,

(a) Derive an expression for the variation in the index of refraction, δn, with temperature, δT. For a single layer within the turbulent region, write the variation in the refractive angle, δR, as a function of δT (see Eq. 2.A.2).

(b) Evaluate δR (arcseconds) for a star at an observed zenith angle of 45° when the fluctuation in temperature, $\delta \mathrm{T} = 0.3\,\mathrm{K}$, and for typical high altitude conditions, i.e. $\mathrm{P} = 600 \times 10^2\,\mathrm{Pa},\ \mathrm{T} = 0°\mathrm{C}$.

(c) Assuming that the combination of different δR from different layers along the line of sight results in a seeing size of $\theta_s \approx 10\,\delta R$, determine the corresponding atmospheric cell size, r_0.

(d) For a wind speed of 20 km/h, how frequently must the deformable mirror be adjusted in an adaptive optics system to correct for seeing at this wavelength?

2.5 Determine whether the dark-adapted human eye is diffraction limited, seeing limited, or pixel-limited (Table G.5).

2.6 (a) Determine the spatial resolutions (arcsec) of the following instruments (assume that seeing for the ground-based instruments is 0.5″): (i) The James Webb Space Telescope (JWST, $D = 6.5\,\mathrm{m},\ D_{\mathrm{eff}} = 5.6\,\mathrm{m}$) at $\lambda 28\,\mu$m. (ii) The James Clerk Maxwell Telescope (JCMT) ($D = 15\,\mathrm{m}$) using the 2nd Submillimetre Common-User Bolometer Array (SCUBA-2) detector ($D_{\mathrm{eff}} = 12.4\,\mathrm{m}$) at $\lambda 450\,\mu\,\mathrm{m}$.

(b) How large would a ground-based optical telescope operating at $\lambda 507$ nm have to be to match the resolution of the JWST? How large would a single dish radio telescope operating at $\lambda 20$ cm have to be?

2.7 Using the information in the caption of Figure 1.2, determine the proper motion (in arcsec/year) of the filaments in Cas A. How long would it take for this proper motion to become measurable, using a radio telescope whose effective diameter is, $D_{\mathrm{eff}} = 32\,\mathrm{km}$?

2.8 A galaxy is face-on, circular in appearance, and subtends an angular diameter of 10″. In which of the following situations would the galaxy be like a point source, i.e. completely unresolved: (a) an amateur 'backyard' 2 inch diameter optical telescope with seeing of 2″, (b) the largest single-dish radio telescope in the world (Arecibo, Puerto Rico) of diameter, 305 m, operating at $\lambda 20$ cm, (c) the 0.85 m Spitzer Space Telescope (launched 2003) operating at the IR wavelength of $\lambda 60\,\mu$m?

2.9 On a dark clear night but with poor seeing ($\theta_s = 4''$), the two planets, Jupiter and Uranus (both near opposition) are viewed by eye. Determine, approximately, the number of independent seeing disks that are present across the face of these two planets and indicate whether they will twinkle or shine.

2.10 From the image and information provided for the radio continuum data in Figure 2.19.a,

(a) list or estimate numerical values of the eight map parameters, as described in Sect. 2.5.

(b) estimate the following values (cgs units) for the galaxy in the $\lambda 20$ cm band : (i) the flux density, (ii) the flux, (iii) the luminosity, (iv) the spectral power.

2.11 A radio telescope operating at $\lambda 6\,\mathrm{cm}$ has a circular beam of $\theta_{\mathrm{FWHM}} = 30''$. If the lowest possible rms noise on a map made using this telescope is $0.01\,\mathrm{mJy\ beam^{-1}}$, could this telescope detect radio emission from the hot intracluster gas whose spectrum is plotted in Figure 8.3?

Appendix: refraction in the Earth's atmosphere

Consider a star whose light impinges on the Earth's atmosphere at an angle, θ_1, from the vertical (Figure 2.A.1). In the simplest approximation, one could consider a ray to pass from the vacuum of space, with an index of refraction, $n_1 = 1$, to an atmosphere which is a uniform slab with index of refraction, n_2. An observer at position, O, would then see the star at a zenith angle of θ_2 in the sky. Thus, the star appears higher in the sky than it actually is. The two angles are related via Snell's Law (Table I.1),

$$\sin(\theta_2) = \frac{1}{n_2}\sin(\theta_1) \tag{2.A.1}$$

The Earth's atmosphere is not uniform, but it is still possible to consider horizontal slabs of atmosphere which are thin enough that, within each, the density is uniform and the index of refraction is constant. The geometry of the refraction in each thin slab is then described by the figure and Eq. (2.A.1) now applies only to the first slab. The slab immediately underneath the first slab would then have an incoming angle of θ_2 and an outgoing angle, θ_3. Thus,

$$\sin(\theta_3) = \frac{n_2}{n_3}\sin(\theta_2)$$

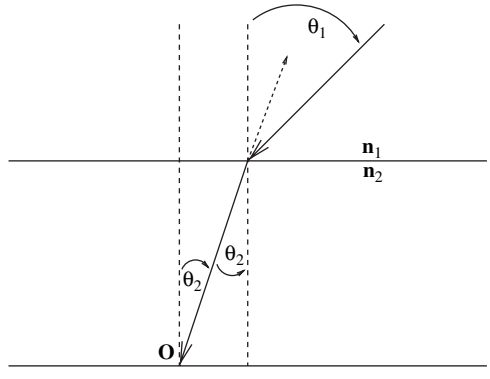


Figure 2.A.1. A ray diagram, showing the direction of the incoming wave front and its refraction in an atmospheric layer (angles exaggerated).

From Eq. (2.A.1), this becomes,

$$\sin(\theta_3) = \frac{1}{\mathrm{n}_3}\sin(\theta_1)$$

Thus, for an arbitrary number of slabs, N,

$$\sin(\theta_N) = \frac{1}{\mathrm{n}_N}\sin(\theta_1)$$

We now identify θ_N with the observed zenith angle, z_o, θ_1 with the true zenith angle, z_t, and n_N with the index of refraction at ground level, n, to write,

$$\sin(z_\mathrm{o}) = \frac{1}{\mathrm{n}}\sin(z_\mathrm{t})$$

Expressing $z_\mathrm{t} = z_\mathrm{o} + R$, where R is the *refractive angle*, using the identity, $\sin(z_\mathrm{o} + R) = \sin(z_\mathrm{o})\,\cos(R) + \cos(z_\mathrm{o})\,\sin(R)$ and assuming that the angle, R, is small so that $\sin(R) \approx R$ and $\cos(R) \approx 1$, we find, for the plane parallel atmosphere,

$$R \approx (\mathrm{n} - 1)\,\tan(z_\mathrm{o}) \tag{2.A.2}$$

Eq. (2.A.2) gives the change in zenith angle (radians) as a function of observed zenith angle and depends on the index of refraction as measured at the position of the observer. The source always appears higher in the sky (smaller zenith angle) than its true position.

A more accurate derivation (Ref. [163]) results in,

$$\begin{aligned} R &= A\,\tan(z_\mathrm{o}) - B\tan^3(z_\mathrm{o}) \\ A &= (\mathrm{n} - 1) \\ B &= 0.001254(\mathrm{n} - 1) - 0.5(\mathrm{n} - 1)^2 \end{aligned} \tag{2.A.3}$$

where R, A and B are in radians. Eq. (2.A.3) is valid[12] for zenith angles $z_\mathrm{o} < 75^\circ$.

The index of refraction, n, is a function of wavelength, λ, and also depends on the temperature, T, pressure, P, and water vapour content of the air. An expression which is valid from the λ230 to λ1690 nm (ultraviolet through infrared) is, for dry air[13],

$$(\mathrm{n} - 1) = \left[\frac{5.792105 \times 10^{-2}}{238.0185 - \left(\frac{1}{\lambda^2}\right)} + \frac{1.67917 \times 10^{-3}}{57.362 - \left(\frac{1}{\lambda^2}\right)}\right]\left[\frac{\mathrm{P}}{\mathrm{P_s}}\frac{\mathrm{T_s}}{\mathrm{T}}\right] \tag{2.A.4}$$

[12]The variation in R due to varying atmospheric scale height with temperature as well as the non-sphericity of the Earth are less than 1 per cent and have therefore been neglected in the expressions for A and B.

[1]

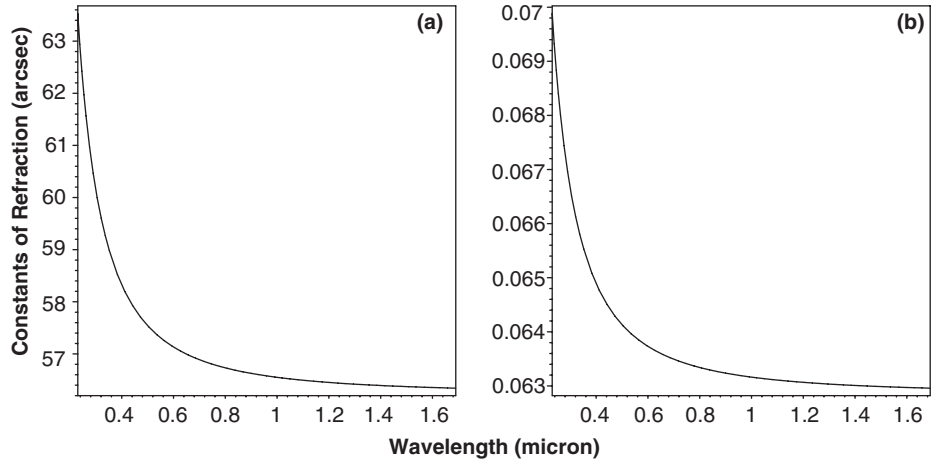


Figure 2.A.2. Plot of the constants, **(a)** A and **(b)** B, for a temperature and pressure of 288.15 K and 1013.25×10^2 Pa, respectively.

for T in K, P in Pa, and λ in μm (Ref. [41], [176]). The reference temperature and pressure are $T_s = 288.15$ K, and $P_s = 1013.25 \times 10^2$ Pa, respectively.

The resulting constants, A and B, are plotted as a function of wavelength in Figure 2.A.2. The resulting refractive angle is plotted against observed zenith angle for a wavelength of 574 nm in Figure 2.A.3.a.

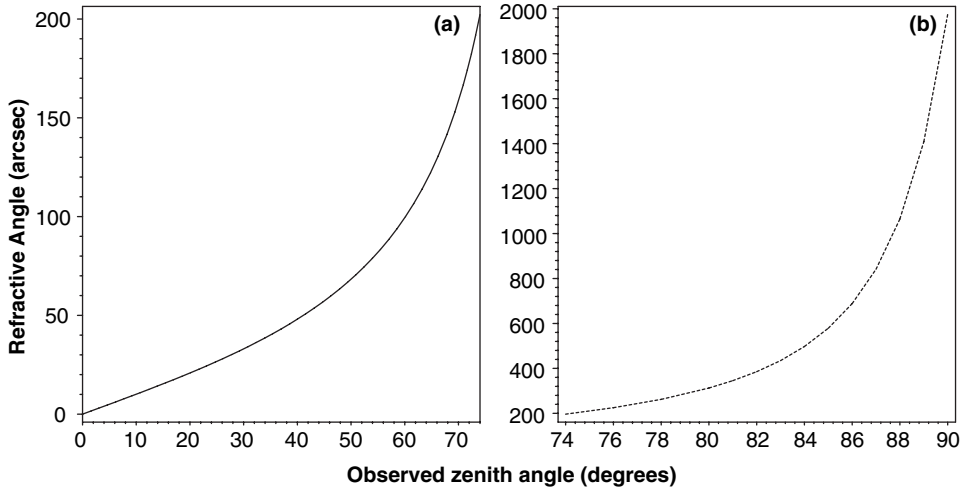


Figure 2.A.3. The refractive angle, R, as a function of observed zenith angle at a wavelength of λ574 nm, for dry air at $T = 288.15$ K and $P = 1013.25 \times 10^2$ Pa. **(a)** Values for zenith angles up to 75° evaluated from Eq. (2.A.3.) **(b)** Values for large zenith angles from Ref. [176]. Depending on atmospheric conditions, there can be significant variations in this plot.

Table 2.A.1. Refractive angle for large zenith angles[a]

z_o(deg)	R (arcsec)	z_o (deg)	R (arcsec)
74	196.49	84	497.25
76	225.00	85	578.72
78	262.20	86	688.25
80	312.78	87	841.19
81	345.52	88	1064.59
82	385.34	89	1408.82
83	434.68	90	1974.35

[a]From Ref. [176], for dry air at sea level and 45° latitude, T = 288.15 K, P = 1013.25 Pa, and λ574 nm.

For zenith angles that are large ($z_o \gtrsim 75°$) the index of refraction becomes more strongly dependent on local atmospheric conditions and modifications must be made to the equations presented here. For example, at sea level, a temperature variation of 65°C can result in a variation in refraction at the horizon by as much as 15′ (Ref. [176]). Corrections for variations in pressure and water vapour content are typically smaller than this, but can still be significant. An example of the refraction at large zenith angle is shown in Fig. 2.A.3.b and the values are given in Table 2.A.1.

PART II

Matter and Radiation Essentials

Radiation and matter can be thought of as different manifestations of energy. Consider, for example, $E = h\nu$ or $E = \mathrm{m}\,c^2$ (Table I.1). These different forms of energy, however, interact with each other in very specific ways. Were it not for this fact, we could discern very little about the Universe around us. In astrophysics, it is radiation that we observe, but it is matter that we most often seek to understand. What is its origin? How does it behave? What is its energy source? By what process does it radiate? How does it evolve with time? Is the radiation that we see even a good indicator of the true content of the Universe?

These are challenging questions and not easily answered. We are not left without some tools, however. The radiation that we detect does indeed provide us with evidence as to the nature of this vast home in which we live. Sometimes, the evidence is strong from a single set of observations that are thorough and well executed. Sometimes it is strong because it comes from completely independent data sets which, when considered together, present a consistent and compelling picture. Sometimes we only have clues and hints as to the direction that the truth lies. Science is a process, not an endpoint, providing us with a snapshot of our natural world. We would always like that snapshot to be brighter, clearer, more detailed or larger, but to have a picture is infinitely better than being left in the dark. Even more encouraging is the fact that we continue to see the picture focussing and refocussing before our eyes.

To understand this process, or to be a part of it, there are some important concepts related to matter and radiation that should first be appreciated. In this section, therefore, we present some 'essentials' of matter and radiation in preparation for a discussion of the interaction between these two forms of energy. It is this matter/radiation interaction which is at the heart and soul of astrophysics.

3
Matter Essentials

3.1 The Big Bang

The Universe, including space, time, energy in its various forms, virtual particles and photons, visible and unseen matter, is believed to have begun in a primordial fireball called, prosaically, the *Big Bang*. It is important to think of the Big Bang, not as an event that has occurred at some point in time and whose result, the Universe, is now expanding into space. Space and time, as we know them, did not pre-exist, but rather came into being with the Big Bang. There is now much evidence for this event that, collectively, presents a compelling case. For example, a variety of observations suggest that the Universe has *evolved*. Since light travels at a finite speed, as we look farther into space, we also look back to earlier epochs in the history of the Universe. This early Universe was very different from what we see today, showing different kinds and admixtures of sources, and objects with properties that differ from those that are local (Figure 3.1). The expansion of the Universe itself (Appendix F) is consistent with this view and the age of the Universe, derived from its cosmological properties (13.7 Gyr, Ref. [16]), is similar to the age of the oldest stars (12.7 to 13.2 Gyr, Ref. [73])[1]. The fact that the night sky is dark is also consistent with a Universe of finite age (Sect. 1.4). At the earliest epoch that it is currently possible to observe, we see the *cosmic microwave background* (CMB), an opaque glow at 2.7 K that fills the sky in all directions (Figure 3.2). The CMB is an imprint of the cooling primordial fireball. Tiny fluctuations in the CMB are now revealing further details about the cosmological parameters that characterize the large-scale properties of the Universe. These fluctuations form the seeds from which galaxies later grow (see Figure 5.7). Finally, the abundances of the light elements, as predicted from nucleosynthesis in the first moments after the Big

[1]The error bar on the age of the Universe is claimed to be less than 1 per cent and on the stellar ages, it is likely less than 5 per cent.

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

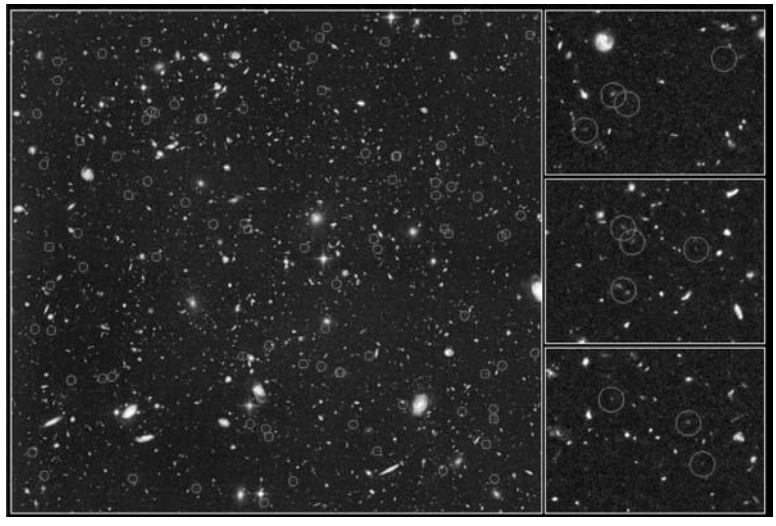


Figure 3.1. The Hubble Ultra-Deep Field, taken from the Hubble Space Telescope in a 10^6 s exposure. The red 'smudges', circled in these images (several blow-ups shown on right) are among the most distant galaxies ever observed optically. The faintest objects are 2.5×10^{-10} times as bright as what can be observed with the naked eye. These galaxies, or rather 'galaxy precursors', are seen as they were at a time when the Universe was only 5 per cent of its current age. They are much smaller than the giant systems we see today and suggest that today's galaxies are formed by smaller systems building up into larger ones. (Reproduced by permission of NASA, ESA, R. Windhorst (Arizona State University), and H. Yan (Spitzer Science Centre, Caltech)) (See colour plate)

Bang, are in agreement with the observed values. Any theory that seeks to supplant this model would have to do a better job of addressing each of these points.

3.2 Dark and light matter

What of the matter produced in the Big Bang? Strangely, most of it is invisible to us. The bulk of the material created in the immediate aftermath of the Big Bang is known as *dark matter* because, to the limits of the available data, this material does not emit light at any wavelength. Its presence is inferred by the effect it has on neighbouring matter that can be seen. For example, galaxies rotate much faster in their outer regions than would be expected from the amount of mass that is inferred from the visible galaxy. This suggests that there is dark matter surrounding galaxies in a distribution that is more extended and more spherical than the visible material (see Figure 3.3). Similar arguments apply to clusters of galaxies. Adding up the mass from each member galaxy as well as any intracluster gas (e.g Figure 8.8) results in a value that is far too low to bind the cluster together. These clusters are believed to be bound because the time it would take for the gas to escape and the galaxies to disperse is quite small if they

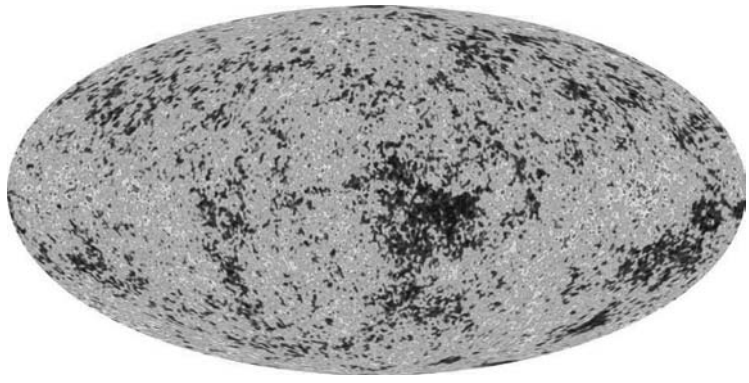


Figure 3.2. Temperature map of the Cosmic Microwave Background (CMB) radiation as measured by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, launched in 2001. This is a *Hammer–Aitoff projection* in which the whole sky is shown in an elliptical projection. The CMB is remnant black body radiation (see Sect. 4.1) from a time 380 000 yr after the Big Bang when the temperature of the Universe cooled to about 3000 K, sufficient for electrons and protons to combine and form neutral atomic hydrogen. At this time, corresponding to a redshift of $z = 1000$ (see Sect. 7.2.2), the Universe went from being opaque to being transparent. The mean observed temperature of the CMB, as we now measure it, is 2.725 K, but tiny fluctuations of temperature are shown in this image of order $+200\,\mu$K (light) to $-200\,\mu$K (dark) (Ref. [16]). The temperature fluctuations are caused by acoustic, or sound waves. The angular scales over which perturbations can be seen provide information that constrains the cosmological parameters of our Universe. (Reproduced courtesy of The WMAP team and NASA)

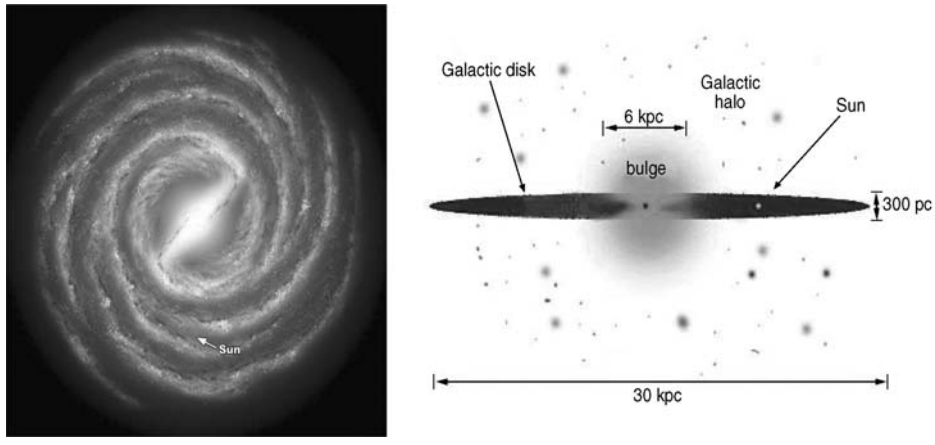


Figure 3.3. Sketch of the Milky Way galaxy face-on (left) and edge-on (right). The Milky Way is a barred spiral galaxy and the Sun is located 8 kpc from the nucleus within the disk. The disk is thin in comparison to its diameter, only about 300 pc thick. It is an active region containing stars, gas and dust, and star formation occurs within it. The stellar distribution becomes more spherical, called the bulge, closer to the centre. Surrounding the disk is a halo of older stars and *globular clusters*, the latter containing typically 10^5 stars each. An even larger distribution of dark matter is believed to encompass this visible material.

are not. Moreover, a study of the bending of light from background sources, called *gravitational lensing* (see description in Sect. 7.3.1) confirms the presence of dark matter in clusters (Figure 7.6) and elsewhere.

Many sensitive searches to detect dark matter have been carried out under the assumption that at least some of the material is *baryonic*. Baryonic matter is any matter containing baryons, which are protons or neutrons[2]. This includes all elements with their associated electrons as well as any object that is thought to have been baryonic in the past[3]. For example, if there are large spherical distributions of dark matter around galaxies, then such a distribution should also be present around our own galaxy, the Milky Way (Figure 3.3). Much observational effort has therefore been expended to measure the gravitational lensing from baryonic objects in the halo of the Milky Way, be they truly dark or just too dim to be directly detectable. The most likely candidates are objects that did not have sufficient mass to become stars (like Jupiter), or the remnants of stellar evolution such as white dwarfs, neutron stars, and black holes (see Sects. 3.3.2 and 3.3.3 for a description of these objects). These have been collectively nicknamed *massive compact halo objects* (MACHOs). The result of such efforts (see Sect. 7.3.2) is that the masses of all MACHOs fall far short of what is required to account for the dark matter.

The failure of observations to detect baryonic dark matter either directly or indirectly as well as other cosmological constraints (see below) have led to the current view that most dark matter consists of unusual matter. Since the dark particles have been so elusive, they must not radiate or interact very easily with other matter, except via gravity. This rules out most known non-baryonic particles. Also, since dark matter does gravitate and has been present since the time the galaxies and stars were forming, it must also have played a part in determining the scale over which structures were established in the Universe. Non-baryonic particles such as neutrinos, therefore, could not make up most of the dark matter because they move at relativistic speeds and the clusters of galaxies that form in a universe dominated by such particles would be too large in comparison to what is observed. These criteria suggest that most of the dark matter must consist of particles not yet observed in Earth-based laboratories or accelerators; such particles are collectively referred to as *weakly interacting massive particles* (WIMPs). Some are predicted by well-established theory, others are more ad hoc. But since classical telescopes have been unable to detect radiation from dark matter, the new 'telescopes' are now WIMP-seeking particle detectors in underground labs[4].

The search for dark matter is no less than a search for most of the mass of the Universe. Unless there is something seriously wrong with our understanding of gravity and current cosmological models, it is estimated that 85 per cent of the matter in the

[2]There are other types of baryons as well but they are unstable, with lifetimes of no more than about 10^{-10} s.
[3]An example of an object that is no longer baryonic is a black hole whose precursor was a massive star (Sect. 3.3.2).
[4]One example is the Project in Canada to Search for Supersymmetric Objects (PICASSO), which is in the labs of the Sudbury Neutrino Observatory (Figure I.3.a), 2 km below ground.

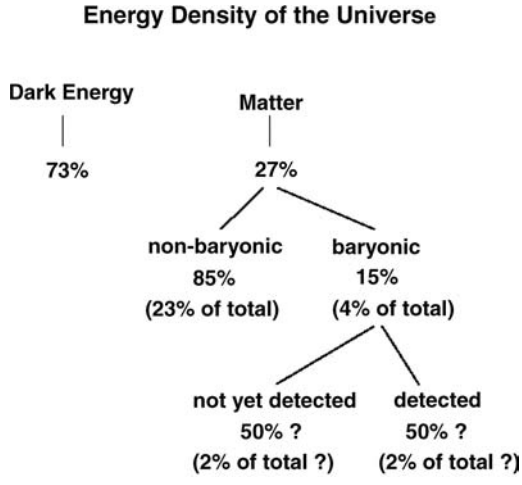


Figure 3.4. Estimates of the components of the energy density of the Universe from a variety of sources (Sect. 3.2). 'Energy density' refers to all forms of energy. 'Dark Energy' is the component, as yet unknown, that causes the apparent acceleration of the Universe. 'Matter' includes all forms of matter, both light and dark, as well as all radiation, the latter being a small fraction at the present time. The values for dark energy and matter (first line) are thought to be known to within a few percentage points. The uncertainty on what fraction of the baryonic matter is considered to have been detected is larger. The percentages of the total are referred to as Ω in Appendix F.

Universe is dark and of unknown nature (see Figure 3.4). Only 15 per cent[5], a value set by knowledge of Big Bang nucleosynthesis, is believed to be baryonic (from Refs. [171], [109], [16]) and potentially visible via radiation. (Note that the energy density of radiation itself constitutes a tiny fraction of the energy density of matter and is included in the matter fractions quoted here.) Of this 15 per cent, however, only half again is accounted for by direct observation, although there is evidence that the remainder may be in hot, intergalactic gas that has been difficult to observe in the past (Ref. [112] and Sect. 8.2.3). Thus, 'missing matter' exists on several levels, from some missing baryonic matter, to the missing non-baryonic matter, the latter presumably in WIMPs.

What is even more astonishing, however, is that recent discoveries (see Appendix F) imply that the Universe is dominated, not by mass, but by a mysterious *dark energy*. All mass, both light emitting matter and dark components, make up only 27 per cent of the total energy content of the Universe. The wonders that we observe in the Universe, from γ-ray bursts, to massive amounts of hot X-ray emitting gas in galaxy clusters, to the myriads of galaxies we see optically in the near and far Universe (Figure 3.1), to cold radio-emitting dust clouds, and all radiation – amount to, at most, a tiny 4 per cent of this total energy (Figure 3.4). The revelation as to what this dark energy and dark matter really are, when it finally comes, promises to revolutionize our understanding of the Universe, and possibly of Physics itself.

[5]This value could be 18 per cent. There is some variation in these quantities, depending on the specific observations.

3.3 Abundances of the elements

3.3.1 Primordial abundance

Of the small fraction of the Universe that makes up all of the stars, galaxies, and other objects that we see today, we still need to account for its constituents. The abundances of the elements can be quantified by their *number fraction* (Figure 3.9) or by their *mass fraction*. The latter is normally expressed as the mass fraction of hydrogen, X, helium, Y, and all heavier elements, Z, respectively,

$$X \equiv \frac{M_H}{M} \quad Y \equiv \frac{M_{He}}{M} \quad Z \equiv \frac{M_m}{M} \tag{3.1}$$

where M_H, M_{He} and M_m are the total masses of hydrogen, helium, and all other elements, respectively, and M is the total mass of the object being considered. Thus,

$$X + Y + Z = 1 \tag{3.2}$$

Since the mass fraction of all elements other than hydrogen and helium tends to be quite low, these are collectively referred to as *metals* and Z is referred to as the *metallicity*.

In the immediate aftermath of the Big Bang, only hydrogen (H), helium (He), and trace amounts of the light elements, deuterium (D), helium-3 (^{3}He) and lithium-7 (^{7}Li) were formed[6]. Although the temperature was extremely high, the expansion of the Universe was so rapid that the density quickly became too low to drive the nuclear reactions required to create heavier elements. The abundances at this time are said to be *primordial* with values (Ref. [114]) of,

$$X = 0.77 \quad Y = 0.23 \quad Z = \text{trace} \tag{3.3}$$

3.3.2 Stellar evolution and ISM enrichment

If the Big Bang could not produce the heavier elements, how then were they formed? The answer lies within stars. As the Universe expanded, inhomogeneities in the matter density resulted in the collapse and formation of the early protogalaxies and the stars within them. The process of star formation involves primarily a gravitational collapse of a molecular gas cloud, but the details of star formation are not entirely understood. Besides gravity, it is likely that turbulence also plays an important role (Ref. [35]).

[6]The numbers specify the atomic weight; for example, normal helium has two protons and two neutrons in the nucleus, whereas helium-3 has two protons and only one neutron. Deuterium is an *isotope* of hydrogen with one proton and one neutron in its nucleus. An isotope of an element has the same number of protons but different numbers of neutrons in the nucleus.

As stars form, the collapse and compression of matter produce an increase in interior temperature and density. If the mass of the object is less than $M_{*_{min}} = 0.08\,M_\odot$, then the interior temperature does not rise sufficiently to ignite *nuclear burning* (*thermonuclear reactions*) in the core. Such objects are referred to as *brown dwarfs* – would-be stars that have never ignited. For higher mass objects, however, once the core temperature rises to a sufficiently high value, (of order $\approx 10^7$ K) the first thermonuclear reactions can begin.

These reactions slowly convert the lighter elements into heavier ones through *fusion*. The energy released by this process both produces the pressure required to halt any further collapse and also causes the star to shine. A current estimate of the star formation rate in the Milky Way is 7.5 stars per year, corresponding to a mass of $4\,M_\odot$ per year (Ref. [49]).

Low mass stars (masses less than about $1.5\,M_\odot$) convert hydrogen into helium in their cores by a series of steps called the *P–P chain* (Figure 3.5) which ultimately turns four protons into a helium nucleus. This is the dominant energy generation mechanism

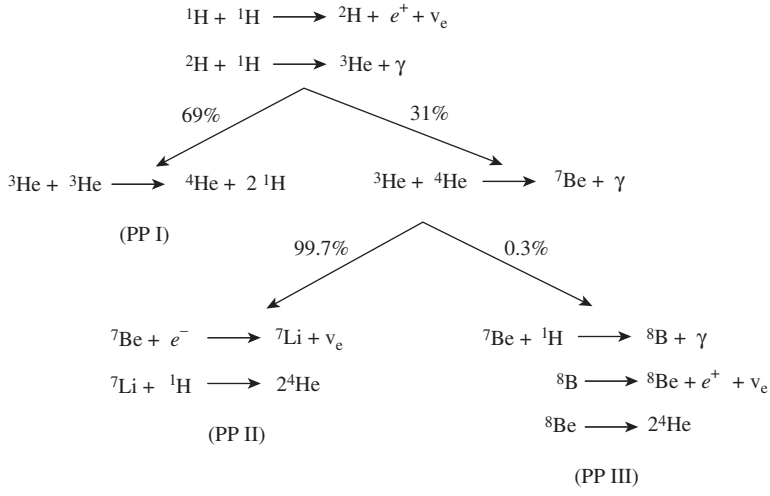


Figure 3.5. The *P–P chain* of nuclear reactions, illustrated here, powers all stars less than about $1.5\,M_\odot$. Note that the intermediate elements created are used up again so the only product is ^{4}He. Two branching arrows indicate that two possible reactions can occur with the probabilities noted. However, the reaction in the PPI branch occurs more quickly than the precursor step to PPII and PPIII so, for the Sun, the production of ^{4}He ends via the PPI branch 86 per cent of the time. Note that the first two reactions need to occur twice in order to drive the PPI reaction. In the Sun, the PPI chain can be summarized as $4H \rightarrow {}^4He + 2e^+ + 2\nu_e$ plus 26 MeV of thermal energy. Of the non-nuclear photons or particles, γ represents a γ-ray photon, e^+ is a positron and ν_e is an electron neutrino. The neutrino in the PPIII chain has historically been an important player in the so-called *solar neutrino problem* in which fewer electron neutrinos from the Sun were observed than predicted by the above steps. This problem was solved in the year 2002 with the definitive detection of different 'flavours' (types) of neutrino by the Sudbury Neutrino Observatory (Figure I.3.a), showing that neutrinos can undergo oscillations and change flavour en route to the Earth (Ref. [106]).

in the Sun, for example, where it takes place in a core region of radius, $\approx 0.2\,R_\odot$. Since the mass of a He nucleus is slightly less than the mass of four free protons, ($4.0026\,u_{\rm amu}$ compared to $4.0291\,u_{\rm amu}$, Table G.2), there is a net mass loss that occurs in the process. This mass loss has an energy equivalent ($\Delta E = \Delta {\rm m}\, c^2$) and it is this energy that powers the stars (Example 3.1, Prob. 3.1). The energy, ΔE, is referred to as the *binding energy* of the nucleus since 'unbinding' the nucleus would require this amount of energy. The longest period of a star's stable lifetime is spent in core hydrogen burning and therefore the timescale set by this process sets the 'lifetime' of a star. It also *defines* the *main sequence* which can be seen as the region from upper left to lower right in the HR diagram in Figure 1.14. On the main sequence, more massive stars are both hotter and more luminous, placing them at the upper left of the figure. More massive stars on the main sequence also burn hydrogen into helium in their cores, but the dominant reactions are via a series of steps that are different from P–P chain. The result, however, is the same. On the main sequence, all stars convert hydrogen into helium in their cores. Data for main sequence stars are given in Table G.7. Main sequences stars are also called *dwarfs* because a star on the main sequence is the smallest it will be compared to later stages in its evolution.

Example 3.1

Estimate the main sequence lifetime of the Sun by assuming that 15 *per cent of its mass is available for core hydrogen burning. Assume that the luminosity is constant with time.*

The available mass is ${\rm M_a} = 0.15\,{\rm X}\,{\rm M}_\odot$, where ${\rm X} = 0.707$ (Eqs. 3.1, 3.4) is the mass fraction of hydrogen in the Sun, so the available number of protons is $N_{\rm a} = 0.11\,{\rm M}_\odot/{\rm m_p}$, where $\rm m_p$ is the mass of a proton. The number of reactions is the number of available protons divided by the number of protons per reaction, $N_{\rm r} = N_{\rm a}/4 = 0.11\,{\rm M}_\odot/(4\,{\rm m_p})$, and the total energy released is the number of reactions times the energy released per reaction, $E = N_{\rm r}\,\Delta E = [0.11\,{\rm M}_\odot/(4\,{\rm m_p})](\Delta {\rm m}\, c^2)$. Using $\Delta {\rm m} = 4.4 \times 10^{-26}$ g and evaluating for the Sun (Table G.3), we find $E = 1.3 \times 10^{51}$ erg. Then the lifetime for constant luminosity is $t = \frac{E}{L_\odot} = 3.4 \times 10^{17}$ s or about 10 Gyr.

The conversion of hydrogen into helium in the core of a star continues until the hydrogen fuel in the core is exhausted. A thick, hydrogen-burning shell then exists around an inert He core. The inert central core, having lost a source of pressure, begins to contract which in turn causes it to heat up. The heating of the core then pumps some extra energy into the hydrogen-burning shell around it. This additional energy is sufficient to produce greater pressure in the shell-burning region which then acts upon the non-burning outer layers of the star, causing them to expand. The resulting increase in the size of the star makes it more luminous and the expansion produces cooling of the outer surface. The surface of the star is what is actually observed, so the star moves upwards and to the right in the HR diagram (Figure 1.14), becoming a *red giant*. This is the cause of the main branch from the main sequence towards the upper

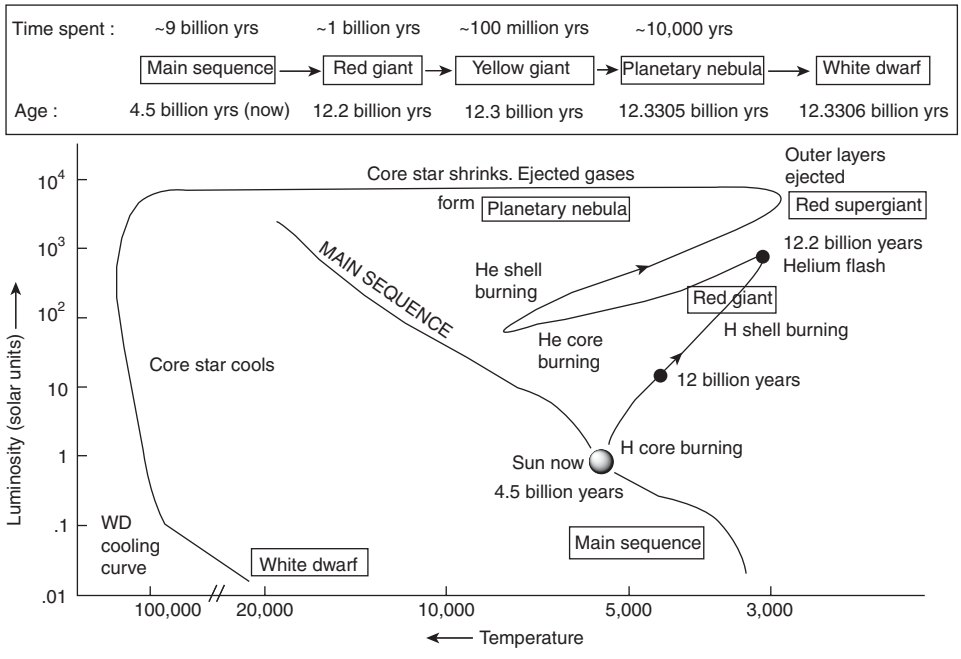


Figure 3.6. Evolutionary track of a 1 $M_\odot$ star in the HR diagram (see Figure 1.14 for an observational analogue). (Adapted from http://skyserver.sdss.org/edr/en/astro/stars/stars.asp)

right in Figure 1.14. Data for giant stars are given in Table G.8 and a theoretical HR diagram with the evolutionary track of a solar-mass star is shown in Figure 3.6.

As the star becomes a red giant, the He core is compressing, its temperature rises, and the core density can become very high. For stars less massive than about 2.3 $M_\odot$ the dense He core becomes *electron degenerate*. This means that the electrons fill the available free energy states, starting from the lowest possible energies and filling each higher energy state sequentially. Only one electron can occupy any given quantum state (the Pauli Exclusion Principle, Appendix C.2) and the occupation of these states becomes independent of temperature. If the atomic nuclei, and the electrons with them, are packed very closely together (small Δx) in the dense core, the Heisenberg Uncertainty Principle (Table I.1) states that the uncertainty in the momentum of the electrons is very high (large Δp). If the uncertainty in the momentum is high, then the momentum itself must be high and so must be the electron velocities. With sufficiently high electron velocities, the electrons themselves can supply enough pressure to hold up the core, dominating over the particle pressure produced by the random motion of nuclei (see Sect. 3.4.1). A typical state of such material is one of extreme density, with the mass of the core containing up to 1.44 $M_\odot$ within a volume approximately equal to the size of the Earth.

For stars with masses between about 0.5 $M_\odot$ and 2.3 $M_\odot$, the core temperature eventually reaches a sufficiently high value (10^8 K) that He burning is ignited. He burning under degenerate conditions occurs quickly (within hours) and is called the *He*

flash, although the rapid flash occurs within the star and therefore is not directly observable. The star, however, takes a sharp change of direction on the HR diagram (see Figure 3.6) as a result. The onset of He burning lifts the degeneracy, the core then stably burns helium, converting it to carbon[7], and the outer layers contract again. Hydrogen burning will still be occurring in a shell around the He-burning core so stars start to become differentiated, with carbon in the core, helium farther out, and hydrogen in the outer layers. A similar evolution results for stars of mass higher than $2.3\,M_\odot$ except that helium burning begins before the core becomes degenerate so there is no internal helium flash. For stars of mass lower than about $0.5\,M_\odot$, the temperature is not expected to become high enough to ignite helium burning. By the red giant phase, these stars have virtually reached the ends of their lives and should expel their outer layers as described below. However, the main sequence lifetimes of such low mass stars are greater than the age of the Universe (Prob. 3.1)! Therefore, we have not yet seen the end products of stellar evolution for stars of such low mass.

The remaining evolution of stars depends on their mass. Stars like our Sun (Figure 3.6) eventually exhaust the helium in the core and evolve to the right again into a *red supergiant* phase (see Table G.9 for data on supergiant stars). The internal and external behaviour mimic the first hydrogen core exhaustion stage but now the carbon core becomes degenerate. The star will then go through a stage of mass loss via strong stellar winds, pulsating instability, and/or other forms of mass loss (see Sect. 5.4.2). Since the greatly expanded outer layers are tenuous and weakly bound gravitationally, they are easily expelled, forming a nebula that can take on a rich variety of morphologies. These are called *planetary nebulae*[8], examples of which are shown in Figure 3.7. The complex nebular morphologies are not completely understood but seem to be related to different phases of mass loss. Circumstellar disks, magnetic fields and the presence of companions may also play a role (Ref. [93]). What remains behind is the hot degenerate core at the centre of the nebula which is now referred to as a *white dwarf*.

Stars more massive than the Sun can continue to go through a slow series of expansions and contractions in response to what is occurring inside them and what temperatures, and subsequent reactions, are achievable. This results in evolutionary tracks on the HR diagram that include repeated motions to the right or left in comparison to what is shown in Figure 3.6. Eventually, all stars with masses less than $\approx 8\,M_\odot$, which amounts to 95 per cent of all stars that have evolved off the main sequence during the lifetime of the Milky Way (Ref. [93], Prob. 3.3), go through the planetary nebula stage, leaving a white dwarf core behind. The white dwarf may contain both carbon and oxygen, but it will never be more massive than $1.44\,M_\odot$, a value called the *Chandrasekhar limit* after the Indian astrophysicist who determined this maximum value (electron degeneracy pressure is insufficient to support a mass that

[7]This is called the *triple alpha process* because three He nuclei, also known as α *particles*, eventually become a carbon nucleus.

[8]The term 'planetary' is used for historical reasons. The first planetary nebulae that were discovered were of small angular size and showed a blue-greenish hue which was similar to the colour of the outer planets, Uranus and Neptune.

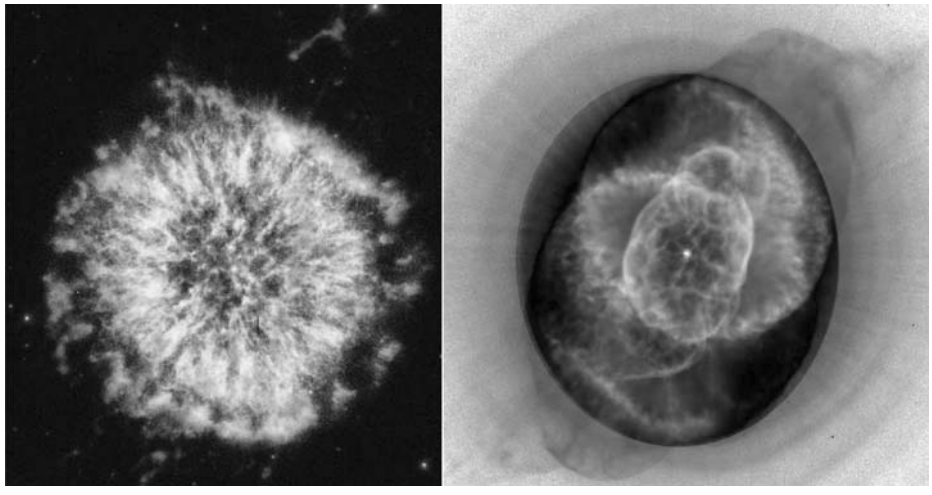


Figure 3.7. Examples of planetary nebulae, the end result of the evolution of low mass stars like our Sun. On the left is NGC 6751 and on the right is NGC 6543 (the Cat's Eye Nebula). The colour scheme is: red = [N II] (λ 6584), green = [O III] (λ 5007 A) and blue = V band. The square brackets mean that the spectral lines are 'forbidden' (see Sect. 9.4.2). (Reproductions by permission. For NGC 6751: Arsen R. Hajian, USNO, Bruce Balick, U Washington, Howard Bond, STScI, Nino Panagia, ESA, and Yervant Terzian, Cornell; For NGC 6543: B. Balick, J. Wilson, and A. R. Hajian) (See colour plate)

is higher than this). As indicated earlier for the degenerate core of a star, a white dwarf is very dense, its mass having been compressed into a size which is only about the size of the Earth.

The material expelled as a planetary nebula will be enriched in heavy elements that have been formed in the interiors of the progenitor stars. From the enriched expelled material, new stars will form. Thus, as generations of stars are 'born' and 'die', the metallicity of the ISM (and of the Universe) increases with time.

3.3.3 Supernovae and explosive nucleosynthesis

A remaining event that contributes to metallicity and is responsible for elements heavier than iron (Fe) is a *supernova explosion*[9]. Stars of mass $\gtrsim 8\,M_\odot$ do not die out gradually as do lower mass stars, but rather explode, spewing their outer layers into the interstellar medium (ISM) at speeds of tens of thousands of km/s. This is called a *Type II supernova*[10], one of the most energetic events in nature (Figure 3.8). A typical supernova kinetic energy is 10^{51} erg. Not only does the internal metal-rich material enter the ISM, but also the explosion itself is so energetic that it drives many more

[9]It has also become apparent that massive stars in a very large red giant phase can spew heavy elements more massive than Fe into the ISM via stellar winds. The heavy elements may dredge up from the interior.

[10]There are two main types of supernovae which occur in different environments and by different processes. We describe only Type II here. The Type I supernovae are more luminous and are used as distance indicators as discussed in Appendix F. They also contribute to the abundances of the heavy metals, including those more massive than Fe.

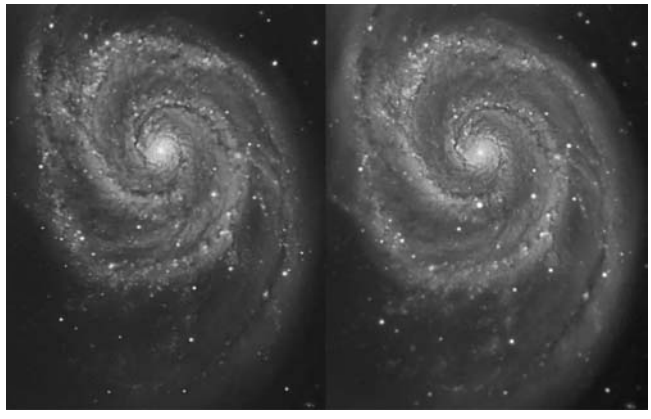


Figure 3.8. The galaxy, M 51, before (left) and after (right) a Type II supernova went off in its disk. The supernova, named SN 2005cs, was discovered in June, 2005 by Wolfgang Kloehr using an 8 inch telescope, and can be seen directly below the nucleus in the spiral arm closest to the nucleus. A supernova explosion takes only a fraction of a second but maximum light is achieved 2 to 3 weeks later. The brightness of a Type II may then plateau or decline in brightness over about 80 days followed by a continuing slow decline thereafter. This galaxy is 9.6 Mpc (3.3×10^6 ly) distant so the supernova detected in 2005 actually occurred 3.3 Myr earlier. (Reproduced by permission of R. Jay GaBany, Cosmotography.com)

reactions, a process called *explosive nucleosynthesis*. It is in this explosive event that elements heavier than Fe are formed.

The reason for the extraordinarily different fate of high mass stars lies in the nature of intranuclear forces. The force that holds the nucleus together is called the *strong force* and it acts on small ($\approx 10^{-13}$ cm) scales between *nucleons*[11], i.e. baryons that are in an atomic nucleus. For light elements with few nucleons, each 'feels' the force of every other nucleon and so as the number of nucleons grows, so does the binding force per nucleon. The higher the binding force per nucleon, the more stable is the element and the greater is the mass deficit when it forms. Eventually, however, the nucleus becomes so large that Coulomb forces (in this case, the electrostatic repulsion between protons) which act on larger scales, become important. Then as the nucleon number grows, the binding energy per nucleon decreases and the nucleus becomes less stable. The consequence of this behaviour is that light elements release energy when undergoing nuclear fusion, such as in stars, and heavy elements release energy during nuclear *fission*, such as currently occurs in nuclear reactors on Earth[12]. The element

[11]The strong force actually acts between *quarks* that are the constituents of nucleons.

[12]The first steps towards creating a *fusion* reactor have been taken in the establishment of ITER (Latin for 'the way'), an international project located in France. Because only low pressures, in comparison to the centre of the Sun, are possible, ITER will operate at 1.5×10^8 K, an order of magnitude higher than the temperature at the centre of the Sun! If successful, future reactors will be like 'mini-Suns', available for power.

at which the binding energy per nucleon is highest and at which this transition occurs is iron[13].

For stars of mass $\gtrsim 8\,M_\odot$, the core temperature achieves a value high enough to initiate fusion reactions of Fe. However, rather than releasing energy during this reaction (called *exothermic*), the reactions now *require* energy (called *endothermic*). With no energy source in the core and available energy going into the endothermic reaction, the core pressure is removed and the core instantly implodes. The result of the core implosion is a remnant, either a *neutron star*, a star so dense that the mass of a Sun or more has been compressed into a volume only tens of km across[14] or a *black hole*, a region of space around a singularity which contains the mass of several Suns[15]. Although the details of supernovae are not perfectly understood, it is believed that the explosion of the outer layers is enabled by shock waves[16] that are triggered by the core collapse. The resulting material, enriched in heavy elements through explosive nucleosynthesis, contributes to the metallicity of the interstellar medium (ISM). The supernova itself can become extremely bright (Figure 3.8), sometimes outshining the light from the entire parent galaxy. Its light then declines in a characteristic exponential fashion with time, though changes in slope can also occur (see also Sect. 9.1.3). Such a plot of a variation in light with time is called a *light curve*. Eventually, the exploded outer layers become visible as a *supernova remnant*, such as the one shown in Figure 1.2.

3.3.4 Abundances in the Milky Way, its star formation history and the IMF

The elemental abundances of our Solar System, expressed as a number fraction, measured from Solar spectral lines (see Figure 6.4) and meteorites, are shown in Figure 3.9. This plot represents the abundances in the local Galactic neighbourhood at the time the Solar System formed. The values thus reflect the origin of the light elements in the Big Bang as well as subsequent evolution or other processes that could alter these relative abundances since that time. Enrichment of heavy elements by previous generations of stars is clearly the most important effect, the current supernova rate for the Milky Way being estimated at about two per century. However, other processes are also involved. For example, high energy cosmic rays (Sect. 3.6) can collide with nuclei, breaking up heavier nuclei into smaller pieces, a process called *spallation*. In addition, the stability of nuclei against radioactive decay will also affect

[13]The most tightly bound element is actually ^{62}Ni, but this element is not very abundant and therefore does not play an important role in supernova explosions.

[14]In such a remnant, the density is so great that electrons have combined with protons to produce a 'sea' of neutrons; the density of a neutron star is approximately equal to the density of a single neutron and the star is held up from further collapse by *neutron degeneracy*, the nuclear version of electron degeneracy. There is some uncertainty about the upper limit to the mass of a neutron star but a reasonable estimate is $3\,M_\odot$.

[15]For larger core masses, there is no known force that can withstand a collapse to this state.

[16]Shock waves are sharp, rapidly moving density discontinuities that are produced when the motions of particles are faster than the speed of sound in the medium.

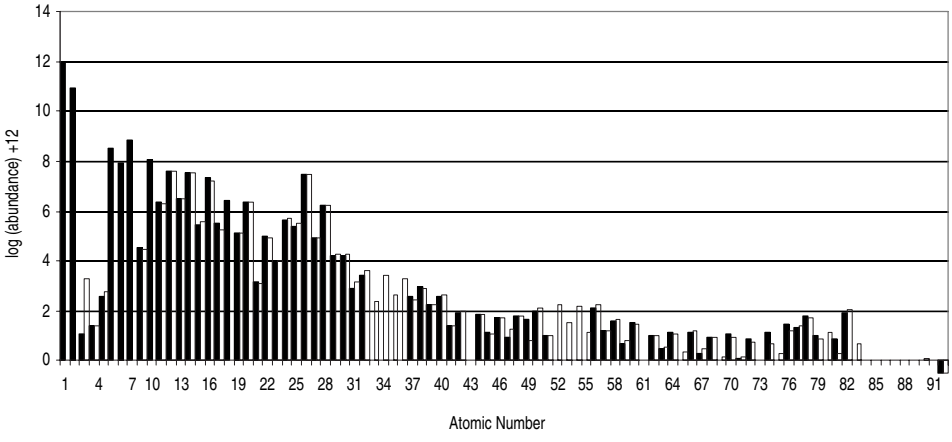


Figure 3.9. Logarithmic plot of elemental abundances in the Solar System as a function of atomic number. Black boxes represent values for the Sun and clear box values are for meteorites. In several places there are no data, for example a measurement for H in meteorites or for short-lived radioactive elements. All abundances have been normalized to hydrogen which has a value of 12 in this plot. For example, a value of 6 on the graph corresponds to an abundance of 10^{-6} with respect to hydrogen. Abundances refer to the numbers of particles, rather than their weights. Data from Ref. [154].

these abundances. Consistent with an origin in the Big Bang and in spite of all subsequent evolution, the abundance of H is still, by far, greater than all the other elements combined, constituting approximately 90 per cent, by number, of the Universe. Normalized to a hydrogen abundance of 1, He constitutes only 8.5 per cent, and Li is a mere 10^{-9} per cent[17]. Even the stable Fe (atomic number = 26) has an abundance only 3×10^{-3} per cent of hydrogen. The most abundant of the metals is oxygen at 6.8×10^{-2} per cent. Given the dominance of hydrogen in our Universe, it is important to understand this simple atom extremely well. A review of the properties of hydrogen is given in Appendix. C.

The corresponding mass fractions for the abundances illustrated in Figure 3.9 (Ref. [44]), referred to as *Solar abundance*, are,

$$\mathrm{X} = 0.707 \pm 2.5\% \quad \mathrm{Y} = 0.274 \pm 6\% \quad \mathrm{Z} = 0.0189 \pm 8.5\% \tag{3.4}$$

From measurements of metallicities in interstellar clouds and other stars, Solar abundances appear to be typical of abundances elsewhere in the disk of our Milky Way.

As has been seen with the production of heavy elements within stars, there are many processes involved in the slow conversion of the primordial abundances of

[17]Li tends to be destroyed in stellar nuclear reactions.

Eq. (3.3) to the Solar abundances of the Galactic disk. The increase in metallicity depends on the amount, rate, and metallicity of material expelled into the ISM by planetary nebulae and supernovae and the heaviest elements originate in supernovae and the winds of massive red giants. Since the end points of stellar evolution depend on stellar mass, it is important, then, to ask how many stars in different mass ranges are actually formed.

The admixture of stars of different masses, when first formed, is called the *initial mass function* (IMF). Various IMFs have been proposed in the astronomical literature, attesting to the difficulty in obtaining this function and its possible variation with environment. Good examples are shown in Figure 3.10. The functional forms for the

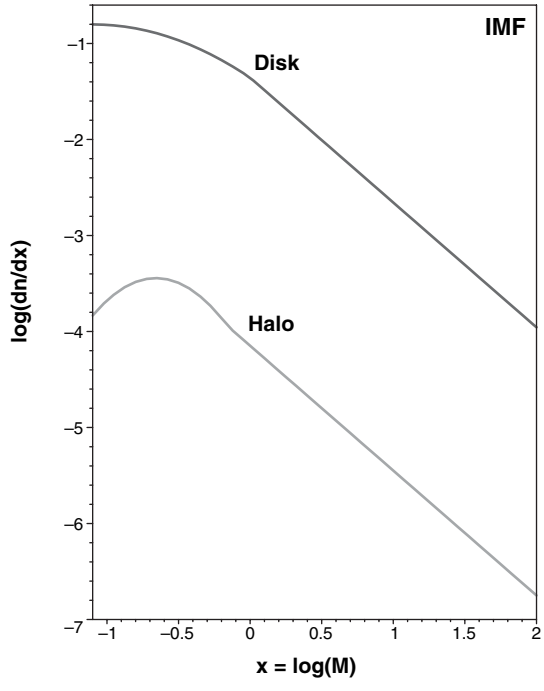


Figure 3.10. The initial mass function (IMF) for the disk and stellar halo of the Milky Way. Masses are in units of $M_\odot$. The IMF, given by $dn/d[\log(M)]$ (pc^{-3} $[\log(M_\odot)]^{-1}$) represents the number density of stars per logarithmic mass interval that have formed over the history of the region being studied. It is determined by counting the number density of stars in a given luminosity interval, applying corrections for observational biases, converting to the number density of stars in a given mass range, and then applying corrections for the number of stars that have evolved, i.e. that are no longer on the main sequence. The functional expressions are provided in Eqs. (3.5) and (3.6). There are some significant error bars (not shown) associated with the fitted parameters, but the plot provides a good idea as to how many stars of different masses form in the Milky Way. There are clearly many more stars per unit volume in the disk than in the halo and many more low mass stars in comparison to high mass stars. Star formation via gravity alone cannot reproduce the details of these plots and it is likely that turbulence plays an important role. From Ref. [35].

two curves, applicable to the Milky Way's disk and halo, respectively, are,

$$\left[\frac{dn}{dx}\right]_{\text{disk}} = 0.158\ \exp\left[-\frac{1}{2}\left(\frac{x+1.10}{0.69}\right)^2\right] \quad \text{M} < 1.0\,\text{M}_\odot$$
$$= 4.4\times10^{-2}\,\text{M}^{-1.3} \qquad \text{M} \geq 1.0\,\text{M}_\odot \tag{3.5}$$

$$\left[\frac{dn}{dx}\right]_{\text{halo}} = (3.6\times10^{-4})\ \exp\left[-\frac{1}{2}\left(\frac{x+0.66}{0.33}\right)^2\right] \quad \text{M} < 0.7\,\text{M}_\odot$$
$$= (7.1\times10^{-5})\,\text{M}^{-1.3} \qquad \text{M} \geq 0.7\,\text{M}_\odot \tag{3.6}$$

where $x = \log(\text{M})$ and M is the stellar mass in Solar mass units. The term, dn/dx, in units of $\text{pc}^{-3}\ [\log(\text{M}_\odot)]^{-1}$, gives the number density of stars per cubic parsec per logarithmic mass interval (Ref. [35]).

Figure 3.10 illustrates that far fewer high mass stars are formed in comparison to those of low mass (Example 3.2, Prob. 3.2). However, high mass stars that end their lives as supernovae also live shorter lives since they burn their energy at a faster rate while undergoing normal hydrogen fusion (Prob. 3.1). As long as the region of the ISM is rich in gas and able to form stars, this means that there will be many successive generations of massive stars forming and dying while low mass stars like our Sun continue to 'simmer' in the stable state of burning hydrogen to helium in their cores. Therefore, even though there are fewer high mass stars, the massive stars will have the greatest influence on the metallicity of the ISM. Since high mass stars are short-lived (of order 10^6 yr), they will also not stray far from their places of origin during their lifetimes. Therefore, the presence of high mass stars and any observational consequences of high mass stars alone, such as HII regions (which only exist around hot massive stars, see Sect. 5.2.2) or supernovae, are taken to be proxies for active star formation. We say that hot high mass stars, HII regions and supernovae are all *tracers* of massive star formation. It is important to note that, once some time has passed, the admixture of stars in a region will not be the mixture of Figure 3.10 because of the different rate of evolution of stars of different masses. Rather, the admixture of stars is determined by noting which stellar mix results in the observed HR diagram for a given region.

Example 3.2

How many disk stars per pc^3 with masses between 0.1 *and* 1 $M_\odot$ *are initially formed in comparison to those between* 10 *and* 100 $M_\odot$*?*

From Eq. 3.5 (see also Figure 3.10) with $x = \log$ M, in the low mass range,

$$n = 0.158\int_{x=-1}^{x=0}\exp\left(-\frac{1}{2}\left[\frac{(x+1.10)}{0.69}\right]^2\right)dx\ =\ 0.106\,\text{pc}^{-3}.$$

In the high mass range,

$$n = 4.4 \times 10^{-2} \int_{x=1}^{x=2} \mathrm{M}^{-1.3}\, dx = 4.4 \times 10^{-2} \int_{x=1}^{x=2} 10^{-1.3x}\, dx \; = 7.00 \times 10^{-4}\, \mathrm{pc}^{-3}$$

Therefore, there are 151 times more stars formed per cubic parsec in the low mass range than the high mass range.

Abundances in our Milky Way are very important probes of its star formation history and, by implication, its global formation and evolution. For example the IMF (Figure 3.10), the quantity of gas present, and knowledge of stellar evolution allow models to be built that predict the metallicity distribution in a given environment now. The modelled result can then be compared to that observed. If one assumes an initially low metallicity and that the amount of gas in a given region of the Galaxy remains at its location without flowing out or other gas flowing in (called the *closed box model*), then one finds that too few stars of low metallicity are observed in comparison to what the models predict[18]. This means that some modification is necessary in the assumptions of the model. It would appear that some additional enrichment of metals is required, possibly from inflowing gas, or possibly from metal-rich gas from previous generations of stars that had stronger stellar winds and/or mass loss than expected.

When we look at the metallicity of stars in the disk in comparison with the halo, another important result emerges. Away from the gaseous, dusty spiral-like disk within which most of the visible material of our Galaxy is found, there is a more spherical halo of stars surrounding the disk in which there is very little gas and no cold dense gas from which stars can form (see Figure 3.3). The metallicities of halo stars vary, but a typical value is $\mathrm{Z} = 0.0005$, less than 3 per cent of the disk value. Clearly, the metallicity in the halo is very low in comparison to the disk and this provides a clue as to the comparative star formation history of these two regions.

The much lower density of stars, as shown by the IMF (Figure 3.10) also attests to the very different environment and history of the halo. Star formation in our Galaxy has mostly occurred and is still occurring in its disk but it is no longer occurring in its halo which now contains the oldest stars in the Galaxy. Although the metallicity in the halo is very low, it is not zero, indicating that some star formation has indeed occurred there in the past. In fact, similar results are found for other galaxies and imply that a period of rapid star formation and enrichment occurred very early in the galaxy formation process when smaller systems were joining to form larger ones (Figure 3.1). The thin galaxy disks that we see today with their beautiful spiral structure (Figures 2.11, 3.8) likely formed later when baryons settled into the central regions. Star formation then continued in the disk, forming a denser stellar environment there (Prob 3.2). A more complete review of galaxy formation can be found in

[18]This is called the *G dwarf problem.*

Ref. [59] and a simulation of the early structures that evolve into galaxies can be seen in Figure 5.7.

Dynamics, global structure and stellar distribution all provide information about the formation of galaxies, but abundance is a key component. There is much that is still not fully understood about the formation process, and the fact that dark matter, of which we know little, is the dominant component is a strongly complicating factor. Nevertheless, even the tiniest of visible constituents and apparently insignificant observations such as metallicity variations can offer just the insights required to piece together a coherent picture of Galactic history.

3.4 The gaseous universe

The two most abundant elements in our Universe, hydrogen and helium, are gases. Only under extreme pressure, such as at the centres of planets like Jupiter, can hydrogen exist as a metallic-like liquid. It follows, then, that most of the Universe is in a gaseous state. Indeed, most of the metals are also gaseous, given the pressures and temperatures under which they may be found, for example, silicon within the Sun, gaseous carbon in interstellar clouds, or iron atoms in hot tenuous gas. Only a tiny fraction of the observed Universe is in the solid state in the form of dust particles and the planets, asteroids, and comet nuclei that we see in our own Solar System or infer to be in others. Once established, the internal tensile strength of solids can keep them in this phase even in the low pressure environment of space. Ices of various molecules such as water, carbon dioxide, and methane, are also a common constituent of solid Solar System material, especially in comets or on cold bodies, distant from the Sun. The liquid phase, however, appears to be rarest. The Earth has a special place in our Solar System in that its temperature and surface pressure allow water to exist in all three phases on its surface. There is now mounting evidence that liquids can flow on the surfaces of other planets and moons, for example, via volcanic flows. Liquid water may leak or be vented from under the icy surfaces of Europa, a moon of Jupiter, and Enceladus, a moon of Saturn. And liquid water *may* have flowed on Mars. However, conditions are not now suitable for long–lived 'lakes' or 'rivers' on these objects. A direct observation of stable surface liquids is still, therefore, elusive. The only other known body on which surface liquid may be present is Titan, another moon of Saturn and this liquid would be methane (see Figure 3.11).

The gases of which the Universe is largely composed can show a wide variety of properties, especially of temperature and density. A cold, dusty dense molecular cloud, for example, is quite different from the hot upper atmosphere or the deep interior of a star. Table 3.1 gives some typical densities and temperatures for various astrophysical gases, and images of a few different kinds of gaseous objects that can be found in the ISM are shown in Figures 1.2, 3.7, and 3.17. Aside from density and temperature, another important difference is the *state of ionization*. Gases can be neutral or ionized or have some partial ionization. A neutral gas is one in which the atoms have their electrons still attached or *bound* to them so that there is no net charge on the atom. In

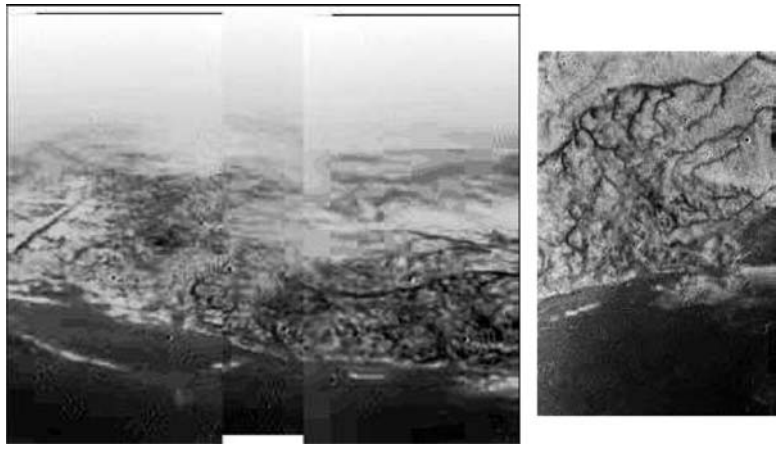


Figure 3.11. On January 14, 2005, the European Space Agency's Huygens space probe took these pictures of the surface of Titan while descending to its surface. Titan is Saturn's largest moon and the one shrouded in a thick, opaque atmosphere that is predominantly nitrogen with ≈ 5 per cent methane (CH_4). Since the surface temperature and pressure on Titan is $T = 94$ K and $P = 1.6$ atm, respectively, and since the triple point (the temperature and pressure at which all three phases can exist) for methane is $T = 91$ K, $P = 0.12$ atm, conditions on Titan appear favourable for methane to exist in the solid, liquid and gas phases. The left image was made from a number of images taken from a height of about 8 km and have a spatial resolution of about 20 m. The 'shoreline' shown in this image separates two regions, both of which are believed to be solid. The right image shows tantalizing evidence that liquids have at some time flowed on the surface, possibly recently and possibly seasonally. Newer images of the north polar region of Titan also suggest that 'lakes of methane' may be present (Ref. [26]). (Reproduced by permission of ESA/NASA/Univ. of Arizona, Brown, D., Hupp, E., and Martinez, C., 2006, NASA News Release: 2006/097)

Table 3.1. Sample densities and temperatures in astrophysical gases[a]

Location	n_e[b] (cm^{-3})	$n_m + n_H$[c] (cm^{-3})	T (K)
Interplanetary Space	$1 - 10^4$	≈ 0	$10^2 - 10^3$
Solar corona	$10^4 - 10^8$	≈ 0	$10^3 - 10^6$
Stellar atmosphere	10^{12}	≈ 0	10^4
Stellar interiors	10^{27}	≈ 0	$10^{7.5}$
Planetary nebulae	$10^3 - 10^5$	≈ 0	$10^3 - 10^4$
HII regions	$10^2 - 10^3$	≈ 0	$10^3 - 10^4$
Interstellar space[d]	$10^{-3} - 10$ (avg. ≈ 0.03)	$10^{-2} - 10^5$ (avg. ≈ 1)	10^2
Intergalactic space	$< 10^{-5}$	≈ 0	$10^5 - 10^6$
Intergalactic HI clouds[e]		$10^{-6} - 10^{-3}$	$10^3 - 10^5$

[a]Ref. [96] unless otherwise indicated.
[b]Electron density. For a pure hydrogen gas that is completely ionized, $n_e = n_p$ where n_p is the proton density.
[c]n_m is the molecular gas density (predominantly H_2) and n_H is the neutral atomic hydrogen (HI) gas density.
[d]The temperature quoted here is typical of interstellar HI clouds. However, there is a wide variety of densities, temperatures and degrees of ionization in the interstellar medium. For example, molecular clouds tend to have low temperatures (5–200 K, typically 30 K) and high densities (10^2–10^5 cm^{-3}, typically 10^4 cm^{-3}). On the other hand, for ionized diffuse intercloud gas at $n_e \approx 0.03\,cm^{-3}$ the temperatures are much higher ($\approx 10^4$ K).
[e]Typical density range for the Ly α forest (see Figure 5.6 and Ref. [129]) for a redshift (Sect. 7.2.2) of $z = 3.1$. The ionized fraction is not easily determined. Values will vary with redshift.

gases with higher states of ionization (partially ionized) some atoms will have lost electrons, becoming *ions* with a net charge. A highly ionized gas is referred to as a *plasma* in astrophysics. Since hydrogen is the most abundant atom, a plasma will have approximately equal numbers of free protons and free electrons, with some variation depending on the admixture of other elements and their state of ionization. The degree of ionization of any given element is specified by Roman numerals after the element name, where the number of electrons removed is given by the Roman numeral minus one. For example, neutral hydrogen is referred to as HI and ionized hydrogen is HII. For metals, there are many more electrons that could leave the atom. Thus, one can have, for example, neutral carbon, (CI), singly-ionized carbon (CII), or triply-ionized carbon (CIV). Many such ions have been observed in various astronomical objects, including some with extremely high states of ionization. For example, iron with all electrons removed but one (Fe XXVI) has been observed in the spectra of some highly energetic X-ray sources (e.g. Ref. [40]), including the Solar *corona*[19] (see Figure 6.7).

Important goals in astrophysics are to determine the densities, temperatures, and ionization states of gases which include, of course, the atmospheres and interiors of stars. With sufficient information, the energetics involved in maintaining these gases in the observed states can then be determined. Armed with this knowledge, it may be possible to piece together a picture of the formation and evolution of the objects. Thus, there is considerable importance in understanding the physics of gases, especially of hydrogen. Appendix C outlines the basic physics of the hydrogen atom and the sections that follow highlight some relevant astrophysics of gases.

3.4.1 Kinetic temperature and the Maxwell–Boltzmann velocity distribution

Consider a gas in which energy has been exchanged between particles via *elastic collisions*. This means that only kinetic energy is exchanged; there are no radiative energy losses, ionizations, or other processes occurring between the particles. In such a case, a velocity distribution is set up such that the number density of particles in each velocity interval is described by the Maxwell–Boltzmann, or Maxwellian velocity distribution. The functional form of this distribution was presented in Eq. (0.A.2) and an example is plotted in Figure 3.12.

It is the Maxwellian velocity distribution that defines the *kinetic temperature*, written as T without a subscript, of a gas. For particles of mass, m, T is related to the average of the squares of the particle speeds, or the *mean-square* particle speed, $\langle v^2 \rangle$,

$$\frac{1}{2}\,\mathrm{m}\,\langle v^2 \rangle = \frac{3}{2}\,k\,\mathrm{T} \tag{3.7}$$

[19]The Solar corona is the faint, low density, outermost region of the Sun's atmosphere, visible to the eye during a total Solar eclipse. Coronal temperatures are typically 10^6 K but can be higher during times of high activity (Sect. 6.4.2).

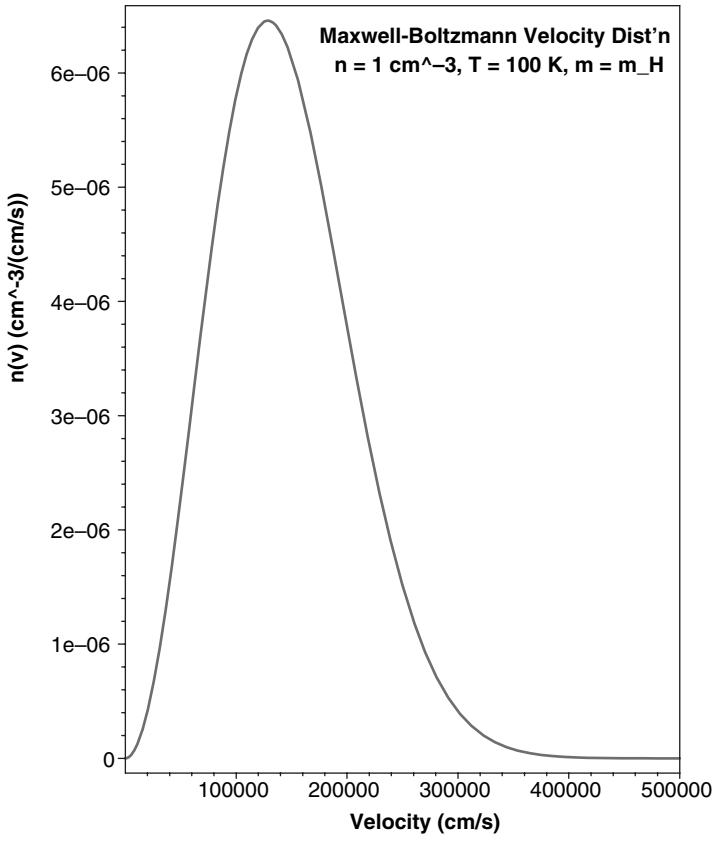


Figure 3.12. Maxwellian velocity distribution for neutral atomic hydrogen at T = 100 K and a total density of 1 cm^{-3}. The ordinate specifies the number of particles in each velocity interval and the abscissa gives the velocity. The area under this curve returns the total density. See Eq. (0.A.2) for its functional form.

The *root-mean-square* particle speed is then $v_{\rm rms} = \sqrt{\langle v^2 \rangle}$. It can be shown (Prob. 3.4) that the *most probable speed* is slightly lower,

$$v_{\rm mp} = \sqrt{2\,k\,\mathrm{T}/\mathrm{m}} \tag{3.8}$$

and the *mean speed* is higher,

$$\langle v \rangle = \sqrt{8\,k\,\mathrm{T}/\mathrm{m}} \tag{3.9}$$

The left hand side of Eq (3.7) represents the mean kinetic energy of the particles, $\langle E_{\rm k} \rangle$, so T is a measure of the mean kinetic energy of particles in a gas. A gas that can be

described by a single kinetic temperature is said to be in *thermal equilibrium*[20]. Such a result would occur if a hot gas and a cold gas were set beside each other in an insulated box. Eventually both gases would reach a single warm temperature.

The Maxwell–Boltzmann distribution can be derived from *statistical mechanics* which applies statistics to the motions of large numbers of particles in a gas in order to predict its macroscopic properties. The resulting equation (Eq. 0.A.2) can be seen to depend on a number of parameters, including the *Boltzmann factor*,

$$e^{-\chi/(k\,\mathrm{T})} \tag{3.10}$$

where χ is an excitation energy. In this case, $\chi = \frac{1}{2}\ \mathrm{m}\ v^2$ represents the kinetic energy of a particle. The Boltzmann factor indicates that there is a lower probability of finding a particle in a given excitation state as the excitation energy increases in comparison to the mean kinetic energy. In addition, however, the Maxwellian velocity distribution also depends on the quantity, $4\,\pi\,v^2$ which represents the surface area of a shell about an origin in three-dimensional 'velocity-space'. As the velocity (or kinetic energy) of a particle increases, the velocity shells become larger, allowing more possible orientations of the velocity vector of a particle. This means that more 'states' are available to be occupied as the velocity increases. These competing effects result in the final Maxwellian distribution that first increases as v^2, reaches a peak then decreases exponentially (Figure 3.12). The constant multiplier in the equation is a normalization factor to ensure that the total number density of particles is correct.

If a gas can be described by a Maxwellian velocity distribution, then there must have been a sufficient number of elastic interactions that this thermal equilibrium could be established. How valid is this assumption for typical astrophysical conditions? For gases in which T is less than about 8×10^4 K, particle encounters are almost always elastic and there will therefore be many elastic collisions before an inelastic encounter occurs (Ref. [160]). The timescale required for establishing an equilibrium temperature, t_{therm}, can be determined by asking how long it would take for a faster test particle (the hotter gas particle) to slow down as it diffuses through the gas of slower particles (the colder gas particles)[21]. It can be shown that this timescale is quite short under astrophysical conditions and depends on the density of the dominant constituent of the gas. For example, in cool ($\mathrm{T} = 100$ K) neutral atomic (HI) clouds in the interstellar medium, the timescale for establishing an equilibrium temperature via H–H collisions is $t_{\mathrm{therm}} = 16/n_{\mathrm{H}}$ yr, where n_{H} is the hydrogen density. If $n_{\mathrm{H}} = 1\ \mathrm{cm}^{-3}$ (Table 3.1) then only 16 years are required, a very short time by astronomical standards. For a fully ionized hydrogen gas at $\mathrm{T} = 10^4$ K and the same density, the equilibrium timescale for electron–proton 'collisions'[22] is only $t_{\mathrm{therm}} = 8.2 \times 10^3$ s, that is, only a few hours. The

[20]The term, 'thermal equilibrium', is sometimes used differently from here, for example, to represent energy balance within stars.

[21]The equilibrium timescale is one half of the slowing down time since the slower moving particles will also speed up. Such a calculation involves, as a starting point, determining the mean time between collisions, an example of which is provided in Sect. 3.4.3.

[22]For coulomb interactions, a collision is considered to be an interaction that significantly affects the trajectory of a particle.

timescale for electron–electron collisions is equivalent to that for electron–proton collisions, provided the electron and proton densities are equal in the gas. Clearly, higher densities facilitate the process of temperature equalization and Coulomb interactions shorten the timescale even more. In even the low density ISM, then, a Maxwellian velocity distribution is expected and therefore it should also present in regions of higher density, such as stellar atmospheres (e.g. densities of order $10^{12}\,\mathrm{cm}^{-3}$).

We therefore assume, in most cases, that astrophysical gases are in thermal equilibrium. Departures can occur, however, especially in gases of very low density. There may also be other physical effects that inhibit the process of temperature equalization. If the gases are physically prevented from diffusing together, thermal equilibrium will not be achieved (for example, magnetic fields can constrain the positions of charged particles in an ionized gas). Particles whose energetics are driven by processes that are intrinsically non-thermal in nature, such as cosmic rays (Sect. 3.6), will not achieve a Maxwellian distribution at all. Collisions of gas particles with dust particles require revisions from the above description as well. Figure 3.13 shows a situation common for an HII region in which dust is mixed with the ionized gas, yet the dust temperature may be two orders of magnitude lower than the kinetic temperature of the gas. Dust temperature, as quoted, is conceptually different from the gas temperature, however, since it does not refer to the random speeds of the dust particles but rather a temperature at which the dust grain radiates; this will be further elucidated in Sect. 4.2.

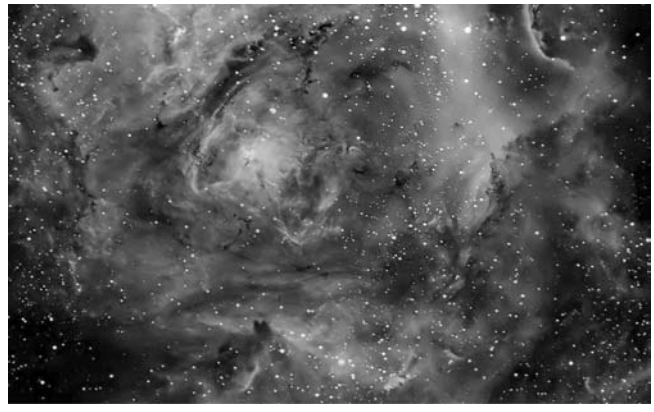


Figure 3.13. This image shows most of the *Lagoon Nebula* (also called Messier 8), which is an ionized hydrogen region (HII region) visible to the naked eye in the constellation of Sagittarius. The nebula is 1.6 kpc away with an average diameter of 14 pc (there is some uncertainty in the distance) and the complete nebula appears as large as the full Moon in angular size. The image shows a complexity of features with colours representing emission from different atoms (oxygen, silicon and hydrogen) and dark bands and filaments representing obscuration by dust. The dust and gas are in close proximity yet the gas temperature is 7500 K whereas dust temperatures are typically much lower, of order $\sim$ 100 K or less. The mean density of electrons in the nebula is $n_{\mathrm{e}} \approx 80\,\mathrm{cm}^{-3}$. (Reproduced by permission of Michael Sherick, Cabirllo Mesa Observatory, 2005) (See colour plate)

3.4.2 The ideal gas

An *ideal gas* is a gas that obeys the *ideal gas law* (or *perfect gas law*, see also Table 0.A.2 and Prob. 0.3),

$$\mathrm{P_p} = n\,k\,\mathrm{T} = \frac{\rho}{\mu\,\mathrm{m_H}}\,k\,\mathrm{T} \tag{3.11}$$

where $\mathrm{P_p}$ is the particle pressure, n is the number density ($\mathrm{cm^{-3}}$) of particles of any mass, k is Boltzmann's constant, T is the temperature, ρ is the mass density of the gas ($\mathrm{g\,cm^{-3}}$), and $\mathrm{m_H}$ is the mass of the hydrogen atom. The quantity, μ, is called the *mean molecular weight*, defined by

$$\mu \equiv \frac{\langle \mathrm{m} \rangle}{\mathrm{m_H}} = \frac{1}{\mathrm{m_H}\,N} \sum_{\mathrm{i}} \mathrm{m_i} \tag{3.12}$$

where m_i is the mass of the ith particle, and N is the total number of particles. The quantity, μ, gives the mean mass of free particles in units of $\mathrm{m_H}$ (Example 3.3). The mean molecular weight can also be expressed in terms of the abundances of the object. For example, the mean molecular weights of a completely neutral and a completely ionized gas are (Prob. 3.5), respectively,

$$\begin{aligned} \mu &= \frac{1}{\mathrm{X} + \frac{\mathrm{Y}}{4} + \frac{\mathrm{Z}}{\bar{\mathrm{A}}_\mathrm{m}}} \quad \text{(neutral)} \\ \mu &= \frac{1}{2\mathrm{X} + \frac{3\mathrm{Y}}{4} + \frac{\mathrm{Z}}{2}} \quad \text{(ionized)} \end{aligned} \tag{3.13}$$

where $\bar{\mathrm{A}}_\mathrm{m}$ is the mean atomic weight of the metals in the gas. The atomic weight of particle i, $\mathrm{A_i}$, is related to its mass, $\mathrm{m_i}$, by $\mathrm{m_i} = \mathrm{A_i\,m_H}$. The second expression assumes that $\bar{\mathrm{A}}_\mathrm{m}/2 \gg 1$. Therefore, for a completely ionized gas with Solar abundance (Eq. 3.4), $\mu = 0.61$. For the particle pressure of electrons alone, $\mathrm{P_e}$, in a fully ionized gas, the number density of electrons, $n_\mathrm{e} = \rho/(\mu_\mathrm{e}\,\mathrm{m_H})$, can be used in Eq (3.11) where,

$$\mu_\mathrm{e} = \frac{1}{\mathrm{X} + \frac{\mathrm{Y}}{2} + \frac{\mathrm{Z}}{2}} \tag{3.14}$$

Example 3.3

Show that the mean molecular weight of a completely ionized pure hydrogen gas is 1/2.

In such a gas, half of the particles are electrons of mass, $\mathrm{m_e}$ and half are protons of mass, $\mathrm{m_p}$. From Eq (3.12), $\mu = \frac{1}{\mathrm{m_H}\,N}\left[\frac{N}{2}\,\mathrm{m_e} + \frac{N}{2}\mathrm{m_p}\right] \approx \frac{1}{\mathrm{m_H}\,N}\frac{N}{2}\mathrm{m_p} \approx \frac{1}{2}$.

The ideal gas law is derived under the assumption that, when energy is exchanged between particles, the interactions are elastic. Thus, if the particles were placed in a small box, the pressure against any side of the box would result from the collective motions of the particles against the wall, with no corrections required for particle–particle interactions. The more closely spaced the particles are, the more likely it will be that non-elastic interactions may occur. Therefore, a common, though not foolproof, way to decide whether such interactions are negligible is to determine the mean separation between particles, $\bar{r}$, and compare this to the particle radius[23], $R_{\rm p}$. The mean separation between particles can be found from $n\,V = N$, where n is the gas density, V its volume, and N the total number of particles, and setting the total number of particles to 1,

$$\bar{r} = \frac{1}{n^{1/3}} \tag{3.15}$$

If $\bar{r} \gg R_{\rm p}$, then the gas may be considered ideal. The caution to this rule of thumb is that other effects may be important, for example, Coulomb interactions between charged particles, or the fact that the effective volume within which a particle can move may be reduced due to finite particle size. Another example is the degenerate cores of low mass stars (Sect. 3.3.2) in which densities are so high that quantum mechanical effects become important.

The perfect gas law is an example of an *equation of state*, that is, an equation that describes how the physical properties of the gas (e.g. pressure, density and temperature) are related at any location. It is important to note that a complete description of pressure includes all contributors, including particle pressure (due to the thermal motion of particles and also free thermal electrons, if present) as well as radiation pressure (Sect. 1.5),

$$\mathrm{P} = \mathrm{P_p} + \mathrm{P_{rad}} \tag{3.16}$$

If electron degeneracy (Sect. 3.3.2) is an important contributor to the pressure, then it must be added as well and, if these other terms are significant in comparison to $\mathrm{P_p}$, then the gas is non-ideal.

3.4.3 The mean free path and collision rate

If a small test particle enters a gas containing larger field particles (the latter initially considered at rest), there will be some probability that it will collide with a field particle. The average distance that the particle travels before such an encounter occurs is called the *mean free path*, $\bar{l}$. The mean free path can be determined by considering a small volume, V, of the gas, as shown in in Figure 3.14. The probability that the test particle will encounter a field particle will depend on the projected area presented by all

[23]Comparing to the particle diameter is more accurate, though factors of two are not that important for most comparisons, as Example 3.4 illustrates.

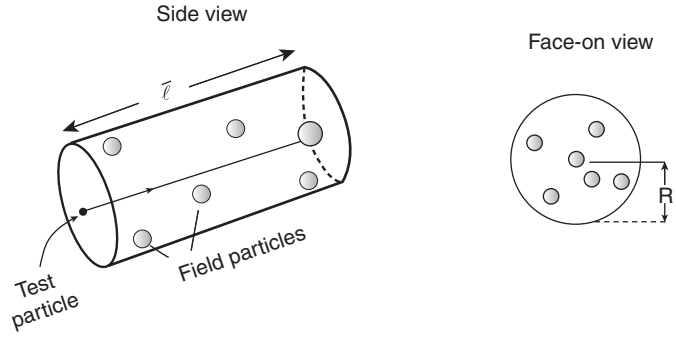


Figure 3.14. A test particle entering a gas containing field particles travels a mean free path (depth of the volume, left) before encountering a field particle. The probability of an encounter depends on the area presented by the field particles over the total area (right). If the entire cross-sectional area of the cylinder is occupied by particles, then the test particle will certainly encounter a field particle.

field particles, A_p, divided by the total area presented by the region of space, A. If we are to ensure that an interaction does occur, then the probability is 1 so,

$$1 = \frac{A_p}{A} = \frac{N\sigma}{\pi R^2} = \frac{nV\sigma}{\pi R^2} = \frac{n\pi R^2 \bar{l}\,\sigma}{\pi R^2} = n\sigma\bar{l} \tag{3.17}$$

Here N is the total number of field particles, σ is the cross-sectional area of each field particle, and the volume is $V = \pi R^2 \bar{l}$ as shown in the figure. Thus, the mean free path is given by,

$$\bar{l} = \frac{1}{n\sigma} \quad \text{cm} \tag{3.18}$$

If the test particle has a velocity, v, this distance can be expressed as $\bar{l} = v\bar{t}$, where $\bar{t}$ is mean time between collisions, thus,

$$\bar{t} = \frac{1}{nv\sigma} \quad \text{s} \tag{3.19}$$

The *collision rate* for this test particle is just the inverse of the mean time between collisions,

$$\mathcal{R} = nv\sigma \quad \text{s}^{-1} \tag{3.20}$$

For a more general case with many test particles and field particles, v is the relative velocity between the test and field particles. For identical particles in a one-temperature gas, the relevant velocity is the mean particle velocity, $\langle v \rangle$, as given in Eq (3.9). If the

test particle is of very low mass in comparison to the field particles, such as collisions between free electrons and protons in an ionized gas[24], then v is just the mean particle speed of the electrons since the protons can be thought of as 'at rest' in comparison. Also, the collision rate, $\mathcal{R}$ (Eq. 3.20) refers to the rate at which a single test particle will encounter another field particle in a gas in which the field particle density is n. If we wish to calculate the total collision rate in a gas due to collisions by all test particles, then we need to multiply $\mathcal{R}$ by the total number of test particles, $N_t = n_t V$, where n_t is the test particle density. The total collision rate per unit volume, then, is,

$$\nu_{tot} = n\, n_t\, v\, \sigma \quad \mathrm{s^{-1}\,cm^{-3}} \tag{3.21}$$

It doesn't matter which particle is considered the test particle and which is considered the field particle as long as the correct values of v and σ are used.

If all particles are neutral and can be approximated as elastically interacting spheres, then σ is just the geometrical cross-sectional area of the particles. However, the cross-sectional area may have to be modified in some circumstances, in which case it is called the *effective cross-section*, or σ_{eff}. For example, σ_{eff} will differ from the geometric cross-section if one or both sets of colliding particles are charged (such as electrons with protons or electrons with electrons) and σ may itself be a function of velocity, $\sigma(v)$. Thus, Eqs. (3.18), (3.19), (3.20), and (3.21) require values of σ_{eff} that depend on the type of colliding particles and velocities under consideration. Often, the quantity, γ_{coll}, where coll refers to the type of collision being considered, is used instead. The quantity, γ_{coll} is the mean *collision rate coefficient* which folds together the quantities, v and σ with appropriate weighting. For example, for collisions between neutral hydrogen atoms, $\gamma_{H\text{-}H} = \langle v \rangle \sigma_{eff}\,\mathrm{cm^3\,s^{-1}}$. Some sample values of γ_{coll} and σ_{eff} for collisions between various types of particles are given in Table 3.2.

The above expressions are put to use in many astronomical situations. Example 3.4 provides a sample calculation for interstellar HI. Later (Example 5.4) we will provide an example for electron – proton collisions that result in recombination to HI. A final important point is that the 'test particle' could be a photon, a point we return to in Example 3.5 and Sect. 5.

Example 3.4

For an interstellar cloud of pure HI at $T = 100\,K$ and $n = 1\,cm^{-3}$, (a) calculate the geometric cross-section, σ_g, for H–H collisions. What is the ratio of σ_{eff}/σ_g? (b) Determine the mean free path, mean time between collisions, and collision rate for an H atom. Also, what is the mean separation between particles. Is this an ideal gas?

(a) Using the Bohr radius, $a_0 = 5.29 \times 10^{-9}\,\mathrm{cm}$ (Table G.2), $\sigma_g = \pi a_0^2 = 8.80 \times 10^{-17}\,\mathrm{cm^2}$. By Eq. (3.9), and setting m to the mass of the hydrogen atom,

[24]See footnote 22 in this chapter.

Table 3.2. Sample collision parameters

Temperature (K)	HI – HI[a]	$\gamma_{\rm H-H}$ ($\rm cm^3\,s^{-1}$)
30		5.1×10^{-10}
100		7.4×10^{-10}
300		10.2×10^{-10}
1000		13.6×10^{-10}
	HI – HI de-excitation[b]	$\gamma_{21\,cm\,line}$ ($\rm cm^3\,s^{-1}$)
30		3.0×10^{-11}
100		9.5×10^{-11}
300		16×10^{-11}
1000		25×10^{-11}
	electron – proton with recomb.[c]	$\alpha_{\rm r}$ ($\rm cm^3\,s^{-1}$)
5000		4.54×10^{-13}
10 000		2.59×10^{-13}
20 000		2.52×10^{-13}
	electron – proton without recomb.[d]	$\sigma_{\rm eff}(\rm cm^2)$
10^4		1.4×10^{-15}
10^5		1.4×10^{-17}
10^6		1.4×10^{-19}
	electron – HI de-excitation[e]	$\gamma_{\rm Ly\,\alpha\,(2p\rightarrow 1s)}$ ($\rm cm^3\,s^{-1}$)
5000		6.0×10^{-9}
10 000		6.8×10^{-9}
20 000		8.4×10^{-9}

[a]Collisions between neutral H atoms. Here $\gamma_{\rm H-H} = \langle v \rangle \sigma_{\rm eff}$ (Ref. [160]).
[b]Collisions between neutral H atoms that result in the de-excitation of the hyperfine ground state (see Appendix C.4).
[c]Collisions between an electron and proton that result in recombination to HI. In this case, $\gamma_{\rm coll}$ is referred to as $\alpha_{\rm r}$, which is the sum of $\alpha_{\rm n,L}$ over all quantum levels, n and L (Appendix C). Case B recombination (see Footnote 6 in Chapter 5 or Sect. 9.4.2) is assumed.
[d]Collision *without* recombination. A collision is taken to be an interaction that significantly affects the trajectory of the electron. Taking $e^2/r_{\rm eff} = \rm m_e\, v^2$ to be the condition for a significant deflection (Ref. [160]), we find, $\sigma_{\rm eff} = \pi\, r_{\rm eff}^2 = \pi\,[e^2/(\rm m_e\, v^2)]^2$. Using Eq. (3.9), this leads to $\sigma_{\rm eff} = 1.4 \times 10^{-7}/\rm T^2$. A more accurate derivation would include the cumulative effects of other particles in the gas and would also use the Maxwellian distribution instead of Eq. (3.9). Therefore, these results are meant to be 'indicative' only.
[e]Collisions that result in the de-excitation of the Lyα line in the permitted transition (see Appendix C.2) from $2p$ to $1s$ only. This has been computed from Eq. (4–11) in Ref. [160] using data from www.astronomy.ohio-state.edu/~pradham/atomic.html. Note that the value for excitation will not be the same as for de-excitation nor will it be the same for the forbidden transition from $2s$ to $1s$.

$\langle v \rangle = 2.57 \times 10^5\,\mathrm{cm\,s^{-1}}$. Then $\langle v \rangle \sigma_g = 2.26 \times 10^{-11}\,\mathrm{cm^3\,s^{-1}}$. From Table 3.2, $\gamma_{\mathrm{H\text{-}H}} = \langle v \rangle \sigma_{\mathrm{eff}} = 7.4 \times 10^{-10}\,\mathrm{cm^3\,s^{-1}}$, so $\sigma_{\mathrm{eff}}/\sigma_g = (\langle v \rangle \sigma_{\mathrm{eff}})/(\langle v \rangle \sigma_g) = 32.7$.

(b) From part (a), $\sigma_{\mathrm{eff}} = 32.7\,\sigma_g = 2.88 \times 10^{-15}\,\mathrm{cm^2}$. Using this value (or '$\gamma_{\mathrm{H\text{-}H}}$' as required) in Eqs. (3.18), (3.19), (3.20), and (3.15), respectively, we find, $\bar{l} = 3.5 \times 10^{14}$ cm (23 AU), $\bar{t} = 1.4 \times 10^9$ s (43 yr), $\mathcal{R} = 7.4 \times 10^{-10}\,\mathrm{s^{-1}}$, and $\bar{r} = 1$ cm. The 'effective radius' of the particle is $R_{\mathrm{Peff}} = \sqrt{\sigma_{\mathrm{eff}}/\pi} = 3.0 \times 10^{-8}$ cm which is $\ll \bar{r}$ so this is an ideal gas.

3.4.4 Statistical equilibrium, thermodynamic equilibrium, and LTE

Even under conditions in which most interactions between particles in a gas are elastic and a Maxwellian velocity distribution exists, there will be a minority of interactions in which kinetic energy is not perfectly conserved and some energy is exchanged in other ways. A good example is when collisions cause the excitation or de-excitation of atoms, a process that involves the transition of electrons from one quantum state to another (see Appendix C). For an excitation, a collision must exchange an energy that corresponds to the energy difference between levels. The probability of such a collisional energy exchange will be different for each transition. Thus, in a Maxwellian velocity distribution, a single value for $v\,\sigma_{\mathrm{eff}}$, as used in Eqs. 3.19 and 3.20, cannot be used to describe the probability that a specific exciting or de-exciting interaction will occur. For transitions between states, i and j, $v\,\sigma_{\mathrm{i,j}} < v\,\sigma_{\mathrm{eff}}$ since only a fraction of all collisions will result in a specific transition. Moreover, excitation can also result from the absorption of a photon and de-excitation can occur spontaneously or via photon interaction.

In general, the populations of energy levels are determined by including all processes that both populate and de-populate any given level. These processes include de-excitations from all higher levels and excitations from all lower levels – collisional, radiative and spontaneous. In a steady state, the transition rate into any level equals the rate out, and this should be true for all levels, a situation that is called *statistical equilibrium*. Equations of statistical equilibrium are set up for each level and involve the density of the particles, the energy density of the radiation field, and coefficients describing collisional, radiative, and spontaneous transition probabilities. The coefficients may themselves be functions of other quantities such as quantum mechanical parameters or temperature. The equations are then solved to determine the populations of the levels for given conditions. This is a complex problem and is not always easily solved. There are conditions, however, for which determining the population of states can be simplified (see next section for a mathematical description). These are when the gas is in *thermodynamic equilibrium* (TE) or in *local thermodynamic equilibrium* (LTE).

If a gas is in *TE*, then the energy in the radiation field is in equilibrium with the kinetic energy of the particles. Such a situation would be set up if a gas were

isolated and confined and no radiation escaped or entered the region. Eventually the radiation and particles would achieve an equilibrium in which energy is exchanged between photons and particles with no net energy gain or loss overall. The radiation field can be characterized by a *radiation temperature*, T_R, which is, in general, a temperature that is used in an equation to calculate parameters of the radiation field such as the specific intensity or flux density (Eqs. 4.1 and 4.2 are examples). In TE, therefore, $T_R = T$ (see also, Sect. 4.1). There is no additional radiative source (which could make $T_R > T$) and no 'sink' to let photons escape (allowing $T_R < T$). The photons that are produced in the gas (represented via T_R) therefore must be tightly coupled to the random motions of the particles (represented by T). For bound states, this means that excitations and de-excitations are *collisionally dominated*, that is, collision timescales are shorter than the timescales associated with photon interactions or spontaneous de-excitations. It is this property that leads to a simplification of the equations of statistical equilibrium.

It is difficult to find an example in nature in which true TE holds since energy will leave all objects eventually (the object will cool), even if very slowly. Probably the best example is the 2.7 K cosmic microwave background radiation (Sect. 3.1). This radiation filled the Universe in the past[25] (no loss of gas or photons from the Universe is possible) and the particles and radiation were in equilibrium. Small variations in temperature are observed with *position*, however, as shown in Figure 3.2.

A more common situation is one of *LTE*, that is, the gas has the properties of TE, but only *locally*. A good example is one in which some radiation can escape or leak out of the gas as, for example, in stars. The interiors of stars have densities that are high enough for the radiation and matter to be considered 'trapped together' locally and are therefore in equilibrium at any given location, yet there is a large scale temperature gradient and a net flux of radiation outwards to the surface. Thus, the radiation temperature is close to the kinetic temperature *over some relevant size scale*. Example 3.5 provides a quantification of this condition.

There are many examples in astrophysics for which LTE (or close to LTE) conditions hold. In LTE, it is still true that the populations of states are collisionally driven, a condition that is more likely to occur when spontaneous de-excitation timescales are long, or the gas is dense and/or hot (see Eq. 3.19). From an astrophysical point of view, the condition, $T_R \approx T$ is very useful, not only because of the simplification of the statistical equilibrium equations, but also because, by observing the radiation, we can infer the kinetic temperature of the gas. We will see this more explicitly in Chapters 4, 8 and 9.

[25]The temperature at the time this occurred was much higher than 2.7 K but we see the lower temperature because of the expansion of the Universe (see Sect. 7.2 and the caption to Figure 3.2).

Example 3.5

Provide an argument showing that the interior of the Sun is in LTE.

We start by computing the mean free path of a photon in the Sun's interior, for which we assume that the gas is fully ionized and adopt a cross-section for interaction to be the Thomson scattering cross-section (see Sect. D.1.1)[26] for which $\sigma = 6.65 \times 10^{-25}\,\mathrm{cm}^2$ (Eq. D.2). The mean density of the Sun (Table G.3) is $\bar{\rho} = \mathrm{M}_\odot/(\frac{4}{3}\,\pi\,{R_\odot}^3) = 1.4\,\mathrm{g\,cm}^{-3}$. The corresponding number density is then $n = \bar{\rho}/\mu\,\mathrm{m_H} = 1.4 \times 10^{24}\,\mathrm{cm}^{-3}$, where we have used $\mu = 0.61$ for a completely ionized gas with Solar abundance (Sect. 3.4.2). Then (Eq. 3.18) $\bar{l} = 1/(n\,\sigma) \approx 1\,\mathrm{cm}$ with these values. How much does the temperature change over this distance? The central temperature of the Sun is $\mathrm{T_c} = 1.6 \times 10^7$ K and the surface temperature is $\mathrm{T}_s = 5781$ K (Table G.3) so the global temperature gradient in the Sun is $\nabla\,\mathrm{T} = \frac{(\mathrm{T_c} - \mathrm{T_s})}{R_\odot} = \frac{1.6\times10^7}{6.96\times10^{10}} = 2.3 \times 10^{-4}\,\mathrm{K\,cm}^{-1}$. Thus, the average temperature change is only 2.3×10^{-4} K over the distance that a photon would travel before being reabsorbed. Since a photon that is released will be absorbed again in a region that is essentially at the same temperature, an LTE condition is satisfied – at least to an accuracy of $\sim 10^{-4}$ K. A more accurate derivation would have to take into account the true temperature gradient corresponding to the adopted density at each interior point, as well as regions that are not fully ionized.

For stellar interiors, the fact that radiation is 'locally trapped' has implications for the transport of energy out of stars. The energy generated in the core does not travel straight to the surface to be radiated away. Rather, the photon will travel a mean free path and then either be absorbed, followed by re-emission of another photon in a random direction, or else be scattered into another random direction. Thus there are many such absorptions, re-emissions, and scatterings as the radiation makes its way from the core to the surface, a process called *diffusion*. As this diffusion proceeds, the mean photon energy also degrades from γ-ray or X-ray energies in the core to UV, optical, or IR photons at the surface, depending on the type of star.

Although the photons near the surface are not the same ones as are originally generated in the core, the process can nevertheless be thought of as a single photon taking a *random walk* from the core to the surface[27] (Figure 3.15) and a diffusion timescale can be determined. Since the mean free path of a photon will vary with radius within the star, we will now consider $\bar{l}$ to be an effective mean over all interior radii. For a random walk, it can be shown that, after taking N random steps of unit length, the straight-line distance travelled is only $\sqrt{N}$. Therefore, for a photon whose mean free path is $\bar{l}$ that takes a random walk towards the surface of a star of radius, $R_\star$, we have

[26] Absorption processes that occur in an ionized gas, such as discussed in Sect. 5.4.1, are ignored, but these will *lower* the mean free path even more than we calculate in this example.

[27] We ignore regions of the star that may be *convective*, that is, regions in which the gas itself has bulk motion up or down, carrying energy with it.

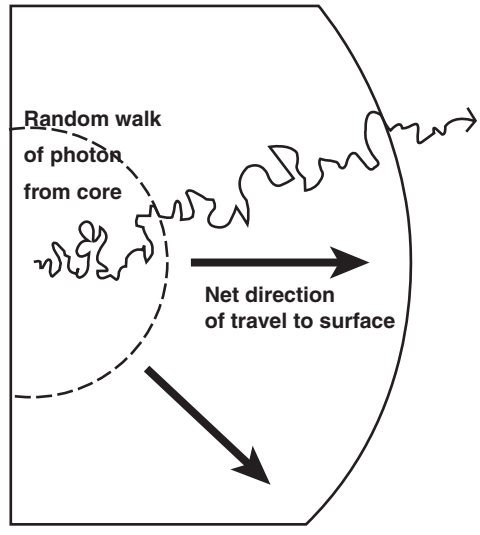


Figure 3.15. In the interiors of stars, photons travel a mean free path before interacting with a particle. The net result is a random walk to the surface.

$R_\star = \sqrt{N}\,\bar{l}$. If the interaction time for scattering or absorption is small in comparison to the time of flight of the photon during each step, then the time for a photon to diffuse from the core to the surface is the distance the photon actually travels divided by its speed (c),

$$t_{\rm d} = \frac{N\bar{l}}{c} = \left(\frac{R_\star}{\bar{l}}\right)^2 \frac{\bar{l}}{c} = \frac{{R_\star}^2}{\bar{l}\,c} \tag{3.22}$$

For the Sun, the diffusion time is of order, 10^7 yr when the duration of the interaction itself is included in the calculation[28] (Prob 3.7). This means that if there is a change in the rate at which energy is generated in the core of the Sun, it would take about 10^7 yr before we even knew about it[29].

3.4.5 Excitation and the Boltzmann Equation

We now assume LTE and return to the issue of the population of energy levels in atoms. For this condition, the equations of statistical equilibrium are much simplified and the

[28]The duration of the interaction is only important for absorption/re-emission interactions; scattering should be almost instantaneous.
[29]As judged by the solar light output. Neutrinos, however, escape freely from the core. They are difficult to detect but observatories like the one shown in Figure I.3.a could detect a change in neutrino output more quickly.

population of states is given by the *Boltzmann Equation*,

$$\frac{N_{\rm n}}{N_1} = \frac{g_{\rm n}}{g_1}\, e^{-(\frac{\Delta E}{k\,{\rm T}})} \tag{3.23}$$

Here, $N_{\rm n}$ is the number of atoms in which electrons are in an energy level, n, and N_1 is the number of atoms with electrons in the ground state, where n is the principal quantum number. ΔE is the energy difference between state, n, and the ground state and $g_{\rm n}$ is the *statistical weight* (see Appendix C.2). For hydrogen, $g_{\rm n} = 2\,{\rm n}^2$. The Boltzmann equation can also be written to compare the populations between any two states, rather than a comparison with the ground state only. If the Boltzmann Equation holds, then LTE conditions are present.

The Boltzmann equation is similar to the Maxwell–Boltzmann velocity distribution equation (Sect. 3.4.1), which should not be surprising given the discussion in the previous section. The populations of states depends on the Boltzmann factor, $e^{-\chi/(k\,{\rm T})}$, χ representing the excitation energy, as well as a quantity that indicates the number of possible states in which a particle could be found. Thus, Eq. (3.23) is like a discrete, or quantum mechanical analogue to the Maxwellian velocity distribution.

There is a difficulty with Eq. (3.23) in that as ${\rm n} \rightarrow \infty$, $g_{\rm n} \rightarrow \infty$ yet $e^{-(\frac{\Delta E}{k\,{\rm T}})} \rightarrow$ constant because, at high excitation, ΔE just becomes the ionization energy of the atom. Thus the Boltzmann equation diverges at very high n. Physically, however, before n reaches values at which the Boltzmann equation diverges, other effects become important that effectively impose an upper limit to the principal quantum number, ${\rm n}_{max}$. For example, the electrons at high n may become stripped off by the effects of neighbouring atoms. Coulomb interactions between the free charged particles and their distribution in the gas can also affect ${\rm n}_{\rm max}$.

Rather than considering how many particles are in an excited state in comparison to the ground state, a more general form of the Boltzmann equation is one which compares the number of particles in any excited state, $N_{\rm n}$, to the total number of neutral particles, $N = N_1 + N_2 + ... + N_{{\rm n}_{max}}$. From Eq. (3.23),

$$\frac{N_{\rm n}}{N} = \frac{g_{\rm n}\, e^{-(\frac{\Delta E_n}{k\,{\rm T}})}}{g_1\, e^{-(\frac{\Delta E_1}{k\,{\rm T}})} + g_2\, e^{-(\frac{\Delta E_2}{k\,{\rm T}})} + ... + g_{{\rm n}_{\rm max}}\, e^{-(\frac{\Delta E_{\rm n_{max}}}{k\,{\rm T}})}} = \frac{g_{\rm n}}{U}\, e^{-(\frac{\Delta E}{k\,{\rm T}})} \tag{3.24}$$

where $U \equiv \sum_{\rm n=1}^{\rm n_{max}} g_{\rm n}\, e^{-(\frac{\Delta E_n}{k\,{\rm T}})}$ is called the *partition function*.

At temperatures below 3500 K, the partition function for hydrogen reverts to the statistical weight of the ground state, i.e. $g_1\, e^{-(\frac{0}{k\,{\rm T}})} = g_1 = 2\,(1)^2 = 2$ since the remaining exponentials in the denominator are very small for low T. For hydrogen at higher temperatures such as in the atmospheres of stars, the partition function has been parameterized by,

$$\log(U) = 0.3013 \; - \; 0.00001\, \log(5040/{\rm T}) \tag{3.25}$$

which is valid over the temperature range, 3500–20 000 K (Ref. [68])[30]. Evaluating this equation shows that, again, $U = 2$ to within 0.1 per cent over this temperature range. Thus, for HI, Eq. (3.24) is essentially equivalent to Eq. (3.23) for temperatures up to 20 000 K. Example 3.6 shows that very high temperatures are required before there are significant numbers of particles in even the first excited state. Such high temperatures, as we shall see in the next section, will ionize the gas.

Example 3.6

Determine the temperature required for the number of particles in the first excited state of hydrogen to be 10 *per cent of the number in the ground state.*

By the Boltzmann Equation (Eq. 3.23), the equation for statistical weight (Eq. C.11), and ΔE (Eq. C.6) expressed in erg with $\mathrm{n}' = 2$ and $\mathrm{n} = 1$ we have, $\frac{N_2}{N_1} = \frac{2(2^2)}{2(1^2)} \exp\left(-\frac{-2.18\times10^{-11}(\frac{1}{2^2}-\frac{1}{1^2})}{k\,\mathrm{T}}\right)$. Setting the LHS to 0.1 and solving, we obtain $\mathrm{T} = 3.2 \times 10^4$ K. Thus, the temperature must exceed 30 000 K in order for there to be a significant number of particles in the first excited state.

When the gas density is low, such as in the ISM, LTE often no longer holds so the Boltzmann Equation, as stated in Eq. (3.23), cannot be used. The rate at which an electron will spontaneously de-excite is determined by the Einstein A coefficient which, for de-excitation from the first excited state to the ground state (Ly α) is $A_{2,1} = 6.3 \times 10^8\ \mathrm{s}^{-1}$ (Appendix C plus notes to Table C.1). The collision rate, however, is of order, $10^{-6}\, n_\mathrm{e}$, where n_e is the number density of free electrons, i.e. the *electron density* (Ref. [160]). Any transition to the first excited state, whether radiatively or collisionally induced, will therefore result in an immediate spontaneous de-excitation for the range of densities observed in interstellar space (Table 3.1). Similar arguments hold for the other excited states of hydrogen (see Table C.1), indicating that hydrogen in the ISM is not in LTE. Moreover, this implies that *neutral hydrogen in the ISM is in the ground state*.

The Boltzmann Equation (Eq. 3.23) *defines* the *excitation temperature*, $\mathrm{T_{ex}}$. i.e. $\mathrm{T_{ex}}$ is *whatever temperature, when put into the Boltzmann Equation, results in the observed ratio of* $\frac{N_\mathrm{n}}{N_1}$ (or whatever two levels are being compared). For a gas in LTE, all levels in the atom can be described by the *same* value of $\mathrm{T_{ex}}$, in which case the excitation temperature is just the gas kinetic temperature. In non-LTE cases, the Boltzmann Equation may still be used provided the excitation temperature, rather than the kinetic temperature, is inserted into Eq. (3.23). There could then be a different value of $\mathrm{T_{ex}}$ between every pair of energy levels in the atom. There are even cases in which $\mathrm{T_{ex}}$ could be a negative value if there are more particles in an excited state in comparison to a lower state. A 'negative temperature' only has meaning in the context of this

[30]Pressure effects may increase the error at the highest temperatures.

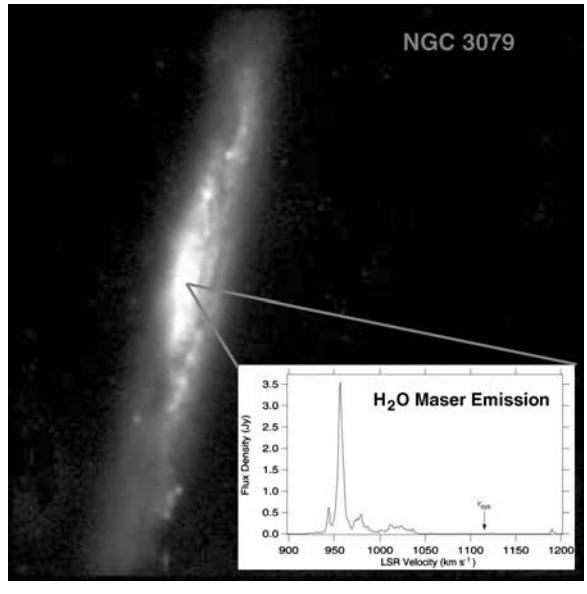


Figure 3.16. The nucleus of the edge-on spiral galaxy, NGC 3079 (Ref. [60]), harbours the most luminous H_2O maser known. The emission is at a rest frequency of 22 GHz, but the *x* axis in the inset spectrum (Ref. [172]) has been converted to a velocity via the Doppler formula (Table I.1). The excitation temperature of this line is negative since, at any time, there are more particles in the upper energy level in comparison to the lower energy level. In astronomical sources, masers are usually associated with molecules such as H_2O or OH which are near sources of pumping such as hot stars or the nuclei of galaxies.

equation. An example of such a situation is an astrophysical *maser*, as illustrated in Figure 3.16[31].

ISM HI clouds are particularly interesting because, although all particles are in the ground state, this level is split into two hyperfine states due to the two possible orientations of the proton and electron spins (see Appendix C.4). The excitation temperature of this line is therefore called the *spin temperature*, T_S. Since the spontaneous de-excitation rate of the upper hyperfine state is only $A_{1,1} = 3 \times 10^{-15}\,\mathrm{s}^{-1}$, unlike the case for Ly α, spontaneous de-excitations are very rare and, for ISM densities, this transition is *collisionally induced*. Since λ21 cm radiation depends on collisions with other neutral particles that have random motions described by the kinetic temperature, for this transition, $T_{ex} \approx T$. Thus, we can say that Eq. (3.23)

[31]Maser stands for 'microwave amplification by stimulated emission of radiation'. Masers are the microwave version of the more common 'lasers' ('l' for 'light') but the principle is the same, i.e. particles are pumped up to a higher energy level that is 'metastable', meaning that the particle can remain in the higher state for a relatively long time so that particles accumulate in the higher state. The particles then cascade downwards to a lower state emitting a very strong signal at a single wavelength (coherent emission).

holds *for this particular transition* and *the HI spin temperature is about equal to the kinetic temperature of the gas*. This is a very useful result that will allow us to compute the temperatures of interstellar HI clouds from the $\lambda 21$ cm line (see Sect. 6.4.3).

3.4.6 Ionization and the Saha Equation

The ionization state of a gas in LTE can also be expressed in a fashion similar to the Boltzmann Equation. For a gas of temperature, T, the number of particles that are in the $K+1$ state of ionization compared to the number of particles that are in the Kth state is given by the *Saha Equation*,

$$\frac{N_{\mathrm{K}+1}}{N_{\mathrm{K}}} = \frac{2\,U_{K+1}}{n_{\mathrm{e}}\,U_K}\left(\frac{2\,\pi\,\mathrm{m_e}\,k\,\mathrm{T}}{h^2}\right)^{3/2} e^{-\frac{\chi_K}{k\mathrm{T}}} \tag{3.26}$$

where $U_{\mathrm{K}+1}$, U_{K} are the partition functions of the $K+1$ and Kth ionization states, respectively, n_{e} is the electron density, $\mathrm{m_e}$ is the electron mass, χ_{K} is the energy required to remove an electron from the ground state of the Kth ionization state, and the other quantities have their usual meanings. Like the Boltzmann Equation, higher temperatures result in a higher excitation state – in this case, a higher fraction of particles that are ionized. However, unlike the Boltzmann Equation, the Saha Equation has a dependence on electron density. If there is a higher density of free electrons, then there is a greater probability that an electron will recombine with the atom, lowering the ionization state of the gas[32]. Regardless of which atom is being considered in the Saha Equation, the electron density must include contributions of free electrons from all atoms. Heavier atoms, for example, may contribute to n_{e} in larger proportion than their abundances (Figure 3.9) would suggest since they have more electrons and since their ionization energies may be lower[33].

Since hydrogen only has one electron to be removed and can only exist in the singly ionized or neutral states, Eq. (3.26) reduces to,

$$\frac{N_{\mathrm{HII}}}{N_{\mathrm{HI}}} = 2.41\times 10^{15}\,\frac{\mathrm{T}^{3/2}}{n_{\mathrm{e}}}\,e^{-\frac{1.58\times 10^5}{\mathrm{T}}} \tag{3.27}$$

where we have used $\chi_{\mathrm{HI}} = 13.6$ eV, $U_{\mathrm{HI}} = 2$ (Sect. 3.4.5), and $U_{\mathrm{HII}} = 1$ since the ionized hydrogen atom is just a free proton and only exists in a single state. It is now straightforward to show that a hydrogen atom tends to become ionized before there are many particles in even the first excited state (Example 3.7).

[32] An exception is if the densities are so high that electron degenerate pressure (Sect. 3.3.2) starts to become important in which case the ionization state remains high even though n_{e} is also high (Prob 3.8).
[33] It is generally easier to strip off an outer electron from an atom that has other electrons around it since the effective nuclear potential is weakened.

Example 3.7

At the surface of a hot star (Ref. [92]) the conditions are: $T = 26\,729$ *K and* $n_e = 9.12 \times 10^{11}\,\mathrm{cm}^{-3}$. *Assuming that LTE holds, determine the fraction of all hydrogen atoms that are ionized and compare this to the result of Example* 3.6.

Using Eq. (3.27),

$$\frac{N_{\mathrm{HII}}}{N_{\mathrm{tot}}} = \frac{N_{\mathrm{HII}}}{N_{\mathrm{HI}} + N_{\mathrm{HII}}} = \frac{\frac{N_{\mathrm{HII}}}{N_{\mathrm{HI}}}}{1 + \frac{N_{\mathrm{HII}}}{N_{\mathrm{HI}}}} = \frac{3.1 \times 10^7}{1 + 3.1 \times 10^7} \approx 1$$

Example 3.6 showed that a temperature of 32 000 K was required to place 10 per cent of all neutral hydrogen atoms in the first excited state. Yet, from this example, it is clear that at an even lower temperature, virtually all hydrogen is ionized. Thus, a hydrogen gas starts to become appreciably ionized at temperatures lower than those required to excite the neutral atom. This is because there are many more possible states available for a free electron, than for a bound electron in the first excited state.

We conclude that, if a hydrogen gas in LTE is neutral, virtually all particles will be in the ground state. Since we have already seen that neutral hydrogen that is not in LTE will also be in the ground state, our conclusion is even firmer.

3.4.7 Probing the gas

In order to probe the nature and properties of a gas, it is necessary to observe the gas in some tracer that is appropriate to its ionization and excitation state. From the previous sections, we know that hydrogen that is neutral will have its electrons in the ground state and therefore no emission lines, such as from the Lyman, Balmer, Paschen, etc. series (Appendix C) would be seen since few electrons are in excited states. Thus, to obtain information on HI such as, for example, the shell shown in Figure 3.17, we require some probe of the ground state itself. This is provided by the λ 21 cm line that is introduced in Appendix C and will be discussed further in Sect. 6.4.3. Another possibility is if a background signal excites the electron from the ground state upwards. The upwards pumping will produce absorption lines since the background photon that would normally pass through the gas is removed from the line of sight (Sect. 6.4.2).

Ionized hydrogen exists wherever there is a source of radiation or collisions energetic enough to eject the electron from the H atom. Examples are HII regions like the ones shown in Figures 3.13 and 8.4 in which an embedded hot star (or stars) provides the radiation field, or planetary nebulae, as in Figure 3.7, in which the radiation field is supplied by the central white dwarf. There are several ways to probe ionized gas. One is to observe the emission that occurs when the free electrons in these nebulae

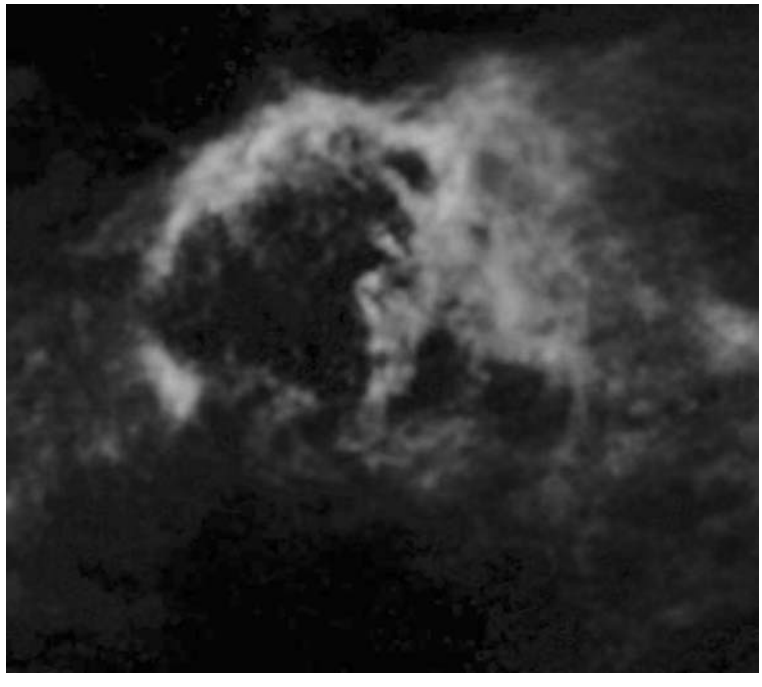


Figure 3.17. An expanding 1.8×10^5 $M_\odot$ HI shell in the low density outer region of the Milky Way galaxy (distance from the Galactic Centre = 23.6 kpc), observed in the λ 21 cm line of neutral hydrogen. This shell is believed to represent the ISM HI that has been pushed outwards after a supernova exploded 4.3 Myr ago and after the emission from the supernova itself (e.g. similar to that shown in Figure 1.2) has faded from view. The radius of the shell is 180 pc, the expansion velocity is 11.8 km s^{-1} and the temperature is 230 K. (Ref. [162]). Composed for the Canadian Galactic Plane Survey by Jayanne English with the support of A. R. Taylor. (Reproduced by permission of J. English)

accelerate near positively charged nuclei (see Sect. 8.2). Another results from electrons recombining with nuclei. Electrons always have some probability of recombining with positively charged nuclei, so continuous ionizations and recombinations occur in a steady state in an ionized gas. This means that emission lines of hydrogen such as the Lyman and Balmer series lines can be observed as electrons cascade down various energy levels, emitting photons in the process. Even though the gas is highly ionized, these transitions between bound states in atoms (meaning the atom is momentarily neutral) can be observed. Such emission lines are called *recombination lines* (see also Figure C.1 and Sect. 9.4). Wherever such hydrogen emission lines are seen, the gas from which these lines originate *must* be ionized.

Gases that are partially ionized are also present in a number of astrophysical conditions such as stellar atmospheres or interiors, or diffuse interstellar regions farther from sources of ionizing radiation than the immediate HII region. Such gases may also be detected by their emission lines, depending on the degree of ionization and the number of particles that are present. For stellar atmospheres, however, an important

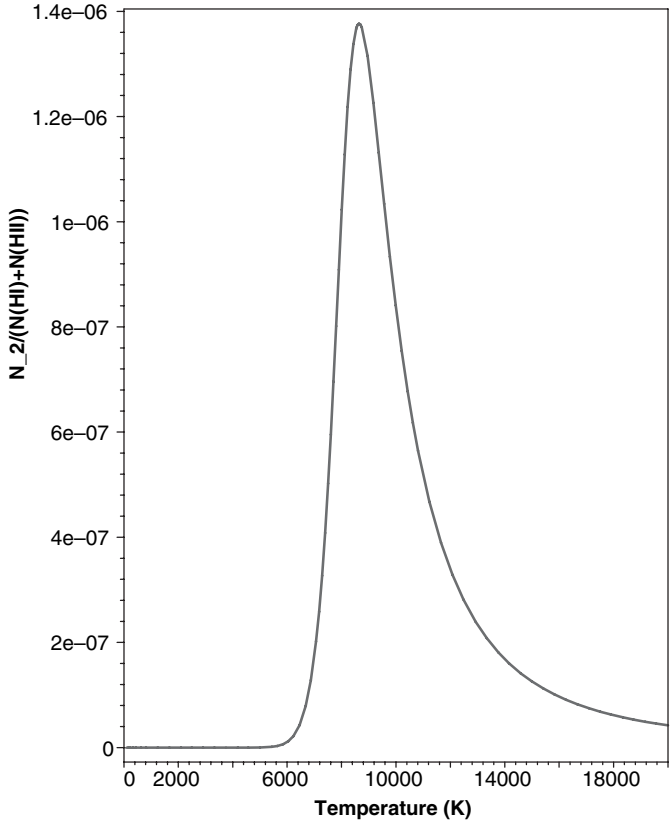


Figure 3.18. Fraction of hydrogen atoms that are in the first excited state, N_2, in comparison to the total number of hydrogen atoms, $N_{HI} + N_{HII}$. The adopted electron density at all temperatures has been taken to be constant at $n_e = 10^{13}$ cm^{-3} (in reality, this will also vary with temperature). This distribution indicates how the strength of the Balmer absorption lines varies with stellar surface temperature. Note that the fraction is still quite low, even at the peak.

probe of the conditions is provided by absorption lines. Absorption lines are formed when the temperature of the background stellar 'surface'[34] is hotter than the partially transparent stellar atmosphere that surrounds it. The background radiation excites the atmospheric atoms and absorption lines are seen, as described above (see also Sect. 6.4.2). Historically, the most important hydrogen absorption lines are those of the Balmer series since they occur in the optical part of the spectrum. However, observing a Balmer absorption line requires that there be a sufficient number of hydrogen atoms in which electrons are in the energy level, n = 2. As we know, few particles are in that state, but a small fraction can still produce an observable line. By considering the Boltzmann and Saha Equations together, it can be shown that the fraction, $N_2/N_{\rm tot}$, reaches a maximum at a temperature around 10^4 K (see Figure 3.18). At lower

[34]The surface occurs at the radius at which the star becomes opaque. See comments in Sect. 6.4.2 and Footnote 12 in Chapter 6. For Solar values, see Table G.11.

temperatures, there are insufficient numbers of particles with electrons in $n = 2$, and at higher temperatures, most of the particles have been ionized. Thus, the strength of the Balmer absorption lines seen in stars reaches a maximum around 10^4 K. The strength of these and other absorption lines in stellar atmospheres forms the basis of the *stellar classification system* in use today. The properties of the various stellar spectral types, which are ordered by temperature, are listed in Tables G.7, G.8, and G.9. Sample spectra, showing how various absorption lines change with spectral type, are illustrated in Figure 6.8. Since stars are at 10^4 K for spectral type, A0V, this is the type for which the Balmer lines are strongest.

3.5 The dusty Universe

Of order 2×10^6 kg of meteoritic dust (also called micrometeorites) in the mass range, 10^{-5} to 10^2 μg descend upon the Earth each year (Ref. [104]). Analysis of these particles and larger meteorites as well as data resulting from space exploration using probes, imagers, and landers, have yielded a wealth of information about the particulate matter in the Solar System. For the first time, we have now collected dust from a comet and returned the particles to the Earth for study (Figure 3.19). Yet, it is not clear to what extent local material resembles the interstellar dust that is scattered throughout the Galaxy. This means that our knowledge of interstellar medium (ISM) dust properties must still be deciphered from the light that is emitted, absorbed, scattered and polarized by this important component of the ISM. In this section, we will look at some of the observational effects of dust and discuss dust properties. Details of the interaction of light with dust will be dealt with in Sect. 5.5.

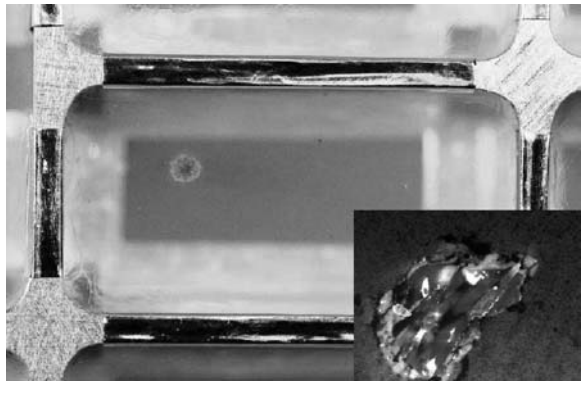


Figure 3.19. The impact of a cometary dust grain onto aerogel (a sponge-like silicon-based solid that is 99.8 per cent air) is shown here, captured when NASA's Stardust spacecraft flew through the dust and gas cloud surrounding Comet Wild 2 on January 2, 2004. Stardust returned to Earth on January 15, 2006. *Inset:* A tiny particle collected by the Stardust spacecraft. This particle is 2 μm across and is made of a silicate mineral called forsterite, known as peridot in its gem form. (Reproduced by permission of NASA/JPL-Caltech and the Stardust Outreach Team)

3.5.1 Observational effects of dust

Almost everywhere we look in the sky, we see some evidence for dust. Dark bands that cut a swath across galaxies (Figure 3.22), thin curves that follow spiral arms (Figure 3.8), filamentary patterns in HII regions (Figure 3.13), patchy dimming of starlight in the Milky Way (Figure 3.20), diffuse glows from scattered light around individual stars (Figure 5.3), and stellar disks hinting of hidden planetary bodies (Figure 4.13) all attest to the presence of small particulate (solid) matter in a wide variety of environments.

Although dust makes up only $\approx$ 1 per cent of the ISM, by mass, its effects are extremely important, especially in the optical and UV where dust strongly scatters and absorbs, obscuring the light from behind. Dust prevents us from seeing the nucleus of our Milky Way galaxy (Figure 3.3) at optical wavelengths and we must rely, instead, on other wavebands to penetrate through this barrier. Indeed, the depth of our view along the plane of the Milky way is quite shallow, being restricted to about a kpc, on average, though there are some lines of sight that are clearer. What is worse, the distribution of dust is so patchy that it cannot be easily modelled or corrected for, except on an object-by-object or location-by-location basis. In the history of our understanding of the Milky Way, dust has been a major impediment to obtaining distances and sizes of various objects, including the large-scale structure of the Galaxy itself. For example, if the absolute magnitude, M, is believed to be known for some object, say via a calibration of other similar objects of known distance, then a measurement of the

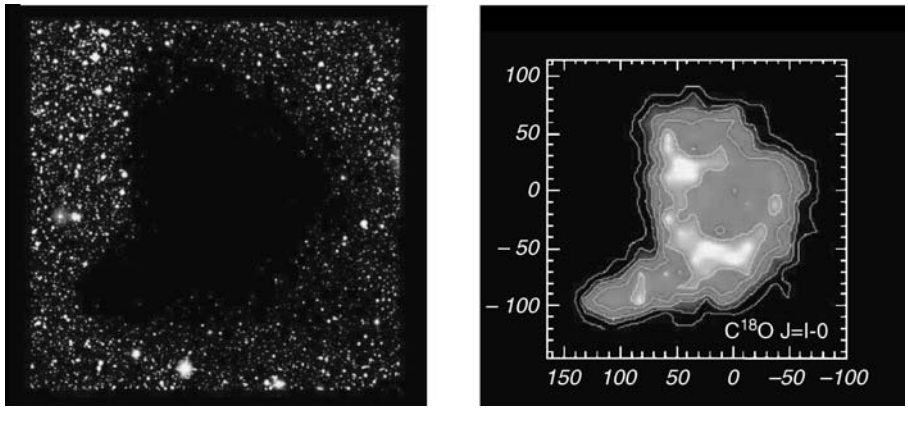


Figure 3.20. Left: At one time, dark regions like this in the sky were thought to be 'holes' – regions in which there were no stars. We now realize that these dark nebulae are dense molecular clouds. The dust makes up only about 1 per cent of the mass of the cloud yet it is the dust that obscures the background starlight from view. This cloud, called Barnard 68, is quite nearby, about 150 pc away and 0.15 pc across. Right: The same cloud is shown by the emission of a spectral line of the carbon monoxide isotope, $C^{18}O$. Most of the mass of the cloud is in molecular hydrogen, H_2, but various other molecules and their isotopes are present, such as CS, N_2H^+, NH_3, H_2CO, C_3H_2, among others. (Lada, C. U., Bergin, E. A., Alves, J. F., and Huard, T. L., 2003, A_pJ, **586**, **286.** Reproduced by permission of the AAS and J. Alves)

apparent magnitude, m, leads to a measurement of distance via Eq. (1.30). If there is dust along a line of sight, however, the object will look dimmer than it otherwise would be, leading to a distance that is erroneously large. Early estimates of the size of the Milky Way were high by at least a factor of 2 for just this reason. In the presence of dust obscuration, therefore, Eq. (1.30) must be re-written to take this into account,

$$\mathrm{V} - M_\mathrm{V} - A_\mathrm{V} = 5 \log d - 5 \tag{3.28}$$

The quantity, A_V, is the *total extinction* or just *extinction*[35] in the V band which gives the amount of dimming, in magnitudes, as a result of scattering and absorption along the line of sight, that is (Eq. 1.26),

$$A_\mathrm{V} = -2.5 \log\left(\frac{f_\mathrm{V}}{f_{\mathrm{V}0}}\right) \tag{3.29}$$

where f_V is the flux density of the object with extinction and $f_{\mathrm{V}0}$ is the flux density in the absence of extinction. Similar equations could be written for the other filter bands. Since the V band extinction per unit distance along the line of sight in the plane is about 1 mag kpc^{-1} (Ref. [180]) large errors in distance will occur if the extinction is ignored. For example, an object at a true distance of 1 kpc will be placed at a distance of 1.6 kpc if dust extinction is not included (Prob. 3.11).

Reddening was introduced in Sect. 2.3.5 in the context of the Earth's atmosphere. Outside of the Earth's atmosphere, the main cause of reddening at optical wavelengths is dust. Reddening is therefore seen in the interstellar medium or in any dusty astrophysical environment. It is due to the fact that dust extinction is more effective at short wavelengths than at long wavelengths, so any object viewed at optical wavelengths will appear redder if seen through a veil of dust. Given this wavelength dependence of extinction, we can define a *selective extinction*, which is a measure of reddening at some wavelength, λ,

$$E_{\lambda-\mathrm{V}} = A_\lambda - A_\mathrm{V} \tag{3.30}$$

It is possible to measure the reddening of an astronomical object via a comparison with another object (a calibrator) that has negligible reddening. For example, if two stars of the same intrinsic type, and therefore the same absolute magnitude, are observed in the V and B bands, the apparent magnitudes of the star of interest and the calibrator, respectively, will be,

$$\begin{aligned} \mathrm{V} &= M_\mathrm{V} + 5 \log(d) + A_\mathrm{V} \quad \text{(star)} \\ \mathrm{B} &= M_\mathrm{B} + 5 \log(d) + A_\mathrm{B} \quad \text{(star)} \end{aligned} \tag{3.31}$$

$$\begin{aligned} \mathrm{V}_0 &= M_\mathrm{V} + 5 \log(d_0) \quad \text{(calibrator)} \\ \mathrm{B}_0 &= M_\mathrm{B} + 5 \log(d_0) \quad \text{(calibrator)} \end{aligned} \tag{3.32}$$

[35] 'Extinction' is also used as a general term that refers to the combined effects of scattering and absorption (see Chapter 5).

where d and d_0 are the distances to the star of interest and the calibrator, respectively. Combining and rearranging these equations, and expressing $(B_0 - V_0)$ as $(B - V)_0$, yields,

$$E_{B-V} = A_B - A_V = (B - V) - (B - V)_0 \tag{3.33}$$

This is an important result because it does not depend on knowing distances which are often difficult to obtain. One need only compare the quantities, $B - V$, for the star of interest and the calibrator star. Differences of magnitudes are also more accurately determined than the magnitudes themselves, in the event that there is some error in the absolute scale being used. It is straightforward, then, to determine $E_{\lambda-V}$ for any λ, provided unreddened (or minimally reddened) calibrators can be found. Note that E_{B-V} is almost always a positive quantity since blue light is more heavily extincted (will have a higher numerical value of the magnitude) than visual light. Thus, by Eq. (3.33), we have a way of determining the selective extinction of an object.

It is less straightforward to determine the total extinction, A_V. However, much effort has gone into determining this value for objects of known distance. As a result, it has been found that the ratio of total to selective extinction is very nearly constant in the Milky Way. This is basically a statement that the same dust that is producing the total extinction is also producing the reddening. The two values are related via (Ref. [180]),

$$R_V = \frac{A_V}{E_{B-V}} = 3.05 \pm 0.15 \tag{3.34}$$

In the absence of other information, if the selective extinction is measured, this relation may be used to find the total extinction which can then be used in Eq. (3.28), as required (Example 3.8).

Example 3.8

An M0III star is measured to have blue and visual magnitudes of $B = 8.76$ and $V = 7.09$. Estimate the distance to this star.

From Table G.8, $(B - V)_0 = 1.57$ for a M0III star. Therefore, the selective extinction of this star is (Eq. 3.33) $E_{B-V} = (B - V) - (B - V)_0 = (8.76 - 7.09) - 1.57 = 1.67 - 1.57 = 0.10$. From Eq. (3.34), $A_V = 0.305$. Using $M_V = -0.2$ for this kind of star (Table G.8), Eq. (3.28) yields a distance of 249 pc.

Historically, since most observational attention has been paid to the optical waveband, the optical extinction properties of dust have dominated the attention of astronomers for the purposes of distance or flux corrections. In other words, the dust has been considered 'in the way'. However, for dust, it is certainly true that 'one man's trash is

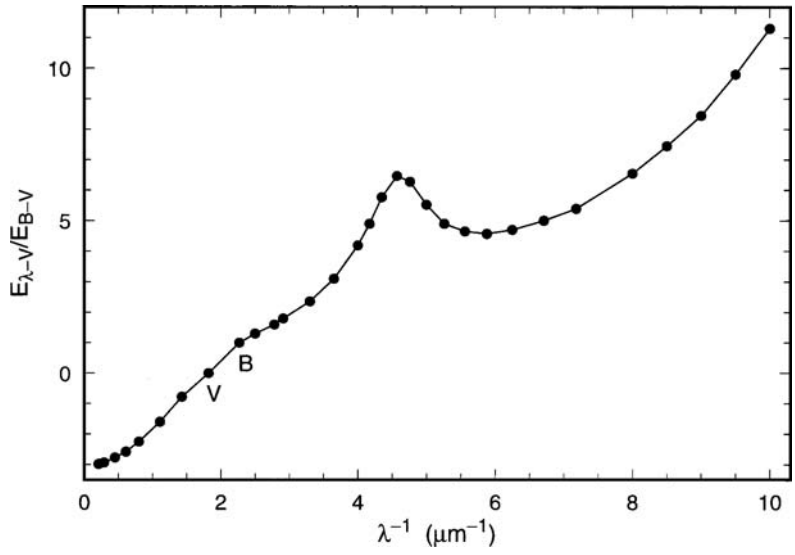


Figure 3.21. The mean extinction curve for the ISM in the optical band from $0.2 \rightarrow 10\,\mu m^{-1}$ ($\lambda 5\,000 \rightarrow \lambda 100$ nm). Frequency increases to the right and the V and B bands are marked. Note that larger $\frac{E_{\lambda-V}}{E_{B-V}}$ corresponds to more extinction, so background interstellar light will be more heavily extinguished at these wavelengths. (Whittet, D.C.B., *Dust in the Galactic Environment*, 2nd Ed., Institute of Physics Publishing)

another man's treasure'. The same dust that blocks our optical view of the cosmos is a crucial component in the complex chemistry of the ISM, it traces magnetic fields, plays an important role in ISM dynamics, must be taken into account in abundance determinations, and contributes to driving stellar evolution. Thus the study of dust is a rich and important sub-field of astronomy in its own right.

Observational data that allow us to probe the characteristics of dust are therefore very important. An example is Figure 3.21 which shows the mean *extinction curve* for the Milky Way in the optical band. The extinction curve is a plot of the selective extinction, $E_{\lambda-V}$ (normalized by E_{B-V}) as a function of λ^{-1}. Since $E_{\lambda-V} = A_\lambda - A_V$, this plot shows how the extinction varies with wavelength. In the optical region, a rough approximation is that $E_{\lambda-V} \propto 1/\lambda$.

Structure in the extinction curve provides clues as to the make-up of interstellar dust. For example, there are several changes in slope and an obvious peak (excess extinction) in the ultraviolet at $4.6\,\mu m^{-1}$ ($\lambda 2175$Å). Such spectral features can be compared to laboratory spectra of vibrational modes within solids in an attempt to determine the type of material responsible. For the $\lambda 2175$Å peak, the best candidate is graphite (Figure 3.23) or a related particle (Ref. [180]). See Example 5.8 for further details on the link between the extinction curve and the optical properties of grains.

In reality, because a complete description of dust requires many parameters, it is difficult to find a unique result using only the extinction curve. This is compounded by the fact that there may be different types and admixtures of grains in different

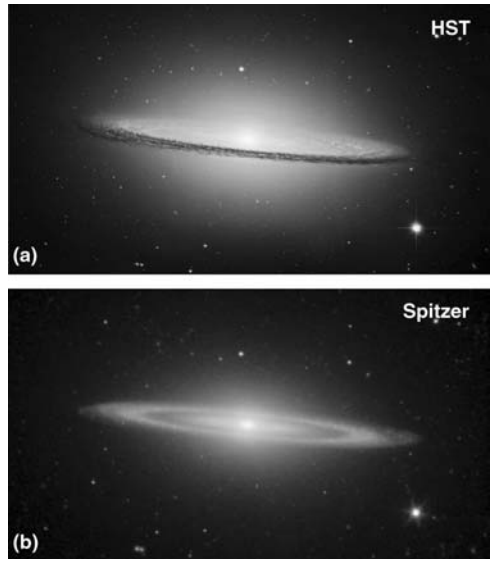


Figure 3.22. The Sombrero Galaxy (M 104) in optical and infrared light. (**a**) A Hubble Space Telescope image reconstructed from a series of images taken in red, green and blue filters. The dust lane obscures the light from the stars behind it. (Credit: Hubble Space Telescope/Hubble Heritage Team) (**b**) False colour image showing $\lambda 3.6\,\mu$m data in blue, $\lambda 4.5\,\mu$m data in green, $\lambda 5.8\,\mu$m data in orange, and $\lambda 8.0\,\mu$m data in red. In this image, blue/green represents starlight primarily from stars cooler than the Sun, while red/orange represents thermal emission from dust at its equilibrium temperature. Now the dust lane is emitting, rather than obscuring, and starlight can be seen through it. [Reproduced by permission of NASA/JPL-Caltech/R. Kennicutt (University of Arizona), and the SINGS Team] (See colour plate)

interstellar environments. A good example is our own Solar System in which solid particles (and planets) become icier farther from the Sun. Fortunately, there are other ways of obtaining additional information about interstellar dust. These include the study of dust scattering and polarization characteristics (Sect. 5.5). In addition, a rich area of study is now the *emission* properties of grains viewed in the infrared part of the spectrum. The same dust that causes extinction in the optical band will radiate in the IR (see Sect. 4.2.1). This is beautifully illustrated by Figure 3.22. With the advent of infrared telescopes such as the Infrared Astronomical Satellite (IRAS) launched in the 1980s, the Spitzer Space Telescope, launched in 2003, and the most powerful of these, the Herschel Telescope, the study of dust has become a growing and maturing sub-discipline in astronomy. Put together, we have an emerging view of dust characteristics, as described in the next sections.

3.5.2 Structure and composition of dust

Large grains, also called *classical grains*, are of order, 0.1 μm in size which is about the size of the particles in smoke. *Very small grains* (VSGs) are of order, 0.01 μm

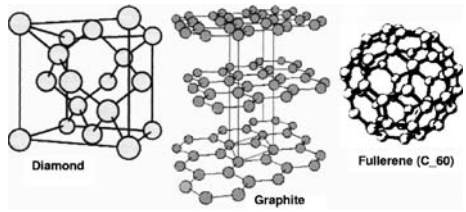


Figure 3.23. Sketch of the three different ordered forms in which carbon can exist in its solid phase. (Whittet, D. C. B., *Dust in the Galactic Environment, 2nd Ed.*, Institute of Physics Publishing)

which is about the size of a small virus. However, a continuum of particle sizes is likely present in a power law distribution such that there are more smaller grains and fewer large grains, i.e. $n(a) \propto a^{-q}$, where $n(a)$ is the number density per unit size range ($\mathrm{cm}^{-3}\,\mathrm{cm}^{-1}$) of grains of radius, a, and q is the power law index that must be fitted to the data. Ignoring some complexities, a function that fits the data reasonably well over the range, $a = 0.005\,\mu\mathrm{m}$ to $a = 0.25\,\mu\mathrm{m}$ is (Ref. [180]),

$$n(a) \propto a^{-3.5} \tag{3.35}$$

Grain composition is more difficult to ascertain since few spectral features are observed from these solid particles. Carbon appears to play an important role and, given its propensity to join together with other carbon atoms, this element can combine in a wide variety of ways, including highly ordered (*crystalline*) structures such as nanodiamonds, graphite, and fullerenes (see Figure 3.23) as well as *amorphous* (irregular) structures. Carbon can also easily form hydrogenated rings, called *polycyclic aromatic hydrocarbons* (PAHs, Figure 3.24), commonly seen on Earth in the soot from automobiles. Some version of PAHs, though possibly not exactly in the same formations, may be in the interstellar medium, as suggested by observed mid-infrared spectral features. PAHs are planar, rather than three dimensional, and so are considered to be large molecules, containing of order 50 atoms, rather than the three-dimensional structures that we identify with solid dust grains. PAHs are therefore at the transition between molecules and dust. Another component of dust is likely *silicates*, a class covering a wide variety of chemical compositions involving SiO_4. These appear to be in amorphous, rather than crystalline, structures and also likely contain magnesium (Mg) and iron (Fe). A third category is *ices*, such as the solid phases of water (H_2O), methane (CH_4) and ammonia (NH_3).

The structure of the grains is less certain. 'Sooty' (carbon rich) mantles over silicate cores is one possibility. Ices may also form mantles about cores since they move more easily between the gas phase and the solid phase. As Figure 3.23 illustrates, dust grains are unlikely to be uniform spheres. It is clear that at least some fraction of grains are elongated, since this explains the fact that scattering from dust results in polarized light, as shown in Figure D.6.

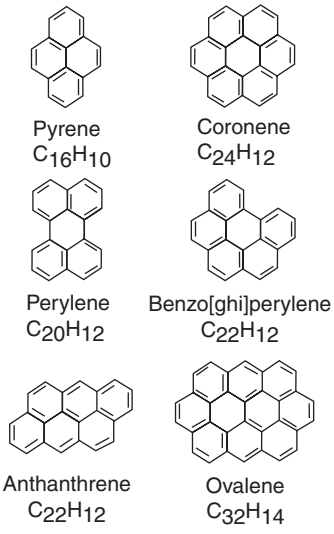


Figure 3.24. Sample configurations of polycyclic aromatic hydrocarbons (PAHs) which may make up the largest molecules in the ISM. (Reproduced using data from http://pubchem.ncbi.nlm.nih.gov/)

3.5.3 The origin of dust

Where do these grains come from? It is well known that the envelopes of stars contain dust. The cool atmospheres of evolved stars appear to be a source of dust since their atmospheres are rich in carbon and oxygen (Sect. 3.3.2) and dust can easily be driven from the extended envelopes of these stars into the ISM via radiation pressure (Sect. 1.5). However, it appears that 'stardust', although important, cannot be the sole source of interstellar dust, since more dust is present in the ISM than can be accounted for by stellar injection alone (Ref. [180]). An additional source of dust now appears to be the metal-rich ejecta of supernovae. For example, the supernova remnant, Cas A, shown in Figure 1.2, contains 2–4 $M_{\odot}$ of dust, which is about 100 times the amount of dust that could have been swept up from the ISM during its expansion. This implies that dust was formed during the supernova event, itself (Ref. [54]).

Observationally, we know that the dust distribution in the Galaxy follows the dense, molecular gas distribution (Figure 3.20). There appear to be complex processes in these regions involving gas-phase chemical reactions, the absorption of particles onto grains, migration and reactions on grain surfaces, destruction of grains in shock waves[36], and/or growth of grains via coagulation or deposition onto mantles. These processes are not yet fully understood, but the careful study of light associated with dust-rich regions is slowly revealing the properties and nature of this important component of our Universe. The intimate link between dust grains and large molecules, especially organic molecules and carbon-based molecules such as PAHs, and related information such as

[36]See Footnote 16 in this chapter.

the presence of amino acids in meteorites, hint that this branch of astrophysics may provide important clues about the origin of life itself.

3.6 Cosmic rays

The existence of *cosmic rays* (CRs) has been known since 1912 when Victor Hess took an electrometer (a device for measuring the presence of charge) to a high altitude in a hot air balloon. Contrary to what was expected, the electrometer discharged more quickly at *higher* altitudes, leading Hess to conclude that the source of the discharge must be from above the atmosphere rather than the Earth itself. For this discovery, he shared the 1936 Nobel Prize for physics. Called cosmic *rays* because they were originally thought to be part of the electromagnetic spectrum, we know today that these are high energy *particles* coming from space. As noted in the Introduction, CRs, together with meteorites and meteoritic dust (see last section) represent the only known *particulate* matter reaching the Earth's surface. As such, they are important carriers of cosmic information beyond that brought to us by light.

Figure I.1.a shows an early photograph of a cloud chamber containing CR tracks. One need only watch a real cloud chamber for a few seconds to see this subatomic world come to life. What is being viewed, however, are not the primary cosmic rays that impinge upon the Earth's atmosphere from above, but rather secondary particles that are created within the atmosphere from *cosmic ray showers*. These occur when a primary cosmic ray collides with an atmospheric nucleus, shattering it and producing secondary particles, some of which will decay and some of which may collide with other atmospheric particles, creating a chain of subsequent events. Various subatomic particles are formed in the process, such as pions, muons, and neutrinos. Secondary cosmic rays are responsible for approximately one third of the naturally occurring radioactivity on Earth.

Since we cannot detect primary particles from our location on the Earth's surface, to understand the origin of cosmic rays, it is necessary either to make corrections for atmospheric interactions or to make measurements from space. In the following sections, we assume that such corrections have been made and discuss those cosmic rays that impinge on the Earth's atmosphere.

3.6.1 Cosmic ray composition

What is the composition of cosmic rays? Approximately 98 per cent of these particles are nucleons (protons and neutrons in nuclei without their electrons, see Sect. 3.3.3) and the remaining 2 per cent are electrons and their positive counterparts, positrons. Of the particles in the energy range, $10^8 - 10^{10}$ eV, 87 per cent are hydrogen nuclei (protons), 12 per cent are helium nuclei (also called *alpha particles*), and 1 percent are heavier nuclei (Ref. [147]). Compared to the elemental abundances of the Solar System and interstellar medium shown in Figure 3.9, CRs are slightly underabundant in

hydrogen (we expect just over 90 per cent, Sect. 3.3) and overabundant, by several orders of magnitude, in the light elements, Li, Be, and B. Moreover, CRs are highly underabundant in electrons since, for a normal plasma, there should be approximately equal numbers of protons and electrons. Thus, the admixture of CR particles is quite different from what we normally see in the Solar System or ISM.

Most compositional anomalies can be explained by models that take into account the interactions of CRs en route to Earth. For example, the over-abundance of light elements can be understood in terms of *spallation* – the breaking up of heavier nuclei into smaller particles via collisions (see also Sect. 3.3.4). It is clear that the ISM acts like another 'atmosphere' which, as before, must be corrected for in order to understand the original composition of these particles at the location where they are first accelerated[37]. Few abundance anomalies remain once the 'ISM atmosphere' is taken into account.

Some departures from Solar abundance do remain, however, and these can provide clues as to the origin and acceleration mechanism of the particles (see also Sect. 3.6.3 below). One is the relative *underabundance of hydrogen*. Supernova ejecta, for example, consist of mostly heavier particles (Sect. 3.3.3) in comparison to hydrogen. Also, acceleration mechanisms may be more effective for interstellar grains (which contain heavy elements) than for particles that are in the gas phase such as hydrogen (Ref. [57]). Another is the element, ^{22}Ne which is in *excess* compared to the Solar value. This element is known to be rich in the spectra of *Wolf–Rayet stars* (very massive, young hot stars, see also Sect. 5.1.2.2 and Figure 5.10), hinting at a possible connection between CRs and massive stars that are rich in heavy metals.

The strong underabundance of electrons compared to nucleons can be understood in terms of the relative energies and momenta of the particles, a subject to which we will return in the next section.

3.6.2 The cosmic ray energy spectrum

Peculiar to CRs are relativistic speeds and consequent high energies. The total energy of a particle whose rest mass is m_0, moving at a speed, v, is,

$$E = m_0 c^2 + T = \gamma m_0 c^2 \tag{3.36}$$

where T is the kinetic energy, $m_0 c^2$ is the rest mass energy, and γ (from Table I.1) is called the *Lorentz factor*, defined as,

$$\gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.37}$$

[37]Because of this, 'primary cosmic ray' is sometimes used to represent the CRs near the acceleration site rather than those that hit the Earth's atmosphere.

If $v \ll c$, a binomial expansion (Eq. A.2) can be used on Eq. (3.37) to find,

$$\gamma \approx 1 + \frac{1}{2}\frac{v^2}{c^2} \tag{3.38}$$

In this case, Eq. (3.36) reduces to,

$$E = \mathrm{m}_0 c^2 + \frac{1}{2}\mathrm{m}_0 v^2 \ (v \ll c) \tag{3.39}$$

which is a more familiar expression showing both the rest mass energy of the particle (first term) and its kinetic energy (second term). The electron and proton rest mass energies are 0.51 MeV and 938 MeV, respectively, corresponding to $\gamma = 1$. Electrons with energies greater than about 0.5 MeV or protons with energies greater than about 1 GeV will therefore be dominated by their kinetic energies. If we consider a 'relativistic speed' to be at least $0.3\,c$, then any particle with $\gamma > 1.05$ is relativistic. We will see, below, that Lorentz factors as high as $\gamma \approx 10^{11}$ have now been measured for some high energy CRs, corresponding to $v = c$ to many decimal places.

The cosmic ray *energy spectrum* is shown in Figure 3.25 (over this range, $E \approx T$), showing a beautiful power law spectrum, with several changes in slope, over 13 orders of magnitude in particle energy! Since this is a log–log plot, a straight line represents a *power law* distribution of energies for these particles which can be described by,

$$J(E) = K E^{-}\Gamma \tag{3.40}$$

where Γ is the *cosmic ray energy spectral index*, E is the energy of a particle and K is a constant of proportionality. The quantity, $J(E)$, in units of particles $\mathrm{s}^{-1}\,\mathrm{m}^{-2}\,\mathrm{GeV}^{-1}\,\mathrm{sr}^{-1}$, is a kind of 'specific intensity[38] for particles' (compare this to a specific intensity for radiation which has units of erg $\mathrm{s}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{Hz}^{-1}\,\mathrm{sr}^{-1}$) and can be integrated to obtain the total number of CRs hitting an object per unit time (Example 3.9). A single value of Γ, as shown by the dashed curve in Figure 3.25, does not fit the data perfectly. There are changes in the slope, two of which are marked 'the knee' and 'the ankle' at energies of 4×10^{15} eV and 5×10^{18} eV, respectively, the value of the ankle being less certain. Except for the very lowest and very highest energies, $\Gamma \approx 2.7$ is a good fit to the data at energies lower than the knee and $\Gamma \approx 3.0$ fits well above it. A value of K for the upper end of the spectrum is provided in Eq. (3.41). See Prob. 3.13 for a value of K at energies below this. Changes in the slope have been scrutinized in some detail because it is thought that they may represent changes in the origin and/or acceleration mechanism of the particles.

[38]Usually 'particles' is left out of the units.

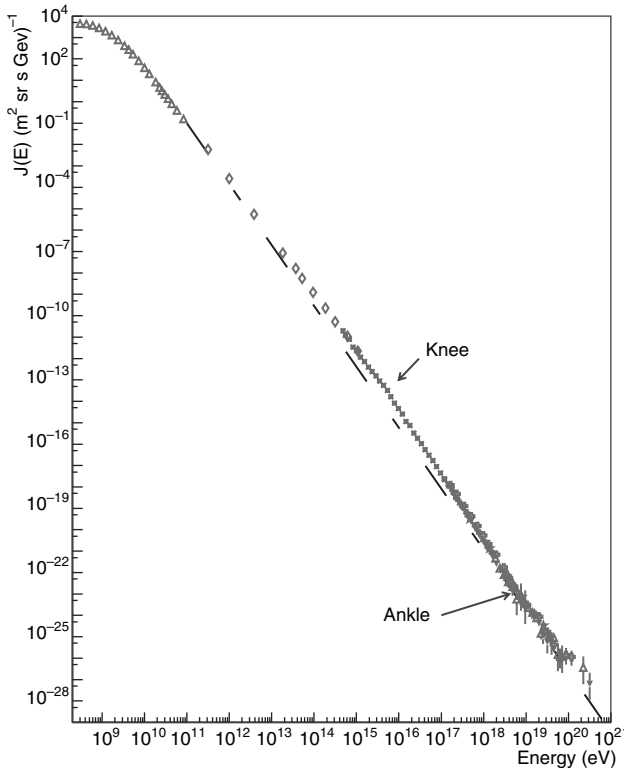


Figure 3.25. Cosmic ray spectrum for particles from 10^8 to 10^{21} eV (Ref. [20]). The ordinate represents $J(E)$ (Eq. 3.40) and the abscissa represents the kinetic energy per particle (Eq. 3.36) for which $E \approx T$ over almost the entire energy range plotted). The dashed line shows a power law fit with a single spectral index. Subtle changes in the spectral slope can be seen in two places: the 'knee' (at $E = 4 \times 10^{15}$ eV) and, to a lesser extent, the 'ankle' ($E = 5 \times 10^{18}$ eV). (Reproduced by permission of Simon Swordy)

Example 3.9

Between the energies of 10^{16} and 10^{19} eV (above the knee), the spectrum of Figure 3.25 *can be fitted by the function,*

$$J(E) = 2.1 \times 10^7 \, E^{-3.08} \ \mathrm{s^{-1} m^{-2} sr^{-1} GeV^{-1}} \tag{3.41}$$

where E is in units of GeV (Ref. [101]).What is the total rate of cosmic rays hitting the Earth's atmosphere in this energy range?
We first integrate over the energy range, $E_1 = 10^7$ GeV to $E_2 = 10^{10}$ GeV,

$$\int_{E_1}^{E_2} J(E)\, dE = \frac{2.1 \times 10^7}{-2.08} \left({E_2}^{-2.08} - {E_1}^{-2.08} \right) = 2.78 \times 10^{-8} \, \mathrm{s^{-1}\, m^{-2}\, sr^{-1}} \tag{3.42}$$

Assuming that CRs are received from all forward angles in the direction of space but not from rearward angles in the direction of the Earth, the integration over solid angle amounts to a multiplication by π steradians (this is equivalent to the geometry of Example 1.2b but for the receiving, rather than the emitting, case). The result is a flux of $f = 8.73 \times 10^{-8}\,\mathrm{s}^{-1}\,\mathrm{m}^{-2}$ at any position at the top of the atmosphere. Finally we integrate over the surface area of the Earth at the top of the atmosphere, which we take to be $4\pi {R_\oplus}^2$, since the radius of the Earth plus its atmosphere is approximately equal to $R_\oplus$ (6.371×10^6 m, Table G.3). The result is $\approx$ 45 million cosmic ray particles s^{-1} for this energy range.

The cosmic ray energy spectrum for electrons is also a power law but it shows some differences with respect to nucleons, especially at low energies because electrons are more strongly influenced by the Solar wind (see next section). It is likely that the electron spectrum is only free of this modulation at electron energies above 10 GeV (Ref. [101]). If the solar modulation is corrected for, the electron spectrum can be fitted by the function (Ref. [34]),

$$J_\mathrm{e}(E) = 412\,E^{-3.44} \quad (3\,\mathrm{GeV} < E < 2\,\mathrm{TeV}) \tag{3.43}$$

where E is the electron energy in GeV and the units of $J_e(E)$ are the same as in Eq. (3.41). The value of the constant of proportionality is less certain than that of the slope, due to difficulties in combining data from many different experiments using different instruments. Thus, Eq. (3.43) indicates that the CR electron spectrum is significantly steeper than the $\Gamma \approx 2.7$ found for protons and heavier nuclei (mainly protons).

This difference in spectral index between electrons and protons can be accounted for by the greater susceptibility of electrons to interactions and energy losses in the ISM. A variety of energy loss mechanisms exist for both electrons and protons, but electrons lose energy more readily via *synchrotron radiation* (see Sect. 8.5) and *inverse Compton scattering* (Sect. 8.6) which do not affect nucleons. Since energy losses are *greater* for higher energy particles, the result is a *steepening* of the CR spectrum. In fact, the spectral indexes, Γ_0, of *both* electrons and protons are believed to be the same initially and in the range, $2.1 \lesssim \Gamma_0 \lesssim 2.5$, depending on the specific acceleration mechanism (Ref. [15]).

We can now return to the problem of the observed low electron fraction of cosmic rays, that is, a measured electron to proton number fraction of about 2 per cent. This value is measured at fixed kinetic energy, T. Although a variety of acceleration mechanisms may be present, several of the better understood mechanisms in the intermediate energy range (e.g. shock acceleration, see next section) give a distribution in *momentum*, p, which is the *same* for electrons and protons except for a constant of proportionality, i.e. $N(p) \propto p^{-\Gamma_0}$, and assume that the *total* number of particles above a fixed lower kinetic energy, T_0, is the same for both electrons and protons. However, at a fixed lower T_0, the electron momentum is less than the proton momentum which means that the electron distribution is being sampled to lower momenta where there are more particles. For the total number of electrons to equal the total number of protons,

the number of electrons must be less than the number of protons at fixed T. Essentially the plot of $N(T)$ for protons is shifted to the right (higher energy) in comparison to the electron spectrum. This argument is described mathematically in the Appendix at the end of this chapter. See also Ref. [147] for further details.

3.6.3 The origin of cosmic rays

What kinds of astronomical sources can accelerate particles to such high energies? For some of the energy range of Figure 3.25, the answer to this question is unknown and the subject of much current research. To accelerate CR particles requires a source or sources of very high energy, possibilities including shock waves associated with supernovae, massive stellar winds, neutron stars, regions around black holes, active galactic nuclei, γ-ray bursts (Sect. 5.1.2.2), and others. Even when the most powerful of sources are considered, accelerating particles to Lorentz factors of order 10^{11} ($E \approx 10^{20}$ eV) strains the limits of known acceleration mechanisms. Aside from the problems of corrections for the atmosphere, ISM interactions and subsequent energy losses with spectral steepening, there is an additional problem in simply identifying the locations of the sources. Cosmic rays easily scatter from magnetic field lines in the galaxy (except at very high energy, see below) and therefore propagate by diffusion, similar to the way a photon would take a random walk out of the Sun (Sect. 3.4.4). Therefore, even if the angle at which a particle impinges on the Earth can be determined, it will not represent the true direction of the source. There are some energy ranges, however, for which the origin of CRs is thought to be understood.

Below about 1 GeV, for example, the CR spectral index changes in Figure 3.25, becoming much flatter. The arrival of these low energy cosmic rays is correlated with Solar activity, indicating that these particles originate from the Sun. A stream of high energy particles called the *Solar wind* continuously leaves the Sun, beginning in the hot, tenuous Solar corona that is shown in Figure 6.6. In addition, bursts of high energy particles also result from activity such as Solar flares (Sect. 6.4.2). Thus, we know that at least one star is responsible for a fraction of the CR flux.

At higher energies, CRs are believed to come from outside of the Solar System with the bulk of the particles originating from sources within the Milky Way. Energetically, ordinary stars or isolated neutron stars in the Milky Way cannot account for the total flux of these higher energy CRs. Supernovae, however, provide sufficient energy. Only 10 per cent of the kinetic energy of all supernovae in the Galaxy would be needed to explain the CRs in the Galaxy at energies up to about the knee of Figure 3.25. This connection has been strengthened by theoretical success at explaining the CR energies and spectral index via acceleration in the shock waves created by supernovae. The detection of synchrotron radiation (see Sect. 8.5), which is emitted by the electron component of CRs, in supernova remnants in the Milky Way provides a further link, as does the detection of TeV γ-rays from supernova remnants. A supernova origin also connects CRs with hot, massive metal-enriched stars since these are the only kinds of stars that produce supernovae (Sect. 3.3.3) and

can also account for the underabundance of hydrogen (Sect. 3.6.1). All evidence considered, supernovae appear to be the dominant source of mid-energy cosmic rays up to the knee.

Once energies above the knee are considered, the situation becomes murkier and the origin of these CRs is not as well understood. At these energies, cosmic rays more readily escape from the galaxy and the admixture of Galactic sources may also change from one in which supernovae dominate. There is also the possibility that this point represents a transition from Galactic to extra-galactic sources. CRs above the ankle are even more mysterious, posing a challenge for modern theory. The fact that particles with energies greater than 10^{20} eV are observed at all is a puzzle of great current interest. If these particles are extra-galactic in origin, they will meet photons from the cosmic microwave background (CMB, Sect. 3.1) while travelling through intergalactic space. Interactions of such high energy particles, mostly protons, with CMB photons can occur via a variety of reactions but, most notably, the formation of pions with a subsequent significant energy loss for the protons. The energy at which this occurs, about 5×10^{19} eV, is called the GZK (for Greisen–Zatsepin–Kuz'min) cutoff. This puts a limit on the region of space within which higher energy particles could have originated, a distance that is approximately 50 Mpc (a few tens to a hundred, depending on energy). A challenge for observers is also that the flux of these particles is extremely low, only about 1 particle per square kilometre per century! On the positive side, at such high energies, these CRs do not scatter so strongly off of magnetic field lines and therefore their apparent angle in the sky should point back more closely to their place of origin. Pinning down the energy source for these highest energy CRs is therefore only a matter of time.

Problems

3.1 (a) Write a single equation for the estimated lifetime of a star on the main sequence, making the same assumptions as in Example 3.1. Evaluate known quantities and express the result in years with the stellar mass and luminosity in units of $M_\odot$ and $L_\odot$, respectively.

(b) Refer to Table G.7 and find the bolometric magnitude and luminosity of a $17 M_\odot$ star. Evaluate the main sequence lifetime of this star, assuming that the formation time of a star and all subsequent evolution off the main sequence are small fractions of the main sequence lifetime. If the star formation is continuous, determine how many generations of $17\,M_\odot$ stars could occur during a single Solar lifetime.

(c) Evaluate the main sequence lifetime of a $0.34\,M_\odot$ star with the same assumptions as in part (b) and compare it to the age of the Universe.

3.2 Approximating stellar mass distributions by the IMFs shown in Figure 3.10 and provided in Eqs. (3.5) and (3.6):

(a) What is the the number density of all stars formed from 0.1 to 100 $M_\odot$ in the disk and in the halo?

(b) Determine the mean separation between stars in both regions of the Galaxy.

(c) Express the mean separations of part (b) in terms of a 'typical' stellar diameter, D_t, for each region.

(d) Assume that the effective diameter for star–star collisions is $10\,D_t$. If the random velocities of stars in the disk are of order $10\,\mathrm{km\,s^{-1}}$, compute the mean time between star–star collisions in the disk. What can you conclude about the probability of such collisions in the Galaxy, excluding stars in dense clusters (see Footnote 6 in Chapter 10).

3.3 For any given collection of stars that are formed in the disk of the Milky Way, determine the fraction of stars that will become supernovae. If star formation continues in the same region until the gas is used up, should the number of supernovae in comparison to the total number of stars in the region, increase, decrease, or stay the same with time? Explain.

3.4 From the Maxwellian distribution (Eq. 0.A.2), derive Eqs. (3.8) and (3.9).

3.5 Show that Eqs. (3.13) result from Eq. (3.1) and Eq. (3.12).

3.6 For the HI shell shown in Figure 3.17, assume that the initial ISM had Solar abundance, uniform density, and $T = 200$ K before the supernova swept up the gas.

(a) Calculate the average density, n, of particles in the pre-supernova (pre-SN) ISM. (You may assume that the mean atomic weight of the metals can be approximated by that of oxygen.)

(b) Determine, for the pre-SN ISM, $\bar{l}$ (AU), $\bar{t}$ (yr), $\mathcal{R}$ (s^{-1}), and $\bar{r}$ (cm). Is this an ideal gas?

3.7 (a) Calculate the time it would take a photon to travel from the core to the surface of the Sun if there were no interactions en route.

(b) Adopting a mean free path of $\bar{l} = 0.3$ cm, determine the time (years) it takes for a photon to diffuse from the core to the surface of the Sun (see Eq. 3.22) when the interaction time is negligible.

(c) Write an expression for the diffusion time of a photon assuming each interaction takes 10^{-8} s and re-evaluate this time for the Sun.

3.8 For the following two ionized regions, determine the mean separation between particles, $\bar{r}$, and the mean free path for a photon, $\bar{l}$. Indicate whether the region should be considered an ideal gas and also whether it is in LTE. For simplicity, assume that

Eq. (3.4) holds for both regions and that the most likely interaction for the photon is scattering from a free electron[39]

(a) The solar core at $R/R_\odot = 0.025$ (Table G.10).

(b) A diffuse region in the Galaxy of diameter, $D = 1$ pc ($T = 10^5$ K, $n_e = 0.001\,\mathrm{cm}^{-3}$).

3.9 (a) Write an expression for the fraction of hydrogen that is in the neutral state, N_{HI}/N_{tot}, for a gas in LTE. Evaluate all constants and express the result as a function of n_e and T.

(b) Consult Table G.11 to determine n_e and T applicable to the Sun's surface (i.e. letting $x = 0$ km) and, assuming LTE, evaluate N_{HI}/N_{tot}.

(c) What can you conclude about the ionization state of hydrogen at the surface of the Sun?

3.10 (a) Write an expression for the temperature, T as a function of energy level, n, for an LTE situation in which the number of hydrogen atoms with electrons in level n is equivalent to the number in the ground state.

(b) Using a spreadsheet or computer algebra software, compute T for the first 300 energy levels of hydrogen. Plot T(n).

(c) Briefly explain or comment on the behaviour of the plot.

3.11 (a) Find the factor by which the distance will be in error if dust extinction is ignored, i.e. find f, such that $d_{true} = d_{err} f$, where d_{true} is the true distance to the object, and d_{err} is the distance calculated without taking A_V into account.

(b) Suppose a F0Ib star is measured to have an apparent magnitude of $V = 4.5$. If $A_V = 0.8$, calculate the distance (pc) to the star, with and without the assumption of dust extinction.

3.12 (a) Determine the ratio of the number of large classical ($0.10 \leq a \leq 0.25\,\mu$m) dust grains per unit volume in comparison to VSGs ($0.005 \leq a \leq 0.01\,\mu$m).

(b) If all of these grains have the same shape and mass density, determine the ratio of the mass of classical grains in comparison to VSGs in some volume of space.

3.13 (a) Write Eq. (3.40) in the same manner as Eq. (3.41) but for the energy range, $10^2 \leq E\,(\mathrm{GeV}) \leq 4 \times 10^6$. That is, ensure that K is evaluated and use the appropriate value of Γ. Assume that Eq. (3.41) applies at an energy of $E = 4 \times 10^6$ GeV.

(b) What is the total rate of CR particles (s^{-1}) in this energy range hitting the Moon? For simplicity, assume that CRs are isotropic and all CRs hitting the Moon are absorbed.

[39]Note, however, that other types of interactions may be important depending on frequency (see Sect. 5.1.1.2).

Appendix: the electron/proton ratio in cosmic rays

We start by adopting a distribution in momentum, p, which is the same for both electrons and protons, but with the possibility of a different constant of proportionality,

$$N_{\mathrm{CR}}(p) = N_{0\mathrm{CR}}\, p^{-\Gamma_0} \tag{3.A.1}$$

where $N_{0\mathrm{CR}}$ is the constant of proportionality and the subscript, CR, refers to the CR species being considered (electrons, e, or protons, p). We also assume that the total number of electrons over all kinetic energy, T, above a non-relativistic threshold, $T_0 = 10$ keV (Ref [147]), equals the total number of protons over the same energy range, i.e. N_{tot}.

For N_{tot} to be the same for both CR species, then we require an integration over the distribution in energy, or the distribution in momentum, to equal this value,

$$N_{\mathrm{tot}} = \int_{T_0}^{\infty} N_{\mathrm{CR}}(T)dT \;=\; \int_{p_{0\mathrm{CR}}}^{\infty} N_{\mathrm{CR}}(p)dp \tag{3.A.2}$$

where $p_{0\mathrm{CR}}$ is the lower momentum limit that corresponds to a fixed lower limit in kinetic energy, T_0, the latter being the same for both electrons and protons. Substituting Eq. (3.A.1) for $N_{\mathrm{CR}}(p)$ and integrating gives,

$$N_{\mathrm{tot}} = \frac{N_{0\mathrm{CR}}}{(\Gamma_0 - 1)\, p_{0\mathrm{CR}}^{\Gamma_0 - 1}} \tag{3.A.3}$$

Notice that, because the power law in momentum is a declining function, the limit as $p \rightarrow \infty$ goes to zero, so it is the *lower* momentum limit, $p_{0\mathrm{CR}}$, that determines N_{tot}. Another way of stating this is that most of the particles are at low momentum (low energy). Therefore, although we ultimately want to compare electron to proton ratios at high energies, it is important for this integration to include the low energy particles. Since both N_{tot} and Γ_0 are the same for both electrons and protons, rearranging Eq. (3.A.3) for the ratio of the electron to proton constants of proportionality gives,

$$\frac{N_{0\mathrm{e}}}{N_{0\mathrm{p}}} = \left(\frac{p_{0\mathrm{e}}}{p_{0\mathrm{p}}}\right)^{\Gamma_0 - 1} \tag{3.A.4}$$

We now require a relation between momentum and kinetic energy in order to relate $p_{0\mathrm{CR}}$ to a given fixed T_0. The kinetic energy is just the total energy, E, less the rest mass energy, so using Eq. (3.36),

$$T = (\gamma - 1)\, \mathrm{m}_{\mathrm{CR}}\, c^2 \tag{3.A.5}$$

where m_{CR} is the rest mass of the CR particle. The momentum of a particle (see Table I.1) is $p_{CR} = \gamma\, m_{CR}\, v$. Using Eq. (3.36) to eliminate the velocity, v, we find,

$$p_{CR} = m_{CR}\, c\, \sqrt{\gamma^2 - 1} \tag{3.A.6}$$

Using Eqs. (3.A.5) and (3.A.6) and eliminating γ, we obtain a relation between kinetic energy and momentum,

$$p_{CR} = \frac{1}{c}\left[T^2 \,+\, 2\,T\, m_{CR}\, c^2\right]^{1/2} \quad (\text{all } T) \tag{3.A.7}$$

$$\approx \frac{E}{c} \qquad (T \gg m_{CR}\, c^2) \tag{3.A.8}$$

Eq. (3.A.8), for which Eq. (3.A.5) reduces to Eq. (3.36) (i.e. $T \approx E$), is what is expected for a relativistic particle[40]. The general expression, Eq. (3.A.7), indicates that, at a fixed kinetic energy for both electrons and protons, the momentum of an electron will be less than the momentum of a proton since $m_e < m_p$. Therefore, $p_{0e} < p_{0p}$, so the electron distribution is being sampled to a lower momentum than is the proton distribution. Examination of Eq. (3.A.4) then requires that $N_{0e} < N_{0p}$. Using Eq. (3.A.7) in Eq. (3.A.4) with the lower limit, p_{0CR}, expressed in terms of T_0, we obtain,

$$\frac{N_{0e}}{N_{0p}} = \left[\frac{T_0 + 2\, m_e\, c^2}{T_0 + 2\, m_p\, c^2}\right]^{\frac{\Gamma_0 - 1}{2}} \tag{3.A.9}$$

Since the lower kinetic energy limit was given as $T_0 = 10$ keV whereas the rest energies, $m_{CR}\, c^2$, of an electron and a proton are much higher at 511 keV and 9.4×10^5 keV, respectively, this ratio reduces to,

$$\frac{N_{0e}}{N_{0p}} = \left[\frac{m_e}{m_p}\right]^{\frac{\Gamma_0 - 1}{2}} \tag{3.A.10}$$

Now we wish to compare the fraction of electrons to protons at high energies where they are typically measured. At fixed high energy, $E \approx T$, and we put Eq. (3.A.10) into Eq. (3.A.1) using Eq. (3.A.8) to find,

$$\frac{N_e(E)}{N_p(E)} \approx \left[\frac{m_e}{m_p}\right]^{\frac{\Gamma_0 - 1}{2}} \tag{3.A.11}$$

Using an initial spectral index of $\Gamma_0 = 2.2$, we find $\frac{N_e(E)}{N_p(E)} \approx 1$ per cent which is approximately what is observed (Ref. [147]).

[40]The momentum of a photon, for example, is expressed in this way (see Table I.1)

4

Radiation Essentials

And God said, Let there be light: and there was light
– Genesis 1:3

4.1 Black body radiation

When a gas is in thermodynamic equilibrium (TE), the absorption and emission rates in the gas are in balance. Such a situation could be set up if the interior walls of a box were maintained at a single temperature, T. If no matter or radiation were permitted to escape from the box, then the gas particles and all radiation within it would eventually reach an equilibrium at this temperature, independent of the kind of material in the box or its shape. It can be shown that the resulting radiation field will be isotropic and that its spectrum (the emission as a function of frequency or wavelength) depends only upon T. Such radiation is referred to as *black body radiation* (sometimes called 'cavity radiation' because of the history of studying such radiation by setting up a 'thermal bath' within a cavity). The resulting specific intensity is described by a particular function called the *Planck function* or *Planck curve* (see the Appendix at the end of this chapter for a derivation) which, in frequency-dependent and wavelength-dependent forms, respectively, are given by,

$$B_\nu(\mathrm{T}) = \frac{2h\nu^3}{c^2}\frac{1}{\left(e^{\frac{h\nu}{k\mathrm{T}}}-1\right)} \qquad \mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}\,sr^{-1}} \tag{4.1}$$

$$B_\lambda(\mathrm{T}) = \frac{2hc^2}{\lambda^5}\frac{1}{\left(e^{\frac{hc}{\lambda k\mathrm{T}}}-1\right)} \qquad \mathrm{erg\,s^{-1}\,cm^{-2}\,cm^{-1}\,sr^{-1}} \tag{4.2}$$

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

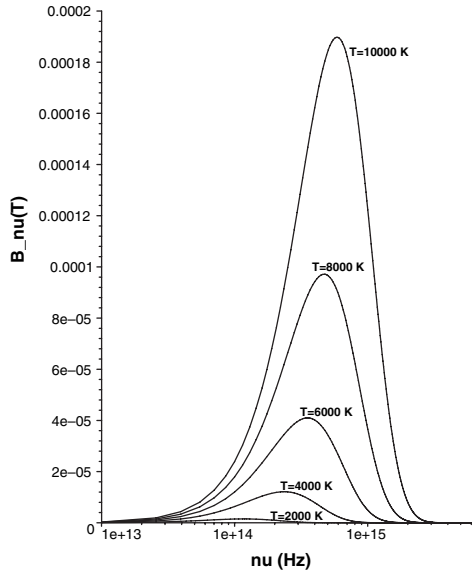


Figure 4.1. Planck curves for temperatures from 2000 to 10 000 K. The ordinate axis is in units of erg $s^{-1}\,cm^{-2}\,Hz^{-1}\,sr^{-1}$. Note that the peaks of the curves shift to higher frequency (smaller wavelength) as the temperature increases, and also that hotter objects have higher specific intensities than cooler objects at *all* frequencies.

where cgs units have been specified. Plots of $B_\nu(\mathrm{T})$ for different temperatures are shown in Figure 4.1 and illustrate that, at higher temperatures, the specific intensity is higher at all frequencies. The two functions are related to each other via,

$$B_\nu\, d\nu = B_\lambda\, d\lambda \tag{4.3}$$

Thus, the integrals under the two curves, from zero to infinity, are equal. However, the functions themselves are different. They have different maxima (Sect. 4.1.3) and Eq. (1.5) must be used to convert from one form to the other. Since $B_\nu(\mathrm{T})$ and $B_\lambda(\mathrm{T})$ are specific intensities, the relations given in Chapter 1 involving specific intensity also apply to the Planck function.

Since T in Eqs. (4.1) and (4.2) is a value that gives the specific intensity of the *radiation*, it is a 'radiation temperature'. However, we noted in Sect. 3.4.4 that for TE, or LTE 'locally', the radiation temperature is equal to the kinetic temperature of the gas. Thus, T in the above equations is written without subscript, indicating that it represents kinetic temperature. Since there must be sufficient interactions between radiation and matter for an equilibrium temperature to be established, this implies that the gas must be *opaque*, as illustrated in Figure 4.2. For an object to be opaque, the mean free path of a photon must be less than the size of the object. The observer sees into the object over a distance about equal to the mean free path. We thus come to the more formal definition of a black body, that is, a black body is an object that is a *perfect*

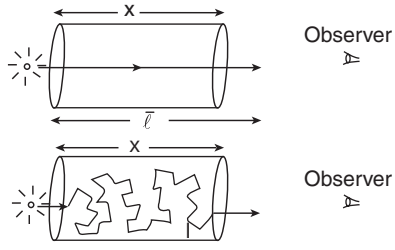


Figure 4.2. In these figures, a single source is shining just behind a gas cloud of temperature, T, as seen by an observer on the right. In the top picture, the mean free path of a photon, $\bar{l}$, is greater than the thickness, x, of the cloud. In such a case, the observer can see through the cloud to the source. In the bottom picture, the mean free path of a photon is much smaller than the cloud thickness. The photon diffuses to the other side (similar to Figure 3.15), coming into equilibrium with the temperature of the cloud. The observer will see into the cloud only as far as a mean free path, i.e only into the outer 'skin' of the cloud, and will measure a Planck spectrum at the temperature of this skin.

absorber. If a photon, regardless of frequency, impinges on a black body, it will neither be reflected back nor pass through unimpeded. The photon will be absorbed, implying that its mean free path is less than the object's size. Any opaque, non-reflecting object at a single temperature is therefore a black body. This could include solids, provided the solid material does not reflect light. For an object to remain at a single temperature, absorptions and emissions must be in balance. Thus a black body must also be a *perfect emitter*. Black bodies are therefore *not* black in colour nor does the word 'black' imply that there is no emission. On the contrary, a black body emits the continuous Planck spectrum whose shape is dictated by its temperature.

As we found for thermodynamic equilibrium (Sect. 3.4.4), identifying a perfect black body in nature is difficult. However, there is one excellent example, that of the cosmic microwave background (CMB) radiation that was shown in Figure 3.2. A Planck spectrum is measured at every position on this map, the slight shifts in temperature at different positions resulting in slightly different Planck curves at different positions. The global mean Planck curve is given in Figure 4.3 and is well fitted by a temperature of 2.73 K.

Stars are also very good examples of black bodies, although some departures from a perfect curve do occur, as shown for the Sun in Figure 4.4. Even though the entire star is not at a single temperature, the interior being hotter than the surface, we cannot see into the interior (Figure 4.2) and, over the depth that can be viewed, the temperature is approximately constant. Since stars have temperatures that correspond to Planck curves that have peaks at or near the optical part of the spectrum, most of the astrophysical radiation that we see with our eyes, whether naked eye or through optical telescopes, is due to stars. This includes the Sun, of course, and it is unlikely to be an accident that our eyes have the greatest sensitivity at just the region over which the Solar Planck curve peaks (Figure 4.5). When we look out into the vast depths of space,

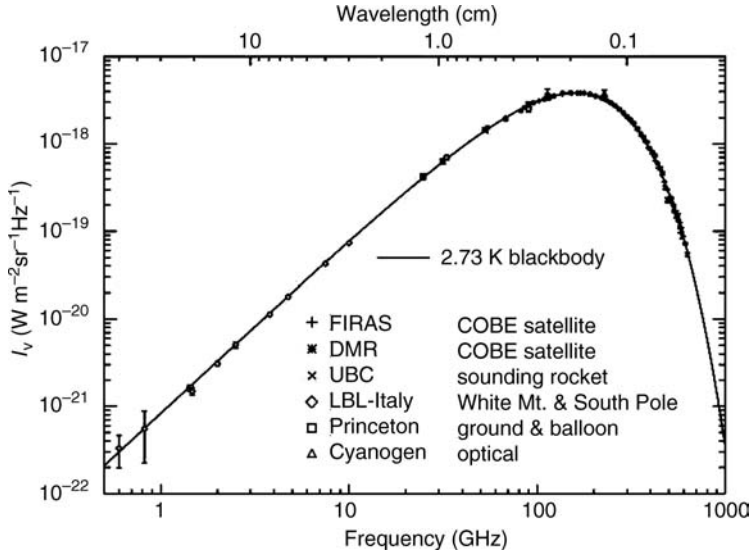


Figure 4.3. Data points for this curve are measurements of the cosmic microwave background radiation from a variety of sources that are listed on the plot. The solid curve is the best fit black body spectrum, corresponding to a temperature of 2.73 K (G.F. Smoot and D. Scott 2002, in *The Review of Particle Physics*, K. Hagiwara *et al.*, Phys. Rev. **D66**, 010001 (2002))

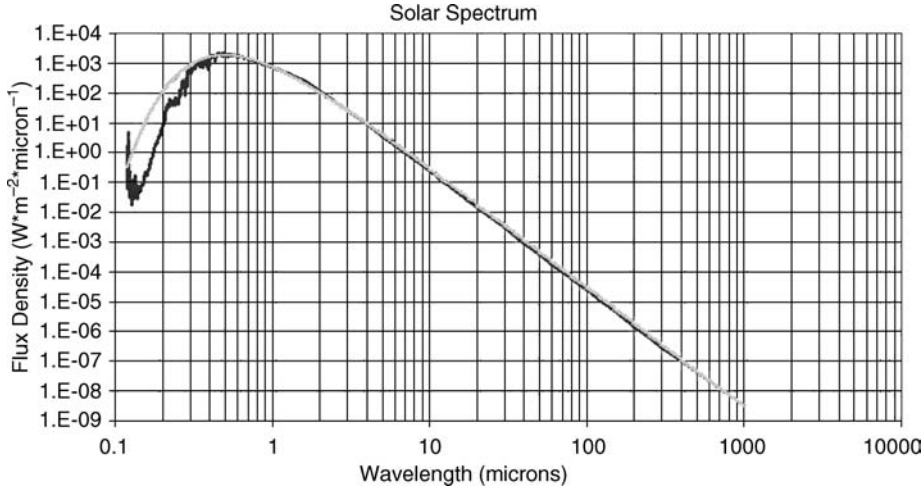


Figure 4.4. The Solar spectrum, shown as a function of increasing wavelength in a logarithmic plot. The plot is of flux density as measured from the distance of the Earth; to convert to specific intensity, a division by $\Omega_\odot$ is required (see Eq. 1.13). The spectrum (black curve) is well fitted by a black body at T = 5781 K (grey shaded curve) but starts to depart significantly in the ultraviolet ($\lambda \lesssim 0.3$, μm due to numerous absorption lines. There are also significant departures in the X-ray and radio regions of the spectrum (not shown) due to Solar activity. These data are consistent with a Solar Constant (cf. Figure 1.6) of 1366.1 W m^{-2}.

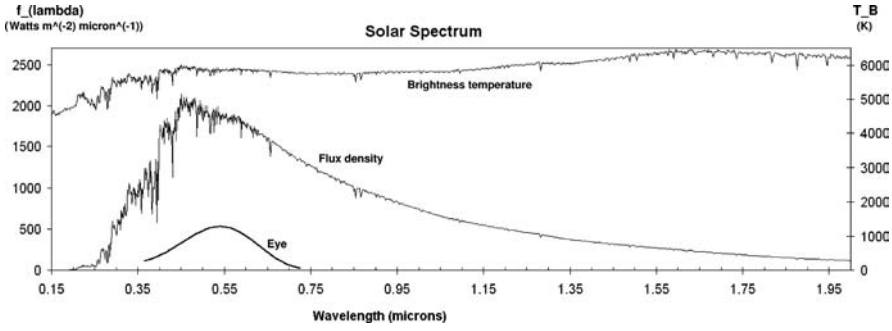


Figure 4.5. The Solar spectrum is again shown (middle curve, left scale) using the same data as in Figure 4.4 except over a more restricted wavelength range and on a linear scale. The brightness temperature (top curve, right scale) is also shown, derived by dividing the flux density by $\Omega_\odot$ (Eq. 1.13) and then using the Planck formula (the wavelength form of Eq. 4.4) to recover T_B. The bottom curve, in arbitrary units, is the daylight photon flux response of the human eye smoothed to 30 nm resolution (see also Figure 2.1). (Eye response reproduced by permission of James T. Fulton, 2005, www.sightresearch.net/luminouseffic.htm)

it is starlight that we see, from our own Galaxy, as shown in Figure 3.3, to nearby galaxies like those of Figure 3.8 and Figure 2.19, to distant galaxy clusters like the one of Figure 7.6. Of all the emission mechanisms that we will discuss (see Part IV), it is black body radiation, by far, that has the most relevance to our visual appreciation of the sky.

If we can measure the specific intensity of an object and *know* it to be a black body, then by Eqs. (4.1) and (4.2), we can determine its temperature. Recall that specific intensity is independent of distance in the absence of intervening matter (see Sect. 1.3), which means that we do not need to know anything else about the object, not even its distance, to find T. Moreover, in principle, only a single measurement at one frequency is required to give this result, although in practice, more measurements are usually needed to ensure that the spectrum is indeed Planckian. The flip side of this coin is that, since the Planck curve depends *only* on temperature, we can discern nothing about the kind or amount of material that emits this radiation by studying its Planck curve alone. Whether the black body is the glowing filament of an incandescent light bulb, 'super black' man-made paint that reflects less than 1 per cent of the light that falls on it, or the interior of a kitchen oven after an equilibrium temperature has been achieved, the same Planck spectrum, dependent on T alone, will be measured.

4.1.1 The brightness temperature

Other than the CMB, most objects that we refer to as black bodies have emission spectra that, like stars, are approximations to the Planck curve. To account for departures from the Planck curve, we define the *brightness temperature* $T_{B\nu}$, at the

observing frequency, ν, to be *the temperature which, when put into the Planck formula, results in the specific intensity that is actually measured at that frequency*,

$$I_\nu = \frac{2h\nu^3}{c^2}\frac{1}{\left(e^{\frac{h\nu}{kT_{B\nu}}}-1\right)} \tag{4.4}$$

The wavelength form can also be used. It is clear that, if the *same* value of $T_{B\nu}$ is measured for all ν, then the object is a true black body. If $T_{B\nu}$ is different for different frequencies, then the object departs from a black body (Example 4.1). The extent to which $T_{B\nu}$ is *not* constant is a measure of how much the object departs from a black body. The brightness temperature of the Sun as a function of wavelength is shown in Figure 4.5. If the Sun were a perfect black body, then the brightness temperature curve would be a straight horizontal line in this plot. This is close to being true over the wavelength region shown, although there are some minor variations. However, in the radio part of the Solar spectrum (not shown), the departures are much more significant. For example, at λ 10 m, $T_{B\nu} \approx 5 \times 10^5$ K, almost two orders of magnitude higher than shown in Figure 4.5, indicating that processes other than black body emission are occurring (e.g. Sect. 8.5)[1].

This definition of brightness temperature can be extended to include *any* astrophysical source, whether or not its emission approximates that of a black body. For example, $T_{B\nu}$ could be specified even for an object whose spectrum is a power law (e.g. Sect. 8.5). The brightness temperature simply provides a convenient way of expressing the radiation temperature of an object, whether or not a physical (kinetic) temperature is implied. Eq. (4.4) or its wavelength equivalent is a straightforward functional relation between specific intensity and brightness temperature. Therefore, to say that an object has a certain brightness temperature is essentially equivalent to saying that it has a certain specific intensity, and vice versa.

Example 4.1

In a particular kind of telescope, the optical path is designed to switch back and forth between pointing at a target in the sky and pointing at a reference black body which is a rotating vane used for calibration. An engineer must first fine tune the calibrator in the lab so that, over λ 2 cm *to* λ 20 cm*, it emits as a true black body at a temperature of* 50 K *to within* $\pm$2 *K. After experimenting with different kinds of non-reflective paints and methods of obtaining a uniform temperature, he measures the specific intensity at the two end points of the wavelength range finding,* $I_{2\,\mathrm{cm}} = 2.57 \times 10^{-5}\mathrm{erg\,s^{-1}cm^{-2}cm^{-1}sr^{-1}}$ *and* $I_{20\,\mathrm{cm}} = 2.24 \times 10^{-9}$ *(in the same units). What is its brightness temperature at each wavelength? Is the calibrator a black body to within the desired tolerance?*

[1]This emission refers to the *radio quiet Sun*. When there are *radio bursts*, the brightness temperature can be up to $10^7\times$ higher than this.

Since the units imply that the λ form of the Planck curve is being used, we rearrange Eq. (4.2) to solve for the brightness temperature,

$$\mathrm{T}_{\mathrm{B}_\lambda} = \left(\frac{h\,c}{\lambda\,k}\right) \frac{1}{\ln\left(1 + \frac{2hc^2}{\lambda^5 I_\lambda}\right)} \tag{4.5}$$

From this equation, we find that the brightnesses temperatures at λ 2 cm and λ 20 cm are, respectively, $\mathrm{T}_{\mathrm{B}_{2\,\mathrm{cm}}} = 50.1$ K and $\mathrm{T}_{\mathrm{B}_{20\,\mathrm{cm}}} = 43.4$ K. The brightness temperature at λ 2 cm is within the desired tolerance but the brightness temperature at λ 20 cm is not. Thus, the calibrator is not yet a black body.

4.1.2 The Rayleigh–Jeans Law and Wien's Law

The *Rayleigh–Jeans Law* and *Wien's Law* are not special 'laws' in a fundamental sense, but are rather approximations to the Planck curve at the low and high frequency (or wavelength) ranges. For example, suppose we observe the Planck curve at a low frequency in comparison to the peak of the curve or, more technically, in the regime, $h\,\nu \ll k\mathrm{T}$ or $\frac{h\nu}{k\mathrm{T}} \ll 1$. Then,

$$B_\nu(\mathrm{T}) = \frac{2h\nu^3}{c^2} \frac{1}{\left(e^{\frac{h\nu}{k\mathrm{T}}} - 1\right)} \approx \frac{2h\nu^3}{c^2} \frac{1}{\left(1 + \frac{h\nu}{k\mathrm{T}} + \ldots - 1\right)} = \frac{2\nu^2 k\mathrm{T}}{c^2} \tag{4.6}$$

where we have used an exponential expansion (Eq. A.3). Note that higher orders are not needed in the expansion because the term $\frac{h\nu}{k\mathrm{T}}$ is less than 1 and is therefore diminishingly small in higher orders. Eq. (4.6), or its wavelength equivalent (Prob. 4.2), is called the *Rayleigh–Jeans Law*. The brightness temperature can also be expressed in this limit by replacing $B_\nu(\mathrm{T})$ with I_ν and T with $\mathrm{T}_{\mathrm{B}\nu}$ in Eq. (4.6).

On the other end of the spectrum, where $\frac{h\nu}{k\mathrm{T}} \gg 1$, the Planck function can be written,

$$B_\nu(\mathrm{T}) = \frac{2h\nu^3}{c^2} \frac{1}{\left(e^{\frac{h\nu}{k\mathrm{T}}} - 1\right)} \approx \frac{2h\nu^3}{c^2} \left(e^{-\frac{h\nu}{k\mathrm{T}}}\right) \tag{4.7}$$

In this regime, the exponential term dominates over ν^3 so the function diminishes rapidly as the frequency increases. Eq. (4.7), or its wavelength equivalent (Prob. 4.2), is called *Wien's Law*.

Examples of the two approximations are shown in Figure 4.6. Example 4.2 indicates when they are valid to within 10 per cent.

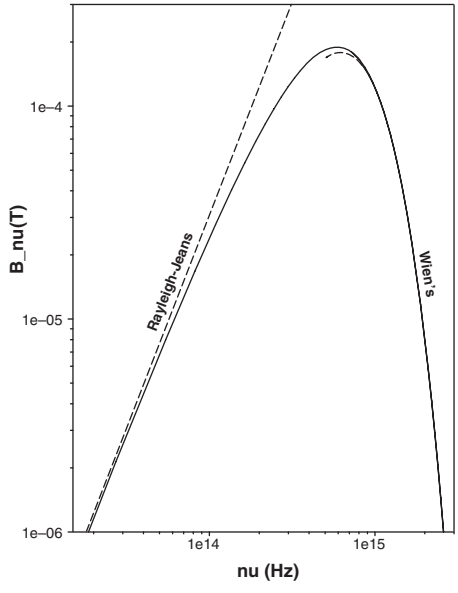


Figure 4.6. A Planck curve, in cgs units, for T = 10 000 K (solid curve) showing the Rayleigh–Jeans and Wien's approximations (both dashed).

Example 4.2

For what values of $\frac{h\nu}{k\mathrm{T}}$ do the Rayleigh–Jeans and Wien's approximations depart from the Planck curve by more than 10 *per cent?*

This can be determined by letting $x = \frac{h\nu}{k\mathrm{T}}$ and setting the ratio of Eq. (4.6) over Eq. (4.1) equal to 1.1, i.e. $(e^x - 1)/x = 1.1$. The solution to this equation is $x = 0.19$. For Wien's Law, we set the ratio of Eq. (4.7) over Eq. (4.1) to 0.9 since Wien's approximation gives values lower than the full Planck equation, $e^{-x}(e^x - 1) = (1 - e^{-x}) = 0.9$. In this case, $x = 2.3$. Therefore, the use of the Rayleigh–Jeans law will give an accuracy of better than 10 per cent, provided $\frac{h\nu}{k\mathrm{T}} < 0.19$ and Wien's law gives better than 10 per cent accuracy if $\frac{h\nu}{k\mathrm{T}} > 2.3$.

4.1.3 Wien's Displacement Law and stellar colours

To find the wavelength, or frequency, of the peak of the Planck curve, we need only differentiate Eq. (4.1), or Eq. (4.2), respectively, and set the result to zero, the result being (cgs),

$$\begin{aligned} \lambda_{\max}\mathrm{T} &= 0.290 \\ \nu_{\max} &= 5.88 \times 10^{10}\,\mathrm{T} \end{aligned} \tag{4.8}$$

This is called *Wien's Displacement Law*[2]. It indicates how the peak of the Planck curve shifts with wavelength or frequency. Note that $\lambda_{\max} \neq c/\nu_{\max}$ since the wavelength and frequency forms of the Planck curve are different functions, as noted in Sect. 4.1. Figure 4.1 illustrates that, as the temperature increases, not only does the specific intensity increase, but also the peak of the curve shifts to higher frequencies as Wien's Displacement Law predicts.

These simple relationships provide very powerful tools for obtaining the temperature of a black body. While only a single measurement of $B_\nu(\mathrm{T})$, in principle, is required to determine T (Sect. 4.1), for many objects, stars in particular, the specific intensity cannot be measured because the object is unresolved (see Figure 1.9.). In such cases, it is the flux density that is measured instead. However, the flux density of the black body, as observed by a telescope, is related to its specific intensity via $f_\nu = B_\nu(\mathrm{T})\,\Omega$, where Ω is the solid angle subtended by the object (Eq. 1.13). Although Ω is not known for an unresolved object, it is assumed not to vary with frequency[3], which means that the shape of $B_\nu(\mathrm{T})$ will be the same as the shape of f_ν. We show an example in Figure 4.5 in which it is the flux density of the Sun, rather than its specific intensity, that is plotted. For the Sun, we know $\Omega_\odot$, but we could easily plot f_ν for any star for which the solid angle is not known. By simply identifying the frequency of the peak of the spectrum and using Wien's Displacement Law, the temperature can be found. Note that this can be done without knowing the distance to the star.

Since most of the emission of any black body occurs around the peak of the curve, Wien's Displacement Law also explains how the colour of a hot opaque object changes with its temperature. Subtle colour differences of stars are quite visible to the naked eye. Examples can be found throughout the sky, but a good contrast is between the reddish star, Betelgeuse in Orion's shoulder and the blueish star Rigel in his foot (see Figure 5.8 for locations). The shift in the peak of the Planck curve provides the basis for the colour index used in the observational HR diagram shown in Figure 1.14 and discussed in Sect. 1.6.3. The ratio of Planck curve values at two different frequencies, or equivalently the ratio of two flux densities, gives a colour index via Eq. (1.32). The colour index of a star of unknown angular size and distance can be compared to (calibrated against) the colour index based on the ratios of Planck curves of known temperature in order to relate colour index to stellar temperature (see Example 4.3). Such a calibration is shown in Figure G.1 of Appendix G. Once a good calibration has been set up, one need only make two measurements of stellar flux density in order to determine the surface temperature. This procedure is even simpler than determining the peak of the spectrum, since to find the peak, a number of measurements would be required. A caveat is that stellar spectra do show departures from a perfect black body as Figure 4.4 illustrates. Thus, temperatures determined by different techniques are not always exactly the same. Also, the calibrations given in Tables G.7, G.8 and G.9 do not include the finer subclasses that are present over a continuum of stellar temperatures.

[2]This should not be confused with 'Wien's Law' of Sect. 4.1.2.

[3]Small changes in Ω can occur with frequency, but these are generally not sufficient to make significant changes in the shape of the spectrum.

Example 4.3

The star, Gomeisa (βCMi) is measured to have $\mathrm{B} = 2.79$ *and* $\mathrm{V} = 2.89$. *Determine its colour index,* $\mathrm{B} - \mathrm{V}$, *the ratio of its specific intensities at the two observing wavelengths,* $B_{\mathrm{B}}(\mathrm{T})/B_{\mathrm{V}}(\mathrm{T})$, *estimate its surface temperature,* T, *determine the wavelength of the peak of its Planck curve,* $\lambda_{\max}$, *and provide a best estimate of its spectral type, given that it is a main sequence star. Assume that there is no dust extinction in the line of sight.*

From the measurements, $\mathrm{B} - \mathrm{V} = 2.79 - 2.89 = -0.10$. Using Eq. (1.32) for colour index, $-0.10 = -2.5 \log\left(\frac{f_{\mathrm{B}}}{f_{\mathrm{V}}}\right) - (-0.601 - 0)$, where we have taken the zero point value from Table 1.1 and are using the wavelength form of the Planck curve since we are being asked for $\lambda_{\max}$. Solving for the ratio of flux densities, $\frac{f_{\mathrm{B}}}{f_{\mathrm{V}}} = 1.9$. This result is the same as the ratio of specific intensities, $\frac{B_{\mathrm{B}}(\mathrm{T})}{B_{\mathrm{V}}(\mathrm{T})}$, because the solid angle subtended by the source is considered constant with wavelength. Referring to the calibration of Table G.7 for main sequence stars, $\mathrm{B} - \mathrm{V} = -0.1$ corresponds to $\mathrm{T} = 11950\,\mathrm{K}$ and a spectral type of B8. By Wien's Displacement Law (Eq. 4.8), the spectrum would peak at $\lambda_{\max} = 243\,\mathrm{nm}$. As a check on the ratio of specific intensities, we can use the above temperature in the Planck equation (Eq. 4.2) and re-compute the ratio. Using the central wavelengths for B and V bands (Table 1.1), we find, $\frac{B_{\mathrm{B}}(11950)}{B_{\mathrm{V}}(11950)} = 1.7$ which is within 10 per cent of the value of 1.9 above and within 5 per cent of it when allowing for a range of temperatures between spectral types B5 and A0 in the table.

4.1.4 The Stefan–Boltzmann Law, stellar luminosity and the HR diagram

The total intensity of a black body can be found by integrating under the Planck curve over all frequencies, or all wavelengths,

$$B(\mathrm{T}) = \int_0^\infty B_\nu(\mathrm{T})\, d\nu = \int_0^\infty B_\lambda(\mathrm{T})\, d\lambda = \mathcal{K}\mathrm{T}^4 \qquad \mathrm{erg\,s^{-1}\,cm^{-2}\,sr^{-1}} \tag{4.9}$$

where,

$$\mathcal{K} = \frac{2(k\pi)^4}{15c^2h^3} \tag{4.10}$$

If we now use the relationship between intensity and flux given in Eq. (1.14) for emission escaping from a surface over 2π sr, then the resulting astrophysical flux through the surface of such a black body is,

$$F_{\mathrm{BB}} = \sigma\mathrm{T}^4 \tag{4.11}$$

where $\sigma = \pi\mathcal{K} = 5.67 \times 10^{-5} \mathrm{erg\,s^{-1}\,cm^{-2}\,K^{-4}}$ is the *Stefan–Boltzmann constant*. Eq. (4.11) is usually written in the more familiar form,

$$F = \sigma \mathrm{T_{eff}}^4 \qquad \mathrm{erg\,s^{-1}\,cm^{-2}} \tag{4.12}$$

This equation defines the *effective temperature*, $\mathrm{T_{eff}}$, which is *whatever temperature results in the flux observed from the object*. Like the brightness temperature, the effective temperature allows for departures from a perfect black body. In the former case, there could be a different brightness temperature at each wavelength, but in the latter case, there is only one effective temperature for any object. Also, although a brightness temperature may be quoted for any object, whether approximating a black body or not, the effective temperature is usually only quoted for objects whose emission resembles a Planck curve, especially stars. Eqs. (4.11) or (4.12) are referred to as the *Stefan–Boltzmann Law*.

Using this flux, it is straightforward to calculate the total (i.e. bolometric) luminosity of the object which, for a sphere of radius, R, is (see Eq. 1.10),

$$L = 4\pi R^2 \sigma \mathrm{T_{eff}^4} \tag{4.13}$$

It is now clear that the luminosity of a star or any black body is a strong function of its temperature. This is why, for any mixture of stars, the luminosity tends to be dominated by hot, high-mass stars (Prob. 4.7).

This equation also helps to clarify the placement of the stars on the HR diagram, an example of which is shown in Figure 1.14. Taking the logarithm of both sides of Eq. 4.13, $\log(L) = 4\log(\mathrm{T_{eff}}) + 2\log(R) + \text{constant}$. The log of the luminosity is related to the absolute visual magnitude (see Eqs. 1.31 and 1.33) which is the ordinate of the HR diagram. The temperature, T_{eff} is related to the colour index, B − V, as discussed in the previous section. The HR diagram, therefore, can be understood as a relationship between the luminosity and temperature of the star.

If all stars were the same size, then a plot of $\log(L)$ against $\log(\mathrm{T_{eff}})$ would yield a straight line of slope, 4. Although Figure 1.14 does not show exactly these axes, it is clear that stars are *not* randomly distributed in the HR diagram. For the main sequence, we see that cooler stars on the right are also much dimmer than the brilliant hotter stars on the left as Eq. (4.13) predicts. In fact, a plot of $\log(L)$ versus $\log(\mathrm{T_{eff}})$ yields a main sequence slope that is steeper than 4 indicating that, *along the main sequence, hotter stars are also larger*. This is shown more quantitatively in Table G.7. Some second order effects are also present[4], causing curvature in the main sequence.

There are many stars in the upper right region of the HR diagram as well. For the same temperature, these stars are much more luminous than their main sequence counterparts. By Eq. (4.13), this means that they must be much larger, hence the designation, 'giants'. A G2III star, for example (Table G.8) has about the same surface temperature as the Sun, yet is ten times larger. The internal structure of

[4]Stellar opacity (see Sect. 5.4.1) can affect the radius of the star, for example.

such stars is different from that of those on the main sequence, as discussed in Sect. 3.3.2. Notice, however, that some very strong conclusions can be drawn about stars by simply plotting their location on the HR diagram and referring to Eq. (4.13). It is from these and other observational starting points that we can actually piece together the evolution of stars.

4.1.5 Energy density and pressure in stars

The approximation that stars are close to black bodies also implies that the radiation field is close to (but not exactly) isotropic at any interior location (see Sect. 6.3.6 for a more accurate description of radiation within stars). Letting the intensity, I, be approximately equivalent to the mean intensity, J, for such a case, we can use Eq. (1.16) to write the energy density (erg cm^{-3}) as,

$$u = \frac{4\pi}{c} I = \frac{4\pi}{c} B = \frac{4\pi}{c} \mathcal{K} \mathrm{T}^4 = a \mathrm{T}^4 \tag{4.14}$$

where we have used Eq. (4.9) and the constant, $a = 7.57 \times 10^{-15}$erg cm^{-3} K^{-4}, is called the *radiation constant* whose accurate value is given in Table G. 2. Thus, the energy density is also a strong function of temperature.

The radiation pressure (dyn cm^{-2}) of such a gas is then given by (Eq. 1.22),

$$\mathrm{P}_{\mathrm{rad}} = \frac{1}{3} a \mathrm{T}^4 \tag{4.15}$$

This radiation pressure must be added to the particle pressure (Eq. 3.11), as described in Eq. (3.16), to determine the total pressure at any location within a star. For low mass stars, like our Sun, the radiation pressure is negligible with respect to the total. However, for hotter more massive stars, the radiation pressure, being a strong function of temperature, becomes dominant. For very high mass stars, for example, radiation pressure can become so important that it affects the stability of stars and contributes to mass loss of the outer layers (e.g. see the discussion of Wolf–Rayet stars in Sect. 5.1.2.2). The importance of radiation pressure for stellar dynamics will be discussed further in Sect. 5.4.2.

4.2 Grey bodies and planetary temperatures

Opaque bodies at a single surface temperature are black bodies if they absorb all incident radiation. However, in many cases, some fraction of the incident radiation will be reflected away. The fraction of light that is reflected is called the *albedo*, A,[5] of the

[5]From the latin, 'albus' for white.

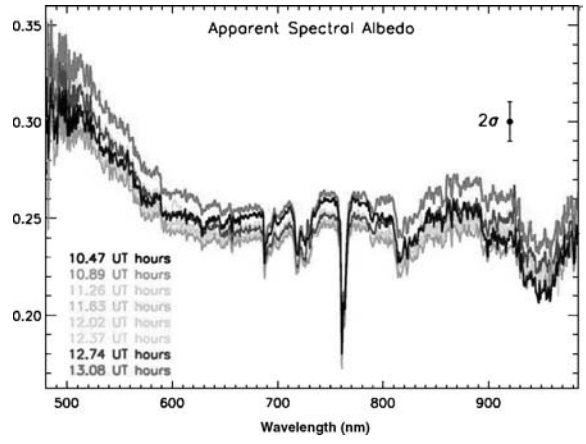


Figure 4.7. Plot of the Earth's albedo as a function of wavelength in the optical band, measured on 19 November 1993 at the times indicated. The albedo increases towards the blue (left) due to Rayleigh scattering (see Sect. 5.1.1.3) and a number of spectral features can be seen, including one at λ 760 nm due to O_2. These wavelength-dependent values are comparable to the Bond Albedo of $A = 0.306$ (Table G.4) which is the value applicable to all wavelengths. (Montañes-Rodriguez, P., *et al.*, 2005, ApJ, **629**, 1175. Reproduced by permission of the AAS)

object, values of which are given in Table G.4 for the planets. A surface that absorbs all incident radiation (a true black body) would have an albedo of 0 and a surface that reflects all incident radiation (such as a 'perfect' mirror) has an albedo of 1. We implicitly considered 'perfect absorption' and 'perfect reflection' previously when dealing with radiation pressure in Sect. 1.5. The introduction of the albedo now permits a quantification of intermediate cases. Objects that do not absorb all incident radiation are called *grey bodies*. The term, 'grey', is meant to imply that the absorption is independent of wavelength (i.e. 'grey' or 'colourless'). This is technically only true for opaque objects that reflect the same fraction of light at any wavelength, i.e. A is independent of λ, but the terminology also tends to be used for any object with $A>0$ (see Figure 4.7)[6].

Planets, asteroids, comets, moons (see Iapetus, Figure 4.8), and dust grains (see also Appendix D.3) are all grey bodies, as are normal opaque objects in every day life like the walls of a room, the surface of a human being, or the hound dog shown in Figure 4.9. *It is the reflected light that allows us to see the object*. Our eyes detect the portion of the reflected light from walls, dogs, or planets that falls into the visual band. *It is the absorbed light that heats the object*. The object will then radiate with a Planck spectrum dictated by its temperature.

[6]This is because, for many such objects, the albedo does not change drastically over the optical part of the spectrum, as Figure 4.7 shows. The albedo in the optical band is most important because this is where the Sun's Planck curve peaks and therefore is the band over which most of the energy is absorbed.

Figure 4.8. Image of Iapetus, a moon of Saturn (equatorial diameter = 1436 km), taken at 'half-moon' phase from the Cassini spacecraft in October, 2004, with a resolution of 7 km. Iapetus is unique in the Solar System because of the strong contrast in albedo between its two hemispheres. This constrast can be seen here as dark and bright sections on the illuminated side of the moon. (The narrow white streak in the dark region is a chain of icy mountains.) The dark section is as dark as coal ($A = 0.03 \rightarrow 0.05$, almost a black body) while the bright section is more typical of Saturn's other icy moons ($A = 0.5 \rightarrow 0.6$). The reason for this contrast is not yet understood. (Reproduced courtesy of NASA/JPL/Space Science Institute)

4.2.1 The equilibrium temperature of a grey body

Grey bodies will absorb a fraction, $1 - A$, of the light that falls on them and it is this portion that heats the object, bringing it to a temperature that remains constant with time, that is, an *equilibrium temperature*. For such an equilibrium temperature to be

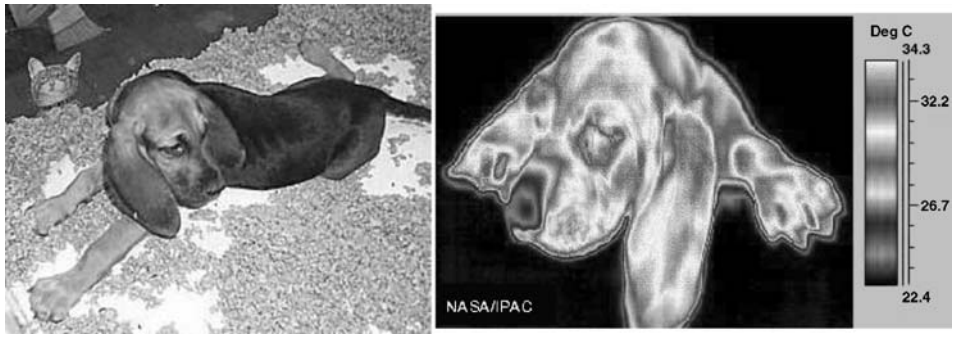


Figure 4.9. This hound dog is a 'grey body'. A fraction of the light that falls on the dog is absorbed and the remaining fraction is reflected. It is the reflected light in the optical part of the spectrum that allows us to see the dog normally as shown on the left. The absorbed light as well as the dog's internal heat generation must be balanced by its emission to maintain a constant temperature at any given position. The emitted spectrum will be a Planck curve which peaks in the IR for the range of surface temperatures shown in the false colour image on the right. (Reproduced by permission of SIRTF/IPAC) (See colour plate)

maintained, the rate at which the grey body radiates must balance the rate at which it absorbs energy. Since a fraction of the incident energy is reflected away, the temperature of a grey body will not come into equilibrium with the temperature of the radiation field but rather with a temperature that corresponds to the absorbed fraction of the incident radiation. The grey body then emits a Planck spectrum at its equilibrium temperature. Example 4.4 shows how the equilibrium temperature of an externally heated grey body can be determined. Emissions from grey bodies in our Solar System have Planck curves that peak in the infrared (Prob. 4.6). Dust that is detected in our Milky Way, such as in the Lagoon Nebula shown in Figure 3.13, as well as the dust that is observed in other galaxies such as in Arp 220 (Figure 4.10) also tend to be at temperatures resulting in $\lambda_{\max}$ in the IR or sub-mm.

There are two cases, however, in which a grey body may have a temperature that departs from the equilibrium temperature expected from absorbed external radiation.

The first is if the object has its own internal source of heat. For living, warm-blooded creatures, for example, chemical reactions turn food into an internal heat source.

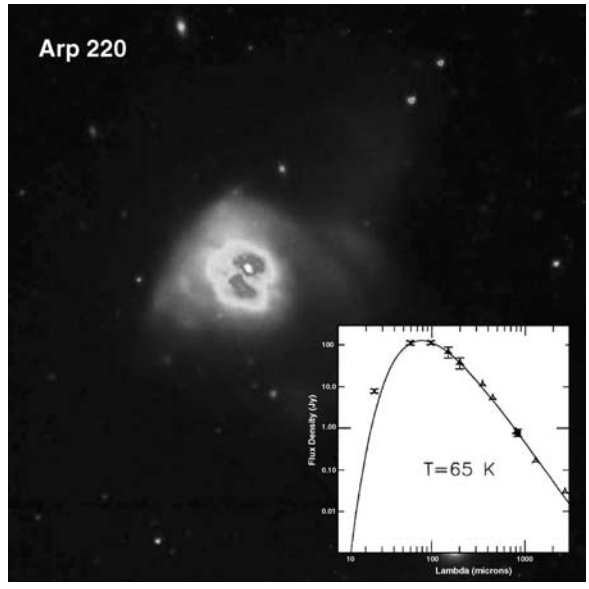


Figure 4.10. Image of the ultra-luminous infrared galaxy, Arp 220. This galaxy is undergoing an intense 'starburst', meaning that it is rapidly forming stars, in this case, at a rate of $100\,M_{\odot}yr^{-1}$. Thus, there are many hot young stars in this galaxy (see Sect. 3.3.4) whose UV photons heat the abundant dust that permeates the system. (An active nucleus is also present.) Most of the galaxy's luminosity is then re-radiated in the IR by dust grains, giving the very high value of $L_{8\rightarrow1000\,\mu m} = 1.6 \times 10^{12}\,L_{\odot}$. Arp 220 is believed to be the remnant of two galaxies that have merged. The inset shows a Planck spectrum that has been fitted to the dust emission from the galaxy. The spectrum is adjusted slightly from a true Planck spectrum to take into account grain properties. [Reproduced by permission of David Sanders. (Adapted from Lisenfeld, U., Isaak, K.G., and Hills, R., 2000, MNRAS, **312**, 433)]

For the outer gas giant planets, Jupiter, Saturn, and Neptune, there are significant internal heat sources[7], possibly remnant heat from their formation (Jupiter) or heat released by the precipitation and slow sinking of helium droplets towards the core (Saturn). Internal heat can also be generated via tidal stresses, such as those that occur within Jupiter's close moon, Io, due to the gravitational forces exerted on it by Jupiter. For the Earth, an internal heat source is the radioactive decay of elements in the crust. However, this source, with a flux of $S_{\text{int}} = 0.06\,\text{W m}^{-2}$ (Ref. [174]), is insignificant in comparison to the incident Solar flux, i.e. that of the Solar constant, $S_{\odot} = 1367\,\text{W m}^{-2}$ (see Sect. 1.2, Figure 1.6).

The second is if the heat from the planet that would normally be radiated away freely is somehow trapped. This is what happens from the *Greenhouse Effect*. Planets such as the Earth and Venus are shrouded in an atmosphere that acts like a blanket, retaining heat within it. This occurs because most incident Solar radiation is at optical wavelengths (see Figure 4.5) whereas the emitted radiation from the surface is in the infrared. The atmosphere is not as transparent in the infrared as in the optical, as the atmospheric transparency curve for the Earth (Figure 2.2) shows. A photon emitted from the surface may be scattered in the atmosphere or re-absorbed and re-emitted, possibly to be absorbed by the surface again. Thus, the atmosphere has the effect of retaining photons for a longer period of time, contributing an extra source of heating. The Greenhouse Effect provides an important and necessary heat source to the Earth, contributing $+33^{\circ}$C to its surface temperature, as Example 4.4 notes. 'Global warming' refers to the much smaller changes, of order less than a degree Celsius, that the Earth has been experiencing over the last century, as Figure 4.12 illustrates.

Example 4.4

Estimate the equilibrium temperature of the Earth. Compare this to the Earth's globally averaged mean temperature of $+14^{\circ}C$ (287 K).

The Earth absorbs radiation over only the cross-sectional area, $\pi {R_{\oplus}}^2$, that faces the Sun (see Figure 4.11 and Example 1.1). The rate at which energy is absorbed is therefore $L_{\text{abs}} = (1 - A)\pi {R_{\oplus}}^2 f$, where f is the flux of the Sun at the distance of the Earth (i.e. the Solar Constant) and A is the Earth's albedo. Expressing the Solar flux in terms of luminosity, $L_{\odot} = 4\pi r^2 f$ (Eq. 1.9), where r is the Earth–Sun distance, the above equation becomes,

$$L_{\text{abs}} = \frac{(1 - A){R_{\oplus}}^2 L_{\odot}}{4r^2} \tag{4.16}$$

The Earth can be approximated as a 'fast rotator', which means that the energy absorbed on the daytime side is quickly equalized over the nighttime surface. This is

[7]These outer planets radiate significantly more energy (of order a factor of 2) than they are absorbing from the Sun. By contrast, the temperature of the planet, Uranus can be explained solely by external Solar heating.

Figure 4.11. A planet receiving light from the Sun intercepts it over its daylight half (left). Being a grey body, a fraction, A (the albedo), of this light is reflected away and a fraction, $1 - A$, is absorbed. Assuming that the absorbed energy is equalized quickly over the planet's surface, the planet then radiates the energy over its entire surface (right) at a rate that balances the energy input rate, thus maintaining an equilibrium temperature (see Example 4.4).

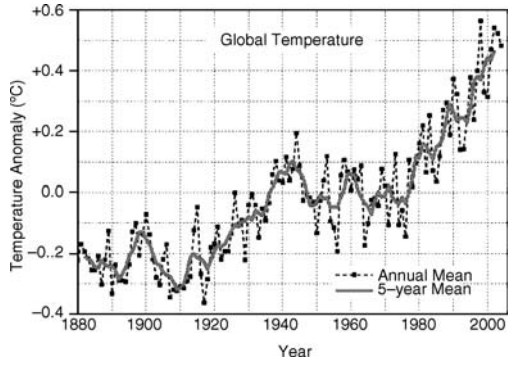


Figure 4.12. Globally averaged temperature anomaly of the Earth from 1880 to 2004, with respect to a zero point of 14°C that is an average over the period, 1951 to 1980 (Reproduced by permission of J. Hansen, from http://data.giss.nasa.gov/gistemp/). The increase in temperature correlates with an increase in carbon dioxide concentration in the atmosphere.

a reasonable approximation because day/night variations in surface temperature (of order 10 K) are much smaller than the temperature itself (287 K). The Earth, being a grey body, then emits a Planck spectrum at a (globally averaged) temperature, T. The rate of energy emission is given by the Stefan–Boltzmann law (Eq. 4.13)[8],

$$L_{\rm em} = 4\pi {R_\oplus}^2 \sigma {\rm T}^4 \tag{4.17}$$

For an equilibrium temperature, we require the absorption rate to equal the emission rate, so equating Eq. (4.16) to Eq. (4.17) and solving for temperature,

$${\rm T} = \left[\frac{(1-A)L_\odot}{16\pi\sigma r^2}\right]^{1/4} \tag{4.18}$$

[8]A multiplicative factor is actually required on the right hand side of this equation to account for the emissive properties of the surface material, but the correction is small for the purposes of this example: approximately 0.9 for solid surfaces and 0.65 for gaseous surfaces (Ref. [75]), leading to corrections to T (Eqn. 4.18) of 0.97 and 0.90, respectively.

Using $A = 0.306$ (Table G.4), $L_\odot = 3.845 \times 10^{33}\text{erg s}^{-1}$, $r = 1.496 \times 10^{13}$ cm (Table G.3), and $\sigma = 5.67 \times 10^{-5}\text{erg cm}^{-2}\,\text{K}^{-4}$ (Table G.2), we find $\text{T} = 254\text{K} = -19°\text{C}$. By Wien's Displacement Law (Eq. 4.8), the peak of the corresponding Planck curve would be in the infrared ($\lambda_{\max} = 11$ μm). Note that Eq. (4.18) is independent of the size of the Earth, $R_\oplus$. Thus, provided the above assumptions are valid, a planet of *any size* placed at the distance of the Earth would have the same equilibrium temperature if it has the same albedo. This equilibrium temperature is considerably colder than the Earth's measured globally averaged temperature of $\bar{\text{T}} = 287\text{K} = +14°\text{C}$ (Ref. [29]), the increase being due to the Greenhouse Effect. It is clear that, without the warming blanket of our atmosphere, the Earth would be a very cold, inhospitable place to live.

Eq. (4.18) in Example 4.4 can be used for any grey body in the Solar system whose illumination and radiation geometry are as shown in Figure 4.11. Changes in the geometry of a grey body will not make large changes in the results, however. For example, a hypothetical asteroid shaped like a cube, rather than a sphere, would have an equilibrium temperature that differs from that given by Eq. (4.18) by a factor of only 1.1. A 'slow rotator' – an object that cannot equalize its temperature between its day/night sides, but rather is hot on one side and cold on the other – will have a hot side temperature that is higher than that of Eq. (4.18) by a factor of 1.2. Therefore, Eq. 4.18 provides a good estimate of the temperature of a grey body that is heated only by the Sun.

If the appropriate external luminosity is used instead of $L_\odot$, Eq. (4.18) can also be used for a grey body in an external system such as dust in the ISM or other galaxies[9], dust rings around stars such as that shown in Figure 4.13 or, as discussed below, planets orbiting stars other than our Sun (*extrasolar planets*).

4.2.2 Direct detection of extrasolar planets

Almost all extrasolar planets (or *exoplanets*) have been discovered by observing the perturbation in the motion of the parent star due to the orbital motion of the planet around it (see Sect. 7.2.1.2). Thus, it is the *stellar emission* that is actually observed and the planet is *inferred* to be present by the dynamical effect it has on the parent star. It is now also possible to infer the presence of exoplanets in spatially unresolved systems via *gravitational microlensing*, a topic that will be further explored in Sect. 7.3.2. However, microlensing is a relatively rare event that requires long term monitoring of many stars. It is of particular interest, then, to see if the 'normal' Planck emission from a planet can be detected directly.

[9]Some modification is required, however, if the grey body is surrounded on many sides by a number of stars, such as would be the case for dust grains in the interstellar medium. Also, the smallest dust grains do not achieve an equilibrium temperature, but rather experience spikes in temperature due to the absorption of individual photons, followed by rapid cooling.

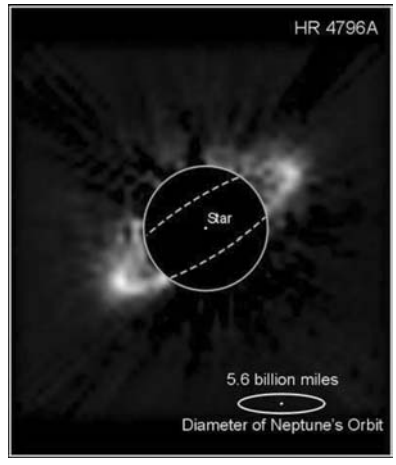


Figure 4.13. A dust disk observed around the star, HR 4796A, by the Hubble Space Telescope using its NICMOS instrument in the near infrared. The brilliant starlight has been blocked out by covering over a region shown by the circle in this picture, revealing the faint disk around it. The central star is located at the centre and dashed curves are interpolations as to where the disk continues through the blocked region. Neptune's orbit is drawn at the bottom right for comparison. (Reproduced by permission of G. Schneider)

There are two challenges to overcome. The first is that an entire planetary system is often unresolved, so the flux density of the planet(s) and star are received together in a single beam or pixel. For example, if a relatively close star at a distance 5 pc had a planet orbiting it at the same distance as Jupiter is to the Sun, then the planet and the star would be separated by only 1 arcsecond. The second challenge is that the star is much hotter (and also larger) than the planet, so the starlight swamps the emission from the planet itself. The total flux density received at any wavelength, for an unresolved system, is the sum of the flux densities from the planet and star together, but the perturbation (increase) in flux due to the planet is very small, even at the infrared wavelengths at which the planet shines brightest, as Example 4.5 shows. Even if a perturbation to the star's spectrum could be measured, it would not necessarily be clear that this perturbation is due to a planet. Perhaps the star has a disk of dust around it, for example, or perhaps the star's spectrum is unusual in some way. Fortunately, there are two classes of extrasolar planet that do lend themselves to direct detection: those for which the planet is very close to the parent star, and those for which the planet is very far away. In both cases, the planet must be fairly massive, of order several Jupiter masses rather than the mass of the Earth.

Such extrasolar planets that are close to the parent star are called *hot Jupiters*. These planets are as massive as Jupiter, but orbit their stars at a very small radius, of order 0.05 AU, and have revolution periods of only a few days. Due to their proximity to the star, these planets will be very hot and bright. In addition, their closeness to the star

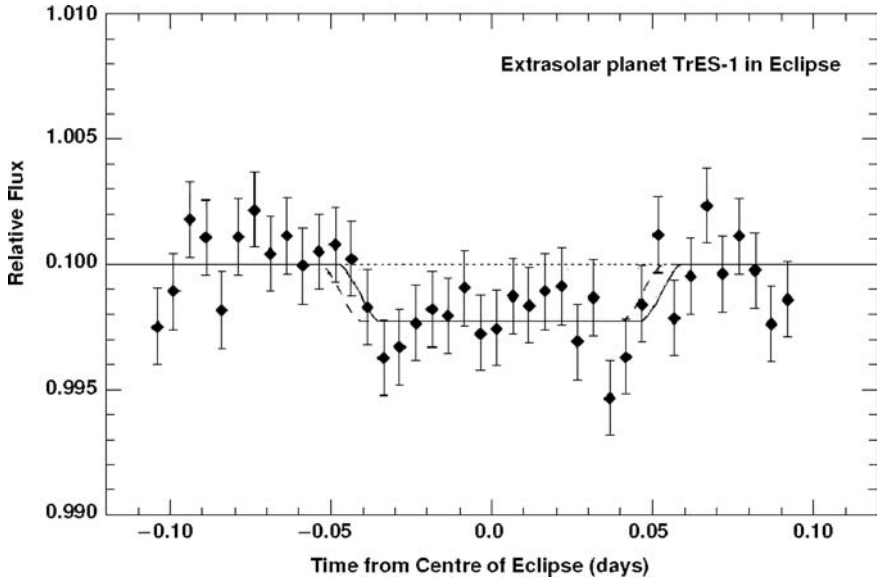


Figure 4.14. Light curve from the extrasolar planet system, TrES-1. At the centre of the plot, the flux is lower because the planet is behind the star. At the right and left, the flux includes both the planet and star. These observations were taken at $\lambda 8\,\mu$m with NASA's Spitzer orbiting telescope (Ref. [36]). They constitute one of the first two observations, published in 2005, in which photons from a planet have been directly and clearly detected (see also Ref. [45]). The planet in this system is one of a class of planets called hot 'Jupiters'. The star is of type, K0V and the planet is a distance, $r = 0.0394$ AU, from it with an albedo of $A = 0.31$ and mass, $M_p = 0.76 M_J$, where M_J is the mass of Jupiter. (Reproduced by permission of D. Charbonneau)

makes it much more likely that they will pass directly in front of or behind the star, in their orbit around it, as seen from our vantage point. The resulting eclipse can make a measurable difference in the light from the system. In such a case, it is clear that the slight change in flux is due to the planet passing in front of or behind the star and not to a dust disk which would always be visible. The first two detections of hot Jupiters in eclipse were made in the year, 2005. The *light curve* (a plot of varying emission with time) of one of these is shown in Figure 4.14.

The second class involves direct imaging. For planets that are very far from their parent star, it is now becoming possible to image the planet directly. The planet appears only as a weak 'pin-prick' of light that is sufficiently separate from the stellar image to be seen as distinct. Such imaging is now becoming possible as detectors improve and observing techniques, such as those that deal effectively with speckles (see Sect. 2.3.2), are put into practice. Two early observations made in 2004 and 2005 are of planetary candidates that are each a few Jupiter masses and at distances of 55 AU and 100 AU from their parent stars, respectively (Ref. [38], Ref. [110]). As telescopes and detectors continue to improve, such detections will soon become routine.

Example 4.5

A Sun-like star at a distance of $d = 5$ pc has an Earth-like planet orbiting 1 AU from it. Determine the excess flux density due to the planet at the wavelength of the peak of the planet's Planck curve. Assume that the planet has no atmosphere and that the system is unresolved.

A Sun-like star has a temperature of 5781 K (Table G.3) and, from Eq. (4.18), we know that an Earth-like planet without an atmosphere (no Greenhouse Effect) will have a temperature of 254 K (Example 4.4). By Wien's Displacement Law for the planet (Eq. 4.8) $\lambda_{\max} = 1.14 \times 10^{-3}$ cm. Using Eq. (4.2) at this wavelength, the star's specific intensity is, $B_s(5781) = 2.53 \times 10^{10}$erg s^{-1} cm^{-2} cm^{-1}sr^{-1} and the planet's specific intensity is $B_p(254) = 4.32 \times 10^{7}$erg s^{-1} cm^{-2} cm^{-1} sr^{-1}. The flux density of each is $f_\lambda = B_\lambda(\mathrm{T})\Omega$ (Eq. 1.13). The solid angle, $\Omega = \pi(R/d)^2$ (Eq. B.3 and B.1), where R is the radius of the object. Using the radius of the Sun (Table G.3) and the Earth (Table G.4), the star's solid angle is $\Omega_s = 6.39 \times 10^{-17}$ sr and the planet's is $\Omega_p = 5.37 \times 10^{-21}$ sr. Then the flux density of the star and planet are, respectively, $f_s = 1.62 \times 10^{-6}$erg s^{-1} cm^{-2} cm^{-1} and $f_p = 2.32 \times 10^{-13}$erg s^{-1} cm^{-2} cm^{-1}. The total flux density detected is $f_s = f_s + f_p \approx f_s$. The planet would therefore make a very tiny perturbation that is only 1.4×10^{-7} times the value of the star's flux density at a wavelength of $\lambda_{\max}$ for the planet. In order to detect such a tiny perturbation, the dynamic range (see Sect. 2.5) of the observations would have to be better than 7×10^6.

Problems

4.1 (a) Verify that Eq. (4.1) can be reproduced when Eq. (4.2) is substituted into Eq. (4.3). (i.e. show that $B_\nu(\mathrm{T}) = B_\lambda(\mathrm{T})(d\lambda/d\nu)$.

(b) Using Wien's Displacement Law, evaluate $\lambda_{\max}$ and $\nu_{\max}$ for a temperature of 6000 K and show that $\lambda_{\max} \neq c/\nu_{\max}$.

(c) Evaluate $B_\lambda(6000)$ at $\lambda = \lambda_{\max}$.

(d) Evaluate $B_\nu(6000)$ at the frequency corresponding to $\lambda_{\max}$. Convert the result to the same units as $B_\lambda(6000)$ and verify that the result is the same as in part (c).

4.2 (a) Write the Rayleigh–Jeans Law and Wien's Law in their wavelength-dependent forms. Will the criteria specified in Example 4.2 for departures from the full Planck curve be the same for the wavelength-dependent functions as for the frequency-dependent functions?

(b) Assuming that the emission can be well-approximated by a Planck curve, determine whether the Rayleigh–Jeans Law or Wien's Law may be used for the following conditions:

(i) an observation of the Sun at λ 912 Å;

(ii) an observation of a dense molecular cloud of temperature, T = 20 K at $\nu =$ 230 GHz;

(iii) an observation of the 2.7 K CMB at λ 21 cm.

4.3 (a) Determine and plot on the same graph, $B_\lambda(T)$ for the eye region (closest to the camera) and the nose of the hound dog shown in Figure 4.9, using the scale in the figure to estimate the temperature.

(b) Determine λ_{max} (μm) for the same eye and nose.

(c) Assuming that photons leaving both the eye and nose can escape from the surface in all directions outwards (over 2π sr), determine the flux density at the surface of the eye and nose.

(d) Is the energy loss rate greater through the eye or through the nose of the hound dog?

4.4 Repeat Prob. 4.3 but, instead of the eye and ear of a hound dog, compare a O8 V star to a K7 V star (Table G.7). Indicate in which part of the spectrum λ_{max} occurs for each.

4.5 In the caption to Figure 4.1, we note that a hot object is brighter than a cool object at *all* frequencies when both emit a Planck spectrum. Yet, observations of the ISM in the infrared show mostly cool dust emission while the hotter stellar emission is negligible in comparison. How can these two statements be reconciled? To answer this question, consider a cubic region of volume, 1pc^3 in the Galaxy some distance, D, away so that any object in the volume can be considered at the same distance. Assume that there is, at most, one star within this volume and that this star is equivalent to the Sun (same radius, R, and effective temperatue, T_{eff}). Assume that all dust grains are spherical, evenly distributed, and that no grain 'shadows' (is directly in front of) any other grain. For the dust, let the temperature, T = 100 K, radius, $a = 0.2\mu$m, density, $\rho = 2.5\, g\,\text{cm}^{-3}$, and the total dust mass in the volume is $2.65 \times 10^{-4}\,\text{M}_\odot$. Find the ratio of the stellar flux to the dust flux at IR (100 μm) and visible (550 nm) wavelengths, assuming both emit perfect planck spectra, and comment on the results.

4.6 Compute the temperature range over which an object would have a Planck curve peak, λ_{max}, in the infrared.

4.7 Assume, for simplicity, that a region of space contains one O8V star, one A0Ib star, 100 G5V stars and 1000 M2V stars. Refer to Tables G.7, G.8, G.9 and,

(a) determine the total stellar mass (units of $\text{M}_\odot$) in the region and indicate which type of star dominates the mass;

(b) determine the total stellar luminosity ($L_\odot$) in the region and indicate which type of star dominates the luminosity.

4.8 (a) Write an expression for the ratio of the flux of a star to the flux of the Sun, $f_\star/f_\odot$, in terms of the radii of these objects and their effective temperatures. Assume that the two are at the same distance and radiate as perfect black bodies.

(b) Using the results of (a), find the flux ratio for the neutron star, J0538+2817, whose radius is $R = 1.68\,\text{km}$ and temperature is $\text{T} = 2.12 \times 10^6\,\text{K}$. Determine $\lambda_{\max}$ for this neutron star and indicate in what part of the spectrum this occurs.

(c) Presently, thermal black body radiation has been measured from the surface of only a 'handful' of neutron stars (Ref. [84])[10]. Based on the results of part (b), explain why this might be the case.

4.9 Compute the equilibrium temperature of Mars. Compare it to the average measured temperature of -63°C and comment on whether Mars has its own internal heat source or experiences a Greenhouse Effect.

4.10 (a) Rewrite Eq. (4.18) for a grey body that is a *slow rotator*, i.e. determine T for the Sun-facing side of a spherical body when the temperature is negligible on the hemisphere opposite to the Sun.

(b) On Iapetus (see Figure 4.8), the bright and dark sections are often separated by no more than a single pixel of width, 7 km. If an astronaut were to take a stroll over such a distance at the bright/dark boundary, by how much would the surface temperature change? Which is the warm side? Assume that Iapetus is a slow rotator and its heat is supplied only by the Sun.

4.11 From the information given in Figure 4.14 for the extrasolar planet system, TrES-1, as well as Table G.7, find the following:

(a) the temperature, T_s, radius, R_s, and luminosity, L_s, of the star ;

(b) the temperature, T_p, of the planet;

(c) the ratio of flux densities of the planet to the star, $f_{\lambda\text{p}}/f_{\lambda\text{s}}$ at the observing wavelength of λ 8 μm;

(d) the ratio of specific intensities, $B_\lambda(\text{T}_\text{p})/B_\lambda(\text{T}_\text{s})$ at λ 8 μm;

(e) the ratio of solid angles, $\Omega_\text{p}/\Omega_\text{s}$;

(f) the radius of the planet, R_p. Express this result in terms of Jupiter radii, R_J.

4.12 (a) Derive an expression for the reflected flux, F_refl, of an extrasolar planet that intercepts light as shown in Figure 4.11a, and reflects it outwards over only the one hemisphere facing the star. The result should be expressed in terms of the albedo, A, the luminosity of the star, $L_\star$, and the distance of the planet to the star, r.

(b) Express Eq. (4.18) in terms of the emitted flux at the planet's surface, F_em, instead of its temperature.

(c) Determine the fraction of stellar flux that is reflected compared to that emitted thermally, $F_\text{refl}/F_\text{em}$. For what value of albedo will the reflected flux equal the emitted thermal flux?

[10]Neutron stars are usually detected as pulsars and it is the stronger *non-thermal radiation* (see Sect. 8.1), rather than their black body radiation, that is usually detected.

Appendix: derivation of the Planck Function

The derivation of the Planck Function (Ref. [76]) proceeds along the same principles as we have already seen for the Maxwell–Boltzmann velocity distribution (Sect. 3.4.1), the Boltzmann Equation (Sect. 3.4.5), and the Saha Equation (Sect. 3.4.6). That is, the probability of finding a particle in any given excitation state depends on the statistical weight (the number of possible ways the particle can fit into the state) as well as the Boltzmann factor, $e^{-E/(k\mathrm{T})}$. There are differences between this development and the previous examples, however. One is that we are now considering 'particles' which are *photons*. Photons are *bosons* (that is, particles with integer *spin*)[11] that have two possible spin states $(+1, -1)$ corresponding to two opposite circular polarizations. Photons do not have to obey the Pauli Exclusion Principle (see Appendix C.2) so the limit to the number of possible states is set only by the *Heisenberg Uncertainty Principle* (HUP, see below). Note also that, since photons are not bound particles like electrons in an atom, we must consider the number of states in free space, or the number of states per unit volume. Another difference is that the Boltzmann factor gives a relative probability, for example, the number of particles in an energy state, a, compared to an energy state, b. However, we now do not wish to compare the number of photons in two particular states, but rather to determine the mean number of photons or the mean energy in any state.

4.A.1 The statistical weight

To consider the number of energy states that are possible in a given volume, we can equally ask how many momentum states per unit volume are possible, since $E = cp$ (Table I.1), where p is the photon momentum and c is the speed of light. The answer is dictated by the HUP which, for distinguishable states in three dimensions, can be expressed as,

$$\delta x\, \delta p_{\mathrm{x}}\, \delta y\, \delta p_{\mathrm{y}}\, \delta z\, \delta p_{\mathrm{z}} \geq h^3 \tag{4.A.1}$$

The particle can be thought of as existing in a six-dimensional *phase space* in which *phase cells* must be separated from each other by a cell of minimum size, (h^3)[12]. The number of possible phase cells is then given by the ratio,

$$\frac{\delta x\, \delta p_{\mathrm{x}}\, \delta y\, \delta p_{\mathrm{y}}\, \delta z\, \delta p_{\mathrm{z}}}{h^3} \tag{4.A.2}$$

[11] Particles such as protons, neutrons, and electrons have half-integer spin.

[12] The expression, $dx\, dp_{\mathrm{x}} \geq \frac{h}{2\pi}$, in Table I.1 is a statement of the uncertainty in the simultaneous measurement of position and momentum in one dimension. The expression given here uses the larger value, h, since this refers to how far apart the phase cells must be to remain distinguishable.

Note that $\delta x\,\delta y\,\delta z$ represents a volume of real space and $\delta p_x\,\delta p_y\,\delta p_z$ represents a 'volume' of momentum-space.

As was discussed for the Maxwell–Boltzmann velocity distribution (Sect. 3.4.1), the momentum vectors of a particle can point in any direction so that any momentum vector of magnitude between p and $p + dp$, will end somewhere on a spherical shell whose volume in momentum-space is $4\pi p^2 dp$. Thus, for this geometry, the number of possible states (the statistical weight) is,

$$g(p)dp = \frac{V\,8\pi p^2\,dp}{h^3} \tag{4.A.3}$$

where the additional factor of 2 is due to the fact that, for photons, the two different spin states represent independent states.

Since we wish to derive a radiative spectrum, it is more useful to express this in terms of frequency. Recalling (Table I.1) that $p = h\nu/c (dp = h\,d\nu/c)$, we finally have the number of momentum (and therefore energy) states per unit volume,

$$\frac{g(\nu)d\nu}{V} = \frac{8\pi\nu^2 d\nu}{c^3} \tag{4.A.4}$$

4.A.2 The mean energy per state

Now that we know the number of possible independent energy states per unit volume for a photon in free space, we need to compute the number of photons we expect to be in any given energy state.

Alhough, in principle, there is no limit to the number of photons per energy state, the relative *probability* of finding n photons of energy, $h\,\nu$, in a cell of frequency, ν, is given by the Boltzmann factor, $e^{-(nh\nu/k\mathrm{T})}$[13]. As indicated above, we do not want the relative probability but rather the absolute probability, P_r, which is the relative probability divided by the sum of all possible relative probabilities (similar to the computation, in Sect. 3.4.5, of the partition function),

$$P_r = \frac{e^{-(nh\nu/k\mathrm{T})}}{\sum_n e^{-(nh\nu/k\mathrm{T})}} \tag{4.A.5}$$

The mean number of photons per cell is then,

$$<N_\nu> = \frac{\sum_n n e^{-(nh\nu/k\mathrm{T})}}{\sum_n e^{-(nh\nu/k\mathrm{T})}} \tag{4.A.6}$$

[13]In fact, there is an additional $\frac{1}{2}h\nu$ of energy in this factor that has been omitted here because it does not change the development nor does it involve photons. It gives the energy of the vacuum.

We now make the substitution, $y = (h\nu)/(k\mathrm{T})$ and note that the *denominator* can be expressed as,

$$\sum_n e^{-ny} = 1 + e^{-1y} + e^{-2y} + e^{-3y} + \ldots = (1 - e^{-y})^{-1} \tag{4.A.7}$$

where we have used the binomial expansion of Eq. (A.2) using $a = 1, x = -e^{-y}$, and $p = -1$. For the *numerator*, we need a sequence of expansions and factorizations,

$$\begin{aligned}
\sum_n ne^{-ny} &= 1e^{-1y} + 2e^{-2y} + 3e^{-3y} + \ldots \\
&= e^{-1y} + (e^{-2y} + e^{-2y}) + (e^{-3y} + e^{-3y} + e^{-3y}) + \ldots \\
&= e^{-1y}\,(1 + e^{-1y} + e^{-2y} + e^{-3y} + \ldots) \\
&+ e^{-2y}\,(1 + e^{-1y} + e^{-2y} + e^{-3y} + \ldots) \\
&+ e^{-3y}(1 + e^{-1y} + e^{-2y} + e^{-3y} + \ldots) + \ldots \\
&= (1 - e^{-y})^{-1}(e^{-1y} + e^{-2y} + e^{-3y} + \ldots) \\
&= (1 - e^{-y})^{-1}(e^{-1y})(1 + e^{-y} + e^{-2y} + \ldots) \\
&= \frac{e^{-y}}{(1 - e^{-y})^2}
\end{aligned} \tag{4.A.8}$$

Substituting Eqs. (4.A.7) and (4.A.8) into Eq. (4.A.6) yields,

$$< N_\nu >= \frac{e^{-y}}{(1 - e^{-y})} \tag{4.A.9}$$

Substituting back for y, dividing numerator and denominator by $e^{-h\nu/k\mathrm{T}}$, and multiplying by the energy, $h\nu$, gives the mean energy per cell,

$$< E_\nu >= \frac{h\nu}{(e^{h\nu/k\mathrm{T}} - 1)} \tag{4.A.10}$$

4.A.3 The specific energy density and specific intensity

The photon energy per unit volume per unit frequency, or *specific energy density* (Sect. 1.4), u_ν, is the mean energy per state (Eq. 4.A.10) times the number of states per unit volume (Eq. 4.A.4),

$$u_\nu d\nu = \frac{8\pi h\nu^3}{c^3}\frac{1}{(e^{h\nu/k\mathrm{T}} - 1)}d\nu \tag{4.A.11}$$

Since the energy density, u_ν, is related to the specific intensity, I_ν, via,

$$u_\nu = \frac{4\pi}{c} I_\nu \tag{4.A.12}$$

(Eq. 1.16 where, for isotropic radiation, I_ν and J_ν are equal), and *defining* the specific intensity of a black body to be $B_\nu(\mathrm{T})$, we finally obtain,

$$B_\nu(\mathrm{T}) = \frac{2h\nu^3}{c^2} \frac{1}{(e^{h\nu/k\mathrm{T}} - 1)} \tag{4.A.13}$$

which is Planck's law. The wavelength form is found by letting $B_\nu d\nu = B_\lambda d\lambda$ (see Eq. 1.5), and is,

$$B_\lambda(\mathrm{T}) = \frac{2hc^2}{\lambda^5} \frac{1}{(e^{hc/\lambda k\mathrm{T}} - 1)} \tag{4.A.14}$$

PART III
The Signal Perturbed

An astronomical signal does not simply emerge from an object and travel through a vacuum unperturbed. Rather, it may travel partially through the object that emitted it or through interstellar clouds that could alter or extinguish it. It travels through space which is itself expanding, and could pass near massive objects that bend its path. Thus, we cannot hope to understand an astronomical object unless we first understand how its signal may have become corrupted en route.

There is another way of looking at this, however. Every time the signal is changed, it may be possible to learn something about the object or medium that produces the change. Thus, an alteration in the signal is not always an unwanted nuisance, as was the case for telescopes and (for astronomers at least) the atmosphere. Instead, the astronomical signal itself might simply be a useful background source that probes the space and matter through which it travels, like a flashlight illuminating the intervening material for us. If the intervening matter neither emitted its own radiation nor had any affect on a background signal (or nearby objects), it would be impossible to learn anything about it.

Thus, we now focus more intently on astrophysical processes that can alter a signal and what insights they might provide either about the signal itself or the altering medium. We will first consider interactions in which the signal has contact with matter (the interaction of light with matter) and then those interactions that are 'non-touching' (the interaction of light with space).

5

The Interaction of Light with Matter

> You see the motes all dancing, as the sun Streams through the shutters into a dark room. . . . From this you can deduce that on a scale Oh, infinitely smaller, beyond your sight, Similar turbulence whirls.
>
> –the Roman poet, Lucretius, 1st Century B. C. (Ref. [102])

The interaction of light with matter spans the boundaries of classical and quantum physics, and of microscopic and macroscopic physics. One approach is to consider how an individual photon or a wavefront might interact with an individual particle. Another is to consider many particles together as representing a smooth medium that collectively alters the intensity and/or the trajectory of an incoming wavefront. Both approaches (and their variants) are valid and, if both exist, the resulting mathematical description of the physical phenomenon must agree. In this chapter, we will consider the interaction of light with matter in some detail, visiting each method along the way.

There are many types of matter that a photon could encounter in its flight across the cosmos – hot ionized hydrogen gas, a cold cloud containing complex molecules, a dust grain or even a planet. Such interactions with matter may produce a change in the photon's path and/or its energy. Although there are many types of matter, interactions with individual particles fall into two categories, *scattering* and *absorption.* In the case of absorption, a photon will disappear completely, its energy going into some other form such as the thermal energy of the gas. For scattering, on the other hand, a photon emerges from the interaction, though not necessarily at the same energy. The combined effects of scattering and absorption, both of which remove photons from the line of sight, are referred to as *opacity* (usually used for gases) or *extinction* (for dust).

The importance of an interaction with a particle depends on the cross-section presented to the photon by the particle (e.g. Example 3.5) as well as the number

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

density of particles that the incoming light beam encounters as it travels through a cloud. Therefore, in the next few sections, we focus on the *scattering cross-section*, σ_s, and the *absorption cross-section*, σ_a (both in units of cm^2) for the interaction of interest. The dependence on particle density is considered later in Sect. 5.4.1. The special case of dust grains is presented in Sect. 5.5.

5.1 The photon redirected – scattering

Scattering is a process by which the direction of an incoming photon changes as a result of an interaction. As we will see, the interpretation of an interaction as 'scattering' rather than 'absorption' is sometimes unclear. This is because, at an atomic or subatomic level, scattering involves the absorption and then re-emission of a photon. However, scattering implies that the re-emission occurs 'immediately', meaning on a short enough timescale that no other processes (e.g. collisions between particles) could interfere and remove the energy supplied by the incident photon. Scattering can be represented by the process: photon → matter → photon.

With a sufficient number of interactions, scattering tends to randomize the direction of the photons. Thus, when an observer looks back along a line of sight, the signal is attenuated (weaker) because many photons will be scattered into directions other than the line of sight, as shown in Figure 5.1a. However, it is also possible for scattering to be preferentially into a specific direction (symmetrically forwards and rearwards, for example) depending on the properties of the scatterers and the wavelength of the radiation. *Reflection*, for example, is just a special case of backwards-directed scattering, or 'back-scattering'.

Depending on the geometry, scattering can also redirect light *into* the line of sight. One example is when a dusty cloud surrounds a star. Photons that are emitted in directions away from the line of sight will be scattered by the dust and a fraction of them will scatter in a direction towards the observer (Figure 5.2.a). A scattering dusty cloud around a star is called a *reflection nebula*, an example of which is shown in Figure 5.3.

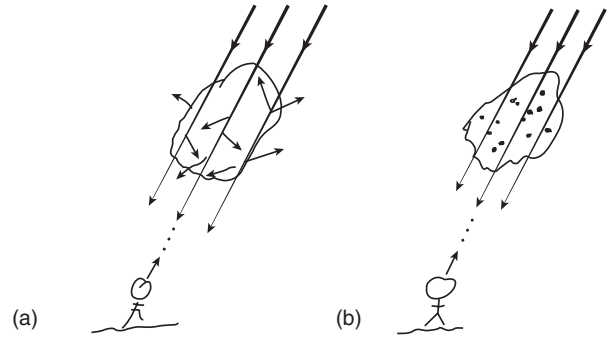


Figure 5.1. Illustration of **(a)** random scattering and **(b)** absorption along a line of sight. The observer is meant to be far from the source as well as the scattering medium in this diagram. Darker arrows convey a more intense signal.

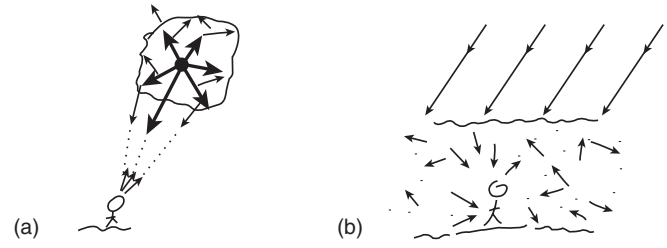


Figure 5.2. Two examples of scattering into the line of sight. **(a)** Dust particles surrounding a star scatter its light randomly, but some of these photons will be directed towards the observer. Thus, the observer sees a faint glow around the star, much like the glow around a streetlight on a misty night. **(b)** When the observer is within a randomly scattering medium, light can enter his eye from all directions even though the original source of light is in one direction only.

A more spectacular example of scattering into the line of sight occurs when there is a sudden increase in the brightness of an object, such as when a supernova (Sect. 3.3.3) or Gamma Ray Burst (Sect. 5.1.2.2) goes off or when a star experiences a bright outburst as can occur in certain types of variable stars. The light, travelling at c, scatters from surrounding dust clouds at specific times after the event dependent on the distance to the cloud. The result is called a *light echo*, an example of which is shown in Figure 5.4. A more familiar example of light scattering into the line of sight occurs when the observer is within the scattering medium itself, as is the case for the Earth's atmosphere. Our sky looks bright even in directions away from the direction of the Sun in the sky, because of such scattering (Figure 5.2.b).

The probability that a photon will undergo a scattering interaction is proportional to the scattering cross-section, σ_s, a quantity that must be derived for each type

Figure 5.3. The reflection nebula around the bright star, Merope, one of the stars in the Pleiades star cluster. The cross on the star is an artifact of the telescope and the other stars in the image are in the background or foreground. (Reproduced courtesy of Yuugi Kitahara, Astronomy Picture of the Day, March 1, 1999)

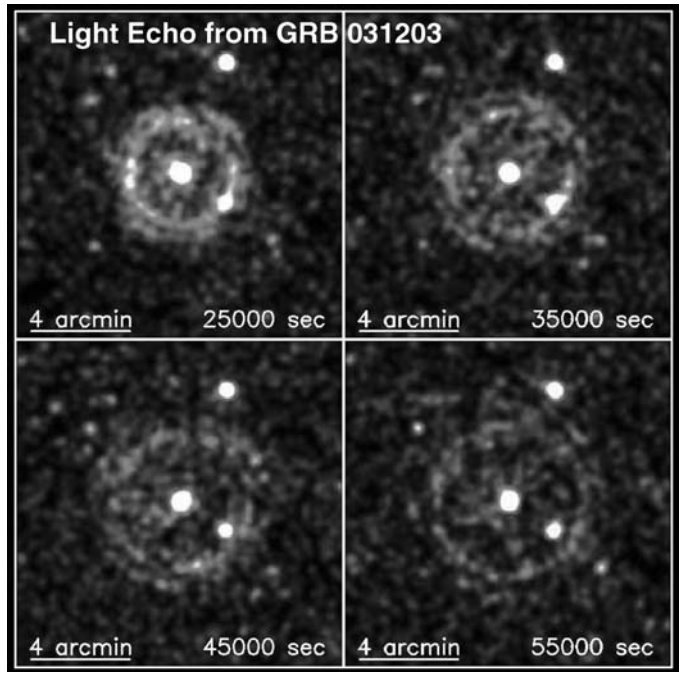


Figure 5.4. The X-ray 'afterglow' of the Gamma-Ray Burst source, GRB 031203, is seen at the centre of these images. The GRB was detected in December/2003 and is in a distant galaxy. The four X-ray images, taken from the XMM-Newton satellite, span the interval from 25 000 s (6.9 h) to 55 000 s (15.3 h) after the GRB was first detected. The rings are caused by X-rays scattering from different dust screens in our own Milky Way that are illuminated sequentially. The larger rings are seen at later times because it takes longer for light to travel from the GRB over the large angle and then to the observer, in comparison to light travelling closer to the GRB line of sight. The images give the appearance of an expanding ring, but are only produced by dust reflections. This is the first observation of a light echo seen at X-ray wavelengths. (Reproduced by permission of Simon Vaughan, Univ. Leicester)

of scatterering process. Since a particle may present a different effective scattering cross-section to photons of different wavelengths, this quantity is often wavelength-dependent. Several scattering cross-sections and further details of some of the scattering processes that are important in astrophysics are given in Appendix D. We provide a summary of them in the following subsections.

An important observational consequence of scattering, as explained in Appendix D and illustrated in Figures D.1 and D.6, is that the scattered emission, depending on viewing geometry, is *polarized*. The degree of polarization, $D_{\rm p}$ (introduced in Sect. 1.7), is the fraction of the light that is polarized. If emission is observed at some position, (x, y), on the sky and some fraction, $D_{\rm p}$, is polarized, then this fraction of the total intensity *may* have been generated at a different location and then scattered by particles at (x, y) into the line of sight. Some care must be exercised in this interpretation,

however, because, as indicated in Sect. 1.7, polarized emission can also result from certain processes intrinsic to the energy generation mechanism. Generally, other available information about the source is also considered in order to decide whether or not the polarization is due to scattering[1].

5.1.1 Elastic scattering

If there is no change in the energy of the photon as a result of the interaction, then the scattering is said to be *elastic*, analogous to a small particle bouncing off a wall without losing kinetic energy. If light is treated as a wave, then such interactions are called *coherent*, meaning that the outgoing energy, and therefore wavelength (recall $E = hc/\lambda$) has not changed from the incoming one. Some examples follow.

5.1.1.1 Thomson scattering from free electrons

Thomson scattering (Appendix D.1.1) is the elastic scattering of light from free electrons in an ionized or partially ionized gas. The scattering is elastic provided the energy of the incoming photon is much less than the rest mass energy of the electron, $E_{\rm ph} \ll m_e c^2$. Otherwise, some energy would be imparted to the electron (see Sect. 5.1.2.2 for that case). Thomson scattering is a very common process in astrophysics and occurs wherever ionized gas is found[2], from regions as diverse as stellar interiors (see Example 3.5) to the intergalactic medium.

A characteristic of Thomson scattering is that the scattering cross-section is independent of wavelength (i.e. it is 'grey', see Eq. D.2). This means that a radio photon is equally as likely to be scattered as an optical or UV photon. Thus, if the degree of polarization, $D_{\rm p}$, is independent of wavelength, it is evidence that Thomson scattering may be occurring. This appears to be the case for the nuclear region of the galaxy, NGC 1068 (see Figure 5.5). This galaxy's *active galactic nucleus* (AGN) is obscured from direct view, but free electrons in the dense ionized gas around the nucleus scatter AGN photons into the line of sight. Moreover, since the scattering is independent of λ, the scattered spectrum looks like the spectrum of the AGN. Thus the ionized gas acts like a mirror that allows us to see into the AGN itself. This is a good example of how a process that might be thought of as 'corrupting' the signal has actually turned into an indispensable tool for studying one of the most powerful types of objects in the Universe (see Figure 1.4 for another example of an AGN).

[1] Examples are the nature and waveband of the spectrum. For example, if polarization is seen at radio wavelengths, some or all of the emission might be due to synchrotron radiation (Sect. 8.5).

[2] Wherever there is ionized gas there will also be free protons. However, the proton scattering cross-section is negligible in comparison to the electron scattering cross-section (see Appendix D.1.1 and Eq. D.2).

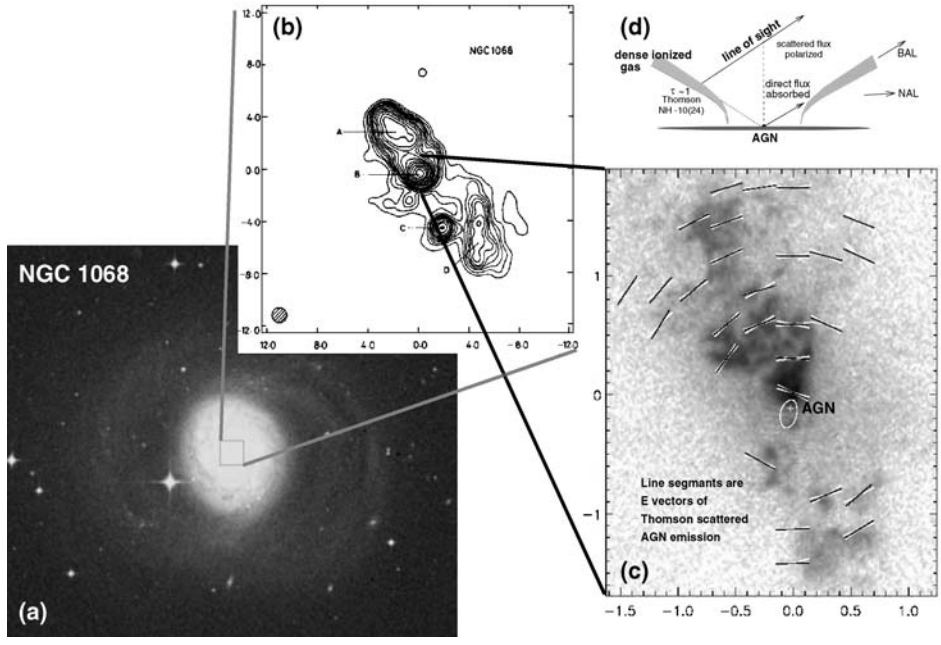


Figure 5.5. **(a)** The galaxy, NGC 1068, belongs to a class of galaxy that has a very active galactic nucleus (AGN), most likely due to a supermassive black hole. (Reproduced courtesy of Pal. Obs. DSS). **(b)** Highly energetic activity around the black hole creates outflow which is collimated (with the help of a circumnuclear disk) into a bipolar radio jet, shown here at $\lambda 73$ cm (Ref. [120]). (Reproduced courtesy of the NASA/IPAC Extragalactic Database). **(c)** A high resolution UV image is shown here in grey scale (Ref [88]). The AGN is marked, and the vectors show the orientation of the electric field of *Thomson scattered UV emission from the AGN*. The fraction of the emission at any position that is polarized is actually scattered AGN emission. The measured degree of polarization, D_p, ranges from 10 per cent to 44 per cent (Eq. 1.34), depending on position. An image at another waveband should show the same value of D_p at any given position in the case of Thomson scattering. (Reproduced by permission of M. Kishimoto). **(d)** A possible model for Thomson scattering of AGN emission (Ref. [58]). The AGN itself is obscured from direct view but electrons in the ionized region scatter the emission into the line of sight. (Reproduced by permission of M. Elvis)

5.1.1.2 Resonance (line or bound–bound) scattering

When the frequency of incoming light matches or is close to the frequency of an energy transition of a bound electron in an atom or molecule, the electron responds strongly and in *resonance* with the forcing of the wave's oscillating electromagnetic field, as outlined in Appendix D.1.2. This is called *resonance scattering* and it is coherent when the bound electron is excited to an upper state by absorbing the photon and then de-excites to the initial level again, re-emitting a photon at an equivalent energy[3] in

[3]The emitted photon may have a slightly different energy since the emission probability is distributed in frequency according to the Lorentz profile (Figure D.2). Other effects will broaden the line even more (see Sect. 9.3) so coherence is not perfect but the shift in energy is small in comparison to the energy of the line itself.

another direction. As with Thomson scattering, resonance scattering results in a polarized signal (Ref. [154]).

An example of a resonance scattering line is the Ly α line (a transition between quantum number n = 1 and n = 2) of hydrogen. This is a UV transition at a rest wavelength of $\lambda 122$ nm with an upper state lifetime of only 10^{-9} s (Table C.1). Since this is a much shorter timescale than the time between collisions under most interstellar circumstances (see Example 5.1), incident UV radiation near the Lyα wavelength is strongly scattered in this line.

Example 5.1

A Ly α photon travels through a pure hydrogen gas of density, $n = 10^3\,cm^{-3}$ and temperature, $T = 10^4$ K. If the gas is mostly ionized and the photon excites an upwards Ly α transition in one of the small fraction of particles that are neutral and in the ground state, will the outcome be scattering or absorption? Consider the permitted $2p \rightarrow 1s$ transition only.

Once the Ly α photon has excited the atom, the electron will spend a time, $\tau_{2,1} = \frac{1}{A_{2,1}} = 1.6 \times 10^{-9}$ s (Eq. D.22, Table C.1 plus notes to table) in the energy level, n = 2 before spontaneously de-exciting to the ground state, n = 1, again. In the mean time, there are collisions between particles and there is a possibility that de-excitation will occur via a collision with a free electron[4] or that collisional excitation will occur to higher levels. Considering only downwards permitted transitions, from Table 3.2 the relevant collision rate coefficient for de-excitation of this transition at the given temperature is $\gamma_{Ly\alpha(2p\rightarrow 1s)} = 6.8 \times 10^{-9}\,\mathrm{cm^3\,s^{-1}}$ so the timescale between de-exciting collisions is (Eq. 3.19) $\bar{t} = 1/[(10^3)(6.8 \times 10^{-9})] = 1.5 \times 10^5$ s. Since $\tau_{2,1} \ll \bar{t}$, the de-excitation will occur spontaneously. Similarly, it can be shown that the downwards collisional timescale via the $2s \rightarrow 1s$ transition, as well as upwards collisional timescales are longer than the spontaneous de-excitation timescale under the conditions described (see also Sect. 8.4). Once spontaneous de-excitation occurs, the re-emitted Ly α photon (after having travelled a mean free path) can excite another particle, followed by spontaneous de-excitation again, and so on. Thus, the photon will be scattered[5].

The Ly α photon scattering cross-section, as developed for resonance scattering in Appendix D.1.4, is given in Table 5.1. However, it is important to note that thermal motions of many atoms in a gas (see Sect. 9.3) will result in a line that is much wider in frequency and has a lower peak scattering cross-section than a line from a single isolated atom at rest. Therefore, a sample cross-section for 10^4 K gas is also given in Table 5.1. Even when one considers a more realistic line width, it is

[4]Collisions with protons are less likely since protons are massive and slow moving in comparison to electrons.
[5]Note that we have neglected the possibility of photo-absorption from the n = 2 level to a higher level, but this will only be important in a strong radiation field.

Table 5.1. Sample photon interaction cross-sections[a]

Type	Description	Wavelength or energy	Cross-section (cm^2)
σ_{T}[b]	Thomson scattering	$\ll 0.51$ MeV	6.65×10^{-25}
σ_{K-N}[c]	Compton scattering	0.51 MeV	2.86×10^{-25}
		5.1 MeV	8.16×10^{-26}
σ_{R}[d]	Rayleigh scattering (N_2)	532 nm	5.10×10^{-27}
	(CO)	532 nm	6.19×10^{-27}
	(CO_2)	532 nm	12.4×10^{-27}
	(CH_4)	532 nm	12.47×10^{-27}
σ_{b-b}[e]	Ly α (natural)[f]	121.567 nm	7.1×10^{-11}
	Ly α (10^4 K)[g]	121.567 nm	5.0×10^{-14}
$\sigma_{HI \rightarrow HII}$[h]	H ionization	13.6 eV	6.3×10^{-18}
σ_{f-f}[i]	free-free absorption	21 cm	2.8×10^{-27}

[a]Cross-sections apply to a single scattering event from a single particle.
[b]Thomson cross-section (Eq. D.2).
[c]Klein–Nishina cross-section for Compton scattering (see Figure D.4).
[d]Rayleigh scattering cross-section for a temperature of 15°C and pressure of 101 325 Pa (Ref. [157]).
[e]Resonance, bound–bound scattering from the line centre.
[f]From the natural line shape using Eq. (D.12) with data from Table C.1. Note that only the permitted transitions have been included (see notes to Table C.1).
[g]As in the previous row but assuming that the line is Doppler broadened (see Sect. 9.3) at the temperature indicated (Ref. [144]).
[h]Photoionization cross-section from the ground state (Eq. C.9).
[i]Free–free absorption cross-section for the conditions: $n_e = 0.1\,cm^{-3}$ and $T = 10^4$ K. The cross-section will vary with these quantities and also decreases with increasing frequency (Ref. [160]).

clear that this scattering cross-section is quite high in comparison to other values in the table.

HII regions (e.g. Figures 3.13, 8.4) are good examples of objects in which Ly α resonance scattering is important. In HII regions, photons at the wavelength of Ly α are numerous, both from Ly α recombination line emission as well as the continuum of the central ionizing star. As explained in Sect. 3.4.7, an HII region is almost entirely ionized but, at any time, recombinations are occurring and some atoms will have electrons in the ground state. Therefore, once a Ly α photon is emitted, it will be reabsorbed by other atoms, exciting the electron into the state, $n = 2$, followed by re-emission to the ground state again, and so on. There will be many such scatterings within the nebula because, not only is the resonance scattering cross-section large (Table 5.1), but also the densities of HII regions are low, minimizing the possibility that a collision could interfere with the scattering process. Even though an HII region has a very low density in comparison to the interior of a star (see Table 3.1), Ly α photons in an HII region take a random walk out of the nebula much like we saw for the interior of the Sun (see Sect. 3.4.4). Because Ly α photons are scattered so strongly, the nebula is actually opaque (see Figure 4.2) at this

wavelength[6] (Example 5.2). Other Lyman lines are similarly scattered, though to a lesser extent than Ly α.

Example 5.2

What is the mean free path for scattering of a Ly α photon in an HII region of density, $n = 100\,cm^{-3}$?

HII regions are typically at a temperature of $T = 10^4$ K (Table 3.1). The scattering cross-section is $\sigma_{b-b} = 5.0 \times 10^{-14}$ cm^2 at this temperature (Table 5.1) so the mean free path of a photon is,

$$\bar{l} = \frac{1}{n\sigma_{b-b}} = \frac{1}{(100)\,(5.0 \times 10^{-14})} = 2 \times 10^{11}\,\text{cm} \tag{5.1}$$

This is less than three times the radius of the Sun and is a very short distance in comparison to a typical HII region which can be hundreds of parsecs in size. Therefore, the Ly α photon will take a random walk in an HII region in order to escape.

Another example of Ly α resonance scattering can be seen in the *Ly α forest*, an example of which is shown in Figure 5.6 (bottom). In the spectra of distant QSOs (Sect. 1.6.4), many lines can be seen on the shorter wavelength (blue-ward) side of the Ly α line that is emitted from the QSO itself. These lines result from absorption and re-emission (resonance scattering) from the Ly α line in many individual intergalactic HI clouds, each at a different distance along the line of sight towards the QSO. (The differing distances correspond to different redshifts and therefore frequencies along the line of sight, as described in Sect. 7.2.) The Ly α forest lines are commonly referred to as 'absorption lines' because what the observer actually sees is removal of the background QSO light. However, the process is one of resonance scattering since, at the densities of these clouds (see Table 3.1), an atom that has absorbed a photon will re-emit in a random direction before it will collide with another particle. The HI clouds that produce most of the observed Ly α forest lines presumably permeate space but are too faint to be observed directly in emission with current telescopes. They are only seen in absorption in front of bright background sources. However, numerical simulations suggest that they form a *cosmic web* similar to that shown in Figure 5.7.

Although hydrogen is the most abundant atom, other elements are present in astrophysical sources and many of these atoms also have resonance lines (see Ref. [154] for examples). They can also play an important role in the transfer of radiation through a source.

[6]The situation in which downward transitions to the n = 1 level are immediately re-absorbed in HII regions is called *Case B. Case A* assumes that all such emission escapes from the nebula without any interaction. However, the density of a Case A nebula would have to be so low that it would be too faint to be detectable. Case B is considered to be the most realistic situation (Ref. [115]). See also Sect. 9.4.2.

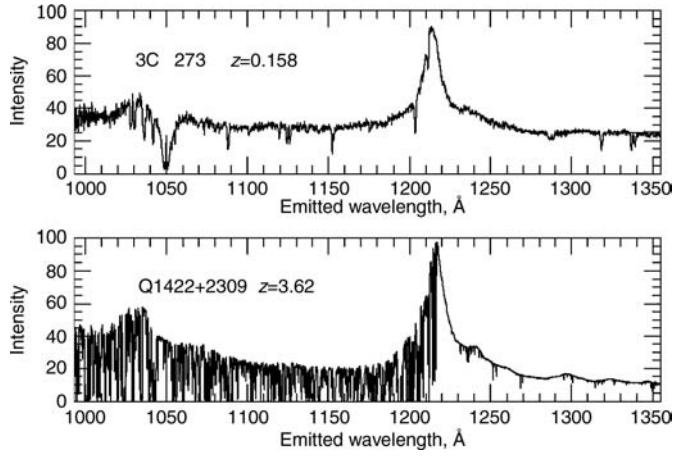


Figure 5.6. Spectra of the quasars, 3C273 (top) and Q1422+2309 (bottom). The quasar, 3C273 is relatively nearby at a redshift of $z = 0.158$ (see Sect. 7.2.2 for a discussion of the expansion redshift) whereas Q1422+2309 is much farther away. The spectra have been aligned so that the Ly α line emitted from the quasars, themselves, is shown at its rest wavelength. In between us and the more distant quasar, Q1422+2309, there are many HI clouds at a variety of distances, creating the many absorption lines seen to the left of the quasar emission. This is called the *Lyman alpha forest*. Most intergalactic gas is, however, ionized. (Reproduced by permission of William Keel)

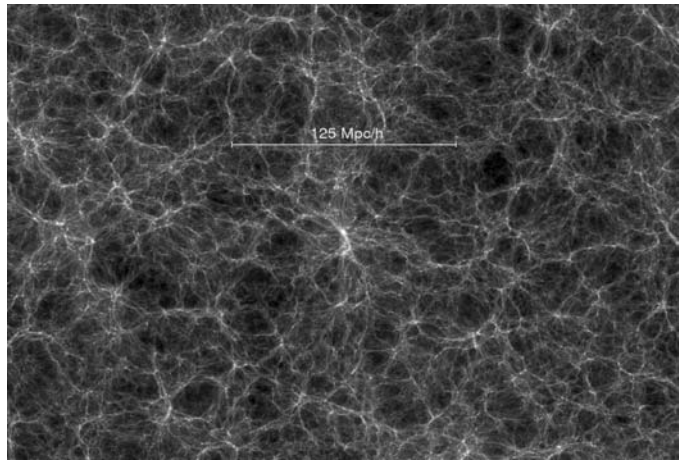


Figure 5.7. One view of the *cosmic web* from the 'Millennium Simulation' which used 10^{10} particles in a computer code to simulate the formation of structure in the Universe (Ref. [161]). The image represents the expected structure of dark matter (see Sect. 3.2) but the structure of the Ly α forest may be similar. This image is for a redshift of $z = 1.4$ (see Sect. 7.2.2), corresponding to a time, 4.7 Gyr after the Big Bang and the unitless quantity, $h = H_0/(100\,\mathrm{kms}^{-1}\,\mathrm{Mpc}^{-1})$ (see Appendix F for information on H_0). (Reproduced by permission of V. Springel and the Virgo Consortium)

5.1.1.3 Rayleigh scattering and the blue sky

The most famous example of Rayleigh scattering is the scattering in the Earth's atmosphere leading to the blue colour of the sky. It was Leonardo da Vinci who, around 1500, first suspected the reason for the blue sky, in particular by his observation that wood smoke looked blue when observed against a dark background (Ref. [79]). The effect was finally described quantitatively in the year 1899 by Lord Rayleigh whose name is now associated with the phenomenon.

Rayleigh scattering occurs when the incoming signal has a wavelength, λ, that is much larger than the resonance wavelengths of a bound electron in an atom or molecule (see Appendix D.1.2). For optical light impinging on particles with UV transitions, this also means that $\lambda \gg$ the size of the scattering particle, as discussed in Appendix D.1.5. Because of the strong dependence of the scattering cross-section on wavelength, $\sigma_s \propto 1/\lambda^4$ (see Eq. D.10, D.13, or D.24), shorter wavelength blue light[7] will scatter more efficiently than longer wavelength red light. Blue light has a wavelength of $\lambda \approx 470$ nm (Table G.5) and, since the most abundant molecules in the atmosphere, namely nitrogen and oxygen, are about 0.3 nm in size, atmospheric scattering is firmly in the Rayleigh regime. Small dust particles also contribute, but the dominant scattering is due to molecules and the sky would still look blue even without the additional contribution by dust.

For a geometry as shown in Figure 5.2b, blue light is more likely to scatter *into* the line of sight of the observer than red light. Thus the yellow sun produces a blue sky to an observer on the Earth. Though less obvious, the nighttime sky is also blue. Although the feebleness of the light at night makes it impossible to discern by eye, a time exposure reveals the colour, as Figure 5.8 shows.

If there were no atmosphere, the daytime sky would be black except at the position of the Sun itself. The fact that the clear sky atmosphere is transparent (Example 5.3) means that most photons travel through it unimpeded with only a small fraction scattered. This is why, on a clear day, the brightness of the Sun is much greater than the brightness of the blue sky.

Example 5.3

What is the mean free path of an optical photon to Rayleigh scattering in the Earth's atmosphere?

The mass density of the atmosphere at sea level (the densest air) is $\rho \approx 1.2 \times 10^{-3}$ gm cm^{-3}. The mean molecular weight of air is $\mu = 28.97$ so the number density of air at sea level is (Eq. 3.11), $n = \rho/(\mu m_H) \approx 2.5 \times 10^{19}$ cm^{-3}, where m_H is the mass of a hydrogen atom. The most likely interaction is with the most abundant molecule, N_2, so its Rayleigh scattering cross-section at 532 nm (Table 5.1) together with Eq. (3.18) yields,

[7] Violet light is scattered even more but our eyes are less sensitive to violet than blue.

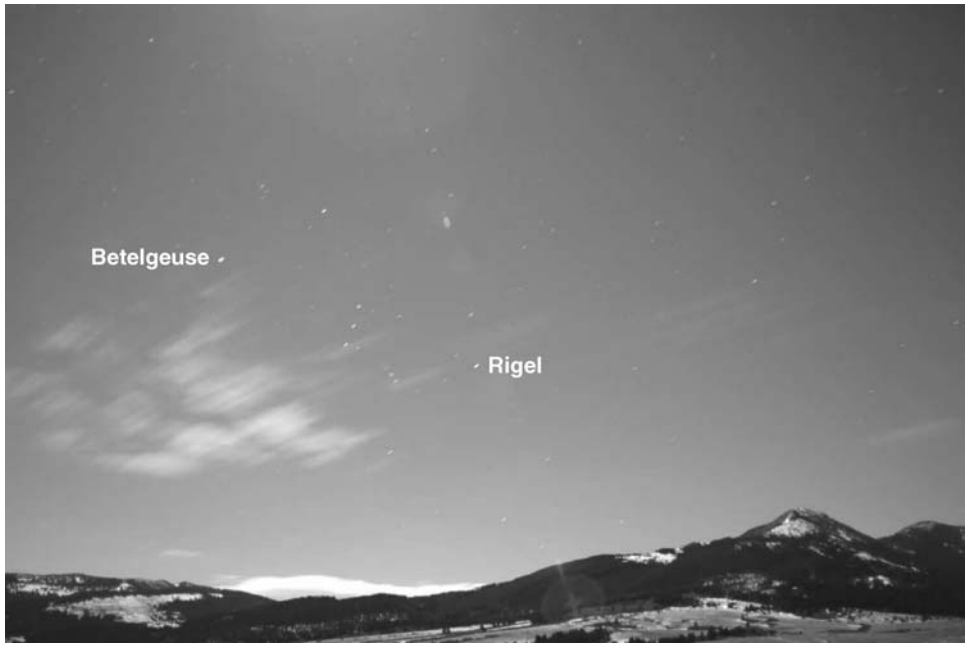


Figure 5.8. A time exposure (ISO 200, f/2.8, 69 s) reveals the blue colour of the night sky in this un-retouched image. (Reproduced by permission of © Joseph A. Shaw, Montana State University Bozeman, Montana). (see colour plate)

$\bar{l} = 1/(n\sigma_R) \approx 80 \times 10^5$ cm. At higher altitude, the density decreases, so the mean free path increases. Thus, $\bar{l} > 80$ km. Recall that this lower limit is well into the ionosphere (Sect. 2.3), confirming what we already know – that the clear sky is optically transparent.

For a geometry as shown in Figure 5.1a, blue light is more likely to be scattered *out* of the line of sight than red light. Thus any light-emitting object above the Earth's atmosphere will be reddened (as first introduced in Sect. 2.3.5) and also dimmer, due to Rayleigh scattering. The Sun is redder than its true colour even when it is high overhead. If the line of sight through the atmosphere is longer, such as when viewing the Sun at sunrise or sunset, (see Figure 2.13), then the redder colour is more enhanced and more obvious to the eye[8]. The same effect can be observed for other objects such as the Moon, planets or stars. Individual scattered photons, however, have the same wavelength as the incident photons, so even though Rayleigh scattering is wavelength dependent, it is still an example of elastic scattering.

Rayleigh scattering produces polarized light in the same way as Thomson scattering (see Figure D.1). Even though the Sun emits light that is unpolarized, for example, its

[8]Scattering (and some absorption) from dust, water vapour and large molecules can also contribute to reddening (see also Prob. 6.1).

scattered light will be polarized at a 90° viewing angle, as can be verified by looking near the horizon with a polarizing filter when the Sun is overhead. Also, like Thomson scattering, Rayleigh scattering provides a way of looking at a source via its 'reflection', although weighted by the $1/\lambda^4$ wavelength dependence. It is thus possible to see the Sun's (weighted) spectrum by pointing a spectrometer at some position in the sky away from the Sun itself. The Sun's Fraunhofer lines[9], for example, can easily be seen in such a fashion. The faint optical light of a reflection nebula (see Figure. 5.3) will also be polarized.

On a final note, the blue sky of the atmosphere can be contrasted with the greyer colour associated with water droplets in clouds. Since water droplets are *not* small with respect to the wavelength of light, scattering from these particles is not in the Rayleigh scattering regime. The wavelength dependence of large particle scattering is flatter than for Rayleigh scattering, hence the greyish colour of clouds, as described in Appendix D.3.

5.1.2 Inelastic scattering

If the scattered photon is of lower energy (longer λ) than the incident photon, then the scattering is said to be *inelastic* or *incoherent.* Several examples follow.

5.1.2.1 Bound–bound inelastic scattering and fluorescence

Inelastic scattering can occur from the bound–bound excitation of an atom to a high energy state followed by cascading to other energy levels between the high state and the initial one. Thus, a high energy photon is degraded into multiple lower energy photons. The emission of light in this cascading process is called *fluorescence.*

There are various possible outcomes for bound–bound scattering, depending on the incident photon energy and the final state of the atom. If the final energy state is the same as the initial one, then the sum of the energies of the emitted photons equals the energy of the incident photon, and the situation is one of pure inelastic scattering (Prob. 5.3). If the final state of the atom is an excited state higher than the initial state, then some energy is lost to the excitation of the atom. Eventually, an atom will always decay spontaneously from any excited state to its ground state but, if an energy-exchanging collision occurs in the mean time, then the energy difference will be lost in the collision. In this case, some of the incident photon energy is turned into lower energy photons and some goes into heating the gas. When energy is 'lost' via excitation or heating, the resulting process is called *absorption* (further considered in Sect. 5.2). The requirement that scattering be inelastic, however, refers only to a comparison between the incoming and outgoing photon energies.

[9]The Sun's Fraunhofer lines are the absorption lines formed in the Sun's *photosphere*, as described in Sect. 6.4.2. See also Figure 6.4.

5.1.2.2 Compton scattering at high energies

Compton scattering (see Appendix D.2) is the scattering of a photon from a free electron in the high energy limit, i.e. when the energy of the incident photon is of order or greater than the rest mass energy of the electron, $E_{\rm ph} \gtrsim {\rm m_e}c^2 = 0.51\,{\rm MeV}$. In Compton scattering, some of the photon energy is imparted to the electron giving it a velocity, as illustrated by Figure D.3. The energy of the scattered photon, $E'_{\rm ph} = hc/\lambda'$, is less than the energy of the incident photon (Eq. D.26), the difference going into the kinetic energy of the electron. The result is the cooling of the radiation field, the scattered photons having longer wavelengths due to loss of energy.

Since 0.51 MeV is in the gamma-ray part of the spectrum (Table G.6), Compton scattering is associated with highly energetic objects, such as active galactic nuclei (AGN), binary stars in which at least one member is a compact object such as a neutron

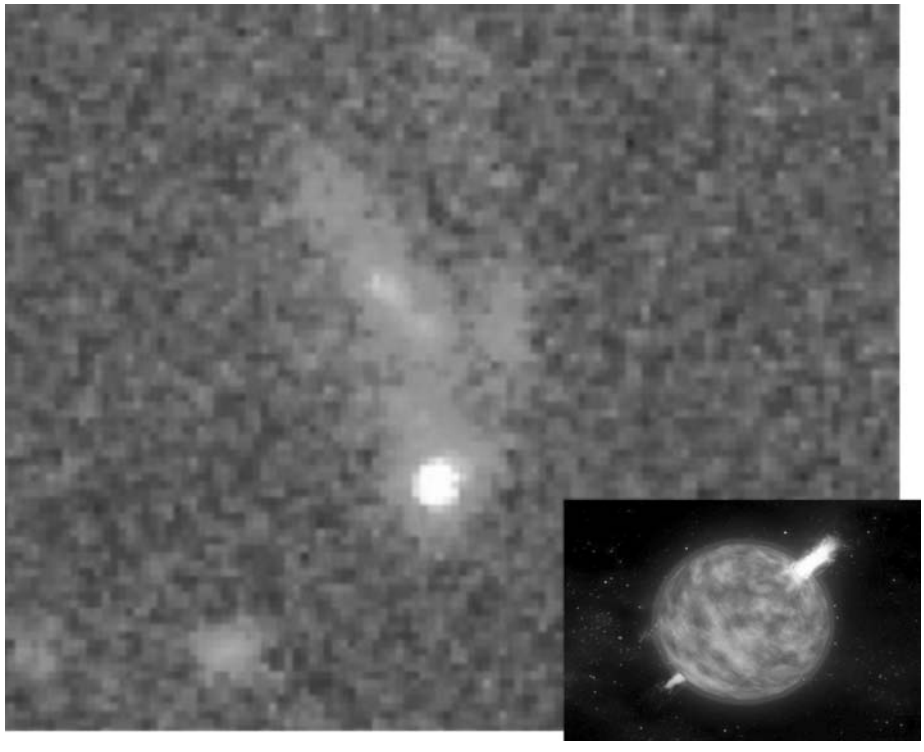


Figure 5.9. The optical afterglow of the powerful gamma-ray burst, GRB 990123, seen here as a bright point source, was caught by the Hubble Space Telescope in this image. The fact that this GRB was found to be associated with a galaxy (faint extended emission) provides powerful evidence that GRBs are located in distant galaxies previously too faint to have been observed. (Reproduced by permission of the HST/GRB collaboration/NASA). The inset shows an artist's concept of an extremely massive star beginning its gravitational collapse with light at gamma-ray energies being beamed into two jets. Compton scattering, among other processes, takes place in the initial phases of the GRB. (Reproduced by permission of NASA/SkyWorks Digital)

star, or in regions around black holes. Other highly energetic sources are the transient *gamma-ray bursts* (GRBs) such as introduced in the light echo example of Figure 5.4. These are bursts of gamma-ray emission of very short duration, from milliseconds to 2 s for the *short duration bursts*, and from 2 s to several minutes for the *long duration bursts*. Discovered by spy satellites in the 1960s, GRBs were found to be associated with distant faint galaxies in the 1990s (see Figure 5.9). GRBs represent the most powerful explosions in the Universe and can be 100 times more powerful than supernovae (Sect. 3.3.3). The long duration events may be related to *hypernovae* – very powerful supernovae that originate from precursor stars more than 20 times the mass of our Sun. A possible precursor is a *Wolf–Rayet star* which is known to have high mass loss via stellar winds (see Figure 5.10). The interaction of the hypernova with wind material may also play a part in forming the GRB. The short duration GRBs likely have a different origin and may be associated with colliding neutron stars or collisions between a neutron star and black hole. It is difficult to explain the high energies of GRBs without requiring that the detected emission be beamed in our direction, as Figure 5.9 suggests (see also Prob. 1.4).

Compton scattering is the dominant scattering process for gamma-ray photons in the range, 0.51 to a few MeV, which is the low energy part of the gamma-ray spectrum. In

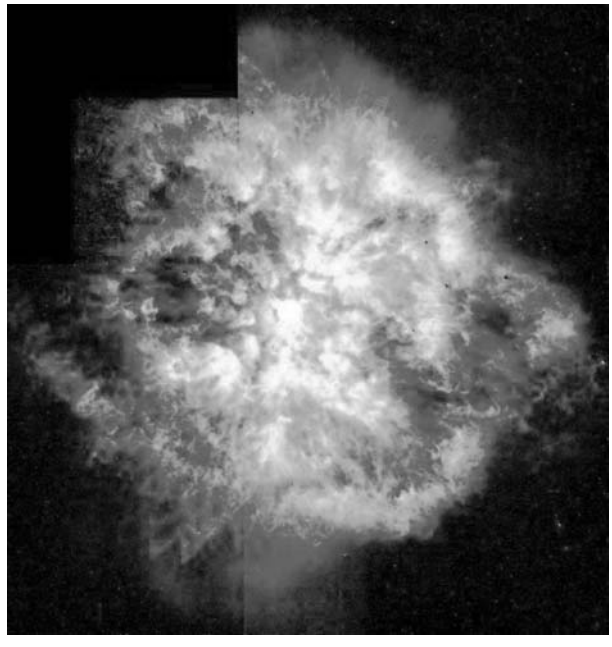


Figure 5.10. The 2.5 pc diameter nebula, M1-67, shown in this H α image from the Hubble Space Telescope, is caused by a massive stellar wind from the Wolf–Rayet star seen at the centre. Wolf–Rayet stars are examples of the youngest, hottest, and most massive stars known and are losing mass via stellar winds at rates of 10^{-5}–$10^{-4} M_\odot$ yr^{-1}. This nebula has a mass of $1.7 M_\odot$ and an age of no more than 10^4 yr, consistent with a very high mass loss rate (Ref. [69]). (Reproduced by permission of Y. Grosdidier et al., U. of Montreal, 1998, ApJ, 506, L127, and NASA)

this regime, the cross-section decreases strongly with increasing photon energy, ensuring that the probability of interaction is low. Thus, gamma-rays in this energy regime are quite penetrating and strong effects are only observed in regions of high density, such as near the source itself (Prob. 5.8). At higher energies, other processes also become important, such as the conversion of a gamma-ray into an electron/positron[10] pair, or the interaction of the gamma-ray with atomic nuclei or even with other photons. Consequently, modelling a highly energetic source, like a GRB, requires that a variety of processes be considered (e.g. Figure 10.1).

Appendix D.2 provides an expression for the energy of the scattered photon when the electron is initially at rest. However, a more general development takes into account the initial kinetic energy of the electron and reveals that the inverse energy exchange can also occur. That is, if the electron is more energetic than the photon (see Eq. D.29 for more exact condition), then the electron can impart an energy to the photon, a process called *inverse Compton scattering* which will be discussed further in Sect. 8.6.

5.2 The photon lost – absorption

Absorption occurs when a photon loses all of its energy (disappears) because of an interaction. In Sect. 5.1.2.1, we gave an example of how a fraction of the energy of an incident photon could be absorbed if a bound–bound absorption occurs, followed by radiative de-excitations to a higher energy state than the initial one. True absorption, however, requires that there is no scattered photon at all; the entire incoming photon energy is lost. Following the symbolism of Sect. 5.1, absorption can be represented as: photon $\rightarrow$ matter. Like the scattering cross-section, the *absorption cross-section* refers to the effective cross-sectional area that a particle presents to an incoming photon for an absorptive process. If a photon is absorbed, energy must still be conserved, so the incoming photon's energy must be converted into some other form. There are primarily two possibilities for the absorption of light by matter: *kinetic energy* gain (which includes heating), or a *change of state*.

5.2.1 Particle kinetic energy – heating

If a photon is absorbed by matter, imparting a motion to it, then the photon energy ($E = h\nu$) is converted into kinetic energy ($E = \frac{1}{2}\mathrm{m}\nu^2$). This occurred for Compton scattering (Sect. 5.1.2.2) but, in that case, a scattered photon remained so the process was categorized as inelastic scattering. We also saw this when we considered radiation pressure (Sect. 1.5). A photon that was completely absorbed by a surface imparted a momentum to that surface. With a sufficiently powerful beam, the surface could be

[10]A positron is the anti-particle to an electron. It has the same mass but a positive charge.

pushed along like a sail. Some of the absorbed photons, however, will contribute to heating the sail. For matter in any phase, heat involves the random internal motions of its constituents and is therefore a form of kinetic energy.

For solids, such as the Solar sail, a dust grain, or a planet, these motions refer to vibrations of the particles within their internal crystalline structure. As was described in Sect. 4.2.1, a dust grain, asteroid or planet that does not move from its position with respect to the light source will eventually achieve and maintain an equilibrium temperature as a result. For dust, it is only the larger grains that actually maintain an equilibrium temperature, since VSGs (see Sect. 3.5.2) cannot retain internal heat and tend to lose it quickly before another photon can be absorbed. The interaction of light with dust is a more complex process and will be considered separately in Sect. 5.5.

If the photon impinges on a gas, then it can contribute to the kinetic energy of individual particles in the gas. Since the randomly moving gas particles share energy via collisions (Sect. 3.4.1), any absorption process that is immediately followed by an energy-exchanging collision will result in the loss of the photon. An example is if a particle experiences a bound–bound excitation followed immediately by a collision (the opposite conclusion to Example 5.1).

In an ionized gas, excess energy from the ionization process goes into the electron kinetic energy (see next section) that is then shared via collisions. Another important absorption process in ionized gases occurs when a free electron is in the vicinity of a positive nucleus so that it feels a small force by the nucleus as it travels past. In such a case, a photon can be absorbed by the electron, giving the electron a higher kinetic energy and different trajectory than it originally had. This is called *free–free absorption*[11]. The free–free cross-section depends on frequency, temperature and density. A sample value is given in Table 5.1.

5.2.2 Change of state – ionization and the Strömgren sphere

Absorbed photons can also produce a change of state, such as the transition from a solid to a gas (*sublimation*), from a neutral gas to an ionized gas (*ionization*), or from a molecular gas to a neutral atomic gas (*dissociation*). These processes could also occur via particle collisions of sufficient energy, so terms such as *photoionization* are sometimes used to distinguish the photo-absorptive process from *collisional ionization.*

Photoionization, in particular, is very common in astrophysics and will occur anywhere that photons are energetic enough to ionize the neutral material that may be nearby. It is useful to remember that there are many constituent elements in the Universe and each has its own ionization energy. These elements can be very important in terms of the heating and cooling balance in various objects (neutral and singly ionized carbon in the ISM is a good example). For the most abundant element, hydrogen, the required energy for ionization from the ground state (its *ionization potential*) is 13.6 eV (Eq. C.5) which is

[11]The inverse of this effect, free–free emission, is discussed in Sect. 8.2

in the UV part of the spectrum (Table G.6). Any photon of equivalent or greater energy, if absorbed by the atom, would ionize it and any excess energy would go into the kinetic energy of the ejected electron, contributing to heating of the ionized gas. The photo-ionization cross-section for hydrogen in its ground state is given in Table 5.1.

Photoionization occurs in a variety of astronomical objects. Some examples are planetary nebulae (Figure 3.7) in which the source of ionizing photons is the central white dwarf, the outer gaseous 'edges' of the disks of galaxies for which the source of ionizing photons is the weak extra-galactic radiation field, the regions around active galactic nuclei (e.g. Figure 5.5) ionized by the AGN itself, and HII regions (Figures 3.13, 8.4) which are ionized by a hot central star or stars. Even the re-ionization of the Universe itself after the 'dark ages' (see Sect. 1.4) is described by this process, although it is not yet clear whether the re-ionization was caused by the first stars or the first AGNs (Prob. 5.10). The quest to understand which sources first 'turned on' in the Universe is currently of much interest and the time period over which this occurred is referred to as the *Epoch of Re-ionization* (EOR).

If an ionized region is not expanding or shrinking and the degree of ionization (the fraction of particles that are in an ionized state) is constant, then the ionization rate must equal the recombination rate and we say that the region is in *ionization equilibrium.* Such an argument is similar to the one used for temperature equilibrium in Sect. 4.2.1. This fact can then be used to determine some physical parameters of the region. In Example 5.4, we do this for the simple case of an HII region around a single star. The star must be a hot, massive, O or B star since only these stars have a sufficient number of UV photons to form HII regions (e.g. Prob. 5.9). In an HII region, the gas is almost completely ionized and we assume that every photon of sufficient energy leaving the star will eventually encounter a neutral atom that it then ionizes. Such a development was first presented by B. Strömgren, and so an HII region around a single star is sometimes called a *Strömgren sphere* which has a radius called the *Strömgren radius.*

Example 5.4

Find the Strömgren radius of an HII region consisting of pure hydrogen of uniform electron density, $n_e = 1000\,cm^{-3}$ *and temperature,* $T = 10^4$ *K around a single central O8V star.*

As described in the text, $N_i = N_r$, where N_i is the number of ionizing photons per second leaving the star and N_r is the number of recombinations per second within the nebula. Beginning with the recombination rate, a recombination can occur whenever an electron 'collides with' a proton. We use the total collision rate per unit volume, ν_{tot} from Eq. (3.2.1) multiplied by the total volume, V of the HII region and use the collision rate coefficient appropriate to recombination, called the *total recombination coefficient* α_r (Table 3.2),

$$N_r = \nu_{tot}\, V = n_e\, n_p\, \alpha_r\, V = n_e^2\, \alpha_r\, \frac{4}{3}\, \pi\, {R_S}^3 \tag{5.2}$$

where n_p is the proton density, taken to be equivalent to n_e for pure ionized hydrogen and R_S is the Strömgren radius. Equating N_r to N_i and rearranging we find,

$$R_S\, {n_e}^{2/3} = \left(\frac{3}{4\,\pi}\,\frac{N_i}{\alpha_r}\right)^{1/3} \tag{5.3}$$

(all cgs units). The RHS of this equation contains information about the exciting star (plus constants[12]) and the LHS contains information about the HII region, often combined into a single parameter called the *excitation parameter*, $\mathcal{U}$,

$$\mathcal{U} \equiv R_S {n_e}^{2/3} \tag{5.4}$$

The excitation parameter has cgs units of cm cm^{-2} but is often expressed in units of pc cm^{-2} (see Sect. 8.2.2, for example). If the exciting star is known, implying that N_i is known (see below), Eq. (5.3) relates the size of the HII region to its density. If either quantity is known, the other can be calculated, although a more accurate derivation would consider non-uniform densities and also take into account some absorption of photons by other elements such as helium, or by dust particles.

The quantity, N_i, can be computed from the ionizing luminosity, L_i (erg s^{-1}), of the star divided by the photon energy, $h\nu$ (erg). The photon energy, however, can have a variety of values provided it is greater than the ionization potential, that is, $\nu \geq \nu_i$, where ν_i is the threshold frequency required to ionize hydrogen. L_i can be determined from an integration of the Planck curve, $B_\nu(\mathrm{T})$ (Eq. 4.1), using Eqs. (1.10) and (1.14),

$$N_i = \frac{L_i}{h\,\nu_{(\geq \nu_i)}} = 4\,\pi^2\,R_*^2 \int_{\nu_i}^{\infty} \frac{B_\nu(\mathrm{T})}{h\,\nu}\,d\nu \tag{5.5}$$

where R_* is the radius of the star. For an O8V star, T = 37 000 K and $R_* = 10 R_\odot = 6.96 \times 10^{11}$ cm (Table G.7). Numerical integration with $\nu_i = 3.28 \times 10^{15}$ Hz yields $N_i = 7.95 \times 10^{48}$ photons s^{-1}. Using this value in Eq. (5.3) along with $n_e = 1000$ and $\alpha_r = 2.59 \times 10^{-13}$ cm^3 s^{-1} (Table 3.2), gives $R_S = 1.93 \times 10^{18}$ cm or 0.63 pc.

As indicated in the above example, this simple development links the stellar type to the size and density of the HII region. If the stellar type is known and the density determined by other means (e.g. Sect. 8.2.2) the angular size can be compared to R_S to find the distance. HII regions have distinct boundaries (Example 5.5) so this measurement is straightforward. If the distance is known, the size can be used to determine the kind of star that may be ionizing the nebula or whether more than one star is required. Although corrections are needed for dust and density variations, the simple result of Eq. (5.3) can still be very useful in placing constraints on the quantities of interest.

[12]The recombination coefficient is not exactly constant but varies only slowly with temperature.

Moreover, the arguments apply to any ionized region. If the source of ionizing radiation is an AGN at the centre of a galaxy, then one need only insert the spectrum of an AGN (a power law) instead of the Planck function in Eq. (5.5).

Example 5.5

Why do HII regions and planetary nebulae have sharp boundaries?

If we assume a density of $n = 100\,\mathrm{cm}^{-3}$ (Table 3.1) and use the cross-section for photoionization in Table 5.1, the mean free path (Eq. 3.18) of a photon to the ionization of hydrogen is $\bar{l} = 1/(n\sigma_{\mathrm{HI}\rightarrow\mathrm{HII}}) = 1/[(100)(6.3 \times 10^{-18})] = 1.6 \times 10^{15}\,\mathrm{cm} \approx 100\,\mathrm{AU}$. Thus, the transition region between an ionized and a neutral gas of this density is about 100 AU. Assuming a spatial resolution of $\approx 1''$ (Sect. 2.3.2) is possible, the HII region or planetary nebula would have to be closer than 100 pc (Eq. B.1) in order for this region to be spatially resolved. The closest known HII region is the Orion Nebula at a distance of 460 pc and the closest planetary nebula is likely the Helix Nebula at a distance of about 140 pc. Therefore all HII regions and planetary nebulae have boundaries that are sharp, that is, they cannot be resolved.

5.3 The wavefront redirected – refraction

Refraction is another example of light interacting with matter. As light passes from one kind of transparent medium to another with a different index of refraction, n, the wavefront changes direction according to *Snell's law* (Table 1.1). An example that we have already seen is the systematic bending of light as it enters the Earth's atmosphere (see Figure 2.A.1 in the Appendix at the end of Chapter 2). Figure 5.11 shows how the bending of a wavefront can be thought of as the result of interactions that slow the speed of light, ν, through the medium with the higher index of refraction[13] (from Table I.1, $\mathrm{n} = c/\nu$). Although other subtleties are involved, this illustration is useful in providing a visualization of the refraction phenomenon from both a wave and particle viewpoint.

Aside from the Earth's atmosphere, refraction can also occur under other astrophysical conditions. For example, two media of the same composition but a different density should have different indices of refraction. Thus, in principle, any gas within which density variations exist should also experience some refractive effects. For neutral gas, the degree of bending turns out to be quite negligible. However, refraction can become important when a signal propagates through an ionized gas (a plasma, see Appendix E) or through a neutral gas that has a significant fractional ionization.

[13] A similar decrease in the speed of light in a medium was considered for the diffusion of photons out of the Sun (see Prob. 3.7), though in that case, the medium was opaque and scattering was in random directions.

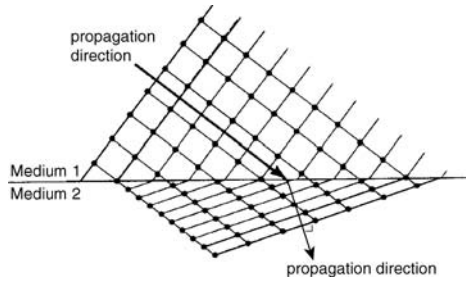


Figure 5.11. This simple diagram is a nice illustration of how refraction can be thought of in terms of both interactions between photons and individual particles, as well as a change in the direction of the wavefront. The arrows show the direction of propagation of the wavefront and the diagonal lines (upper left to lower right) indicate the direction of propagation of the photons. In this illustration, the photons themselves do not change direction when they interact with particles (consistent with scattering that is preferentially along the line of sight). The speed of light is slower in medium 2 because there are more interactions in this denser medium. (Reproduced courtesy of Shu, F. H., *The Physical Universe, An Introduction to Astronomy*, University Science Books)

For a plasma, we can express the index of refraction using its definition (Table I.1)[14] together with Eq. (E.7), and writing the result in terms of the frequency, ν, of the incoming wave,

$$\mathrm{n} = \left[1 - \left(\frac{\nu_\mathrm{p}}{\nu}\right)^2\right]^{1/2} \tag{5.6}$$

where $\nu_\mathrm{p} = \omega_\mathrm{p}/(2\pi)$ is the *plasma frequency*, ω_p being the angular plasma frequency (see Appendix E). Note that we require $\nu > \nu_\mathrm{p}$ for n to be real, implying that a signal with a frequency less than ν_p cannot propagate through the plasma. Such a signal is reflected, rather than transmitted (see Appendix E for examples.) On the other hand, if $\nu \gg \nu_\mathrm{p}$, then $\mathrm{n} \approx 1$ in which case there is no bending. Eqs. (5.6) and (E.6), and a binomial expansion for small ν_p/ν yield,

$$\mathrm{n} = 1 - n_\mathrm{e}\left[\frac{0.0063}{\left(\frac{\nu}{\mathrm{MHz}}\right)}\right]^2 \tag{5.7}$$

where n_e is the electron density.

Refraction will be greatest for media in which n differs the most from 1. Therefore, from Eq. (5.7), we can see that refraction is most important at low frequencies. Since the lowest radio frequencies observed are seldom less than several hundred MHz, this

[14] As indicated in the table, the index of refraction can be written as a complex number for cases in which some fraction of the incoming light is absorbed. For the astrophysical plasmas discussed here, a complex index of refraction is not needed.

expression shows that departures of the index of refraction from unity and therefore changes in the angle of propagation of the signal are very small unless electron densities exceed $\sim 10^7\,\mathrm{cm}^{-3}$ (for example, the coronas and atmospheres of stars). However, both a sufficient electron density as well as some specific geometry, such as sheets or filaments, are required for refraction from ionized gas to be important in any systematic sense. Note that Eq. (5.6) implies that $\mathrm{n} \leq 1$, (since $\nu > \nu_\mathrm{p}$ is required for propagation) so an intervening plasma can act like a *diverging lens*[15] in front of a background signal. If we repeated the atmospheric bending example given in the Appendix to Chapter 2 for the ionosphere, for example, the bending of light would be away from the normal in Figure 2.A.1.

From the point of view of astrophysical gases, refraction is mostly observed in terms of its random effects. Even in the very low density ISM ($n_\mathrm{e} \approx 10^{-3}\,\mathrm{cm}^{-3}$), for example, refraction can be important for background point sources such as pulsars or distant point-like radio-emitting quasars. As the signal travels through regions of varying density in the ISM en route to the Earth, even a very small change in angle due to refraction can move the signal in and out of a direct path to the detector on Earth. This, along with other effects such as time variable scattering, causes the brightness of the signal to fluctuate with time, an effect called *interstellar scintillation.* This effect, a kind of *space weather*, is similar to the scintillation (Sect. 2.3.4) that we have already seen for the Earth's atmosphere, so the ISM acts like a *cosmic atmosphere.* Since the refractive index is frequency dependent (Eq. 5.7), the ISM is also *dispersive*, like a prism. That is, if a broad spectrum of wavelengths are present in the incident beam (which is the case for pulsars and AGN) then the lower frequency light will be bent more than the higher frequency light.

Pulsars, in particular, tend to move rapidly so, statistically, the component of their velocity that is transverse (i.e. perpendicular to the line of sight) will be rapid as well (of order $100\,\mathrm{km\,s}^{-1}$). Therefore, a background pulsar will 'illuminate' different parts of the ISM with time. Studying the properties of the resulting signal as a function of both time and frequency provides important information on the properties of the ISM and ISM turbulence. Scintillation due to refraction, for example, is variable on times scales of days for a pulsar at a distance of 1 kpc moving at $100\,\mathrm{km\,s}^{-1}$ and an ISM electron density fluctuation of order a factor of 2. The implied size scale for the perturbing cloud is then $\sim 1.5 \times 10^{12}$ cm (Ref. [146]), or one tenth of an astronomical unit. If dispersive properties of the ISM are considered, size scales of 10^{10} cm or smaller can be studied. Estimates for the size scales of interstellar density fluctuations that have been studied in this way range from about 10^7 cm to 10^{18} cm and the number density of clouds follows the power law $N \propto D^{-11/3}$, where D is the cloud diameter. Thus, analysing the perturbation in the signal from a background source is a powerful tool for studying the richness of the structure of the ISM on size scales that would otherwise be inaccessible to direct imaging (Prob. 5.13).

[15] From the definition of n, this implies that $\nu > c$, but here ν is the phase velocity (see footnotes to Table I.1) which can be greater than c without violating Special Relativity since it is the group velocity that carries information.

5.4 Quantifying opacity and transparency

5.4.1 Total opacity and the optical depth

As we have seen from the previous sections, there are many possible ways in which a photon can interact with matter. A key parameter that helped to determine whether an interaction might occur was the mean free path, $\bar{l}$ (Eq. 3.18) already discussed in a variety of contexts for both particle–particle collisions as well as photon–particle collisions (e.g. Examples 3.4, 3.5, 5.2, 5.3). Although the mean free path is a 'representative' path length, not every photon (or particle) will travel exactly this distance before interacting. In addition, we have discussed gases in terms of whether they are 'transparent' (mean free path is greater than the size of the gas, such as the Earth's clear sky) or 'opaque' (mean free path is less than the size of the gas, such as the Sun), as illustrated in Figure 4.2. We now wish to quantify these terms. That is, we need to consider what fraction of the incoming beam will be scattered out of the line of sight or absorbed and what fraction will be transmitted through the gas. Such an approach will allow the characterization of cases that are intermediate between being fully opaque and fully transparent, such as a cirrus cloud in front of the Sun, a mist around a street lamp, or a partially transparent HI cloud in the ISM. Since we wish to fold together all processes that remove photons from a line of sight, scattering and absorption are considered together as a total *opacity* (or *extinction* which is used for dust, see Sects. 3.5.1, 5.5).

We now introduce a unitless parameter called the *optical depth*, τ_ν, such that an infinitesimal change in optical depth, $d\tau_\nu$, along an infinitesimal path length, dr, is given by,

$$d\tau_\nu = -\sigma_\nu \, n \, dr = -\kappa_\nu \, \rho \, dr = -\alpha_\nu \, dr \tag{5.8}$$

where σ_ν is the *effective cross-section* for the interaction (cm^2), κ_ν ($\mathrm{cm}^2\,\mathrm{g}^{-1}$) is the *mass absorption coefficient* and α_ν (cm^{-1}) is the *absorption coefficient,* at a given frequency. The quantities, n, and ρ are the particle density (cm^{-3}) and mass density ($\mathrm{g\,cm}^{-3}$), respectively. The term, 'absorption', in the above context, is meant to include all processes that remove photons from the beam (as in Figure 5.1.a and b)[16]. These terms could be separated into a true absorption and a scattering portion, separately, if desired. The negative signs in Eq. (5.8) indicate that, while the coordinate, r, increases in the direction that the incoming beam is travelling, the optical depth is measured *into* the cloud as measured from the observer (see Figure 6.2). The total optical depth is then determined from an integration over the total line of sight through the source. Using the version with the mass absorption coefficient,

$$\tau_\nu = -\int_l^0 \kappa_\nu \, \rho \, dr \approx \kappa_\nu \, \rho \, l \tag{5.9}$$

[16]The coefficients could also include corrections for scattering *into* the line of sight.

where we integrate from the near to the far side of the cloud and l is the line of sight thickness of the cloud. The latter approximation is often used for situations in which the density and mass absorption coefficient can be considered constant through the cloud.

The optical depth provides a quantitative description of how transparent or opaque a gas is. If $\tau_\nu < 1$, then the probability that the photon will be scattered or absorbed is < 1 and the cloud is *optically thin* or transparent (Figure 4.2 top). If $\tau_\nu > 1$, then the probability that a photon will be scattered or absorbed is high and the cloud is *optically thick* (Figure 4.2 bottom). If $\tau_\nu = 1$, then the cloud is 'just' optically thick and the mean free path is equal to the line of sight thickness of the cloud. The optical depth can be thought of as the number of mean free paths through an object. The shape of the function, τ_ν, when plotted against frequency shows how the effective absorption cross-section varies with frequency for a cloud of give density (Eq. 5.8). From Eq. 5.8 it is clear that the optical depth depends on (a) the type of material, (b) the frequency of the radiation, (c) the density of material and (d) the line of sight distance. It is possible, for example, for a small dense cloud to have the same optical depth as a large diffuse cloud of the same material. It is also possible for two clouds of the same size and density, but different material, to have very different optical depths. Example 5.6 further elaborates.

Example 5.6

Consider two ionized clouds of pure hydrogen, each with the same fractional ionization of 99.9 *per cent, density of* 10^{-2} *cm*$^{-3}$ *and temperature of* $T = 10^4$ *K. Cloud* 1 *has a line of sight thickness of* $l = 1$ *pc and Cloud* 2 *has* $l = 100$ *pc. Compare the optical depth of Cloud* 1 *at a wavelength of* $\lambda = 121.567$ *nm to the optical depth of Cloud* 2 *at a wavelength of* $\lambda = 200$ *nm. Assume that only scattering processes are important (see Table* 5.1*) and ignore scattering into the line of sight.*

Cloud 1: Most particles are ionized, so electron scattering must be considered. In addition, however, the incoming photons are exactly at the wavelength of the resonance scattering of the Ly α line (Table C.1) so we also need to consider whether Ly α scattering will be important for those few particles that are neutral (all of which we assume are in their ground states, Sect. 3.4.6). Then, from Eqs. (5.8) and (5.9),

$$\tau_\nu = \sigma_\nu n l = \sigma_{\rm T} (0.999) n l + \sigma_{{\rm Ly}\,\alpha} (0.001) n l \tag{5.10}$$

$$= (6.65 \times 10^{-25}) (0.999) (10^{-2}) (3.09 \times 10^{18}) \tag{5.11}$$

$$+ (5.0 \times 10^{-14}) (0.001) (10^{-2}) (3.09 \times 10^{18}) \tag{5.12}$$

$$= 2 \times 10^{-8} + 1.5 = 1.5 \tag{5.13}$$

where we have used the Thomson scattering and Ly α cross-sections from Table 5.1. This cloud is optically thick, with an opacity dominated by Ly α scattering from a small fraction of neutral particles.

Cloud 2: In this case, the incoming photons are not at the frequencies of any resonance lines so we need only consider Thomson scattering,

$$\tau_\nu = (6.65 \times 10^{-25})(0.999)(10^{-2})(100)(3.09 \times 10^{18}) = 2 \times 10^{-6} \tag{5.14}$$

This cloud, although 100 × larger than the previous one, is highly optically thin.

As indicated above, the total mass absorption coefficient of Eqs. (5.8) and (5.9) must include all processes that may be occurring. We can therefore write a general expression for κ_ν that takes this into account,

$$\kappa_\nu = \kappa_{bb\nu} + \kappa_{bf_\nu} + \kappa_{es} + \kappa_{ff_\nu} + \ldots \tag{5.15}$$

where the subscripts indicate, bb: bound–bound processes such as line absorption and scattering, bf: bound–free processes such as ionization, ff: free–free processes such as free–free absorption, es: electron scattering (Thomson scattering), and whatever other processes might be occurring[17]. Fortunately, not all processes are occurring in any given environment. In fact, in many situations, a single process may dominate all others, as suggested in Example 5.6 for the second cloud.

In the case of stars, all of the processes listed in Eq. (5.15) do have to be considered. This is not an easy or straightforward process since some of these quantities, themselves, depend on density and temperature and can also be highly frequency dependent. Bound–bound opacities, $\kappa_{\rm bb}$, for example, will be significant at frequencies corresponding to the quantum transitions in atoms and zero otherwise. For the bound–free and free–free absorption coefficients, it has been found that they can be approximated by functional forms that depend on mass density, ρ, and temperature, T (see Ref. [152] for further details),

$$\overline{\kappa_{\rm bf}} = 4.34 \times 10^{25}\, {\rm f_{bf}}\, ({\rm Z})\, (1 + {\rm X}) \frac{\rho}{{\rm T}^{3.5}} \qquad {\rm cm^2 g^{-1}} \tag{5.16}$$

$$\overline{\kappa_{\rm ff}} = 3.68 \times 10^{22}\, {\rm g_{ff}}\, (1 - {\rm Z})\, (1 + {\rm X}) \frac{\rho}{{\rm T}^{3.5}} \qquad {\rm cm^2 g^{-1}} \tag{5.17}$$

where X, Y, and Z represent the stellar composition (Sect. 3.3.1). The correction factors, $\rm f_{bf}$ and $\rm g_{ff}$, are of order $\approx 0.01 \rightarrow 1$ and ≈ 1, respectively[18]. The 'overline' indicates that the quantity is a *Rosseland mean opacity* for the process being considered, which is a statement that a weighted average over frequency has already been carried out. For example, $\overline{\kappa_{\rm ff}}$ can be derived from the free–free absorption coefficient

[17]There could be other sources of opacity in special environments. For example, some very high energy processes such as Compton scattering or the absorption of a high energy photon in an atomic nucleus are possibilities, or absorption processes specific to extremely dense environments such as those that are degenerate (see Sect. 3.3.2).
[18]The factor, $\rm f_{bf} = g_{bf}/t$, where $\rm g_{bf}$ is the mean bound–free Gaunt factor for the various types of particles (of order 1, see also Eq. C.9), and t is called the *guillotine factor* which is of order 1 to 100. The factor, $\rm g_{ff}$ is the free–free Gaunt factor (also of order 1) averaged over frequency. See Figure 8.2 for frequency-dependent values (Sect. 8.2).

given in Eq. (8.3), with appropriate weighting (see Sect. 8.2). Any opacity that has the functional form $\propto \rho/T^{3.5}$ is referred to as a *Kramers' law.* Although these functions are approximations, they do show how the opacities depend on density and temperature. The bound–free opacity is dominated by metals in the interior of the Sun since hydrogen and helium are mostly ionized. However, the free–free opacity is dominated by free electrons from hydrogen and helium because these elements are far more numerous than the metals (Ref. [152]). Note that, as the temperature increases, the bf opacity decreases because more atoms are already ionized and therefore there are fewer bound electrons to release from the atoms. The ff opacity also decreases because, at higher temperature, the free electrons are moving faster, on average, and have a lower probability of absorbing an incident photon given the shorter time spent near nuclei.

For electron (Thomson) scattering in regions of complete ionization (Ref. [144]),

$$\kappa_{es} = 0.200\,(1 + X) \quad \text{cm}^2\text{g}^{-1} \tag{5.18}$$

(Prob. 5.16), where a mean over frequency is not necessary since Thomson scattering is frequency independent (Appendix D.1.1).

In stars in which there are few heavy metals, the most important processes are free–free absorption and electron scattering. In stars with significant metallicity and in the cooler outer layers of stars in which there are more bound electrons, κ_{bf} will dominate. The most important source of opacity in the outer layers of the Sun, for example, is bound–free absorption from the H^- ion. It is clear that the opacity of stars is a very sensitive function of density, temperature, and composition, and can vary strongly between the deep interior and the surface. See Ref. [70] for opacity values that are consistent with the Solar interior model of Table G.10.

Determining the opacities in stars and elsewhere is an extremely important pursuit in astrophysics and many astronomers have spent their careers calculating and improving upon the published values. Without a knowledge of opacities, we could not understand how the signal is affected as it travels through matter, be it stellar, interstellar, atmospheric, or any other material (this is pursued in greater detail in the next chapter). Without a knowledge of opacities in stellar interiors, moreover, we could not relate the luminosities that are observed at the surfaces of stars to the sources of energy generation at their cores, nor could other stellar parameters be accurately determined as a function of radius, such as densities, temperatures, ionization states, and others. In fact, the very *structure* of a star is intimately connected with its opacity and the variation of opacity with radius. As we will see in the next section, the opacity can have important consequences for stellar *dynamics* as well.

5.4.2 Dynamics of opacity – pulsation and stellar winds

If an incident beam passes through a gas with little probability of interaction ($\tau_\nu \ll 1$) then that beam will have little or no dynamical effect on the gas. However, if the probability of absorption is very high ($\tau_\nu \gg 1$) then, as indicated in Sects. 1.5 and

5.2.1, the photons exert a pressure against the gas, as if hitting a wall. Thus, anywhere that the opacity is sufficiently high, dynamical effects can become significant. The following are two important examples of the dynamics of opacity.

The first occurs in certain types of *variable stars* (stars that vary in brightness) that do so because they are radially expanding and contracting. *Cepheid variables* are examples of this, named after the proto-type star, δ Cephei. These stars are essential tools in obtaining distances in the Universe because they are bright enough to be seen over large distances and also because their pulsation periods (which are easily measured and vary from about 1 to 100 days) are related in a known way to their intrinsic luminosities. If the apparent magnitude of a Cepheid is measured and the intrinsic luminosity is inferred from the pulsation period, then the distance can be found using Eq. (1.30). Objects like this, whose intrinsic luminosities are known, are called *standard candles.* Since Cepheids are very bright and can be seen in galaxies other than our own Milky Way, they are used to measure distances out to about 25 Mpc.

If a small perturbation (compression) occurs in a layer within a star, both the density and temperature will increase slightly. From Kramers' laws of Eqs. (5.16) and (5.17), the temperature has a greater effect than the density so the mass absorption coefficient normally decreases. Since $\tau_\nu = \kappa_\nu \rho$ (Eq. 5.9), a decrease in κ_ν in the perturbed layer offsets the increase in ρ so that there is little change in the optical depth. Thus, stars tend to be dynamically stable to such perturbations. However, there are certain regions within stars that are called *partial ionization zones.* These are at depths corresponding to temperatures at which hydrogen and helium (the latter corresponds to a deeper layer) become ionized. In these layers, if there is a small compression, instead of the temperature increasing as it did before, energy goes into ionization instead, which is a change of phase. The effect is similar to one in which energy is continually added to ice, raising its temperature to the melting point at which time the energy goes into the ice-to-water phase change rather than an increase in temperature. Since the temperature rise in the stellar ionization zones is now very small, the opacity, described by Kramers' Laws, *increases* with the increasing density of the perturbed layer. The optical depth then also increases. Photons from the deeper interior then exert a greater pressure against the perturbed, higher opacity layer, causing an expansion. The opposite occurs during expansion as electrons recombine, releasing energy, increasing the temperature and therefore decreasing the optical depth again. Thus, during compression the partial ionization zone absorbs heat and during expansion the layer releases heat, acting like a *heat engine.* During expansion, the layer may overshoot, that is, the velocity given to the layer by radiation pressure perturbs it from gravitational (or *hydrostatic*) equilibrium and therefore it will fall back again gravitationally, causing the cycle to repeat. This process is called the *kappa mechanism* after the mass absorption coefficient, which acts like a valve, within the star.

Not every star experiences such pulsation. Stars that are too hot have ionization zones too close to the surface in regions in which the density is too low for the kappa mechanism to be effective. Stars that are too cool have deep convection zones which disrupt the kappa mechanism. Thus, only a certain region of the HR diagram

(Figure 1.14) is amenable to this pulsation mechanism. This region is referred to as the *instability strip.*[19] Stellar pulsation can result in mass loss and, if the outer layers of the pulsating star are extended and of low density, such as in a red giant star, then mass loss is more likely. Opacity-driven dynamical effects, such as the kappa mechanism, therefore play a crucial role in the evolution of stars (see Sect. 3.3.2).

The second important example involves stars, or other objects, of very high luminosity. Most stars, like our Sun, are stable because their internal pressure provides support to balance their inwardly pulling gravity. The internal pressure is the sum of all internal pressures (see Eq. 3.16) which, for low mass stars like our Sun, is dominated by particle pressure given by the perfect gas law (Eq. 3.11). However, in stars that are more massive, hotter, and more luminous, as described in Sect. 4.1.5, the radiation pressure (Eq. 4.15) dominates. In very high mass stars, the radiation pressure is so strong that the star's gravity is insufficient to retain the outer gaseous layers and they are blown off in *stellar winds* (Figure 5.10). This, in fact, puts a theoretical upper limit on the masses of stars. In Sect. 3.3.2, we noted that a lower limit to the mass of a star is set by the inability of low mass stars to ignite nuclear burning in the core. Now we have an argument for an upper limit to the mass of a star. Although other processes can also contribute to stellar mass loss (such as magnetic fields) the simple argument of radiation pressure leads to a theoretical upper limit that is very close to the observed upper limit of stellar masses (see Example 5.7). The highest luminosity that any object can have before radiation pressure starts to 'blow the object apart' is called the *Eddington Luminosity* or *Eddington Limit.* Our Sun's luminosity is 'sub-Eddington' because it is not shining at the maximum luminosity possible for an object of its mass. However, we should not see any objects that are 'super-Eddington', because they would not be stable. It has been suggested that some super-Eddington examples exist, but these may be related to special circumstances such as short time scales, inhomogeneous gas distributions, or unusual geometries.

Example 5.7

Estimate the highest stellar mass possible before radiation pressure starts to cause mass loss.

Consider a star with an outer layer of thickness equivalent to a photon's mean free path, $\bar{l}$ (taken to be small in comparison to the star's radius). Then a typical photon that passes through this region will be absorbed and $\tau = \kappa_\nu \rho \bar{l} = 1$. The pressure in such a layer is directed outwards and we can then use Eq. (1.23) with $\theta = 0$, to find the radiation pressure on this outer layer,

$$\mathrm{P_{rad}} = \frac{f}{c} = \frac{L}{4\pi r^2 c} \tag{5.19}$$

[19]Note that there are a variety of different kinds of variable stars and also a variety of pulsation mechanisms. The instability strip shown in Figure 1.14 is applicable only to Galactic Cepheids [Ref. 166a].

where f is the star's flux, L is its luminosity, r is its radius, and c is the speed of light. The force per unit area due to gravity acting inwards on this layer is,

$$\mathrm{P_{grav}} = \frac{\mathrm{F}}{4\,\pi\,r^2} = \frac{1}{4\,\pi\,r^2}\,\frac{G\,\mathrm{M_*\,m}}{r^2} = \frac{G\,\mathrm{M_*}(\rho\,V)}{4\,\pi\,r^2\,r^2} = \frac{G\,\mathrm{M_*}\rho\,4\,\pi\,r^2\,\bar{l}}{4\,\pi\,r^2\,r^2} \tag{5.20}$$

where m is the mass of the layer, ρ is its density and $V = 4\pi r^2\bar{l}$ is its volume. For balance, we equate Eq. (5.19) to Eq. (5.20) and replace $\rho\bar{l}$ with $1/\kappa_\nu$ to give,

$$L_{\mathrm{Ed}} = \frac{4\,\pi\,c\,G\,\mathrm{M_*}}{\kappa_\nu} \tag{5.21}$$

where L_{Ed} is the Eddington Luminosity. That is, if $L > L_{\mathrm{Ed}}$, then the star will start to lose mass through radiation pressure. It is therefore is unlikely that stars with luminosities greater than L_{Ed} exist. For the very high temperature, low density outer layers of the most luminous, massive stars, the dominant source of opacity should be electron scattering. Therefore, using Eq. (5.18) with $\mathrm{X} = 0.707$ (Eq. 3.4), expressing the mass in units of $\mathrm{M_\odot}$ and the luminosity in $L_\odot$,

$$L_{\mathrm{Ed}} = 3.8 \times 10^4\,\frac{\mathrm{M_*}}{\mathrm{M_\odot}} \quad L_\odot \tag{5.22}$$

It is now possible to compare this luminosity to the highest luminosities that are actually observed for stars. The most massive stars known (Ref. [103]) have $\mathrm{M} \approx 136\ \mathrm{M_\odot}$, for which $L_{\mathrm{Ed}} = 5.2 \times 10^6\ L_\odot$ (Eq. 5.22). These stars are observed to have $M_{\mathrm{bol}} \approx -11.9$ which, using Eq. (1.31), corresponds to an observed luminosity of $L_{\mathrm{obs}} = 4.5 \times 10^6\ L_\odot$. Thus the observed upper limit is in good agreement with the Eddington value.

The Eddington limit applies to objects other than stars as well. For example, the Eddington luminosity argument is often applied to active galactic nuclei (AGN) that exist at the centres of some galaxies (e.g. Figures 1.4, 5.5). AGN activity is believed to be powered by a *supermassive black hole*, that is, black holes in the range, 10^5–10^{10} $\mathrm{M_\odot}$. Although no light can escape from the black hole itself, the material that is surrounding it, either accreting onto the black hole or rapidly orbiting it, is in a region in which the gravitational energy is extremely high. This material is heated to very high temperatures, producing radiation and its resulting pressure. Just as for stars, the radiation pressure must not exceed that produced by the gravity of the black hole itself or else the material will be blown away. It is believed that most AGN are radiating at or close to their Eddington limits (within a factor of ≈ 3, Ref. [90]). Thus, Eq. (5.22) provides a way of relating an *observable AGN luminosity to the mass of an unobservable black hole.* Typical values of luminosity for powerful AGNs are in the range, $10^{45} \rightarrow 10^{47}$ erg $\mathrm{s^{-1}}$, corresponding to black hole masses of $\approx 10^7 \rightarrow 10^9\ \mathrm{M_\odot}$ at the centres of galaxies.

5.5 The opacity of dust – extinction

The interaction of light with dust grains is complex since dust grains have irregular shapes, come in a variety of compositions with individual grains also being of non-uniform composition, span a range of sizes, and some fraction of them may be aligned with the magnetic field (see Sect. 3.5). As a result, details of the interaction may involve scattering, absorption, diffraction, and polarization.

A convenient starting point is to assume that the grains are spherical, span a range of sizes in comparison to the incident light, and have uniform composition (though the composition may differ for different grains). An extensively used theory for dust is called *Mie Theory* which folds the absorption and scattering characteristics of a spherical grain into a *complex* index of refraction, as described more fully in Appendix D.3. The result is a determination of how efficiently light interacts with the grain as a function of wavelength and grain size. This is measured by the *extinction efficiency factor*, $Q_{\text{ext}\lambda}$, of a dust grain (the sum of scattering and absorption components) as a function of $x = 2\pi a/\lambda$ where $Q_{\text{ext}\lambda}$ is the ratio of the effective to the geometric extinction cross-section of a grain, a is the grain radius and λ is the wavelength of the incident light. A plot of $Q_{\text{ext}\lambda}$, including its absorption and scattering parts, is shown in Figure D.5 for one specific grain composition. The strongest interaction occurs when the wavelength of the light is of order the grain size.

Using Eqs. (5.9), (5.8), and (D.30), the optical depth of a set of grains, assuming that there is no change in properties as a function of line-of-sight distance, is,

$$\tau_\lambda = n_{\text{d}}\, \sigma_\lambda\, l = n_{\text{d}}\, Q_{\text{ext}\lambda}\, \pi\, a^2\, l \tag{5.23}$$

where n_{d} is the number of grains per unit volume, σ_λ is the effective grain cross-section (equivalent to $C_{\text{ext}\lambda}$ in Appendix D.3) and l is the line-of-sight depth of the region.

Once $Q_{\text{ext}\lambda}$ is determined for a variety of different grains, it is then possible to calculate a theoretical extinction curve for various admixtures of grain sizes and compositions (see Example 5.8 for a simplified example, and Prob. 5.18) and compare the result with the observed extinction curve shown in Figure 3.21. The admixture which produces the closest fit is more likely to represent the dust that is actually present.

Example 5.8

Assuming that all grains are identical and possess the properties illustrated by Figure D.5, describe how Figure D.5 can be related to the observed extinction curve shown in Figure 3.21.

From Eqs. (1.13) and (3.29), the absorption in some waveband, λ, can be written,

$$A_\lambda = -2.5 \log\left(\frac{I_\lambda}{I_{\lambda 0}}\right) \tag{5.24}$$

where I_λ is the specific intensity of the object undergoing extinction and $I_{\lambda 0}$ is the specific intensity in the absence of extinction. Anticipating a result from Sect. 6.2 for the transmission of light through a medium that absorbs, but does not emit optical light (Eq. 6.9),

$$\frac{I_\lambda}{I_{\lambda 0}} = e^{-\tau_\lambda} \tag{5.25}$$

Combining the above two equations,

$$A_\lambda = 1.086\,\tau_\lambda \tag{5.26}$$

This result shows that an optical depth of one is approximately equivalent to an extinction of one magnitude[20] Using Eqn. (5.23),

$$A_\lambda = 1.086\,n_{\rm d}\,Q_{{\rm ext}\lambda}\,\pi\,a^2\,l \tag{5.27}$$

From this equation, Eq. (3.33), and our assumption that there is no change in dust properties along the line of sight,

$$\frac{E_{\lambda-{\rm V}}}{E_{\rm B-V}} = \frac{Q_{{\rm ext}\lambda} - Q_{\rm extV}}{Q_{\rm extB} - Q_{\rm extV}} \tag{5.28}$$

The only wavelength-dependent quantity of the right hand side of this equation is $Q_{{\rm ext}\lambda}$. Since we have taken the grain size to be constant, the abscissas of the $Q_{{\rm ext}\lambda}$ plot (Figure D.5) and the extinction curve (Figure 3.21) both represent $1/\lambda$. Therefore, if all of our assumptions are correct, then the *shape* of the curve of $Q_{{\rm ext}\lambda}$ (except for some scaling) should be the same as the shape of the extinction curve. A visual comparison indeed shows that there is some similarity. However, there are also significant differences, leading to the conclusion that all dust grains (if described by figure D.5) are not identical.

Problems

5.1 What would be the scattering cross-section of a photon from a free proton (see Eq. D.2)? In an ionized gas, what fraction of scatterings will be due to free protons in comparison to free electrons?

5.2 (a) From the relationships in Appendix D.1, show that the Einstein $A_{\rm j,i}$ coefficient between energy levels, j and i can be written,

$$A_{\rm j,i} = \frac{0.667}{\lambda^2_{\rm i,j}}\,\frac{g_{\rm i}}{g_{\rm j}}\,f_{\rm i,j} \tag{5.29}$$

[20] This would be true for any medium (dust or gas) that does not, itself, produce emission at the same wavelength as the incident light.

where $\lambda_{i,j}$ is the wavelength of the transition, g_j, g_i, are the statistical weights of the upper and lower states, respectively, and $f_{i,j}$ is the absorption oscillator strength. Verify that this relationship gives the Einstein $A_{j,i}$ coefficients of the Ly γ and H γ lines given in Table C.1.

(b) Show that the cross-section given by Eq. (D.12) can be written,

$$\sigma_s = \frac{\lambda_{1,2}^2}{2\pi}\left(\frac{g_2}{g_1}\right) \tag{5.30}$$

for the centre of the Ly α line in the absence of any other line broadening mechanisms, where $\lambda_{1,2}$ is the wavelength of the line centre. Verify that this result gives the natural bound–bound cross-section given in Table 5.1.

5.3 For a scattering process in which a Ly γ photon is absorbed by hydrogen followed by the emission of an H β and then Ly α photon, show that the absorbed energy is equal to the total emitted energy.

5.4 Consider resonance scattering from bound electrons in the hydrogen atom (Appendices D.1.2, D.1.3, and D.1.4) in which the line widths are given by the natural line width[21].

(a) Which of the spectral lines in Table C.1 has the widest frequency response to photon scattering?

(b) Under typical interstellar conditions, almost all neutral hydrogen atoms are in their ground states (see Sects. 3.4.5, 3.4.6). Assuming that there are an equal number of photons at all frequencies impinging upon a hydrogen atom in the ISM, which transition is the most probable and why?

5.5 Consider Rayleigh scattering from bound electrons (Sect. 5.1.1.3).

(a) In the Rayleigh scattering limit for any atom or molecule, how much greater (i.e. by what factor) will the scattering cross-section be for blue light in comparison to red light (Table G.5)?

(b) Calculate the total Rayleigh scattering cross-section, σ_R, for the hydrogen atom at the two wavelengths adopted in part (a) using data from Table C.1 and Eq. (D.24). Consider only the lines for which the Rayleigh limit is applicable. Determine the ratio of cross-sections for the two wavelengths and comment as to why there might be a difference with the results of part (a).

5.6 In Example 5.3, the mean free path of a photon in the Earth's atmosphere was found to be greater than 80 km. Above this altitude, the atmosphere is largely ionized (the ionosphere). From the information in Sect. 2.3, calculate the mean free path of a photon to Thomson scattering in the ionosphere. Is the ionosphere more or less transparent to optical photons than the atmosphere at sea level?

[21] Note that the natural line width is much narrower than observed in nature (see Sect. 9.3).

5.7 (a) What is the mean value of the increase in wavelength, $\overline{\Delta\lambda}$, for Compton scattering in an isotropic radiation field?

(b) Show that, after N Compton scatterings of equal angle, θ, each time, the final photon energy, E_N, can be written in terms of the initial photon energy, E_i, via,

$$E_\mathrm{N} = \frac{E_\mathrm{i}}{1 + \frac{NE_i}{\mathrm{m_e}\,c^2}(1 - \cos\,\theta)} \tag{5.31}$$

(c) If a photon of energy, 5.1 MeV is repeatedly Compton scattered at an angle of 60°, how many scatterings are required to reduce its energy to 0.51 MeV?

5.8 What is the mean free path of a 0.51 MeV photon to Compton scattering through (a) the intergalactic medium of mean density, $n = 10^{-4}\,\mathrm{cm}^{-3}$ and (b) through a Wolf–Rayet star of mass, $20\,\mathrm{M}_\odot$, and radius, $20\,\mathrm{R}_\odot$? Comment on how easily this gamma-ray photon can escape from the star and how easily it can travel through intergalactic space in the absence of other interactions.

5.9 Determine the Strömgren radius of the resulting H II region if the Sun were to enter an H I cloud of the same density as the H II region described in Example 5.4 i.e. let n_e after ionization $= n_\mathrm{H}$ before. Assume that the H II region temperature is also the same. Based on the result, qualify the statement: 'only O and B stars create H II regions'. [Hint. Either $B_\lambda(\mathrm{T})$ or $B_\nu(\mathrm{T})$ may be used. Consider whether either the Rayleigh–Jeans Law or Wien's Law might be used to simplify the integration.]

5.10 Suppose a spherical, pure H I primordial galaxy of uniform density is present in the dark ages (Sect. 5.2.2) before any stars or AGN exist. The galaxy has a mass and radius of $\mathrm{M_G} = 10^{10}\,\mathrm{M}_\odot$ and $R_\mathrm{G} = 10$ kpc, respectively. If an AGN of ionizing luminosity, $L_\mathrm{i} = 10^{46}$ erg s^{-1} suddenly 'turns on' at the centre of this galaxy, would the entire galaxy become ionized and, if so, would photons leak out into the intergalactic medium? If not, what fraction of the galaxy would be ionized? For simplicity, assume that all ionizing photons from the AGN are exactly at the energy required to ionize hydrogen.

5.11 Consider the development given in Appendix E and write the plasma frequency equation (Eq. E.6) for ions of charge, Ze, instead of electrons. Determine the ratio of the plasma frequency oscillation period of ions compared to electrons for a pure hydrogen gas.

5.12 (a) Compute the electron plasma frequency, ν_p, for a typical maximum density of $n_\mathrm{e} = 10^6\,\mathrm{cm}^{-3}$ in the Earth's ionosphere (Sect. 2.3).

(b) Look up the FM, AM, and shortwave radio band ranges from some convenient reference. Indicate which frequencies in these bands, emitted by a ground-based transmitter, could reflect off of the ionosphere and therefore be detected by a radio receiver that is over the physical horizon.

(c) Referring to Figure E.2 of a Type III radio burst, estimate the maximum and minimum electron densities in the plasma between the Sun and the Earth encountered by this Solar flare.

(d) Indicate over what range of frequencies the radio burst of part (c) could have been detected from the Earth's surface. How much of the emission shown in Figure E.2 could have been detected from the Earth's surface?

5.13 What would be the angular size (arcsec) subtended by the smallest interstellar cloud described in Sect. 5.3 if it were at the fairly nearby distance of 50 pc? Determine how large a radio telescope (let $\lambda = 20$ cm) would be required to image such a cloud directly. Comment on the technique of using interstellar scintillation for probing these size scales in comparison to direct imaging.

5.14 Compute optical depths (see Table 5.1) for the following cases and indicate whether the cloud or medium is optically thick or optically thin:

(a) 5.1 MeV photons travelling 100 Mpc through intergalactic space of density of $n_e = 10^{-4}\,\mathrm{cm}^{-3}$.

(b) An incident beam containing 13.6 eV photons impinging on an ISM H I cloud of density, $n_H = 1\,\mathrm{cm}^{-3}$ and thickness, $l = 1$ pc.

(c) Incident photons at $\lambda 500$ nm in an upper layer of a star that is 100 km thick with a density, $\rho = 10^{-8}\,\mathrm{g\,cm}^{-3}$ and a total mass absorption coefficient of $\kappa_{500\,\mathrm{nm}} = 0.83\,\mathrm{cm}^2\,\mathrm{g}^{-1}$.

5.15 Consult Table G.11 and explain why the Sun's limb appears sharp to the unaided eye (assuming no glare as would be the case if the Sun is viewed behind a thin, partially transparent cloud).

5.16 Using the Thomson scattering cross-section, $\sigma_T = \sigma_{es}$, derive Eq. (5.18). Eqs. (5.8) and (3.14) will be useful.

5.17 From the information given in Table G.10 or Figure G.2, estimate κ_{bf}, κ_{ff}, and κ_{es} at positions, (a) $R = 0.1\,R_\odot$, (b) $R = 0.5\,R_\odot$, and (c) $R = 0.9\,R_\odot$, within the Sun. Adopt $f_{bf} = 0.1$ and $g_{ff} = 1$. Comment on the relative importance of these three processes (bound–free absorption, free–free absorption, and electron scattering) as a function of depth.

5.18 Estimate $\frac{E_{\lambda-V}}{E_{B-V}}$ for a wavelength, $\lambda 210$ nm, assuming that all dust grains are identical with radii of $a = 0.1$ μm and have the properties shown in Figure D.5 (see Example 5.8). Compare the result to the observed value shown in Figure 3.21. Estimate how much this would change if all grains had radii a factor of 2 larger and comment on the results.

6
The Signal Transferred

Behold! Only 34 circles are required to explain the entire structure of the universe and the dance of the planets!

–Nicolaus Copernicus in *Commentariolus* (Ref. [64]).

6.1 Types of energy transfer

The fact that we see light in the Universe means that the energy contained in that light – a tiny fraction of what was actually generated – has been transported from its point of origin to our eyes or our telescopes. With no intervening matter, the energy transferred is just that contained in photons diminished by distance (e.g. Eq. 1.9). With intervening matter, as we have seen, photon–particle or particle–particle interactions[1] can transfer or 'share' energy. We have so far considered how such interactions might divert a photon out of (or possibly into) the line of sight. We now wish to consider the net transfer of energy *along a path* as a beam of light moves through matter. This will allow us to relate the interactions that are occurring at a microscopic level to the light that we actually observe. In the process, we can obtain information about the material within which energy transport is occurring.

Energy can exist in several forms. *Kinetic energy*, which is the energy of motion, is called 'heat' when applied to random particle motions as discussed in Sect. 3.4.1. The energy of a photon and the energy equivalent of matter are given by $E = h\nu$ and $E = \mathrm{m}\,c^2$, respectively (Table I.1). Example 3.1 used the latter expression to explain the luminosity of the Sun. *Potential energy* exists only with reference to a 'field'. In a gravitational field, for example, a higher object 'contains' more gravitational potential energy than a lower one. If let go, the potential energy of the object will be converted into the kinetic energy of motion. In an electric field, two like charges near each other

[1]As mentioned in Sect. 5.1.2.2, photon–photon collisions are also possible, but they are a less common mechanism for transferring energy in astrophysics.

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

contain more electrical potential energy than two like charges separated by a large distance. Again, if let go, the potential energy will be converted to kinetic energy as the like charges repel each other. An emitted photon resulting from the transition of an electron from a higher quantum state to a lower one can be thought of as the conversion of electrical potential energy into radiative energy. In discussing energy transport, therefore, it is useful to keep in mind the kinds of energy exchanges that can occur.

The transport of energy along a path through matter can occur in four ways: *(a) thermal conduction, (b) convection, (c) radiation,* and *(d) electrical conduction*. The first three are routinely cited as heat transport mechanisms in non-astronomical objects. *Thermal conduction* involves collisions between particles, and therefore the sharing of particle kinetic energy, with no net bulk motion of the material. If one side of a material is hot and the other cold, eventually collisions between adjacent particles will cool the hot side and warm the cool side. *Convection* involves the physical motion of a fluid (liquid or gas) and only exists in the presence of a gravitational field. In a gas, hotter pockets ('cells') move upwards into cooler regions via buoyancy and start a circulation or 'boiling' motion[2]. Energy is thus transported by the physical motion of hot gas upwards. *Radiation* is taken to mean the transport of energy via emission and eventual absorption again of photons, as we considered at length in Chapter 5. *Electrical conduction* refers to the motion of electrons through a medium in which the nuclei are fixed. It is rarer in astrophysics but it is an important mechanism for transporting energy in white dwarfs. Electrical currents can also exist in some regions in which there are strong magnetic fields, such as regions near Sunspots.

In our own Sun (see Figure 3.15), radiative transfer and convection are the important mechanisms for transporting energy from the core to the surface. In principle, thermal conduction can also occur, but its efficiency is low in comparison to the other mechanisms. Electrical conduction will not be important until the Sun evolves off the main sequence and the Solar core becomes dense enough to be considered a white dwarf precursor (Sect. 3.3.2). Presently, radiative diffusion dominates out to about 70 per cent of the Solar radius while, in the outer 30 per cent, energy is primarily transferred via convection which dominates when the temperature *gradient* is high[3]. The *granulation* seen at the surface of the Sun (Figure 6.1) is a beautiful illustration of convective cells. In other stars, the same energy transfer mechanisms are occurring although the interior locations and relative importance of radiation and convection may differ. Stars of lower mass than the Sun have deeper convective zones, for example, whereas high mass stars may have no convection in the outer layers, but instead have convective cores.

Radiative transfer is arguably the most important energy transport mechanism in astrophysics. It involves the kinds of absorptions, scatterings, and re-emissions of

[2]Convection is a complex process that does not lend itself well to a detailed mathematical description. An approach that considers how far a particular cell will rise before its temperature matches its surroundings, and that calculates the resulting transfer of energy is called *Mixing Length Theory*. See Ref. [33] for more details.

[3]More specifically, convection occurs when $d\ln(\mathrm{P})/d\ln(\mathrm{T}) < \gamma/(\gamma-1)$, where P is the pressure, T is the temperature, and $\gamma \equiv C_\mathrm{P}/C_\mathrm{T}$ is the ratio of specific heats, where C_P and C_T are the amount of heat required to raise the temperature of 1 g of material 1 K held at constant pressure, and constant temperature, respectively.

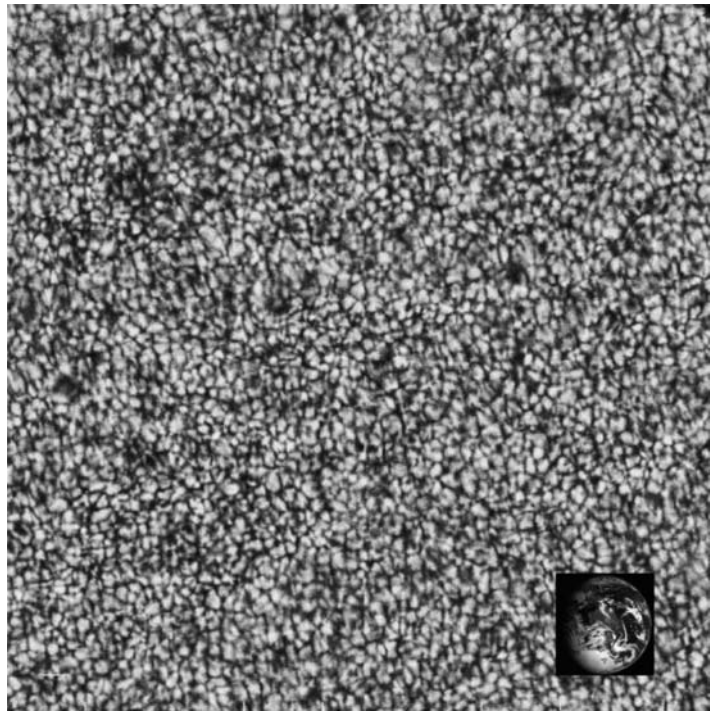


Figure 6.1. This granulation pattern illustrates convection in the outer layers of the Sun. Brighter regions are hotter and darker regions are cooler. This section is 91 000 km × 91 000 km in size and the image was taken at a wavelength of λ403.6 nm using the Vacuum Tower Telescope on Tenerife in the Canary Islands. (Reproduced by permission of W. Schmidt, Kiepenheuer-Institut für Sonnenphysik) *Inset:* The Earth, to scale, as taken from the Galileo Spacecraft. (Reproduced by permission of NASA/Goddard Space Flight Centre)

photons that have been described in some detail in Chapter 5 and earlier with respect to photons traversing the interior of the Sun (Sect. 3.4.4). Beyond its importance in stars, this is also how radiation from any background source transfers through intervening material that may lie along its path. This includes the signal from a quasar in the distant Universe, light from an interstellar cloud, or starlight passing through the Earth's atmosphere. Moreover, the intervening material itself may not only be a source of opacity, but also a source of radiation. In the next section, then, we focus on radiative transfer which takes into account the propagation of a signal through absorbing, scattering and emitting matter.

6.2 The equation of transfer

Let us consider an incoming signal of specific intensity, I_ν, passing through a cloud. By 'cloud', we mean any gaseous region containing matter, including sections of larger

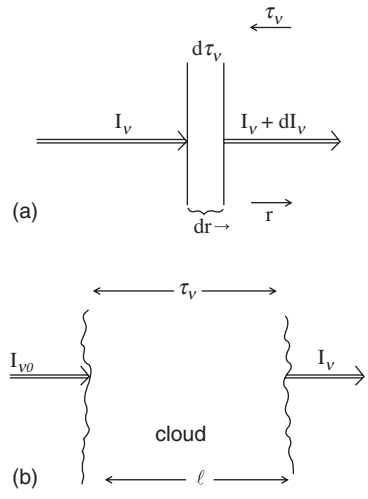


Figure 6.2. **(a)** Illustration of the change in intensity of a beam of light as it passes through material of infinitesimal thickness, *dr*, or optical thickness, $d\tau_\nu$. The outgoing intensity may be higher or lower than the incident intensity. **(b)** As in (a), but showing the background incident beam and the total thickness, *l*, or total optical thickness, τ_ν, of the cloud. The development of Sect. 6.2 does not require the beam to be perpendicular to the cloud, as shown here.

clouds, layers in the interiors of stars or their atmospheres, or whatever other gaseous region is of interest. The cloud has some total thickness, *l*, and the direction, *r* (see Figure 6.2) is in the same direction as the incoming beam. As the beam passes through a small thickness of width, *dr*, its specific intensity will change from I_ν to $I_\nu + dI_\nu$, where,

$$dI_\nu = dI_{\nu\,\text{loss}} + dI_{\nu\,\text{gain}} \tag{6.1}$$

The first term includes all scattering and absorption processes that remove photons from the line of sight at the frequency being observed (from now on, referred to collectively as 'absorptions'). The second term contains all processes that add photons to the beam, such as all emission processes from the matter itself at the frequency being observed[4]. Note that dI_ν could be positive or negative. Using the absorption coefficient, α_ν (cm^{-1}) introduced in Eq. (5.8), and introducing the *emission coefficient*, j_ν ($\text{erg s}^{-1}\,\text{cm}^{-3}\,\text{Hz}^{-1}\,\text{sr}^{-1}$), Eq. (6.1) can be written,

$$dI_\nu = -\alpha_\nu I_\nu\, dr + \text{j}_\nu\, dr \tag{6.2}$$

Recall that the absorption coefficient represents the mean number of absorptions per centimetre through a material that would be experienced by photons (the inverse, $1/\alpha_\nu$, being a photon's mean free path). Therefore, if there are more incident photons, there will be more absorptions, which is why the loss term is a function of I_ν. The emission coefficient, on the other hand, is a quantity that represents all

[4]We ignore scatterings into the line of sight from other angles in this development.

radiation-generating mechanisms per unit volume of material and it does not depend on the strength of the incident beam[5].

Rearranging Eq. (6.2), using Eq. (5.8) and recalling that r increases in the direction of the incident beam while τ_ν increases into the cloud from the point of the view of the observer, we arrive at two forms of the *Equation of Radiative Transfer*,

$$\frac{dI_\nu}{dr} = -\alpha_\nu I_\nu + \mathrm{j}_\nu \qquad \text{(Form 1)} \tag{6.3}$$

$$\frac{dI_\nu}{d\tau_\nu} = I_\nu - \frac{\mathrm{j}_\nu}{\alpha_\nu} = I_\nu - \mathrm{S}_\nu \quad \text{(Form 2)} \tag{6.4}$$

where

$$\mathrm{S}_\nu \equiv \mathrm{j}_\nu/\alpha_\nu \tag{6.5}$$

is called the *Source Function* and has units of specific intensity ($\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}\,sr^{-1}}$). The Source Function is a convenient way of folding in all the absorptive and emissive properties of the cloud into one term.

The solution of the equation of radiative transfer links the observed intensity, I_ν, to the properties of the matter, since τ_ν contains information on the type and density of the intervening material through Eq. (5.8). This link between the observed light and the physical properties of matter is at the heart and soul of astrophysics and provides the key to deciphering the emission that is seen. We will now lay the groundwork for this process by showing the solution for some specific cases. Since Eqs. (6.3) and (6.4) are equivalent, we will adopt whichever form is most convenient. It is also helpful to remember that, for a single physical situation, the Equation of Transfer could be written many times, once for every frequency for which data are available (although frequency-averaging is sometimes carried out as described in Sect. 5.4.1). For example, matter that 'absorbs only' at one frequency could 'emit only' at a different frequency (e.g. dust at optical and infrared wavelengths, respectively, see the Sombrero Galaxy shown in Figure 3.22).

6.3 Solutions to the equation of transfer

6.3.1 Case A: no cloud

If there is no intervening cloud, then there is no absorption or emission, i.e. $\alpha_\nu = \mathrm{j}_\nu = 0$. Using Eq. (6.3),

$$\frac{dI_\nu}{dr} = 0 \Rightarrow I_\nu = \text{constant} \tag{6.6}$$

[5] An exception is *stimulated emission* which is a process by which an incoming photon causes an atom to de-excite without the absorption of the incident photon. This is an emission process but, since it depends on the strength of the background radiation, if it is occurring, it can be incorporated into the loss term as a correction factor. It is typically important only in very strong radiation fields.

Thus we come to the same conclusion as first presented in Sect. 1.3, that is, the specific intensity is constant and independent of distance in the absence of intervening matter.

6.3.2 Case B: absorbing, but not emitting cloud

If the cloud absorbs background radiation at some frequency, but does not emit at the same frequency then (Eq. 6.4) $\mathrm{j}_\nu = \mathrm{S}_\nu = 0$ and,

$$\frac{dI_\nu}{d\tau_\nu} = I_\nu \tag{6.7}$$

Integrating from the front to the rear of the cloud,

$$\int_{I_\nu}^{I_{\nu 0}} \frac{dI_\nu}{I_\nu} = \int_0^{\tau_\nu} d\tau_\nu \tag{6.8}$$

where $I_{\nu 0}$ is the initial specific intensity when the beam first enters the cloud (the background intensity), I_ν is the emergent intensity on the near side of the cloud, and τ_ν is the total optical depth of the cloud along the line of sight being considered. The solution is,

$$I_\nu = I_{\nu 0}\, e^{-\tau_\nu} \tag{6.9}$$

We see that the intensity diminishes exponentially with optical depth.

This equation was used to describe the extinction of optical light when travelling through a dusty medium (Example 5.8). It is also relevant to the opacity of the Earth's atmosphere which is an absorbing, but not emitting medium at optical wavelengths[6]. Thus, Eq. (6.9) allows us to quantify the atmospheric reddening that was discussed in Sects. 2.3.5 and 5.1.1.3 (see Prob. 6.1).

6.3.3 Case C: emitting, but not absorbing cloud

Using Eq. (6.3) with $\alpha_\nu = 0$, we find,

$$I_\nu = I_{\nu 0} + \int_0^l \mathrm{j}_\nu\, dr \tag{6.10}$$

Unless we are modelling the cloud in some detail, in many cases an assumption that the emissive properties of the cloud do not change along a line of sight (i.e.

[6]Since the atmosphere does show some optical emission lines, this development applies to frequencies that are off the line frequencies.

j_ν independent of r) is adequate for estimating the needed physical parameters of the cloud. For such cases, Eq. (6.10) can be written,

$$I_\nu = I_{\nu 0} + j_\nu \, l \tag{6.11}$$

In either case, an expression for j_ν is required in order to relate the emission to the source parameters. Expressions for j_ν for several emission processes are provided in Chapter 8.

6.3.4 Case D: cloud in thermodynamic equilibrium (TE)

In thermodynamic equilibrium (Sect. 3.4.4), the radiation temperature is constant and equal to the kinetic temperature, T, everywhere in the cloud. The specific intensity, I_ν, is therefore given by the Planck function, $B_\nu(\mathrm{T})$ (see Sect. 4.1) at every point in the cloud. Thus, there is no intensity gradient ($dI_\nu/dr = 0$, I_ν = constant) and (Eq. 6.3) becomes,

$$0 = -\alpha_\nu B_\nu(\mathrm{T}) + j_\nu \Rightarrow B_\nu(\mathrm{T}) = \frac{j_\nu}{\alpha_\nu} \tag{6.12}$$

Eq. (6.12) is known as *Kirchoff's Law* and is another way of stating that emissions and absorptions are in balance in such a cloud[7]. By the definition of the Source Function (Eq. 6.5), for a cloud in TE we therefore have,

$$I_\nu = S_\nu = B_\nu(\mathrm{T}) \tag{6.13}$$

If such a cloud existed in reality, a photon that impinged on it would simply be absorbed and 'thermalized' to the temperature of the cloud. However, as indicated in Sect. 3.4.4, a cloud in true thermodynamic equilibrium is difficult to find. The best example is the cosmic background radiation.

6.3.5 Case E: emitting and absorbing cloud

When the cloud is both emitting and absorbing, the general solution for a case in which the emission and absorption properties do not change along a line of sight[8] is (Eq. 6.4),

$$I_\nu = I_{\nu 0}\, e^{-\tau_\nu} + S_\nu \left(1 - e^{-\tau_\nu}\right) \tag{6.14}$$

[7] More accurately, $\frac{j_\nu}{\alpha_\nu} = {n_\nu}^2 B_\nu(\mathrm{T})$, where n_ν (≈ 1) is the index of refraction.
[8] See Ref. [143] for the derivation of this result and also for the case in which the properties change with line-of-sight distance.

This expression allows us to see the various contributions to the observed specific intensity. The first term on the RHS is identical to Eq. (6.9) and indicates that a background signal will be attenuated exponentially by a foreground cloud of optical depth, τ_ν. The second term includes two contributions. S_ν is the added specific intensity from the cloud itself, and $-\mathrm{S}_\nu\, e^{-\tau_\nu}$ accounts for the cloud's absorption of its own emission.

Let us now consider the two extremes of optical depth: opaque $(\tau_\nu \gg 1)$ and transparent $(\tau_\nu \ll 1)$. Then Eq. (6.14) reduces to,

$$I_\nu = \mathrm{S}_\nu \qquad (\tau_\nu \gg 1) \tag{6.15}$$

$$I_\nu = I_{\nu 0}\,(1 - \tau_\nu) + \mathrm{S}_\nu \tau_\nu \qquad (\tau_\nu \ll 1) \tag{6.16}$$

$$= I_{\nu 0}\,(1 - \tau_\nu) + \mathrm{j}_\nu\, l \tag{6.17}$$

where we have used an exponential expansion (Eq. A.3) for the case, $\tau_\nu \ll 1$. In Eq. (6.17), l is the line-of-sight distance through the cloud and we have used the definition of the Source Function (Eq. 6.5) and the equations relating optical depth to the absorption coefficient (Eqs. 5.8, 5.9), assuming there is no variation in these quantities through the cloud.

6.3.6 Case F: emitting and absorbing cloud in LTE

In LTE, as described in Sect. 3.4.4, the temperature of a pocket of gas is approximately constant over a mean free photon path (see Example 3.5 for the interior of the Sun) and therefore, over this scale, the emissions and absorptions must be in balance. We therefore associate the Source Function with the Planck function,

$$\mathrm{S}_\nu \equiv \frac{\mathrm{j}_\nu}{\alpha_\nu} = B_\nu(\mathrm{T}) \tag{6.18}$$

Since there is a net flux of radiation through the cloud, however, I_ν is not constant so there is a non-zero intensity gradient. Then following Eq. (6.14),

$$\boxed{I_\nu = I_{\nu 0}\, e^{-\tau_\nu} + B_\nu(\mathrm{T})\,(1 - e^{-\tau_\nu})} \tag{6.19}$$

The above equation is, arguably, the most important equation in radiation astrophysics and can be applied to any *thermal cloud* (i.e. one whose emission in some way depends on its temperature), provided the gas is in LTE. As indicated in Sect. 3.4.4, moreover, even if a cloud is not in LTE, sometimes an assumption of LTE is a useful starting point and can provide initial, approximate parameters for the cloud. We therefore explore some of the implications of the LTE solution in the next section.

6.4 Implications of the LTE solution

6.4.1 Implications for temperature

Eq. (6.19) leads to some important conclusions in specific circumstances. Following the development in Sect. 6.3.5, we can write down the result for the optically thick and optically thin extremes,

$$I_\nu = B_\nu(\mathrm{T}) \qquad (\tau_\nu \gg 1) \tag{6.20}$$

$$I_\nu = I_{\nu 0}\,(1 - \tau_\nu) + B_\nu(\mathrm{T})\,\tau_\nu \qquad (\tau_\nu \ll 1) \tag{6.21}$$

$$= I_{\nu 0}\,(1 - \tau_\nu) + \mathrm{j}_\nu\, l \tag{6.22}$$

Eq. (6.20) is a confirmation that an optically thick cloud in LTE emits as a black body. A background source, if present, does not contribute to the observed specific intensity (although it might contribute to heating the cloud) since the foreground cloud completely blocks its light. Stars are good examples of this case (Sect. 4.1). *The only information obtainable from an optically thick source in LTE is the temperature* since the Planck curve is a function of temperature only. We can see into such a source only to a depth of the mean free path which is less than the depth of the source itself (see Figure 4.2). We have no way of obtaining any information from the far side of an optically thick cloud and therefore no way of obtaining other information about it, such as its density.

On the other hand, Eqs. (6.21) and (6.22) show that, for an optically thin source, the observed intensity depends on the intensity of the background source, if present, since the background source can be seen through the cloud. Moreover, Eq. (6.21) indicates that, *for an optically thin source, the observed intensity depends on* both *the temperature and density of the foreground cloud* because $B_\nu(\mathrm{T})$ is a function of temperature and τ_ν is a function of density via Eq. (5.8) or (5.9). The signal that we detect has probed every depth through the cloud in such a case. The emission coefficient of Eq. (6.22) will also be a function of both density and temperature (see, for example, Sect. 8.2).

If we now consider a cloud *without* a background source ($I_{\nu 0} = 0$), the observed specific intensity is the Planck intensity for an optically thick source (Eq. 6.20) and the observed specific intensity is the Planck intensity *diminished by* the optical depth for an optically thin source (Eq. 6.21). Therefore, regardless of the value of τ_ν, it will always to be true that,

$$I_\nu \leq B_\nu(\mathrm{T}) \tag{6.23}$$

In Sect. 4.1.1, we found that the specific intensity of any signal, I_ν, can be represented by a brightness temperature (Eq. 4.4) and similarly the Planck curve can be represented by the kinetic temperature. Therefore, Eq. (6.19) could be written with I_ν and $I_{\nu 0}$ expressed as functions of $\mathrm{T}_{B\nu}$ and $\mathrm{T}_{B\nu 0}$, respectively, and $B_\nu(\mathrm{T})$ could be written as a

function of T. As temperature increases, so does the respective specific intensity. Then, following Eq. (6.23) for the case with no background source, we find,

$$T_B \leq T \tag{6.24}$$

Eq. (6.24) is a simple, yet important result because it means that, for sources in LTE without background emission, *the observed brightness temperature places a lower limit on the temperature of the source*. The brightness temperature can be determined for any source by simply inserting the observed value of I_ν into Eq. (4.4). Once T_B is known, the true kinetic temperature must be the same or higher.

6.4.2 Observability of emission and absorption lines

Whether a spectral line is observed in absorption or emission in the presence of a background source is an important clue as to whether the region in which the line formed is hotter or colder than the background. This can be determined from Eq. (6.19), as Example 6.1 outlines.

Example 6.1

Under what conditions will a spectral line be seen in absorption or emission?

We will take a simple case in which a foreground cloud is capable of interacting with the background signal and therefore forming a spectral line at only one specific frequency (with a small line width) as Figure 6.3 illustrates. The cloud is perfectly transparent to the background signal at all frequencies off the line. If the specific intensity at the line frequency is less than the value of the background, then there is an *absorption line* as in Figure 6.3.a, and if the line specific intensity is greater than the background, then an *emission line* is seen as in Figure 6.3.b. Taking the absorption case and using Eq. (6.19) for the frequency of the line centre,

$$I_\nu = I_{\nu 0}\, e^{-\tau_\nu} + B_\nu(T)\,(1 - e^{-\tau_\nu}) < I_{\nu 0} \tag{6.25}$$

where $I_{\nu 0}$ is the specific intensity of the background continuum, $B_\nu(T)$ is the Planck function at the temperature of the cloud, and τ_ν is the optical depth of the cloud at the frequency of the line centre[9]. Rearranging this equation and repeating the process for an emission line leads to,

$$B_\nu(T) < I_{\nu 0} \qquad \text{(absorption line)} \tag{6.26}$$

$$B_\nu(T) > I_{\nu 0} \qquad \text{(emission line)} \tag{6.27}$$

[9]Note that, since a spectral line has some intrinsic width (see Sect. 9.3) the optical depth varies over the line.

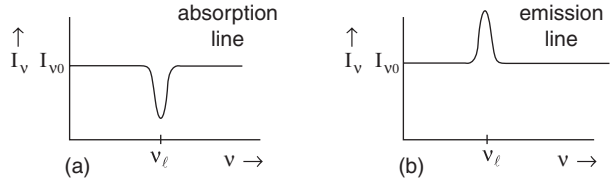


Figure 6.3. In this example, a cloud, such as is sketched in Figure 6.2, is capable of forming a single spectral line at a frequency, ν_l. If the line is lower than the background level, $I_{\nu 0}$, it is an absorption line **(a)** and if it is higher than the background, it is an emission line **(b)**. Eqs. (6.28) and (6.29) give the conditions for the two cases. Note that the *x* axis could be plotted as a function of frequency, wavelength, or velocity (for the latter, see Eqs. 7.3 and 7.4).

From the Planck function and the definition of brightness temperature, this becomes,

$$\mathrm{T} < \mathrm{T}_{B\nu 0} \qquad \text{(absorption line)} \tag{6.28}$$

$$\mathrm{T} > \mathrm{T}_{B\nu 0} \qquad \text{(emission line)} \tag{6.29}$$

Therefore, if the temperature of the foreground cloud is less than the brightness temperature of the background source, then the line will be seen in absorption (Figure 6.3.a) and if the temperature of the foreground cloud is greater than the brightness temperature of the background source, then the line will be seen in emission (Figure 6.3.b). If there is *no* background source, then effectively, $\mathrm{T}_{B\nu 0} \approx 0$ and the line is always seen in emission even if the cloud is cold[10]. Also, if $\mathrm{T} = \mathrm{T}_{B\nu 0}$, then *no line will be seen*.

The Sun's lower atmosphere is an environment within which a rich array of absorption lines are formed, as can be seen in the Solar spectrum of Figure 6.4. These lines are called *Fraunhofer lines* and the fact that they are seen in absorption, rather than emission, indicates that they are formed in a region that is cooler than the background[11]. The structure and properties of the Solar atmosphere are shown in Figure 6.5. and details of the Solar *photosphere* are given in Table G.11. The photosphere is a region from just below the surface (the surface is taken to be where the optical depth is unity[12]) to a height of about 500 km at which the temperature is lowest. Above the photosphere is the *chromosphere* ('sphere of colour' after its first identification as a thin reddish ring during Solar eclipses, see Figure 6.6) which is the region in which the temperature rises again, but hydrogen still remains predominantly neutral. The chromosphere also contains much activity such as *prominences* which are denser features, often loop-like, that are seen on the limb of the Sun and are related to the magnetic field structure. As Figure 6.5.a suggests, it is in the upper photosphere and chromosphere that absorption lines will form, the hotter background radiation being supplied by the Sun's surface whose spectrum is given by the Planck curve.

[10]Actually the minimum value is $\mathrm{T}_{B\nu 0} = 2.7\,\mathrm{K}$ which is set by the CMB (Figure 4.3).
[11]A more accurate comparison would take into account departures from LTE that occur in these regions.
[12]More accurately, a stellar 'surface' occurs at the depth from which most photons originate, on average. This actually corresponds to an optical depth of $\tau = 2/3$ (see Ref. [33] for details). For the Sun, however, the difference is only 12 km and will be neglected. Sometimes the word 'photosphere' is used to denote just this surface, rather than the region shown in Figure 6.5.

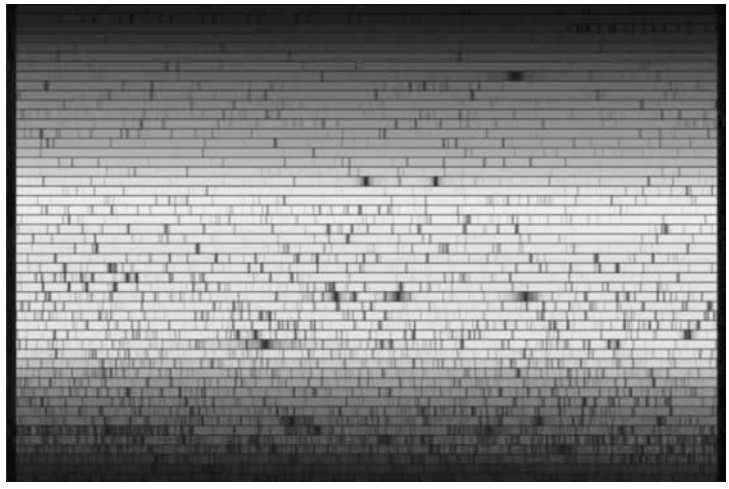


Figure 6.4. This spectrogram reveals the Sun's Fraunhofer lines, shown as dark vertical bands superimposed on background colour. The wavelength increases from left to right along each strip which covers 60 Å. There are 50 strips with increasing wavelength from bottom to top, for a total spectral coverage from 4000 Å to 7000 Å. An intensity trace across these strips would result in a spectrum similar to that shown in Figure 4.5 centre (with proper calibration of the background brightness). This image has been created from a digital atlas based on observations taken with the McMath–Pierce Solar Facility on Kitt Peak National Observatory in Arizona. (Reproduced by permission of NOAO/AURA/NSF) (see colour plate)

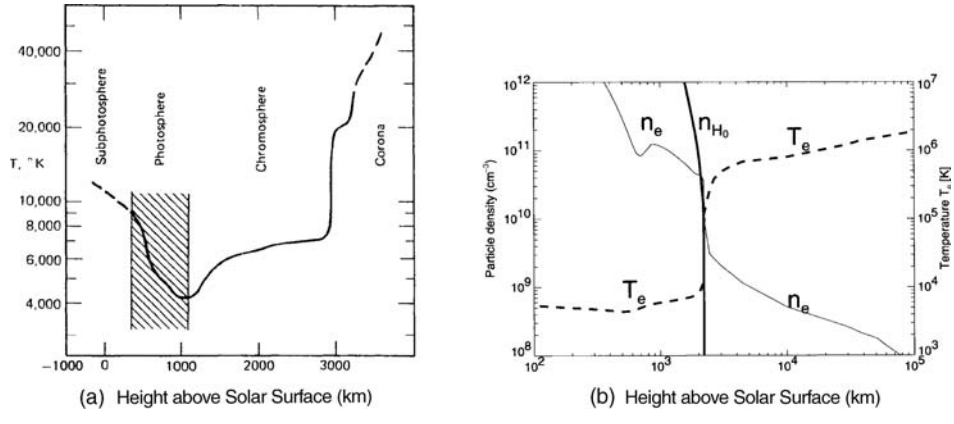


Figure 6.5. Plots of physical conditions in the Sun's atmosphere as a function of height. A height of 0 corresponds to the Solar surface which we take to be at $\tau = 1$. **(a)** Temperature as a function of height, with labelling showing the parts of the Sun from below the photosphere to the lower corona. (Adapted from Ref. [68]). **(b)** Density and temperature as a function of height over a larger region from the upper photosphere to the corona. Note the change in hydrogen ionisation state at a height of about 2000 km called the *transition region*. This marks the boundary between the chromosphere (hydrogen mostly neutral) and the corona (hydrogen mostly ionised). (Reproduced by permission of M. J. Aschwanden)

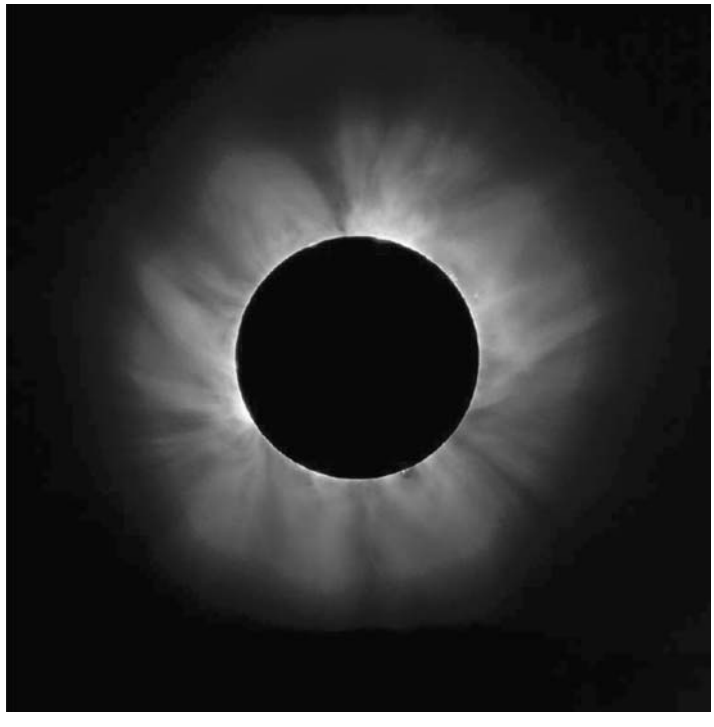


Figure 6.6. This Solar eclipse image shows us glimpses of the reddish narrow chromosphere as well as the extended hot corona with its delicate filaments. The discrete features at the Solar limb are prominences and are reddish because of H α emission. The optical corona is too faint to be seen in normal daylight and is only visible by eye during an eclipse. (Reproduced by permission of L. Viatour and the GNU Free Documentation License) (see colour plate)

The Sun also provides us with a good example of emission lines. As the temperature of the Solar atmosphere continues to rise with height, hydrogen becomes more fully ionised, marking the boundary between the chromosphere and Solar *corona*. The corona has a typical temperature of $\approx 10^6$ K but can be an order of magnitude higher or more during *Solar flares*, which are extremely high energy (up to 10^{32} erg) outbursts lasting minutes to hours. The corona extends well beyond the photosphere, as Figure 6.6 shows. With temperatures higher than the Solar surface, any spectral lines in the corona would be seen in emission above the black body spectrum from the surface. Since the density is low, however, these emission lines are weak. Optical emission lines, for example, do not stand out sufficiently in comparison to the background photospheric planck curve to be seen. Instead they are more easily observed at positions in the corona outside of the Solar disk where there is no background. At such high temperatures, however, emission lines of highly ionized species are present and these can occur at UV or X-ray wavelengths. Figure 8.11, for example, shows a UV emission line image of a large *coronal loop*. X-ray emission lines are also observed and can be seen against the Sun's disk, since the Sun's black body photospheric emission is negligible, in comparison, at X-ray wavelengths. Figure 6.7 shows an emission line from a Solar flare, although note that other continuum-emitting processes are also present in the spectrum.

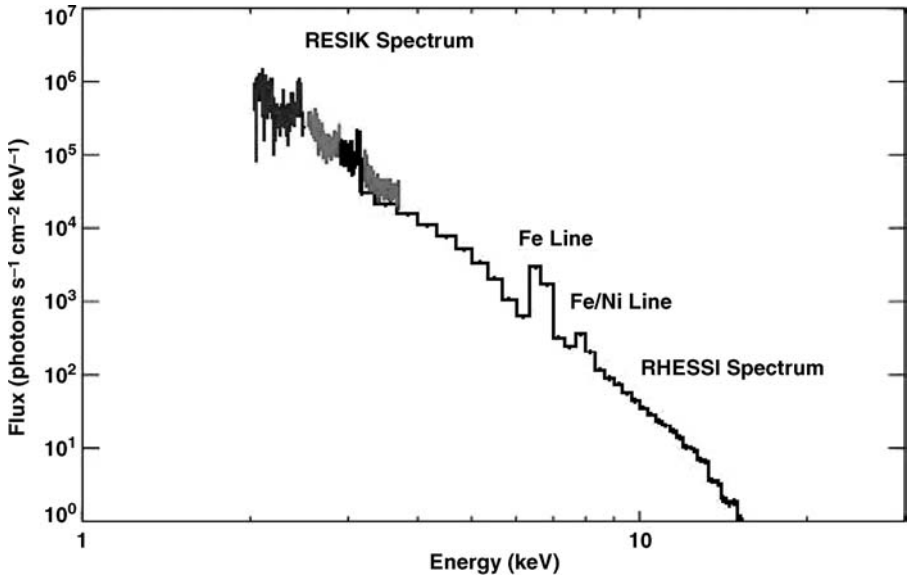


Figure 6.7. Two X-ray emission features at 6.7 and 8.0 keV are seen in this X-ray spectrum of a Solar flare. Each feature consists of a number of spectral lines that are unresolved, mainly from FeXXIV, FeXXV, FeXXVI, and some ionised Ni. The underlying continuum spectrum is not from Solar black body radiation which is very weak at X-ray wavelengths, but rather from optically thin Bremsstrahlung and free–bound recombination radiation (see Sects. 8.2, 8.3). This spectrum was obtained by the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) telescope, launched in 2002 (NASA). To the left are data from the RESIK spectrometer aboard the Russian CORONAS-F Solar telescope launched in 2001. (Reproduced by permission of K. J. H. Phillips and the RHESSI team)

As in the Sun, the spectra of other stars also provide us with important information about their physical parameters. Absorption lines, in particular, are observed in the spectra of other stars, as Figure 6.8 shows. The change in absorption line strength with stellar spectral type can be understood in terms of a temperature sequence, as quantified in Tables G.7, G.8, and G.9. An analysis of the strengths of various lines and their ratios further leads to tighter constraints on the temperature and density of stellar atmospheres (see Sect. 9.4.2). Thus, absorption lines provide a wealth of information about stars, including the elemental abundances discussed in Sect. 3.3.4.

6.4.3 Determining temperature and optical depth of HI clouds

As indicated in Sect. 3.4.5, any interstellar cloud that is neutral has all hydrogen particles in the ground state. Therefore, the only detectable spectral line from hydrogen in a neutral cloud is from the spin-flip transition of the ground state (see Appendix C.4) resulting in the $\lambda 21$ cm HI line. The lifetime of the upper state of this transition is very

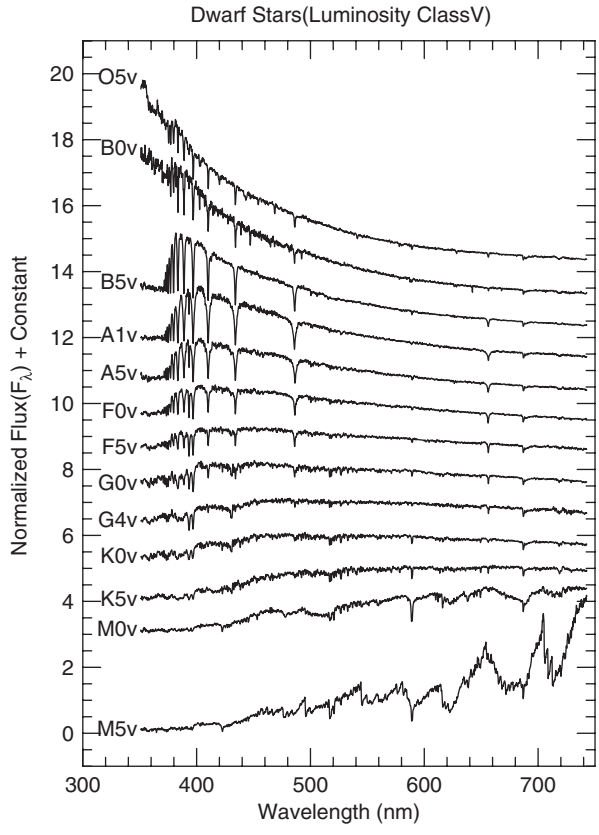


Figure 6.8. This plot shows a comparison of spectra for various types of stars of luminosity class, V (see Table G.7 for related stellar data). The shape of the background emission in each case roughly follows the shape of the Planck curve corresponding to the temperature of the star's photosphere. The locations and depths of the absorption lines provide a unique 'fingerprint' for the types and amounts of elements that are present. The Balmer lines, H α (λ656.3 nm), H β (λ486.1 nm) and H γ (λ434.0 nm) can be easily seen and, in this plot, are strongest at spectral type, A1V. (Reproduced by permission of Richard Pogge, with data from Jacoby, G.H., Hunter, D.A., and Christian, C.A., 1984, ApJS, **56**, 257)

long, of order 10^7 years, so the line is considered to be *forbidden*, as described in Appendix C.4, and transitions are more likely to result from collisions (Prob. 6.4a). The fact that collisions are driving the line excitation means that LTE applies for this line, though not for the atom as a whole (see Sects. 3.4.4, 3.4.5, and 9.2) so the LTE solution to the Equation of Transfer may be used. Moreover, the λ21 cm line is always observed in the Rayleigh–Jeans limit ($h\nu \ll k$T, Prob. 6.4b). Thus the specific intensity is directly proportional to the brightness temperature (Sect. 4.1.2) and we can rewrite Eq. (6.19) as,

$$\mathrm{T}_{B\nu} = \mathrm{T}_{B\nu 0}\, e^{-\tau_\nu} + \mathrm{T}\,(1 - e^{-\tau_\nu}) \qquad \text{(with background source)} \qquad (6.30)$$

$$\mathrm{T}_{B\nu} = \mathrm{T}\,(1 - e^{-\tau_\nu}) \qquad \text{(no background source)} \qquad (6.31)$$

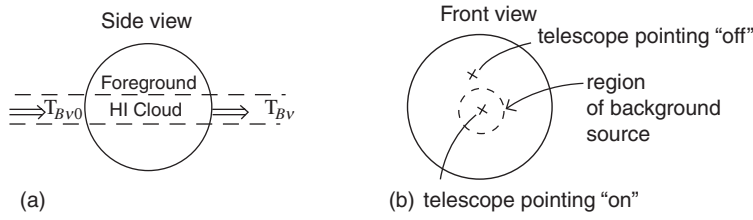


Figure 6.9. Geometry showing how a background radio-emitting source can be used to probe a foreground HI cloud. **(a)** The emission from a background source of brightness temperature, $T_{B\nu 0}$, impinges on a foreground HI cloud and emerges at a brightness, $T_{B\nu}$ on the near side of the cloud. **(b)** The angular size of the background source is shown with a dashed curve. When the telescope points at the cloud directly along the line of sight to the background source, it measures a brightness temperature, $T_{B\nu}(\mathrm{ON})$. When it points on the cloud but away from the direction of the background source, it measures $T_{B\nu}(\mathrm{OFF})$.

where $T_{B\nu}$ is the observed brightness temperature, $T_{B\nu 0}$ is the brightness temperature of the background source, and T and τ_ν are the kinetic temperature (taken to be equal to the spin temperature, T_S, of the gas[13], see Sect. 3.4.5) and optical depth of the HI cloud, at a frequency, ν, in the line, respectively. For the sake of completeness and for future reference, note that, in the case of no background source, Eq. (6.31) reduces to,

$$T_{B\nu} = T \qquad (\tau_\nu \gg 1) \tag{6.32}$$

$$T_{B\nu} = T\,\tau_\nu \quad (\tau_\nu \ll 1) \tag{6.33}$$

(using Eq. A.3) which shows more explicitly that the brightness temperature will always be less than the kinetic temperature (see Eq. 6.24) for a gas in LTE.

If we wish to know *both* the temperature and optical depth of an HI cloud, a single measurement of $T_{B\nu}$ is insufficient. However, if an HI cloud covers a background source and also extends over a larger region where there is no background source, then *both* Eqs. (6.30) and (6.31) can be used. We then can solve these two equations for the two unknowns, T and τ_ν. Fortunately, nature has provided many such examples, since most HI clouds we wish to understand are nearby, being in the ISM of the Milky Way, in comparison to the background sources which are typically radio-emitting quasars in the distant Universe (see Figure 6.10 for an example). Figure 6.9 shows the relevant geometry (though note that foreground HI clouds are generally irregular in shape). When the telescope is pointed at the cloud directly in front of a background quasar which has a brightness temperature higher than the temperature of the cloud, then an absorption line against the background continuum is seen, as in Figure 6.3.a. The emergent brightness temperature in this case can be called, $T_{B\nu}(\mathrm{ON})$. When the telescope is pointing at the cloud but immediately *adjacent* to the background source, then an emission line is seen without any continuum and the emergent brightness

[13]The spin temperature, T_S, is often retained in these equations instead of T in case the assumption of LTE breaks down.

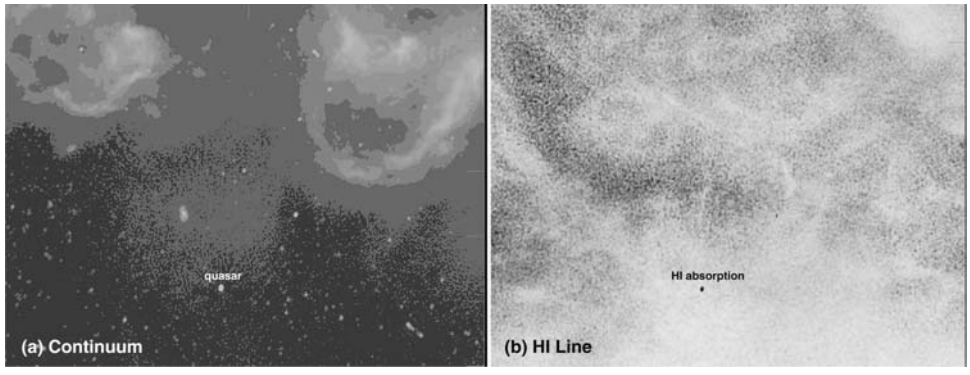


Figure 6.10. Two images of the same region near the plane of the Milky Way taken at $\lambda 21$ cm. Tick marks separate 0.5° (about the diameter of the full Moon) on the sky, showing what a large region of the sky these images cover. **(a)** Radio continuum emission at a wavelength just off the HI line. The two large structures at the top are HII regions in our Milky Way whereas the myriad point or point-like sources are quasars in the distant Universe, one of which is labelled and has a brightness temperature of $T_{B\nu 0} = 471$ K. **(b)** HI emission (and some absorption) from neutral HI clouds in our Milky Way. This image, which is at one particular frequency, ν, in the line, has already had the continuum subtracted (i.e. it represents $T_{B_\nu} - T_{B_{\nu_0}}$ at every position). The value at the location of the absorption feature is -126 K and the average value immediately adjacent to it is 55 K. (Reproduced by permission of The Canadian Galactic Plane Survey (CGPS) team, Taylor, A.R. *et al.*, 2003, AJ, **125**, 3145).

temperature is $T_{B\nu}(\mathrm{OFF})$. Subtracting Eq. (6.31) from Eq. (6.30) for these two cases and rearranging yields,

$$\tau_\nu = -\ln\left[\frac{T_{B\nu}(\mathrm{ON}) - T_{B\nu}(\mathrm{OFF})}{T_{B\nu 0}}\right] \tag{6.34}$$

Since $T_{B\nu 0}$ can be obtained by measuring the continuum of the background source just *off* the line frequency, the optical depth of the HI cloud can be found. We take the optical depth here to refer to the value at some frequency in the line. (See Chapter 9 for more information on line shapes and widths.) Once τ_ν is known, either Eq. (6.30) or (6.31) can then be used to obtain T.

Although nature has provided many examples of background sources behind HI clouds, one also needs to be aware of some subtleties. For example, this process will work only if the temperature and optical depth of the cloud are the same in the ON and OFF positions. Cloud temperatures do not change dramatically from place to place (consistent with our LTE assumption), but τ_ν depends on both the density and line-of-sight distance through the cloud (Eq. 5.9). Both of these quantities may vary with position and therefore a number of OFF measurements should be made as close as possible to the ON position and the results averaged. This development also assumes the presence of only a single HI cloud along a line of sight whereas there may be several. The fact that clouds are moving at different velocities, however, ameliorates

this problem somewhat. Since the frequency of the HI line will be Doppler-shifted (Sect. 7.2.1) by different amounts for each cloud, this helps in treating them as discrete objects along a given line of sight. Another issue is the possibility of variations within the spatial resolution of the telescope. For example, the telescope's beam in the ON position may have a larger angular size than the background quasar, in which case some fraction of the beam may be 'filled' with a signal that has no background. A more fundamental problem is that the ISM is known to contain not only HI that is cool ($T \approx 100\,K$), fairly dense ($n_{HI} \approx 1\,cm^{-3}$) and absorbing, called the *cold neutral medium (CNM)*, but also hot (of order 8000 K), diffuse ($\approx 0.1\,cm^{-3}$) HI, called the *warm neutral medium (WNM)*. Contributions from the hot HI to the beam affect the results although, because the hot component is less dense, the cold component has greater effect. For example, if half of the beam is occupied by the CNM at $T = 100\,K$ and half by the WNM at $T = 8000\,K$, this method will result in a temperature of $T = 200\,K$ (Ref. [81]), an error of a factor of 2.

With these caveats in mind, this method is still a very powerful tool for probing the physical conditions in HI clouds. More sophisticated analyses that take the hot contribution into account find that the cold component has temperatures that range from about $20 \rightarrow 125\,K$ with a median temperature of 65 K, with warmer clouds more common than colder ones (Ref. [47]). Optical depths take on a range of values depending on the cloud size and density. A real example of HI in the Milky Way is shown in Figure 6.10.

Problems

6.1 (a) Assume that the Earth's atmosphere can be approximated by a uniform density plane parallel slab (e.g. one layer of Figure 2.A.1) and modify Eq. (6.9) so that the specific intensity, I_ν, is a function of the specific intensity above the atmosphere, $I_{\nu 0}$, the vertical optical depth, $\tau_{\nu(z=0)}$ and the zenith angle, z.

(b) Use the result of part (a) to write an expression for the ratio of observed specific intensities at two different frequencies, I_{ν_1}/I_{ν_2}, for a source whose intrinsic spectrum is flat (i.e. the true specific intensity is the same at all frequencies).

(c) For clear sky optical depths of $\tau_{blue\,(z=0)} = 0.085$ and $\tau_{red\,(z=0)} = 0.060$, plot a graph of I_{red}/I_{blue} as a function of zenith angle over the range, $0^\circ \leq z \leq 75^\circ$ and comment on the result.

6.2 An H α emission line of specific intensity, $I_\nu = 2.7 \times 10^{-4}\,erg\,s^{-1}\,cm^{-2}\,Hz^{-1}\,sr^{-1}$ at the line centre is observed from an ionised gas that is in LTE. Find a lower limit to the temperature of this gas.

6.3 Indicate whether, in the following cases, a spectral line would be seen in emission, in absorption, or not at all, and briefly explain why. In each case, the vantage point is the surface of the Earth.

(a) A region of the Sun's photosphere of temperature, $T = 5000\,K$, measured in a direction towards the centre of the Sun's disk.

(b) A Solar prominence of temperature, $T = 5000\,K$, seen in the chromosphere during a Solar eclipse.

(c) The Earth's atmosphere at night.

(d) A gas cloud at a temperature of 2.7 K with no discrete background source.

(e) An HII region in front of a background quasar. The quasar has a brightness temperature of $T_B = 100\,K$.

(f) As in (e) but the quasar has a brightness temperature of $T_B = 80000\,K$.

6.4 (a) Compare the collisional timescale for de-excitation of the $\lambda 21$ cm line in a typical interstellar HI cloud to the mean time for spontaneous de-excitation of this line (Tables 3.1, 3.2, and C.1 may be of help) and confirm that the $\lambda 21$ cm line emission is more likely to be collisionally, rather than spontaneously de-excited.

(b) Show that the $\lambda 21$ cm line is always observed in the Rayleigh–Jeans limit.

6.5 (a) From the information given in Figure 6.10 (see also Sect. 6.4.2), find the temperature, T, and optical depth, τ_ν, of the HI cloud at the position of the labelled quasar. Is the cloud optically thick or optically thin?

(b) Assuming that the depth of the cloud along the line of sight is 5 pc, find the number of absorptions per cm, α_ν, in this cloud at the frequency of the line.

(c) Assuming that the cloud remains the same density, how deep (pc) would it have to be for it to become optically thick?

(d) What cloud temperature would be required for an HI emission line to be seen at the position of the quasar?

7

The Interaction of Light with Space

> [The] cause of gravity is what I do not pretend to know . . . That gravity should be innate inherent & essential to matter so [that] one body may act upon another at a distance through a vacuum [with]out the mediation of any thing else . . . is to me so great an absurdity . . . Gravity must be caused by an agent acting constantly . . . but whether this agent be material or immaterial is a question I have left to [the] consideration of my readers.
>
> –Isaac Newton, in letters to Richard Bentley (Ref. [179]).

We now wish to consider how light interacts with space. Matter is not ignored in this process since it is matter that emits light and also affects the light's path. However, the interaction is not one of direct contact, such as we saw for scattering and absorption previously. If the signal did not intercept a single intervening particle or another photon, and even if we had perfect measuring instruments that did not alter it, the signal could still be perturbed from the state it was in when emitted. How can a signal be altered as it travels through a vacuum? In the next sections, we consider some possibilities. But first, it is helpful to consider what is meant by space, or rather *space–time*.

7.1 Space and time

Einstein's *Special Theory of Relativity* is based on the postulate that the laws of physics are the same and the speed of light, c, is the same, regardless of the speed of the observer. Indeed, the concept of *space–time* is a result of the intimate link between space and time via the constancy of c. His *General Theory of Relativity* later expanded these concepts to provide a geometrical description of space–time, its relation to gravitating masses, and its dynamic character, as we shall soon describe.

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

A geometrical description of space–time requires the use of some coordinates. Two points can have the same spatial coordinates, x, y, z, and therefore a distance, 0, between them. However, if two objects are at these coordinates at different times, then they are separated by a time, t, and will have no interaction with each other. This seems quite obvious, but the issue is much more acute in cosmology. Because of the finite speed of light and large distances in the Universe, we see the galaxies where they were in the past, not where they are now. The more distant the galaxy, the farther in the past we see it because it has taken longer for the light to reach us. Thus we are separated from these galaxies by *both* space and time. It is more useful, therefore, to describe a separation between *events in space–time*, rather than a separation between points in space or between instants in time. A mathematical measure of a distance between events in some 'space' is called a *metric*. By incorporating the appropriate physics, it is possible to write metrics for specific circumstances. For example, a metric that describes a *homogeneous* (meaning the same at every location), *isotropic* (the same as viewed in any direction) and expanding universe is called the *Robertson–Walker metric*. This metric is important in the derivation of the Hubble Relation described in Appendix F. A metric that describes the curvature of space about a spherically symmetric stationary mass (see below) is called the *Schwarzschild metric*. The Schwarzschild metric is used in the development of the relevant equations for the gravitational redshift (Sect. 7.2.3) and gravitational lenses (Sect. 7.3.1). We do not present the various metrics, mathematically, in this text (see Ref. [33] for some examples). However, even without a mathematical description, it is possible to describe and visualize some of the concepts of relativity. Central to this is the concept of space–time as an 'entity'. For the moment, let us consider only space.

The fact that matter moves through space is well known since this is part of our day-to-day experience. Not obvious, however, is the fact that space can 'move matter', that is, the trajectory of a particle will depend on the specific geometry of the space through which it travels. Such effects are too small to be measurable on day-to-day human scales, but are very important cosmologically. The 'vacuum of space', even when devoid of all normal matter, radiation, magnetic or electric fields, still has properties as well as a residual vacuum energy. Indeed, particles can appear out of this vacuum for very brief periods of time. These are called *virtual particles* and cannot be measured directly. Thus space is not empty, but is an entity that can expand, curve, or distort.

A helpful visualization is to represent three-dimensional space as a two-dimensional surface that can stretch and deform (Figure 7.1). In this picture, a photon or particle that moves through three-dimensional space is represented as travelling along the surface of the sheet. If the sheet is perfectly flat, then two parallel light beams sent out over its surface will remain parallel indefinitely. This is called flat, or Euclidean space. If the sheet had some global curvature, then initially parallel beams would either converge (positive curvature like a sphere) or diverge (negative curvature like a saddle).

On the largest cosmological scales, the best available evidence suggests (Ref. [142], Ref. [159]) that space is flat, like the regions far from the central depression in Figure 7.1. This does not mean, however, that space is static. A flat sheet can stretch outwards and still remain flat as it carries any embedded particles along with it.

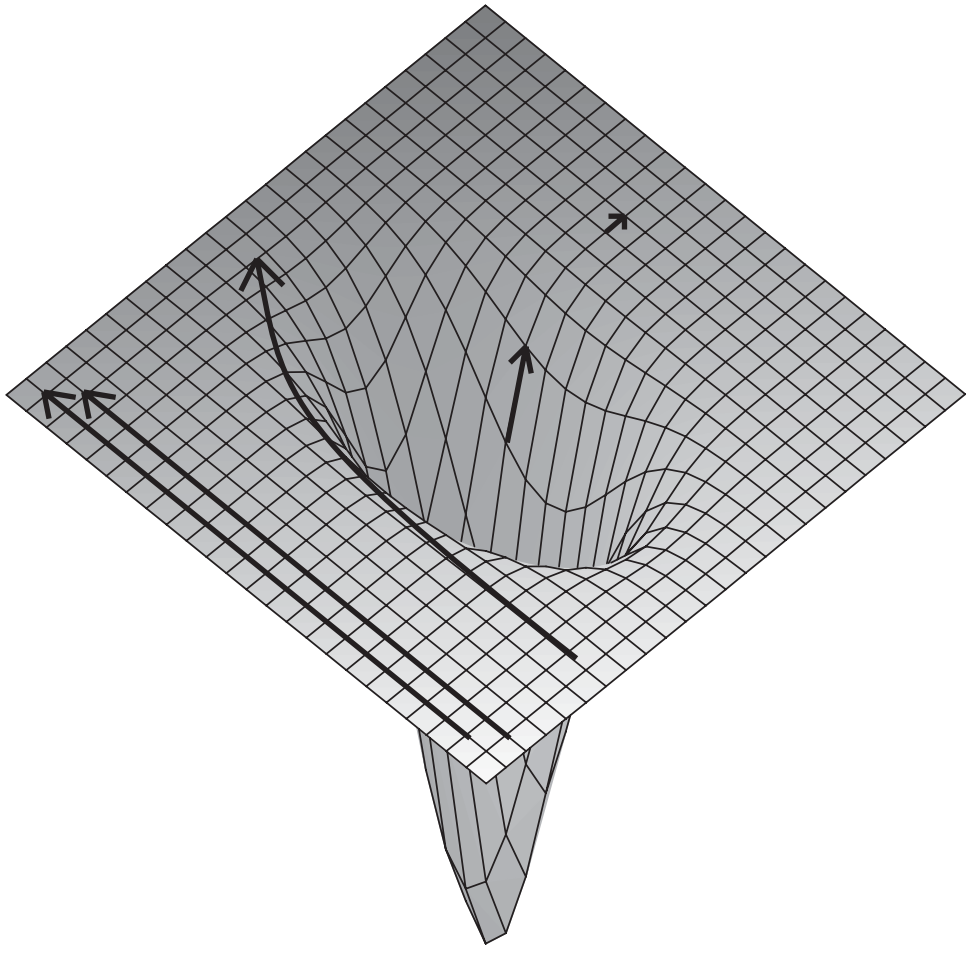


Figure 7.1. Conceptualization of three-dimensional space as a two-dimensional surface that can be curved by a gravitating mass. The two parallel lines at left represent parallel light beams that remain parallel in flat space. This is what our Universe is believed to be like on large scales. The curve represents light that travels near the mass and is deflected. This is what produces gravitational lensing (Sect. 7.3). The two short arrows represent a light ray that leaves the surface of the massive object in a radial direction, helping to illustrate the gravitational redshift (Sect. 7.2.3).

Similarly, Euclidean space can expand and carry the galaxies with it, a property that is further described in Appendix F.

On small local scales near any mass, space is not flat but curved, and the curvature is greatest for the most strongly gravitating masses. This is analogous to a sheet with a heavy ball at the centre (Figure 7.1). Any photon travelling near the mass will experience a change in direction as it moves through this curvature. This is called *gravitational bending* and will be discussed in Sect. 7.3.1.

Including the element of time again, a profound consequence of relativity, now experimentally verified, is *time dilation*. A photon travelling near a mass is travelling a greater distance through curved space in comparison to a photon that is farther away (Figure 7.1). Since a local freely-falling observer near the mass must measure the same speed of light, c, in comparison to an observer farther away, it follows that a unit of time near the mass is greater than a unit of time farther away. In a sense, time and space 'stretch' together. If the observer near the mass sends out a signal at regular one second intervals according to his clock, a distant observer will measure the arriving signals as being greater than one second apart. Each observer's clocks are advancing 'normally' in their own frame of reference, but not when compared to each other. Similarly, the observer near the mass will age more slowly in comparison to the observer farther away. A consequence of this fact is that a photon that leaves a mass will experience a *gravitational redshift*. This will be discussed in Sect. 7.2.3.

7.2 Redshifts and blueshifts

If a signal is emitted at a frequency, ν_0, it may be observed at a frequency, ν, that is shifted from the emitted frequency. Since $\lambda = c/\nu$, the observed wavelength will also be altered from the emitted value, the latter called the *rest wavelength*, λ_0. If λ has increased in comparison to λ_0, the light is said to be *redshifted* and if it has decreased, the light is *blueshifted*. For a photon, $E = h\,c/\lambda$, so redshifted and blueshifted photons have lower and higher energies, respectively.

Since, on cosmological scales, observed wavelengths are redshifted (Sect. 7.2.2), changes of wavelength are often described by a parameter called the *redshift parameter* or just *redshift*, z, defined by,

$$z \equiv \frac{\Delta\lambda}{\lambda_0} = \frac{\lambda - \lambda_0}{\lambda_0} \;\; => \;\; z + 1 = \frac{\lambda}{\lambda_0} \tag{7.1}$$

Thus, z describes the fractional change in the wavelength. An observed increase in wavelength corresponds to positive z and a decrease in wavelength to negative z. There are three possible origins for such a shift, as will now be described.

7.2.1 The Doppler shift – deciphering dynamics

7.2.1.1 Kinematics – the source velocity

As a source moves through space, its velocity, v (the *space velocity*), can be expressed in terms of a component that is directed towards or away from the observer, called the *radial velocity*, v_r, and a component that is in the plane of the sky, called the *transverse velocity*, v_t (see Figure 7.2a). Then $v = \sqrt{v_\mathrm{r}^2 + v_\mathrm{t}^2}$.

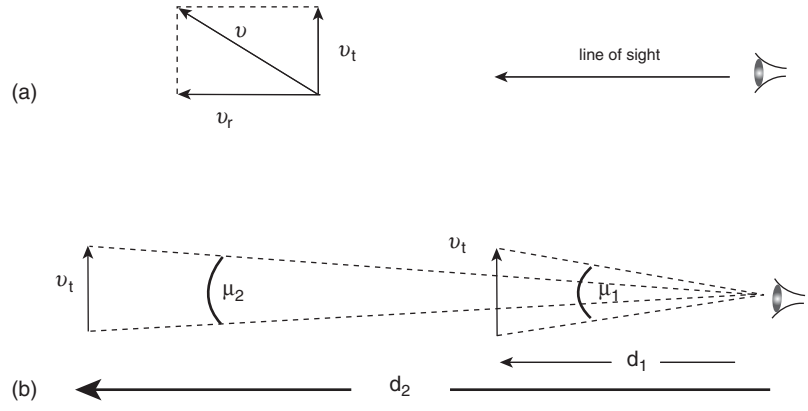


Figure 7.2. **(a)** An object's space velocity, v, can be represented as the vector sum of the transverse velocity in the plane of the sky, v_t, and the radial velocity in the line of sight, v_r. **(b)** The transverse velocity can be determined from the proper motion if the distance is known, i.e. $v_t = \mu\, d$. A nearby object (at d_1 in the figure) will have a higher proper motion than a distant object (at d_2) for the same v_t.

It is the radial velocity that produces a *Doppler shift*[1] as the source moves through space. A motion away from the observer results in a redshift and a motion towards the observer gives a blueshift. Note that v_r is the radial velocity of the source *with respect to the observer*. If both the source and observer are deemed to have motions with respect to some other reference centre or *rest frame*, then the radial velocity producing the Doppler shift is the difference in radial velocity components between the source and observer, each measured with respect to the rest frame[2]. For example, motions can be measured with respect to the Sun (as if the Sun were at rest), with respect to an imaginary point moving in a perfect circle about the centre of the Galaxy at the location of the Sun (called the *Local Standard of Rest*, LSR), or with respect to the centre of the Milky Way.

From the Doppler shift formulae of Table I.1 and Eq. (7.1),

$$z + 1 = \left(\frac{1 + \frac{v_r}{c}}{1 - \frac{v_r}{c}}\right)^{1/2} \qquad z = \frac{v_r}{c} \quad (v_r \ll c) \tag{7.2}$$

The second expression, which is the approximation for radial velocities much less than the speed of light, is all that is needed for most cases (Prob. 7.1). This simple relation is

[1]There is also a Doppler shift for transverse motion when speeds are relativistic. This is due to time dilation and will not be considered further here.

[2]This is different from the Doppler shift for sound which requires a medium within which to propagate and therefore has an absolute rest frame. Thus, the source velocity and the observer velocity must be treated separately when dealing with the Doppler shift for sound.

very powerful because it directly relates an observable quantity, z, to a physical parameter of the source, ν_r, independent of any other knowledge of the source, including its distance. All that is necessary is to ensure that some spectral feature, usually a spectral line (Prob. 7.2) can be identified, as Example 7.1 illustrates.

Example 7.1

The H α line, emitted from a location near the centre of the Andromeda galaxy, M 31, is observed at a wavelength of $\lambda(H\alpha) = 655.624$ nm. Write an expression for the radial velocity of a galaxy, assuming $v_r \ll c$ and determine the radial velocity of the Andromeda galaxy.

From Eqs. (7.1) and (7.2), the radial velocity is given by,

$$v_r = \left[\frac{\lambda - \lambda_0}{\lambda_0}\right] c = \frac{\Delta\lambda}{\lambda_0} c \tag{7.3}$$

Eq. (7.3) can also be written in terms of frequency,

$$v_r = \left[\frac{\frac{c}{\nu} - \frac{c}{\nu_0}}{\frac{c}{\nu_0}}\right] c = \left[\frac{\nu_0 - \nu}{\nu}\right] c = \frac{-\Delta\nu}{\nu} c \tag{7.4}$$

We need only one of these equations and, given $\lambda_0(\mathrm{H}\alpha) = 656.280$ nm from Table C.1, as well as the above value for $\lambda(\mathrm{H}\alpha)$, Eq. (7.3) leads to $v_r = -300\,\mathrm{km\ s^{-1}}$. Thus, the Andromeda galaxy is approaching us.

The problem of obtaining the transverse velocity is less easily solved. If the object is relatively nearby, then it may be possible to measure its angular motion in the plane of the sky over time (Figure 7.2b). This is called the *proper motion*, μ, and it is typically quite small, being measured in units of arcseconds per year. If the distance to the object is known, then Eq. (B.1) can be used to convert the proper motion to a transverse velocity (km s^{-1}) with proper conversion of units (e.g. $v_t = \mu d$ will give v_t in km s^{-1} for μ in rad s^{-1} and d in km). This kind of measurement is only possible for the nearby stars (Prob. 7.3). For more distant objects, statistical arguments can sometimes be used. For example, stars in a cluster may be moving with random motions in which case we might expect $v_r \approx v_t$, on average, with respect to the cluster centre. In other cases, the velocity may be inferred based on geometry. In a worst case scenario, v_t is simply accepted to be unknown and v_r places a lower limit on the space velocity. However, for any arbitrary configuration in the sky, the transverse velocity is likely to be of the same order of magnitude as v_r unless the object happens to be in a 'special' geometry of moving very close to the plane of the sky or line of sight.

7.2.1.2 *Dynamics – the central mass*

As early as 1619, Johannes Kepler realized that the orbital motions of the planets were governed, not by the mass of the planet itself, but by the mass of the Sun. In modern form, Kepler's third law[3] states,

$$\mathcal{T}^2 = \frac{4\,\pi^2\,a^3}{G\,(\mathrm{M}+\mathrm{m})} \tag{7.5}$$

where $\mathcal{T}$ is the period of the orbit, G is the universal gravitational constant, and a is the semi-major axis of the elliptical orbit (see Appendix A.5). Eq. (7.5) indicates that, for $\mathrm{m} \ll \mathrm{M}$, the period depends only on the mass of the object being orbited. Since the period, in turn, depends on the *velocity* of the orbiting body and its separation from M, it is sufficient to measure the velocity and separation of the orbiting object to find the interior mass. This is a very important and powerful concept in astronomy because it is relatively straightforward to measure velocities via Doppler shifts (Example 7.1). Thus, the orbiting body or bodies act like test particles that probe the central mass about which they move. Moreover, this method can be generalized to include extended mass distributions, rather than the simple point-like mass of the Sun at the centre of the Solar System, as we show below.

It is frequently of interest, then, to determine a velocity, not with respect to us, but with respect to another centre of rest. Restricting ourselves to the simpler case of circular orbits[4] the important quantity is the circular velocity, v_{c}, about the centre. For example, the Earth's orbital motion is about the Sun and the Sun's orbital motion is about the centre of the Milky Way galaxy. Similarly, we might want to consider the orbital motion of extra-solar planets about their parent star or the motions of stars about the centre of another galaxy. If the entire external system (central mass plus orbiting bodies) has some radial velocity with respect to the observer (us), this motion is called the *systemic velocity*, v_{sys}.

Rotating *thin disks* are examples of extended mass distributions for which the orbits of individual particles can be closely approximated by circles. They are frequently observed in many contexts in astronomy and have a common geometry. By *thin*, we mean that the disk thickness is much less than the disk diameter. A thin disk would consist of individual particles, each with its own rotational speed, $v_{\mathrm{c}}(R)$, at different radii, R, from the centre. Examples are the Solar System-sized dusty disks around stars (Figure 4.13), the galaxy-sized dusty disks in elliptical galaxies (Figure 3.22), and hot, gaseous disks (called *accretion disks* since their material accretes onto the central

[3]Kepler's first law states that the orbits of the planets are ellipses with the Sun at one focus, and the second law indicates that the planets 'sweep out' equal areas of the ellipse in equal units of time. The latter point refers to the fact that the planets move fastest when they are near perihelion and slowest near aphelion (see Table A.1).

[4]Over a period of time, elliptical orbits tend to evolve towards circularity, which corresponds to the lowest energy orbit. There are many examples of circular or nearly-circular motions in astronomy.

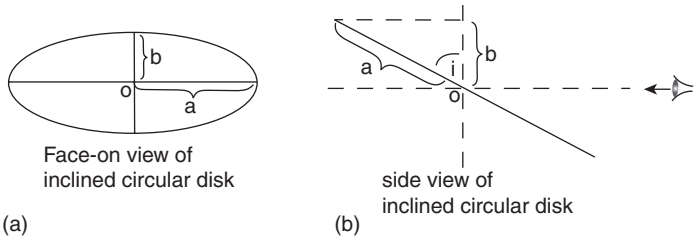


Figure 7.3 **(a)** A circular galaxy that is inclined to the line of sight will look elliptical in projection. The galaxy's true radius is *a*. When inclined, its apparent semi-major axis is *a* but its semi-minor axis is *b*, as shown. **(b)** This side view shows the minor axis of the inclined galaxy of (a) and its inclination.

object) around compact objects like neutron stars and black holes. *Spiral galaxies* can also be modelled as thin disks, the individual particles being all those objects making up the disk such as stars, gas, dust, and nebulae (see Figure 3.3). These include the nearly face-on galaxies like M 51 (Figure 3.8) and NGC 1068 (Figure 5.5), galaxies of intermediate inclination like NGC 2903 (Figures 2.11, 2.18, and 2.19) and galaxies that are edge-on to the line of sight like NGC 3079 (Figure 3.16). The *inclination* of a galaxy is measured with respect to the plane of the sky, so that a face-on galaxy has $i = 0^\circ$ and a galaxy that is edge-on to the line of sight has $i = 90^\circ$. A thin circular disk looks elliptical in projection and the inclination can therefore be determined by the geometry of the ellipse, as shown in Figure 7.3,

$$\cos(i) = \frac{b}{a} \tag{7.6}$$

where b is the semi-minor axis as projected in the plane of the sky and a is the galaxy's semi-major axis. The quantities, a and b, can be expressed in linear units (e.g. cm, kpc) or angular units (e.g. radians, arcminutes), see Eq. (8.1), as long as the same units are used for a and b.

The geometry of a rotating disk and how it relates to the observed Doppler shifts of spectral lines is shown for a galaxy disk in Figure 7.4. For any radius, R, along the major axis[5]

$$v_\mathrm{r} = v_\mathrm{sys} + v_\mathrm{c}(R)\sin(i) \tag{7.7}$$

At the galaxy's centre, the circular velocity is zero, so the radial velocity measured at the galaxy's centre is just v_sys, as Example 7.1 implied. Since i is known from Eq. 7.6, $v_\mathrm{c}(R)$ can be found for any position, R, along the major axis, provided a spectral line can be identified and v_r determined for that position.

[5]Off the major axis, the equation is $v_\mathrm{r} = v_\mathrm{sys} + v_\mathrm{c}(R)\cos(\theta)\sin(i)$. The angle, θ, is measured in the plane of the galaxy with a vertex at the galaxy's centre. It is the angle between the major axis and the line between the galaxy centre and the point off the major axis that is being considered.

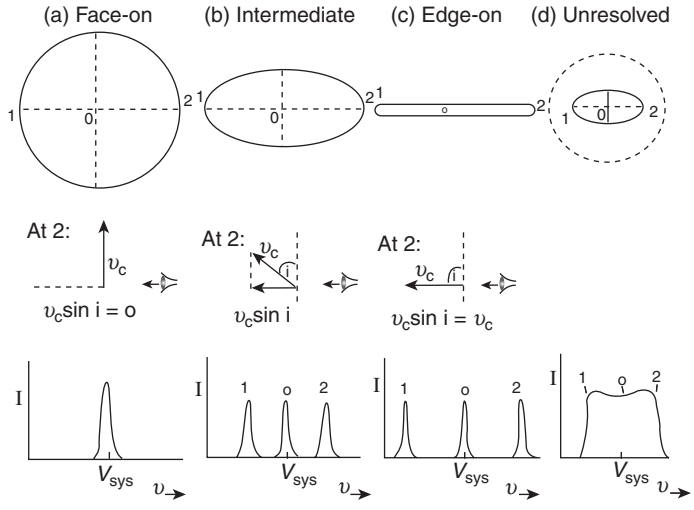


Figure 7.4 **(a)** The top image shows a face-on galaxy ($i = 0°$) that is rotating in the plane of the sky. Points 1 and 2 at the ends of the disk are marked, as is point *o* at the centre. Below this, the middle diagram shows a side view, indicating that there is no component of v_c projecting onto the line of sight. The bottom view indicates the spectrum from such a galaxy. Since there is no radial motion with respect to the centre, all points lie at the same radial velocity which is just the systemic velocity, v_{sys}. **(b)** As in (a) but for a galaxy of intermediate inclination. In this case, there is a radial velocity component with respect to v_{sys} so points 1 and 2 are separated from point *o* in the bottom spectrum. **(c)** As in (a) but for an edge-on disk ($i = 90°$). Points 1 and 2 are separated by the maximum amount in the spectrum. **(d)** Example of a galaxy of intermediate inclination that is unresolved spatially. The dashed circle in the top image shows the size of the resolution element. In this case, all points in the galaxy are received at once so the spectrum shows a continuous distribution of velocities centred on v_{sys}. The maximum values of v_r on either side correspond to points 1 and 2

It is even possible to determine $v_c(R)$ for our own Galaxy, although the geometry is somewhat more complicated (and will not be described in detail here) because of our unique position within the Milky Way (see Figure 3.3). With a rotating disk model, it is possible to use the observed radial velocity of an object together with a knowledge of $v_c(R)$ to obtain the distance to the object. This is an important way of obtaining distances, called *kinematic distances*, within the Milky Way[6]

The development described above for obtaining $v_c(R)$ for a rotating disk is very powerful because it allows us to determine the dynamics of galaxies or any other object for which circular velocities can be derived. If a stable rotational configuration has been

[6]The geometry is such that two distances are determined, one corresponding to the near side and one to the far side of the Galaxy. Thus, there is a distance ambiguity that remains in this analysis. If other independent information about the object is available, it can often be brought to bear on the problem so that one of the two distances can be discarded in favour of the other.

established, this means that the gravitational force, F_G can be equated to the centripetal force, F_c (see Table I.1) which, re-arranging and solving for central mass becomes,

$$M(R) = f\frac{R\,[v_c(R)]^2}{G} \tag{7.8}$$

Here, $M(R)$ is the mass *interior to* the point, R from the centre and G is the Universal gravitational constant. The quantity, f, is a geometrical correction factor to account for the fact that the distribution of interior mass may be flattened rather than spherical. For a spherical mass distribution (like a spiral galaxy with a dark matter halo, see below) and for a situation in which virtually all mass is contained in a central point (like the Solar System), $f = 1$. If we want to obtain the total mass of an object, such a calculation should be done at the outermost measurable point so as to include as much interior mass as possible. For values expressed in common astronomical units, the above equation becomes,

$$\frac{M}{M_\odot} = 2.32 \times 10^5\, f \left[\frac{R}{\text{kpc}}\right] \left[\frac{v_c}{\text{km s}^{-1}}\right]^2 \tag{7.9}$$

where the dependence of mass and velocity on R is taken to be understood.

Eq. 7.9 can be used for any circularly rotating disk consisting of individual freely orbiting particles, or any single circularly orbiting body *whether the interior mass is observed or not*. All that is required is that a spectral line be identified and measured to obtain v and the distance to the object be known in order to convert an angular distance along the major axis to R. Note that the spectral line is being emitted by an object whose own mass does not enter into Eq. (7.9). Thus, as indicated earlier, the moving object is a tracer of the mass that it orbits. For example, the mass of a black hole can be found from the accretion disk orbiting it, the mass of an unseen massive planet can be found from the orbit of the central star about the common centre of mass (a small but measurable perturbation, Sect. 4.2.2), the mass of a galaxy, including the parts that we can see as well as any dark matter halo interior to the measured point can be found by measuring the Doppler shifts of the outermost points on the major axis, as Example 7.2 describes.

Determining the *mass distribution* of a galaxy can be achieved by measuring the circular velocity, v_c, as a function of R. A plot of $v_c(R)$ is called a *rotation curve* and it is observationally found that v_c is approximately constant with R in the outermost parts of galaxies. This means that the mass continues to rise with R in these regions ($M \propto R$, from Eq. 7.9, when $v_c \approx$ constant) even though the *light* is seen to fall off exponentially. This is one of the strongest arguments for dark matter in and around galaxies. In the other extreme, in which virtually all mass is contained in the central point, all test particles 'feel' the central mass regardless of their distance. This implies that $v_c \propto 1/\sqrt{R}$ (since $M \approx$ constant in Eq. 7.9). An example of this case is the Solar System, for which $M = M_\odot$. Substituting $a = R$ and $v_c = 2\pi R/\mathcal{T}$ in Eq. (7.5) leads to the same proportionality for v_c.

Example 7.2

A spiral galaxy at a distance of 20 Mpc and $v_{sys} = 525$ km s^{-1} has semi-major and semi-minor axes of $a = 5.0'$ and $b = 2.3'$, respectively. The λ 21 cm line of HI at the farthest measurable point along the major axis has an observed wavelength of λ 21.155190 cm. Find the mass of the galaxy interior to the measured point.

From Eq. (7.6), the inclination is $i = \cos^{-1}(b/a) = \cos^{-1}(2.3/5.0) = 62.6°$. The outermost measurable point along the major axis (5′) is at a distance (by Eq. B.1 and converting to kpc and radians) of $R = (20 \times 10^3)\,(1.45 \times 10^{-3}) = 29.1$ kpc. The radial velocity of this point, from Eq. (7.3) and using $\lambda_0 = 21.106114$ (Table C.1), is 697.1 km s^{-1}. Therefore, this side of the galaxy is *receding with respect to its centre*. From Eq. (7.7) and known values of inclination and systemic velocity, we find a circular velocity of $v_c = 193.8$ km s^{-1}. Finally, adopting a spherical mass distribution (f = 1), Eq. (7.9) yields $M = 2.5 \times 10^{11} M_\odot$.

For situations in which the whole rotating disk is spatially unresolved, the emission from all parts of the disk are received in a single resolution element, as shown in Figure 7.4.d. All radial velocities represented in the galaxy are then blended into a single broad feature as indicated in the spectrum, although the shape may vary depending on the distribution of material in the disk and inclination. If it is known from some other measurement that the object is indeed a rotating disk, then it is still possible to obtain the central mass by noting that the central velocity of the spectrum should correspond to v_{sys} and the extreme points of the spectrum should correspond to the radial velocity at the maximum value of R as the figure indicates (Prob. 7.6).

Considering the geometry of a rotating disk is of great use in obtaining the interior mass over a large range of size scales, from dusty systems around individual stars to entire galaxies. In principle, however, this kind of dynamical study can be applied to any gravitationally bound system that isn't necessarily in the geometry of a thin disk. The stars in elliptical galaxies (see Figure 3.22), for example, are in a more spherical distribution that results from individual stellar orbits over a wide variety of angles. These stars are moving in individual orbits that are dictated by the gravitational potential of the elliptical galaxy as a whole. A statistical analysis of such stellar motions then leads to a determination of the mass – both light and dark mass – of the galaxy. Going to a significantly larger scale, by studying the motions of many individual galaxies that are in gravitationally bound galaxy clusters (see e.g. Figure 8.8), the mass of the cluster can also be found. Thus, a tiny Doppler shift in the wavelength of light can lead to a knowledge of light and dark mass, including the largest known masses in the Universe.

7.2.2 The expansion redshift

As early as 1914, it was realized through the work of the American astronomer, Slipher, that galaxies preferentially showed redshifts, rather than blueshifts. As more data were

acquired, it became clear that almost all galaxies showed redshifts. Edwin Hubble later determined the distances to a number of galaxies and, combining his results with the measurements of Slipher, found that galaxies at larger distances also showed larger redshifts. This is known as the *Hubble Relation*, a modern version of which is shown in Figure F.2. The slope of the curve closest to the origin is called the *Hubble constant*, H_0.

The Hubble Relation is a direct consequence of the expansion of the Universe (see Figure F.1). A photon that leaves a distant galaxy must travel through a universe that is expanding. As a result, the wavelength of the observed photon will be 'stretched out' by this expansion in comparison to its wavelength in the rest frame in which it is emitted. Unlike the Doppler shift, this redshift does not result from galaxies moving *through* space. Rather, *the expansion of space itself carries the galaxies with it*, consistent with the properties of space discussed in Sect. 7.1. More distant objects emit light from regions in which the expansion is greater with respect to us and therefore they have higher redshifts. Since redshifts are easily obtained but distances are not, the calibrated Hubble Relation is used to find the distances to galaxies and quasars, assuming that the galaxy or quasar is far enough away that its redshift is dominated by the expansion of the Universe rather than its peculiar velocity through space (Example 7.3). If this is true, then a redshift, z, can be converted into a distance, d, via Eq. F.1 or, for larger redshifts ($z \gtrsim 0.1$), via Eq. F.16. At redshifts higher than $z = 1$, a more complex expression is required (Appendix F.3).

Example 7.3

Estimate a typical distance beyond which a galaxy's redshift is dominated by the expansion of the Universe rather than its peculiar motion through space.

From Eq. (7.2) and Eq. (F.1), we wish to know the distance, d, at which,

$$c\,z\,(\text{Expansion}) = H_0\,d > v_\text{r}(\text{Doppler}) \tag{7.10}$$

We will take the cluster of galaxies, Abell 2218 (see Figure 7.6), as a typical example of the largest, most massive structures in the Universe and therefore the ones capable of producing the highest galaxy velocities through space. A typical velocity of a galaxy in this cluster is $v_\text{r} = 1370\,\text{km s}^{-1}$. Using $H_0 = 71\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ (Table G.3), we find $d > 19.3\,\text{Mpc}$. In terms of extragalactic sources (those outside of the Milky Way), this is not very distant. It is about the same distance as the *Virgo Cluster*, which is the closest substantial cluster of galaxies to us. Therefore, beyond approximately 20 Mpc, the redshift is dominated by the expansion of the Universe and peculiar motions through space can be neglected unless they are unusually high.

The expansion redshift is defined by considering how much the Universe has expanded between the time the light has been emitted, t_em, and the time that it is

observed, $t_{\rm obs}$. The fractional increase in the wavelength of light due to expansion, z, is then equal to the fractional increase in the size of the Universe between these two times,

$$z = \frac{a(t_{\rm obs}) - a(t_{\rm em})}{a(t_{\rm em})} => \quad z + 1 = \frac{a(t_{\rm obs})}{a(t_{\rm em})} \tag{7.11}$$

where $a(t)$ is a dimensionless *scale factor* that describes how the scale ('size') of the universe changes with time. A galaxy with a measured redshift of $z = 6$, for example, emitted its light when the Universe was only 1/7th of its current size. In Appendix F we use this information to derive the Hubble Relation (see Figure F.2) and also to consider how the observable properties of this relation relate to the energy and mass density of the Universe that were presented in Sect. 3.2.

7.2.3 The gravitational redshift

Since any mass will distort space–time around it, a photon that leaves the surface of a mass must emerge from a region of curved space–time. Such a photon is redshifted as a result. This gravitational redshift is a result of time dilation near the gravitating mass. Imagine a signal, such as repeated short pulses of light, emitted at regular time intervals from a region near the mass. A distant observer will measure these time intervals to be longer than the intervals he would measure in his own rest frame for the same process. Thus the observed frequency of these pulses ($\nu \propto 1/t$), as measured by the distant observer, will be lower. If the signal is now a wave of some frequency, rather than individual pulses, then the observed wavelength will be longer, or redshifted, in comparison to the rest wavelength.

It is sometimes useful to look at this in a more classical fashion such as would be done for a particle with a mass, rather than a photon. That is, it takes energy for a photon to climb out of a gravitational potential well. Since the emergent photon's energy decreases during this process, its wavelength must increase. A gravitational blueshift is also possible for an incoming photon as it gains energy approaching a mass.

An expression for the gravitational redshift is derivable from the Schwarzschild metric, the result being,

$$z + 1 = \frac{1}{\sqrt{(1 - \frac{r_{\rm S}}{r})}} \quad r_{\rm S} < r \tag{7.12}$$

where z, measured by a distant observer, is the redshift of a photon that originates at a distance, r, from the centre of the mass[7] and $r_{\rm S}$ is the Schwarzschild Radius, defined by,

$$r_{\rm S} \equiv \frac{2\,G\,\mathrm{M}}{c^2} \tag{7.13}$$

[7] At $r \leq r_{\rm S}$, the coordinate system of an external observer breaks down and we do not consider this case further.

The Schwarzschild Radius is the radius of a black hole, i.e. the radius of a region of space within which the gravity is so strong that not even light can escape[8]. Any mass, M that is *completely* contained within a radius, r_S, will be a black hole. (See Sect. 3.3.3 for information on the formation of stellar mass black holes.)

From Eq. (7.12), it can be seen that the gravitational redshift will not be a strong effect unless the photon leaves an object from a position, r, that is close to its Schwarzschild Radius. Any object that is not a black hole has a radius that is larger than its Schwarzschild Radius (usually *much* larger), so the gravitational redshift is negligible for most objects (Example 7.4). However, it can become important for very compact or collapsed objects like neutron stars and the light-emitting regions surrounding black holes (Prob. 7.12, see also Example 8.3).

Example 7.4

Compare the magnitude of the gravitational redshift to the Doppler shift from a Solar-type star that is receding from us at a velocity of $v_r = 20\,km\ s^{-1}$.

From Eq. (7.2), $z\,(\text{Doppler}) = v_r/c = (20/3.00 \times 10^5) = 6.7 \times 10^{-5}$. From Eqn. (7.13) for a $1\,M_\odot$ star, $r_S = (2\,G)\,(1.99 \times 10^{33})/c^2 = 2.95 \times 10^5$ cm. A photon leaves such a star at a radius, $r = R_\odot = 6.96 \times 10^{10}$ cm, so using Eq. (7.12), $z\,(\text{Gravitational}) = 2.1 \times 10^{-6}$, which is more than an order of magnitude smaller than the Doppler shift.

7.3 Gravitational refraction

Just as a photon which leaves a mass must travel through curved space – time to escape, a photon from a background source that passes near a mass en route to us must also pass through curved space – time. If the background light source, the mass and the observer are sufficiently aligned, then the observer may see an image or images of the source displaced from its true position in the sky. The mass is called a *gravitational lens* since it acts like a convex lens, refracting the light rays towards it.

A variety of image geometries may result, depending on the alignment and distances of the source and lens as well as the source light distribution and lens mass distribution. Several examples are shown in Figure 7.5 and Figure 7.6.

7.3.1 Geometry and mass of a gravitational lens

The geometry of a gravitational lens is shown in Figure 7.7a. The light path is actually curved, but it can be represented by a single *bending angle*, ϕ, defined as the difference

[8]The Schwarzschild Radius is accurately derived by considering General Relativity. However, a simplified way in which to understand the equation is to set the escape velocity from the object equal to the velocity of light, i.e. $v_{esc} = \sqrt{2GM/r} = c$, from which Eq. (7.13) follows.

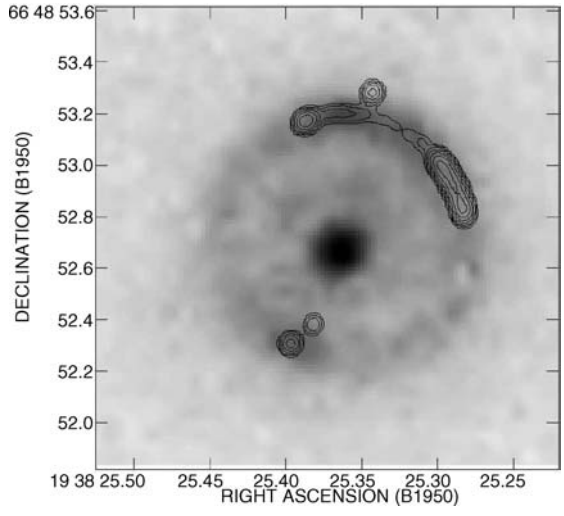


Figure 7.5 The Einstein Ring, B1938+666. The lensing mass (a galaxy) is at the centre. Contours show the 5 GHz radio emission obtained using the Multi-Element Radio-Linked Interferometer (MERLIN) in Great Britain and the grey scale image shows near-IR emission obtained using the Hubble Space Telescope (adapted from Ref. [87]). (Reproduced by permission of L. J. King, University of Manchester)

between the initial and final directions. It has therefore been drawn as two straight ray paths in Figure 7.7a with a single bend. The bending angle, ϕ, was determined by Einstein for an isolated lens and can be derived from the Schwarzschild metric. Assuming a small bending angle, the result is,

$$\phi = \frac{2\,r_S}{b} = \frac{4\,G\,M_L}{b\,c^2} \qquad \phi = 1.75'' \frac{\left[\frac{M_L}{M_\odot}\right]}{\left[\frac{b}{R_\odot}\right]} \tag{7.14}$$

where b, the *impact parameter*, is the distance of closest approach to the mass and r_S is the Schwarzschild Radius (Eq. 7.13) of a lens of mass, M_L. The left equation provides the result in radians (all cgs units) and the right in arcseconds for inputs in Solar units. This result is exactly twice that expected from Newtonian physics[9] and catapulted Einstein to fame when the larger value was confirmed by observations of the displacements of background stars during the solar eclipse of 1919. The potential power of gravitational lensing as an astronomical tool, however, was realized by S. Refsdal in the 1960s (Ref. [149]).

There are two important differences between a gravitational lens as implied by Eq. (7.14) and one made of glass (or other refracting material):

- No focal length: For a glass lens, incoming rays are bent the least near the lens centre (zero at the centre) and greatest near the lens edge, allowing light to intersect at a

[9]Space curvature is not expected in Newtonian physics, but the photon has an 'equivalent mass' via $E = h\,\nu = m\,c^2$ and could be treated as a small particle to find a deflection angle in a Newtonian model.

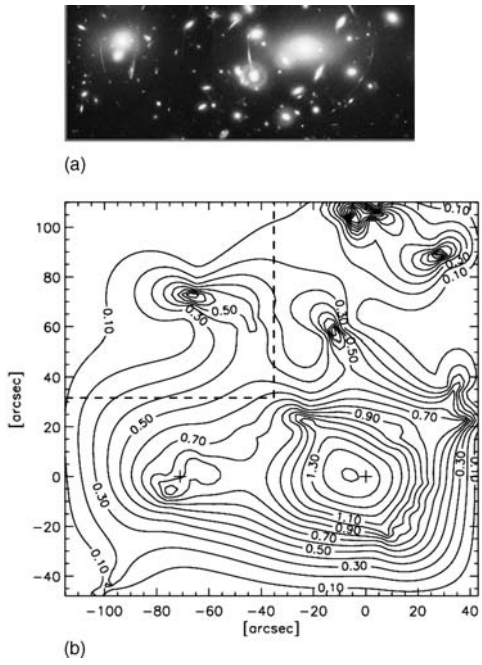


Figure 7.6 **(a)** An optical view of the galaxy cluster, Abell 2218, showing the astonishing warping of space–time via thin arcs that are images of gravitationally lensed background sources. There are over 100 arcs in this cluster which is at a redshift of $z = 0.175$. The velocity dispersion of galaxies in the cluster is 1370 km s^{-1} (Ref. [188]). The x axis spans 127″ and the y axis spans 64″. (Reproduced by permission of W.Couch (University of New South Wales), R. Ellis (Cambridge University), and NASA.) **(b)** A reconstructed mass model over a larger field of view as in (a) from Ref. [1]. The scale is approximately the same as in (a), the crosses marking the positions of the two brightest regions in (a). The mass distribution, which includes both light and dark matter, is much more spread out than the visible galaxies (Reproduced by permission of P. Saha and the AAS)

single point in the focal plane. This is accomplished by machining the surfaces to the desired curvature. For a gravitational lens, on the other hand, the bending angle is inversely proportional to the impact parameter, so light is bent less at larger distances from the lens centre. Thus a gravitational lens has no focal length.

- Wavelength independence: The speed of light in a medium depends on interactions between light and the particles of the medium and such interactions depend on the wavelength of the light. Thus, for a glass lens, the index of refraction ($n = v/c$, Table I.1) is wavelength dependent, blue light being bent more than red light. However, the speed of light in a vacuum is constant. Therefore Eq. (7.14) has no wavelength dependence and a gravitational lens is always *achromatic*. X-rays, IR or radio waves are all bent through the same angle, provided the background source emits in these bands and the emission is coming from the same location. Note how the radio and near-IR emission occurs at the same angular distance from the lensing galaxy in Figure 7.5, for example.

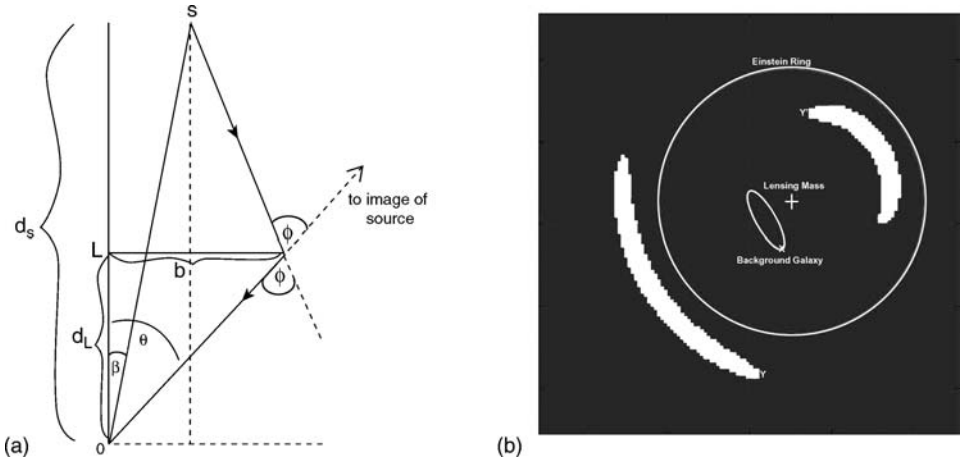


Figure 7.7. **(a)** Geometry of a gravitational lens (with angles exaggerated) when both the source and lens are points. A point source at S is a distance, d_S, and a true angle, β, from the observer at O. A point lens, L, of mass, M_L, is a distance, d_L, from the observer. The light path, shown here by two straight lines with arrows, is actually a curved path whose bending can be characterized by the gravitational bending angle, ϕ. The impact parameter, b, is the closest distance of the light path to the lens, L. The observer will see an image at an angle, θ, from L. **(b)** Face-on view of a gravitational lens in which both the source and lens are no longer points. Here a background elliptical galaxy is lensed by a closer spherically symmetric mass distribution. The observer sees the two arcs shown. The location of the background galaxy, the lensing mass and its corresponding Einstein Ring are indicated. The point, X, on the source, at an angle, β from the lens, has two images: Y, at an angle θ_1 from the lens, and Y' at θ_2 from the lens. Adapted from Ref [111].

From the simplified geometry of Figure 7.7a which assumes that both the source and lens are points, we can calculate the observed displacement of a background object. Since the bending angle is actually very small, the impact parameter of the curved path that the light actually follows is about equal to the impact parameter of a path represented by the straight lines shown. From this geometry, Eq. (7.14) and the assumption of small angles, it can be shown (Prob. 7.13) that the observed angle of the image, θ, satisfies a quadratic equation, known as the *lens equation*,

$$\theta^2 - \beta\,\theta - \theta_E^2 = 0 \tag{7.15}$$

where,

$$\theta_E^2 = \frac{4\,G\,M_L}{c^2}\left(\frac{1}{d_L} - \frac{1}{d_S}\right) \tag{7.16}$$

The latter expression contains the information about the lens mass and the source and lens distances[10]. The lens equation has two roots corresponding to two image

[10]For very large *cosmological* distances, a distance can be defined in several ways. For gravitational lensing, the *angular diameter distance* should be used (Ref. [149]) which is the ratio of the object's physical transverse size to its angular size (in radians), i.e. the distance implied by Eq. (B.1).

locations,

$$\theta = \frac{\beta}{2} \pm \frac{1}{2}\sqrt{\beta^2 + 4\,\theta_{\rm E}^2} \tag{7.17}$$

Since the value of the square root will always be greater than β, there will be one positive root (which we call θ_1) corresponding to the image on the same side of the lens as the source, and one negative (θ_2) corresponding to an image on the other side of the lens. Both these angles can be measured, provided the lensing mass location is observed. Then the true position of the source, which is not observable, can be found from the sum,

$$\theta_1 + \theta_2 = \beta \tag{7.18}$$

and $\theta_{\rm E}$ can be determined from the total angular separation between the two images,

$$\theta_1 - \theta_2 = \sqrt{\beta^2 + 4\,\theta_{\rm E}^2} \tag{7.19}$$

A special case occurs if the source and lens are perfectly aligned ($\beta = 0$). In this case, $\theta_1 = -\theta_2 = \theta_{\rm E}$. There is no preferred plane, and instead of two individual images, a ring is formed around the lens called the *Einstein Ring* of radius, $\theta_{\rm E}$. An astronomical example is shown in Figure 7.5.

Determining $\theta_{\rm E}$ is very important since this quantity contains the important astronomical information. The distances to both the source and lens can be found from the Hubble Relation (Appendix F) if their redshifts are measured, the source redshift being the same as the redshift of any of its images. Then Eqs. (7.16), (7.18), and (7.19) can be combined to obtain the lens mass,

$$\mathrm{M_L} = -\frac{c^2\,\theta_1\,\theta_2}{4\,G}\left(\frac{d_{\rm L}\,d_{\rm S}}{d_{\rm S} - d_{\rm L}}\right) \qquad \mathrm{M_L} = (1.22\times 10^8)\,\theta_1|\theta_2|\left(\frac{d_{\rm L}\,d_{\rm S}}{d_{\rm S} - d_{\rm L}}\right)\mathrm{M}_\odot \tag{7.20}$$

The first equation is in cgs units with angles in radians, and the second requires θ in arcseconds and distances in Mpc. The significance of $\mathrm{M_L}$ is that it includes *all* mass, both light and dark. Thus, *gravitational lenses provide a means for measuring even the mass that cannot be seen*. We now have an independent method of obtaining masses, other than via the Doppler shifts and Newtonian mechanics described in Sect. 7.2.1.2.

In reality, sources and lensing masses will not be point-like. Figure 7.7b shows an example with an offset background elliptical galaxy of uniform brightness and a spherically symmetric lens. The background galaxy can be thought of as consisting of many individual point sources, each of which creates two images as before (see points X, Y, and Y' in the diagram). In general, a gravitational lens distorts the image which, in this case, is clearly no longer elliptical. Because a lens directs light that would normally not be seen towards the observer (consider the case of the Einstein ring, for example), the result is an *amplification* of the signal (see also Sect. 7.3.2). The flux of a

background source will actually be greater as a result of the lensing than it would be in the absence of the lens.

With a variety of possibilities for background light distributions and lensing mass distributions, it is not difficult to imagine that rather complex images can result, as Figure 7.6a illustrates. One must then *model* the system, putting in various mass distributions until the observed images are reproduced (see Figure 7.6b). Although some assumptions may be required, such modelling can place strong constraints, not only on the total mass of the lens, but on the spatial distribution of its mass as well.

7.3.2 Microlensing – MACHOs and planets

There are situations in which the bending angles are so small that the entire system is unresolved. The background source images and lens are collectively seen as a point source, making it virtually impossible to tell that gravitational lensing is occurring in any static orientation. However, the lensing mass will have a proper motion (Sect. 7.2.1.1) with respect to the background source and observer. As the lens passes between the source and observer, there will be an increase and then decrease in the observed flux of the background source due to the amplification of the signal during the lensing phase. This is called *microlensing*. The duration of the microlensing event is approximately the time over which the background source is within the Einstein radius of the foreground lens (Prob. 7.16). A characteristic light curve results, such as that shown for one star lensing another star in Figure 7.8.

It is important to be able to identify such a light curve as being due to a microlensing event and not simply to a variable star or a supernova, both of which show light curves that increase, then decrease in brightness. Aside from the shape of the curve, it is possible to separate true microlensing events from other variable phenomena by the frequency dependence of the curve. The light curves of variable stars and supernovae result from processes that involve the transfer of photons through matter which has some opacity. As we have seen earlier (e.g. Sects. 5.4.2, 6.2) opacity is highly frequency dependent and, as a result, there tend to be differences, frequency to frequency, in the light curves of variable stars and supernovae. Since microlensing is achromatic, however, the light curves should be identical at different frequencies. It is thus possible to identify true microlensing events this way[11].

The detection of microlensing relies on the chance alignment of objects at any time, an event which is intrinsically unlikely. However, modern telescopes, especially robotic ones with customized software, have the ability to monitor millions of objects and routinely extract the lensing candidates. In addition, if fields are chosen for which there are many possible background objects, then the statistics are much better. These approaches have been taken in a variety of experiments. For example, microlensing events have been observed when MACHOs in the halo of the Milky Way (see Sect. 3.2)

[11] In practise, the process is somewhat more difficult because each pixel receives the light of many stars. Thus, the problem is approached statistically.

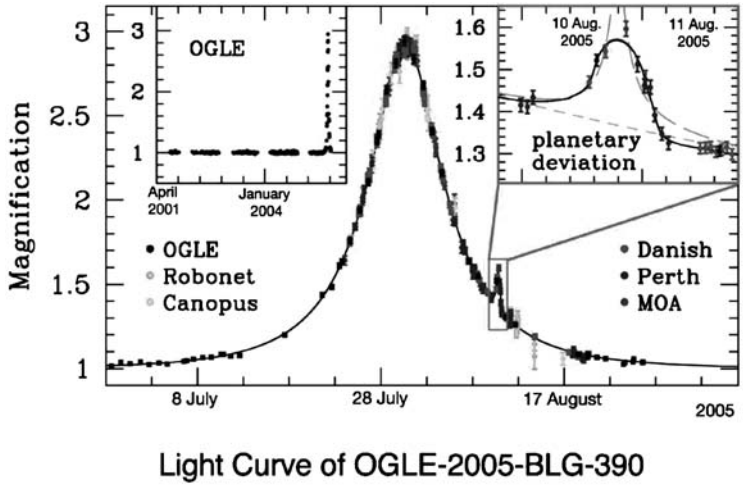


Figure 7.8 Characteristic light curve of a microlensing event. Both the lens and the background source are stars in our Milky Way. The background star is a G4 III giant star at a distance of 8.5 kpc. The lensing star is an M dwarf of mass, $M = 0.22M_{\odot}$, at a distance of 6.6 kpc. The left inset shows the light curve, as monitored over 4 years time. The small bump on the right, blown up in the right inset shows the smaller peak due to a planet, called OGLE-2005-BLG-390Lb, orbiting the lensing star. The planet's mass is $M_{pl} = 5.5M_{\oplus}$ and is a distance, $r = 2.6$ AU from its parent star. This is the third exoplanet discovered using microlensing techniques. The discovery was made by an international team (Ref. [13]) involved in three microlensing projects: Probing Lensing Anomalies NETwork (PLANET) + RoboNet, Optical Gravitational Lensing Experiment (OGLE), and Microlensing Observations in Astrophysics (MOA). (Reproduced by permission of ESO/PLANET/RoboNet, OGLE, and MOA)

pass in front of the *Large Magellanic Cloud* which is a nearby galaxy. It is studies like this that have led to the conclusion that there aren't enough MACHOs to account for the dark matter that is expected to be in the Milky Way's halo (Ref. [4], Ref. [3], Ref. [5] and Sect. 3.2). Figure 7.8 shows another exciting application of microlensing. The main light curve results from a background star in the bulge of the Milky Way being lensed by a foreground star. However, a small perturbation on the right of the curve, blown up in the right inset, shows another peak, this time due to a planet only 5.5 times the mass of the Earth, that orbits the lensing star. Thus, microlensing is now detecting planets with masses approaching that of the Earth, extending the range of possible detections beyond those that can be explored by other methods (see Sect. 4.2.2).

7.3.3 Cosmological distances with gravitational lenses

Gravitational lensing on very large cosmological scales can also provide independent information on the distance to a source for the special case in which the source brightness is varying with time. An example is when the background source is a quasar, most of

which show time variability. When the source is offset from the lens centre and two images at θ_1 and θ_2 are seen, one on either side of the lens (see Figure 7.7), the path length on one side of the lens will be different from the path length on the other side. A change in brightness will therefore be observed in the image that has the shorter path length before the image that has the long path length. The time delay between the brightness change in the two images is related to the path difference. With measurements of θ_1 and θ_2 and a knowledge of the lens distance, the source distance, d_S, can be determined. Quasars are in the distant Universe where distances are normally determined by using the Hubble Relation (see Figure F.2), so a method like this is very important in providing another measure of distance. Since the redshift of the quasar can be measured (it is the value obtained from the image), this technique has been used to provide an independent determination of the Hubble constant, H_0, with reasonable results. However, uncertainties about the distribution of mass (which must include light and dark matter) in the lens has meant that the effectiveness of this technique is currently limited.

7.4 Time variability and source size

A simple, but powerful argument related to fluctuations in the brightness of a source has proven to be extremely useful in constraining source sizes which otherwise would be too small to measure (see Figure 7.9). The size of the region that is variable must be,

$$d \leq c\,\Delta t \tag{7.21}$$

where Δt is the timescale of the variability. This condition is a requirement of causality. If it were *not* the case, then the light that is emitted from a varying source at point A (see figure) would not be able to travel across the source to point B in a time, Δt. Point B, on the other side of the source, would therefore not 'know about' the variation at A so could not be in synchronicity with it. If points on a source are varying independently, then no global variation would be detected. This is similar to arguments presented earlier as to why planets shine and stars 'twinkle' (see Sect. 2.3.4). If variability *is* observed, then Eq. (7.21) places a limit on the size of the varying source.

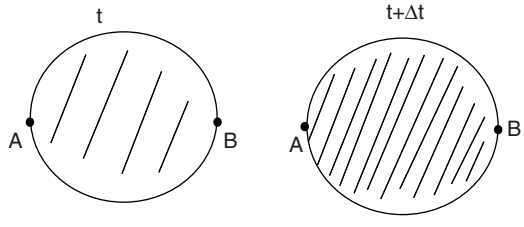


Figure 7.9 The size of a variable source must be less than the light travel time across it. At time, t (left), the source is dim and, at a time, $t + \Delta t$ (right), the source has brightened. All parts of the source can vary together, provided, $d \leq c\,\Delta t$.

Some sources are variable over a variety of timescales. It can be seen from Eq. (7.21) that the shortest measured timescale puts the tightest restrictions on source size. X-ray variability is particularly important because the fact that X-rays are observed indicates that the source is associated with high energy phenomena. X-rays tend to be emitted near very compact objects that have high gravitational fields, such as white dwarfs and neutron stars, or active galactic nuclei that are thought to be powered by supermassive black holes (e.g. Sect. 5.4.2). The shortest variability timescale then provides a limit on the size of the emitting object (Prob. 7.17). If information about the mass of an object is available from dynamics (Sect. 7.2.1.2) and information about the size of the object is available from time variability, then strong constraints can be placed on the type of object that is present, even if it is spatially unresolved.

Problems

7.1 How high would the radial velocity of an object have to be before the low velocity approximation for the Doppler shift differs from the accurate expression (see Eq. 7.2) by more than 10 per cent? (The solution may be obtained 'by hand' or via computer algebra software.)

7.2 In principle, any spectral feature for which the rest wavelength, λ_0, is known could be used to measure the radial velocity, v_r, of an object. Let us take the peak of the Planck curve of a star as the 'spectral feature'. The star has a radial velocity of $v_r = 50\,\mathrm{km\ s^{-1}}$ and we assume that its temperature, $T_0 = 5800\,\mathrm{K}$, has been determined by some other method.

(a) Find the rest wavelength, λ_0, for the peak of the Planck curve of this star and the observed, Doppler-shifted wavelength, λ, of this peak.

(b) Find the equivalent Doppler-shifted temperature, T corresponding to the new peak, λ.

(c) Compute the specific intensity of the star at the Doppler-shifted wavelength, λ, if there were no radial velocity (i.e. $B_\lambda(T_0)$). Compute the specific intensity of the moving star at the same wavelength (i.e. $B_\lambda(T)$). What is the change (in per cent) in specific intensity at the wavelength, λ, as a result of the star's motion?

(d) Assume that your measuring instruments can detect changes of about 3 per cent or larger and indicate whether the change indicated in part (c) could be detected.

(e) Compute the Doppler shift, $\Delta\lambda$, of the H α line for the same star. If the width of the H α line is $\Delta\lambda = 0.4$Å, can the H α Doppler shift be easily measured or not?

(f) Comment on the suitability of the Planck curve peak, in comparison to a spectral line, for measuring the radial velocities of stars.

7.3 How close does a star, moving with $v_t = 20\,\mathrm{km\ s^{-1}}$, have to be in order for its proper motion to be $\mu = 1''\ \mathrm{yr^{-1}}$ (Sect. 7.2.1.1)? Express the result in parsecs and in units of the distance of the Sun from the centre of our Galaxy (see Figure 3.3). Do you expect proper motions to be used to obtain v_t for a large fraction of the stars in the Galaxy?

7.4 From the information related to the following test particles, find the mass of the central object ($M_\odot$):

(a) The Earth in its orbit about the Sun.

(b) A roughly spherical cluster of galaxies for which the outermost galaxies are approximately 1 Mpc from the centre of mass and the maximum measured radial velocities are $\approx$ 300 km s^{-1} with respect to v_{sys} of the cluster.

(c) A globular cluster of stars for which the average radial velocity with respect to ν_{sys} is $v_r = 2.8$ km s^{-1}. The radius corresponding to this average radial velocity is 30 pc. Ignore star-star interactions and assume that the cluster can be represented by a spherical gravitational potential.

(d) A white dwarf orbiting an unseen object. The radius of its orbit is 10 AU and the circular velocity is 21 km s^{-1}.

7.5 Refer to Example 7.2 and, using the information given in Figures 2.11 and 2.20, determine the mass of NGC 2903.

7.6 A galaxy's nucleus harbours a supermassive black hole of mass, $M = 10^8\,M_\odot$, surrounded by an edge-on accretion disk that is spatially unresolved at X-ray wavelengths. The full width of a spectral line of iron, whose central energy is $E_{line} = 6.4$ keV, is due entirely to disk rotation and is 2.0 keV.

(a) What is the maximum circular velocity of this disk?

(b) What is the distance of the accretion disk, at its maximum velocity, from the centre of the black hole (AU)? Should this correspond to the inner edge of the accretion disk or the outer edge?

(c) Express the radius of part (b) in units of Schwarzschild Radii.

7.7 (a) Write an expression for the relative error in d_L that results from using Eq. (F.1) for the Hubble Relation instead of Eqn (F.16) and evaluate it for the values of q_0 used in the two curves shown in Figure (F.2). The expression will be a function of z.

(b) Determine this relative error for the two values of q_0 for (i) a galaxy with recessional motion of $c\,z = 5000$ km s^{-1}, (ii) a galaxy at $z = 0.1$ and (iii) a quasar at $z = 0.6$.

7.8 For $z < 1$ and $|q_0| < 1$, verify that the equation for the comoving coordinate, Eq. (F.10), results from Eq. (F.9). Note that, in this regime, any terms including $(t_0 - t)^3$ or higher orders can be neglected.

7.9 (a) Starting with the definition of redshift (Eq. 7.1), express the quantity, $(z + 1)$, in terms of the frequency of a wave, rather than its wavelength and, with this result, show that Eq. (F.14) follows.

(b) Supernova 1995 K, at a redshift of 0.479, was observed to have a light curve that is time dilated (Ref. [97]). If nearby supernovae of the same type normally show a light curve

width (full width at half maximum) of about 25 days, what was the width of the light curve of SN1995K?

7.10 (a) Starting with Eq. (F.16), derive an expression for the slope of the curve of the Hubble Relation shown in Figure F.2 (i.e. $d(cz)/d(d_L)$) as a function of z. Confirm that the slope reduces to H_0 for $z \rightarrow$ o.

(b) Measure the slope directly from Figure F.2 (dark solid curve) at two values of z and estimate H_0 and q_0 using the expression from part (a).

(c) Discuss the reasons for any differences between the results of part (b) and the values of q_0 and H_0 that have actually been plotted (see caption).

7.11 For a flat universe in which there is a cosmological constant, there is an early time at which the matter term dominates the behaviour of the expansion and a later time at which the expansion is dominated by Λ . For the epoch at which the energy densities of each component are equal ($\epsilon_M = \epsilon_\Lambda$), find numerical values for the following (assume $H_0 = 71\,\mathrm{km\ s^{-1}\,Mpc^{-1}}$, $\Omega_{M0} = 0.3$, $\Omega_{\Lambda 0} = 0.7$, and $q_0 = -0.55$):

(a) the density parameters, Ω_M and Ω_Λ, (b) the deceleration parameter, q, (c) the cosmological constant, Λ, (d) the Hubble parameter, H (km s^{-1} Mpc^{-1}), (e) the density, ρ, (f) the scale factor in comparison to the current value, a/a_0, (g) the redshift, z, and (h) the lookback time, $t_0 - t$ (Gyr).

7.12 Repeat Example 7.4 for a neutron star of mass, $M = 2\,M_\odot$, and radius, $r_{ns} = 20\,\mathrm{km}$.

7.13 Derive the lens equation (Eq. 7.15) from the geometry of Figure 7.7 and Eq. 7.14. (Recall that tan $\alpha = \alpha$, for small angle, α.)

7.14 (a) Write the Einstein Ring angle equation (Eq. 7.16) for a case in which the background source is very distant in comparison to the lens. Express the result in units of arcsec with constants evaluated, for input masses in $M_\odot$ and distances in pc.

(b) Using the result of part (a), compute θ_E (arcseconds) for the following cases: (i) the Sun, (ii) a neutron star of mass, $1.5\,M_\odot$, at a distance of 60 pc, (iii) a galaxy of mass, $10^{11}\,M_\odot$ at a distance of 10 Mpc, and (iv) a cluster of galaxies of mass, $10^{14}\,M_\odot$ at a distance of 20 Mpc.

(c) Using the same equation and the information given about Abell 2218 in Figure 7.6a, make a rough estimate of the mass of the bright subcluster visible on the right hand side of Figure 7.6a.

7.15 (a) Consider four very distant background quasars that are in the true configuration of a square on the sky of length 2″ on a side. One of these quasars is perfectly aligned with the centre of a foreground supermassive black hole of mass $M_{bh} = 4 \times 10^9\,M$ which is at the core of a galaxy of distance, $d_L = 1\,\mathrm{Mpc}$. Plot the images of the four stars that result from being gravitationally lensed by the black hole. The result of Prob. 7.14a may be useful.

(b) Suppose these four quasars instead delineated the four corners of a uniform brightness background source. Sketch on the plot the resulting images.

7.16 (a) Write an expression for the duration of a microlensing event, t_L, in terms of the transverse velocity of the lens, v_t, the distance to the lens, d_L, and the Einstein radius, θ_E.

(b) Beginning with the result of (a), re-express t_L (in days) in terms of d_L (kpc), d_S (kpc), M_L ($M_\odot$), and v_t (km s^{-1}).

(c) Evaluate t_L (days) for a lensing mass which is a 0.5 $M_\odot$ white dwarf in the Milky Way halo a distance of 5 kpc away with a transverse velocity of 220 km s^{-1}. The background star is in the Large Magellanic Cloud a distance 50 kpc away.

(d) From the information given in the caption of Figure 7.8, evaluate t_L (days) for the lensing star and compare the result to the observed curve duration. Assume that the transverse velocity (net of background star, lens and observer) is 220 km s^{-1}. Comment on the result.

7.17 An X-ray binary system consists of a normal star and either a white dwarf, neutron star, or black hole companion, in which material from the star is accreting onto the collapsed object. In some X-ray binaries, X-ray variability on timescales as short as a millisecond have been observed. If the timescale is this short, what is the maximum size of the source (km)? Is it possible to rule out any of the types of companions for such sources?

PART IV
The Signal Emitted

We have seen how a signal can be altered by our instruments, our atmosphere, and the matter and space that it encounters en route to our detectors. Each of these steps provided important information about the signal, about the intervening matter along and near its path, and even about the large-scale structure of our Universe. It is now time to turn our attention to the source, itself. How is the original signal actually emitted? This is an important question because, encoded in the signal, is information about the properties of the source. An understanding of the various processes involved in the emission of light will provide us with the keys to unlock these secrets.

Fortunately, there are only a limited number of ways in which charged particles couple to electromagnetic radiation. This means that, although the number of the objects in the Universe is incomprehensibly large, there are actually only a small number of light-emitting mechanisms that are responsible for all of the cosmic radiation that we see. Some of these mechanisms operate only at very high energies and so are relegated to a subset of exotic objects. Others are more common and widespread. Generally, emission processes fall into two categories: *continuum radiation*, emission that occurs over a broad spectral region, and monochromatic or *line radiation*, emission that occurs at a discrete wavelength with a narrow width in frequency. It is these processes that we now wish to consider.

8
Continuum Emission

A scientist must also be absolutely like a child. If he sees a thing, he must say that he sees it, whether it was what he thought he was going to see or not.
– *So Long, and Thanks for all the Fish*, by Douglas Adams

Continuum radiation is any radiation that forms a continuous spectrum and is not restricted to a narrow frequency range, the latter applying to quantum transitions in atoms and molecules. By *emission*, we mean a process that starts with matter and ends with the creation of a photon. This could involve an interaction between a charged particle and another charged particle, a charged particle and an electric or magnetic field, or a charged particle with another photon, the end result being a photon. Following the symbolism of Sects. 5.1 and 5.2, emission is represented as: matter $\rightarrow$ (matter or photon or field) $\rightarrow$ photon. Thus, even though scattering can be thought of as absorption and re-*emission* (see Sect. 5.1), the starting point is a photon so we do not consider it in this chapter[1].

Some commonly observed astrophysical continuum processes are included in this chapter, an important exception being black body radiation which was encountered in Chapter 4. The goal, for each type of continuum, is to obtain the intensity and *spectrum* of the signal, a process that involves determining the *emission coefficient*, j_ν (introduced in Eq. 6.2), and/or the *absorption coefficient*, α_ν Eq. (5.8), together with the appropriate solution of the Equation of Radiative Transfer (see Sect. 6.3) to find the specific intensity, I_ν. Other quantities such as flux density, luminosity, etc. follow from I_ν, as described in Chapter 1. To have an expression for I_ν is to have an essential astrophysical tool since it means that properties of the emission, and indeed of the source itself, can be characterized. To this end, it is useful to note that, for a uniform source, f_ν and L_ν will have the same spectral shape as I_ν.

[1]A decision as to 'starting point' is really a decision as to choice of rest frame. See additional comments in Sect. 8.6.

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

8.1 Characteristics of continuum emission – thermal and non-thermal

For continuum emission, it is helpful to remember that an accelerating charged particle that is not bound to an atom or molecule will radiate. Since we want to know the spectrum (a function of frequency, ν), it is also helpful to recall that frequency is inversely proportional to time. For example, an oscillating electron (Appendix D.1.1) will emit radiation at a frequency $\nu = 1/\tau$, where τ is the period of oscillation; an electron that has a brief 'collision' with an ion of duration, Δt, will emit a photon within some bandwidth $B \approx 1/\Delta t$. Thus, short timescales are related to high frequencies. In any gas in which there are free charged particles, there are many particles, many accelerations, and many timescales involved. Emission over a range of frequencies – a continuous spectrum – will result. As we will see, this range can be quite large, spanning many orders of magnitude in frequency-space. The Planck curve (Sect. 4.1), for example, takes on all frequencies between zero and infinity, although emission at some frequencies is much more probable than at others. Thus a characteristic of continuum emission from astronomical sources is that it is *broad-band*[2].

Other characteristics of continuum emission depend more specifically on the type of radiation being considered. These fall into two categories – *thermal emission* and *non-thermal emission* – and a few comments can be made for each.

Thermal emission is any process for which the observed signal is associated with a system whose states are populated according to a Maxwell–Boltzmann velocity distribution Eq. (0.A.2). Since this distribution defines the kinetic temperature T, for all thermal emission, I_ν may be a function of various parameters but it will certainly be *a function of the temperature*, T. The emitted radiation must therefore be related to random particle motions in the thermal gas. This can be achieved if the particles interact with each other via collisions. As indicated in Sect. 3.4.1, for Coulomb interactions, a collision is an encounter that is sufficiently close that it alters the trajectory of a charged particle. For a collisional process, we can apply the LTE solution to the Equation of Radiative Transfer (Eq. 6.19, see also Sect. 3.4.4) which, as we found in Sect. 6.4.1, implies that $T_B \leq T$ Eq. (6.24). In astrophysics, we rarely see temperatures above $\approx 10^{7-8}$ K, so an observed brightness temperature *higher than* this range is a clue that the process is likely non-thermal (see below). Therefore, a characteristic of thermal emission is *low brightness temperatures*. Thermal emission is also intrinsically *unpolarized*. Since there is no particular directionality to velocities in a thermal gas, there is no particular directionality to the resulting radiation. If polarization is observed, then either some other process is polarizing the signal after it is emitted (e.g. scattering by electrons or dust, see Appendix D) or else the emission is non-thermal.

Non-thermal emission is everything else. If I_ν is independent of T, then the process is non-thermal. Examples are when the emission depends on the acceleration of a charged particle in a magnetic field or an electric field, or when the emission depends on

[2]See, however, comments in Sect. 8.5.1.

collisions with other particles whose velocities are *not* Maxwellian (for example, a power law). Depending on the details of the specific emission process, the brightness temperature could be low or high and there may or may not be polarization of the resulting signal. We now examine these processes in more detail.

8.2 Bremsstrahlung (free–free) emission

Bremsstrahlung, which is derived from the German words for 'braking' and 'radiation', occurs in an ionized gas when a free electron travels through the electric field of a positively charged nucleus. The electron is 'braked' by the electrostatic attraction of the nucleus, feeling a Lorentz force, $\vec{F}_e = e\vec{E}$ (for zero magnetic field, Table I.1) as it passes by (Figure 8.1). The electron is accelerated (its direction is changed) by this 'collision' with the nucleus, and will therefore radiate. The process can also be thought of as the scattering of electrons in the electrostatic field of the nucleus. The resulting radiation is equivalently referred to as *free–free emission*. It is the inverse of the *free–free absorption* process discussed in Sect. 5.2 in which an electron absorbs a photon in the vicinity of a nucleus.

8.2.1 The thermal Bremsstrahlung spectrum

Derivations of thermal Bremsstrahlung emission can be found in Refs. [143], [101] and [79] and will not be repeated here. However, it is useful to offer a more qualitative description of this process in order to see how the functional dependences result.

One first considers a single electron in an encounter with a single nucleus. Given the mass difference between these particles, the nucleus can be considered at rest. Since we

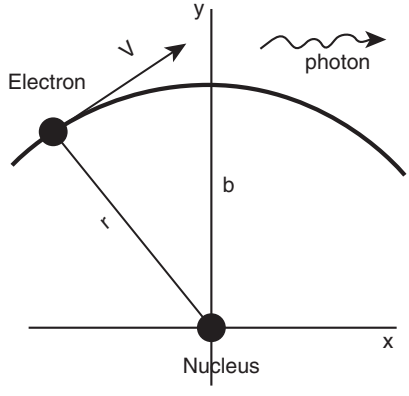


Figure 8.1. Geometry showing thermal Bremsstrahlung (free–free) emission resulting from an electron travelling by a nucleus of charge, Ze, where Z is the atomic number. The heavy curve shows the electron's path. An instantaneous velocity vector is shown for the electron at a position, $\vec{r}$, as well as an x,y coordinate system and the impact parameter, b, which is the distance of closest approach.

are dealing with a free–free process, the electron must not recombine with the nucleus. Rather, the kinetic energy of the electron, $E_{\rm k} = (1/2)\mathrm{m_e}v^2$, must be greater than the potential energy at the distance of closest approach to the nucleus, $E_{\rm P} = Ze^2/b$, where b (as we saw with gravitational lenses, Sect. 7.3.1) is the impact parameter and Z is the nuclear charge (atomic number). For a large range of astrophysical conditions, the change in angle is small and the change in electron speed is negligible. The energy of the emitted photon is then much less than $E_{\rm k}$ of the electron (Prob. 8.1) and the interaction can be treated as an elastic collision (Sect. 3.4.1) of duration, $\Delta t \approx b/v$. The power radiated by an accelerating charged particle is given by the *Larmor formula* (see Table I.1) which can be integrated over the duration of the encounter and converted to a dependence on frequency, ν, instead of time[3]. With the aid of some geometry and a knowledge of scattering angles for elastic collisions between charged particles, the result is an expression for the energy radiated by an electron of speed, v, and impact parameter, b.

The next step is to consider the radiation from an ensemble of particles that are all interacting with nuclei in this fashion. For many electrons at a single fixed velocity, a variety of impact parameters are possible so an integration over impact parameter is required. A reasonable minimum value, $b_{\rm min}$, is one at which the electron would be captured, and a maximum, $b_{\rm max}$, would be the value at which the perturbation is so small that virtually no change in trajectory results. In a real plasma, corrections to these limits are required, but they are helpful in understanding how the integration over impact parameter might proceed. Electrons that travel closest to the nucleus will experience the shortest duration interactions. Therefore, $b_{\rm min}$ determines the upper limit to the emitted frequency. Since the electron cannot come arbitrarily close to the nucleus, there should be a fairly sharp emission *cutoff* at high frequency for electrons of a given velocity.

The final step is to integrate over a distribution of electron velocities. Any velocity distribution that is thought to be present could be introduced at this point (e.g. a power law distribution or even relativistic velocities). However, we are here considering the most commonly considered Bremsstrahlung emission in astrophysics, that of *thermal* Bremsstrahlung which requires the Maxwellian velocity distribution Eq. (0.A.2). The final result is the *free–free emission coefficient* (cgs units of erg $\mathrm{s^{-1}\ cm^{-3}\ Hz^{-1}\ sr^{-1}}$),

$$\mathrm{j}_\nu = \frac{8}{3}\left(\frac{2\pi}{3}\right)^{1/2}\frac{Z^2e^6}{\mathrm{m_e}^2c^3}\left(\frac{\mathrm{m_e}}{k\mathrm{T_e}}\right)^{1/2} n_{\rm i}\, n_{\rm e}\, \mathrm{g_{ff}}(\nu, \mathrm{T_e})\, e^{-\left(\frac{h\nu}{k\mathrm{T_e}}\right)} \tag{8.1}$$

$$= 5.44 \times 10^{-39}\left(\frac{Z^2}{\mathrm{T_e}^{1/2}}\right) n_{\rm i}\, n_{\rm e}\, \mathrm{g_{ff}}(\nu, T_{\rm e})\, e^{-\left(\frac{h\nu}{k\mathrm{T_e}}\right)} \tag{8.2}$$

where $n_{\rm i}$, $n_{\rm e}$ are the ion and electron densities, respectively, $\mathrm{T_e}$ is the electron temperature, and Z is the atomic number. The quantity, $\mathrm{g_{ff}}(\nu, \mathrm{T_e})$, is a correction factor (for example, it includes corrections to $b_{\rm min}$ for quantum mechanical effects) called the *free–free Gaunt factor*[4] and is a slowly varying function of frequency and temperature. The Gaunt factor takes on different functional forms depending on temperature and

[3] A proper treatment requires the application of a mathematical function called a *Fourier Transform*. A Fourier analysis allows one to relate processes in the time domain to the observed frequency domain.

[4] This quantity is sometimes designated with a bar, $\bar{\mathrm{g}}_{\rm ff}$, because it has been averaged over a velocity distribution.

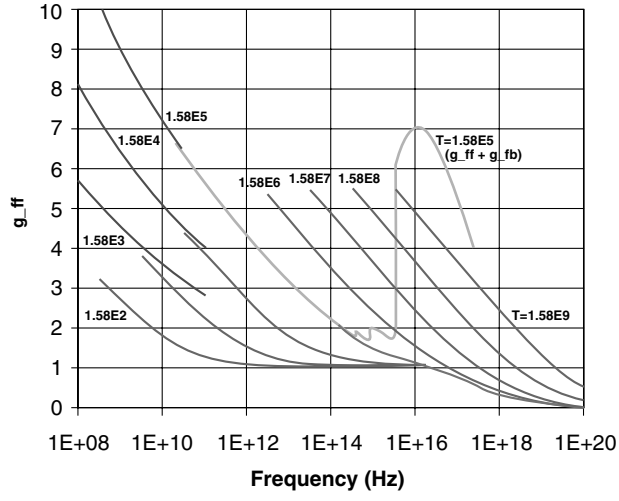


Figure 8.2. Value of the free–free Gaunt factor, $g_{ff}(\nu, T_e)$ (taking $Z = 1$) for a wide variety of conditions. Each curve is labelled with a temperature. Dark solid curves are accurate values taken from Ref. [166]. The three shorter curves at the left are the radio wavelength approximations given by Eq. (8.11). The grey curve is the combined free–free plus free–bound Gaunt factor for the temperature, 1.58×10^5 K from Ref. [28]. Therefore, $g_{fb}(\nu, T_e)$ for this temperature is represented by the excess above the smooth curve of $g_{ff}(\nu, T_e)$ below it.

frequency regime (e.g. see Ref. [28]) and is plotted for a wide range of parameters in Figure 8.2. As indicated when Kramers' Law for absorption was introduced (Eq. 5.17), the value of the free–free Gaunt factor is close to unity.

Since the shape of the thermal Bremsstrahlung spectrum follows that of the emission coefficient for optically thin radiation (see Eq. 6.22 and below), it is worthwhile pausing to consider the frequency dependence of j_ν (Eq. 8.1 or 8.2). At very high frequencies (specifically, $h\nu > kT_e$), the exponential term ensures that the emission coefficient cuts off rapidly, that is, electrons of (most probable, Eq. 3.8) energy, kT_e, cannot emit photons of energy $h\nu > kT_e$. If this cutoff frequency can be observed, then the temperature of the gas can be determined without the requirement of any other measurement (Prob. 8.2). Aside from this high frequency cutoff, the only frequency dependence is within the Gaunt factor. Therefore, the spectral shape of the Gaunt factor gives the spectral shape of the emission coefficient up to the cutoff. The Gaunt factor shows little variation with frequency for an object of given temperature and, as we shall see below, *the optically thin thermal Bremsstrahlung spectrum is a flat spectrum.*

The emission coefficient, j_ν, tells us the emissive behaviour of a pocket of gas without allowance for its internal absorption. In order to derive the full spectrum, I_ν, we also need an expression for the absorption coefficient, α_ν, as required by the Equation of Radiative Transfer (Eqs. (6.3) and (6.4)). In the absence of a background source, the absorption coefficient tells us how much *self-absorption* is occurring in the gas, that is, how effectively the gas absorbs its own radiation. Since the thermal Bremsstrahlung mechanism arises from a collisional process, the LTE equations can be used (see Sect. 3.4.4). This means that the absorption and emission coefficients are related via the

Planck function, $B_\nu(\mathrm{T}) = \mathrm{j}_\nu/\alpha_\nu$ Eq. (6.18), so α_ν and therefore the optical depth, τ_ν, (Eqs. 5.8 and 5.9) can be immediately obtained for a line of sight, l. Inserting Eq. (8.1) for j_ν,

$$\tau_\nu = \int \alpha_\nu \, dl = \int \frac{\mathrm{j}_\nu}{B_\nu(\mathrm{T})} \, dl$$

$$= \left(\frac{4\,e^6}{3\,\mathrm{m_e}\,h\,c}\right)\left(\frac{2\,\pi}{3\,k\,\mathrm{m_e}}\right)^{1/2} \mathrm{T_e}^{-1/2}\, Z^2\, \nu^{-3}\,(1 - e^{-\frac{h\nu}{k\mathrm{T_e}}})\, \mathrm{g_{ff}}(\nu, \mathrm{T_e}) \int n_\mathrm{e} n_\mathrm{i}\, dl \qquad (8.3)$$

$$\approx (3.7 \times 10^8)\, \mathrm{T_e}^{-1/2}\, Z^2\, \nu^{-3}\,[1 - e^{-\frac{h\nu}{k\mathrm{T_e}}}]\, \mathrm{g_{ff}}(\nu, \mathrm{T_e})\, \mathcal{EM} \qquad (8.4)$$

Ref. [143], where $n_\mathrm{e} \approx n_\mathrm{i}$ is assumed[5] and, since the temperature and Gaunt factor usually do not change by large factors along a line of sight for any given cloud, these quantities are taken outside of the integral. The quantity, $\mathcal{EM}$, is called the *emission measure*, defined by,

$$\mathcal{EM} \equiv \int {n_\mathrm{e}}^2\, dl \approx {n_\mathrm{e}}^2\, l \qquad (8.5)$$

The emission measure[6] contains information about the number density of particles along the line of sight. The approximation of Eq. (8.5) is applicable if the density is approximately constant over l. However, even if the density varies, the emission measure still provides a useful approximation of the density.

The frequency dependence of the expression for τ_ν in Eq. (8.3) is of interest. Since there is little emission when $h\,\nu \gg k\,\mathrm{T_e}$, we restrict our interest to the regime, $h\,\nu \ll k\,\mathrm{T_e}$. Here $\exp[-(h\,\nu)/(k\,\mathrm{T_e})] \approx 1 - (h\,\nu)/(k\,\mathrm{T_e})$ Eq. (A.3) so the term in parentheses $\propto \nu$. Since the Gaunt factor has only a weak dependence on frequency, it can be considered roughly constant, so $\tau_\nu \propto \nu^{-2}$. This means that the opacity of the cloud *increases at low frequencies*.

The final spectrum, in the absence of a background source, is obtained by inserting the Planck Function (Eq. (4.1)) and the above expression for optical depth (Eq. (8.3)) into the LTE solution of the Equation of Transfer Eq. (6.19) which, due to the complexity of the expression, we leave in its original form here,

$$I_\nu = B_\nu(\mathrm{T_e})\,(1 - e^{-\tau_\nu}) \qquad (8.6)$$

Results for two different plasmas, one applicable to 10^4 K gas from an HII region and one applicable to 10^6 K gas in a cluster of galaxies, are shown in Figure 8.3.

[5]He has an ionization energy of $E_\mathrm{i} = 24.6$ eV and so is ionized only in a 'harder' radiation field than hydrogen for which $E_\mathrm{i} = 13.6$ eV. If HeII is present, it contributes only one electron per ion, consistent with our assumption. HeIII, which contributes two electrons per ion, is negligible in HII regions (Ref. [145]). However, if the situation arises that He is fully ionized, then $n_\mathrm{e}/n_\mathrm{i} = 1 + 2\,N_\mathrm{He}$, where N_He is the fractional abundance, by number, of He. For Solar abundance, $N_\mathrm{He} = 0.085$ (Sect. 3.3.4), so a 'worst case' error made by this assumption is 17 per cent.
[6]Sometimes the Z^2 of Eq. (8.4) is included in the definition of emission measure.

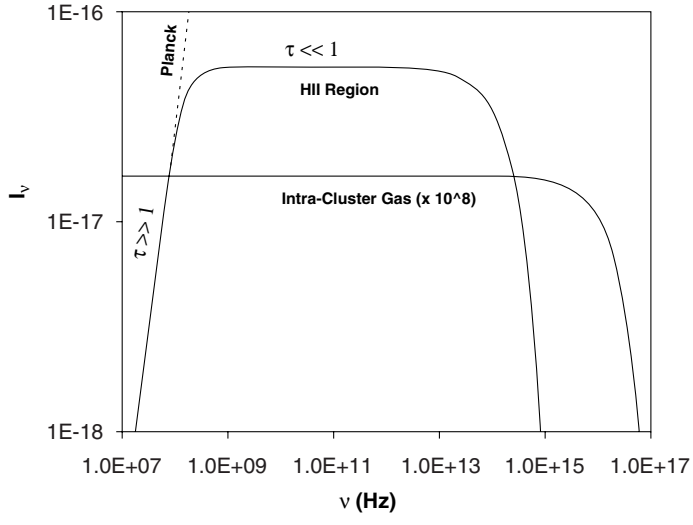


Figure 8.3. Two examples of the thermal Bremsstrahlung spectrum with ordinate (I_ν) in cgs units. For the HII region, the parameters are: $n_\mathrm{i} = n_\mathrm{e} = 100\,\mathrm{cm}^{-3}$, $\mathrm{T_e} = 10^4$ K, $l = 5$ pc, $Z = 1$, $g_\mathrm{ff} \approx$ constant $= 6.5$. For the hot intracluster gas: $n_\mathrm{i} = n_\mathrm{e} = 10^{-4}\,\mathrm{cm}^{-3}$, $\mathrm{T_e} = 10^6$ K, $l = 1$ Mpc, $Z = 1$, $g_\mathrm{ff} \approx$ constant $= 1.0$, and the specific intensity has been multiplied by a factor of 10^8 to appear on the plot. For the H II region, the optically thick and optically thin regions are marked as well as the Planck curve at an equivalent temperature. The high frequency exponential cutoffs are evident for both curves but the optically thick self-absorbed region for the intracluster gas occurs at a frequency lower than shown on the plot.

The flat nature of the thermal Bremsstrahlung spectrum is quite evident in this figure.

As we did in Sect. 6.4.1, we can look at the optically thick and optically thin part of the spectrum. In the absence of a background source see Eqs. (6.20) and (6.22)

$$I_\nu = B_\nu(\mathrm{T}) \quad (\tau_\nu \gg 1) \tag{8.7}$$
$$I_\nu = \mathrm{j}_\nu\, l \qquad (\tau_\nu \ll 1) \tag{8.8}$$

In Figure 8.3, Eq. (8.7) applies to the region where the spectrum turns down at low frequency, becoming that of a black body. In this limit the measured specific intensity is a function only of temperature. As the frequency increases, the optical depth decreases, and the spectrum then departs from the Planck curve. Eq. (8.8) applies to higher frequencies when $\tau_\nu \ll 1$ and the spectrum follows that of the emission coefficient (Eq. 8.1) whose frequency response is that of the Gaunt factor, being flat and then cutting off exponentially at high frequency. In this part of the spectrum, the emission depends on both the temperature and density. In the flat part of the spectrum (from Eqs. 8.2 and 8.8, and 8.5) the specific intensity, $I_\nu \propto {n_\mathrm{e}}^2\, l/{\mathrm{T_e}}^{1/2} \propto \mathcal{EM}/{\mathrm{T_e}}^{1/2}$, so the source brightness is much more sensitive to density than temperature.

Now that we know the spectrum of free–free emission and how it depends on the parameters of the source, we can use the *observed* specific intensity of the source to obtain the source parameters. From the Bremsstrahlung spectrum alone, this could be accomplished by making a measurement at low frequency in the optically thick regime to obtain T_e and then making another measurement in the optically thin regime to obtain n_e. Note that the line-of-sight distance through the source must also be known, but this is often assumed to be about equal to the source diameter, the latter obtained from the source distance and apparent diameter (Eq. (B.1)). Once the density is found, the hydrogen mass, M_{HII}, or total gas mass, M_g of the ionized region can be found from,

$$M_{HII} = m_H n_e V = X M_g \tag{8.9}$$

where m_H is the mass of the hydrogen atom, V is the volume of the ionized region, and X is the mass fraction of hydrogen (X = 0.707 for Solar abundance, Eq. 3.4). Note that, although dust may play an important role in intercepting ionizing photons, it contributes negligibly to the total mass (Sect. 3.4.1).

Free–free emission is most commonly seen in the radio part of the spectrum when the emission is from gas at $\approx 10^4$ K such as HII regions around hot young stars, and in the X-ray part of the spectrum when the gas is at $\approx 10^{6-7}$ K such as diffuse gas in clusters of galaxies or halos around individual galaxies. As Figure 8.3 illustrates, the spectrum is broader than these wavebands alone but free–free emission is most easily isolated from other processes in these wavebands. We will therefore elaborate on radio and X-ray waveband observations in the next sections.

8.2.2 Radio emission from HII and other ionized regions

Regions of ionized gas, such as HII regions (see Figure 8.4), emit by thermal Bremsstrahlung and properties of these regions can be derived through observations of their radio emission. Radio observations are exceedingly important in astrophysics because radio waves are not impeded by interstellar dust. While optical observations must contend with extinction (Sect. 5.5), radio wavelengths are much longer than interstellar grain sizes so can pass through without interaction (recall that Rayleigh scattering $\propto 1/\lambda^4$ when $\lambda \gg$ the particle size, Appendix D.3). Thus, measurements in this waveband can provide information about the source without having to make often uncertain corrections for extinction. In some cases, especially for HII regions that are embedded in dusty molecular clouds or behind many magnitudes of extinction in the plane of the Milky Way, the region cannot be seen at all optically and observations at radio wavelengths may be the only way of detecting them (see, for example, Figure 9.6).

For any HII region at radio wavelengths, $h\nu \ll k T_e$ (Prob. 8.4) and certain simplifications can be made to the thermal Bremsstrahlung equations. The Planck curve, for example, can be represented by the Rayleigh–Jeans approximation

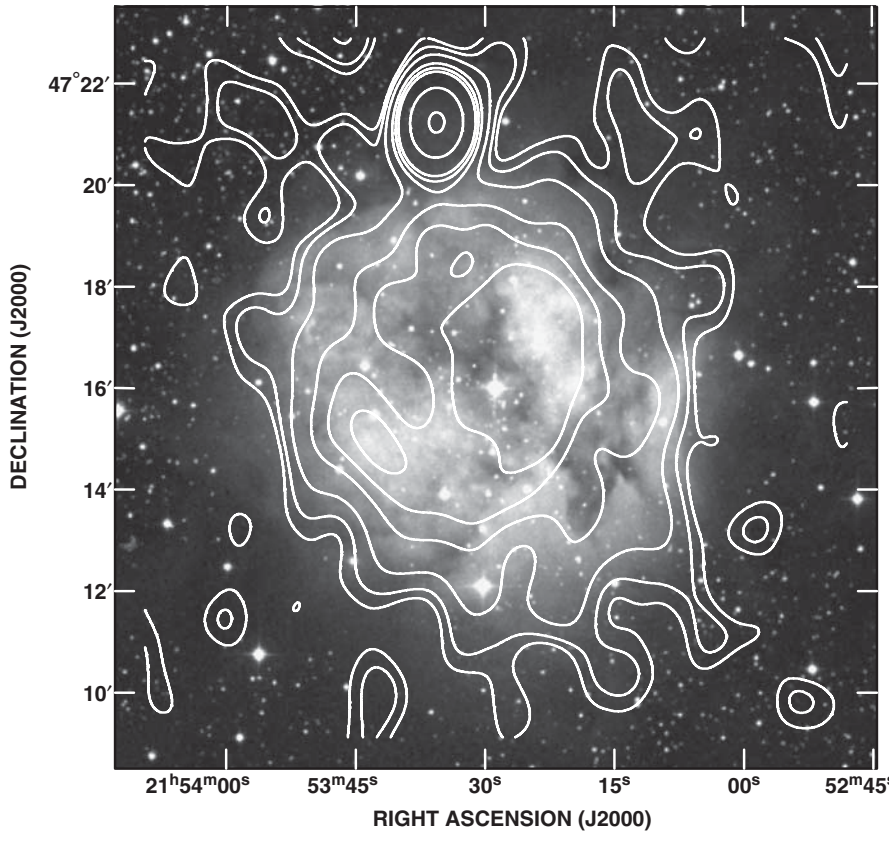


Figure 8.4. The HII region, IC 5146, at a distance of $D = 0.96$ kpc. The greyscale image is an R-band optical image. (Reproduced courtesy of the Palomar Observatory-STS$_c$ I Digital Sky Survey, Cal Tech). The contour overlays have been made from $\lambda 92$ cm data of the Westerbork Northern Sky Survey (WENSS, Ref. [134]). Contour levels are at 5, 10, 20, 40, 55, 70, 159, and 300 mJy beam^{-1} and the map peak is 318 mJy beam^{-1}. The beam is Gaussian with a FWHM of $\approx 80''$. The total radio flux at this wavelength is $S_{\lambda 92\,\mathrm{cm}} = 1.44$ Jy and the electron temperature is $\mathrm{T_e} = 9000$ K. The ionizing star, BD+46°3474, can be seen at the centre. The strong point source at the top of the image is an unrelated background source.

(Eq. (4.6)), so the *optically thick* emission at low frequency can be represented by in (cgs. units),

$$I_\nu = B_\nu(\mathrm{T_e}) = \frac{2\,\nu^2}{c^2}\,k\,\mathrm{T_e} \propto \nu^2 \quad (\tau_\nu \gg 1) \tag{8.10}$$

For *optically thin* emission, we require a value for the Gaunt factor. When $\mathrm{T_e} < 9 \times 10^5$ K, which will be the case for HII regions, the Gaunt factor can be represented analytically by (Ref. [96]),

$$\mathrm{g_{ff}}(\nu, \mathrm{T_e}) = 11.962\,\mathrm{T_e}^{0.15}\,\nu^{-0.1} \tag{8.11}$$

Since the frequency dependence of the optically thin spectrum follows the frequency dependence of the Gaunt factor (Eq. (8.2)), Eq. (8.11) explicitly shows the flat nature of the optically thin spectrum. With these assumptions and expanding the term in brackets in Eq. (8.4), the latter expression for optical depth can be rewritten in common astronomical units as,

$$\tau_\nu \;=\; 8.24 \;\times\; 10^{-2} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{-1.35} \left[\frac{\nu}{\mathrm{GHz}}\right]^{-2.1} \left[\frac{\mathcal{EM}}{\mathrm{pc\,cm^{-6}}}\right] \tag{8.12}$$

The brightness temperature of an optically thin source can now be written by using Eq. (8.12) with Eq. (6.33) to find,

$$\mathrm{T_{B\nu}} \;=\; 8.24 \;\times\; 10^{-2} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{-0.35} \left[\frac{\nu}{\mathrm{GHz}}\right]^{-2.1} \left[\frac{\mathcal{EM}}{\mathrm{pc\,cm^{-6}}}\right] \quad (\tau_\nu \ll 1) \tag{8.13}$$

Using Eq. (8.13) with the Rayleigh–Jeans relation (Eq. 4.6) again, we can see the frequency dependence of the specific intensity,

$$I_\nu \;=\; \frac{2\,\nu^2}{c^2}\, k\,\mathrm{T_{B\nu}} \;\propto\; \nu^{-0.1} \quad (\tau_\nu \ll 1) \tag{8.14}$$

It is sometimes more useful to measure the total flux density, f_ν, of an HII region which is related to the specific intensity via $f_\nu = I_\nu\,\Omega$ (Eq.(1.13)). Together with the Rayleigh–Jeans equation, the brightness temperature can then be expressed in terms of source flux density and Eq. (8.13) can be re-arranged to solve, explicitly, for various parameters of an HII region. These are the electron density, n_e, the hydrogen mass, $\mathrm{M_{HII}}$, the emission measure, $\mathcal{EM}$ (defined in Eq. 8.5), and the excitation parameter, $\mathcal{U}$ (see Eq. (5.4)). For the specific geometry of a spherical HII region of uniform density, the results are (Ref. [151]),

$$\left[\frac{n_\mathrm{e}}{\mathrm{cm^{-3}}}\right] \;=\; 175.1 \left[\frac{\nu}{\mathrm{GHz}}\right]^{0.05} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{0.175} \left[\frac{f_\nu}{\mathrm{Jy}}\right]^{0.5} \left[\frac{D}{\mathrm{kpc}}\right]^{-0.5} \left[\frac{\theta_\mathrm{sph}}{\mathrm{arcmin}}\right]^{-1.5} \tag{8.15}$$

$$\left[\frac{\mathrm{M_{H\,II}}}{\mathrm{M_\odot}}\right] \;=\; 0.05579 \left[\frac{\nu}{\mathrm{GHz}}\right]^{0.05} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{0.175} \left[\frac{f_\nu}{\mathrm{Jy}}\right]^{0.5} \left[\frac{D}{\mathrm{kpc}}\right]^{2.5} \left[\frac{\theta_\mathrm{sph}}{\mathrm{arcmin}}\right]^{1.5} \tag{8.16}$$

$$\left[\frac{\mathcal{EM}}{\mathrm{pc\,cm^{-6}}}\right] \;=\; 8920 \left[\frac{\nu}{\mathrm{GHz}}\right]^{0.1} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{0.35} \left[\frac{f_\nu}{\mathrm{Jy}}\right] \left[\frac{\theta_\mathrm{sph}}{\mathrm{arcmin}}\right]^{-2} \tag{8.17}$$

$$\left[\frac{\mathcal{U}}{\mathrm{pc\,cm^{-2}}}\right] \;=\; 4.553 \left\{ \left[\frac{\nu}{\mathrm{GHz}}\right]^{0.1} \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{0.35} \left[\frac{f_\nu}{\mathrm{Jy}}\right] \left[\frac{D}{\mathrm{kpc}}\right]^{2} \right\}^{1/3} \tag{8.18}$$

where[7] $\theta_{\rm sph}$ is the angular diameter of a spherical HII region as seen on the sky and D is its distance. A measurement of flux density at a radio frequency in the optically thin limit, with a knowledge of the temperature and distance, are sufficient to obtain these quantities (Prob. 8.5). Table 3.1 provides some typical results. The excitation parameter (see Eq. 5.4) can then be related to the properties of the ionizing star or stars, as was described in Example 5.4. Although many HII regions are not perfectly spherical, their line-of sight distances are not usually significantly different from their diameters, leading to only minor adjustments in the above equations.

Historically, this approach has most often been applied to HII regions but, in principle, the above equations for the radio regime apply to *any* discrete ionized region, provided $T_e \ll 9 \times 10^5$ K.

Planetary nebulae, for example, which are ionized gaseous regions formed by the ejecta of dying stars (Sect. 3.3.2), have properties that are not very different from HII regions. They are ionized by a central white dwarf and have similar temperatures with (on average) higher densities. Radio observations of free–free emission from these nebulae have also been carried out, revealing their properties (see Table 3.1). Planetary nebulae have historically been a more challenging target than HII regions, however, because of their comparatively smaller sizes (less than a pc).

Another example is ionized gas in the Milky Way that is not in discrete regions, but rather spread out throughout the ISM. This is diffuse, low density (e.g. ≈ 0.01 cm^{-3} though it varies) gas and therefore the emission from this ISM component is weak. It is called either the *warm ionized medium* (WIM, more often used for the Milky Way) or the diffuse ionized gas (DIG, more often used when observed in other galaxies). The WIM/DIG is likely caused by ionization of interstellar HI by hot stars that are close to the boundaries of their parent clouds so that ionizing photons leak out of the surrounding HII regions into the general ISM. The above equations would require a modification for geometry if the line of sight through the gas is significantly different from the diameter, but otherwise are still applicable. The challenge in extracting information about this component in the radio regime, however, is that the free–free radio emission tends to be contaminated by synchrotron emission (see Sect. 8.5), requiring that the synchrotron fraction, if known, be subtracted first. Other methods for probing this diffuse component are also available, fortunately, such as observing the signals from background pulsars in the case of the Milky Way (Sect. 5.3) or measuring optical line emission (Sect. 9.4.2).

8.2.3 X-ray emission from hot diffuse gas

The X-ray wave band between about 0.1 keV and 5 keV (the 'soft' X-ray band, corresponding to $\nu \approx 10^{16} \rightarrow 10^{18}$ Hz) is dominated by emission from hot ($\approx 10^{6-8}$ K),

[7] The equation for electron density, n_e, includes contributions from singly ionized He, if present, so the hydrogen mass, $M_{\rm HII}$, computed from Eq. (8.16) should be *reduced* by a correction factor, $1/[1 + N({\rm HeII})/N({\rm HII})]$, to account for this, where $N({\rm HeII})/N({\rm HII})$ is the ratio of abundances of singly ionized He to ionized hydrogen. For Solar abundance, this ratio should be less than 8.5 per cent, depending on how many He atoms are ionized, and therefore we have left the correction factor out of Eq. (8.16).

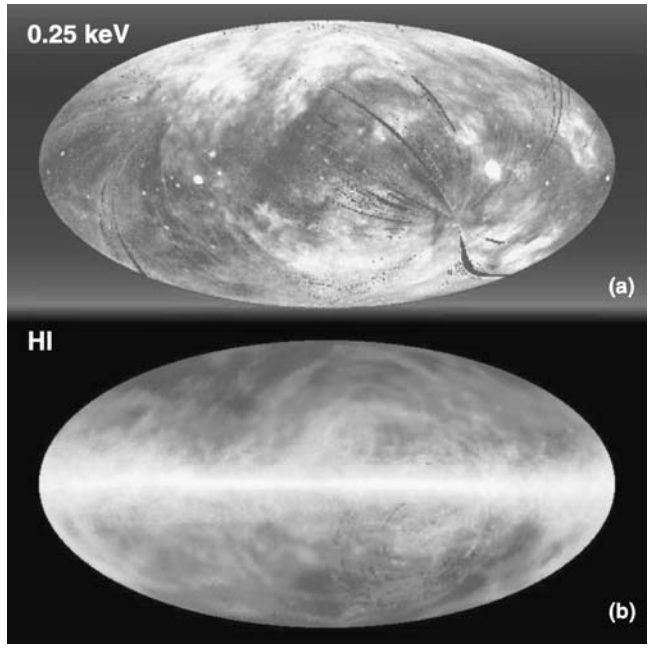


Figure 8.5. Illustration of the *anti-correlation* between the soft X-ray emission and total HI distribution in the Galaxy, shown over the whole sky in this Hammer–Aitoff projection (see Figure 3.2). The centre of the map is the direction towards the nucleus of the Milky Way, called the Galactic Center (GC), and the right and left edges correspond to the *anti-centre* direction, 180° from the GC. The top and bottom of the maps correspond to the North Galactic Pole (NGC) and South Galactic Pole (SGP), respectively, and the plane of the Milky Way runs horizontally through the centre. See Figure 3.3 for a map of the Galaxy as it would be seen from the outside. **(a)** The 0.25 keV soft X-ray band, observed with the ROSAT satellite (Ref. [158]). The long curved streaks are artifacts. Most of the emission is seen towards the Galactic poles because much of the X-ray emission along the plane of the Galaxy has been absorbed in neutral hydrogen clouds. **(b)** All-sky map of the total (summed over all velocities) neutral hydrogen emission in the Milky Way (Ref. [46]). Most of the emission is along the Galactic Plane.

low density (10^{-4} to 10^{-2} cm^{-3}) gas that emits via thermal Bremsstrahlung. In this band, the gas is always optically thin (Prob. 8.6). Since $h\nu \not\ll k\,\mathrm{T_e}$, the exponential term in Eqs. (8.1) and (8.2) is not approximately 1 as it was for radio observations of HII regions. At 1 keV, for example, the gas temperature would have to be higher than 10^8 K for $h\nu \ll k\,\mathrm{T_e}$. Therefore the exponential term in these equations is important and the emission is observed on or near the high frequency exponential tail (Figure 8.3). This also means that the electron temperature can be determined (Sect. 8.2.1) by a fit to the shape of the spectrum in this region.

A challenge of observing in the soft X-ray band is that these photons are very easily absorbed in the neutral hydrogen clouds that are abundant in the ISM[8]. This is beautifully illustrated by Figure 8.5 which shows the entire sky in the soft X-ray band and in HI. The

[8]The absorption is mainly by bound electrons in the heavy elements (e.g. oxygen) that are within these HI clouds.

X-ray map is dominated by free–free emission from hot gas, though contributions from other emission processes are also present. Most of the X-ray emission is seen in the direction of the Galactic poles because, along the Galactic plane, X-rays are absorbed in the abundant HI clouds of the disk of the Galaxy (see Figure 3.3 for a sketch of the Milky Way). A correction for this absorption must first be applied, therefore, before an accurate X-ray intensity can be found. Since we have a good map of HI for our Galaxy, such a correction is routinely carried out, though it adds to the error bar of the result.

Once a correction for interstellar absorption has been made, it is common to quote an X-ray luminosity (in erg s^{-1}) for the object. This can be obtained by converting the specific intensity (Eq. 8.8) to a luminosity (Prob. 8.7). Although thermal Bremsstrahlung emission is expected to dominate in the soft X-ray band, some corrections may still be required to account for contributions to the luminosity from emission mechanisms other than thermal Bremsstrahlung alone. Possibilities include free–bound emission (Sect. 8.3) or line emission. Assuming that the emission represents only the free–free component, the luminosity of a uniform density, constant temperature plasma with $Z = 1$ and $n_e \approx n_i$ is,

$$L_X = 6.84 \times 10^{-38} \frac{{n_e}^2}{{T_e}^{0.5}} V \int_{\nu_1}^{\nu_2} g_{ff}(\nu, T_e)\, e^{-\frac{h\nu}{kT_e}}\, d\nu \tag{8.19}$$

where ν_1, ν_2 are the lower and upper frequencies of the band in which the emission is observed and V is the volume occupied by the emitting gas.

Taking the Gaunt factor to be a constant ($g_{ff}(\nu, T_e) \approx g_X$) and using this equation and Eq. (8.9), the electron density and mass of the ionized hydrogen gas[9] can be found,

$$\left[\frac{n_e}{cm^{-3}}\right] = (1.55 \times 10^{-19} f_X) \left[\frac{L_X}{erg\, s^{-1}}\right]^{0.5} \left[\frac{V}{kpc^3}\right]^{-0.5} \left[\frac{T_e}{K}\right]^{-0.25} \tag{8.20}$$

$$\left[\frac{M_{H\,II}}{M_\odot}\right] = (3.81 \times 10^{-12} f_X) \left[\frac{L_X}{erg\, s^{-1}}\right]^{0.5} \left[\frac{V}{kpc^3}\right]^{0.5} \left[\frac{T_e}{K}\right]^{-0.25} \tag{8.21}$$

where f_X is a unitless function given by,

$$f_X = \left[g_X \left(e^{-\frac{E_1}{kT_e}} - e^{-\frac{E_2}{kT_e}}\right)\right]^{-0.5} \tag{8.22}$$

$E_1 = h\nu_1$ and $E_2 = h\nu_2$ being the lower and upper energies of the band over which the Bremsstrahlung spectrum is observed, respectively. It is common to express the temperature, not in Kelvins, but rather in energy units (keV) for a more straightforward comparison to the observational energy band. For example, a temperature of 10^7 K corresponds to $kT_e = 1.38 \times 10^{-9}$erg $= 0.863$ keV.

The interstellar absorption illustrated by Figure 8.5 is usually considered to be an undesirable effect that weakens the emission that we wish to observe and introduces

[9]Similar comments as given in Footnote 7 of this chapter apply here. However, the error may be larger, depending on metallicity, since at higher temperatures, more electrons may be released from heavy metals.

error into the result. Surprisingly, however, absorption due to HI clouds can actually *benefit* our understanding of the X-ray sky. The velocity of any object that emits a spectral line can be determined via the Doppler shift of the line (Sect. 7.2.1). For HI clouds in the Milky Way, these velocities can be translated into distances by adopting a model for the rotation of the Galaxy (Sect. 7.2.1.2). Since HI emits by the $\lambda 21$ cm line (Sect. 3.4.5, Appendix C) it is possible to place constraints on the distances to various HI clouds in the Galaxy. This is not true of continuum emission for which we cannot obtain velocities (e.g. Prob. 7.2). When continuum emission is observed, it could be coming from anywhere along a line of sight – from nearby hot gas or from the distant Universe. However, a detailed comparison of maps, like those in Figure 8.5, for which HI cloud distances are known, can help to determine how much X-ray emission is coming from gas in the foreground or background. If an HI cloud is obscuring the X-ray emission, for example, then the X-ray continuum must originate from behind the cloud, placing a limit on the distance to the X-ray-emitting gas. This effect is called *shadowing* and it has been used to help us understand the hot gas distribution in the Solar neighbourhood.

The results of this kind of analysis are not perfect, given the placements of clouds and the fact that sometimes uncertain corrections are required for X-ray emission from our own Sun. However, together with other information and analysis, a picture has emerged that the Sun is located in a *Local Hot Bubble* (LHB) in the Milky Way, as illustrated by Figure 8.6. The bubble is irregular in shape but appears to be elongated in the vertical

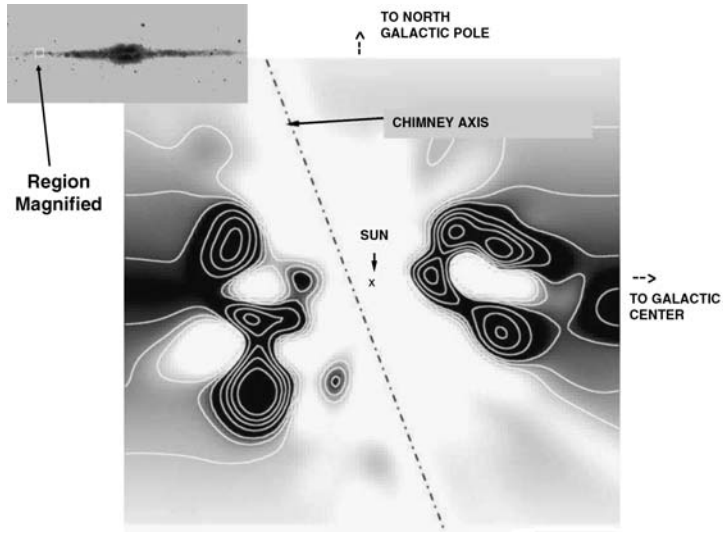


Figure 8.6. Diagram of the *Local Hot Bubble* (LHB) within which the Solar System is located. The top left image shows the region of the Galaxy that has been blown up for the bottom image. The LHB, of temperature, 10^6 K, and density, $\approx 10^{-3}$ cm^{-3}, is the irregularly shaped white region around the Sun extending of order 50–200 pc in size. Contours and shading are related to density. Various features are shown, including the chimney axis, which is the approximate direction that hot gas might be venting into the halo (Ref. [95]). (Reproduced by permission of Barry Welsh)

direction, roughly perpendicular to the plane. Such vertical low density regions are seen elsewhere in the Milky Way and also in other galaxies and are called *chimneys*. As the name implies, these are conduits or tunnels in cooler, denser ISM gas through which hot gas may flow into the Galaxy's halo[10]. They are likely produced by supernovae and stellar winds whose collective outflows have merged into larger structures.

In a number of spiral galaxies, the collective effects of the venting of hot gas have produced an observable soft X-ray halo around the galaxy. These are typically observed in spirals that are edge-on or close to edge-on to the line of sight so that the halo can be clearly seen independent of the disk. They also support a picture in which the disk–halo interface of a galaxy is a dynamic place through which outflows occur, and possibly inflows as well if cooler gas descends again, like a fountain. There may, however, be other reasons for the presence of X-ray halos around galaxies. The galaxy, NGC 5746, for example does not have sufficient supernova activity in its disk to produce the X-ray halo that is observed around it (see Figure 8.7). It may be that hot gas around this galaxy has been left over from the galaxy formation process itself (Ref. [119]). Example 8.1 provides an estimate of the density and mass of this halo.

Example 8.1

Assuming that the soft X-ray halo of NGC 5746 (Figure 8.7) is entirely due to thermal Bremsstrahlung, estimate the electron density and hydrogen mass of this halo.

From the figure caption, $T_e = 6.5 \times 10^6$ K and $L_X = 4.4 \times 10^{39}$ erg s^{-1}. The radius of the emission is about 20 kpc leading to a volume of $V \approx 3.35 \times 10^4$ kpc^3, assuming spherical geometry. The lower and upper bounds of the observing band are $E_1 = 0.3$ and $E_2 = 2$ keV, respectively. The mid-point of this band ($E = 1.15$ keV) corresponds (via $E = h\nu$) to $\nu = 2.8 \times 10^{17}$ Hz. From Figure 8.2, the Gaunt factor is then $g_X \approx 0.7$. By Eq. 8.22, this gives $f_X = 1.6$. Putting the above values into Eqs. (8.20) and (8.21) yields, $n_e = 0.002$ cm^{-3} and $M_{HII} = 1 \times 10^9 M_\odot$. These results are only approximate because this halo is likely to have a significantly varying density with radius. The assumption of pure thermal Bremsstrahlung must also be considered carefully (see Sects. 8.3 and 8.4).

In clusters of galaxies, immense reservoirs of hot gas have been detected between the galaxies, visible via their X-ray emission. An example is the galaxy cluster, Abell 2256, shown in Figure 8.8. This *intracluster gas* is the dominant baryonic component of mass in the cluster. In Abell 2256, for example, the total mass

[10]It is not always clear whether hot gas is currently flowing into the halo through individual chimneys or whether they are simply structures that may be relics or result from other factors such as magnetic fields. Therefore, the cooler 'walls' of such features that extend above the plane have also been called *worms* which has no implication for hot gas outflow.

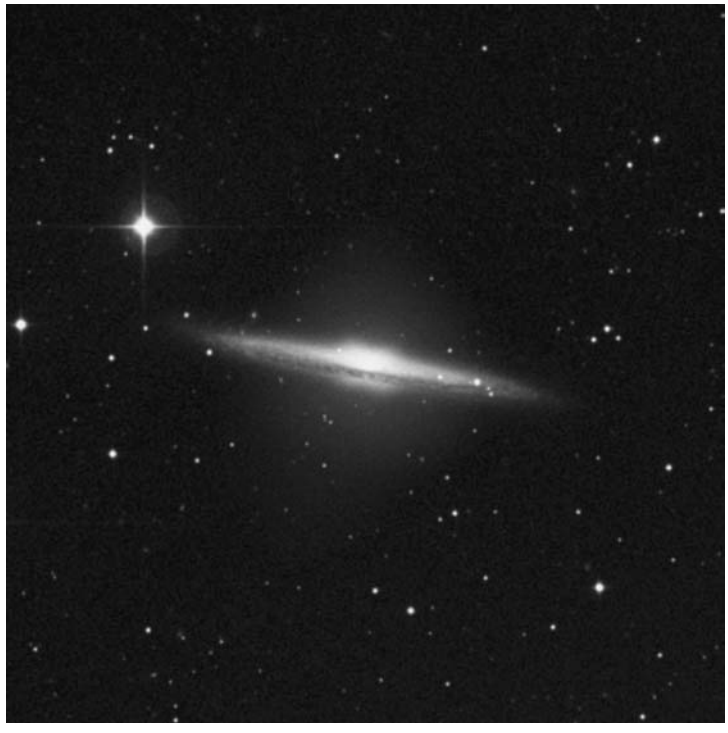


Figure 8.7. The galaxy, NGC 5746, is a typical spiral in the optical, displaying a flat disk cut by a dust lane. Surrounding this galaxy, however, is a halo of X-ray emission, shown in blue. The X-ray emission is due to thermal Bremsstrahlung from hot gas at $T_e = 6.5 \times 10^6$ K that extends to about 20 kpc from the disk. The X-ray luminosity is $L_X = 4.4 \times 10^{39}$ erg s^{-1} in the 0.3 – 2 keV waveband (Ref. [119]). (Reproduced by permission. X-ray: NASA/CXC/U. Copenhagen/K.Pedersen et al; Optical: Palomar DSS) (See colour plate)

contained in the intracluster gas is 18 times greater than the total mass contained in all the stars of all the galaxies in the cluster (Ref. [141])! Most of this gas is likely a remnant of the cluster formation process, although a small fraction may have come from supernova outflows from the galaxies. The latter contribution accounts for a small (less than Solar) heavy metal enrichment (Sect. 3.3.2). The total mass of the cluster is even greater than that of the gas, however, consisting of both light and dark matter components. The total (light + dark) mass of the cluster may be found via gravitational lensing (Sect. 7.3.1), via an analysis of the motions of the individual galaxies (Sect. 7.2.1.2) or by an assumption of *hydrostatic equilibrium* for the gas. The latter means that the gas, as a whole, is gravitationally held in place and not 'evaporating' away. With this assumption, it is possible to calculate how much mass is required to hold gas of a given temperature in place. The results of such analyses indicate that the gas mass to total mass fraction in clusters of galaxies is $\approx$15–20 per cent (e.g. Prob. 8.8). This is in agreement with the baryon fraction expected from Big Bang nucleosynthesis (Sect. 3.2), suggesting that there may be

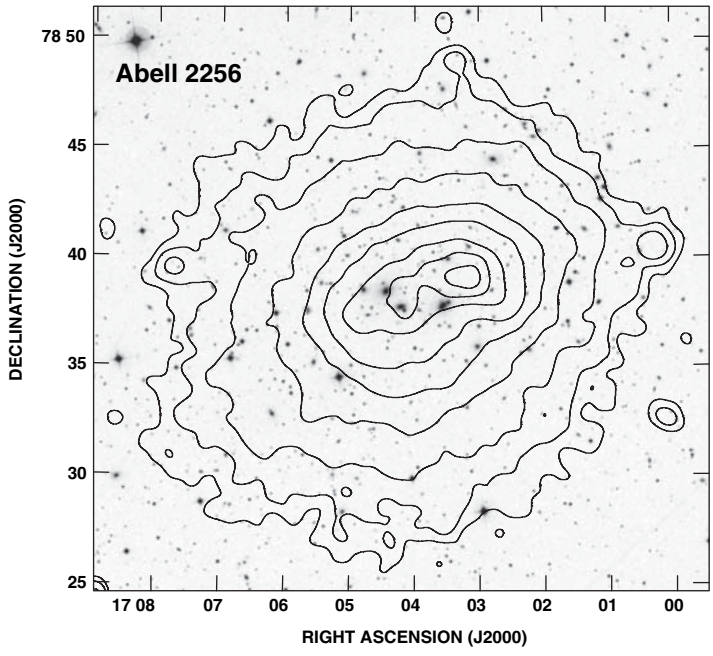


Figure 8.8. The cluster of galaxies, Abell 2256 (redshift, $z = 0.0581$), is shown in the optical (negative grescale) and in X-rays in the observing band, 0.1–2.4 keV (contours). The gas is emitting at $k\,T_e = 7.5$ keV and the X-ray luminosity is $L_{0.1\rightarrow2.4} = 3.6 \times 10^{44}$ erg s^{-1} (Ref. [56]), of which 78 per cent is due to thermal Bremsstrahlung (Ref. [22]). The cluster radius is $R = 2.0$ Mpc (Ref. [141]). The total mass of the cluster, including all light and dark matter, is $M = 1.2 \times 10^{15}\ M_\odot$ (Ref. [130]). All values have been adjusted to $H_0 = 71$ km s^{-1} Mpc^{-1}. The optical image is in R band (Reproduced courtesy of the Palomar Observatory-STS_c I Digital Sky Survey, Cal Tech). X-ray data were obtained from the Position Sensitive Proportional Counter (PSPC) of the Röntgen Satellite (ROSAT)

no 'missing baryons' in rich clusters of galaxies that have X-ray emitting gas. The problem of missing non-baryonic matter is, however, still present.

8.3 Free–bound (recombination) emission

The free–bound process, or *recombination*, has already been discussed in the context of the ionization equilibrium that is present in HII regions (Sect. 3.4.7 and Example 5.4) in which the ionization rate must balance the recombination rate. We now wish to consider recombination as an emission mechanism.

Recombination involves the capture of a free electron by a nucleus into a quantized bound state of the atom. Therefore, free–bound radiation must only occur in ionized gases. This now introduces the possibility that *both* free–bound and free–free radiation may occur from the same ionized gas, complicating the analysis of such regions. The free–bound process is essentially the same as described for free–free emission except

for the end state, and in fact the free–bound emission coefficient can be written in the same fashion as Eq. (8.2) (Ref. [96]),

$$\mathrm{j}_\nu = 5.44 \times 10^{-39} \left(\frac{Z^2}{\mathrm{T_e}^{1/2}} \right) n_\mathrm{i}\, n_\mathrm{e}\, \mathrm{g_{fb}}(\nu, T_\mathrm{e})\, e^{-(\frac{h\nu}{k\mathrm{T_e}})} \tag{8.23}$$

Note that this emission coefficient is identical to that of free–free emission except for the use of the *free–bound Gaunt factor*, $\mathrm{g_{fb}}$, so an understanding of the relative importance of free–bound emission compared to free–free emission reduces to an understanding of the relative importance of the two Gaunt factors.

Computation of the free–bound Gaunt factor, $\mathrm{g_{fb}}(\nu, \mathrm{T_e})$, requires sums over the various possible final states in atoms and taking into account a Maxwellian distribution of velocities for the intitial states. In the limit, if the final state has a very high quantum number that approaches the continuum, then $\mathrm{g_{fb}} = \mathrm{g_{ff}}$. Gaunt factors have been determined for a range of temperatures, frequencies, and metallicities by various authors (e.g. Refs. [28], [107]). In Figure 8.2 we show one result, the *sum* of $\mathrm{g_{fb}} + \mathrm{g_{ff}}$, for the temperature, $\mathrm{T_e} = 1.58 \times 10^5$ K (grey curve). The free–bound Gaunt factor is therefore represented by the excess over the smooth $\mathrm{g_{ff}}$ curve below it. Note that there are sharp *edges*, corresponding to transitions into specific quantum levels of energy, E_n (see Eq. C.5), followed by smoother distributions to the right (higher frequencies) of the edges. This behaviour can be understood by considering conservation of energy as electrons are captured, the energy difference going into the emitted photons (Prob. 8.9).

For low frequencies (i.e. $h\nu \ll k\,\mathrm{T_e}$), Figure 8.2 shows that the total Gaunt factor is just the free–free value, implying that recombination radiation is negligible in comparison to thermal Bremsstrahlung in such a gas. This is typical of other temperatures as well, validating our assumption (Sect. 8.2.2) that the radio continuum emission from HII regions and planetary nebulae is due entirely to thermal Bremsstrahlung. In this regime, the electron is perturbed only slightly from its path. However, once $h\nu \geqslant k\,\mathrm{T_e}$, $\mathrm{g_{fb}}$ may become significant. For the temperature, 1.58×10^5 K, plotted on the graph, this corresponds to $\nu \geqslant 3 \times 10^{15}$ Hz which is in the UV part of the spectrum. For HII regions and planetary nebulae with temperatures closer to $\sim 10^4$ K (Table 3.1), the corresponding frequency regime is $\nu \geqslant 2 \times 10^{14}$ Hz which is in the near-IR and optical part of the spectrum. Thus, depending on temperature and frequency band, the recombination continuum can be a significant or even dominant component to the total continuum. As the temperature increases, however, the relative contribution of $\mathrm{g_{fb}}$ decreases. The free–free Gaunt factor has a weak dependence on $\mathrm{T_e}$ (see Eq. 8.11 or Figure 8.2), increasing as $\mathrm{T_e}$ increases. However, $\mathrm{g_{fb}}$ has a stronger dependence on temperature and also *decreases* as $\mathrm{T_e}$ increases. At higher temperatures, the electrons will have higher speeds, on average, and are less likely to be captured. Therefore at very high temperatures, free–free emission dominates the continuum. A final important note is that a proper calculation of Gaunt factors requires all elements to be taken into account. For example, at temperatures below 10^6 K, recombinations to C V, C VI, O VII, O VIII and N VII are important at high frequencies.

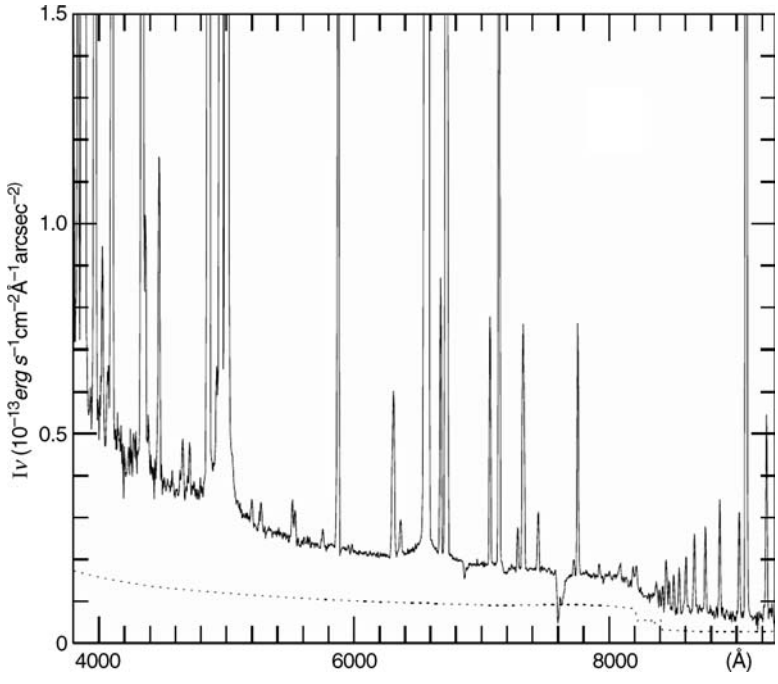


Figure 8.9. A spectrum of an HII region in the optical waveband. Many strong emission lines can be seen superimposed on the underlying continuum. The dashed curve shows the *combined* continuum from both free–bound radiation and two-photon emission. The remainder of the continuum is due to scattered light from dust. The spectrum is of a region in the 30 Doradus nebula, also called the *Tarantula Nebula*, which is in the *Large Magellanic Cloud*, a companion galaxy to the Milky Way. (Reproduced by permission of Darbon, S., Perrin, J.-M., and Sivan, J.-P., 1998, A&A, **333**, 264)

A real spectrum of a combined free–free and free–bound continuum is shown for a Solar flare at X-ray wavelengths in Figure 6.7. Several spectral lines are also visible in that spectrum. Figure 8.9 shows another example, this time for an HII region at optical wavelengths. In this spectrum, the optical continuum consists of free–bound radiation, two-photon emission (see next section) and scattered light from dust. In addition, a great many optical emission lines are present so a measurement of the continuum must be made at frequencies between the lines. These figures demonstrate that it is sometimes difficult to isolate emission from a single process alone. They are examples of *complex spectra* for which *modelling* is required to understand the various contributions. Complex spectra will be discussed further in Sect. 10.1.

8.4 Two-photon emission

Another important contribution to the continuum in low density ionized regions at some frequencies is a process called *two-photon emission*. Two-photon emission

occurs between *bound* states in an atom but it produces *continuum* emission instead of line emission. It occurs when an electron finds itself in a quantum level for which *any* downwards transition would violate quantum mechanical selection rules (Appendix C) and the transition is therefore highly *forbidden* (e.g. Appendix C.4). The electron cannot remain indefinitely in an excited state, however, and there is some probability, although low, that de-excitation will occur. The most likely de-excitation may not be the usual emission of a single photon, but rather the emission of *two* photons. These two photons can take on a range of energies whose sum is the total energy difference between the levels. Two-photon emission will occur as long as the electron is not removed from the level via a collisional energy exchange first. Therefore, it occurs in lower density ionized nebulae in which the time between collisions that would depopulate the problematic level is longer than the two-photon lifetime.

The hydrogen atom provides a good example of two-photon emission. We have already seen that the Ly α line (a recombination line, Sect. 3.4.7), which results from a de-excitation of an electron from the n = 2 to n = 1 levels, has a very high probability (Table C.1). However, the n = 2 quantum level consists of both $2s$ and $2p$ states (Table C.2). The transition $2p \rightarrow 1s$ is a permitted transition with $A_{2p \rightarrow 1s} = 6.27 \times 10^8\,s^{-1}$, but the $2s \rightarrow 1s$ transition is highly forbidden. The forbidden transition will occur via two-photon emission with a rate coefficient of $A_{2s \rightarrow 1s} = 8.2\,s^{-1}$, considerably lower than the $2p \rightarrow 2s$ transition, but still high enough for this process to be important in nebulae of densities $< 10^4\ cm^{-3}$ (Ref. [155]).

Since the sum of the energies of the two photons must equal the energy difference between the two levels, then so must the frequencies ($E = h\nu$),

$$\nu_1 + \nu_2 = \nu_{\mathrm{Ly}\alpha} \tag{8.24}$$

where ν_1 and ν_2 are the frequencies of the two photons and $\nu_{\mathrm{Ly}\alpha}$ is the frequency of Ly α. The spectrum of photon frequencies can take on any value as long as Eq. (8.24) is satisfied, but the most probable configuration is that the two photons have the *same* frequency. Therefore, the probability distribution of photon frequencies is symmetric and centred on the midpoint frequency which, for Ly α, is $\nu_{0,2\mathrm{ph}} = 1.23 \times 10^{15}$ Hz ($\lambda 243$ nm). This is in the ultraviolet part of the spectrum with a width that stretches to $\nu = 0$ and $\nu = \nu_{\mathrm{Ly}\alpha}$, the probability being lower as the photon frequency departs from $\nu_{0,2\mathrm{ph}}$. Note that since the *number* of photons with frequencies above $\nu_{0,2\mathrm{ph}}$ is equal to the number below, the *energy distribution* (which describes the spectrum) is actually skewed to higher energies.

The strength of two-photon emission depends on the number of particles in the n = 2 state of hydrogen. This, in turn, depends on the recombination rate to this level which is a function of both density and temperature. Therefore, two-photon emission, although quantum-mechanical in nature, can be thought of as thermal emission and the density dependence is the same as for free–free and free–bound emission, i.e. $j_\nu \propto {n_e}^2$. The emission coefficient decreases with increasing temperature, described by a somewhat more complex function. Further details can be found in Ref. [96].

As noted in Sect. 8.3, two-photon emission is present in the optical continuum of the HII region shown in Figure 8.9. It tends to be a smaller contributor to the spectrum than free–free or free–bound emission except at frequencies near its peak. For example, at $T_e = 10^4$ K, two-photon emission is about 30 per cent higher than the sum of free–bound and free–free emission at $\nu_{0,2ph} = 1.23 \times 10^{15}$ Hz. However, at $\nu = 3.0 \times 10^{14}$ Hz it is about six times weaker and, once the frequency departs significantly from $\nu_{0,2ph}$, two-photon emission is negligible (Ref. [27]). However, again, all species must be considered when computing the continuum from an ionized gas. Two-photon emission is also possible from He as well as heavier species, for example. In hotter gas, two-photon emission from highly ionized species such as N VI, Ne X, Mg XI, Si XIV, SX VI, Fe XXV, and others cannot be ignored if these elements are present. For gas between 10^6 and 10^7 K, two-photon emission can again dominate within certain parts of the observing band between 0.1 and 1 keV (Ref. [131]).

8.5 Synchrotron (and cyclotron) radiation

Synchrotron radiation results from *relativistic electrons* (those with speeds approaching the speed of light) moving in a *magnetic field*. The source of relativistic electrons is the electron component of the cosmic rays that were discussed in Sect. 3.6. Synchrotron radiation is widespread in the Milky Way and other galaxies and is one of the most readily observed continuum emission processes in astrophysics. There is some similarity between the Bremsstrahlung radiation discussed in Sect. 8.2 and synchrotron radiation. In the former case, the electron is accelerated in an electric field, $\vec{E}$, and in the latter case, the electron is accelerated in a magnetic field, $\vec{B}$. Both represent 'scattering' of a charge in a field. Therefore, synchrotron emission is sometimes referred to as *magnetic Bremsstrahlung* or *magneto-Bremsstrahlung* radiation.

To understand the emission, it is helpful to recall that a charged particle moving at a velocity, $\vec{v}$, in a magnetic field, $\vec{B}$, experiences a Lorentz force which, for an electron of charge, e, and negligible electric field is $\vec{F}_e = e\left(\frac{\vec{v}}{c} \times \vec{B}\right)$ (see Table I.1). Such a force exists whether or not the electron is relativistic. The direction of the force is perpendicular to *both* $\vec{v}$ and $\vec{B}$ and the magnitude of the force is,

$$F_e = \frac{e\,v}{c} B \sin\phi = \frac{e\,v}{c} B_\perp \tag{8.25}$$

where ϕ is the angle between $\vec{v}$ and $\vec{B}$, called the *pitch angle*, and $B_\perp = B \sin\phi$. Figure 8.10 illustrates this geometry. If $\vec{v}$ is perpendicular to $\vec{B}$ ($\sin\phi = 1$), the electron experiences the maximum force and will circle about the field line. If the electron moves only parallel to $\vec{B}$ ($\sin\phi = 0$), then it experiences no force and will not radiate. If the initial velocity is in an arbitrary direction, then the electron will spiral about the field line. Over a period of gyration (the time for a loop), the speed of the

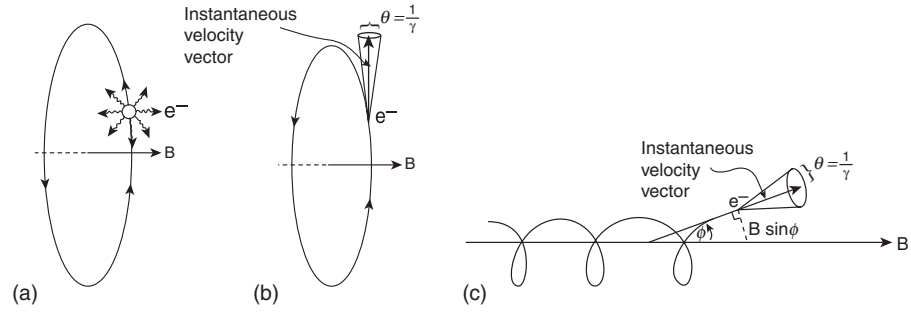


Figure 8.10. Illustration of the radiation given off by an electron with instantaneous velocity, $\vec{v}$, encircling magnetic field lines of strength, B. **(a)** Non-relativistic electron whose velocity is perpendicular to the direction of the magnetic field. The force acting on the electron is perpendicular to both $\vec{v}$ and $\vec{B}$, causing the electron to circle about the field lines. The resulting *cyclotron radiation* is emitted in all directions. **(b)** Relativistic electron whose motion is perpendicular to the magnetic field. The force on the electron is in the *same* direction as in (a) but the radiation given off (*synchrotron radiation*) is highly beamed in the forward direction into a cone of angular radius, $\theta = 1/\gamma$, where γ is the Lorentz factor (see Eq. 3.37). **(c)** More realistic situation in which an electron has components of its velocity both parallel and perpendicular to the direction of B. The *pitch angle*, ϕ, is the angle between $\vec{v}$ and $\vec{B}$.

electron does not change due to this force[11], but its change of direction constitutes an acceleration and it will therefore radiate. Note that *any* charged particle, including protons and other ions, experience this behaviour although, as we will see, the synchrotron emission from ions is negligible in comparison to that of electrons.

If the magnetic field within a plasma is strong enough that it dynamically affects charged particles, it is called a *magnetized plasma*. A more specific criterion is if a charged particle completes at least one gyration before it interacts with (collides with) another particle, a condition that may be different between electrons and ions. In a magnetized plasma, the gyrating charged particle is 'coupled' to the field and therefore, in a sense, trapped by it. If the field lines curve, for example, the gyrating particle will follow along this curvature. Any emission that may be given off by such trapped particles can then be used to *map out* the magnetic field configuration. A good example is the prominences seen on the Sun, as shown in Figure 6.6 or the coronal loop of Figure 8.11. These particular images show recombination line emission (rather than synchrotron) associated with the ionized gas (Sect. 3.4.7). Such emission traces out the magnetic fields that loop out of the Sun, forming the delicate filamentary structures shown. In this way, trapped charges result in the 'illumination' of the otherwise invisible magnetic fields. Similarly, synchrotron emission allows us to map out other cosmic magnetic fields in locations where the electrons are relativistic.

[11]Eventually the loss of energy due to radiation will slow the particle, but this occurs over a timescale that is very long in comparison to a gyration time.

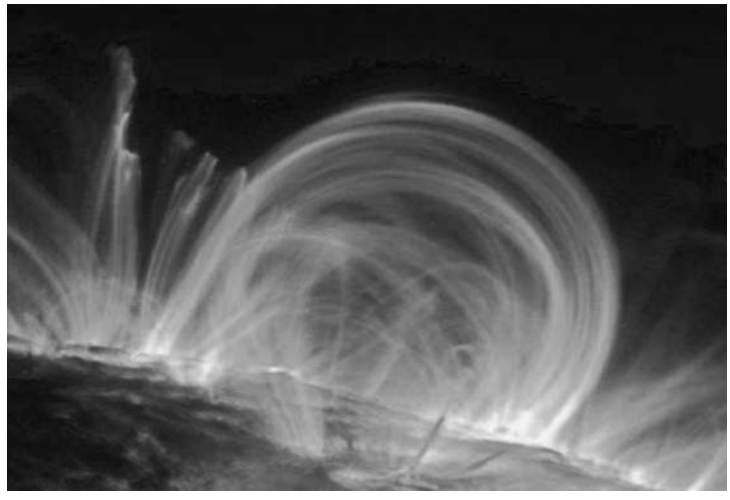


Figure 8.11. A *coronal loop* can be seen at the edge of the Sun. This image, taken with the Transient Region and Coronal Explorer (TRACE) satellite, shows emission at 171 Å from Fe IX and Fe X ions that trace out delicate filamentary and loop-like magnetic field lines in the Solar atmosphere. This emission is characteristic of plasmas at $T_e = 6 \times 10^5$ K in the upper transition region of the atmosphere (see Figure 6.5). (Reproduced by permission. TRACE is a mission of the Stanford-Lockheed Institute for Space Research, part of the NASA Small Explorer program)

Radiation that results specifically from motion about the magnetic field will depend on the strength of the field, B. Therefore, not only do observations of such emission provide a means of determining the presence and orientation of fields, but they also provide a means of obtaining the magnetic field strength (e.g. Table 8.1). We will see how this dependence occurs in the next sections. Before considering the details of

Table 8.1 Sample magnetic field strengths[a]

Object	B (G)
Interstellar space	10^{-6}
Interplanetary space	$10^{-6} - 10^{-5}$
Solar corona	$10^{-5} - 100$
Planetary nebulae	$10^{-4} - 10^{-3}$
H II region	10^{-6}
Pulsar (surface)	10^{12}
Supernova remnants (SNRs)	$10^{-5} - 10^{-2}$
Earth	0.31
Jupiter	4.28
Saturn	0.22
Uranus	0.23
Neptune	0.14

[a]Ref. [96] except for the planets and SNRs. (Ref. [12]) Planetary fields are surface equatorial values (but note that the value is variable with position and time).

synchrotron radiation, however, it is helpful to start with its non-relativistic, and less frequently observed, counterpart – *cyclotron radiation*.

8.5.1 Cyclotron radiation – planets to pulsars

When a non-relativistic electron gyrates about the magnetic field, cyclotron radiation is emitted at the frequency of this gyration, called the *gyrofrequency*, ν_0. For circular motion, the acceleration always points to the centre of the circle and the electric field vector of the emitted radiation follows that of the acceleration vector as seen by a distant observer (see Ref. [101], pp. 62–66 for details on the emission of an accelerating charge). If seen from a direction *along* the field line, $\vec{\mathrm{E}}$ will rotate with time, giving rise to circular polarization of the emission. Seen from the side, the electric field vector will oscillate linearly (like a dipole), giving rise to linear polarization. From an intermediate angle, the polarization is elliptical. A characteristic of this kind of emission, therefore, is that it is intrinsically *polarized* and high degrees of polarization (D_p, see Sect. 1.7) have been measured in sources in which cyclotron radiation has been observed. Examples are the planets that have magnetic fields – the Earth, Jupiter, Saturn, Uranus, and Neptune – for which D_p up to 100 per cent has been measured (Ref. [187]).

For spiral motion, the electron has both a velocity component that is parallel to the magnetic field and therefore unaffected by it, $v_\parallel$, as well as a perpendular component, $v_\perp = v \sin\phi$. The Lorentz force (Eq. 8.25) can then be equated to the centripetal force,

$$\mathrm{F_e} = \frac{e\,v}{c}\,B_\perp = \mathrm{m_e}\,\frac{{v_\perp}^2}{r_0} \tag{8.26}$$

where r_0 is the orbital radius that is perpendicular to the field, called the *radius of gyration* or *gyroradius*. For protons or other charged particles, the appropriate mass and charge must be used (Prob 8.10). Eq. (8.26) can be solved for r_0,

$$r_0 = \frac{\mathrm{m_e}\,v_\perp\,c}{e\,B} = \frac{v_\perp\,\mathcal{T}}{2\,\pi} = \frac{v_\perp}{2\,\pi\,\nu_0} \tag{8.27}$$

where $\mathcal{T} = 1/\nu_0$ is the period of gyration and we have used the relation between velocity, distance and time ($v_\perp = 2\,\pi\,r_0\,/\mathcal{T}$). The gyrofrequency then follows from Eq. (8.27),

$$\nu_0 = \frac{e\,B}{2\,\pi\,\mathrm{m_e}\,c} \quad => \quad \left[\frac{\nu_0}{\mathrm{MHz}}\right] = 2.8\left[\frac{B}{\mathrm{Gauss}}\right] \tag{8.28}$$

Note that the electron gyrofrequency is independent of velocity and therefore independent of the kinetic energy of the electron ($E_k = \frac{1}{2}\,\mathrm{m_e}\,v^2$) when $v \ll c$. Even if there is a range of particle velocities, as is normally the case, all particles will emit at the same frequency for the same magnetic field strength, though the radius of gyration will be different. Since the gyrofrequency is single-valued, it really represents

monochromatic, or line radiation (see Chapter 9), rather than continuum radiation, with a width that can depend on various line widening mechanisms (e.g. Sect. 9.3). However, the physical process is fundamentally different from that of 'normal' spectral lines which result from quantum transitions within atoms and molecules. Example 8.2 provides a numerical example.

Example 8.2

Determine the radius of gyration, period and gyrofrequency of a typical electron within the warm ionized medium (WIM) of the Milky Way, in which the magnetic field strength is $B \approx 3\,\mu G$. *Can we expect to observe the resulting cyclotron radiation?*

From Sect. 8.2.2, the WIM consists of ionized gas with $T_e \approx 10^4$ K and $n_e \approx 0.01\,cm^{-3}$. From Eq. (8.28) with the given magnetic field strength, we find, $\nu_0 = 8.4\,Hz$ so $\mathcal{T} = 1/\nu_0 = 0.12\,s$. A 'typical' electron might be expected to have the mean electron speed in a Maxwellian velocity distribution which, for a gas at 10^4 K, is $v = 1.1 \times 10^8\,cm\,s^{-1}$ (see Eq. 3.9). Therefore, letting $v \approx v_\perp$ and using Eq. (8.27), the radius of gyration is $r_0 = 2.1 \times 10^6\,cm$ or 21 km. A gyrofrequency of 8.4 Hz is extremely low and the emission could not be observed from the ground because it would not propagate through the Earth's ionosphere (see Figure 2.2). Moreover, the plasma frequency in the WIM is (see Eq. E.9) $\nu_p = 8.9\,n_e^{1/2} = 0.89\,kHz$, more than $100\times$ higher than the cyclotron frequency, so this radiation would not even escape from the WIM. Therefore, this emission cannot be observed.

The importance of Eq. (8.28) is that the magnetic field strength can be immediately obtained if the cyclotron frequency can be measured. The problem, however, is that this frequency is very low, as demonstrated by Example 8.2. The cyclotron frequency must be higher than the plasma frequency for the radiation to escape, favouring objects with high magnetic fields (see Table 8.1) and/or low electron densities (Prob. 8.10). Also, the magnetic field must be higher than about 3.5 G in order for this emission to be detectable from the ground because of the Earth's $\approx 10\,MHz$ ionospheric cutoff (Figure 2.2). Thus, the number of objects for which cyclotron emission has so far been observed is limited to those with strong fields, or those for which satellite or space probe data are available, or both. Examples are pulsars, the Sun, and the planets that have magnetic fields (see Table 8.1).

For low velocity electrons ($v \lesssim 0.03\,c$) only a single gyrofrequency at ν_0 is observed, as we have seen (Ref. [14]). In a thermal distribution of particles, this corresponds to $T_e \lesssim 10^6$ K. In hotter gases or in non-thermal distributions with higher velocities, however, not only is the *fundamental frequency*, ν_0, observed, but also weaker *harmonics* at integer multiples of the fundamental[12]. The observation of multiple 'spectral lines' at the fundamental frequency and its harmonics helps to confirm that

[12]See Footnote 3 in this chapter.

the emission is actually due to the cyclotron process. Such lines have been observed in emission in some regions and absorption in others, depending on the presence and strength of the background radiation field. Example 8.3 provides an example.

Example 8.3

Three absorption lines have been observed in the X-ray spectrum of a pulsar, named 1E1207.4-5209, at energies of 0.7, 1.4 *and* 2.1 *keV (Ref. [*21*]). If the gravitational redshift from this neutron star is* $z_G = 0.2$*, what is the strength of its magnetic field?*

These three lines are equally spaced in frequency and are presumed to represent the cyclotron fundamental frequency and its first and second harmonics. The fundamental gyrofrequency is then (via $E = h\nu$) $\nu_0 = 1.69 \times 10^{17}$ Hz. For a gravitational redshift of $z_G = 0.2$, the true frequency (Eq. 7.1 with $\lambda\nu = c$) is $\nu_0 = \nu(1 + z_G) = 2.03 \times 10^{17}$ Hz. Then using Eq. (8.28), $B = 7.1 \times 10^{10} G$.

Other cyclotron spectra may not be quite as straightforward to interpret as is implied by the neutron star of Example 8.3. For example, the magnetic field strength may vary significantly within the region being observed, resulting in a number of fundamental cyclotron frequencies, each corresponding to a different field strength. For a smoothly varying field, the result will be emission that is spread out over a corresponding frequency range, giving *continuum*, rather than line emission. This can be seen in Figure 8.12 which shows the low frequency radio spectra of the planets.

Low frequency planetary radio emission comes from a region around the planets called the *magnetosphere* which is the region within which charged particles are more strongly affected by the magnetic field of the planet than by other external fields. A diagram of Jupiter's magnetosphere is shown in Figure 8.13. This topology, that of a *magnetic dipole* swept by the Solar wind (Sect. 3.6.3), is similar for the other planets that have magnetic fields. Jupiter's field is both larger and stronger than those of the other planets and it also has an additional component related to its Galilean[13] satellite, Io, whose volcanic outbursts supply additional plasma into a torus, called the *Io plasma torus*, around the planet.

The magnetic field strength varies with position within the magnetosphere. The strongest magnetic fields are near the poles of the planet, as indicated by a higher density of field lines in Figure 8.13. From Figure 8.12, Jupiter's maximum gyrofrequency is $\nu_0 \approx 40$ MHz which, by Eq. (8.28), gives a maximum magnetic field strength of $B \approx 14$ G, corresponding to a location approximately where the magnetic field lines that pass through the Io plasma torus intersect the planet (Ref. [187]). However, its equatorial surface field is $B \approx 4$ G (Table 8.1). The radio emission also varies with time as the planet rotates, as the various satellites orbit through the magnetosphere, as new material is injected into the magnetosphere, and as the Solar

[13]The four largest moons of Jupiter, Io, Europa, Ganymede, and Callisto, are called the *Galilean satellites*, after their discoverer.

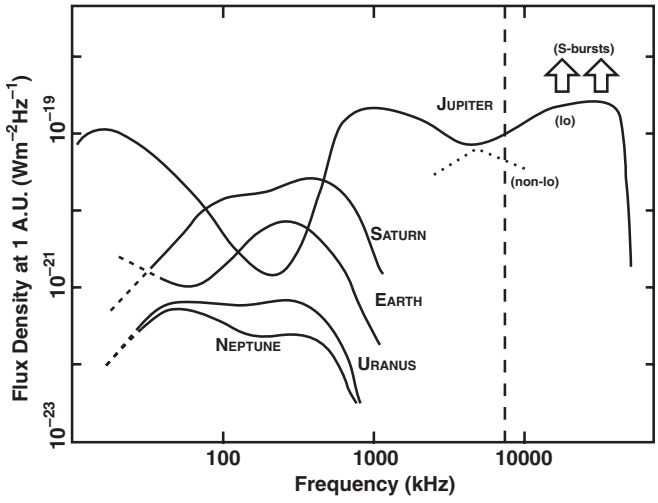


Figure 8.12. Low frequency radio spectra of the planets that have magnetic fields. The vertical dashed line marks approximately the Earth's ionospheric cutoff. Jupiter has the strongest emission, having a higher magnetic field (Table 8.1) and a larger magnetospheric cross-section for interaction with Solar wind particles. The interaction of Jupiter's magnetosphere with its inner Galilean moon, Io, is responsible for higher frequency emission from this planet as well as the 'S-bursts' which refer to short duration radio bursts. During radio bursts, peak flux densities can be $> 10 \times$ higher than shown here and, for Jupiter, $> 100 \times$ higher. The flux density scale assumes that the radio emission has been measured at a distance of 1 AU from each planet (Adapted from Ref. [187]).

wind varies. The result is strong intermittent *radio bursts*. Depending on the type of stellar system, similar radio bursts might be detectable from planets orbiting other stars as well (Prob. 8.11), providing a possible additional probe of extrasolar planets.

An accurate description of the radio spectra shown in Figure 8.12 involves complex modelling that takes a variety of effects into account. For example, for particles that are more energetic or mildly relativistic, the cyclotron frequencies and line widths are functions of electron energy. Therefore, the shape of the spectrum depends on the electron energy/velocity distribution which tends *not* to be Maxwellian (i.e. it is not thermal). At the low frequencies shown for the planets, the electron distribution is actually *inverted*. Rather than the *declining* power law distribution that we saw for cosmic ray nucleons (Figure 3.25) and electrons (Eq. 3.43), there are instead more high energy electrons than low energy ones. This is because electrons with high pitch angles (see Figure 8.10c), which tend to be those of higher energy, are more easily reflected back along the field lines when they enter the 'bottleneck' region of high magnetic field strength and high plasma densities near planetary poles. Electrons with smaller pitch angles are more likely to suffer collisions in these regions and are not reflected. Therefore, the low energy particles are selectively removed. Other effects, such as resonances between the electron gyrofrequency and ambient radiation frequencies from other emitting electrons are also important and can result in very intense, low frequency

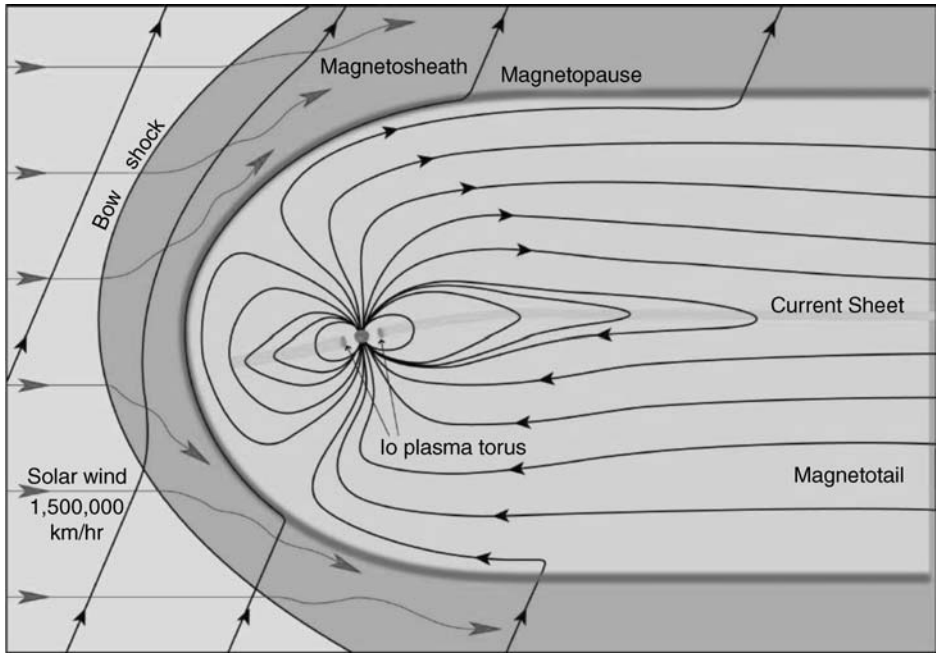


Figure 8.13. Diagram of Jupiter's *magnetosphere*. The planet, itself, is the small dot near the centre. The curves with arrows show the field lines which come closer together (stronger fields) near the poles of the planet. Note the Io plasma torus as well as the asymmetric shape of the field due to its interaction with the Solar wind. (Adapted from http://pluto.jhuapl.edu/science/jupiterScience/magnetosphere.html)

radio emission[14]. Scattering of electrons from waves in the magnetic field lines (called *Alfvén waves*), plasma waves, and other processes can also affect the shape of the spectrum. Synchrotron radiation is also produced in the magnetosphere of Jupiter.

Those particles that do leak through to the upper atmosphere in the regions of the magnetic poles can collide with and ionize atmospheric atoms and molecules. When the electrons recombine, the subsequent downwards-cascading bound–bound transitions (i.e. fluorescence, see Sect. 5.1.2.1) can produce visible light. For the Earth, the result is the *aurora borealis* (northern lights) in the northern hemisphere and the *aurora australis* (southern lights) in the southern hemisphere. These are most often seen after a Solar flare (Sect. 6.4.2), demonstrating the sensitive interaction between the Earth's magnetosphere and the Solar wind.

8.5.2 The synchrotron spectrum

There are a number of important ways in which the synchrotron spectrum differs from that of cyclotron emission. As we did for thermal Bremsstrahlung radiation (Sect. 8.2),

[14]These effects produce what is called the *electron cyclotron maser.*

we will describe 'semi-qualitatively' the factors that are involved in deriving the spectrum. More detailed descriptions can be found in Refs. [138], [143], [14], and others.

Firstly, the mass of a relativistic particle is greater than its rest mass by a factor, γ, where γ is the Lorentz factor given by Eq. (3.37). The energy of a relativistic electron is therefore, $E = \gamma\, m_e\, c^2$ Eq.(3.36). The *relativistic gyroradius* and *relativistic gyrofrequency* must take this into account, so using Eqs. (8.27) and (3.36) and setting $v_\perp \approx c$ (note, all in cgs units),

$$r = \gamma\, r_0 = \frac{\gamma\, m_e\, c^2}{e\, B} = \frac{E}{e\, B} = 2.2 \times 10^9\, \frac{E}{B} \tag{8.29}$$

and using Eqs (8.29) with (3.36),

$$\nu = \frac{\nu_0}{\gamma} = \frac{e\, B}{2\,\pi\,\gamma\, m_e\, c} = 2.3\, \frac{B}{E} \tag{8.30}$$

and the period, $\mathcal{T} = 1/\nu$. Eqs. (8.29) and (8.30) reduce to Eqs. (8.27) and (8.28), respectively when the electron is non-relativistic (i.e. $\gamma \approx 1$). Note that these equations now *do* depend on electron energy, unlike their cyclotron counterparts. Since Lorentz factors can be extremely high (Sect. 3.6), the above two equations indicate that the relativistic gyroradius is much higher than the cyclotron radius and the relativistic gyrofrequency is much lower than the cyclotron gyrofrequency. At first glance, a very low value of ν might suggest that synchrotron-emitting frequencies would be too low to measure. However, this is not the case, as we will see below. Although Lorentz factors up to $\gamma = 10^{11}$ have been measured for nucleons, there are very few particles with such high values (see Figure 3.25) and a 'typical' electron energy is lower (Example 8.4).

Example 8.4

Determine the gyroradius, the gyrofrequency, and the period of an electron with $\gamma = 10^4$ in the ISM. What is the speed of the electron compared to the speed of light?

For $\gamma = 10^4$, the electron energy is (Eq. (3.36)) $E = \gamma\, m_e\, c^2 = 8.18 \times 10^{-3}$ erg (5.1 GeV). For a typical ISM magnetic field of 3μG, we find (Eq. 8.29) $r = 6.0 \times 10^{12}$ cm, or approximately 86 × the radius of the Sun. The gyrofrequency (Eq.(8.30)) is $\nu = 8.4 \times 10^{-4}$ Hz, so its period is $\mathcal{T} = 20$ min. By Eq. (3.37), $v/c = 0.999999995$.

The second important difference with cyclotron radiation is that, because of the relativistic motion of the electron, the *transformations* that need to be made between the rest frame of the electron (the frame moving with $v_\parallel$ of the electron) and the rest frame of the distant observer (the 'lab frame') are significant. These transformations[15] have a number of consequences.

[15]The required transformations, called *Lorentz transformations* for velocity and time, are extended to also include a transformation for acceleration.

For example, in the rest frame of the electron, the power emitted is given by the Larmor formula (Table I.1) which involves the electron's acceleration. A transformation of this acceleration to the lab frame, however, results in a total power that is boosted by a factor, γ^2. Moreover, the distribution of this power with angle also changes. In the electron's rest frame, the emission is over a very wide angle (see Figure 8.10a). However, a distant observer will see the emission beamed into a narrow cone (Figure 8.10b). The opening angle of the cone depends on the particle's energy such that higher energy particles (higher γ) have narrower cones, the angular radius of the cone (see Figure 8.10c) being,

$$\theta = \frac{1}{\gamma} \tag{8.31}$$

The direction of the beam is the direction of the instantaneous velocity vector of the particle, so as the particle spirals about the field, the cone direction rotates with it and a distant observer will only see emission when (and if) the cone sweeps by his line of sight. The opening angle is very small for a highly relativistic electron (Example 8.5) so the time over which this cone sweeps by the observer is small in comparison to the larger gyration period, resulting in brief, repeated *pulses* of emission.

To obtain the *synchrotron spectrum* of a relativistic gyrating electron requires a conversion from the time to the frequency domain as well as the final important departure from the cyclotron case – the need for a Doppler shift (Sect. 7.2.1) since the observer is only seeing emission when the particle is moving towards him. As before, the time/frequency conversion requires a more detailed analysis[16]. As we saw for thermal Bremsstrahlung radiation (Sect. 8.2.1), however, the shortest time period (the pulse duration) sets the maximum frequency as seen by the observer. This maximum, called the *critical frequency*, above which there is negligible emission, is given by,

$$\nu_{\rm crit} = \frac{3}{2}\,\gamma^2\,\nu_0\,\sin\,\phi = \frac{3\,e}{4\,\pi\,{\rm m_e}\,c}\,\gamma^2\,B_\perp \quad => \quad \left[\frac{\nu_{\rm crit}}{\rm MHz}\right] = 4.2\,\gamma^2\left[\frac{B_\perp}{\rm Gauss}\right] \tag{8.32}$$

where we have used Eq. (8.28). Thus, the maximum frequency is, in fact, very high (see Example 8.5) and much higher than the relativistic gyrofrequency. Most of the energy, however, is actually emitted at a frequency somewhat lower than the critical frequency, i.e. at $\nu_{\rm max} = 0.29\,\nu_{\rm crit}$. The longest time period, which is related to the gyration period, determines the fundamental frequency, ν_f. The spectrum will contain ν_f and all its harmonics, where,

$$\nu_f = \frac{1}{\gamma}\,\frac{\nu_0}{\sin^2\,\phi} \quad => \quad \left[\frac{\nu_f}{\rm MHz}\right] = \frac{2.8}{\gamma\,\sin^2\,\phi}\left[\frac{B}{\rm Gauss}\right] \tag{8.33}$$

[16]See Footnote 3 in this chapter.

using Eq. (8.28) again. The fundamental frequency is very low and its harmonics are so closely spaced (Example 8.5) that the synchrotron spectrum, *unlike* the cyclotron spectrun, is essentially *continuous*. The end result is a very broad-band continuous spectrum that peaks at $\nu_{\max}$ and becomes negligible at $\nu > \nu_{\rm crit}$. Since the spectrum spans frequencies that are typically much greater than the plasma frequency, even small magnetic fields produce observable synchrotron radiation if a supply of relativistic electrons is present. It can be shown that the power emitted by a single relativistic particle is a strong function of the particle rest mass, i.e. $P \propto 1/{\rm m}^4$. Therefore, the emission from relativistic protons, which are more abundant than electrons (see Sect. 3.6.1) is negligible.

Example 8.5

For the same conditions as given in Example 8.4*, and assuming that the relativistic electron has pitch angle of* $\phi = 45^\circ$*, determine the total opening angle of the emission cone, the critical frequency, and the spacing of the harmonics for this particle.*

For $\gamma = 10^4$ (Eq. 8.31), $2\,\theta = 2/\gamma = 2 \times 10^{-4}$ rad or 0.01°. From Eq. (8.32), $\nu_{\rm crit} = 890\,{\rm MHz}$. The harmonic spacing (Eq. 8.33) is every $\nu_f = 1.7 \times 10^{-9}\,{\rm MHz}$.

The final step is to consider an ensemble of particles. Unlike thermal Bremsstrahlung emission, the electron velocities do *not* follow a Maxwellian distribution (synchrotron radiation is *non-thermal*). These relativistic particles comprise the electron component of cosmic rays and so form a *power law distribution* in energy (see, e.g. Eq. 3.43) which can be expressed in the form,

$$N(E) = N_0\,E^{-\Gamma} \tag{8.34}$$

where $N(E)$ (cgs units of ${\rm cm}^{-3}\ {\rm erg}^{-1}$) is the number of cosmic ray electrons per unit volume per unit interval of electron energy, N_0 (cgs units of ${\rm erg}^{\Gamma-1}\ {\rm cm}^{-3}$) is a constant (Ref. [116]) and Γ is the energy spectral index of the cosmic ray electrons (see Sect. 3.6.2). An integral over electron energy will return the total volume density of cosmic ray electrons, $N_{\rm CRe}$,

$$N_{\rm CRe} = \int_{\rm E_{min}}^{E_{\max}} N(E)\,dE \tag{8.35}$$

where $E_{\min}$, $E_{\max}$ are the minimum and maximum electron energies, respectively, that are present in the relativistic gas.

Assuming that the overall distribution of electrons follows a power law as described above and possesses an isotropic velocity distribution, the final emission coefficient,

Table 8.2 Pacholczyk's constants[a]

$c_1 = 6.27 \times 10^{18}$		
Γ	$c_5(\Gamma)$	$c_6(\Gamma)$
0.5	2.66×10^{-22}	1.62×10^{-40}
1.0	4.88×10^{-23}	1.18×10^{-40}
1.5	2.26×10^{-23}	9.69×10^{-41}
2.0	1.37×10^{-23}	8.61×10^{-41}
2.5	9.68×10^{-24}	8.10×10^{-41}
3.0	7.52×10^{-24}	7.97×10^{-41}
3.5	6.29×10^{-24}	8.16×10^{-41}
4.0	5.56×10^{-24}	8.55×10^{-41}
4.5	5.16×10^{-24}	9.24×10^{-41}
5.0	4.98×10^{-24}	1.03×10^{-40}
5.5	4.97×10^{-24}	1.16×10^{-40}
6.0	5.11×10^{-24}	1.34×10^{-40}

[a]Ref. [116]. Note all quantities are in cgs units. Γ is the spectral index of the electron energy distribution given by Eq. (8.34).

j_ν, and absorption coefficient, α_ν, can be determined. Provided $\Gamma > 1/3$, these are,

$$j_\nu = c_5(\Gamma)\, N_0\, B_\perp{}^{\frac{\Gamma+1}{2}} \left(\frac{\nu}{2\,c_1}\right)^{-\frac{(\Gamma-1)}{2}} \tag{8.36}$$

$$\alpha_\nu = c_6(\Gamma)\, N_0\, B_\perp{}^{\frac{\Gamma+2}{2}} \left(\frac{\nu}{2\,c_1}\right)^{\frac{-(\Gamma+4)}{2}} \tag{8.37}$$

The constant, c_1 and the functions, $c_5(\Gamma)$ and $c_6(\Gamma)$ are collectively (and loosely) called *Pacholczyk's constants* and are provided in Table 8.2. The absorption coefficient is needed to account for the absorption of photons by the relativistic electrons within the synchrotron-emitting cloud, a case called *synchrotron self-absorption*. The *Source Function* Eq. (6.5)[17] and optical depth (Eqs. 5.8, 5.9) are then, respectively,

$$S_\nu = \frac{j_\nu}{\alpha_\nu} = \frac{c_5(\Gamma)}{c_6(\Gamma)}\, B_\perp{}^{-\frac{1}{2}} \left(\frac{\nu}{2\,c_1}\right)^{\frac{5}{2}} \tag{8.38}$$

$$\tau_\nu = \alpha_\nu\, l = c_6(\Gamma)\, N_0\, B_\perp{}^{\frac{\Gamma+2}{2}} \left(\frac{\nu}{2\,c_1}\right)^{\frac{-(\Gamma+4)}{2}} l \tag{8.39}$$

where l is the line-of-sight distance through the synchrotron-emitting source and we have assumed no change in properties along the line of sight for Eq. (8.39). Since Γ is *positive* for a typical cosmic ray energy distribution, Eq. (8.39) indicates that the *optical depth increases with decreasing frequency*. Therefore a synchrotron-emitting cloud becomes more opaque at lower frequencies.

[17]Note that this is *not* an LTE situation so we cannot use Eq. (6.18) for the Source Function.

As we have done before, these equations can be put into the solution to the Equation of Radiative Transfer to determine the *synchrotron spectrum*. Following the equations in Sect. 6.3.5,

$$I_\nu = \mathrm{S}_\nu \left(1 - e^{-\tau_\nu}\right) \qquad (\text{all } \tau_\nu) \tag{8.40}$$

$$I_\nu = \mathrm{S}_\nu \propto \nu^{\frac{5}{2}} \qquad (\tau_\nu \gg 1) \tag{8.41}$$

$$I_\nu = \mathrm{j}_\nu \, l \propto \nu^{-\frac{(\Gamma-1)}{2}} \propto \nu^{\alpha} \qquad (\tau_\nu \ll 1) \tag{8.42}$$

where

$$\alpha \equiv -\frac{(\Gamma - 1)}{2} \tag{8.43}$$

is the *frequency spectral index* or just *spectral index* (not to be confused with the absorption coefficient which is distinguished by the subscript, ν). In Eqs.(8.41) and (8.42), we have made a comparison to Eqs. (8.38) and (8.36), respectively, to see how the spectrum varies with frequency. Synchrotron spectra for several values of Γ determined from Eq. (8.40) are shown in Figure 8.14 along with the observed spectrum

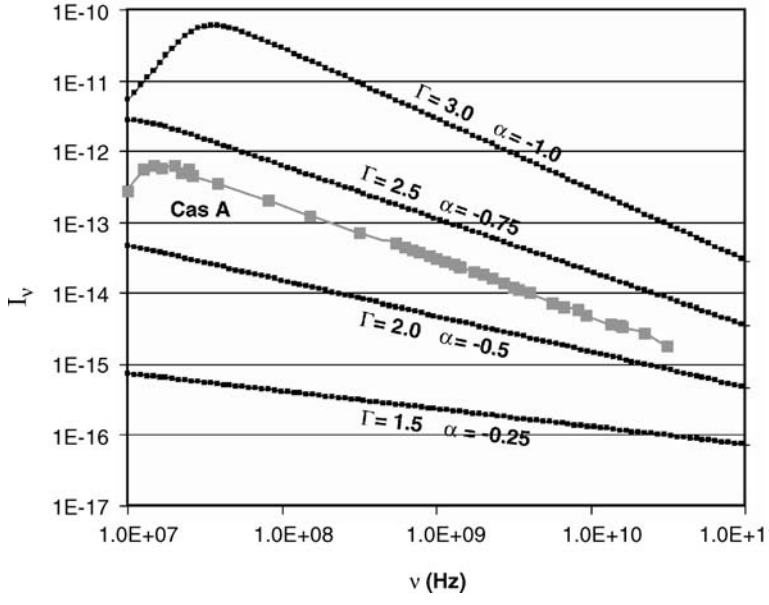


Figure 8.14. Examples of the non-thermal synchrotron spectrum with ordinate (I_ν) in cgs units. All curves with small black boxes are labelled with their electron energy spectral index, , and the corresponding observed frequency spectral index, α, given by Eq. (8.43). These curves have been calculated from Eq. (8.40) using (in Eqs. 8.38 and 8.39): $N_0 = 10^{-10}$ erg^{r-1} cm^{-3}, $B = 100\,\mu$G, $l = 10$ pc and values of the Pacholczyk's constants from Table 8.2 that correspond to the value of Γ used. The grey curve with larger boxes is the observed spectrum of the supernova remnant, Cas A (see image in Figure 1.2), using data from Ref. [11] and an angular diameter of 4.0′

of the strongest radio source in the sky (aside from the Sun), the supernova remnant, Cas A. Note that power laws are straight lines in this log–log plot.

8.5.3 Determining synchrotron source properties

Like any spectrum, that of synchrotron emission allows us to link the received radiation to properties of the source. For simplicity, it is helpful to consider the optically thick and optically thin parts of the spectrum separately.

As indicated by Eq. (8.41), if the emission is *optically thick*, the spectrum *rises* with frequency ($\propto \nu^{5/2}$). This can be seen for the curve with $\Gamma = 3.0$ in Figure 8.14 at the low frequency end of the spectrum. In this part of the spectrum, the emission is described by Eq. (8.38) which is a function only of $B_\perp$ and some constants. This physical situation is not so different from that of an optically thick thermal source for which only the temperature can be determined. Since we cannot 'see' all the way through an optically thick synchrotron source, we cannot detect all the particles that are present in the relativistic gas. Therefore, the brightness of the source is independent of particle density. An important consequence is that *a single measurement of emission from an optically thick synchrotron source leads to a determination of the strength of the perpendicular magnetic field*. For a statistical sample of sources, on average we expect $B_\perp \approx B$ to order of magnitude, so $B_\perp$ is sometimes simply referred to as the magnetic field strength. In reality, it is sometimes difficult to find a source that shows a true optically thick synchrotron spectrum, however. It is the compact sources, such as active galactic nuclei, that are most often optically thick and these tend to consist of a number of CR populations with different power law energy distributions and optical depths that, together, make a complex spectrum. For other sources, the optically thick turnover may be at too low a frequency to be observed, or else may be caused by free–free absorption from contaminating ionized gas (see Sect. 8.2.1), a situation that is likely to be the case for the supernova remnant, Cas A (Ref. [85], see Example 8.6). If a true synchrotron self-absorption spectrum can be identified, however, the magnetic field strength follows (Prob. 8.13).

As indicated by Eq. (8.42), if the emission is *optically thin*, the spectrum *decreases* as a power law ($\propto \nu^\alpha$), several of which are shown in Figure 8.14. The power law slope in frequency space, quantified by the spectral index, α, is a characteristic of optically thin synchrotron emission and is a consequence of the power law slope of the electron energy distribution, Γ, the two being linked by Eq. (8.43). In the optically thin limit, the emission is described by Eq. (8.42) with j_ν given by Eq. (8.36). Thus, the emission depends on *both* the particle density (via the constant, N_0) and the magnetic field strength.

Example 8.6

The supernova remnant (SNR), Cas A (Figure 1.2), has an internal magnetic field of order a few mG (Ref. [8]). At what frequency will this SNR become optically thick to its own synchrotron radiation?

For the source to be optically thick implies that $\tau_\nu = 1$ so we need to solve Eq. (8.39) for $\nu_{\tau=1}$. We let $B_\perp = 2 \times 10^{-3}$ G and, to obtain Γ, we measure the spectral index, α, from

the optically thin part of Figure 8.14, finding $\alpha = -0.77$. Using Eq. (8.43), this gives $\Gamma = 2.54$. From Table 8.2, Pacholczyk's constants for this value of Γ are, $c_1 = 6.27 \times 10^{18}$, $c_5 = 9.51 \times 10^{-24}$, and $c_6 = 8.09 \times 10^{-41}$ (assuming linear interpolation). For the line of sight distance, we assume that the depth through Cas A is approximately equal to its diameter. From the caption to Figure 1.2, this is $l = 4$ pc. We now have all quantities required in Eq. (8.39) except N_0.

To find N_0, we must look at the optically thin part of the spectrum. From Figure 8.14, we take the point at $\nu = 10^{10}$ Hz and measure a value of $I_\nu \approx 4 \times 10^{-15}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}. Using the applicable equation for the optically thin limit (i.e. Eq. 8.42 together with Eq. 8.36 for j_ν) gives, $N_0 = 2.01 \times 10^{-13}$ erg$^{1.54}$ cm^{-3}.

Finally, putting all required values into Eq. (8.39) and solving for frequency gives, $\nu_{\tau=1} = 8.3$ MHz which is just to the left of the frequency range shown in Figure 8.14[18]. Therefore, a more likely explanation for the observed higher frequency turnover is the presence of absorbing ionized thermal gas.

As suggested above, for many sources, observations of pure synchrotron emission in the optically thick limit are not always possible, whereas *optically thin* emission is readily observed. This, then, poses a problem because emission of optically thin radiation depends on both the magnetic field strength and the CR particle density. We cannot obtain one without knowing the other. Moreover, the particle density, $N_{\rm CRe}$ Eq. (8.35), does not really give us the most important property of the relativistic gas. There may be few CR electrons present, but the energy associated with those electrons and other CR nuclei can be very high. If we wish to identify the original source that is producing the relativistic gas, then some determination of the *energy* contained in the relativistic gas and associated magnetic field is necessary.

The *total energy* represented by all CR particles and magnetic fields, is $U_{\rm T} = U_{\rm CR} + U_{\rm B}$, where $U_{\rm CR}$ includes both electrons as well as heavier nuclei and $U_{\rm B}$ is the energy contained in the magnetic field. Since it is the CR electron component only that emits synchrotron radiation, the heavy particle component must be estimated. It is common to represent the total CR energy, then, as $U_{\rm CR} = (1 + {\rm k})\, U_{\rm e}$, where k is a constant representing the ratio of energy contained in heavy nuclei compared to electrons, and $U_{\rm e}$ is the energy of the CR electrons only. The constant, k is not always well known and may vary in different environments, but is thought to be in the range, 40–100.

The energy contained in the magnetic field increases with magnetic field strength (see the equation for magnetic energy density in Table I.1). On the other hand, the energy contained in the cosmic ray particles *decreases* with magnetic field strength. This is a consequence of the fact that particles emit radiation more strongly (see Eq. 8.36), therefore losing energy faster, in higher magnetic fields. Thus, there is a *minimum* in the function, $U_{\rm T}$ when plotted against B. A conservative assumption is

[18]For a pure synchrotron spectrum, the frequency of the peak of the spectrum does not exactly coincide with the frequency at which the spectrum becomes optically thick but, for Cas A, the corresponding peak would be about 10^7 MHz.

therefore commonly made that the total energy in a synchrotron-emitting source is at or near this minimum. The minimum energy values also turn out to be very close to those of *equipartition of energy*, i.e. the condition in which the energy of the magnetic field is equal to the energy in CR particles. By making the assumption of minimum energy, we introduce one more constraint into the calculations which allows us to extract information on *both* the magnetic field strength and CR particle energy from observations of synchrotron radiation in the optically thin limit. This method is widely used to obtain magnetic field strengths and energies or energy densities in synchrotron-emitting objects. Even with the minimum energy assumption, the derived energies can be extremely high, especially for distant extra-galactic radio sources (see next section). Further details involving the minimum energy criterion and development can be found in Ref. [116].

Like cyclotron emission, synchrotron emission is intrinsically *polarized* for a magnetic field direction that is uniform. For synchrotron emission, it can be shown that the maximum degree of polarization is $D_p \approx 70$ per cent for $2 \lesssim \Gamma \lesssim 4$ (Ref. [177]). In fact, a detection of polarization is the best way of being sure that the emission is indeed synchrotron. As indicated above, a power law slope is characteristic of optically thin synchrotron emission, but the power law spectral index, α, depends on the energy injection spectral index, Γ (Eq. 8.43) which may vary depending on the specific acceleration mechanism. In addition, some unresolved sources (like active galactic nuclei) contain a mixture of different sources which result in the superimposition of various values of α. The net result is that, although most optically thin spectra are similar to those shown in Figure 8.14, it is possible for the observed value of α to be flat or even inverted (positive). If a flat slope were to exist, then this would mimic a thermal Bremsstrahlung slope. The solution to the dilemma is a polarization observation. Since thermal Bremsstrahlung emission is never polarized, a clear polarization detection decides the issue.

As is sometimes the case in nature, real situations may not be so clear cut. Most synchrotron sources *do* show polarization. However, sources can contain both uniform and random components to their magnetic fields. The interstellar medium in the Milky Way, for example, has both uniform and random magnetic field components, each of which is of order, a few μ Gauss (Table 8.1). In other sources, the magnetic field lines may become 'tangled' and there may be reduced (or no) directionality to the field within the size scale of the telescope's beam. This lowers the degree of polarization. Thus, as indicated in Sect. 1.7, even strong radio jets tend to have $D_p \lesssim 15$ per cent. A clear detection of polarization is therefore indicative of synchrotron radiation but lack of polarization does not rule it out. Very high brightness temperatures, however, are an additional clue that the emission is synchrotron (see Sect. 8.1).

8.5.4 Synchrotron sources – spurs, bubbles, jets, lobes and relics

Most synchrotron-emitting sources are observed in the optically thin limit at low frequencies where the emission is strongest (Figure 8.14). Although routine

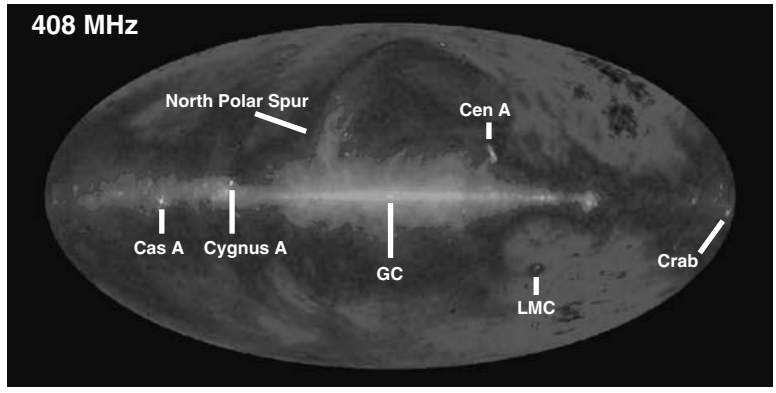


Figure 8.15. All-sky map of synchrotron emission at $\nu = 408\,\mathrm{MHz}$ in the same projection as shown in Figure 8.5. (The map is in false colour with brightest to dimmest represented by white/yellow through purple/deep blue and finally red.) Most emission from our own Milky Way is seen along the Galactic plane as well as from spurs (e.g. the North Polar Spur) extending away from the plane. The Galactic Centre (GC) is marked, as well as the supernova remnants, Cas A and the Crab Nebula. Most of the discrete sources away from the plane are extra-galactic, including the Large Magellanic Cloud (LMC, a nearby galaxy), and the radio galaxies, Cygnus A and Centaurus A (for the latter, the extended, double-lobed shape can be discerned) (Reproduced by permission of the Max-Planck-Institut für Radioastronomie) (See colour plate)

measurements are now made for the stronger sources at optical and X-ray wavelengths, synchrotron radiation is still most readily detected at *radio frequencies*. Since radio waves do not suffer from dust extinction (see Appendix D.3), this also means that synchrotron radiation is quite easily detected, passing through the ISM of the Milky Way without significant attenuation. Unlike cosmic ray particles which are easily scattered in the ISM (see Sect. 3.6.3), radio photons come directly from the location at which they are generated, helping us to identify their origin.

Figure 8.15, for example, shows an all-sky image of $\nu = 408\,\mathrm{MHz}$ emission which contains predominatly synchrotron radiation. As discussed in Sect. 3.6.3, supernovae are likely candidates for the acceleration of most Galactic cosmic rays at energies below the 'knee' of Figure 3.25 and their observed spectral indices of $-0.8 \lesssim \alpha \lesssim -0.2$ imply $1.4 \lesssim \Gamma \lesssim 2.6$ (Eq. 8.43) which spans the range, $2.1 \lesssim \Gamma_0 \lesssim 2.5$, believed to describe the initial CR energy spectrum (Sect. 3.6.2). Two supernova remnants, Cas A and the Crab Nebula, are marked in the figure, and the *North Polar Spur* is part of a local, 250 pc diameter *superbubble* which may also have been formed by one or more supernovae (Ref. [182]). This superbubble is likely related to the Local Hot Bubble discussed in Sect. 8.2.3.

Although supernova remnants are important sources of synchrotron emission, they are by no means the only sources. In the realm of high energy astrophysics, there are many objects with sufficient energies to accelerate electrons and sufficiently strong magnetic fields to scatter them. In the Milky Way, these include shocks formed by massive stellar winds, stellar jets, some binary stars, pulsars, and others. Synchrotron emission is even observed from Jupiter and the Sun. Cosmic ray electrons that leak into the ISM from discrete acceleration sites will also emit synchrotron radiation, since

magnetic fields are ubiquitous in the disk of the Milky Way. Typical ISM magnetic field strengths of only a few μ Gauss are sufficient to generate observable emission. Thus, much diffuse emission can be seen along the plane of the Galaxy in Figure 8.15.

Extra-galactic sources include active galactic nuclei (AGN) that are powered by supermassive black holes (see Sect. 5.4.2 and Figure 5.5) and the bipolar jets that they generate (e.g. Figure 1.4). The classic *extra-galactic double-lobed radio sources* were the first discovered in the radio band and display two radio lobes at the ends of the jets, as can be discerned for the radio galaxy, Centaurus A, in Figure 8.15. The powerful radio galaxy, Cygnus A, is also identified in Figure 8.15 and is shown in more detail in Figure 8.16. The inconspicuous elliptical galaxy at the centre, whose optical light comes from stars, belies the true nature of its powerful core. The total energy contained in its magnetic fields and cosmic ray particles, using the minimum energy assumption (see previous section), is a spectacular 10^{60} erg (Ref. [123]). By comparison, the total kinetic energy of a single supernova is $\sim 10^{51}$ erg. Not only are the particles that emit synchrotron radiation relativistic, but the jets that feed the two lobes also have bulk motions (meaning that the entire synchrotron-emitting plasma is moving in bulk) that are relativistic, with velocities between $0.4\,c$ and $\sim 1\,c$. The minimum energy magnetic field strength in the radio lobes is, on average, 50 μG.

AGN activity is a strong contender for the origin of magnetic fields and cosmic rays in the intergalactic medium of clusters of galaxies from which synchrotron emission has also been detected. In some cases, regions of synchrotron emission are seen

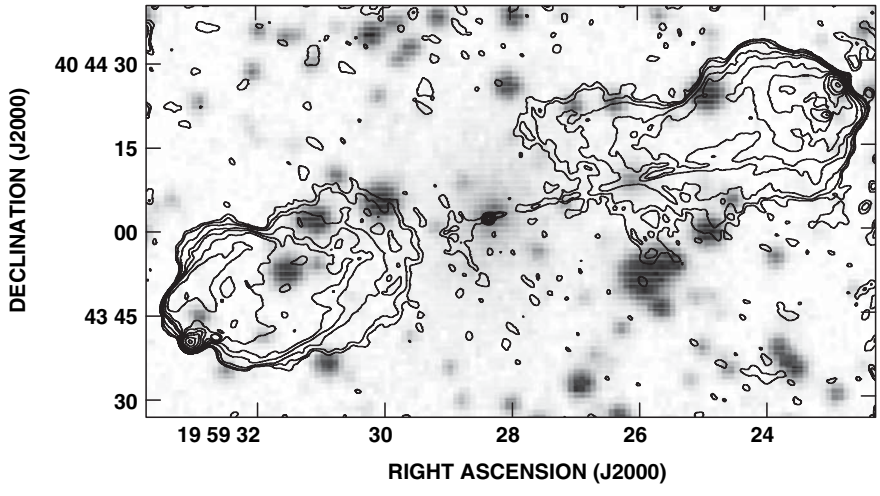


Figure 8.16. The double-lobed extra-galactic radio source, Cygnus A (redshift, $z = 0.0562$), shown in contours of $\nu = 4.9$ GHz emission, over the optical image (shown with black as highest intensity) of the same field of view. The faint greyish region at the centre of the image is the distant elliptical galaxy that harbours a bright AGN at its centre. (Other black spots are stars in our own Milky Way.) Two jets emerge from the AGN in opposite directions, ending in two large bright radio lobes on either side of the galaxy. The total flux density of the source is 213 Jy and the effective FWHM of a single lobe is $\theta_{\rm FWHM} \approx 10"$.

in the outskirts of galaxy clusters (for an optical image of a cluster of galaxies, see Figure 8.8). These regions are called *relics*, or sometimes *fossils* or *ghosts*. The origin of this emission is not entirely clear since these regions do not appear to be directly connected to currently active galaxies within the cluster. Rather, the emission may be generated within shocks that formed during the formation of the cluster and/or during later cluster mergers. An older, existing population of cosmic ray particles that were ejected from radio galaxies in the past could be reaccelerated in such shocks.

8.6 Inverse Compton radiation

Inverse Compton (IC) radiation is somewhat different from the other emission discussed in this chapter because it results from the interaction of a particle with a photon. However, since the emission begins with matter and ends with a photon (i.e. matter $\rightarrow$ photon $\rightarrow$ photon) we include it here. As indicated in Sect. 5.1.2.2 and further detailed in Appendix D.2, Compton scattering occurs when a high energy photon scatters inelastically from a free electron, boosting the kinetic energy of the electron at the expense of the photon energy. IC radiation, often referred to as *Inverse Compton scattering*, is the reverse of this process, as its name implies. It occurs when a high energy electron inelastically scatters off of a low energy photon, the result being a loss of kinetic energy for the electron and a gain of energy (a *blueshift*) of the photon. The scattering will go in this direction if the electron energy is greater than the photon energy in the centre-of-momentum rest frame of these two 'particles'. For a highly relativisitic electron and low energy photon, this rest frame is that of the moving electron. For such a case, it can be shown that the frequency of the outgoing radiation, as seen by a distant observer, is,

$$\nu_{\rm IC} \approx \gamma^2 \nu \quad (\gamma h \nu \ll {\rm m_e} c^2) \tag{8.44}$$

where γ is the Lorentz factor Eq. (3.37) of the electron and ν is the frequency of the photon before being scattered[19]. For example, for a relativistic electron with $\gamma = 10^4$, a radio photon, at a typical frequency of $\nu = 1\,{\rm GHz}$, would be 'up-scattered' to $\nu_{\rm IC} = 10^{17}\,{\rm Hz}$ which is in the *X-ray* part of the spectrum (see Table G.6). The result is a redistribution of photon energies with fewer photons in the radio and more photons in the optical or X-ray (depending on γ) than would otherwise be the case. In the presence of a magnetic field, relativistic electrons will also emit synchrotron radiation, so the addition of IC losses implies that the electrons will lose energy faster and the radiated output over all ν will be higher than from synchrotron emission alone.

Any synchrotron-emitting source contains a plentiful supply of high-energy electrons and emits low energy (e.g. radio) photons and these are also the required ingredients for IC emission. Therefore, IC radiation, when observed, will be seen in the same kinds of sources that emit synchrotron radiation. Because the relativistic

[19]Note that an electron moving with a Lorentz factor, γ, will 'see' a photon of energy, $\gamma h \nu$ in its own rest frame.

electrons have a power law distribution of energies, this IC radiation is *non-thermal*. Moreover, the IC spectrum is also a power law and so is difficult to distinguish from that of synchrotron radiation. However, not every synchrotron-emitting source is an IC source as well. Synchrotron radiation depends on the strength of the magnetic field, B, whereas IC radiation depends on the number of low energy photons that are available for up-scattering. Thus, for a given number and distribution of relativistic electrons, wherever B is strong and the radiation field is weak, synchrotron radiation will dominate, and wherever the radiation field is strong and B is weak, IC radiation will dominate. To be more precise, it can be shown that the fraction of the total radiative luminosity of a source due to IC radiation, $L_{\rm IC}$, in comparison with the total radiative luminosity due to synchrotron radiation, $L_{\rm s}$, is (Ref. [155]),

$$\frac{L_{\rm IC}}{L_{\rm s}} = \frac{u_{\rm rad}}{u_{\rm B}} \approx \left[\frac{\mathrm{T_{B\,max}}}{10^{12}}\right]^5 \left[\frac{\nu_{\rm max}}{10^{8.5}}\right] \tag{8.45}$$

where $u_{\rm rad}$ is the energy density of the radiation field (see Sect. 1.4) and $u_{\rm B}$ is the energy density of the magnetic field (see Table I.1). The quantities, $\mathrm{T_{B\,max}}$ (K), and $\nu_{\rm max}$ (Hz), are the brightness temperature and frequency, respectively, of the source at the peak (the turnover) of the synchrotron spectrum, assuming that only synchrotron self-absorption is producing the turn-over. Example 8.7 shows how the brightness temperature and frequency dependence in this equation results.

Example 8.7

Show how the functional dependence on T_B and ν in Eq. (8.45) *is obtained.*

From Eq. (1.16), the energy density of the radiation field, $u_{\rm rad} \propto I$, where I is the intensity of the source already integrated over frequency, so $u_{\rm rad} \propto I_\nu\,\nu$, where I_ν is the specific intensity. Using the definition of brightness temperature (Eq. 4.4) expressed in the Rayleigh–Jeans limit Eq. (4.6) since most of the photons will be in the radio part of the spectrum and $h\,\nu \ll k\,\mathrm{T_B}$, we have, $I_\nu \propto \mathrm{T_B}\,\nu^2$, which leads to, $u_{\rm rad} \propto \mathrm{T_B}\,\nu^3$.

Now, from Table I.1, $u_{\rm B} \propto B^2$. In the optically thick limit, the magnetic field strength is related to the specific intensity via Eqs. (8.41) and (8.38) and this dependence should still be approximately correct at the peak of the curve where the spectrum turns over. Letting $B \approx B_\perp$, and rearranging Eq. (8.38) to solve for B gives $B \propto \nu^5/{I_\nu}^2$. Thus, $u_{\rm B} \propto \nu^{10}/{I_\nu}^4$. Now, converting the specific intensity to a brightness temperature as we did above, then, $u_{\rm B} \propto \nu^2/{\mathrm{T_B}}^4$.

Finally, from the above two results, the ratio, $u_{\rm rad}/u_{\rm B} \propto (\mathrm{T_B}\,\nu^3)/(\nu^2/{\mathrm{T_B}}^4) \propto {\mathrm{T_B}}^5\,\nu$ which is consistent with Eq. (8.45).

The importance of Eq. (8.45) is in the extremely strong dependence of IC losses on the brightness temperature of the source. IC radiation should not be very important until the brightness temperature approaches 10^{12} K at the reference frequency given in the equation (e.g. Prob. 8.14). Once brightness temperatures exceed 10^{12} K, however, a

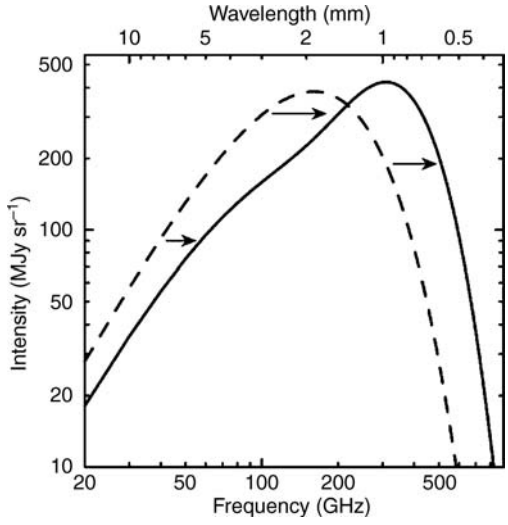


Figure 8.17. Illustration of the Sunyaev–Zeldovich effect, showing the undistorted CMB spectrum (dashed curve) and the spectrum after CMB photons have passed through the ionized gas in a cluster of galaxies (solid curve). The cluster has been made 1000 $\times$ more massive than a typical cluster to exaggerate the effect. (Reproduced by permission of John Carlstrom, Carlstrom, J. E., Holder, G. P., and Reese, E. D., 2002, ARAA, **40**, 643)

situation called the *Compton catastrophe*, the electrons should lose energy so rapidly from IC losses (of order days, Ref. [132]) that no sources with such high brightness temperatures should actually exist. Thus Eq. (8.45) suggests a natural upper limit of $T_B \approx 10^{12}$ K for source brightness temperatures[20]. In fact, this limit appears to be approximately correct when compared to observations. There are sources in which higher brightness temperatures are observed, but these are *moving* sources which were not considered when deriving Eq. (8.45). If the entire synchrotron-emitting plasma has a bulk motion that is moving relativistically towards the observer, such as in a jet pointed towards the observer, then the emission will be brighter than if the plasma were stationary, a situation called *Doppler beaming*.

The most famous example of inverse Compton scattering is the *Sunyaev–Zeldovich effect* (S–Z effect). This is the IC up-scattering of 2.7 K CMB photons by very hot (but non-relativistic) electrons in clusters of galaxies, shifting the CMB photons to higher frequencies (see Prob. 8.15). An abundant supply of ionized gas, and therefore hot electrons, is present in some galaxy clusters, as illustrated in Figure 8.8, and background photons will pass through such clusters en route to us. The result is a distortion of the 2.7 K CMB spectrum. Rather than the 'perfect' black body that we saw in Figure 4.3, a distorted spectrum, such as shown in Figure 8.17 results. Although the effect is typically quite small ($\Delta T < 1$ mK), it has now been measured in many galaxy clusters. The importance of the effect is that the observed shift depends on the electron density and temperature (n_e, T_e). Since X-ray measurements of thermal

[20]Other considerations related to equipartition of energy may make the upper limit somewhat smaller, of order $T_B \approx 10^{11}$ K (see Ref. [132]).

Bremsstrahlung emission also depend on n_e and T_e, the S–Z effect provides an independent check on these values. More importantly, if X-ray data are available and T_e is known (Sect. 8.2.3), combining all information allows the determination of the *distance* to the cluster. Knowledge of the distance and total gas mass (via Eq. 8.9) leads to crucial data on the cosmological parameters (see Sect. 8.2.3, Figure 3.4).

Problems

8.1 (a) Determine the root-mean-square speed, v_{rms}, and corresponding kinetic energy, E_k, of an electron in an HII region of electron temperature, $T_e = 10^4$ K.

(b) If the electron lost all of its kinetic energy (came to a stop) during an encounter with a nucleus without recombining, what would be the frequency of the resulting emission and in what part of the spectrum does this frequency occur?

(c) Compare the frequency of part (b) with a typical thermal Bremsstrahlung photon at a radio frequency of $\nu = 5 \times 10^9$ Hz from an HII region. Based on this comparison, what fraction of an electron's kinetic energy has actually been lost during an encounter with a nucleus?

8.2 The spectrum of an HII region is observed to have an exponential cutoff such that the emission declines by a factor, $1/e$ (its *e-folding* value), at a wavelength of $\lambda 2.4$ μm. What is the temperature of this HII region?

8.3 With the help of a spreadsheet or computer algebra software, determine and plot the spectrum of free–free emission from the Lagoon Nebula shown in Figure 3.13. Ensure that the low frequency optically thick part of the spectrum and the high frequency cutoff are shown. For simplicity, assume that the ionized gas consists of pure hydrogen and that the Gaunt factor is constant with frequency.

8.4 (a) Show that $h\nu \ll k T_e$ for any HII region observed at radio wavelengths.

(b) Show that Eq. (8.12) results from Eq. (8.4).

8.5 For the HII region, IC 5146, whose radio emission, as shown in Figure 8.4, is in the optically thin limit, find the following:

(a) The frequency at which the HII region becomes optically thick.

(b) The electron density, n_e, ionized hydrogen mass, M_{HII}, total gas mass, M_g, and excitation parameter, $\mathcal{U}$.

(c) The radius of its Strömgren sphere (assume that the effects of dust are negligible).

(d) The number of ionizing photons per second, N_i, from the star, BD+46°3474.

8.6 Using the parameters of the hot intracluster gas given in Figure 8.3, determine the frequency at which the emission becomes optically thick. In what part of the spectrum does this occur and is it observable from the ground?

8.7 Show that the luminosity within a frequency band, ν_1 to ν_2, is given by Eq. (8.19) for optically thin thermal Bremsstrahlung radiation from ionized hydrogen gas of constant temperature and density.

8.8 For the cluster of galaxies, Abell 2256, whose soft X-ray emission is shown in Figure 8.8, determine or estimate the following:

(a) Its distance, D (Mpc).

(b) The volume, V (kpc^3), of the X-ray emitting gas, assuming that it occupies the entire volume of the cluster.

(c) The gas temperature, T_e (K).

(d) The mean electron density, n_e (cm^{-3}).

(e) The ionized hydrogen mass, M_{HII} ($M_\odot$).

(f) The fraction, M_{gas}/M_{tot}, where M_{gas} is the total gas mass and M_{tot} is the total light+dark mass of the cluster.

8.9 (a) Write an expression for the conservation of energy for an electron which makes a transition from an initial free state in an ionized hydrogen gas to a final bound state, showing that the difference in energy corresponds to the energy of the emitted photon.

(b) Compute the frequency of an emitted photon for a transition to the $n = 1$ *and* $n = 2$ levels of hydrogen when (i) the initial electron velocity is zero, and (ii) the initial electron velocity is the most probable velocity in a gas at $T_e = 1.58 \times 10^5$ K.

(c) Compare the results of part (b) to the curve of g_{fb} shown in Figure 8.2 and explain the shape of this curve.

8.10 Cyclotron emission from electrons moving at $v \approx 2 \times 10^4$ km s^{-1} in a region of the magnetosphere of Jupiter is observed at the gyrofrequency, $\nu_0 = 540$ kHz.

(a) Determine the magnetic field in this region.

(b) Find the gyroradius of the electrons. Is this larger or smaller than the wavelength of the emitted radiation?

(c) Write an expression for ν_0/ν_p in terms of B and n_e with constants evaluated. What is the maximum possible electron density of this region?

(d) What is the gyrofrequency of a proton, ν_{op}, in the same region? Repeat the first part of step (c) above using ν_{op} and determine the maximum possible electron density that would allow cyclotron emission from a proton to escape from the region.

8.11 Suppose a Jupiter-like planet is orbiting a Sun-like star and that this stellar system is at a distance of 10 pc from the Earth. The system is observed at $\nu = 30$ MHz and is unresolved. Adopt T_B (30 MHz) $= 5 \times 10^5$ K (Sect. 4.1.1).

(a) What is the flux density of the star (cgs units) as measured at the Earth? Assume that the star is *not* undergoing any radio bursts.

(b) From the information given in Figure 8.12, what is the flux density of the planet (cgs) as measured at the Earth? Assume that the planet *is* undergoing a radio burst, and has a modulating moon like Io.

(c) Comment on the prospects of measuring radio bursts from extrasolar planets at low frequencies and the possibility of detecting extrasolar planets via their radio emission assuming that the telescope is sensitive enough to detect the stellar radio emission.

8.12 Determine the relativistic gyroradius (kpc) of a CR proton that has the highest energy shown in Figure 3.25 in the ISM of the Milky Way. Compare this to the distance of the Large Magellanic Cloud (see Prob. 7.16 for data) and comment on the ability of galaxies, in general, to retain such high energy cosmic rays.

8.13 In 1993, a supernova (SN1993J) was observed in the nearby galaxy, M 81, which is a distance $D = 3.6$ Mpc from us. At a time, 273 days after the observed explosion[21], the following parameters (approximated from Ref. [122]) described the radio emission from this source: (i) a rising spectrum ($I_\nu \propto \nu^{5/2}$) at low frequencies with a flux density of $f_\nu = 72$ mJy at a frequency of $\nu = 1.43$ GHz, (ii) a falling spectrum ($I_\nu \propto \nu^{-1}$) with a flux density of $f_\nu = 25$ mJy at a frequency of $\nu = 23$ GHz, (iii) a radius of $r = 0.0123$ pc, and (iv) a minimum energy cutoff to the electron energy spectrum of $E_{\rm min} = 46$ MeV. Determine the following parameters for this supernova (all in cgs units):

(a) The electron energy spectral index, Γ.

(b) The solid angle subtended by the source, Ω.

(c) The brightness temperature at $\nu = 1.43$ GHz, T_B.

(d) The strength of the perpendicular magnetic field, $B_\perp$.

(e) The constant of the electron energy spectrum, N_0.

(f) The number density of cosmic ray electrons, $N_{\rm CRe}$.

8.14 The total flux density of Cygnus A (see Figure 8.16) is $f_\nu = 2.19 \times 10^4$ Jy at $\nu = 12.6$ MHz, divided roughly equally between the two radio lobes. Assuming that this flux density represents the peak of the spectrum of Cygnus A and that synchrotron self-absorption is the only absorption process that is producing the spectral turnover, determine the ratio of $L_{\rm IC}/L_{\rm s}$ for one radio lobe. How important is inverse Compton radiation in comparison to synchrotron radiation for this source?

8.15 Verify that the condition in parentheses in Eq. (8.44) is satisfied for the case of CMB photons and hot intracluster gas such as is shown in Figure 8.8.

[21] The actual explosion occurred 11.7 million years prior to its observation in 1993 since M 81 is 3.6 Mpc = 11.7 million light years away.

9
Line Emission

Have you seen anything so beautiful?

– C. V. Raman, pointing to the evening sky (Ref. [128])

By the end of the 19th century, the German physicist, Max Planck, had explained the continuous black body spectrum by allowing light to exist as a photon, or 'particle' (see the Appendix at the end of Chapter 4). His theoretical development introduced the constant, h, now called Planck's constant. The concept of quantized light was further strengthened when Einstein, in 1905, successfully explained the *photoelectric effect* – the release of electrons in some metals and semiconductors via the action of incident light – by treating light as a particle. Still not understood, however, was the series of lines seen in the spectrum of the hydrogen atom. It was the Danish physicist, Niels Bohr, who brought the concept of quantization to the realm of the atom. He proposed, in 1913, that electrons did not radiate unless they made a transition from one state to another. The difference in energy between these states could be related to the emitted energy of a photon, according to $\Delta E = h\nu$. This profound shift from the classical view of requiring orbiting electrons to radiate, to one in which radiation results only from transitions between states, won Bohr the Nobel Prize in 1922. It also set the scene for Heisenberg, Sommerfeld, Schrödinger, Pauli and others to further develop and refine the new field of quantum mechanics.

This chapter is devoted to an understanding of the observed *spectral lines* from atoms and molecules that result from *discrete* internal changes in energy. While the details of quantum theory are beyond the scope of this book, an introduction to quantum mechanical principles for the hydrogen atom has been provided in Appendix C. Much of the groundwork has been laid out earlier. For example, we have seen that the population of bound states in an atom under LTE conditions is described by the Boltzmann Equation (Sect. 5.1.1.2, 5.1.1.3, 5.1.2, and Appendixes D.1.2, D.1.3, D.1.4, and D.1.5. Bound–bound absorption was mentioned in Sect. 5.2.1

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

and recombination lines were introduced in Sect. 3.4.7. The conditions for forming absorption and emission lines were described in Sect. 6.4.2, and the information that can be obtained from detecting the $\lambda 21$ cm line of hydrogen in both emission and absorption was presented in Sect. 6.4.3. In Sect. 7.2, we also discussed how the relatively simple observation of a redshift can lead to some very powerful conclusions, from information about black hole masses to an understanding of the nature of our expanding Universe.

We focus now on what spectral lines can be formed in atoms and molecules and how these spectral lines provide us with information about the sources that emit them. Although a line spectrum can form from pure cyclotron emission (Sect. 8.5.1), we will restrict the discussion to spectral lines that are formed from discrete quantum transitions in atoms and molecules as well as their ions and isotopes, that is, *bound–bound transitions*. We also focus on the line *emission* process and therefore will deal with 'downwards' (higher to lower energy) transitions, ignoring the possible presence of a background source[1]. The latter can be handled by considering the appropriate solution to the Equation of Transfer, as described in Chapter 6.

9.1 The richness of the spectrum – radio waves to gamma rays

What kinds of bound–bound transitions are possible? These depend on the quantizations that occur within atoms and molecules The possibilities are: *electronic transitions*, *vibrational transitions*, *rotational transitions*, and *nuclear transitions*, all of which are quantized.

9.1.1 Electronic transitions – optical and UV lines

Electronic transitions occur when an electron changes from a higher to lower energy state in atoms and molecules or their ions and isotopes. These are the familiar transitions that have already been discussed in various contexts. For example, a rich variety of spectral lines from many different species was seen in the optical spectrum of an H II region in Figure 8.9.

The electronic transitions of hydrogen are discussed in Appendix C. If the hydrogen is ionized, a recombination line spectrum is observed (Sect. 3.4.7) as electrons continuously recombine with protons and then cascade down various bound–bound levels. The highest frequency line that is possible from the hydrogen atom results from the largest energy difference between bound–bound states. This would be close to the ionization energy (see Eq. C.7) of 13.6 eV for which the corresponding wavelength, $\lambda 91.18$ nm, is in the ultraviolet part of the spectrum. At the other extreme, transitions between very high quantum levels result in radio wave emission since the energy levels

[1]The 2.7 K CMB background will always be present, but we will assume that its brightness is negligible, in comparison to the line.

of hydrogen become progressively closer together with increasing n (see Figure C.1). Thus, line emission from hydrogen progressively shifts in frequency across the electromagnetic spectrum from the UV to the radio band, depending on which principal quantum numbers are involved. At higher quantum numbers, however, the transition probabilities are lower (e.g. see values of $A_{j,i}$ in Table C.1), so it is common to refer to electronic transitions in atoms to be 'typically' in the optical or UV parts of the spectrum.

Electronic transitions occur in other atoms and in molecules as well. When all species are taken into account, spectral lines are seen across the entire electromagnetic spectrum from the X-ray to radio bands. X-ray lines have been observed from the highly ionized iron ions in a Solar flare, for example (Figure 6.7). Larger, heavier atoms can have higher energy transitions and also more of them, since there are more electrons associated with such elements. Put together, a wide array of possible electronic transitions are available over a broad waveband, though only a subset of these may be excited at any time, depending on the physical conditions in the gas. Spectral lines are therefore probes of the physical conditions in the gas, as we will soon see. At the very least, the identification of spectral lines as belonging to certain species provides information as to which elements are present in the gas.

9.1.2 Rotational and vibrational transitions – molecules, IR and mm-wave spectra

In molecules, there are two additional types of transitions that can occur, besides electronic ones. These are *rotational transitions* and *vibrational transitions*. Rotation and vibration are illustrated for a diatomic molecule in Figure 9.1, but it should be noted that there are no exact quantum mechanical analogues to these motions. Just as we saw for an electron in 'orbit' about the nucleus (Appendix C.1) and for the 'spin' of an electron (Appendix C.2), the rotation and vibration of a molecule are convenient models that allow us to picture these new energy levels and to derive their energies from a classical starting point.

If a molecule rotates, it has some energy associated with that rotation. A good classical approximation is that of a rigid rotator which, for a diatomic molecule, would be as if the molecule were a dumbbell with a nucleus at either end. As illustrated in Figure 9.1.a, the rotation is about either of two principal axes[2]. If the two nuclei have different atomic weights, then the molecule will have a permanent electric dipole moment (Table I.1) due to the way in which charge is distributed in the molecule. For example, in the CO molecule, the oxygen atom has a slight negative charge and the carbon atom has a slight positive charge. In the classical picture, as the molecule rotates, so does the dipole moment and this would produce electromagnetic radiation

[2]Rotation about the third axis is not energetically important since the moment of intertia, I, is very small. For any object consisting of individual particles, $I = \sum m_i r_i^2$, where m_i is the mass of particle, i, and r_i is its distance from the axis of rotation.

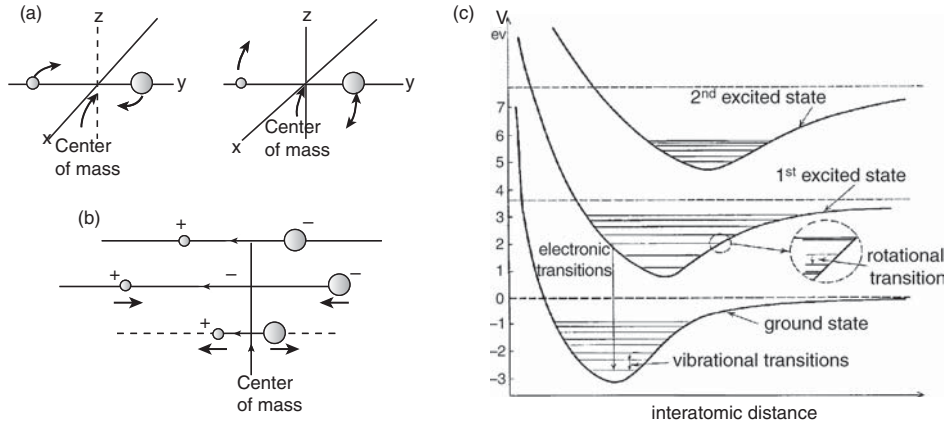


Figure 9.1. Molecular transitions in a diatomic molecule. **(a)** Illustration of rotation. Rotation is in the *x–y* plane at left, and in the *y–z* plane at right. The rotation is quantized and a spectral line is emitted only when there is a transition between two different rotational states. **(b)** Illustration of vibration. The motion resembles that of a simple harmonic oscillator. **(c)** A sample energy level diagram showing the molecule's potential energy (including all nuclei and electrons) as a function of separation between nuclei. Electronic, vibrational and rotational transitions are shown.

continuously. However, just as we saw for 'orbiting' electrons, quantum mechanically, this is not the case. The rotation is quantized and radiation is emitted only when a transition occurs from one rotational state to another. It is nevertheless necessary for an electric dipole to be present in order for a rotational change to produce a rotational spectral line. Molecules like H_2, N_2, and O_2, for example, do not have electric dipole moments since their charge is equally distributed, and therefore these molecules also display no rotationally-induced electric dipole radiation. These molecules may have electric quadrupoles, however[3], and if so, much weaker (by many orders of magnitude) rotational lines can result from changes in the quadrupole moment. As can be seen in the potential energy diagram of Figure 9.1.c, rotational transitions correspond to the smallest energy changes in molecules and therefore these lines have the lowest frequencies, typically in the far-IR, sub-mm or mm parts of the spectrum.

Aside from rotation, a molecule can also vibrate, as illustrated in Figure 9.1.b and a classical analogue is that of simple harmonic motion. As with rotation, the vibrational states are quantized and a spectral line is emitted only if there are changes between these states. The close-to-parabolic shape of simple harmonic motion can be seen in the potential energy diagram of Figure 9.1.c. If the positive nuclei are separated by a small distance, there is little room for electrons between them and the force between the two protons is repulsive. At a large distance, more electrons will occupy the internuclear

[3]Taking a simplified case in which the electrons and protons are along a straight line, the electric dipole moment is $p_1 = \sum q_i x_i$, where q_i is the charge of particle, i, and x_i is its position. The electric quadrupole moment is then $p_2 = \sum q_i x_i^2$. Even though the electric dipole moment may be zero, the electric quadrupole moment may be non-zero, depending on the configuration.

space and the force will be attractive. Thus, the molecule can oscillate over different internuclear separations if perturbed. If the separation becomes too great (to the right on the figure) the molecule will dissociate, but if the separation is close to the minimum of the curve for any given electronic state, then it is stable[4]. Transitions between vibrational states are of much higher energy (typically in the IR) than those between rotational states, the molecule vibrating perhaps 1000 times during a single rotation.

In reality, a single transition may involve a change in *both* vibrational and rotational states and these are called *ro-vibrational transitions*. The changes in quantum numbers must obey the same quantum mechanical selection rules as if they occurred individually. The total change in energy is the sum of the changes in the vibrational and rotational energies, $\Delta E_{\text{ro-vib}} = \Delta E_{\text{vib}} + \Delta E_{\text{rot}}$, but this will be dominated by ΔE_{vib} with ΔE_{rot} small in comparison. Observationally, for a given vibrational transition, all possible allowed rotational transitions also occur and therefore the vibrational line is broken into discrete components with spacing of the rotational steps, i.e. the rotational spectrum is superimposed on the vibrational line. Since the rotational steps are very closely spaced, this spreads out the vibrational 'line' into a spectral *band*. Similarly, it is possible for a transition to occur that involves a change in all three electronic, vibrational, and rotational quantum numbers in one transition. In such a case, the total energy change is the sum of all three, with the electronic energy change dominant.

From the above discussion, it is clear that the observed spectrum of a single molecule will, in general, be far more complex than the spectrum of a single atom. A molecular cloud, moreover, will contain a variety of different kinds of neutral molecules as well as charged molecules and molecules containing isotopes, each of which has its own spectrum. Put together, molecular spectra tend to be so rich that *spectral surveys* are undertaken to search for and identify the lines in molecular clouds. Figure 9.2 shows a dramatic example. The Orion KL[5] nebula contains a massive, young stellar cluster as well as abundant dust and molecular gas. The molecular gas reaches densities of $n \approx 10^7$ cm^{-3} and temperatures up to T $\approx$ 200 K in the core of the cluster (Ref. [98]). These conditions differ from those of typical interstellar molecular clouds that are far from star-forming regions ($n \approx 10^3$ cm^{-3}, T $\approx$ 20 K, see Table 3.1). Thus, the excitation conditions are such that more lines are seen in the spectrum of Orion KL than would be the case in colder ISM clouds. Figure 9.2 shows only a tiny fraction of the lines that are actually present in this nebula. Another survey of Orion KL over the frequency range, 790 GHz (0.38 mm) to 900 GHz (0.033 mm), for example, reveals approximately 1000 spectral lines of which about 90 per cent have been identified (Ref. [43]).

How many different molecules have actually been detected in astronomical objects? Table 9.1, which provides a list to date, indicates that more than 130 molecules have so

[4]The minimum energy state is not exactly at the minimum of the curve. Even the ground state will have a slight vibration associated with it.

[5]'KL' stands for 'Kleinmann–Low'.

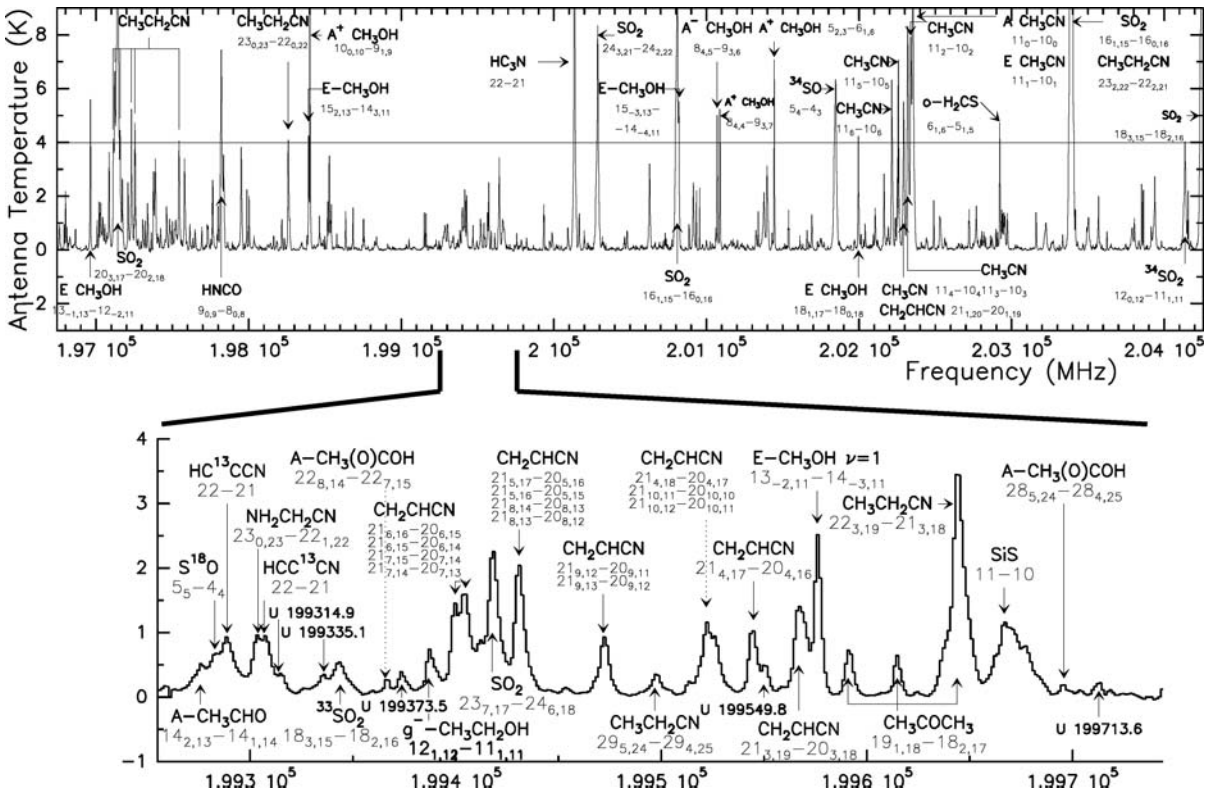


Figure 9.2. Spectrum of the molecular cloud, Orion KL, at a distance of 450 pc, over a small frequency range, 197 GHz (1.52 mm) to 204 GHz (1.47 mm) (top), obtained using the 30 m diameter Institut de Radio Astronomie Millimétrique (IRAM) telescope. The continuum has already been subtracted. Many more lines outside of this frequency range are also present in this source. A blow-up of a fraction of this frequency range (bottom) further reveals the richness of the spectrum. The lines have been identified by comparison with a catalogue of known spectral lines of over 1250 species. 'U' stands for 'unidentified'. See Refs. [169] and [170]. (Reproduced by permission of Asimolar, Tercero, B., Cernicharo, J., and Pardo, J. R., 2006, in preparation)

Table 9.1. Molecules detected in astronomical sources[a]

2	3	4	5	6	7
H_2	C_3*	c-C_3H	C_5*	C_5H	C_6H
AlF	C_2H	l-C_3H	C_4H	l-H_2C_4	CH_2CHCN
AlCl	C_2O	C_3N	C_4Si	C_2H_4*	CH_3C_2H
C_2**	C_2S	C_3O	l-C_3H_2	CH_3CN	HC_5N
CH	CH_2	C_3S	c-C_3H_2	CH_3NC	CH_3CHO
CH^+	HCN	C_2H_2*	H_2CCN	CH_3OH	CH_3NH_2
CN	HCO	NH_3	CH_4*	CH_3SH	c-C_2H_4O
CO	HCO^+	HCCN	HC_3N	HC_3NH^+	H_2CCHOH
CO^+	HCS^+	$HCNH^+$	HC_2NC	HC_2CHO	
CP	HOC^+	HNCO	HCOOH	NH_2CHO	
SiC	H_2O	HNCS	H_2CNH	C_5N	
HCl	H_2S	$HOCO^+$	H_2C_2O	l-HC_4H*(?)	
KCl	HNC	H_2CO	H_2NCN	l-HC_4N	
NH	HNO	H_2CN	HNC_3	c-H_2C_3O	
NO	MgCN	H_2CS	SiH_4*	H_2CCNH(?)	
NS	MgNC	H_3O^+	H_2COH^+		
NaCl	N_2H^+	c-SiC_3			
OH	N_2O	CH_3*			
PN	NaCN				
SO	OCS				
SO^+	SO_2				
SiN	c-SiC_2				
SiO	CO_2*				
SiS	NH_2				
CS	H_3^+*				
HF	H_2D^+, HD_2^+				
SH*	SiCN				
HD	AlNC				
FeO(?)	SiNC				
O_2(?)					
CF^+					
SiH(?)					

8	9	10	11	12	13
CH_3C_3N	CH_3C_4H	CH_3C_5N	HC_9N	C_6H_6*(?)	$HC_{11}N$
$HCOOCH_3$	CH_3CH_2CN	$(CH_3)_2CO$	CH_3C_6H	$C_2H_5OCH_3$(?)	
CH_3COOH	$(CH_3)_2O$	$(CH_2OH)_2$			
C_7H	CH_3CH_2OH	CH_3CH_2CHO			
H_2C_6	HC_7N				
CH_2OHCHO	C_8H				
l-HC_6H*(?)	$CH_3C(O)NH_2$				
CH_2CHCHO(?)					
CH_2CCHCN					

[a]See the table of S. Thorwirth from data provided by S. Martin at http://www.ph1.uni-koeln.de/vorhersagen/molecules/main_molecules.html. Numbers at the tops of the columns specify the number of atoms. A * specifies molecules that have been detected by their ro-vibrational spectrum. A ** specifies molecules that have been detected by electronic spectroscopy only. A (?) indicates a probable detection.

far been identified. The largest in this table is $HC_{11}N$, containing 13 atoms. However, as shown in Figure 3.24, if the nature of polycyclic aromatic hydrocarbons (PAHs, Sect. 3.5.2) is confirmed, the true number is higher and so is the number of atoms contained within the molecules.

9.1.3 Nuclear transitions – γ-rays and high energy events

Probing still deeper into the heart of an atom's structure brings us to the nucleus within which further quantization occurs. If a nucleus finds itself in an excited state as a result of a *radioactive decay* or high energy collision, then it may de-excite with the emission of a γ-ray photon at a specific energy. Various decay chains and daughter products are possible, depending on the specific nucleus and its stability. A mathematical description of nuclei and their energy levels is more difficult than has been discussed for the electronic, vibrational and rotational transitions. Nevertheless, many nuclear transitions are well known. Table 9.2 lists some that are important in astrophysics, indicating the relevant decay chains and energies of the emitted γ-rays. Those that have been detected are noted in bold-face type. Different isotopes of the same species produce lines at different energies, so it is possible to distinguish, not only between the element that is producing the line, but also which isotope is involved.

Gamma-rays have the highest energies known for light (Table G.6) so the original energy required to excite the nucleus must also have been of the same order. Thus, γ-ray emission is associated with high energy events. Examples are solar flares, novae and supernovae (Sect. 3.3.3), and spallation of nuclei (Sects. 3.3.4, 3.6.1) by the high energy cosmic rays that permeate the Galaxy. Massive stellar winds from Wolf–Rayet stars (Sect. 3.6.1 and Figure 5.10) can also deposit nuclei that have been created via nucleosynthesis in their stellar interiors into the ISM. Energies from nuclear transitions

Table 9.2. Some important nuclear γ-ray Lines[a]

Isotope	Mean lifetime[b]	Decay chain[c]	Energy[d](keV)
^{7}Be	77 days	$^{7}\mathrm{Be} \rightarrow {}^{7}\mathrm{Li}$	478
^{56}Ni	111 days[e]	$^{56}\mathrm{Ni} \rightarrow {}^{56}\mathrm{Co} \rightarrow {}^{56}\mathrm{Fe} + e^{+}$	**847, 1238**
			2598, 1771
^{57}Co	1.1 years	$^{57}\mathrm{Co} \rightarrow {}^{57}\mathrm{Fe}$	**122, 136**
^{22}Na	3.8 years	$^{22}\mathrm{Na} \rightarrow {}^{22}\mathrm{Ne} + e^{+}$	1275
^{44}Ti	85 years	$^{44}\mathrm{Ti} \rightarrow {}^{44}\mathrm{Sc} \rightarrow {}^{44}\mathrm{Ca} + e^{+}$	**1157, 78, 68**
^{26}Al	1.04×10^{6} years	$^{26}\mathrm{Al} \rightarrow {}^{26}\mathrm{Mg} + e^{+}$	**1809**
^{60}Fe	2.2×10^{6} years	$^{60}\mathrm{Fe} \rightarrow {}^{60}\mathrm{Co}$	**1173, 1332**

[a] Ref. [48].
[b] This 'lifetime', τ, is related to the half-life, $t_{1/2}$, by $t_{1/2} = \tau \ln 2$. In the event of two decays, the longer time is given.
[c] Not all possible decay routes are shown – just those that are relevant in the production of the γ-rays indicated. Decays that result in the emission of a positron (e^{+}) are also indicated.
[d] Values in bold-face have been detected.
[e] The $^{56}\mathrm{Ni} \rightarrow {}^{56}\mathrm{Co}$ part of this chain has a lifetime of only 8.8 days.

are typically in the MeV range which, although considered high energy, are at the lower end of the full γ-ray spectrum observed in nature.

Supernovae are a particularly important source of γ-ray lines since the nucleosynthesis that occurs during a supernova explosion leaves many nuclei in excited and unstable states. The γ-rays that result from de-exciting nuclei have been shown to drive supernova light curves which, like radioactivity, decay exponentially with time (Sect. 3.3.3). For example, the bolometric (Sect. 1.1) light curve of Supernova 1987A[6], which occurred in the nearby galaxy, the Large Magellanic Cloud, decayed with a half-life of 77 days, nicely corresponding to the known 77 day half-life (111 day lifetime[7]) of the decay of ^{56}Co to ^{56}Fe. The γ-ray photons that are emitted during this decay interact with the surrounding gaseous material via Compton scattering (Sect. 5.1.2.2) and, with further interactions, the energy is converted to the observed bolometric luminosity[8]. SN1987A has now been observed for 20 years and the rate of decline of its light curve has changed as the quantity of ^{56}Co has declined and other isotopes with longer half-lives have become more important sources of γ-ray photons. The current light curve is well-matched to the decay rate of ^{44}Ti, although a direct detection of the associated γ-ray lines has yet to be made (Ref. [48]).

Looking at the Milky Way, we can also see γ-ray line emission from Galactic supernovae. In the young ($\approx$ 312 yr old) nearby ($d = 3.4$ kpc) supernova remnant (SNR), Cas A (Figure 1.2), the 1157 keV γ-ray line from ^{44}Ti has indeed been detected (Ref. [78]), even though this isotope has a half-life of only 59 yr (lifetime of 85 yr, Table 9.2). From the line flux, it has been possible to measure the total mass of ^{44}Ti ($\approx 2 \times 10^{-4}$ $M_\odot$) produced in the supernova explosion. Such studies not only contribute to our knowledge of elemental abundances (see Figure 3.9) but also provide a better understanding of the processes and conditions that occur during the SN explosion itself. Spontaneous nuclear de-excitations, themselves, are not sensitive to the temperature and density of the surrounding material (they depend on nuclear properties), but the rate at which nucleosynthesis can form these nuclei will be sensitive to physical conditions.

The ^{44}Ti γ-ray lifetime is not that different from the expected mean time between supernova explosions in our Milky Way (about one every 50 yr, Sect. 3.3.4). If this line is detected and originates in supernovae, then the sources from which the line is observed should appear discrete. On the other hand, if a γ-ray line from a longer lived isotope of supernova origin is observed, then its distribution should describe the distribution of Galactic supernovae *integrated* over a timescale corresponding to the lifetime of the line. This is the case for the 1.809 MeV line from ^{26}Al which has a lifetime of a million years (Table 9.2). Over this period of time, with a mean supernova rate of one per 50 yr, we might expect to see the integrated γ-ray luminosity from approximately 20 000 supernovae (SNe). These SNe should originate from different

[6]This supernova, discovered in 1987 from Chile by Ian Shelton of the University of Toronto, Canada, has played a pivotal role in testing supernova models.

[7]The 'lifetime', τ, is related to the half-life, $t_{1/2}$, by $t_{1/2} = \tau \ln 2$.

[8]This process is reminiscent of the diffusion and energy degradation of photons in the Sun as they travel from the core to the surface (see Sect. 3.4.4).

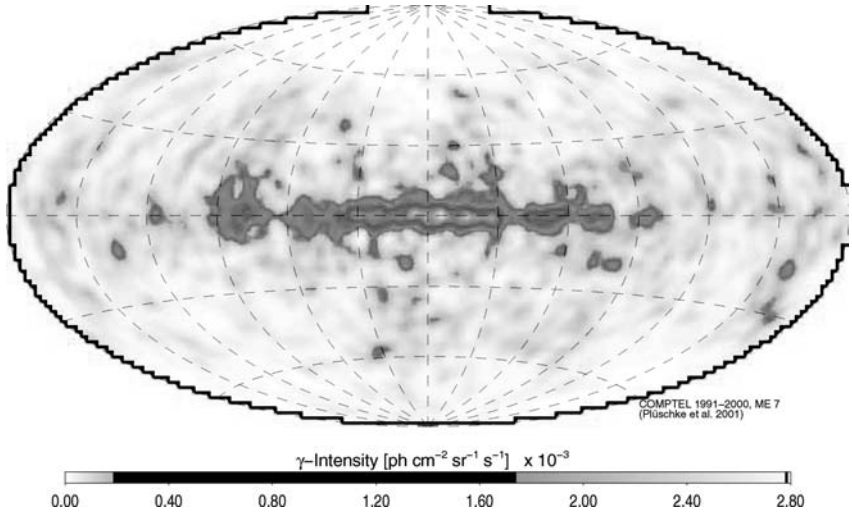


Figure 9.3. Map of the ^{26}Al 1.809 MeV γ-ray line emission in the Galaxy shown in a Hammer–Aitoff projection (e.g. see Figure 3.2 or Figure 8.5) with the Galactic Centre at the map centre. This map (see Ref. [125]) represents 9 yr of data collected by NASA's Compton Gamma Ray Observatory using the imaging Compton Telescope (COMPTEL). (Reproduced by permission of R. Diehl, Plueschke *et al.*, 2001, in *Exploring the Gamma-Ray Universe* (4th Integral Workshop), Eds. A. Gimenez, V. Reglero and C. Winkler, ESA SP-459, Noordwijk, 56–58)

locations in the Galaxy, leading to distributed, rather than discrete, emission. Moreover, since the lifetimes of the massive progenitor stars are also of order, 10^6 yr, the observed ^{26}Al emission will only be seen in locations where massive star formation has recently occurred. This line is therefore a tracer of massive star formation. We have encountered such tracers before (e.g. H II regions, Sect. 3.3.4), but compared to optical observations, γ-rays have an advantage in that they are highly penetrating and can travel through the ISM virtually unhindered.

A plot of the ^{26}Al distribution in the Milky Way is shown in Figure 9.3. Not all the emission in this figure results from supernovae alone. Between 20 and 50 per cent may originate from Wolf–Rayet stars (Ref. [117]), but since Wolf–Rayet stars, themselves, are massive and short-lived, the bulk of the ^{26}Al emission still traces sites of massive star formation in the Galaxy. Doppler shifts of this line have also provided information on motions of the various sources (Sect. 7.2.1). Even the broad-scale rotation of the Galaxy has been measured from this line emission (Ref. [49]).

9.2 The line strengths, thermalization, and the critical gas density

In Chapter 8, we provided the emission coefficient, j_ν, for continuum radiation and then found the specific intensity, I_ν, by applying the relevant solution to the Equation of

Transfer. Other quantities, such as flux density or luminosity, followed. The approach for spectral lines is essentially the same, but with the additional constraint that the line is present over a well-defined frequency interval. In order to characterize some *line strength*, then, we could either refer to the peak value of the line, or we could refer to the integral over the line. Both of these quantities are physically meaningful but, since the integral includes the total number of particles that are contributing to the line, we will use the integral to specify the 'line strength'. The line strength is determined by the *downwards transition rate* of whatever mechanism is producing the line, as well as the *number of particles that are in the upper level*.

Downwards transitions can occur spontaneously with a rate given by the Einstein A coefficient for a given transition (e.g. Table C.2), or can be induced via a collision, whose rate depends on the collision cross-section for that particlar transition as well as various physical parameters such as the density and temperature of the gas (see Eqs. 3.19 and 3.7)[9]. Example 5.1 provided a comparison of these two rates for the Ly α line of hydrogen under one set of conditions, showing the tendency of this line to undergo a spontaneous de-excitation, given the atom's short natural lifetime in the $n = 2$ state. On the other hand, the λ 21 cm line of H I is collisionally induced, since it has a longer spontaneous de-excitation timescale in comparison to a typical collision time. Since there is a different Einstein A coefficient and a different collision cross-section for each line, the probability of a downwards transition must be considered, spectral line by spectral line, for any given species under a set of physical conditions.

As discussed in Sects. 3.4.4 and 3.4.5, if *all* transitions of an atom are collisionally-induced, then the gas is in LTE. We now allow for the possibility that some transitions may be collisionally-induced and some may not – a non-LTE situation. However, if a *specific* transition is collisionally-induced, then it can be stated that *the line is in LTE* (see also arguments in Sect. 8.2.1). In such a case, the Boltzmann Equation (Eq. 3.23) holds for the individual line being considered. The line is then said to be *thermalized* and the LTE solution of the Equation of Transfer (Eq. 6.19) applies for the line.

There is some importance in knowing whether a line results from a collisional or a spontaneous downwards transition because, as outlined in Sect. 3.4.4, only if collisions are somehow involved can the observed line radiation tell us something about the gas kinetic temperature. The gas density at which the downwards collisional transition rate equals the downwards spontaneous rate for a given line is called the *critical gas density*[10], n^*. These rates depend on temperature as well, but more weakly than density. Some sample values are given in Table 9.3.

From Table 9.3, the Ly α line, a recombination line which occurs in an ionized environment, will never be collisionally de-excited under typical interstellar conditions since the density of interstellar material (see Table 3.1) does not achieve such high

[9] Downwards transitions can also be radiatively induced (called stimulated emission), but typically only if the radiation field is strong. More specifically, a radiatively-induced downwards transition will occur if the timescale for this process is shorter than the timescales for both collisionally-induced and spontaneous downwards transitions. Although this situation is not as common in astrophysics, stimulated emission must be included under some circumstances. See also Footnote 5 in Chapter 6.

[10] The term, *critical density* is more commonly used, but this term also has another meaning in cosmology, so we do not employ it here.

Table 9.3. Sample critical gas densities[a]

Species	Wavelength or transition	Colliding species	Temperature (K)	$n^*(cm^{-3})$
H I[b]	Ly α ($2p \rightarrow 1s$)	electron	10^4	9.2×10^{16}
H I[c]	21 cm	H I	100	3×10^{-5}
OIII[d]	493.26 nm	electron	10^4	7.1×10^5
CO[e]	2.6 mm ($J = 1 \rightarrow 0$)	H_2	30	2×10^3

[a]The critical gas density of the colliding species for the transition listed is determined by setting $\gamma_{j,i}\, n^* = A_{j,i}$, where $\gamma_{j,i}$ is the collisional rate coefficient ($cm^3\ s^{-1}$) for a transition from upper state, j, to lower state, i, and $A_{j,i}$ (s^{-1}) is the Einstein A coefficient for that transition.
[b]Using γ from Table 3.2 and A_{ji} from Table C.1.
[c]Using $\gamma = 9.5 \times 10^{-11}$ (Ref. [160]) and A_{ji} from Table C.1.
[d]Using data from www.astronomy.ohio-state.edu/~pradham/atomic.html with Eq. 4–11 from Ref. [160] and A from Ref. [160].
[e]From data in Ref. [160]. J is the total angular momentum quantum number.

values. This does not mean that collisions are never important in an ionized hydrogen gas, however, since other recombination lines, especially at high quantum numbers, may indeed be collisionally-induced at lower densities, as we shall see in Sect. 9.4.1. Other kinds of transitions in hydrogen can also be collisionally induced. For example, a transition between two n = 2 levels ($2s \rightarrow 2p$) becomes collisionally induced at a density of 1.5×10^4 cm^{-3} (Ref. [115]). This is higher than the density of most interstellar ionized gas but can occur around active galactic nuclei. As for interstellar H I clouds, these have typical densities $n_H > 0.1$ cm^{-3}, so the H I λ 21 cm line, with a critical gas density of 10^{-5} cm^{-3}, should always be thermalized. This is also the case for the listed CO line, since most molecular clouds have densities, $n_{H_2} > 10^3$ cm^{-3}.

Given the large range of possible critical densities and the number of spectral lines that are observationally accessible (Sect. 9.1), it is sometimes possible to *choose* the spectral line or lines that best probe the conditions of interest. A well-designed observational program will consider these points before time is spent at the telescope. We will explore how spectral lines can probe the physical conditions of the gas in Sects. 9.4 and 9.5.

9.3 Line broadening

In Appendix D.1.3, we indicated that a spectral line cannot be arbitrarily narrow. It has a *natural line width* which is dictated by the time the particle spends in the upper energy level before spontaneously de-exciting. The line shape, in the absence of other effects, is *Lorentzian*, with its shape given by the *line shape function*, also called the *line profile*, $\Phi_{\mathcal{L}}(v)$ (Eq. D.14). Since all transitions have this quantum mechanical limitation, all spectral lines will have a natural line width.

In real systems, however, the measured line width is always much greater than the natural line width. This means that one or more additional line broadening mechanisms must also be present and be sufficiently important to dominate the quantum mechanical effects. Various line broadening mechanisms are known. In astrophysics, however, the

most important ones are *Doppler broadening* and *pressure broadening*, of which the former is the most common and most easily linked to the physical state of the gas. In the next two sub-sections, therefore, we describe Doppler broadening in some detail and then discuss pressure broadening and related effects more qualitatively.

9.3.1 Doppler broadening and temperature diagnostics

Doppler broadening is due to the collective motions of the particles that are emitting the spectral line (recall that $\Delta\lambda/\lambda_0 = v/c$, where v is a radial velocity, Eq. 7.3). Even if a gas cloud (meaning any gaseous system that is emitting the line) has no net motion towards or away from the observer ($v_{\rm sys} = 0$, see Sect. 7.2.1.2) the individual particles within the cloud will still have some motion, be it thermal, turbulent, or perhaps from internal systematic motions such as rotation, expansion or contraction or motions related to shock waves. If each particle within the cloud has a different motion, then the various radial velocity components of these motions will produce different Doppler shifts. The net result will be a superposition of all the individual Doppler shifted lines, resulting in a broadened line. If $v_{\rm sys} = 0$, then the widened line will be centred on its rest frequency, and if $v_{\rm sys} \neq 0$, then the line is centred at the Doppler-shifted frequency of the line for that $v_{\rm sys}$ (Eq. (7.4), or Eq. (7.3) if plotted by wavelength). We have already seen an example of this for a rotating galaxy (see Figure 7.4.d). When emission from the entire galaxy is detected within a single beam, the galaxy's rotation causes the line profile to be spread out over a wide frequency range, or a wide velocity range if the x axis is expressed in velocity units. Also, a rapidly expanding supernova remmant, if unresolved, should have a line profile whose width reflects the maximum velocities of both the advancing and receding sides of the shell. The *expansion velocity* is then half of the line width (Prob. 9.4).

Not every cloud may be expanding or contracting or have other peculiar motions associated with it, but spectral lines that are formed in a gas at some temperature, T, will always experience at least *thermal line broadening* due to the Doppler shifts of particles in a Maxwellian velocity distribution (Eq. (0.A.2), Figure 3.12). Taking a simple case in which the cloud has no systemic velocity, an optically thin spectral line (see below) from such a cloud has a *Gaussian* shape centered on the rest frequency, $\nu_{\rm i,j}$, as illustrated in Figure 9.4b (lower curve).

The Gaussian line shape function ($\rm Hz^{-1}$), or profile, is described by,

$$\Phi_{\mathcal{D}}(\nu) = \frac{1}{\sqrt{2\,\pi}\,\sigma_{\mathcal{D}}}\; exp\left(-\frac{(\nu-\nu_{i,j})^2}{2\,{\sigma_{\mathcal{D}}}^2}\right) \tag{9.1}$$

where,

$$\sigma_{\mathcal{D}} = \frac{1}{\sqrt{2}}\frac{\nu_{i,j}}{c}\left(\frac{2\,k\,\mathrm{T}}{\mathrm{m}}\right)^{1/2} = \frac{1}{\sqrt{2}}\frac{\nu_{i,j}}{c}\,b = \frac{1}{\sqrt{2}\,\lambda_{i,j}}\,b \tag{9.2}$$

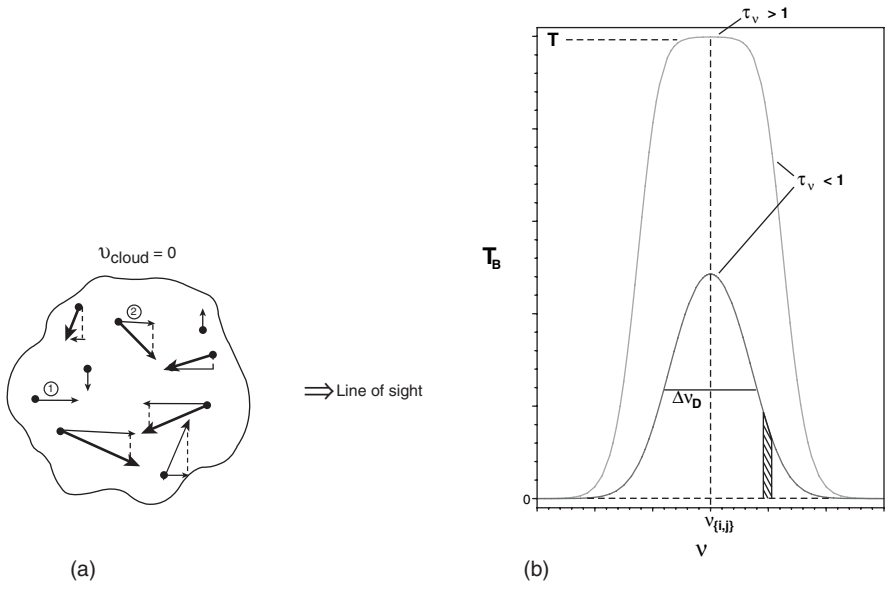


Figure 9.4. The thermal motions in a gas cloud that is globally at rest form a Gaussian-shaped spectral line, if the line is optically thin. **(a)** Sketch of a gas cloud at rest, showing some of the internal particles (dots) and their velocity vectors (heavy arrows). The thin arrows show the components of velocity along the line of sight that broaden the spectral line. The two labelled particles have identical radial velocities and so will contribute to the line intensity at the same frequency. **(b)** Two spectral lines plotted on a brightness temperature scale. The lower dark curve shows the Gaussian line shape given by Eq. (9.1) with the FWHM labelled. The hatched area (with width exaggerated) corresponds to all particles that have equal radial velocities, like those of particles 1 and 2 in (a). The grey higher curve shows the same line, but for a higher density cloud which has resulted in optically thick emission. In that case, $T_B = T$ near the line centre.

T is the gas kinetic temperature (K), m is the mass (g) of the particle emitting the spectral line, b (cm s^{-1}) is the *velocity parameter* which is defined by Eq. (9.2), i.e. $b = (2\,k\mathrm{T}/\mathrm{m})^{1/2}$, and we have used the relation, $\lambda\,\nu = c$.

The full width at half maximum (FWHM) intensity, $\Delta\,v_{\mathcal{D}}$, is related to $\sigma_{\mathcal{D}}$ and is given (in units of Hz) by,

$$\Delta\,v_{\mathcal{D}} = 2.3556\,\sigma_{\mathcal{D}} = 7.14 \times 10^{-7}\,v_{i,j}\left(\frac{\mathrm{T}}{\mathrm{A}}\right)^{1/2} = \frac{2.14 \times 10^{4}}{\lambda_{i,j}}\left(\frac{\mathrm{T}}{\mathrm{A}}\right)^{1/2} \qquad (9.3)$$

where we have used Eq. (9.2), set $\mathrm{m} = \mathrm{A}\,\mathrm{m_H}$, with $\mathrm{m_H}$ the mass (g) of the hydrogen atom, A (unitless) is the atomic weight, and the constants have been evaluated. The value of $\sigma_{\mathcal{D}}$, if it is desired, can also be obtained from Eq. (9.3). Note that higher

frequency lines have broader widths in frequency, which reflects the fact that $\Delta v = v_0\, v/c$ from the Doppler shift. The peak of the line is,

$$\Phi_{\mathcal{D}}(v_{\rm i,j}) = \frac{0.9394}{\Delta v_{\mathcal{D}}} \tag{9.4}$$

and the integral over all frequencies (a unitless quantity, like Eq. D.18) is,

$$\int_{-\infty}^{\infty} \Phi_{\mathcal{D}}(v)\, dv = 1 \tag{9.5}$$

Eq. (9.3) provides a relation between the line width in frequency, $\Delta v_{\mathcal{D}}$, which is a straightforward quantity to measure, and the temperature of the gas, assuming that thermal broadening is the dominant line broadening mechanism. It is sometimes more useful to have a measure of the line width in velocity units, however, especially since many spectra are plotted as a function of velocity, rather than frequency or wavelength (see also Prob. 9.3). Upon applying the Doppler formula to the right hand side of Eq. (9.3), the velocity line width is (cm s^{-1}),

$$\Delta v_{\mathcal{D}} = 2.14 \times 10^4 \left(\frac{\rm T}{\rm A}\right)^{1/2} \tag{9.6}$$

Eq. (9.6) is now independent of the specific spectral line being observed. This must be the case, since any given atom, which could have many transitions at different frequencies, will move at only one velocity.

The thermal line width is larger than the natural line width (e.g. Prob. 9.3) and is always present in a thermal gas. However, there may be other Doppler motions (e.g. from rotation, turbulence, etc. as indicated above) which could widen the line even more, and these often dominate the thermal motions. If turbulence is present and the turbulent motions are also Gaussian, then it can be shown that the resulting profile, a convolution (Appendix A.4) of the two functions, will still be Gaussian, but wider. If other motions are present, the profile may no longer be Gaussian, as we saw for Figure 7.4.d. Thus, for Doppler broadening, the shape and width of the line profile provides information as to the internal velocities or temperature of the cloud.

Allowing for the possibility of additional internal motions, then, the *observed* FWHM of the line, $\Delta v_{\rm FWHM}$ or $\Delta v_{\rm FWHM}$, will be greater than or equal to the thermal line width,

$$\Delta v_{\rm FWHM} \geq \Delta v_{\mathcal{D}}, \qquad \Delta v_{\rm FWHM} \geq \Delta v_{\mathcal{D}} \tag{9.7}$$

Together with Eqs. (9.3) or (9.6), this shows that the measured line width places an *upper limit on the temperature of the gas*. This is a useful result because it means that we need only plot the spectral line and measure its width, without doing any other

calculations, in order to determine an upper limit to T. In Sect. 6.4.1, moreover, we noted that the peak of a spectral line in LTE provided a *lower limit* to the gas temperature. Thus, provided these conditions are met, the gas temperature range can be immediately constrained by simply plotting the spectrum and taking two relatively straightforward measurements, as Example 9.1 describes.

Example 9.1

The λ 21 cm line of hydrogen emitted from an interstellar cloud has a peak specific intensity of $I_\nu = 4.03 \times 10^{-17}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} and a line width of $\Delta\, \nu_{\rm FWHM} = 15.0$ kHz. Find the range of possible temperatures of the cloud.

The λ 21 cm line is in LTE (Sect. 9.2, Sect. 6.4.3) and therefore, by Eq. (6.24), $T_B \leq T$. Since $h\,\nu \ll k\,T$ for this line (Sect. 6.4.3), we can use the Rayleigh–Jeans formula (Eq. 4.6) to convert the specific intensity, I_ν, at the frequency of the line centre ($\nu_{i,j} = 1420.4$ MHz) to a brightness temperature, finding $T_B = 65$ K for the line peak. Then using the given value of $\Delta\, \nu_{\rm FWHM}$ in Eq. (9.3) with $\lambda_{i,j} = 21.106$ cm (Table C.1) and $A = 1$, we find an upper limit to the temperature of $T = 219$ K. We conclude that the temperature of this cloud is in the range, $65 \leq T \leq 219$ K.

It is worth noting that the thermal line width applies to any transition in the gas, no matter how it is produced (e.g. collisionally, radiatively, or spontaneously) because it is the velocity distribution of the particles that is determining this width (not the physics of the spectral line). Thus, the upper limit to T applies to any optically thin line. On the other hand, the condition, $T_B \leq T$ will only be true for a line in LTE because the peak of the line must be related to the motions of particles in the gas for this condition to have any meaning.

We have so far considered only optically thin lines. It is also of interest to consider what a line in LTE will look like if there is a large density of particles that contribute to the line. For a specific transition in a given cloud, the optical depth equation (Eq. 5.8) indicates that τ_ν increases linearly with increasing density. Therefore, at some high density, the line should become optically thick ($\tau_\nu > 1$). From the LTE solution to the Equation of Radiative Transfer (Eq. 6.19), assuming no background source, then $I_\nu = B_\nu(T)$ at frequencies for which $\tau_\nu > 1$. This also means that $T_B = T$ at these frequencies (see Sect. 4.1.1). Since more particles contribute to the line centre than its wings, the result is a flat-topped profile, as illustrated by the higher curve in Figure 9.4.b[11]. The presence of flat-topped profiles, therefore, is a diagnostic for optically thick thermalized lines. Provided we can be sure that no other effects are

[11] *Flat-bottomed* profiles also result from optically thick *absorption* lines. In such a case, all of the background continuum emission is removed by the foreground line emission at and near the central line frequency. A quantity called the *equivalent width*, W, is used to measure the strength of an absorption line for any τ_ν, i.e. $W \equiv \int [(I_{\nu 0} - I_\nu)/I_{\nu 0}]\, d\nu$, where $I_{\nu 0}$ is the specific intensity of the background continuum and I_ν is the specific intensity measured in the line. The integral is over the line.

producing unusual profiles[12] the peak brightness temperature of such a line will give the gas temperature.

9.3.2 Pressure broadening

The second important line broadening mechanism in astrophysics is *pressure broadening* which we consider here only qualitatively. Pressure broadening actually includes a number of possible effects which become important when densities are higher, such as in denser stellar atmospheres, and involve interactions between a radiating atom and other particles around it. The interacting particles essentially perturb the radiating atom in such a way as to cause a small shift in the frequency of an emitted spectral line. For an ensemble of particles, each shift may be different, resulting in a widened line.

There are a number of approaches that have been taken to understand this process, an important one being to treat the atom as a radiating harmonic oscillator, as is done in Appendix D.1.2. Collisions (impacts) with surrounding particles then effectively 'interrupt' the emitted radiation, introducing a perturbation on the phase and amplitude of the oscillation. The collision itself does not produce a transition, but perturbs the transition that is taking place. This interruption in time results in a wider frequency response (recall the reciprocal relation between time and frequency). The collision may be with protons, electrons, or neutral particles, and the result is different, depending on the type of impactor. This approach has been widely used and the line broadening is therefore called *collision broadening*. It can be shown that collision broadening, for each type of impactor, leads to a *Lorentzian* profile (see Eq. D.14), but with the quantum mechanical radiation damping constant, $\Gamma_{i,j}$, replaced by a *collisional damping constant*, $\Gamma_{Pi,j}$. The latter is a function of temperature, density, the specific line being considered, and the nature of the impacting particle, and therefore the Lorentzian FWHM, Δv_P, also depends on these quantities (see Eq. D.17).

Since we know that thermal line broadening, which has a Gaussian profile, is always present and its width is much greater than the natural line width, then the presence of an *observed* Lorentzian profile suggests that collision broadening is occurring and is dominating. This, then, provides information about the environment within which the line is forming. In reality, the various line widths are difficult to model and compare with observation, especially since collision coefficients are not known for every type of encounter and spectral line. However, some lines have been successfully reproduced. For example, the widths of certain strong sodium lines in the Solar atmosphere can be adequately explained by impacts with neutral hydrogen (Prob. 9.3). If *both* collision broadening and thermal Doppler broadening are important (one does not strongly dominate the other), then both the Gaussian and Lorentzian profiles are incorporated into the line profile. The resulting shape is called a *Voigt profile*.

[12] A spherically symmetric expanding wind which has a cut-off in radius, for example, can produce flat-topped profiles, even if optically thin.

An interaction may be considered an 'impact' if the duration of the collision is much less than the time between collisions. However, if the impact duration approaches the time between collisions, then the effect may be considered continuous. For example, perturbations can occur from the collective quasi-static effects of surrounding ions. These ions produce a net electric field which causes the energy levels of a radiating atom to split and /or shift. Since the energy levels of many particles are no longer at one fixed value but are spread out, the resulting spectral line will also acquire a width. This effect is called *Stark broadening* because the shifts in energy levels result from the application of an external electric field, an effect known as *the Stark effect*. (The analogy for a magnetic field is the Zeeman effect, as discussed in Appendix C.3.) The resulting line profile may no longer be Lorentzian[13].

For some of the above interactions, the different approaches could be considered different ways of describing similar effects. For example, an impact of a free electron with a bound electron can be thought of as an encounter in which the trajectory of the free electron changes significantly (see Footnote 22 in Chapter 3). As the free electron encounters the atom, it briefly applies Stark broadening to the atom. Thus, line broadening due to electron or proton collisions might also be referred to as Stark broadening. Similarly, the net electric field produced by surrounding ions can be thought of as the collective effects of many slow impacts. Because of this, as well as the fact that these mechanisms all tend to be important in high pressure environments, the terms pressure broadening, collisional broadening and Stark broadening are sometimes used without distinction in the literature, especially the first two. More information on these mechanisms can be found in Ref. [68].

9.4 Probing physical conditions via electronic transitions

In principle, it is straightforward to write down the emission coefficient of a spectral line that results from a spontaneous downwards transition from upper level, j, to lower level, i,

$$\mathrm{j}_\nu = \frac{n_\mathrm{j} A_{\mathrm{j,i}} h \nu}{4\pi} \Phi(\nu) \tag{9.8}$$

where n_j (cm^{-3}) is the number density of atoms having electrons in level, j, $A_{\mathrm{j,i}}$ (s^{-1}) is the Einstein A coefficient, h is Planck's constant, ν (Hz) is the frequency of the line, and $\Phi(\nu)$ (Hz^{-1}) is the line shape function describing the relevant line broadening mechanism (e.g. Eq. 9.1). A dimensional analysis of this equation will verify that the units are erg s^{-1} cm^{-3} Hz^{-1} sr^{-1}, as required for j_ν (Sect. 6.2).

In practise, however, it is often a challenge to arrive at a full description of the intensity of a spectral line. For example, the population of the upper state, n_j, must be

[13]Stark broadening can be of two types, called *linear* and *quadratic*, depending on the strength of the frequency shift with particle separation. The line profile shape is different for the two types.

found, and collisional and radiative de-excitations, as well as ionization equilibrium must also be included, if important. The full equations of statistical equilibrium may therefore be required (see Sect. 3.4.4) in order to compute emission coefficients for the various lines in the atom. To compute the observed spectral line intensities, I_ν, the emission coefficient must be used in the appropriate solution to the Equation of Radiative Transfer (Sect. 6.3) which requires that accurate optical depths be determined. The optical depths will vary from line to line and will also vary, according to some line profile shape, within each line. The type and extent of the line broadening mechanism is also important. For example, a line that would normally be optically thick could be optically thin if the cloud is undergoing large-scale turbulence or other motions that spread out the line and lower the line peak. This, in turn, affects the radiative transfer through the cloud. Finally, the spectral lines may sit atop an underlying continuum (e.g. Figure 8.9).

Fortunately, however, some conditions exist for which simplifications can be made without badly compromising the results. We will now look at some of these and consider what the observed spectral lines can tell us about the physical state of the gas. We will start with high quantum numbers and work 'down the quantum number ladder', focussing on three main transitions or types of transitions: *radio recombination lines*, *optical recombination lines*, and *the* λ 21 *cm line of H I*. Recombination lines will be seen in any ionized gas, including H II regions, planetary nebulae, and diffuse ionized gas in the interstellar medium. The λ 21 cm line will be seen in neutral H I clouds (Sect. 3.4.7). As before, we will focus mainly on the hydrogen atom, but there are many other constituents besides hydrogen that emit their own spectral lines, each providing further diagnostics of the physical state of the gas.

9.4.1 Radio recombination lines

At high quantum numbers (e.g. $n \gtrsim 40$), the energy levels of the hydrogen atom become very close together (see Figure C.1), so transitions between these levels result in radiation at radio wavelengths. These transitions represent the upper quantum number limit to the recombination line spectrum (designated Hnα for a transition from principal quantum number $n+1$ to n, Hnβ for $n+2$ to n, etc., see Appendix C.1) and are called *radio recombination lines* (RRLs). At high n, hydrogen transition probabilities are low, so RRLs tend to be weak in comparison to lower quantum number optical recombination lines (Sect. 9.4.2). Nevertheless, many such observations have been made, not only of hydrogen RRLs, but of many other species as well (see Figure 9.5). RRLs from quantum numbers as high as $n = 766$ have been measured from carbon, for example, corresponding to frequencies as low as 14 MHz. For hydrogen, RRLs have so far been observed at lower n but, in principle, it is believed that quantum numbers up to $n \approx 1600$ are possible[14] (Ref. [66]). Using Eq. (C.3), the

[14]Such high quantum number transitions are more likely to occur in gas that is very low density, cooler, and with little background radiation. See Ref. [66] for further information.

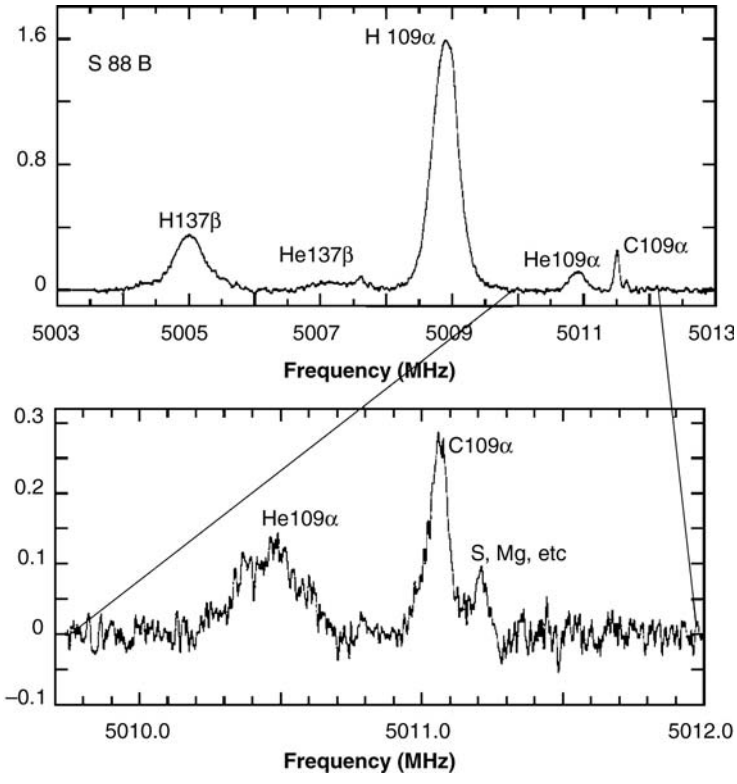


Figure 9.5. The radio recombination line spectrum of the H II region, Sharpless 88 B, near $\nu = 5$ GHz using the Arecibo telescope (see Figure 2.6a). The region that has been observed is circular on the sky of angular diameter, $\theta_b = 59$ arcsec. The continuum has already been subtracted. Notice the presence of various weaker recombination lines, besides the H109 α line of hydrogen. (Reproduced by permission of Yervant Terzian)

diameter of such an atom would be an astonishing 0.3 mm! This is larger than the thickness of a human hair, and is three million times larger than the size of the hydrogen atom in its ground state.

As previously indicated (see Sect. 8.2.2), radio waves are not hindered by dust and can therefore can be seen over great distances. If the optical emission of an H II region is obscured by extinction, radio (or long wavelength infrared) observations may be the only way of detecting that an H II or other ionized region is present at all (see Figure 9.6). Any ISM component that suffers from heavy extinction can be similarly probed at radio wavelengths. *Ultra-compact (UC) H II regions*, for example, represent a class of H II regions which are almost always heavily obscured, optically. These are small ($\lesssim 0.1$ pc), very dense ($\gtrsim 10^4$ cm^{-3}) H II regions around O or B stars that are still embedded within their natal dusty molecular clouds (recall Sect. 5.2.2, see also Ref. [42]). They therefore represent a relatively early stage in the development of massive star forming regions and long wavelength data are required to probe their characteristics. RRLs also provide an

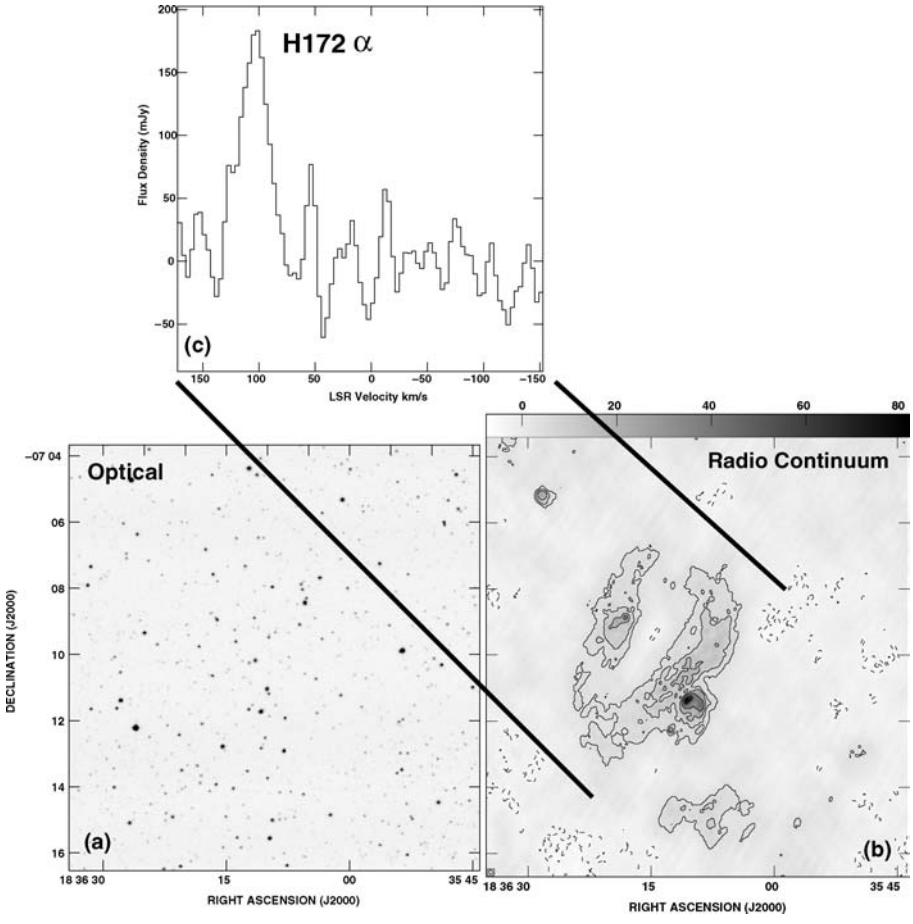


Figure 9.6. The massive star forming region, G24.8+0.1, located in the plane of the Milky Way (see Figure 3.3), is estimated to be at a distance of $D = 6.5$ kpc from us and would be completely invisible were it not for observations at long wavelengths. **(a)** This optical image shows only nearby stars since the optical emission from G24.8+0.1, itself, is heavily obscured by foreground dust. **(b)** This image, taken at $\nu = 1281.5$ MHz using the Giant Metrewave Radio Telescope (GMRT) in India, shows the same region as in (a). Here we see complex radio continuum emission containing a number of HII regions. Most of the emission is shell-like, but a reasonable assumption is that it occupies a volume which is equivalent to a sphere of diameter, 4′. The radio continuum flux coming from this volume is $f_{1282\,\mathrm{MHz}} = 7.7$ Jy. **(c)** The H172 α recombination line, at the same frequency, from the region shown in (b) is displayed in this image. The underlying continuum has already been substracted from this spectrum. LSR refers to Local standard of Rest (see Sect. 7.2.1.1). (GMRT images Reproduced by permission of N. Kantharia)

opportunity to probe the diffuse ionized gas (DIG, Sect. 8.2.2) over large distances in the Galaxy (e.g. Ref. [40]). Together with a model of galaxy rotation, RRL Doppler shifts can be used to obtain distances to the line-emitting gas (Sect. 7.2.1.2), something that cannot be achieved by continuum measurements alone.

For RRLs, several assumptions can usually be made that simplify the analysis of these lines. Firstly, RRLs are weak lines and are observed in the optically thin limit. Secondly, at high quantum numbers, the spontaneous downwards transition rate is a very strong function of quantum number, i.e. $A_{\mathrm{n+1}\rightarrow\mathrm{n}} \propto \mathrm{n}^{-6}$ (Ref. [50]), so the time-scale for spontaneous de-excitation becomes very large (e.g. see the H109α line value in Table C.1 for which $t \approx 10^3$ s). This means that RRLs tend toward LTE conditions (Sect. 9.2)[15]. And thirdly, at radio wavelengths, $h\,\nu \ll k\,\mathrm{T}$ so the Rayleigh–Jeans approximation to the Planck function (Eq. 4.6) can be used. Thus (as was the case for thermal Bremsstrahlung emission observed at radio wavelengths, see Sect. 8.2.2) we can use Eq. (6.33) for the line, $\mathrm{T}_{\mathrm{B}\nu} = \mathrm{T_e}\,\tau_\nu$ where, in this case, τ_ν will be a function of the line shape, among other parameters.

We now assume that the line shape is Gaussian (applicable to thermal broadening alone or thermal broadening with Gaussian turbulence, see Sect. 9.3.1) in order to write an expression for the optical depth at the centre of the line, τ_L. For transitions between quantum levels, n, and n + Δn, this is given by (Ref. [55]),

$$\tau_\mathrm{L} = 1.01 \times 10^4\, Z^2\, \Delta\mathrm{n} \left[\frac{f_{\mathrm{n,n+\Delta n}}}{\mathrm{n}}\right] \mathrm{T_e}^{-2.5}\, e\left(\frac{\chi_\mathrm{n}}{\mathrm{n}^2\, k\, \mathrm{T_e}}\right) \Delta\nu^{-1} \mathcal{EM}_\mathrm{L} \tag{9.9}$$

where Z is the atomic number, $f_{\mathrm{n,n+\Delta n}}$ is the oscillator strength for transition, n $\rightarrow$ n + Δn (see Appendix D.1.3), χ_n (erg) is the ionization potential (Sect. 5.2.2) of level n, $\Delta\nu$ (kHz) is the FWHM of the Gaussian line, and $\mathcal{EM}_\mathrm{L}$ (cm^{-6} pc) is the emission measure of the line given by Eq. (8.5), where it is assumed that the ion and electron densities are roughly equal. Taking $\Delta\mathrm{n} = 1$, $Z = 1$, $(f_{\mathrm{n,n+1}})/\mathrm{n} \approx 0.1908$ for large n when $\Delta\mathrm{n} = 1$ (Ref. [177]), and noting that the value of the exponential approaches 1 for large n, then Eq. (9.9) becomes, for the centre of a Hnα RRL,

$$\tau_\mathrm{L} = 1.92 \times 10^3 \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{-2.5} \left[\frac{\Delta\nu}{\mathrm{kHz}}\right]^{-1} \left[\frac{\mathcal{EM}_\mathrm{L}}{\mathrm{cm^{-6}\,pc}}\right] \tag{9.10}$$

The brightness temperature of the centre of an optically thin line can then be written (Eq. 6.33 and above),

$$\mathrm{T_L} = 1.92 \times 10^3 \left[\frac{\mathrm{T_e}}{\mathrm{K}}\right]^{-1.5} \left[\frac{\Delta\nu}{\mathrm{kHz}}\right]^{-1} \left[\frac{\mathcal{EM}_\mathrm{L}}{\mathrm{cm^{-6}\,pc}}\right] \tag{9.11}$$

As we have seen before, emission from an optically thin source is a function of *both* the temperature and density, the latter via the emission measure. Thus, an observation of a single RRL cannot lead to information on one of these quantities without knowing the other.

[15]Typically, if n > 168 n_e, where n_e is the electron density, LTE can be assumed (Ref. [50]). However, stimulated emission (see footnote 5 in Chapter 6) can also affect the line strength. See Ref. [137] for information as to when stimulated emission and non-LTE conditions may apply.

There is additional information associated with RRL observations, however, that aids our efforts in finding both the temperature and density of the gas. The same ionized gas that produces RRLs will also emit continuum radiation. At radio wavelengths, this continuum is due to thermal Bremsstrahlung (Sect. 8.2) for which the optical depth and brightness temperature were given by Eqs. (8.12) and (8.13), respectively. Thus, the RRL sits on top of continuum emission[16] and T_L is a value that is measured from the level of the continuum 'upwards'. Usually, the continuum emission is first subtracted from the total line+continuum before the measurement is made (e.g. Figure 9.5). If the continuum emission is also optically thin[17], then the line-to-continuum brightness temperature ratio at the line centre, r, can be formed by dividing Eq. (9.11) by Eq. (8.13) or, equivalently, dividing Eq. (9.10) by Eq. (8.12). The result is (Ref. [138]),

$$r = \frac{T_L}{T_C} = \frac{\tau_L}{\tau_C} = 2.33 \times 10^4 \left[\frac{T_e}{K}\right]^{-1.15} \left[\frac{\Delta v}{kHz}\right]^{-1} \left[\frac{v}{GHz}\right]^{2.1} \tag{9.12}$$

where the subscripts, L and C, refer to the line and continuum, respectively.

In practise, we cannot measure the value of the continuum at exactly the line centre (e.g. Figure 6.3b), but we can measure it just adjacent to the line, as was done for the λ 21 cm line in Sect. 6.4.3. In Eq. (9.12), we have assumed that $\mathcal{EM}_L/\mathcal{EM}_C = 1$ which would be the case for a pure hydrogen ionized gas[18]. Thus, the line-to-continuum ratio can be used to find the electron temperature of the ionized gas explicitly. This is a very useful result because, if the assumptions are correct, a determination of T_e requires only a measurement of a single line together with its continuum, and the result is *independent of distance*. The electron density can then be found from either the line or continuum measurement, provided the distance is known, as Example 9.2 outlines.

Example 9.2

The H93 α recombination line has been measured (Ref. [2]*) within a* $5'' \times 5''$ *square region centred on the ultra-compact H II region, G45.12 (D = 6.4 kpc). The observing frequency is* $\nu = 8.044$ *GHz, the line FWHM is* $\Delta v = 1497$ *kHz, and the flux density in the continuum and line centre are* $f_C = 1.72$ *Jy and* $f_L = 0.059$ *Jy, respectively. Assume that the line and continuum, both optically thin, are formed in the same physical region and find* T_e, τ_L, τ_C, $\mathcal{EM}$, *and* n_e *of this region.*

[16]The condition for emission lines which requires that the line-emitting gas be hotter than the background continuum (see Sect. 6.4.2) does not apply in this case because there is no 'background'; both the line and continuum come from the same 'pocket' of gas.

[17]At very low frequencies, corresponding to very high quantum numbers, the line will be optically thin but the continuum may be optically thick (see Eq. 8.12).

[18]As noted in Footnote 7 of Chapter 8, singly-ionized helium will contribute free electrons which will increase the continuum emission, but not the line emission, as derived. To account for ionized helium, the right hand side of Eq. (9.12) should by multiplied by the factor, $1/[1 + N(\text{HeII})/N(\text{H II})]$.

From the given values, $f_L/f_C = 0.059/1.72 = 0.0343$, but using Eqs. (1.13), (4.4), and (4.6), $f_L/f_C = T_L/T_C$, where T_L and T_C are the brightness temperatures of the line and continuum, respectively. We called this ratio, r, in Eq. (9.12), so $r = 0.0343$. Using the given values of Δv and v, Eq. (9.12) allows us to find the electron temperature, resulting in, $T_e = 9198$ K.

The angular size of the observed $5'' \times 5''$ square region is $\Omega = 5.876 \times 10^{-10}$ sr, upon converting units. At a distance of $D = 6.4$ kpc, the angular diameter corresponds to (Eq. B.1) $l = 0.155$ pc which we also take to be the line of sight distance through the source. Thus, the geometry is that of a 'box' for this example. The optical depths, τ_L and τ_C, can be found from $T_{B_v} = T_e\,\tau_v$ (Eq. 6.33) which requires that individual brightness temperatures be computed (only the ratio is known, from above). Converting the flux densities to specific intensities (Eqs. 1.13 and 1.8), yields $I_L = 1.00 \times 10^{-15}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} and $I_C = 2.93 \times 10^{-14}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} for the line and continuum, respectively. Converting from specific intensities to brightness temperatures (Eqs. 4.4 and 4.6) gives $T_L = 50.3$ K and $T_C = 1475$ K for the line and continuum, respectively. With the known value of T_e, this gives $\tau_L = 0.0055$ and $\tau_C = 0.16$.

To compute the $\mathcal{EM}$, either the line (Eqs. 9.10 or 9.11) or continuum emission (Eqs. 8.12 or 8.13) can be used[19], both of which lead to $\mathcal{EM} = 34.7 \times 10^6$ cm^{-6} pc. We now use Eq. (8.5) with the value of l from above to obtain $n_e = 1.5 \times 10^4$ cm^{-3}.

Note that the calculated values are averages that are representative of the region as a whole. The density, for example, likely varies considerably within the region. A more accurate calculation for this source, which includes non-LTE effects, leads to $T_e = 8300$ K and $n_e = 1.8 \times 10^4$ cm^{-3} (Ref. [2]), i.e. errors of order 10 per cent in temperature and a 15 per cent in density have resulted from an LTE assumption.

As Example 9.2 has pointed out, some error is introduced into the calculations by making the assumptions outlined in this section. The magnitude of this error will vary, depending on the RRL adopted and the physical conditions in the gas (see Ref. [153] for examples). Thus, Eq. (9.12) can be used to obtain an initial estimate of the electron temperature, but a full non-LTE approach will yield more accurate results (Prob. 9.6).

It is interesting that, at very low frequencies, some RRLs can be seen in *absorption* against the brighter Galactic background, the latter due primarily to synchrotron radiation that is stronger at lower frequencies (see Eqs. 8.42, 8.43). As an example, the absorption line of singly ionized carbon, C575 α has been observed at $v = 32.4$ MHz (Ref. [83]).

9.4.2 Optical recombination lines

As we step down the 'ladder' of quantum numbers, we eventually reach the more familiar Balmer lines of hydrogen which are observed in the optical part of the

[19]Eq. (8.17) is not applicable for this example since it uses a different geometry from what has been assumed.

spectrum (see Table C.1). Given the lengthy history of optical observations in comparison to any other waveband (Sect. 2.1), it is not surprising that images in the Balmer lines, especially the red H α line, are plentiful. Just as we saw for the RRLs, optical recombination lines are formed in ionized gas, and many beautiful and striking images of H II regions, planetary nebulae, and other ionized regions have been obtained in the H α line as well as in emission lines from other species. Figures 3.7, 3.13, 5.10, and 6.6 provide some examples. The spectrum of an H II region was also shown in Figure 8.9, illustrating the rich abundance of lines available for study. What can these lines tell us about the physical conditions in the ionized gas?

For hydrogen in the relatively low density interstellar medium, optical recombination lines are not in LTE. The lifetimes of electrons in their various states are very short (see the Einstein A coefficients in Table C.1) and downwards transitions will be spontaneous. This means that the line emission coefficient, j_ν, can be expressed by Eq. (9.8). For optically thin lines, then, in the absence of a background source, the specific intensity of a line simply requires a multiplication of j_ν by the line-of-sight depth through the cloud, l, as indicated by Eq. (6.17). The problem, then, reduces to one of finding the number of particles in the upper state, n_j, that can be used in Eq. (9.8). This task is less trivial and requires a consideration of the recombination of free electrons with protons, non-LTE conditions, and quantum mechanical selection rules as electrons cascade down various levels in the atom. The result, integrated over the line (erg s^{-1} cm^{-3} sr^{-1}), is,

$$\int_\nu \mathrm{j}_\nu \, d\nu = \frac{1}{4\,\pi}\, h\, \nu_{\mathrm{j,i}}\, \alpha_{\mathrm{j,i}}\, n_\mathrm{e}\, n_\mathrm{i} \approx \frac{1}{4\,\pi}\, h\, \nu_{\mathrm{j,i}}\, \alpha_{\mathrm{j,i}}\, {n_\mathrm{e}}^2 \tag{9.13}$$

where $\nu_{j,i}$ is the frequency of the line centre, $\alpha_{\mathrm{j,i}}$ is the *effective recombination coefficient for transition, $j \rightarrow i$* (cm^3 s^{-1}), n_i, n_e are the ion and electron densities, respectively, and the latter approximation applies if the ion and electron densities are equal which we take to be the case. The recombination coefficient has appeared before when we considered the ionization equilibrium of an H II region (see Example 5.4) for which α_r included recombinations to all levels of the atom. In the above case, $\alpha_{\mathrm{j,i}}$ will be different for different transitions and is *defined* such that $\alpha_{\mathrm{j,i}}\, n_i\, n_\mathrm{e}$ is the total number of photons per second per cubit centimetre in transitions from level j to i. Since recombination is a collisional process, the effective recombination coefficients, as we saw for α_r (see Table 3.2), are functions of temperature. Several values are provided in Table 9.4.

We can now use Eq. (9.13) with Eq. (6.17) to obtain the intensity (erg s^{-1} cm^{-2} sr^{-1}) over the line, assuming $n_\mathrm{i} \approx n_\mathrm{e}$,

$$\int_\nu I_\nu \, d\nu = \frac{1}{4\,\pi} h\, \nu_{\mathrm{j,i}}\, \alpha_{\mathrm{j,i}}\, n_\mathrm{e}^2\, l = 2.46 \times 10^{17}\, h\, \nu_{\mathrm{j,i}}\, \alpha_{\mathrm{j,i}}\, \mathcal{EM} \tag{9.14}$$

where we have expressed the emission measure, $\mathcal{EM}$ (see Eq. 8.5) in the usual units of cm^{-6} pc. As usual for thermal gas in the optically thin limit, the emission is a function of temperature (via α_ji) and density (via $\mathcal{EM}$).

Table 9.4. Effective recombination coefficients[a]

	Temperature (K)		
$\alpha_{j,i}$ ($cm^3\ s^{-1}$)	5 000	10 000	20 000
$\alpha_{3,2}$ (H α)	2.21×10^{-13}	1.17×10^{-13}	5.97×10^{-14}
$\alpha_{4,2}$ (H β)	5.41×10^{-14}	3.03×10^{-14}	1.61×10^{-14}

[a]Case B recombination has been assumed (see Sect. 9.4.2). H α values are from Ref. [160] and H β values are from Ref. [115]. For the latter reference, means for the densities, 10^2 and 10^4 cm^{-3} are given.

We now come to an interesting point. The effective recombination coefficients are well known and have been tabulated for a wide range of conditions. Also, for any two lines in a given atom, the emission measure should be the same (see also Sect. 9.4.1). This means that the *ratio* of the intensities of any two lines should be a function of frequency and the effective recombination coefficient, only. Forming this ratio from Eq. (9.14).

$$\frac{\int_\nu I_\nu \, d\nu(j \rightarrow i)}{\int_\nu I_\nu \, d\nu(k \rightarrow m)} = \frac{\nu_{\mathrm{j,i}}\, \alpha_{\mathrm{j,i}}}{\nu_{\mathrm{k,m}}\, \alpha_{\mathrm{k,m}}} \tag{9.15}$$

where $j \rightarrow i$ and $k \rightarrow m$ represent the two different line transitions.

Thus, the various line ratios can be calculated and tabulated as a function of temperature, since $\alpha_{\mathrm{j,i}}$ is a function of temperature. In principle, then, measuring the line ratio of any two such lines can provide us with the temperature by comparison with the tabulated values. Since the ratio is formed from the integrals over two lines, the result is not dependent on the type of line broadening that is present, as was the case for RRLs for which a line/continuum ratio was formed. Thus, we have yet another temperature diagnostic for H II regions or other interstellar ionized gas clouds. The density could then be found via Eq. (9.14) with a knowledge of T_e and distance to the object. In practice, as we will now describe, there are a few complications to this simple approach using the Balmer lines of hydrogen that need to be considered.

If all hydrogen lines were optically thin (a situation called *Case A*), then there should be no density dependence to $\alpha_{j,i}$ or to the line ratios of Eq. (9.15). However, as pointed out in Sect. 5.1.1.2, the Lyman lines of hydrogen, which emit in the UV, tend to be trapped in ionized nebulae (a more realistic and density-dependent condition called *Case B*). For example, after several scatterings, there is some probability that a Ly β line (a transition from n = 3 to n = 1) will be degraded into an H α (n = 3 to n = 2) and then a Ly α (n = 2 to n = 1) line. Even though the Balmer lines themselves are optically thin, the Balmer line ratios will still be affected. Thus, the values of $\alpha_{\mathrm{j,i}}$ given in Table 9.4 are for Case B and a sample line ratio calculation for this case is provided in Example 9.3. Expected line ratios are tabulated for this case in Table 9.5. The corrections are not large. Over the range of temperature and density listed, the maximum difference between Case A and Case B is only 3 per cent and, as can be seen in the table, the maximum error introduced by changing the density a factor of 100 is even less.

Table 9.5. Sample line ratios for the hydrogen balmer series[a]

	T_e (K)					
	5 000		10 000		20 000	
	n_e (cm^{-3})		n_e (cm^{-3})		n_e (cm^{-3})	
Line ratio	10^2	10^4	10^2	10^4	10^2	10^4
H α/H β	3.04	3.00	2.86	2.85	2.75	2.74
H γ/H β	0.458	0.460	0.468	0.469	0.475	0.476
H δ/H β	0.251	0.253	0.259	0.260	0.264	0.264
H ϵ/H β	0.154	0.155	0.159	0.159	0.162	0.163

[a]Ref. [50]. Case B recombination has been assumed (see Sect. 9.4.2).

Example 9.3

Compute the expected line ratio, H α/H β, for an H II region of temperature, 10^4 *K.*

From Eq. (9.15), we can write,

$$\frac{\mathrm{H}\,\alpha}{\mathrm{H}\,\beta} = \frac{\nu_{3,2}}{\nu_{4,2}}\frac{\alpha_{3,2}}{\alpha_{4,2}} = \frac{\lambda_{4,2}}{\lambda_{3,2}}\frac{\alpha_{3,2}}{\alpha_{4,2}} = \left[\frac{486.132}{656.280}\right]\left[\frac{1.17\times10^{-13}}{3.03\times10^{-14}}\right] = 2.86 \qquad (9.16)$$

where we have used $c = \lambda\,\nu$ and data from Tables 9.4 and C.1.

Of greater concern is that the dependence of the Balmer line ratios on temperature is actually rather weak. Looking at Table 9.5, a change in temperature of a factor of two results in changes in line ratios of only a few per cent – in other words, about the same as changing the assumption from Case A to Case B. The line ratios as well as model calculations for $\alpha_{j,i}$ would have to be measured and calculated, respectively, to very high accuracy for these ratios to provide useful temperature diagnostics.

The solution is to use other line ratios that show stronger temperature dependences and for which all lines are optically thin. Fortunately, such lines do exist, the most commonly used ones being associated with [OIII] and [NII]. The square brackets refer to the fact that the relevant lines are *forbidden* (e.g. see Appendix C.4). For forbidden lines, the lifetime in the upper state is relatively long yet, in interstellar space, the densities are still low enough that emission results from spontaneous transitions. These lines were first discovered in nebulae and the substance responsible for them was originally named 'nebulium' since the same lines could not be seen in the higher density gases of Earth-based laboratories in which collisionally-induced transitions dominated. The unknown lines were later identified as being due to forbidden [OIII] and [NII] transitions. It is useful to form ratios using three

lines for each species, designated R_O and R_N for [OIII] and [NII], respectively (Ref. [96]),

$$R_O = \frac{I(\lambda 4959) + I(\lambda 5007)}{I(\lambda 4363)} = \frac{7.73 \exp[(3.29 \times 10^4)/T_e]}{1 + 4.5 \times 10^{-4} [n_e/(T_e)^{1/2}]} \quad ([OIII]) \qquad (9.17)$$

$$R_N = \frac{I(\lambda 6548) + I(\lambda 6583)}{I(\lambda 5775)} = \frac{6.91 \exp[(2.50 \times 10^4)/T_e]}{1 + 2.5 \times 10^{-3} [n_e/(T_e)^{1/2}]} \quad ([NII]) \qquad (9.18)$$

where the ratio is most sensitive to temperature in comparison to density (Prob. 9.7).

These are not the only optical diagnostics of ionized regions. The line-to-continuum ratio can also be used, as was done for RRLs (though this ratio involves more complexity given the different processes that contribute to the optical continuum), and other line ratios, for example, from [SII] and [OII], are particularly sensitive to electron density, n_e. Thus, by carefully choosing the appropriate lines, a variety of optical observations can successfuly probe the physical conditions in H II regions, planetary nebulae, and the diffuse ionized interstellar gas.

There remains one more complication that must be considered when dealing with optical observations – that of obscuration by dust. Even if the spectral lines are optically thin, there will still be dust throughout ionized regions, as illustrated in Figures 3.13 and 8.4. If the extinction (Sect. 3.5.1) were independent of wavelength, then in forming a line ratio, the effect would be the same for both lines and the ratio would not be affected. However, extinction is wavelength dependent. Optical transitions that fall into the blue part of the spectrum will be affected by dust more than transitions in the red part of the spectrum, hence the term 'reddening' to describe this effect. As Figure 3.21 showed, $E_{\lambda-V} \propto 1/\lambda$, roughly, in the optical part of the spectrum, but there can be variations depending on location, so a 'location-specific' correction is desirable.

One way to deal with this problem is to use lines, if possible, that are very close together in frequency (for example, the [OII] $\lambda 3727/\lambda 3726$ ratio) so that extinction is approximately the same for the lines. Another approach, however, is to *use* line ratios to correct for the reddening. Since the Balmer lines, for example, are not strongly dependent on temperature and density, significant departures from the expected ratios of Table 9.5 are more likely to be due to reddening than to intrinsic variations in nebular properties. Corrected intensities would then be used in equations such as Eq. (9.17) or (9.18).

Thus, the rich spectra seen in the optical waveband yield the properties of the nebulae, from information on the elements that they contain, to their densities and temperatures, and even their dust content. These are all bits and pieces of a broader mosaic to which we can appeal in an attempt to form a coherent picture of star formation and evolution, such as described in Chapter 3.

9.4.3 The 21 cm line of hydrogen

In 1944, the Dutch astronomer, Jan Oort, turned to a young H. C. van de Hulst and said, "... by the way, radio astronomy can really become very important if there were at least one line in the radio spectrum." (Ref. [186]). Within the year, van de Hulst had predicted the presence of the λ 21 cm line of H I and, by 1951, three independent teams (Dutch, Australian, and American) successfully detected this line, each publishing their results in the same issue of the scientific journal, *Nature*.

The λ 21 cm line originates from the bottom rung of the hydrogen atom's ladder. Resulting from a tiny electron that flips its spin, the corresponding transition is between two hyperfine levels of the ground state (see Appendix C.4). Not being hindered by dust, this radio emission can be seen clear across the Galaxy or, for that matter, through any other galaxy or intergalactic space. Oort's prediction was prophetic. The importance of this apparently insignificant quantum event cannot be overstated. It is largely through observations in the λ 21 cm line that we know about the frothy nature of the neutral interstellar medium (e.g. Figures 3.17, 6.10), the large-scale structure and rotation of the Milky Way (see discussion in Sect. 7.2.1.2), the structure of the ISM in other galaxies, the rotations of other galaxies (e.g. Figure 2.20) and implied presence of associated dark matter (Sect. 3.2), tidal interactions between galaxies (see Figure 9.7, and the expansion of the Universe itself (Sect. 7.2.2). Here, we concentrate on relating the observed λ 21 cm line to the mass of H I.

We have already seen that essentially all hydrogen atoms in a neutral cloud will have their electrons in the ground state (Sect. 3.4.5). We have also seen how the temperature of H I gas can be determined via the absorption and emission of the λ 21 cm line when a

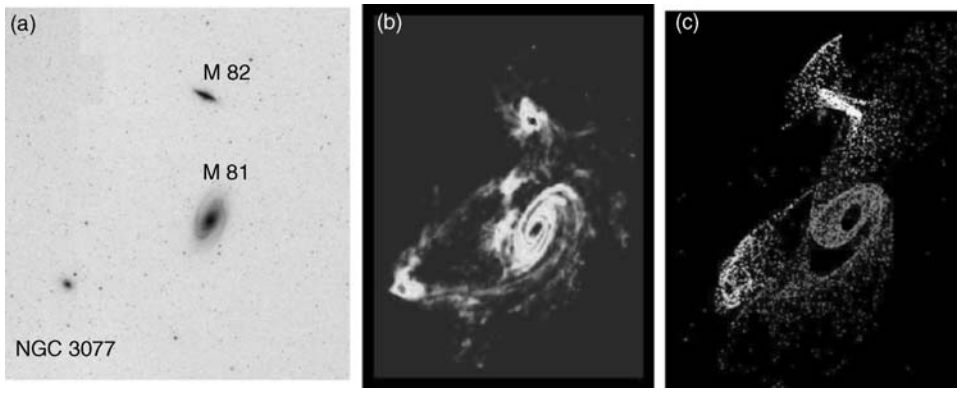


Figure 9.7. **(a)** Optical image of the M 81 Group (distance, $D = 3.6$ Mpc). **(b)** This HI image clearly shows that the three galaxies are interacting, with tidal bridges between the galaxies and other intergalactic neutral hydrogen gas. For M 81, the average column density is $\mathcal{N}_{\rm HI} \approx 4.0 \times 10^{20}$ cm^{-2} within an elliptical region of semi-major $\times$ semi-minor axes, 20×15 arcmin. **(c)** This numerical simulation has reproduced the interaction between the galaxies. The gas originally associated with each galaxy is represented by a different colour. Refs. [184], [185]. (Reproduced by permission of M. Yun). (See colour plate)

background source is present (Sect. 6.4.3). Since this is a forbidden line, the lifetime in the upper state is long and, under typical astrophysical conditions, the line is collisionally excited and therefore in LTE (Sect. 6.4.3). This means that the Boltzmann Equation (Eq. 3.23) can be used to describe the relative populations of the two hyperfine levels. Since $h\,\nu \ll k\,\mathrm{T}$ for this line, the Boltzmann factor becomes 1 so the Boltzmann Equation gives an expression for the population of states that is simply the ratio of their statistical weights.

$$\frac{N_2}{N_1} = \frac{n_2}{n_1} = \frac{g_2}{g_1} = \frac{3}{1} \tag{9.19}$$

where the subscripts, 2, and 1, refer to the upper and lower hyperfine energy levels, respectively and we have also expressed this ratio in terms of the ratio of number densities, n_2/n_1. Therefore, the number density fraction of all particles in the upper and lower states, respectively are,

$$\frac{n_2}{n_{\mathrm{H\,I}}} = \frac{3}{4} \qquad \frac{n_1}{n_{\mathrm{H\,I}}} = \frac{1}{4} \tag{9.20}$$

where $n_{\mathrm{H\,I}}$ is the total number density of neutral particles.

The optical depth is then (Eqs. 5.8, 5.9),

$$\tau_\nu = \sigma_\nu \int_l n_1\, dl = \frac{1}{4}\,\sigma_\nu \int_l n_{\mathrm{H\,I}}\, dl = \frac{1}{4}\,\sigma_\nu\,\mathcal{N}_{\mathrm{H}} \tag{9.21}$$

where σ_ν is the effective cross-section (cm^2) (presumed not to vary along the line of sight) dl is an element of line-of-sight distance through the cloud (cm), and the *column density*, $\mathcal{N}_{\mathrm{H\,I}}$ (cm^{-2}) of hydrogen is defined by,

$$\mathcal{N}_{\mathrm{H\,I}} \equiv \int_l n_{\mathrm{H\,I}}\, dl \tag{9.22}$$

Thus, the column density of some species is the number density of that species integrated over the line of sight.

For the λ 21 cm line, it can be shown that the effective cross-section is given by (Ref. [138]),

$$\sigma_\nu = \frac{h\,c^2}{8\,\pi\,\nu_{1,2}\,k}\,\frac{g_2}{g_1}\,A_{2,1}\,\frac{\Phi_\nu}{\mathrm{T_S}} = 1.04 \times 10^{-14}\,\frac{\Phi_\nu}{\mathrm{T_S}} \tag{9.23}$$

where $\nu_{1,2}$ is the frequency of the transition, $A_{2,1}$ is the Einstein A coefficient for the transition (see Table C.1), Φ_ν is the line shape function (Eq. 9.1 being one example),

and T_S is the spin temperature (see Sect. 3.4.5) of the gas, presumed to equal the kinetic temperature for a line in LTE.

We can now put Eq. (9.25) into Eq. (9.21) and integrate over the line,

$$\int_v \tau_v \, dv = \frac{1}{4} \left(1.04 \times 10^{-14}\right) \frac{\mathcal{N}_{\mathrm{H\,I}}}{T_S} \int_v \Phi_v \, dv \tag{9.24}$$

By definition, the integral over the line shape function is 1 (see Eqs. 9.5 and D.18, for example), so Eq. (9.24) can be rearranged to solve for the H I column density,

$$\mathcal{N}_{\mathrm{H\,I}} = 3.85 \times 10^{14} \, T_S \int_v \tau_v \, dv \tag{9.25}$$

T_S is presumed to be constant along a line of sight[20] but no such assumption is necessary for the density. When the $\lambda\,21$ cm line spectrum is plotted, the Doppler relation (Eq. 7.4) is often applied so that velocity, rather than frequency, is along the x axis (e.g. Figure 6.3). In such a case, Eq. (9.25) becomes,

$$\mathcal{N}_{\mathrm{H\,I}} = \left(1.82 \times 10^{18}\right) T_S \int_v \tau_v \, dv \tag{9.26}$$

where $\mathcal{N}_{\mathrm{H\,I}}$ is still in cgs units of cm^{-2} but the velocity is in km s^{-1}. We now use Eq. (6.31) to substitute for τ_v,

$$\mathcal{N}_{\mathrm{H\,I}} = -\left(1.82 \times 10^{18}\right) T_S \int_v \ln\left[1 - \frac{T_B(v)}{T_S}\right] dv \tag{9.27}$$

Notice that $T_B(v)$ is an *observed* quantity, so the column density can now be found, provided there is some knowledge of T_S. Also note that, since $T_B \leq T_S$ (Eq. 6.24), the logarithmic term will be negative, so the column density is positive, as required.

As we have done before, let us consider the optically thick and optically thin limits of Eq. (9.27). If the line is optically thick at some velocity, v, then $T_B(v) = T_S$ (Eq. 6.32), and the logarithm cannot be evaluated. In such a case, the column density of the cloud cannot be obtained because we can only see into the front of the cloud (see Figure 4.2). On the other hand, if the line is optically thin, then Eq. (6.33) can be substituted into

[20]The variation in temperature within some region of the ISM is usually less than the variation in density. However, should the spin temperature vary significantly along the line of sight, then the appropriate temperature to be used in these equations is a harmonic mean temperature, weighted by HI gas density, n. More explicitly, $\left\langle \frac{1}{T_S(v)} \right\rangle = \left[\int_l \frac{n_v(l)}{T_S(l)} \, dl\right] / \left[\int_l n_v(l) \, dl\right]$, where l refers to position along the line of sight and v is the velocity (linked to frequency via the Doppler formula, Eq. 7.4) of the gas. See Ref. [138] for further details.

Eq. (9.26) to find,

$$\left[\frac{\mathcal{N}_{\rm H\,I}}{\rm cm^{-2}}\right] = (1.82\times10^{18}) \int_v \left[\frac{\rm T_B(\mathit{v})}{\rm K}\right] \left[\frac{d\mathit{v}}{\rm km\,s^{-1}}\right] \qquad (\tau_\mathit{v} \ll 1) \tag{9.28}$$

where we explicitly indicate the units. Now the column density depends only on the observed brightness temperature integrated over the line, without the need to know the gas temperature. It is also independent of distance. This is an important result and is used widely since λ 21 cm lines are very often optically thin. Once the column density is known, the volume density and/or H I mass can be found, provided the distance is known, as Example 9.4 indicates.

Example 9.4

The H I line emission from the galaxy, NGC 2903*, is optically thin and subtends a solid angle of* $\Omega_{\rm s} = 1.8\times10^{-5}$ *sr. This emission, averaged over* Ω_s *, and then integrated over all line velocities, is* $\int I_\mathit{v}\,d\mathit{v} = 0.0458$ *Jy beam*$^{-1}$ *km s*$^{-1}$ *(Figure* 2.20 *shows how this emission is distributed in the data cube), where the beam solid angle is* $\Omega_{\rm b} = 5.21\times10^{-9}$ *sr. Using information given in Figure* 2.11 *and the solution to Prob.* 7.5*, find the H I mass of NGC* 2903 *and what fraction of the total mass this represents.*

To use Eq. (9.28), we require the integral over the line to be in units of K km s^{-1}. From the definition of the Jy (Eq. 1.8) and the beam solid angle, $\Omega_{\rm b}$, we find, $\int I_\mathit{v}\,d\mathit{v} = 8.79\times10^{-17}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} km s^{-1}. Then using the Rayleigh–Jeans relation (Eq. 4.6) to convert to a brightness temperature at the line centre frequency, we have, $\int_{\rm v} {\rm T_B}\,d\mathit{v} = 142$ K km s^{-1} as an average over $\Omega_{\rm s}$. Putting this into Eq. (9.28) yields an average column density of $\mathcal{N}_{\rm H\,I} = 2.58\times10^{20}$ cm^{-2}.

Now if this column density is multiplied by the *area* of the source, the result will be the total number of neutral hydrogen atoms in the galaxy. The area is given by $\sigma_{\rm s} = r^2\,\Omega_{\rm s}$ (Eq. B.2), where r is the distance to the galaxy ($r = 8.6$ Mpc, see Figure 2.11), so $\sigma_{\rm s} = 1.27\times10^{46}$ cm^2. Therefore, the total number of H I atoms is $N = \mathcal{N}_{\rm H\,I}\,\sigma_{\rm s} = 3.27\times10^{66}$. Multiplying this by the mass of an H I atom and dividing by the mass of the Sun (Tables G.2, G.3), yields ${\rm M_{H\,I}} = 2.7\times10^9\ {\rm M_\odot}$.

From Prob. 7.5, we found a total mass for this galaxy of ${\rm M_{tot}} \approx 1.6\times10^{11}\ {\rm M_\odot}$. Therefore, the H I mass constitutes about 2 per cent of the total mass.

A number of the steps in Example 9.4 can be folded into a single equation for the H I mass. The result is,

$$\left[\frac{\rm M_{H\,I}}{\rm M_\odot}\right] = 2.35\times10^5 \left[\frac{D}{\rm Mpc}\right]^2 \int_v \left[\frac{f_\mathit{v}\,d\mathit{v}}{\rm Jy\,km\,s^{-1}}\right] \tag{9.29}$$

where f_v is the total flux density of the source at some velocity in the line and D is the distance to the source, in the specified units. Eq. (9.29) is commonly used to obtain the H I mass of an object when the flux density is known at each velocity in the line.

9.5 Probing physical conditions via molecular transitions

Stars are formed from dense, dusty molecular clouds in the interstellar medium and, as Figure 9.2 showed, many molecular lines can be excited in the warm, dense molecular envelopes around new-born stars or in the nearby environment. Farther from local sites of star formation, however, are extended molecular clouds that permeate the disk of the Milky Way. These may also include embedded star formation, but the extended clouds (for example, *giant molecular clouds*, or GMCs, which may be 100 pc in size) have conditions that are less amenable to the excitation of so many lines. Typical temperatures and densities are T = 20 K and $n = 10^3$, both of which are considerably lower than the highest temperatures and densities found in the Orion KL cloud (Sect. 9.1.2). Nevertheless, many lines are still accessible and provide probes of the molecular conditions, especially rotational transitions which do not require as much energy to be excited as do vibrational ones.

Hydrogen is the most abundant element in the Universe and it is not surprising the most abundant molecule, by far, is H_2[21]. Determining the mass of H_2 gas is equivalent to finding the mass of the molecular cloud, since the quantities of all other molecules and dust are much less. However, because H_2 has no permanent electric dipole moment, it cannot emit dipole radiation from rotational transitions (Sect. 9.1.2). Moreover, even the much weaker quadrupole transitions require relatively high energies to excite. For example, the first rotational state above the ground state for H_2 requires temperatures of 510 K which is much higher than in the deep interstellar medium. The conclusion is that the most abundant molecule cannot be observed in the state in which it is most commonly found.

How then can we observe interstellar molecular clouds? The alternative is to consider the next most abundant molecule, which is carbon monoxide.

9.5.1 The CO molecule

The CO molecule has an electric dipole moment (Sect. 9.1.2) and, as indicated in Table 9.3, the critical density of the CO($J = 1 \rightarrow 0$) rotational line at λ 2.6 mm, at a molecular cloud temperature of 30 K is approximately the same as the typical density that is observed in such clouds. This implies that this rotational CO line is collisionally excited, and the species with which it collides will be the most abundant molecule, H_2. The λ 2.6 mm line is a strong line, is easily observed, and in fact, is often optically thick. Other CO lines may also be present, depending on excitation conditions (for example, there are lines at λ 1.3 mm, and λ 0.87 mm), and lines of the $C^{18}O$ (e.g. Figure 3.20) and ^{13}CO isotopes, which are weaker and optically thin, may also be

[21]The process of formation of this and other molecules is still uncertain. The current view is that molecules are formed in the presence of a catalyst, probably on the surfaces of dust grains, after which they evaporate into the gas phase again.

observable. Therefore, a variety of CO lines provide important diagnostics of interstellar molecular clouds. Moreover, the dynamics of the cloud can also be accessed via the Doppler shifts of these lines.

Since the density of the molecular cloud is essentially the density of its constituent H_2 molecules, much work has been expended in an attempt to determine the density of H_2 that is required to excite the various CO lines to their observed values, especially the CO($J = 1 \rightarrow 0$) line, which is most readily observed[22]. This process is not straightforward because, unlike the λ 21 cm line for which the quantity of H I can be determined directly from a line of the same atom, the observed CO lines are used to *infer* the quantity of the colliding species, H_2. For example, non-LTE conditions may apply and optical depth effects are important. The weaker, optically thin isotopic lines may be used, but they are not always observed and isotopic ratios may not be as well known as desired. Moreover, unknown amounts of carbon may be tied up in grains rather than existing only in the gas phase. Variations in the environmental conditions can also affect the CO-to-H_2 relation. If the metallicity is low, for example, the CO lines will be weaker for a given H_2 density. A high UV radiation field or cosmic ray flux will also be important because the two molecules can be dissociated, but CO tends to be dissociated more readily than H_2.

All things considered, there are many potential pitfalls in this process. Yet when the molecular mass is determined via the assumption that the cloud is gravitationally bound (such a cloud is said to be *virialized*) and compared to measures of its mass via a line ratio analysis, the conclusion is that the relationship between H_2 density and CO($J = 1 \rightarrow 0$) line strength is essentially linear. The relationship is therefore expressed as,

$$\left[\frac{\mathcal{N}_{\mathrm{H_2}}}{\mathrm{cm^{-2}}}\right] = X \int_v \left[\frac{\mathrm{T_B[CO}(J = 1 \rightarrow 0)]}{\mathrm{K}}\right] \left[\frac{dv}{\mathrm{km\,s^{-1}}}\right] \tag{9.30}$$

where Eq. (9.30) applies only to the CO ($J = 1 \rightarrow 0$) λ 2.6 mm line and X is the sought-after conversion factor. For the Milky Way, the 'standard' conversion factor, X, is in the range,

$$X = (2.3 \rightarrow 2.8) \times 10^{20} \quad \mathrm{cm^{-2}\ (K\,km\,s^{-1})^{-1}} \tag{9.31}$$

and values in this range, in the absence of other information, are widely applied. It is well-known that X increases if the metallicity decreases, the variation in X being about a factor of 5 for a metallicity variation of a factor of 10 (Ref. [183]). When galaxies of different morphological type are taken into account, X may span a range from $(0.6 \rightarrow 10) \times 10^{20}$ molecules $\mathrm{cm^{-2}\,(K\,km\,s^{-1})^{-1}}$ (Ref. [24]).

The CO molecule has now been observed in a wide variety of Galactic environments as well as in other galaxies. The same molecule that, in excess, is an unwelcome

[22]The wavelength of this line is long enough that it does not need to be observed from a high altitude (see Figure 2.2).

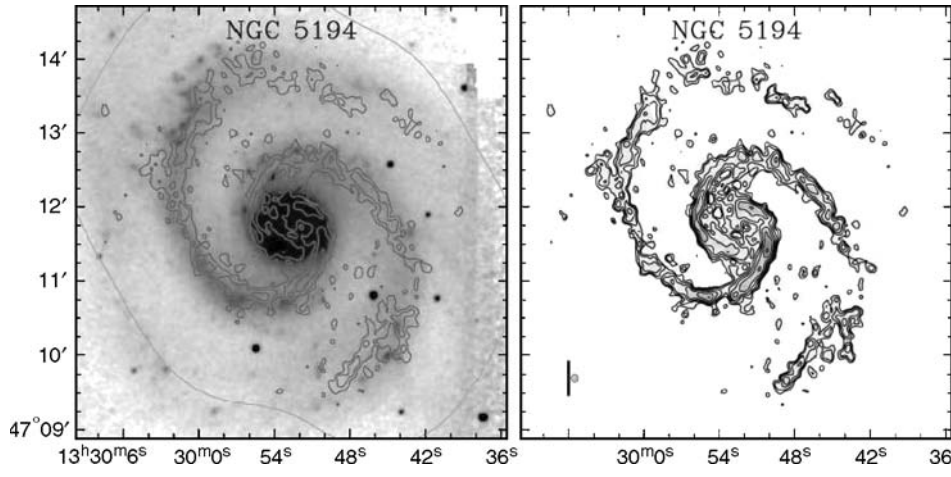


Figure 9.8. The galaxy, NGC 5194 (alias M 51 or the Whirlpool Galaxy, see also Figure 3.8) is shown in both these images, from Ref. [133]. At left, is a K-band (Table 1.1) image in greyscale superimposed with contours of CO ($J = 1 \rightarrow 0$) emission. The right image shows CO emission only using the same contours as at left. The data were taken as part of the Berkeley–Illinois–Maryland Association Survey of Nearby Galaxies (BIMA-SONG). For the both images, the first contour is at 5 Jy beam^{-1} km s^{-1} and, for the right image, each higher contour is 1.585 times higher than the previous one. The beam has major and minor axes of $5.8'' \times 5.1''$, respectively. The bar on the right image is 1 kpc in length. (Reproduced by permission of M. Regan)

constituent of smog on the Earth, 'lights up' the otherwise invisible molecular clouds in our own and other galaxies, providing us with a probe of the nature and properties of these cold clouds. Moreover, since stars are formed within molecular clouds (Sect. 3.3.2), CO and other molecules allow us to trace stellar birthplaces. The molecular gas distribution, in general, follows the light distribution of young stars in galaxies[23], as Figure 9.8 shows. The CO emission in the spiral galaxy, M 51 (see also Figure 3.8) is particularly obvious along the spiral arms where shock waves[24] have compressed the molecular gas, enhancing star formation in these regions. Although the details of star formation are still not well understood, observations of molecular lines such as from CO promise to unlock these secrets in the future, as telescope resolution and sensitivity improve.

Problems

9.1 Compute the frequencies of the following hydrogen transitions and indicate in which part of the electromagnetic spectrum the resulting spectral lines would be found, (a) Ly δ, (b) H_9, (c) Bϵ, (d) H110 α, (e) H239γ.

[23]There are differences on small scales when the distribution is scrutinized in detail, however.
[24]See Footnote 16 in Chapter 3.

9.2 (a) Beginning with $\Phi_{\mathcal{D}}(\nu)$ (Eq. 9.1), derive the line shape function, $\Phi_{\mathcal{D}}(v)$, where v is the velocity of the particles in a gas cloud. Express the result as a function of the velocity parameter, b, assume that the cloud, itself, is at rest, and that all $v \ll c$.

(b) From the result of part (a), find an expression for the line width (FWHM) and confirm that it is the same as Eq. (9.6).

(c) Verify that $\int_{-\infty}^{\infty} \Phi_{\mathcal{D}}(v)\, dv = 1$.

9.3 (a) Find the ratio of natural to thermal line width, $\Delta v_{\mathcal{L}}/\Delta v_{\mathcal{D}}$, for the Balmer H$\alpha$ emission line from a pure hydrogen cloud and typical ISM conditions.

(b) A sodium line (λ 589 nm) in the Solar atmosphere at a temperature of 7500 K is observed to have a line width of 10 nm. Find the thermal to collisional line width ratio, $\Delta \lambda_{\mathcal{D}}/\Delta \lambda_{\rm P}$ for this line.

9.4 The ^{44}Ti line measured from the Cas A supernova remnant (Figure 1.2) is centred at 1152 ± 15 keV and has a FWHM of 84.6 ± 10 keV (Ref. [78]). What is the velocity of expansion of this supernova remnant, as measured by the ^{44}Ti ejecta? Does this result agree or disagree with the maximum known expansion velocity of $\approx 10\,000$ km s^{-1} for this SNR measured from higher resolution observations?

9.5 (a) With the help of a spreadsheet or other computer software, compute the frequencies of all hydrogen radio recombination lines (RRLs), Hnα, where $50 \leq {\rm n} \leq 300$. Compute the thermal line widths of each of these lines, $\Delta v_{\mathcal{D}}$, assuming that they originate in a typical H II region.

(b) The Very Large Array (VLA) radio telescope[25] can be set to any frequency in 'L' band over a range $1240 \leq \nu \leq 1700$ MHz. How many Hnα RRLs fall within L band?

(c) Once the telescope's receiver has been set to a given frequency, ν_0, within L band, observations at any one time can only be carried out within a sub-band which has some bandwidth, ΔB, centred at ν_0, consisting of $N_{\rm ch}$ individual channels of width, $\Delta \nu_{\rm ch}$. Bandwidth possibilities include – *Sub-band 1*: $\Delta B = 50$ MHz, $N_{\rm ch} = 8$, $\Delta \nu_{\rm ch} = 6250$ kHz; *Sub-band 2*: $\Delta B = 25$ MHz, $N_{\rm ch} = 16$, $\Delta \nu_{\rm ch} = 1562.5$ kHz; *Sub-band 3*: $\Delta B = 6.25$ MHz, $N_{\rm ch} = 64$, $\Delta \nu_{\rm ch} = 97.656$ kHz; *Sub-band 4*: $\Delta B = 1.5625$ MHz, $N_{\rm ch} = 256$, $\Delta \nu_{\rm ch} = 6.104$ kHz. Given these options, what is the highest number of Hnα RRLs that can be observed at any one time and still be spectrally resolved (recall Sect. 2.5)?

9.6 From the information in Figure 9.6 about the ionized gas in the G24.8+0.1 star forming region, do the following:

(a) Estimate Δv(kHz) from the line profile.

(b) Estimate the electron temperature, T_e, adopting the assumptions given in Sect. 9.4.1. Correct this value for non-LTE and other effects by increasing it by a factor of 1.5.

[25]The VLA is operated by the National Radio Astronomy Observatory. The National Radio Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. (USA). Quoted values are 2006 numbers.

(c) Examine and compare Eq. (8.13) with Eq. (9.11) and suggest reasons as to why it might better to use the continuum measurement to calculate EM, rather than the line measurement (assuming that the continuum is optically thin).

(d) From the continuum flux density, the corrected electron temperature from (b) above, the adopted spherical geometry, and assuming uniform density, determine $\mathcal{EM}$, n_e, and $M_{H\ II}$.

(e) Assuming Solar abundance, determine the total mass of the ionized gas.

9.7 (a) Plot a graph of the optical line ratios, R_O, and R_N, as a function of electron temperature over the range, $5\,000 \leq T_e \leq 20\,000$, for two values of electron density, $n_e = 10^2$ and $n_e = 10^4$ cm^{-3}. Use a linear scale for T_e and a logarithmic scale for the ratio.

(b) Comment on the ability of these line ratios to distinguish between electron temperatures and electron densities over the ranges plotted.

(c) Indicate what the minimum dynamic range (Sect. 2.5) would have to be in order to measure the ratio, R_O, of an H II region with a temperature, $T_e = 5\,000$ K.

9.8 Use the information in the question of Example 9.4 to find the H I mass of NGC 2903 using Eq. (9.29) instead of Eq. (9.28). For simplicity, assume that Ω_s is about the same at every velocity in the line.

9.9 From the information of Figure 9.7, find the H I mass ($M_\odot$) of M 81.

9.10 A spherical Galactic H I cloud at a distance of $D = 1$ kpc has an angular diameter of $\theta = 1.2$ arcmin. This cloud is known to have a $\lambda\,21$ cm line width that is due to thermal motions only, at a temperature of $T = 85$ K. The flux density of the cloud at the peak of the line is $f_{max} = 0.1$ Jy. For this cloud, find the following:

(a) The line FWHM, Δv (km s^{-1}).

(b) Its H I mass, M_{HI} ($M_\odot$).

(c) Its diameter, d (pc), and volume, V (cm^3).

(d) Its number density, n (cm^{-3}).

(e) Its brightness temperature, T_B (K), at the line peak.

(f) Its optical depth, τ, at the line peak.

9.11 Assuming that the isolated contour at the location, RA = $13^h\,30^m\,07.5^s$, DEC = $47^\circ 13' 16''$ in Figure 9.8, represents a real molecular cloud in the galaxy, estimate the molecular gas mass ($M_\odot$) of this cloud. Assume that the Rayleigh–Jeans relation holds.

PART V
The Signal Decoded

As the previous chapters have shown, considerable effort is required to extract the basic parameters of astronomical objects. The signal must be defined, measured, corrected for instrumental and atmospheric effects, possibly corrected for the effects of interstellar particles and fields, and understood at a physical level that is deep enough to relate the observed emission to the properties of the source. Now that the basics are all in place, how do we actually separate one type of process from the other, and how do we piece together reasonable scenarios so that some scientific insight may be gleaned, beyond a simple description of the astronomical source? The next chapter attempts to address these questions.

10
Forensic Astronomy

> . . .a list of observations, a catalogue of facts, however precise, does not constitute science. The reason is that they tell no story. Only facts arranged as narrative qualify as science.
>
> –Nobel Laureate, John Polanyi (Ref. [126])

In this chapter, our goal is to put together the bits and pieces of information that we have gleaned from earlier chapters and, as the prologue suggests, discover a story. The signal may come to us 'coded' but, with knowledge of the physics of emission processes and interactions that occur en route to our telescopes, we may be able to decipher these codes and gain insight into the nature of the astronomical objects that have both puzzled and fascinated us. Our tools are not just the sophisticated instruments that detect the signal, but also the equations that relate the light that we observe to parameters of the source. It is sometimes possible to go even further. By considering as much evidence as is available, the origin, evolution, energy source, lifetime, relation to other sources, or other information that goes beyond basic properties, might be deduced. But first, we need to disentangle the mixed messages that are sometimes sent to us from space.

10.1 Complex spectra

10.1.1 Isolating the signal

Sometimes we are lucky enough that a signal comes to us uncorrupted and also unmixed with other types of signals. An example is the $\lambda\,21$ cm line from an isolated interstellar HI cloud. Not only does the radio emission travel through interstellar space unperturbed by dust, but no other significant emission at or near that wavelength is emitted by the cloud. It is then straightforward to interpret the result. Often, however,

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

this is not the case. Even if the signal has not been altered in its long journey towards Earth, or even if all corrections have been applied so that the signal is as close as possible to the state in which it was emitted, the result may still include a mixture of different kinds of radiation.

Consider, for example, the spectrum of NGC 2903 that was shown in Figure 2.3. The total flux density ($f_\nu = I_\nu \Omega$, Eq. 1.13) of the continuum emission from this spiral galaxy is plotted as a function of frequency in this figure. At the lowest frequencies up to $\nu = 10^{10}$ Hz, the radio continuum emission declines as a power law with increasing frequency. From our knowledge of emission processes at radio wavelengths, this suggests that the radio spectrum is dominated by synchrotron emission due to cosmic ray electrons in this galaxy. As the frequency increases, we see a broad peak resembling a Planck curve and, since this peak is in the infrared, this suggests that the emission is dominated by the collective contributions of dust grains (Prob. 10.1). Above $\nu = 10^{14}$ Hz, another broad peak is seen, but is less well delineated, given the number and spread of data points in this part of the plot. This peak extends through the optical part of the spectrum and we can infer that it is dominated by the collective Planck curves from different types of stars. Finally, several low intensity points between 10^{17} and 10^{18} Hz (the X-ray band) indicate that high energy and/or temperature emission is also present, likely from a variety of sources.

In fact, the NGC 2903 spectrum is quite typical of what is seen for any normal spiral galaxy, including the Milky Way. From our knowledge of continuum processes, it is possible to provide a rather rough interpretation of the global spectrum, as described above. However, there are many objects in a galaxy, each emitting and thereby providing information that we would like to obtain. The emission from individual objects is hidden in plots like Figure 2.3, being swamped by the collective emission of many objects[1]. Spectral line emission has also been left out of the figure although the lines might have been plotted, had the data been available. If a dominant component can be identified, such as the infrared peak from dust, we are still left with many questions. What range of grain sizes and temperatures are represented? Is most of the dust associated with star-forming regions, with old stars, or with molecular clouds? Answering such questions helps us to grapple with deeper issues, such as the origin of the dust. We therefore need to *isolate* the dust (or other) emission from the more complex global signal.

An obvious approach is to isolate the data *spectrally*. Since observations in different frequency bands require different observational techniques, isolating the signal spectrally is in fact a byproduct of any observation. This immediately limits the number of emission processes that are important in the band. Black body radiation from stars is negligible at centimetre wavelengths, for example. Further narrowing the band may also help. For example, observations of soft X-rays are more likely to detect emission from thermal hot gas (Sect. 8.2.3) whereas hard X-rays are more likely to pick up individual energetic sources, such as X-ray emitting pulsars. Since each radiation

[1]An exception might be if one type of object clearly dominates the spectrum, such as a very strong active galactic nucleus at the core of a distant galaxy, or a supernova that has recently gone off.

process has its own characteristic spectrum, it is possible to 'tune' the observations to focus on a specific type or types of emission. In the narrow band limit, one tunes to a specific spectral line.

The data can also be isolated *spatially*. High spatial resolution allows the targetting of specific regions or objects for careful study. In external galaxies that are not too distant, individual HII regions, globular clusters, bright stars, and interstellar clouds have all been spatially isolated and studied. Although it may not be possible to isolate individual objects in more distant galaxies, as long as the galaxy itself is resolved, one may study a region within it whose size scale is that of the resolution used for the observations. An example has been shown in Figure 2.20, in which a data cube for the $\lambda\,21$ cm emission from NGC 2903 is presented. This line has been isolated both spectrally and spatially, although individual HI clouds have not been resolved.

Even with much effort devoted to spectral and spatial resolution, however, it is still possible for a signal to contain a mixture of different types of emission. It is not always possible, for example, to arbitrarily improve the spatial resolution of a telescope (see Chapter 2) and, even if a specific object can be isolated, there may be foreground or background emission along the line of sight. Different types of radiation may also be emitted from the same object or spatial region. We must then consider whether there are other ways to isolate a given signal.

A possibility presents itself if the signals are so different that they can easily be recognised. An example is when spectral lines sit on top of continuum emission, such as the radio recombination lines discussed in Sect. 9.4.1 or the X-ray emission lines seen in a Solar flare (Figure 6.7). The continuum and line can then be dealt with separately, both providing useful information about the source. Another is if the signal consists of spectral lines that are Doppler-shifted with respect to each other (Sect. 7.2.1.2). Various clouds along a given line of sight can then be separated because their relative velocities prevent them from occurring at the same frequency. HI and molecular interstellar clouds in our own Galaxy can often be isolated this way. A more dramatic example is the separation of intergalactic clouds in the Universe due to the expansion of space, as shown by the Lyman α forest in Figure 5.6.

In spite of every best effort, the fact remains that some signals simply cannot be isolated from others, either because the resolution is insufficient, because there are many signals from a line of sight at the same frequency, or because several types of signals are mixed at the source. For such cases – and there are many – we must *build a model*.

10.1.2 Modelling the signal

The process of building a model spectrum involves guessing at a set of physical parameters for the source, including any constraints that might already be known, calculating a model spectrum using these parameters using equations such as those given in Chapters 8 and 9, and then comparing the model spectrum to the real one.

Input parameters might include temperature, density, source size, fraction of emission due to various processes, or whatever else is required to calculate a spectrum. If the model and real spectra differ from each other, the inputs to the model must be adjusted and the process repeated until the model spectrum matches the real one as closely as possible. If a good match can be found, it is then presumed that the input parameters provide a good description of the source. The problem is often considered to be solved at this point, although concerns about *uniqueness* should still be kept in mind. That is, it is possible that more than one set of input parameters could reproduce the observed spectrum equally well. For a thorough analysis, the set of input parameters will be carefully and sequentially adjusted (this is called 'searching through parameter-space') to see if other combinations might also reproduce the spectrum.

An example of a model was shown in Figure 8.9 for the optical continuum emission from an HII region. This model included contributions from free–bound radiation and two-photon emission, as labelled in the figure. The difference between the calculated spectrum and the observed spectrum was then attributed to scattered light from dust.

A more complex example of modelling is illustrated by Figure 10.1 which shows the total γ-ray emission from the inner region of our Galaxy obtained from a variety of γ-ray instruments (upper data points in the plot). At γ-ray energies, a wider variety of astrophysical processes than have so far been described in this text can occur. From about 10 to 10^5 MeV, the data are well fitted by a model (solid curve) that describes diffuse (as opposed to point source) emission, including contributions from (a) γ-rays produced by neutral pion (π^0) decay[2], where the pions result from collisions between high energy cosmic rays (mostly CR protons) and atomic nuclei of the ISM (mostly protons), (b) Inverse Compton emission (Sect. 8.6) from the scattering of relativistic CR electrons off of stellar photons, (c) Bremsstrahlung emission from relativistic CR electrons scattering from interstellar gas (primarily hydrogen)[3], and (d) a small contribution from extragalactic background (EB) sources (Refs. [164], [165]). In addition, a narrow peak at 0.511 MeV is due to electron–positron annihilation. However, the emission at energies less than 10 MeV are not well fitted by this model. At the lower energies, there are likely contributions to the γ-ray flux from individual point sources which were not included in the model of the diffuse emission. A possibility (Ref. [164]) is *anomalous X-ray pulsars* (AXPs) – pulsars with extremely powerful magnetic fields. One of the problems plaguing γ-ray instruments has been their low spatial resolution (Sect. 2.2.1) which makes it difficult to pinpoint individual sources. This problem will become less important as instrumental resolution improves.

[2]A neutral pion is a subnuclear particle. It is a type of *meson* which is a particle that consists of a *quark–antiquark pair*. *Quarks* (of which there are different types) are the basic 'building blocks' of matter; for example, protons and neutrons are each composed of three quarks, but the mixture of the types of quarks are different for the two particles. The neutral pion is just one type of particle that can result from a high energy collision. It has a rest mass of 135 MeV and its most likely decay, after a mean lifetime of only 8×10^{-17} s, is into two γ-rays.

[3]This is called *relativistic Bremsstrahlung* and the scattering can be off of protons or other nuclei that still have bound electrons. The spectrum of relativistic Bremsstrahlung emission is not like the thermal Bremsstahlung spectrum discussed in Sect. 8.2 due to the relativistic nature of the interaction and because CR electrons do not follow a Maxwell–Boltzmann velocity distribution.

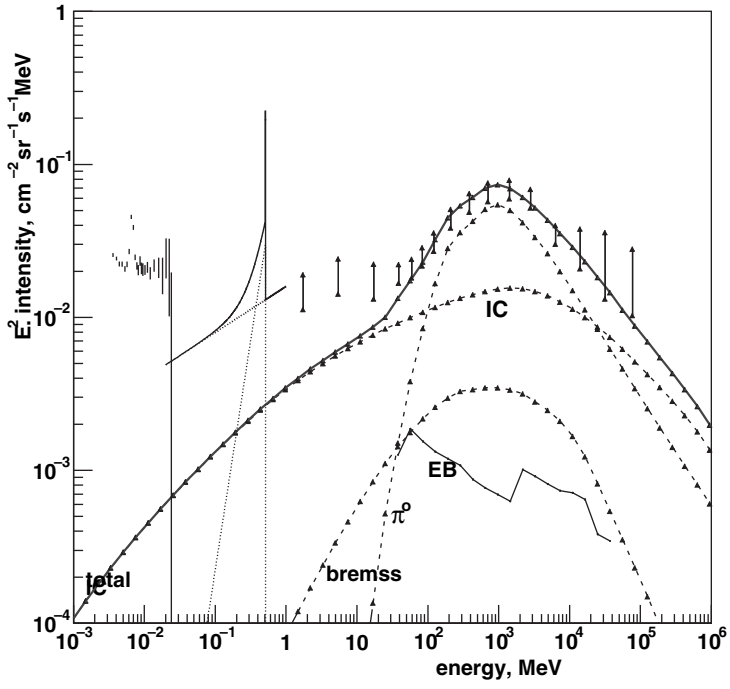


Figure 10.1. The γ-ray spectrum of the inner Galaxy (Ref. [164]) showing data (upper points) from a variety of γ-ray instruments. The 'intensity' (number of photons per unit time per unit area per unit solid angle per unit interval of energy) has been multiplied by E^2 on the ordinate axis. The solid curve shows the total modelled spectrum which fits the high energy part of the observed spectrum and includes contributions from Inverse Compton (IC) emission, neutral pion (π^0) decay emission, relativistic Bremsstrahlung (bremss) emission, and the extragalactic background (EB), each of which is labelled. The low energy points are not well fitted by this model (see text). The data points at energies below about 2×10^{-2} MeV (leftmost on diagram) are somewhat high, likely because they were obtained over a slightly different Galactic longitude range. (Reproduced by permission of Strong, A.W., *et al.*, 2005, A&A, **444**, 495)

To illustrate how the modelling process works, we need to choose a spectral band in which there are fewer emission processes at work than we have just seen in the γ-ray band of Figure 10.1. A good choice is the radio spectrum, for which there are only two main contributors to the continuum in normal galaxies: thermal Bremsstrahlung emission and non-thermal synchrotron emission[4]. Figure 10.2 presents a calculated radio spectrum, typical of a normal spiral galaxy, which shows the total emission as well as both the thermal and non-thermal components that contribute to the total. Throughout most of the plot, the non-thermal component dominates but, at high frequencies, the thermal component dominates. While this radio spectrum is typical for a normal galaxy (see Figure 2.3), the specific fraction of emission contributed by

[4]However, a dust continuum may be present at the highest radio frequencies.

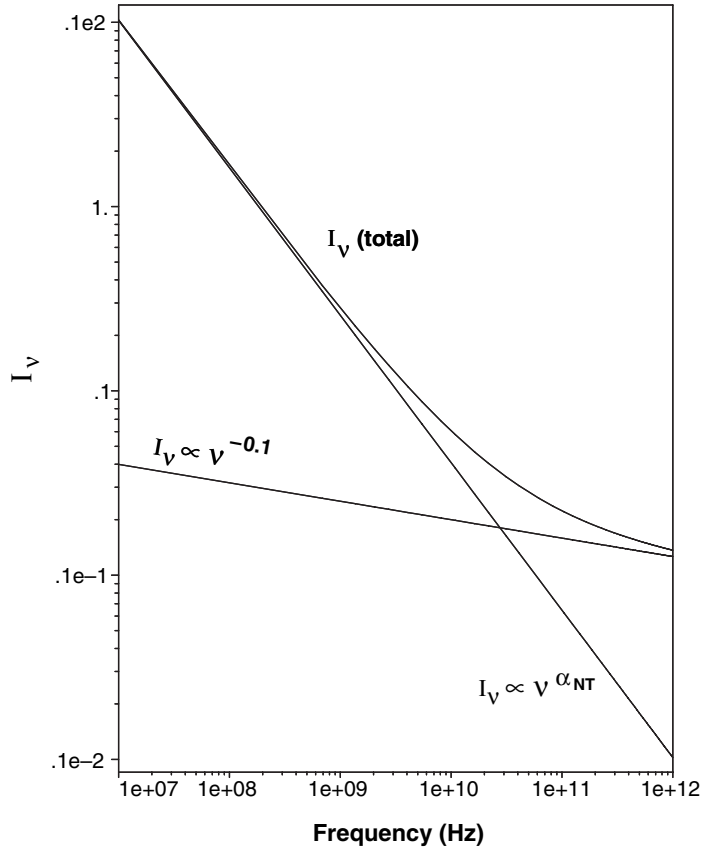


Figure 10.2. A model radio spectrum, showing the total observed emission ($I_\nu(\text{total}) \propto \nu^{\alpha}$), and its two contributors, a thermal Bremsstrahlung component ($I_\nu \propto \nu^{-0.1}$) and a non-thermal synchrotron component ($I_\nu \propto \nu^{\alpha_{NT}}$). The observed spectral index, α, varies with frequency.

thermal and non-thermal components can vary between galaxies and the value of $\alpha_{\rm NT}$, which is related to the cosmic ray energy spectrum (Eq. 8.43), can also differ between galaxies. How, then, can we determine the thermal/non-thermal contributions and extract $\alpha_{\rm NT}$ from the observations? Example 10.1 suggests an approach.

Example 10.1

Assume that the total continuum emission from a galaxy at radio wavelengths consists only of thermal Bremsstrahlung radiation whose spectrum can be described by $I_\nu(\mathsf{T}) = a\,\nu^{-0.1}$ *(Eq.* 8.14*) and non-thermal synchrotron radiation whose spectrum is* $I_\nu(\mathsf{NT}) = b\,\nu^{\alpha_{NT}}$ *(Eq.* 8.42*), both optically thin. Describe a way to determine the constants, a, b, and* α_{NT}*, when only the total emission is measured.*

The total emission can be represented by the sum of the thermal and non-thermal components (see also Figure 10.2),

$$I_\nu(\text{total}) = I_\nu(\text{T}) + I_\nu(\text{NT}) = a\,\nu^{-0.1} + b\,\nu^{\alpha_{\text{NT}}} \tag{10.1}$$

There are three unknowns, a, b, and α_{NT}, and therefore three equations are required. These equations can be set up by measuring $I_{\nu_i}(total)$ at three different frequencies, ν_i, and writing Eq. (10.1) for each frequency. The set of three equations can then be solved for a, b, and α_{NT}. Note that a similar equation could be written for the flux density (Prob. 10.2) since $f_\nu = I_\nu\,\Omega$ (Eq. 1.13) and we take Ω to be constant.

The constant, a, is related to the physical quantities associated with thermal Bremsstrahlung emission via Eqs. (8.13) and (8.14) so, depending on what other information is available, it may then be possible to extract information about the thermal part of the source. Similarly, b and α_{NT} are related to the non-thermal parameters by Eqs. (8.36) and (8.42).

With the help of computer software, this concept of modelling can be extended to 'build' spectra of nebulae, galaxies, or whatever other system is of interest. For example, there are many stars within any given resolution element in a typical optical observation of a galaxy. Extracting knowledge of the admixture of different kinds of stars requires modelling the total spectrum by including different numbers of stars of various spectral types (e.g. Prob. 10.3) until a good match results. This approach is called *population synthesis* and provides important information about the galaxy. For example, if the spectrum can be modelled without including any hot massive stars, then the galaxy is not undergoing any significant star formation at the present time (Sect. 3.3.4). On the other hand, if many hot massive stars are required to describe the spectrum, then the galaxy is an active star-forming system. Extending this modelling concept to include HII regions, dust, or other contributors to the spectrum is also possible and is currently actively pursued by researchers interested in knowing the make up of galaxies.

As a final comment, we note that the process of model-building is not restricted to spectra. The *spatial distribution* of light can also be modelled, something that is routinely carried out over a variety of scales, from planetary systems, to large scale structure in the Universe. Modelling spectra, however, is a particularly productive approach, since there are only a limited number of known emission mechanisms and corresponding spectral behaviour.

10.2 Case studies – the active, the young, and the old

How can we build links between objects and events in astrophysics and what scenarios can be pieced together? A good approach is to focus on specific objects or regions and apply the astrophysical principles that we have already seen. To this end, we now present *case studies* of three very different objects.

10.2.1 Case study 1: the Galactic Centre

Eight kpc from the Sun in the direction of the constellation, Sagittarius (the Archer) is the location of the Galactic Centre (GC, see also Figure 3.3). At optical wavelengths, the GC is obscured by many magnitudes of extinction from dust along the Galactic plane. However, at infrared, radio and hard X-ray wavelengths, which are not grossly affected by dust, it is possible to catch a glimpse of the heart of our Galaxy. 'Galactic Centre' is loosely used to refer to the very nucleus of the Milky Way as well as the vicinity around it, of order a few hundred pc, within which the properties differ markedly from the rest of the Galaxy. Figure 10.3 shows two views at radio wavelengths. These images show a rich field containing many sources. Some sources are unique to the GC, such as the magnetically aligned *threads* and the outflow-related Ω shaped Galactic Centre radio lobe (GCL). Both non-thermal (synchrotron) and thermal (Bremsstrahlung) radio emission are present.

The GC harbours a richer variety of sources and processes than even Figure 10.3 suggests. In this region, there is hard X-ray emission, X-ray and IR flaring, strong organised magnetic fields, high stellar densities, high gas pressures and temperatures,

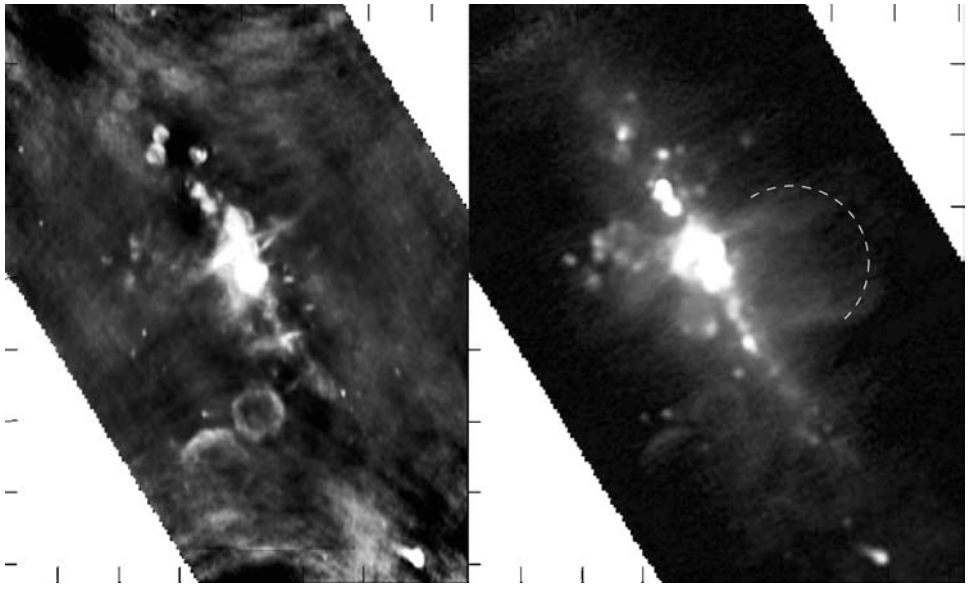


Figure 10.3. The Galactic Centre at two different radio frequencies. In each image, the Galactic plane runs from top left to bottom right. On the left is the λ 90 cm image from the Very Large Array radio telescope in New Mexico, USA (resolution, 2′, Ref. [25]). There are many supernova remnants in this image, as well as narrow *threads* of emission, likely due to aligned magnetic fields. On the right is the λ 3 cm image from the Nobeyama telescope in Japan (resolution, 3′, Ref. [72]). The Galactic Centre radio lobe, with a diameter of about 200 pc, forms an Ω shape perpendicular to the Galactic plane delineated here with a dashed curve. This feature is brighter at λ 3 cm than at λ 90 cm and therefore has a flat spectral index. (Reproduced by permission of C. Brogan)

and high velocity dispersions. Given that the density of the Galaxy, as a whole, increases towards the centre, it is not surprising that the properties of the GC are more extreme than elsewhere in the Galaxy. In addition, however, there is also an 'activity level', as attested to by the presence of very high energy radiation and features, like the GCL, that are suggestive of outflows. Other outflow features, seen over a variety of scales, include bipolar X-ray lobes centred on the nucleus, loops of molecular gas (Ref. [61]), and an extended column of γ-ray emission above the nucleus from electron–positron annihilation referred to as the *annihilation fountain*. Although the origin of some features (such as the annihilation fountain) is still debated, put together, the collective observations imply that the GC harbours an active galactic nucleus (AGN), similar to those observed in other galaxies and discussed at various times throughout this text (e.g. see Figs. 5.5 and 8.16). The AGN is located at the nucleus of the Milky Way – a source called Sgr A*[5]. This is the closest AGN to us and its proximity allows us to study it with a linear resolution that surpasses anything possible for external galaxies.

The AGN paradigm requires the presence of a 'driver' for the high energies that are associated with it. Sometimes referred to as a *central engine*, sometimes a *monster*, the object that is believed to be powering the AGN is a *supermassive black hole*. Although the black hole itself does not emit light, the strong gravitational field around it (see Sect. 7.1) provides the energies required to drive the other observed phenomena (e.g. Sect. 5.4.2). For the Milky Way AGN, a black hole mass of a few million solar masses is implied. What is the evidence for the existence of this black hole? Is it an argument based only on the need for a high energy source or are there any other lines of evidence? Example 10.2 outlines the case.

Example 10.2

Figure 10.4 (Ref. [150]) shows, on the left hand side, a λ 2 μm (near-IR) view of a star field in the GC that is only 2 arcsec on a side. For a distance to the GC of $D = 8$ kpc, the corresponding linear scale is $d = 0.078$ pc (Eq. B.1). Almost all of the stars in this image belong to a dense stellar cluster that is near the Galactic core, since the probability of finding a disk star (one that is closer to us along the line of sight) in such a small field of view is small (Prob. 10.4). A star, S2, is labelled as well as the position of Sgr A*.

The proper motion of S2 has now been measured and, in fact, traced over a 10 year period so that its orbit about Sgr A* has been precisely determined, as shown in the right hand image. The velocity of S2 is highest at closest approach to Sgr A*[6], indicating that the interior mass is concentrated at this position. The true orbit, which has some inclination with respect to the plane of the sky, is an ellipse (Appendix A.5) with Sgr A* at one focus. The apparent (observed) orbit is also an ellipse, but Sgr A* is no longer at the focus of the apparent orbit because of the projection onto the plane of the sky. This fact can be used, with some geometry, to determine the true orbit. For S2, it is found that the inclination

[5]Surprisingly Sgr A* itself is actually underluminous for an AGN of its mass.
[6]Recall Kepler's laws from Sect. 7.2.1.2.

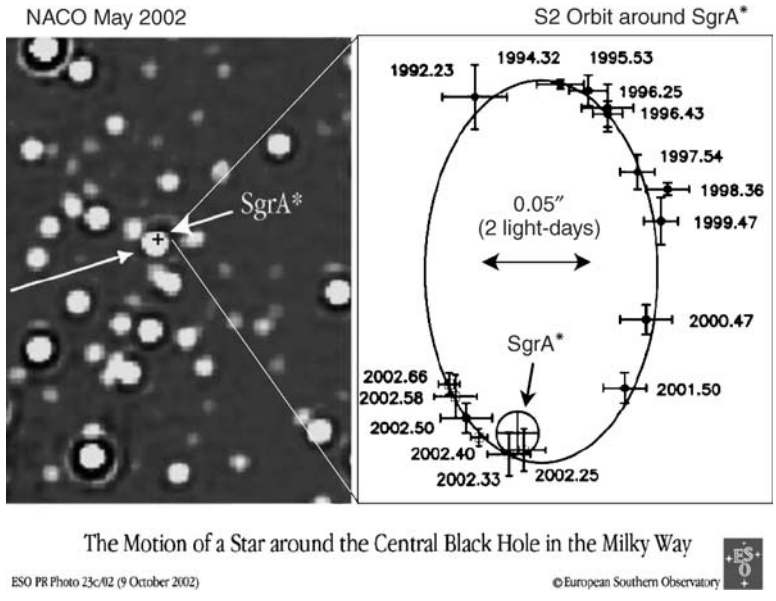


Figure 10.4. Left: the star field around the nucleus of our Galaxy, shown in this 2.1 μm (near-IR) image over a field of view, 2 arcsec on a side with an angular resolution of 0.06 arcsec. The nucleus, located at the position of Sgr A* and a star called S2, whose light is blended with Sgr A*, are both marked. Right: The elliptical orbit of S2 about Sgr A* (as seen on the sky) is shown, after having been observed over the course of 10 yr. Notice that the speed of the orbit is greatest at closest approach. (Reproduced by permission of ESO)

is $i = 46°$, the eccentricity is $e = 0.87$, the semi-major axis is $a = 5.5$ light days $(1.42 \times 10^{16}\,\mathrm{cm})$, and the orbital period is $\mathcal{T} = 15.2$ yr.

The values of a and $\mathcal{T}$ determined from the orbit, together with Eq. (7.5), yield a mass interior to the orbit of, $\mathrm{M} = (3.7 \pm 1.5) \times 10^6\ \mathrm{M}_\odot$. The separation of S2 from the nucleus at closest approach is only 17 light-hours (124 AU). The only known object that could contain this much mass in such a small volume is a supermassive black hole. A stellar cluster or any other conceivable types of object would require a much larger volume.

Since the time of this definitive observation, additional constraints have been placed on the black hole. From Doppler shifts of spectral absorption lines in the orbiting star, the star's radial velocity, v_r, has been determined and, together with the measured transverse motion on the sky, v_t, this has allowed a determination of its true space velocity (see Figure 7.2). In addition, orbits for more stars have now been determined. The result has been a more accurate and precise black hole mass of $\mathrm{M_{bh}} = (3.7 \pm 0.2) \times 10^6\ \mathrm{M}_\odot$ (Ref. [63])[7]. In addition, an X-ray flare has been observed from Sgr A* which showed some variability on a timescale of only 200 s (Ref. [127]). This places a limit on the size of the black hole of $d < 0.4$ AU, via Eq. (7.21). Thus, 4 million Solar masses are within a region that is smaller than the orbit of Mercury! The conclusion that the Milky Way harbours a supermassive black hole at its centre seems inescapable.

[7]The result is distance dependent. For a GC distance of 8.5 kpc (Table G.3), $\mathrm{M_{bh}} = (4.4 \pm 0.2) \times 10^6\ \mathrm{M}_\odot$.

10.2.2 Case study 2: the Cygnus star forming complex

In the constellation of Cygnus (the Swan), within the disk of the Milky Way and about 1.8 kpc distant, is the *Cygnus star forming complex*. A picture of this region, from combined radio and far-infrared images, is shown in Figure 10.5. The enormous angular size of this region is truly impressive. The diameter of the field of view covers fully 10° of sky, about 20 times the diameter of the full moon. If our eyes were sensitive to the wavebands shown, it is clear that the sky would look nothing like the star filled firmament that we see each night.

The Cygnus star forming complex is a vast stellar nursery in which stars are born from dense molecular gas, live their lives, and die, recyling their chemically enriched material back into the interstellar medium. In the process, winds from hot massive stars (e.g. Sect. 5.4.2), radiation, and explosive energy from supernovae compress, heat, ionize, and disrupt the surrounding gas clouds. For a star-forming region in the disk of the Milky Way, this is one of the most active. How can such a picture about this region emerge? For this, we need to draw on a number of resources, including the information contained in Figure 10.5, data from other wavebands, and a knowledge of stellar

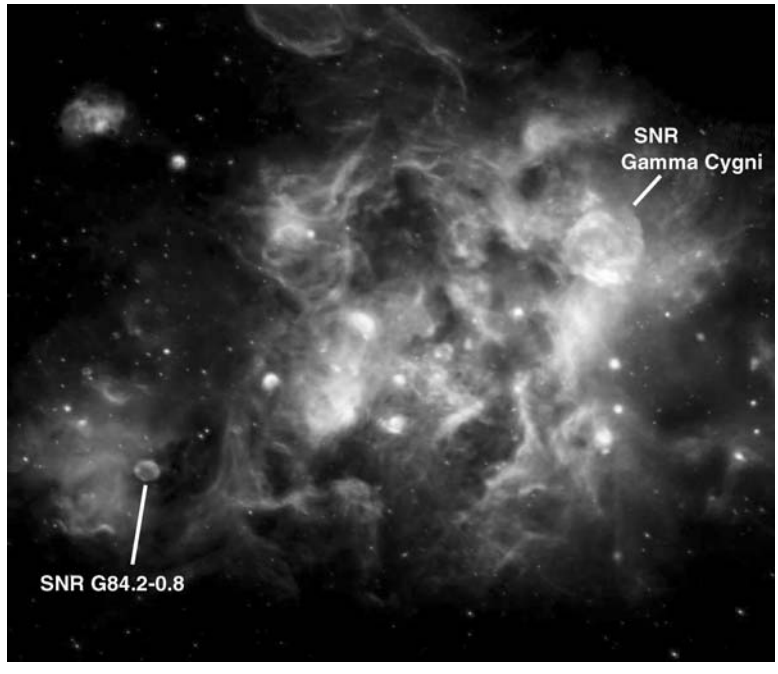


Figure 10.5. A complex star forming region in our Milky Way in the direction of the constellation, Cygnus. The image diameter spans 10° of sky. In this image, the colour rose represents emission at λ74 cm, green is λ 21 cm emission, turquoise is λ60 μm emission, and blue is λ25 μm emission, all continuum emission. Two supernova remnants are labelled. [Credit: Composed for the Canadian Galactic Plane Survey by Jayanne English (CGPS/U. Manitoba) with the support of A. R. Taylor (CGPS/U. Calgary)] (see color plate)

evolution that has been painstakingly pieced together from both observation and theory (see Sect. 3.3). We begin with a discussion of Figure 10.5 given in Example 10.3.

Example 10.3

In Figure 10.5, four different wavelengths of *continuum* emission are represented in *false colour*: the radio wavelengths, $\lambda\,74\,\mathrm{cm}$ $(\nu = 408\,\mathrm{MHz})$ in *rose/red* and $\lambda\,21\,\mathrm{cm}$ $(\nu = 1420\,\mathrm{MHz})$ in *green*, and the far-infrared wavelengths, $\lambda\,60$ μm in *turquoise* and $\lambda\,25\mu\mathrm{m}$ in *blue*. Careful quantitative analysis of the individual images that contribute to Figure 10.5 is needed to properly extract the information that it contains. Nevertheless, it is possible to piece together some information, with just a careful visual examination. The mix of colours of the various sources, for example, provides us with clues about the emission processes, and therefore about the most likely sources.

Any object that is reddish or yellow-red, depending on other emission in the vicinity, must be strongest at the longest radio wavelength, $\lambda\,74\,\mathrm{cm}$. Since optically thin synchrotron emission becomes stronger at longer wavelengths, this suggests that the reddish sources are steep-spectrum, optically thin synchrotron emitting sources. Two such circular features are labelled in the figure. The source, SNR γ Cygni, so named because it lies near the star, γ Cygni, spans 1° of sky. As the name indicates, this is a supernova remnant, which we could infer from its non-thermal radio spectrum and bubble-like shape. A second labelled SNR, G84.2-0.8, is also quite distinct and shows the *edge-brightened* morphology expected from a shell of material.

Throughout the image are also many reddish (in false colour) point sources. These may look like stars, but they are clearly not, since their reddish colour indicates that they emit most strongly at the longest radio wavelength, again consistent with a synchrotron spectrum. Stellar continuum spectra, by contrast, are Planck curves that peak from near-IR to UV wavelengths (Tables G.7, G.8, and G.9, and Eq. 4.8) and stellar emission would be negligible at all wavelengths shown in the figure. The red point sources are more broadly distributed in the image (more obvious in a wider field of view), rather than clustered in the Cygnus star-forming complex, and they are unresolved. This information suggests that these are background sources, likely quasars, in the distant Universe.

As the emission shifts to green $(\lambda\,21\,\mathrm{cm})$ and blue, more blending is seen between colours and sources. The green emission emphasizes the contribution from flat-spectrum thermal Bremsstrahlung emission (see Figure 10.2), implying the presence of ionized gas (HII). In fact, ionized gas is abundant in this complex, both as diffuse emission throughout the region as well as discrete cocoon-like 'knots' in the map. The existence of HII regions requires the presence of hot O and B stars, as described in Sect. 5.2.2, and also means that there is sufficient interstellar gas available to be ionized.

Finally, the emission from the far-infrared wavelengths, shown in turquoise and blue, permeates the region. At these wavelengths, we are seeing collective Planck curves from dust (e.g. Sect. 10.1.1). At $\lambda\,60\,\mu\mathrm{m}$ and $\lambda\,25\,\mu\mathrm{m}$, the characteristic temperatures (Eq. 4.8) are $\mathrm{T} = 48\,\mathrm{K}$ and $\mathrm{T} = 116\,\mathrm{K}$, respectively, both far too cool to represent photospheric emission from stars. Some very blue 'knots' of emission are also seen, possibly very small dusty regions around forming stars.

Thus, in this image alone, we see supernova remnants, abundant interstellar gas and dust, and HII regions. We can also infer the presence of hot, O and B stars, although they are not seen directly. From our knowledge of stellar evolution, we know that HII regions, O and B stars and supernovae are all tracers of recent star formation (Sect. 3.3.4) since the lifetimes of O and B stars are so short (of order 10^6 yr). They should also, therefore, still be associated with the molecular material from which they formed. Moreover, since dust and molecular gas are correlated spatially in the Milky Way (e.g. see Figure 3.20), this also suggests that molecular gas should be present throughout this region.

The rudimentary treatment of Example 10.3 should never take the place of a proper quantitative analysis. Yet, with some careful scrutiny and an understanding as to which emission processes dominate in various wavebands, it is sometimes possible to extract useful information. Are there any other data, then, that support the conclusions of Example 10.3?

The γ Cygni SNR has been detected at X-ray and γ-ray wavelengths (Ref. [173]), confirming the high energy nature of this source. Its expansion velocity has also been measured to be $v_{\rm exp} = 900\,{\rm km\,s^{-1}}$ and the result of the expansion analysis indicates that the supernova exploded some 5000 yr ago. This is a very recent event, by astronomical standards.

The molecular gas, inferred to be present, has been directly observed and is extensive throughout the region in a *giant molecular cloud (GMC) complex*. From CO observations (see Sect. 9.5.1), the estimated total molecular gas mass is $M_{\rm GMC} = 5 \times 10^6\,{\rm M}_\odot$ (Ref. [148]), making it one of the most massive molecular cloud complexes known in our Galaxy.

Finally, the hot O and B stars themselves, assumed to be responsible for the thermal radio emission, have also been directly observed. Several star clusters containing O and B stars, called *OB associations* are embedded in the cloud. One of them, the Cygnus OB2 association, is one of the most massive and richest clusters in the Galaxy, containing 2600 cluster member stars of which about 100 are O or B stars (Ref. [86]). The total mass of this cluster alone is is $4-10 \times 10^4\,{\rm M}_\odot$.

There is little doubt that this complex plays host to a rich array of processes, of which star formation and the evolution of massive stars are key elements. While the details of star formation are not completely understood, complexes like the Cygnus cloud provide astrophysical 'laboratories' that may reveal the secrets of star formation in the future.

10.2.3 Case study 3: the globular cluster, NGC 6397

In the southern hemisphere constellation of Ara (the Altar), located out of the disk of our Galaxy is NGC 6397, one of the closest globular clusters to us ($D = 2.6\,{\rm kpc}$). The globular clusters are spherical in appearance, typically contain a few hundred thousand stars each and have a typical absolute magnitude of $M_{\rm V} = -7.3$. There are about 100

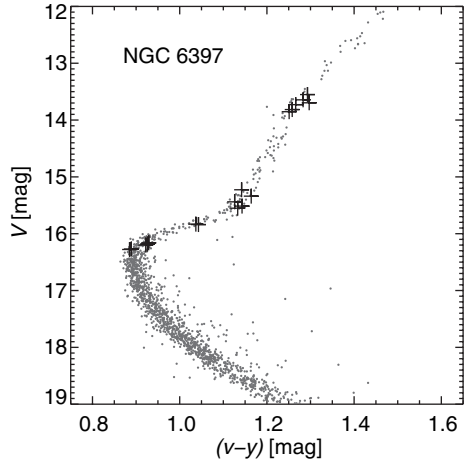


Figure 10.6. Colour-magnitude diagram of the globular cluster, NGC 6397 (Ref. [91]), taken with the Danish 1.54 m telescope in La Silla, Chile. Crosses mark specific stars that were targetted for spectroscopy in this study. (Reproduced by permission of Korn, A., 2006, astro-ph/0610077, The Messenger, **125**, 2006, 6–10)

known that dot the halo region and bulge of our Milky Way (see Figure 3.3). These clusters are believed to be the oldest stellar systems in the Universe, and NGC 6397 is one of the oldest, at an age of 13.4 Gyr (Ref. [67]). This is only marginally less than the age of the Universe (13.7 Gyr, Sect. 3.1), and the two values are in agreement if typical error bars of 0.8 Gyr are taken into account. Thus, these sytems must have formed in the very early Universe and would have been among the first types of objects that took part in the assembly of massive galaxies like the Milky Way. They provide the means of dating the Milky Way, itself, whose age (13.6 Gyr old) is taken to be the age of the oldest clusters plus some allowance of time for the clusters to form. What are the characteristics of this sytem that lead to these conclusions? We begin, in Example 10.4, with a discussion of the colour-magnitude (CM) diagram of NGC 6397, shown in Figure 10.6.

Example 10.4

It is instructive to compare the CM diagram of the globular cluster, NGC 6397 (Figure 10.6), with that of a collection of stars in the gaseous dusty disk of the Milky Way in the Solar neighbourhood (Figure 1.14). If the mix of stars in the globular cluster were the same as in the disk, there should a close resemblance between the figures[8]. What we find, however, is that the two CM diagrams differ significantly from each other.

[8]Note that the filter bands of the two figures differ, but this does not change the argument.

First of all, the globular cluster CM diagram lacks any trace of hot, luminous massive stars which we know, from their short lifetimes, to be young. All stars on the main sequence occur below a *turn-off point* which is the location that the main sequence turns off into the red giant branch (see Sect. 3.3.2). If any recent star formation had been occurring in NGC 6397, then there should be evidence of hot, massive stars in this diagram, and they should have been be easily detected, being more luminous than the other detected stars on the lower main sequence[9]. From a knowledge of stellar evolution, we know how long a star can remain on the main sequence before it exhausts its fuel and turns off to the red giant branch. This fact can be used to determine how long it has been since stars formed in the cluster. Model colour-magnitude diagrams for a range of ages have been constructed from theory and they are then compared to the data, the best-fit result giving the cluster age. This is called *cluster fitting*.

Secondly, the turn-off point is quite well defined and narrow, unlike the broad turn-off region in Figure 1.14. For NGC 6397, the turn-off corresponds to a stellar temperature of $T_{\rm eff} = 6254\,{\rm K}$ (Ref. [91]). If the stars that formed long ago had formed continuously over some period of time, then different stars of different temperatures would now be turning off the main sequence, thickening the turn-off position. A sharp turn-off point, then, indicates that all the stars in the cluster were formed in a relatively short period of time. In other words, the cluster itself has a unique age.

Other evidence, besides the cluster fitting technique described in Example 10.4, also points to the conclusion that globular clusters are old systems in which star formation ceased long ago. There is currently no atomic or molecular gas in globular clusters and therefore no supply of material from which stars could form. Images of NGC 6397 (e.g. Figure 10.7) reveal only a population of old stars and stellar remnants. White dwarfs, for example, are the relics of stars that have long since left the main sequence and come to the end of their nuclear burning lifetimes. A significant population of white dwarfs has been detected in NGC 6397 to an extremely faint magnitude limit (see Figure 10.7). Since the luminosity of a white dwarf declines with time as it cools, the location of the white dwarf *cooling curve* on the CM diagram also allows the cluster to be dated, with consistent results.

Other creative dating techniques have also been applied. For example, observations of the element, beryllium, in cluster members are useful. Beryllium is neither produced in the Big Bang, nor by nucleosynthesis in stars in any significant quantity. The source of its production is by spallation from cosmic rays (Sects. 3.3.4, 3.6.1). Over a period of time, the quantity of beryllium is expected to increase in any gas from which the cluster originally formed. Therefore a measure of the beryllium abundance in the cluster provides a way of timing when the cluster formed, with some assumptions related to calibration. Beryllium absorption lines in the atmospheres of two stars in NGC 6397

[9]This cluster does, however, contain a population of *blue stragglers*, which are stars that are unexpectedly blue, given the age of the cluster (Ref. [10]). The most likely explanation for blue stragglers is that they have been formed by stellar collisions in the dense cores of these clusters.

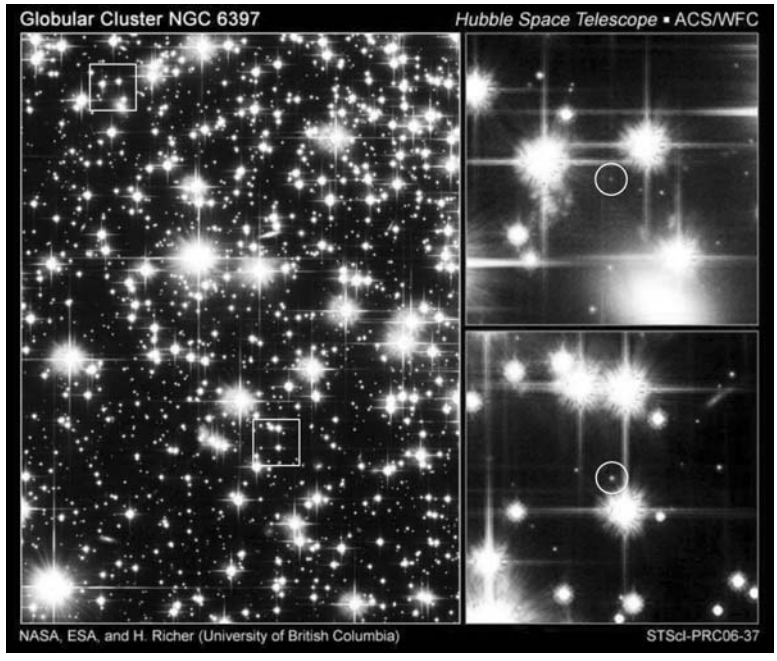


Figure 10.7. Hubble Space Telescope view of the centre of the globular cluster, NGC 6397, shown with exquisite detail. This image reaches the faintest magnitudes ever observed for low mass stars and white dwarfs. Two squares mark fields that are blown up at the right. In the image at the top right, a faint white dwarf star is circled. This white dwarf is at 28th magnitude and is blueish in colour, indicating its hot surface temperature. In the lower right image, a faint red dwarf star, at 26th magnitude, is circled. Close scrutiny will reveal faint background galaxies that are seen through the cluster. [Reproduced by permission of NASA, ESA, and H. Richer (University of British Columbia)]

have now been measured, resulting in an age estimate that is in agreement with the other values (Ref. [118]).

Another important clue as to the age of a globular cluster is its very low metallicity (Sect. 3.3.1). For NGC 6397, for example, the metallicity is only 1/100th of the value observed in the Sun. Since we know that successive generations of star formation will increase the metallicity over time, this implies that globular clusters formed early and that successive generations of stars have not occurred within them. As pointed out in Sect. 3.3.4, however, it is important to note that the metallicity is not zero. This means that some metal enrichment took place even before the formation of the globular clusters. A population of massive stars that rapidly exploded as supernovae could have accomplished this, but the exact nature of this population is unknown[10].

[10] This population is referred to as 'Population III stars'.

All indicators are that there was an early rapid burst of star formation at which time the globular clusters formed, and star formation then ceased within these systems. Current data suggest that the globular cluster formation epoch began within about 1.7 Gyr of the Big Bang, lasted for 2.6 Gyr, and ended 10.8 ± 1.4 Gyr ago (Ref. [67]). The spatial distributions and velocities of globular clusters in the spherical halo of the Milky Way agree with this scenario, pointing to an early formation period in low metallicity gas, before material settled into the spiral disk in which we now *do* see active star formation.

It is clear that globular clusters contain information whose implications go far beyond the boundaries of the cluster, itself. These systems provide clues as to the conditions that were present in the very early Universe at the time that our Milky Way was just forming. NGC 6397 has, in addition, proven to be a laboratory for the study of the lowest mass stars. The faint end of the stellar main sequence in NGC 6397 has been completely detected, showing that the main sequence has a low luminosity cut-off (see Figure 10.7). From these observations, a low mass cut-off of $0.083\,M_{\odot}$ has been determined (Ref. [135]). This is in beautiful agreement with the theoretical low mass limit of $M_{*_{min}} = 0.08\,M_{\odot}$ required for the nuclear burning discussed in Sect. 3.3.2. It also hints that there may be a population, yet to be detected, of sub-stellar brown dwarfs in which nuclear burning has not, and never will, occur.

10.3 The messenger and the message

In 1610, Galileo Galilei published his *Sidereus Nuncius* ('Sidereal Messenger' or 'Starry Messenger') which was the first book to be published based on observations of the sky through a telescope. In it, he described the fact that the Moon is not perfectly spherical but has an irregular surface, the fact that Jupiter has four moons (today called the 'Galilean satellites') which orbit about it, and that many nebulosities, such as the apparently fluid light of the Milky Way, consist of a myriad of stars that are simply unresolved by eye. In this brief report, he upset an Aristotelian world view that held the heavens to be 'perfect' and showed that the Earth was not at the centre of every orbit. Instead of deferring to past authority, he was willing, to paraphrase the Prologue to Chapter 8, to 'see a thing, and say that he sees it'. For this he suffered and, for this, we have gained, for by so doing, he laid the foundations of modern science.

The 'signal' that we have tracked over the scope of this text has not been imbued with meaning, though the meaning may indeed be there for those who seek it. What *has* been assumed, implicitly, is that it is actually possible to measure a signal – a message – and discern something concrete about the objects that have emitted it or altered it along the way. For this, we require theoretical models that, if not perfect, at least provide as close a description as possible of our physical world.

The achievements and power of this approach are impressive. The spins of some electrons in a distant galaxy change their orientation, and we learn that the galaxy's

mass is one hundred billion times that of the Sun. Electrons alter their course slightly while passing by nuclei, and we find that the hot gas in a cluster of galaxies far outweighs the mass of the visible stars. Atoms in the atmospheres of stars show absorption lines that are shifted slightly in frequency, and we conclude that a supermassive black hole lurks at the core of the Milky Way. Slight warps in space–time perturb photons travelling through a cluster of galaxies implying that dark matter dominates all other material. And glimmers of light from the dawn of time hint of the origin and evolution of the Universe on vast cosmological scales. Perhaps in the future, such a world view may be superseded by something even more perfect and profound. For now, the signals that link our detectors to the cosmos continue to yield their secrets.

Problems

10.1 The infrared peak in the spectrum of NGC 2903 is due to the collective contributions of many different types of grains over a variety of temperatures. Figure 2.3, however, suggests that the bulk of the grains might be described by one 'characteristic' temperature. Estimate this temperature from the figure.

10.2 At the three frequencies, 4850 MHz, 2700 MHz, and 1465 MHz, the radio continuum flux densities of a galaxy are, respectively, $f_{4850} = 3.94\,\text{Jy}$, $f_{2700} = 5.37\,\text{Jy}$, and $f_{1465} = 7.69\,\text{Jy}$.

(a) Write three equations in three unknowns that show the explicit contributions of thermal Bremsstrahlung and non-thermal synchrotron radiation to the total emission (see Example 10.1, and note that cgs units are not necessary).

(b) With the help of computer algebra software, solve for the three unknown constants. [Hint: Include the restriction, $\alpha_{\text{NT}} < -0.15$.] Specify the units of each of the constants.

(c) What fraction of the total emission is due to non-thermal emission at $\nu = 4850\,\text{MHz}$ and at $\nu = 1465\,\text{MHz}$?

10.3 With the help of a spreadsheet or other computer software, calculate the shape of the total spectrum of a small galaxy that consists only of three different types of stars: 10^6 M5V stars, 5000 A0V stars, and 20 K0Ib stars (see Tables G.7, G.9). Since we can only measure the flux density of stars, not their specific intensity (e.g. Figure 1.9), the black body curves of the stars need to by multiplied by R^2, where R is the stellar radius [$f_\nu \propto R^2 B_\nu(\text{T})$, Eqs. 1.10, 1.14] in order to see the shape of the spectrum. Plot each of the three stellar components as a function of frequency as well as the total. You may wish to change the admixture of stars to see how the result varies.

10.4 If we take the number density of disk stars in the Milky Way to be approximately $n = 0.1\,\text{pc}^{-3}$, how many disk stars (on average) should be visible in the left image of Figure 10.4?

10.5 Of the distinct features in Figure 10.3, identity one that is likely to be an HII region, and explain why.

10.6 For the following radio sources, indicate whether the emission is due to synchrotron radiation, thermal Bremsstrahlung radiation, or whether it is possible to tell. Provide an explanation.

(a) A source that has a steep, negative power law spectrum.

(b) A source that has a flat spectrum and is polarised.

(c) A source that has a flat spectrum and is not polarised.

10.7 Indicate whether the following lines are more likely to have originated in an HI cloud or an HII region and provide a reason. (When 'absorption' is indicated, assume that the required background source is available.)

(a) The $\lambda\,21$ cm line in absorption.

(b) The $\lambda\,21$ cm line in emission.

(c) The Lyα line in absorption.

(d) The Lyα line in emission.

(e) The Hα line in absorption.

(f) The Hα line in emission.

10.8 Suggest a possible object or objects that could produce spectra with the following characteristics (each part represents a different object or objects):

(a) There are many spectral lines in the mm-wave part of the spectrum. The underlying continuum shows a slope that suggests a Planck curve which peaks in the infrared.

(b) The spectrum contains a broad Planck-like peak in the near-infrared and optical. No other significant continuum is observed.

(c) The optical emission resembles that of a star and is polarised.

(d) The X-ray continuum shows a declining spectrum with increasing frequency. Some X-ray emission lines are seen.

(e) The radio spectrum is flat and has emission lines.

(f) An emission line is observed at a wavelength of $\lambda\,21.232$ cm.

(g) Numerous optical absorption lines are observed against a Planck continuum.

10.9 Using the concept of minimum energy (see Sect. 8.5.3), 'hot spots' in the radio lobes of Cygnus A have been found to have lifetimes (at $\nu = 89$ GHz) of $t = 3 \times 10^4$ yr. From the information given in Figure 8.16, determine whether relativistic particles in the hot spots

could have been accelerated at the nucleus of the galaxy or whether they were accelerated in the lobe itself.

10.10 Visit the *Astronomy Picture of the Day* (APOD) website at http://antwrp.gsfc.nasa.gov/apod/astropix.html and, as we did in Example 10.3, attempt to identify the type of emission that is displayed in an image that interests you. Some questions to ask might be:

(a) Does this picture represent a single waveband or is it a composite?

(b) Does the image show continuum or line emission?

(c) Is the emission thermal or non-thermal in nature?

(d) If a spectrum in the displayed waveband were available to you, what information could you obtain about the source?

10.11 Visit the *Multi-wavelength Milky Way* website at http://adc.gsfc.nasa.gov/mw/. For each of the wavebands shown, (a) identify the important emission mechanism(s), and (b) provide examples of the types of sources that contribute to the observed emission.

10.12 A point source of magnitude, $V = 18$, is observed in a galaxy at a distance of $D = 30$ Mpc. Is this source an individual star, a globular cluster, or a supernova?

Appendix A
Mathematical and Geometrical Relations

The German mathematician Carl Friedrich Gauss suggested planting avenues of trees in the barren plains of Siberia to form a giant right-angled triangle. He was attempting to signify to any watching Martians that there was life on Earth intelligent enough to appreciate the wonders of geometry.

–*The Science of Secrecy,* by Simon Singh

A.1 Taylor series

A Taylor series is an expansion about a point, a,

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a) + \ldots \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots \qquad \text{(A.1)}$$

where $f'(x) \equiv \frac{df}{dx}$, $f''(x) \equiv \frac{d^2f}{dx^2}$, etc. If $a = 0$, then the expansion is called a Maclaurin series.

A.2 Binomial expansion

For any value of p, integer or non-integer, positive or negative, one can write,

$$(a+x)^p = a^p + p\,a^{(p-1)}\,x + \frac{p\,(p-1)}{2!}\,a^{(p-2)}\,x^2 + \ldots + x^p \qquad \text{(A.2)}$$

For example, if $x \ll 1$, then $(1+x)^{1/2} \approx 1 + \frac{1}{2}x$

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

A.3 Exponential expansion

$$e^x = 1 + x + \frac{1}{2!}\,x^2 + \ldots + \frac{1}{n!}\,x^n \tag{A.3}$$

A.4 Convolution

When two functions are *convolved* together, they are effectively 'blended' with each other. Convolutions arise in a wide variety of situations in which the signal is observed with an instrument that imposes some limitation upon what can be detected. For example, we can consider the situation in which a source on the sky is observed with an optical telescope. The true brightness distribution of this source is arbitrarily sharp since it is not hindered by the spatial resolution of our instruments. When we observe it, however, the true brightness distribution is convolved, at every position on the sky, with the point spread function (PSF, see Sect. 2.3.2) of the telescope, normalized so that the total flux of the source is preserved. The crisp image, such as that shown in Figure 2.11a, becomes blurred, as suggested in Figure 2.11b, because of this convolution. Thus, convolution 'smooths' the source brightness distribution.

Mathematically, if $h(x)$ is the convolution of the two functions, $f(x)$ and $g(x)$, in one dimension, x, then,

$$h(x) = f(x) * g(x) \equiv \int_{-\infty}^{\infty} f(u)\, g(x-u)\, du \tag{A.4}$$

the operator, $*$, meaning convolution. Convolutions are commutative, such that, $f(x) * g(x) = g(x) * f(x)$. The result of the convolution, $h(x)$, is a function of position along the one-dimensional coordinate (x). The right hand side of Eq. (A.4) indicates that one of the functions, say the one-dimensional PSF, is shifted across the one-dimensional source brightness distribution with a multiplication at each shift and then the result is integrated over all shifts. For a real situation in which there are two coordinates, x and y, the convolution must be carried out in two-dimensions.

A.5 Properties of the ellipse

It is useful to review the properties of an ellipse since orbital motion follows the path of an ellipse. In addition, circular disks, such as that of a spiral galaxy or proto-planetary disk, become ellipses when projected onto the sky. The resolution (beam) of a diffraction-limited telescope (see Sect. 2.2.4) may also be elliptical on the sky.

An ellipse is defined as that set of points that satisfies the equation, $r + r' = 2\,a$, where a is a constant called the *semi-major axis*. The distances, r and r' are distances measured from a point in the set to the two *focal points*, $\mathcal{F}$ and $\mathcal{F}'$ of the ellipse,

Table A.1. Properties of the ellipse

Description	Expression
Area of the ellipse	$A = \pi\, a\, b$
Relation between a and b	$b^2 = a^2\,(1 - e^2)$
Distance of closest approach to $\mathcal{F}$[a]	$r_{\min} = a\,(1 - e)$
Distance of farthest approach to $\mathcal{F}$[a]	$r_{\max} = a\,(1 + e)$
Distance from a point on the curve to $\mathcal{F}$[b]	$r = a(1 - e^2)/(1 + e\cos\theta)$

[a] For an object orbiting the Sun, $r_{\min}$ is called *perihelion* and $r_{\max}$ is called *aphelion*. Similarly, closest approach to a star would be *periastron* and farthest is *apastron*, or for a galaxy, *perigalacticon* and *apogalacticon*, respectively.
[b] The angle, θ, is measured around the ellipse from $\theta = 0$ at $r = r_{\min}$.

respectively. In the case of orbital motion in which a small mass, m is orbiting a large mass, M, one focal point, $\mathcal{F}$, will be at the location of M and the other focal point $\mathcal{F}'$ will be empty. The distance of each of the two focal points from the centre of the ellipse is the quantity, ae, where e is called the *eccentricity*. If $e = 0$, then the ellipse reduces to a circle and the two focal points become coincident at the centre of the circle. If $e = 1$ then the ellipse 'opens up' and becomes a parabola. Thus, $0 \leq e < 1$ is a requirement for an ellipse, the circle being a special case, with highly elongated ellipses having larger values of e. Table A.1 provides additional properties.

Appendix B
Astronomical Geometry

Astronomy lends itself well to the use of spherical coordinates. For example, stars and planets are spheres and many other sources are approximated as spheres. The sky itself is a spherical 'dome', and photons from astrophysical sources typically escape in all directions, passing through the imaginary surfaces of spheres. Thus, it is important to understand this geometry.

B.1 One-dimensional and two-dimensional angles

The arc length of a circle, s, is related to the angle, θ, via (see Figure B.1a),

$$s = r\,\theta \tag{B.1}$$

for θ in radians and s in the same units as r. Over an entire circle, θ subtends $2\,\pi$ radians so the full circumference of the circle is $s = 2\,\pi\,r$. When measuring astronomical sources from Earth, distances are so large that an arc length, s, and line segment, d, are virtually indistinguishable for small θ (Figure B.1a, Prob. 1.2) so Eq. (B.1) can be used to determine the projected linear size of a source if the distance is known in most cases. By 'projected', we mean projected onto the plane of the sky as shown, for example, in Figure B.1b.

By analogy, the surface area of a section of a sphere, σ, is related to the two-dimensional *solid angle*, Ω, via,

$$\sigma = r^2\,\Omega \tag{B.2}$$

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978 0-470-01305-2(HB) 978-0-470-01306-9(PB)

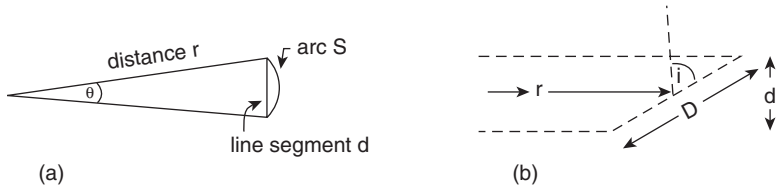


Figure B.1. **(a)** For angles, θ, that are very small (here exaggerated), the arc length, s, is equal to the linear segment, d, in the plane of the sky. **(b)** Example of a non-spherical source and its projection in the plane of the sky. In this case, one dimension of a flat object such as a galaxy projects into the sky plane such that $d = D \cos i$, where i is the angle of inclination.

for Ω in units of *steradians* (sr) and σ in the same units as r^2. For a given surface area (σ constant), it is clear that the solid angle falls off as $\frac{1}{r^2}$. Over an entire sphere, Ω subtends $4\,\pi$ sr so the full surface area of the sphere is $\sigma = 4\,\pi\,r^2$. Again, if Ω is small, σ approximates the projected surface area of a real object on the sky.

For spherical sources which are circular in projection and subtend a small angle on the sky, a useful expression relating the one-dimensional angle subtended by the source, θ, to its two-dimensional solid angle, Ω, is easily obtained. Since the area of the circle on the sky is $\sigma = \pi(\frac{d}{2})^2 \approx \pi(\frac{s}{2})^2$, combining this with Eqs. (B.1) and (B.2) yields,

$$\Omega = \frac{\pi}{4}\,\theta^2 \tag{B.3}$$

for θ in radians and Ω in steradians. If the source in the sky is elliptical rather than circular, in projection, then,

$$\Omega = \frac{\pi}{4}\,\theta_1\,\theta_2 \tag{B.4}$$

where θ_1 and θ_2 are the angular major and minor axes, respectively, of the ellipse.

B.2 Solid angle and the spherical coordinate system

Cartesian (x, y, z) and spherical (r, θ, ϕ) coordinate systems are shown in Figure B.2 where the unit sphere is shown. The origin of the spherical coordinate system can, of course, be placed wherever needed, for example, at the centre of a star, the surface of a star, or at a detector on the Earth. The coordinate angle, θ, is customarily measured downwards from the positive z direction. If the origin is placed at the surface of the Earth, then the positive z direction is towards the zenith and the angle, θ, is called the *zenith angle*. In spherical coordinates, an infinitesimal solid angle is,

$$d\Omega = \sin\theta\,d\theta\,d\phi \tag{B.5}$$

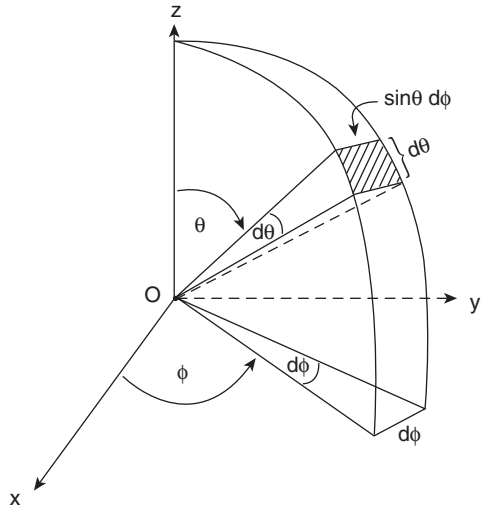


Figure B.2. The spherical coordinate system (r, θ, ϕ) showing the unit sphere ($r = 1$), the angles, θ, ϕ, and their infinitesimal increments, $d\theta$ and $d\phi$, respectively. The origin is at *O*. An infinitesimal solid angle, $d\Omega$ is shown as the hatched region with sides of length, $d\theta$ and $\sin\theta\, d\phi$.

which relates the solid angle to the individual independent angles, θ and ϕ. Eq. (B.5) is especially helpful in integrations over angles that are large. We can check to ensure that an integration of $d\Omega$ over all angles yields 4π sr, for example,

$$\int_{\Omega} d\Omega = \int_0^{2\pi} \Big[\int_0^{\pi} \sin\theta\, d\theta \Big] d\phi = \int_0^{2\pi} \Big[-\cos\pi + \cos 0 \Big] d\phi = \int_0^{2\pi} 2\, d\phi = 4\pi \tag{B.6}$$

Appendix C
The Hydrogen Atom

C.1 The hydrogen spectrum and principal quantum number

Hydrogen is the most abundant element in the Universe (see Sect. 3.3) and is also the simplest, with a single proton for the nucleus and a single bound electron in its neutral, non-isotopic form. A useful description of this atom can be obtained by first assuming that the electron is in uniform circular orbit about the nucleus. The Coulomb force holding the electron to the nucleus is thus balanced by the centripetal force of the electron's motion,

$$F = \frac{e^2}{r^2} = \frac{\mathrm{m_e}\, v^2}{r} \tag{C.1}$$

where r is the separation between the electron and the proton, $\mathrm{m_e}$ is the electron mass, e is the charge on the electron/proton, and v is the electron's speed.

Classically, since any accelerating charged particle should radiate, one would expect continuous emission from such an orbiting electron. However, this is not observed. Instead, discrete emission lines at specific frequencies are seen. This puzzle led to the development of *quantum mechanics*, the simplest approximation of which is the *Bohr model* of the atom. In this model, the *angular momentum* of a bound electron, L, is *quantized*, meaning that it can take only discrete, rather than continuous, values. The electron would not radiate when in this allowed state. Radiation results only when an electron makes a transition from one state to another. As strange as these assumptions seem, they describe atoms and their observed spectra very well.

The quantized angular momentum is given by,

$$L_\mathrm{n} = \mathrm{m_e}\, v\, r_\mathrm{n} = \mathrm{n}\, \frac{h}{2\,\pi} \tag{C.2}$$

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

where h is Planck's constant and n is an integer called the *principal quantum number*[1]. Combining Eqs. (C.1) and (C.2) and solving for $r_{\rm n}$ yields the quantized radius,

$$r_{\rm n} = {\rm n}^2 \, \frac{h^2}{4\,\pi^2\,{\rm m_e}\,e^2} = {\rm n}^2\,a_0 = {\rm n}^2\,5.3\times 10^{-9} \qquad {\rm cm} \tag{C.3}$$

where a_0 is called the *Bohr radius* (Table G.2) and gives the size of the hydrogen atom when n = 1, i.e. when the atom is in its *ground state*.

The energy of a bound electron is the sum of its kinetic and potential energies. For a non-quantized case,

$$E = E_k + E_p = \frac{1}{2}\,{\rm m_e}\,v^2 - \frac{e^2}{r} \tag{C.4}$$

where we have used the usual convention that the potential energy is zero at a separation of infinity. Putting Eqs. (C.2) and (C.3) into Eq. (C.4) yields the quantized energy,

$$E_{\rm n} = -\frac{1}{{\rm n}^2}\,\frac{{\rm m_e}\,e^4\,2\,\pi^2}{h^2} = -\frac{2.18\times 10^{-11}}{{\rm n}^2} \quad {\rm erg} = -\frac{13.6}{{\rm n}^2} \quad {\rm eV} \tag{C.5}$$

Thus the energy of a bound electron is negative. From Eqs. (C.3) and (C.5), it is clear that quantization of angular momentum implies quantization of the radius and energy as well.

When an electron makes a transition from a higher to lower energy state, then a photon is emitted that carries away the excess energy. The energy difference between states is,

$$\Delta E = E_{\rm n'} - E_{\rm n} = -13.6 \quad \left(\frac{1}{{\rm n'}^2} - \frac{1}{{\rm n}^2}\right) {\rm eV} \tag{C.6}$$

where n′ is the higher energy quantum level and n is the lower level. If a transition is from a higher to lower energy state (ΔE positive), then a photon is emitted and an *emission line* is observed. If an atom is perturbed by a collision or a photon which has the energy of an allowed transition, then this energy can be absorbed by the electron and it will make a transition from a lower to a higher energy state (ΔE negative). An *absorption line* may then be observed. More information as to when emission and absorption lines are observed is given in Sect. 6.4.2.

Equating the energy difference of Eq. (C.6) to the energy of an emitted photon $\Delta E = h\,\nu = h\,c/\lambda$, and evaluating the constants leads to a photon frequency and wavelength, respectively, given by,

$$\begin{aligned} \nu &= 3.29\times 10^{15}\left(\frac{1}{{\rm n}^2} - \frac{1}{{\rm n'}^2}\right) \quad {\rm Hz} \\ \frac{1}{\lambda} &= R\left(\frac{1}{{\rm n}^2} - \frac{1}{{\rm n'}^2}\right) \quad {\rm cm^{-1}} \end{aligned} \tag{C.7}$$

[1]The same equation results by expressing the electron mass in terms of its de Broglie wavelength (Table 1.1) and requiring an integral number of wavelengths around a circular orbit, i.e. $n\,\lambda = 2\pi\,r_n$.

where R is the *Rydberg constant for hydrogen*,

$$R = R_\infty \left(\frac{\mathrm{M}}{\mathrm{M} + \mathrm{m_e}}\right) = 109\,677.6 \quad \mathrm{cm}^{-1} \tag{C.8}$$

where M is the mass of the nucleus (for hydrogen this is just the mass of a proton, $\mathrm{M} = \mathrm{m_p}$) and R_∞ (Table G.2) is the Rydberg constant for an 'infinitely' heavy nucleus in comparison to the electron.

Table C.1. Line data for the hydrogen atom[a]

Designation	Transition[b] *i–j*	$\lambda_{\mathrm{i,j}}(\mathrm{nm})$[e]	$f_{i,j}$[c]	$A_{\mathrm{j,i}}$[d] (s^{-1})
Ly α[f]	1–2	121.567	0.4162	4.699×10^8
Ly β	1–3	102.572	0.07910	5.575×10^7
Ly γ	1–4	97.254	0.02899	1.278×10^7
$\mathrm{Ly_{limit}}$	1–∞	91.18		
H α	2–3	656.280	0.6407	4.410×10^7
H β	2–4	486.132	0.1193	8.419×10^6
H γ	2–5	434.046	0.04467	2.530×10^6
H δ	2–6	410.173	0.02209	9.732×10^5
H ϵ	2–7	397.007	0.01270	4.389×10^5
$\mathrm{H_8}$	2–8	388.905	0.008036	2.215×10^5
$\mathrm{H_{limit}}$	2–∞	364.6		
P α	3–4	1875.10	0.8421	8.986×10^6
P β	3–5	1281.81	0.1506	2.201×10^6
P γ	3–6	1093.81	0.05584	7.783×10^5
$\mathrm{P_{limit}}$	3–∞	820.4		
B α	4–5	4051.20	1.038	2.699×10^6
B β	4–6	2625.20	0.1793	7.711×10^5
B γ	4–7	2165.50	0.06549	3.041×10^5
$\mathrm{B_{limit}}$	4–∞	1458.4		
H109α[g]	110–109	5.985 cm		7.0×10^{-4}
HI	1–1[h]	21.106114 cm		2.876×10^{-15}
Deuterium I	1–1[h]	91.5720 cm		4.65×10^{-17}

[a] From Ref. [44] unless otherwise indicated.
[b] The upper level is j and the lower level is i, where i and j are indices representing the principal quantum numbers, unless otherwise indicated.
[c] Absorption oscillator strengths (see Appendix D.1.3) for transitions from lower level i to upper level j.
[d] Average Einstein A coefficient for transitions from upper level j to lower level i. The average value means that the particles are assumed distributed in their substates (determined by the orbital angular momentum quantum number, l) according to the statistical weights of those substates.
[e] Units are nm unless otherwise indicated.
[f] The Einstein A coefficient for the permitted transition is $A_{2p\rightarrow 1s} = 6.27 \times 10^8\ \mathrm{s}^{-1}$ (Ref. [50]). See Sect. 8.4 for a discussion of the forbidden transition.
[g] The wavelength is from Eq. (C.7) and the Einstein A coefficient from $A_{\mathrm{n+1}\rightarrow\mathrm{n}} = 1.167 \times 10^9/\mathrm{n}^6$ Ref. [50].
[h] Transition between the hyperfine splitting of the ground state.

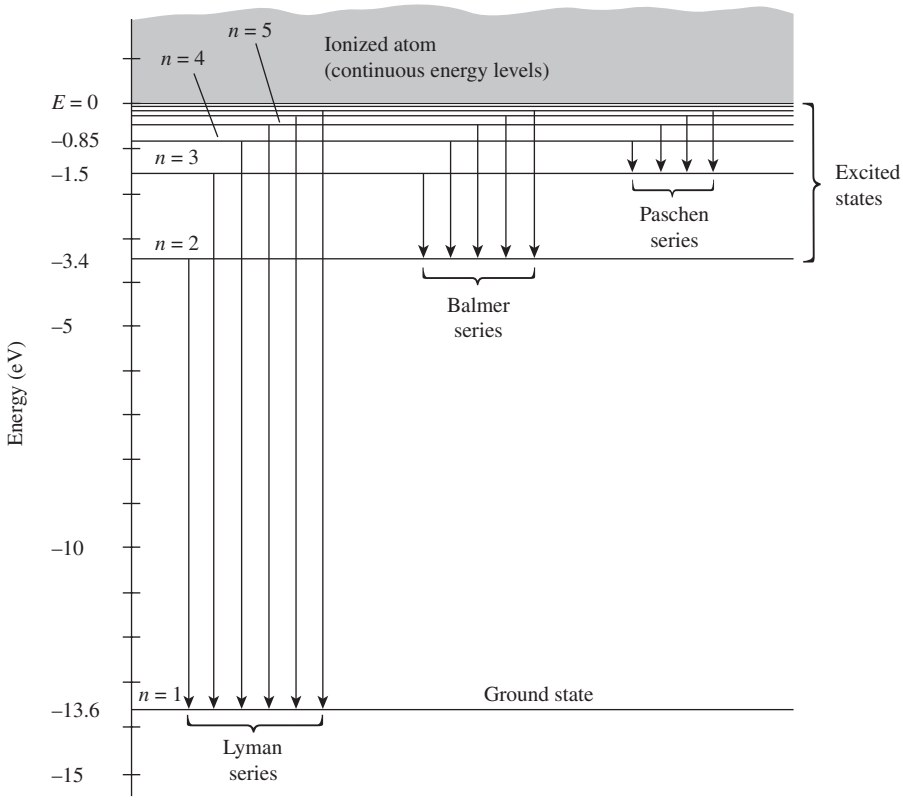


Figure C.1. Energy level diagram for hydrogen, showing the various series seen in the hydrogen spectrum.

Eq. (C.7) predicts the spectrum of the hydrogen atom, some lines of which are listed in Table C.1 and an energy level diagram showing a number of transitions is shown in Figure C.1. Table C.1 also gives the oscillator strength (see Appendix D.1.3), as well as the Einstein *A* coefficient for the transition. The Einstein *A* coefficient indicates the spontaneous downwards transition rate that is expected as a result of intrinsic properties of the atom. The inverse of this quantity indicates the time that the electron can remain in the higher energy level before spontaneously de-exciting in the absence of any external perturber, either radiative or collisional. Series of transitions are named according to a common lower energy level. For example, the *Lyman series*, designated *Ly* in Table C.1, denotes transitions between the ground (n = 1) state and all higher levels. The *Balmer series*, denoted H, famous because the lines occur mainly in the visible part of the spectrum, indicate transitions between n = 2 and higher levels. Other series are the Paschen (P, n = 3) and Bracket (B, n = 4) series. The transitions between closest principal quantum numbers is called α, the next up, β, and so on. Thus, H α corresponds to the transition between n = 3 and n = 2. This Greek letter nomenclature continues to very high energy states where the series is designated simply by H and the principal quantum number of the common lower state. For example, H50α and H250β would designate transitions between n = 51 and n = 50, and between n = 252 and n = 250, respectively.

At higher values of n, the energy levels become closer together (Figure C.1, Table C.1, Eq. C.7) and transitions between more closely spaced energy levels result in photons of lower energy and lower corresponding frequency. Thus, transitions such as the H250α line will be in the radio part of the spectrum, the Brackett series results in infrared emission, the Balmer series lines are in the optical and Lyman series in the ultraviolet. The highest possible energy transition is between $\mathrm{n} = 1$ and $\mathrm{n}' = \infty$, essentially the continuum rather than a bound state. From Eq. (C.6), it can be seen that this energy is 13.6 eV. This is also, therefore, the energy required to ionize the hydrogen atom. The threshold photoionization cross-section (i.e. the value applicable to a photon with *just* sufficient energy for ionization) for hydrogen from a quantum level, n is given by (Ref. [44]),

$$\sigma_{\mathrm{H}\rightarrow\mathrm{HII}} = \frac{8\,h^3\,\mathrm{g\,n}}{3\sqrt{3}\,\pi^2\,\mathrm{m_e}^2\,c\,e^2} = 7.907\times10^{-18}\,\mathrm{n\,g}\quad \mathrm{cm}^2 \tag{C.9}$$

where g is called the *Gaunt factor* given in Table C.2. This value gives the effective area that the hydrogen atom presents to an incoming ionizing photon.

Table C.2. Quantum numbers for the first five energy levels of hydrogen

n[a]	l[b]	m_l[c]	m_s[d]	Designation[e]	Number of possible states[f]	g_n[g]	g[h]
1	0	0	$\pm1/2$	1*s*	2	2	0.80
2	0	0	$\pm1/2$	2*s*	2	8	0.89
	1	−1, 0, 1	$\pm1/2$	2*p*	6		
3	0	0	$\pm1/2$	3*s*	2	18	0.92
	1	−1, 0, 1	$\pm1/2$	3*p*	6		
	2	−2...2	$\pm1/2$	3*d*	10		
4	0	0	$\pm1/2$	4*s*	2	32	0.94
	1	−1, 0, 1	$\pm1/2$	4*p*	6		
	2	−2...2	$\pm1/2$	4*d*	10		
	3	−3...3	$\pm1/2$	4*f*	14		
5	0	0	$\pm1/2$	5*s*	2	50	0.95
	1	−1, 0, 1	$\pm1/2$	5*p*	6		
	2	−2...2	$\pm1/2$	5*d*	10		
	3	−3...3	$\pm1/2$	5*f*	14		
	4	−4...4	$\pm1/2$	5*g*	18		

[a]Principal quantum number. This is an integer starting at 1.
[b]Orbital angular momentum quantum number. $l = 0, 1 \ldots (\mathrm{n}-1)$, where ... means 'add 1'.
[c]Magnetic quantum number, $m_\mathrm{l} = -l \ldots + l$.
[d]Spin quantum number, $m_\mathrm{s} = \pm1/2$.
[e]Designation of the state for a given l. The number specifies the principal quantum number and the letter designates the value of l, where the letters are *s*, *p*, *d*, *f*, and *g* for $l = 0,\ 1,\ 2,\ 3,$ and 4, respectively.
[f]The number of possible states for a given l. Since there are $2\,l + 1$ values of m_l and two possible spins for each, there are $2(2\,l + 1)$ states for any l.
[g]Statistical weight for the energy level. This is the number of possible states for a given n, $g_\mathrm{n} = \sum_{l=0}^{n-1} 2(2\,l + 1) = 2\,\mathrm{n}^2$.
[h]Gaunt factor for the level (Ref. [44])

C.2 Quantum numbers, degeneracy, and statistical weight

The Bohr model is adequate to explain the energy levels of the hydrogen atom and Eq. (C.7) is a good predictor of the observed spectral lines listed in Table C.1. However, deeper scrutiny reveals some fine structure that cannot be explained using the principal quantum number alone. In addition, if the atom is placed within a magnetic field, the energy levels split into sublevels. To explain this behaviour, a full quantum mechanical development is required in which four quantum numbers are needed to describe the state of the electron: the principal quantum number, n, the orbital angular momentum quantum number, l, the magnetic quantum number, m_l, and the electron spin quantum number, m_s. These numbers and their relationships to each other are given for the first five energy levels of hydrogen in Table C.2.

The orbital angular momentum of the electron is now given by,

$$L = \sqrt{l(l+1)}\,\frac{h}{2\pi} \tag{C.10}$$

instead of Eq. (C.2) of the Bohr model. Rather than considering the electron to be 'orbiting' the nucleus, l gives the three-dimensional shape of a probability distribution about the nucleus within which the electron can be found. The electron spin, which can either be $+1/2$ ('up') or $-1/2$ ('down'), represents the angular momentum of the electron alone. Although it may be helpful to visualise electron spin in the classical sense like a spinning top, m_s should more accurately be thought of simply as an intrinsic property of the electron. The magnetic quantum number indicates the quantized projection of the angular momentum vector along a preferred axis, z, the latter being that of an externally applied magnetic field.

A set of quantum numbers that can be used instead of those of Table C.2 are n, l, j, and m_j, where n and l are the same as before, but j is the total angular momentum quantum number and m_j is the magnetic quantum number corresponding to j. The quantum number, j, includes both orbital angular momentum and the electron's spin together, and m_j is the projection of j onto the z axis[2]. This system does not contain any new information beyond n, l, m_l, and m_s and can be derived from this set of quantum numbers. However, it can be useful to consider the electron's orbital angular momenum and spin angular momenum together, especially when considering fine structure, as we do in the next section.

Table C.2 shows that the electron can be in a number of possible states for any given principal quantum number. The number of states at a given n is called the *statistical weight* or the 'degeneracy' of the level and is given by,

$$g_\mathrm{n} = 2\,\mathrm{n}^2 \tag{C.11}$$

[2]The quantun number, j takes values $j = l \pm 1/2$, unless $l = 0$ in which case $j = 1/2$. The quantum number, m_j takes values from $-j$ to $+j$ resulting in $2j + 1$ values.

The *Pauli Exclusion Principle* states that no two electrons can have the same set of quantum numbers. Thus, if there were two electrons in the atom, such as in the negative hydrogen ion, H^-, which is found in abundance at the surface of the Sun, then they cannot occupy the same state.

A spectral line is thus formed when the electron makes a transition from one set of quantum numbers into another. However, not every possible change in quantum number can occur or can occur with equal probability (see $A_{i,j}$ of Table C.1). Certain *selection rules* apply, for example, $\Delta l = \pm 1$. Under certain circumstances, however, highly *forbidden* lines can be seen, as we shall see in Sect. C.4. There are no selection rules for Δ n, hence, all possible Δ n are permitted.

C.3 Fine structure and the Zeeman effect

Fine structure is the splitting of energy levels by a small amount due to relativistic effects as well as the fact that electron spin couples with the orbital angular momentum, forming a total angular momentum vector with quantum number, j. This leads to a modification of the energy equation (Eq. C.5) by a small amount of order $m_e\, c^2\, \alpha^4$, where $\alpha = \frac{1}{137}$ is the dimensionless *fine structure constant*. The result is a *lowering* (more negative) of the energy of the n = 1 level ($j = 1/2$) by,

$$\Delta E = \frac{m_e\, c^2\, \alpha^4}{8} = 2 \times 10^{-4} \quad \text{eV} \tag{C.12}$$

and a lowering of the $j = 1/2$ and $j = 3/2$ states in the n = 2 level by different amounts resulting in a splitting of this level into two sublevels separated by,

$$\Delta E = \frac{m_e\, c^2\, \alpha^4}{32} = 4.5 \times 10^{-5} \quad \text{eV} \tag{C.13}$$

(Ref. [100]).

If the atom is placed in a magnetic field, B, with direction, z, the total angular momentum vector precesses about the field direction, called classically *Larmor precession*, and there is a *magnetic moment* in the atom whose quantization is given by, $\mu = \sqrt{j(j+1)}\,\mu_B$, where $\mu_B = \frac{e}{2\, m_e\, c}\,\frac{h}{2\,\pi} = 9.24 \times 10^{-21}\,\text{erg}\,\text{G}^{-1}$ is called the *Bohr magneton*. It is the magnetic quantum number, m_j, that describes the projection of the total angular momentum along the z axis. The energy state will split into sub-levels whose separation depends on the strength of the field. If a transition between energy states occurs such that $\Delta m_j = 0$ or $\Delta m_j = \pm 1$, then, respectively, the resulting line will be unshifted, or there will be a shift in the frequency of the spectral line according to,

$$\nu = \nu_0 \mp \frac{\mu_B}{h}\, B = \nu_0 \mp 1.4 \times 10^6\, B \quad \text{Hz} \tag{C.14}$$

where ν_0 is the frequency of the line in the absence of the field. In a gas in which many such transitions are occurring, then three lines can be observed instead of one[3]. This is called the *Zeeman effect*[4] and, if measurable, is a very important way to obtain the strength of the magnetic field in an object. Since interstellar magnetic fields have strengths of order a few 10^{-6} G, a typical shift for interstellar lines is of order a few Hz. Most of the lines given in Table C.1, have frequencies of 10^{14} to 10^{15} Hz, so this is a negligible shift. However, if the magnetic field is stronger or a lower frequency line is chosen, then the Zeeman effect becomes measurable. For example, the Zeeman effect is the primary way in which we know the magnetic field in Sunspots which have strong fields of order 1000 ($\Delta\,\nu = 1.4$ MHz) to 4000 ($\Delta\,\nu = 5.6$ MHz) G in comparison to a globally averaged Solar field of about 1 G. Zeeman splitting has also been measured in the HI λ 21 cm line (see below) in regions in which the interstellar magnetic field is stronger than average.[5]

C.4 The λ 21 cm line of neutral hydrogen

There is one more effect that has not yet been considered, that of the intrinsic angular momentum of the nucleus, called *nuclear spin* and designated I. For the normal hydrogen atom in which the nucleus is a proton, $I = 1/2$ and for deuterium, $I = 1$. Just as the electron's orbital angular momentum and spin angular momentum could be combined into a total angular momentum quantum number, j, so now the nuclear spin can be included with the total angular momentum to form a total angular momentum quantum number for the whole atom, called F. This coupling is only important for the ground state of hydrogen where the electron is closest to the nucleus.

The result is *hyperfine splitting* of the ground state. For hydrogen, it is helpful to think of this classically as either the proton and electron spinning in the same sense ($F = 1$), or the proton and electron spinning in opposite senses ($F = 0$), the former orientation having the higher energy. The statistical weights of these two states are given by $g = 2F + 1$. Hence the spin-parallel state has a statistical weight of $g_u = 3$ and the spin-anti-parallel state has a statistical weight of $g_l = 1$, where the subscripts, u, and l refer to the upper and lower energy levels, respectively.

The λ 21 cm transition is sometimes called the *spin-flip* transition since de-excitation involves this process. The transition from $F = 1$ to $F = 0$ is a *forbidden* transition meaning, in this case, that it violates the $\Delta\, l = \pm 1$ selection rule. This does not mean that it cannot occur, only that it is extremely rare in comparison to other lines (see the

[3]The three lines are polarized with the unshifted line polarized in a different sense than the shifted lines. If the lines are viewed from an angle perpendicular to the magnetic field direction, then all three lines are observed, but if the viewing angle is along the field direction, then the two shifted lines are seen but not the unshifted line.
[4]There is also an *anomalous Zeeman effect* which is line splitting by amounts different from Eq. (C.14). In general Eq. (C.14) must be modified to include a function of the various quantum numbers (called the *Landé g factor*). See Ref. [143].
[5]The Zeeman effect is not restricted to the hydrogen atom. It is regularly observed in other atoms as well as in molecules.

Einstein coefficient in Table C.1). In very low density gas, some fraction of the hydrogen will have time to spontaneously de-excite and emit a photon before interacting with another particle. For most ISM densities, however, it is collisions between particles that will cause the line to be excited or de-excited. Since there is so much atomic hydrogen in space, it is possible to observe the resulting spectral line, the *λ 21 cm line of hydrogen*. This radio line is exceedingly important in astrophysics providing us with an enormous amount of information, from galaxy dynamics to cosmology, as is described further in Sect. 9.4.3. It is astonishing that one of the most powerful probes of our Universe rests with a tiny, insignificant hyperfine energy shift in the simplest of all atoms.

Appendix D
Scattering Processes

Scattering involves a brief interaction between light and matter that results in a change of direction of incoming photons, as Figure 5.1.a illustrates. The observed light intensity will diminish because only a fraction of the incoming photons are scattered into the same line of sight as the original beam. If the light is scattered with equal probability into all 4π sr, then the scattering is isotropic[1]. If the outgoing photons have the same energy as the incoming photons, then the scattering is said to be *elastic* and, if some energy is lost, the scattering is *inelastic*, similar to the treatment of colliding particles in the discipline of Kinematics.

To understand some scattering processes, however, it is helpful to treat the signal as a wave because the scatterer may be a charged particle that is responding to the oscillating electric field of the incoming light. Elastic scattering is called *coherent scattering*[2] in this model, with the outgoing energy, and therefore wavelength ($E = h\,c/\lambda$), of the scattered wave remaining unchanged. Inelastic scattering then corresponds to *incoherent* scattering.

In either approach, the result is the same and the fraction of light that is lost from the line of sight is directly proportional to the particle's effective scattering cross-section, σ_s. It is therefore of some importance to determine σ_s for different types of particles and scattering circumstances. A few of these, important to astrophysics, will now be described.

[1] In reality, scattering may not be isotropic, but is often symmetric in the forward/reverse directions. A mathematical function that describes the strength of the scattering with angle is called the *phase function*.
[2] 'Coherent' classically implies that the phase of the outgoing light is also the same as that of the incoming beam. This is true for coherent scattering except in the case of resonance (see Appendix D.1.4 and Ref. [77]).

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

D.1 Elastic, or coherent scattering

D.1.1 Scattering from free electrons – Thomson scattering

An example of elastic scattering is *Thomson scattering* which occurs when light scatters from free electrons. For the photon to be scattered elastically, its energy must be much less than the rest mass energy of the electron (otherwise some of the energy is lost to the electron, as will be described in Appendix D.2), i.e.,

$$E_{\rm ph} \ll {\rm m_e}\, c^2 = 8.2 \times 10^{-7} \quad {\rm erg} = 0.51 \quad {\rm MeV} \tag{D.1}$$

where $\rm m_e$ is the mass of the electron and c is the speed of light. Since this energy occurs in the γ wave part of the spectrum (Table G.6), Thomson scattering can occur for any light frequency up to at least soft X-rays before energy exchange starts to become important. When the wave impinges on an electron, the wave's oscillating electric field, $\vec{\rm E}$, induces an oscillation of the electron at the same frequency due to the electric field component of the Lorentz force, $\vec{\rm F} = e\,\vec{\rm E}$ (see Table I.1), where e is the charge on the electron. An oscillatory motion is also an accelerating motion, and any accelerating charged particle, in this case an electron, will radiate. The result is *electric dipole radiation*. The electron can be thought of as absorbing a wave and re-emitting it again at the same frequency but over a wide range of angles which, for Thomson scattering, is symmetric in the forwards and backwards directions with less emission towards the sides. The electron oscillates in phase with the incoming wave, and no energy is lost as a result of the interaction.

Note that the oscillating magnetic field of the incoming wave could also have initiated an oscillation of the electron via the magnetic component of the Lorentz force, $\vec{\rm F}_{\rm B} = e\,\frac{\vec{v}}{c} \times \vec{B}$ (Table I.1). However, for electromagnetic radiation, the magnitude of the electric field vector equals the magnitude of the magnetic field vector, $|\vec{\rm E}| = |\vec{B}|$, in cgs units. In order for the magnetic oscillation to have the same amplitude as that induced by the electric field, the electron velocity would have to equal the velocity of light. The oscillating magnetic field can therefore be neglected for Thomson scattering.

The probability of a photon interacting with an electron is determined by the *Thomson cross-section*, which is the area presented by the electron to an incoming photon[3],

$$\sigma_{\rm T} = \frac{8\,\pi}{3}\, r_{\rm e}^2 = \frac{8\,\pi\, e^4}{3\,{\rm m_e^2}\, c^4} = 6.65 \times 10^{-25} \quad {\rm cm^2} \tag{D.2}$$

where $r_{\rm e}$ is the classical electron radius, given in Table G.2. Note that this cross-section is independent of wavelength so the fraction of scattered emission compared to incident emission is the same for all wavelengths. Eq. (D.2) would also apply to a free proton with $\rm m_e$ replaced by the proton mass, $\rm m_p$, but it is clear that the proton cross section would lower by a factor of $({\rm m_e}/{\rm m_p})^2 \approx 10^{-7}$ and therefore negligible in

[3]The 'effective area' for an interaction may not be exactly the 'geometrical' area (see Sect. 3.4.3).

comparison to the electron cross-section. Therefore, it is the free electrons in an ionized gas that cause the scattering, rather than the free protons.

If the incoming signal is unpolarized, composed of many plane waves with a random orientation of $\vec{E}$ perpendicular to the direction of propagation, then the scattered waves from a distribution of free electrons will be polarized in a direction 90° from the direction of the incoming signal. This is because an oscillating electron can only produce a transverse wave, as shown in Figure D.1. Thus, Thomson scattering polarizes light.

D.1.2 Scattering from bound electrons I: the oscillator model

Scattering from bound electrons is a more complex situation (see Ref. [143] or Ref. [79] for details) since there are potentially many scatterers (many bound electrons) and many natural frequencies in the atom or molecule (many possible quantum transitions). This problem has been approached by first treating a single bound electron as a classical harmonic oscillator like a mass on a spring, and adjusting the results as appropriate to a real quantum mechanical system later. The model is motivated by the fact that, as with Thomson scattering, the electron should oscillate in response to the impinging oscillating electromagnetic wave, yet the electron is also anchored, via a Coulomb force, to its parent atom or molecule like a mass would be anchored via a connecting spring to a wall.

Just like a mass on a spring, a single bound electron will oscillate at a natural frequency if it is perturbed from its equilibrium configuration. This natural frequency can be expressed either as an angular frequency, ω_0, or a frequency, ν_0, where $\omega_0 = 2\,\pi\,\nu_0$. Since the electron is oscillating (and therefore accelerating), it also emits radiation. This emitted radiation provides a *radiation reaction force* which damps the oscillations so that the electron eventually settles down into its equilibrium state again. Thus, in response to a single perturbation, the electron behaves as a *damped harmonic oscillator*. The damping is exponential with time, t, such that, averaged over a single oscillation, the energy declines as $e^{-\Gamma_{\rm cl}/t}$. Here, $\Gamma_{\rm cl}$ is called the classical *damping constant*,

$$\Gamma_{\rm cl} \equiv {\omega_0}^2\,\tau_{\rm e} = (2\,\pi)^2\,\nu_0^2\,\tau_{\rm e} \tag{D.3}$$

where,

$$\tau_{\rm e} \equiv \frac{2\,e^2}{3\,{\rm m_e}\,c^3} = 6.3 \times 10^{-24} \quad {\rm s} \sim \frac{r_{\rm e}}{c} \tag{D.4}$$

is approximately the time for light to travel a distance equal to the classical electron radius, $r_{\rm e}$. Therefore, the classical damping constant is,

$$\Gamma_{\rm cl} = \frac{8\,\pi^2\,e^2\,\nu_0^2}{3\,{\rm m_e}\,c^3} = 2.47 \times 10^{-22}\,\nu_0^2 \qquad {\rm s}^{-1} \tag{D.5}$$

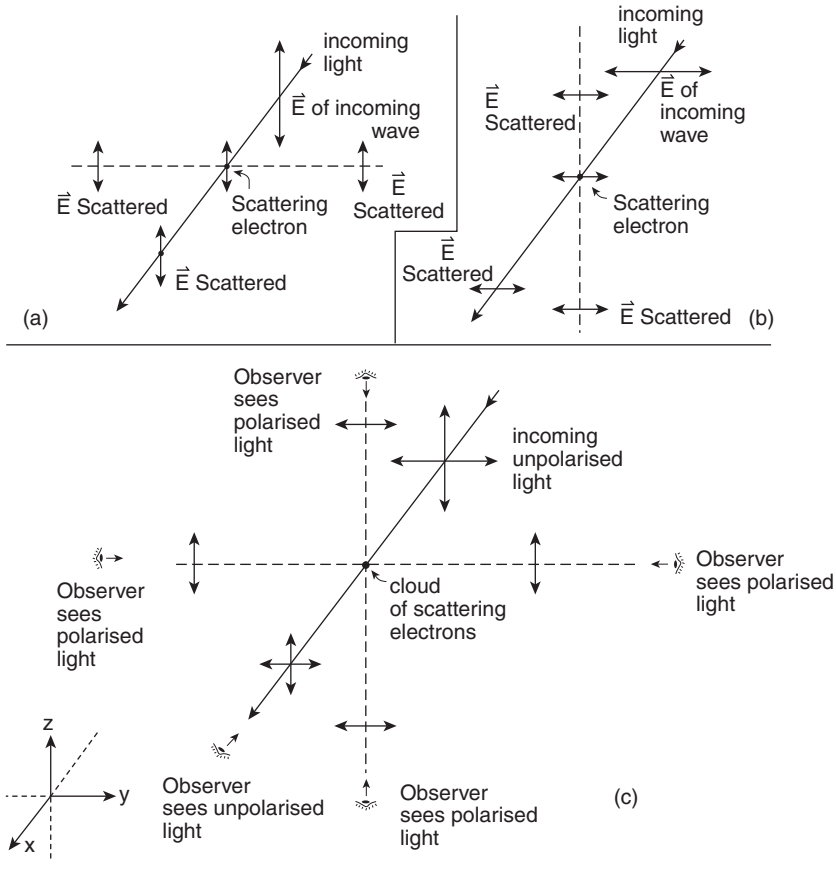


Figure D.1. Geometry showing how unpolarized light can be polarized as a result of Thomson scattering or Rayleigh scattering. The three-dimensional $x - y - z$ axis is shown at lower left and, in each panel, the incoming light is propagating in the positive x direction with its electric field vector, $\vec{E}$, perpendicular to the direction of propagation. The electron will oscillate in the direction of oscillating $\vec{E}$, emitting a scattered electromagnetic wave. (**a**) In this case, $\vec{E}$ of the incoming wave, and therefore the motion of the electron, oscillates in the $\pm z$ direction. The outgoing scattered radiation due to an oscillation in z will be only in the $x - y$ plane. Three of the scattered $\vec{E}$ vectors are shown on the positive x and positive and negative y axes but they could have been drawn at any point in the $x - y$ plane. (**b**) In this case, the incoming wave's $\vec{E}$ lies in the $\pm y$ direction, resulting in scattered radiation in the $x - z$ plane only. Three of the scattered $\vec{E}$ vectors are shown. (**c**) This case illustrates the situation in which there are many incoming waves being scattered off many electrons. The incoming wave is unpolarized, meaning that there are many $\vec{E}$ vectors all of which are perpendicular to the direction of propagation, though only the two in the $\pm y$ and $\pm z$ directions are shown. An observer looking back along the direction of propagation will continue to see unpolarized scattered light. However, any observer perpendicular to the direction of propagation would see a polarized signal, as illustrated.

The lifetime of the oscillation, is called the *damping time*,

$$\tau_{\text{damp}} = 1/\Gamma_{\text{cl}} \tag{D.6}$$

An oscillation corresponding to an optical light frequency of $\nu_0 \approx 6 \times 10^{14}$ Hz, for example, would give a damping time of $\tau_{\text{damp}} \approx 10$ ns. This is much longer than the period of oscillation itself, $T_0 = 1/\nu_0 \approx 10^{-15}$ s so there will be many oscillations before the energy is dissipated. The case of the damped harmonic oscillator is the classical analogue of spontaneous line emission, a subject to which we will return in the next section (Appendix D.1.4).

If the perturber is now an incident electromagnetic wave, then instead of a single perturbation, the electron is continuously forced by the applied oscillating electric field of the wave. It now behaves as a *forced, damped harmonic oscillator.* The resulting scattering cross-section in this harmonic oscillator model is a function of frequency,

$$\sigma_s(\omega) = \sigma_{\text{T}} \frac{\omega^4}{\left(\omega^2 - \omega_0^2\right)^2 + \left(\omega_0{}^3\, \tau_{\text{e}}\right)^2} \tag{D.7}$$

where $\omega = 2\,\pi\,\nu$ is the angular frequency of the perturbing wave. The forcing frequency may be similar to, higher, or lower than the natural frequency, and the type of scattering that results depends on which regime is being considered. These are:

(**a**) $\omega \gg \omega_0$:

If the incoming wave's frequency is much greater than the natural frequency then Eq. (D.7) reduces to the constant,

$$\sigma_s = \sigma_{\text{T}} \tag{D.8}$$

so the situation approximates that of Thomson scattering. This is because the period of the natural oscillation, $T_0 = 1/(2\,\pi\,\omega_0)$, is much greater than the forcing time $T = 1/(2\,\pi\,\omega)$, of the incident radiation. The bound electron does not have time to oscillate naturally in the presence of such rapid forcing so the electron simply responds to the external force as if it were 'free' of its natural binding to the atom.

(**b**) $\omega \approx \omega_0$:

If $\omega \approx \omega_0$, then $\left(\omega^2 - \omega_0^2\right) \approx (\omega - \omega_0)(\omega + \omega_0) \approx (\omega - \omega_0)\, 2\omega_0$ and Eq. (D.7) becomes,

$$\sigma_s(\omega) = \frac{\sigma_{\text{T}}}{2\,\tau_{\text{e}}} \frac{(\Gamma_{\text{cl}}/2)}{(\omega - \omega_0)^2 + (\Gamma_{\text{cl}}/2)^2} \tag{D.9}$$

where we have used Eq. (D.3). This describes *resonance scattering* and occurs when the incoming photon has a frequency approximately equal to the frequency at which the

system would naturally oscillate. The function has a peak[4] at ω_0 and declines as ω departs from ω_0. This resonance is similar to other common systems in which a forcing frequency is similar to a natural frequency, such as pushing a child on a swing. The scattering cross-section achieves its highest values for this case.

(**c**) $\omega \ll \omega_0$:

In this case, the perturbing electric field appears 'frozen' in comparison to the rapid natural oscillation of the electron, so the perturbing force acts like a static force. Eq. (D.7) then becomes (noting that, from Eq. (D.3), $\omega_0^3 \tau_e = \omega_0 \Gamma_{cl} \ll \omega_0^2$)

$$\sigma_s(\omega) = \sigma_T \left(\frac{\omega}{\omega_0} \right)^4 = \sigma_T \left(\frac{\lambda_0}{\lambda} \right)^4 \tag{D.10}$$

where we have used $\omega = 2\pi c/\lambda$ and $\omega_0 = 2\pi c/\lambda_0$. This is called *Rayleigh scattering* and has a strong dependence on wavelength, λ.

D.1.3 Scattering from bound electrons II: quantum mechanics

We now wish to extend the classical oscillator model into the realm of quantum mechanics. Classically, the accelerating, oscillating electron radiates and this radiation produces the damping. Quantum mechanically, the 'natural frequencies' in the atom are identified with the transitions between quantum states, so that ν_0 becomes $\nu_{i,j}$, i and j representing the lower and higher energy states, respectively. Since an atom has many natural frequencies that correspond to many different electron transitions governed by quantum mechanical rules, the oscillator model must also be modified to take this into account. This is achieved via a correction factor called the *absorption oscillator strength*, $f_{i,j}$. The oscillator strength is a unitless parameter that can be thought of as the number of oscillators (electrons) that respond to incident radiation at frequency, ν. Values of $f_{i,j}$ for the hydrogen atom and corresponding transition frequencies can be found in Table C.1. With these modifications and expressing the result as a function of ν instead of ω, cases a, b, and c from the previous section become respectively

$$\sigma_s = \sigma_T \quad \text{cm}^2 \qquad \nu \gg \nu_{i,j} \tag{D.11}$$

$$\sigma_s(\nu) = \frac{\pi e^2}{m_e c} f_{i,j}\, \phi_{\mathcal{L}}(\nu) \quad \text{cm}^2 \qquad \nu \approx \nu_{i,j} \tag{D.12}$$

$$\sigma_s(\nu) = \sigma_T f_{i,j} \left(\frac{\nu}{\nu_{i,j}} \right)^4 \quad \text{cm}^2 \qquad \nu \ll \nu_{i,j} \tag{D.13}$$

[4] We have neglected a frequency shift in the peak of the line which is negligible in this classical development but can be important in a quantum mechanical treatment. It is called the *Lamb shift*.

where,

$$\Phi_{\mathcal{L}}(\nu) = \frac{1}{\pi} \frac{\Gamma_{\mathrm{i,j}}/(4\,\pi)}{(\nu - \nu_{\mathrm{i,j}})^2 + [\Gamma_{\mathrm{i,j}}/(4\,\pi)]^2} \qquad \mathrm{Hz}^{-1} \tag{D.14}$$

$$\Gamma_{\mathrm{i,j}} = \Gamma_{\mathrm{i}} + \Gamma_{\mathrm{j}} \qquad \mathrm{Hz} \tag{D.15}$$

and

$$\Gamma_{\mathrm{i}} = \sum_{k<i} A_{\mathrm{i,k}} \qquad \Gamma_{\mathrm{j}} = \sum_{k<j} A_{\mathrm{j,k}} \tag{D.16}$$

We have also used Eqs. (D.2) and (D.5). $A_{\mathrm{i,k}}$, $A_{\mathrm{j,k}}$, are the Einstein A coefficients between the lower state, i and all states, k, below it, and the upper state, j, and all states, k, below it, respectively (Refs. [68], [77]), i.e. the spontaneous de-excitation rates for the two energy levels[5]. These equations are equivalent to those expressed earlier for the classical oscillator model except for the introduction of the oscillator strength and the modification of the damping constant from its classical value, Γ_{cl}. Values of absorption oscillator strengths and Einstein A coefficients for hydrogen can be found in Table C.1.

D.1.4 Scattering from bound electrons III: resonance scattering and the natural line shape

For resonance scattering (Eq. D.12), each photon that is absorbed from a lower to an upper state will result in an emission again from the upper to the initial lower state. The scattering cross-section is then proportional to the function, $\Phi_{\mathcal{L}}(\nu)$ given by Eq. (D.14), which is called the *Lorentz profile*. It has a peak, $\Phi_{\mathcal{L}}(\nu_{\mathrm{i,j}})$, and a full width at half maximum (FWHM), $\Delta\nu_{\mathcal{L}}$, of, respectively,

$$\Phi_{\mathcal{L}}(\nu_{\mathrm{i,j}}) = \frac{4}{\Gamma_{\mathrm{i,j}}} \qquad \Delta\nu_{\mathcal{L}} = \frac{\Gamma_{\mathrm{i,j}}}{(2\,\pi)} \tag{D.17}$$

where $\Gamma_{i,j}$ is given by Eqs. (D.15) and (D.16), and is normalized such that the integration over all frequencies (unitless) is equal to 1,

$$\int_0^\infty \Phi_{\mathcal{L}}(\nu)\, d\nu = 1 \tag{D.18}$$

[5]Eqs. (D.16) refer to spontaneous emission only, in the absence of any further perturbation. If the incident radiation field is strong, then two more terms must be added: one to account for the absorption of photons while the particle is in the upper state resulting in transitions to higher energy levels, and one to account for *stimulated emission* which is a process by which the incident photon induces an electron in the upper state to de-excite to the lower state without the absorption of the incident photon.

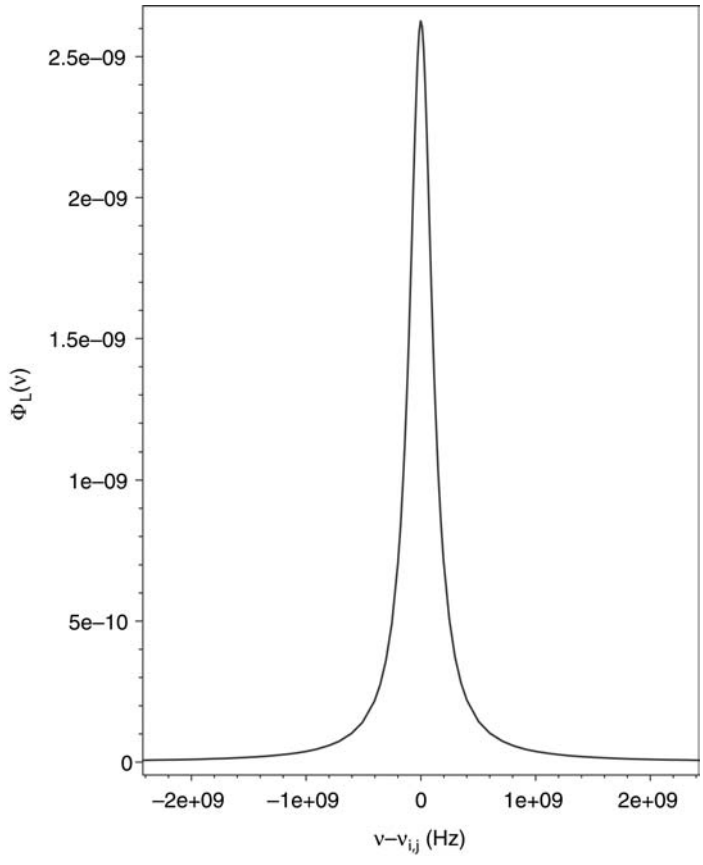


Figure D.2. The Lorentz profile, $\Phi_{\mathcal{L}}(\nu)$, plotted as a function of $(\nu - \nu_{\mathrm{i,j}})$ for the Ly α line, using the classical damping constant of Eq. (D.5).

An example is shown in Figure D.2 and illustrates that this case is not perfectly coherent since there is some probability that the photon may be emitted at a slightly different frequency than it was absorbed, the probability distribution being described by the Lorentz profile.

The damping constant, $\Gamma_{i,j}$, has a value that is somewhat larger than the classical damping constant for an oscillator and takes into account the 'fuzziness' of each of the upper and lower energy states. Using the Energy–Time form of the Heisenberg Uncertainty Principle, the relation between photon energy and frequency (both Table I.1), and the value of the line with, $\Delta\nu$, from the Lorentz profile (Eq. D.17) we have, for a transition between states, i, and j,

$$\Delta E_{i,j} = \frac{h}{2\pi}\frac{1}{\Delta t_{i,j}} = h\,\Delta\nu_{i,j} = h\,\frac{\Gamma_{i,j}}{2\pi} \tag{D.19}$$

Thus, the quantum mechanical damping constant is associated with the uncertainty in time for the transition,

$$\Gamma_{\mathrm{i,j}} = \frac{1}{\Delta\, t_{\mathrm{i,j}}} \tag{D.20}$$

The lifetime of an electron in an upper state, j, meaning the time that the electron would remain in state j before spontaneously de-exciting to *any* lower state is,

$$\tau_{\mathrm{j}} = \frac{1}{\Gamma_{\mathrm{j}}} = \frac{1}{\sum_{k<j} A_{\mathrm{j,k}}} \quad \mathrm{s} \tag{D.21}$$

where we have used Eq. (D.16), and the lifetime of an electron in upper state, j, before making a transition to a *specific* lower state, i, is (Ref. [96]),

$$\tau_{\mathrm{j,i}} = \frac{1}{A_{\mathrm{j,i}}} = \frac{1}{3\,\Gamma_{\mathrm{cl}} f_{\mathrm{j,i}}} \quad \mathrm{s} \tag{D.22}$$

Eq. (D.22) allows us to see the relationship between the Einstein A coefficient (the de-excitation rate from state j to i) and the classical damping constant, Γ_{cl}, given by Eq. (D.5) (but substituting $\nu_{\mathrm{i,j}}$ for ν_0). Note that the *emission oscillator strength*, $f_{\mathrm{j,i}}$ is a quantity whose magnitude will not, in general, be the same as that of the *absorption oscillator strength*, $f_{\mathrm{i,j}}$, because the number of possible states (i.e. the number of possible ways that an electron can 'fit' into any given quantum level) is not necessarily the same for the two levels. The number of possible states, i.e. the *statistical weight*, must therefore be taken into account. The two oscillator strengths are related via,

$$g_{\mathrm{j}} f_{\mathrm{j,i}} = g_i f_{\mathrm{i,j}} \tag{D.23}$$

where g_{j} is the statistical weight of level, j, and g_{i} is the statistical weight of level, i. Values of absorption oscillator strengths and statistical weights for the hydrogen atom are provided in Table C.1.

The process of resonance scattering involves absorption followed by re-emission by bound–bound electrons. If we consider absorption and re-emission separately, the Lorentz profile also describes the frequency response of the quantum mechanical system for each process individually. For any spontaneous de-excitation, for example, a bound–bound *emitting* transition (upper level j to lower level i) in an atom is described by the Lorentz profile which is also called the *natural line shape*. The Lorentzian line width, $\Gamma_{\mathrm{i,j}}/(2\,\pi)$ (Eq. D.17), is also called the *natural line width* or the *damping width* and it is the width of a spectral line *in the absence of any other line broadening mechanisms*. As described in Sect. 9.3, other effects will always broaden the line to a value greater than the natural line width.

D.1.5 Scattering from bound electrons IV: Rayleigh scattering

The Rayleigh scattering equation (Eq. D.13) is equivalent to the result from the classical oscillator model except for the presence of the absorption oscillator strength, $f_{\mathrm{i,j}}$. Since there are various transitions, $\nu_{\mathrm{i,j}}$, that are possible in an atom or molecule, there should be many different regimes in which $\nu \ll \nu_{\mathrm{i,j}}$, depending on the transition, i, j, being considered. However, it is possible to identify sets of lines such that $\nu \ll \nu_{\mathrm{i,j}}$ for all of them. A good example is Rayleigh scattering in the Earth's atmosphere whose prominent lines occur mostly in the UV (e.g. for N_2, O_2 and H_2O) whereas the Solar spectrum peaks in the optical. With this in mind, the total quantum mechanical Rayleigh scattering cross-section can be determined by considering a sum over the various lines. The result is (Ref. [96]),

$$\sigma_{\mathrm{R}} \approx \sigma_{\mathrm{T}} \left(\sum_k f_{1,\mathrm{k}} \left[\frac{{\nu_{1,k}}^2}{\nu^2} - 1 \right]^{-1} \right)^2 \quad \mathrm{cm}^2 \tag{D.24}$$

where $f_{1,k}$ is the absorption oscillator strength from quantum level 1 to k. Note that this equation still gives $\sigma_{\mathrm{R}} \propto \nu^4 \propto 1/\lambda^4$, just like Eqs. (D.10) and (D.13).

Rayleigh scattering is a good example of how a physical phenomenon can be approached in more than one way, as described in the introductory comments of Chapter 5. Other than this 'bottom-up' scenario of counting oscillators in atoms and molecules, the same result can be achieved by considering a molecule as a single particle with some charge distribution. The regime corresponding to $\nu \ll \nu_{\mathrm{i,j}}$ (or equivalently $\lambda \gg \lambda_{\mathrm{i,j}}$), now becomes $\lambda >>$ the particle size[6]. If the molecule has a net separation of charge (such as in the H_2O molecule) it is said to be an *electric dipole*. However, incoming light with an oscillating electric field, $\vec{\mathrm{E}}$, can induce an oscillating electric dipole moment, $\vec{p}$ (see Table I.1), even in molecules that do not normally have a permanent electric dipole moment, according to $\vec{p} = \alpha \vec{\mathrm{E}}$ where α is called the *polarizability* of the molecule and is a measure of how easy it is to distort the charge distribution. Thus, this scattering can be thought of as the response of the whole particle to the incoming oscillating electric field. Another approach is to consider all particles together in a gas, in which case Rayleigh scattering is due to fluctuations or inhomogeneities in the *dielectric constant* of the gas, the latter being a measure of how easily a material responds to an impinging electric field[7]. The dielectric constant, in turn, is related to the index of refraction, n, of the material. Thus, σ_{R} is sometimes written in terms of n or in terms of the polarizability, α. These approaches are essentially equivalent[8] and give the same result as the oscillator model, in particular, the $1/\lambda^4$

[6]Recall the simple case of the Bohr atom in which the particle size is related to the quantized wavelength, $\mathrm{n}\,\lambda = 2\,\pi\,r_{\mathrm{n}}$ (see Appendix C, first footnote).

[7]The dielectric constant is the ratio of the *permittivity*, ϵ, of the material in comparison to the permittivity of free space, ϵ_0. Recall (Table I.1) that $\epsilon_0 = 1/(4\pi)$ in cgs units.

[8]See Ref. [121], for example, which shows the relationship between the index of refraction and oscillator strengths.

dependence of $\sigma_{\rm R}$. For mathematical expressions involving these other quantities, see Ref. [96].

Rayleigh scattering is coherent and results in a scattered wave whose angular distribution is the same as for Thomson scattering, that is, forward–backwards symmetric. Figure D.1, introduced to explain the polarization caused by Thomson scattering, also applies to Rayleigh scattering.

D.2 Inelastic scattering – Compton scattering from free electrons

Inelastic scattering occurs when a scattered photon loses some fraction (< 1) of its energy. The outgoing scattered photon will then have a longer wavelength because of the loss of energy. There are several examples of this (see Sect. 5.1.2), but here we consider only *Compton scattering* which occurs when an incoming photon has approximately as much or more energy as the rest mass energy of a free electron,

$$E_{\rm ph} \gtrsim {\rm m_e}\, c^2 = 8.2 \times 10^{-7} \quad {\rm erg} = 0.51 \quad {\rm MeV} \tag{D.25}$$

If so, then some of the photon's energy will be imparted to the electron on 'impact'. The situation is shown in Figure D.3 for an electron initially at rest. In this case, conservation of energy and momentum yields,

$$E'_{\rm ph} = \frac{E_{\rm ph}}{1 + \frac{E_{\rm ph}}{{\rm m_e}\, c^2}\,(1 - \cos\theta)} \qquad (\theta \neq 0) \tag{D.26}$$

where $E_{\rm ph} = h\,c/\lambda$ is the energy of the incident photon, $E'_{\rm ph} = h\,c/\lambda'$ is the energy of the scattered photon, and θ is the scattering angle. For low photon energies

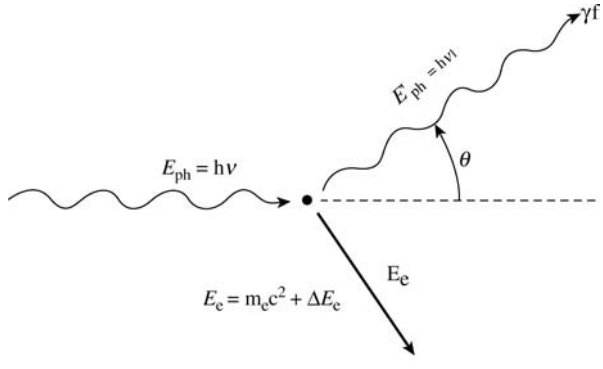


Figure D.3. Geometry for Compton scattering when the electron is initially at rest.

($E_{\rm ph} \ll {\rm m_e}\, c^2$), this expression reduces to $E'_{\rm ph} = E_{\rm ph}$ which is just elastic, or Thomson scattering (Appendix D.1.1); the scattered photon has the same wavelength as the incident photon and the electron remains at rest. For very high photon energies ($E_{\rm ph} \gg {\rm m_e}\, c^2$) and large scattering angles ($\cos\theta \approx 0$), $E'_{\rm ph} \approx {\rm m_e}\, c^2$ so the scattered photon is of order the rest mass energy of the electron, regardless of the energy (or wavelength) of the incident photon. The energy imparted to the electron after impact will be $\Delta E_{\rm e} = E_{\rm ph} - E'_{\rm ph}$.

The net result is the 'cooling' of the radiation field as radiative energy is transferred into kinetic energy of the electrons. The scattered photons will be redshifted with an increase of wavelength,

$$\Delta\,\lambda = \lambda' - \lambda = \lambda_{\rm C}(1 - \cos\theta) \tag{D.27}$$

where,

$$\lambda_{\rm C} = \frac{h}{{\rm m_e}\, c} = 2.426 \times 10^{-10} \quad {\rm cm} \tag{D.28}$$

is the *Compton wavelength* for electrons. The Compton wavelength provides a value for the change in wavelength after a single scattering (Prob 5.7).

For Compton scattering, the scattering cross-section is no longer the Thomson cross-section, but rather the *Klein–Nishina* cross-section, $\sigma_{\rm KN}$, which includes quantum and relativistic effects. The Klein–Nishina cross-section is smaller than the Thomson cross-section (Figure D.4) and therefore scattering becomes less efficient as photon energy increases.

Since Compton scattering occurs in regions of high energy photons, it is possible for electrons in such regions to have high energies as well. If so, their velocities, which could be relativistic, cannot be ignored. Conservation of energy and momentum, therefore, must be applied in the centre-of-momentum frame of reference to derive the general solution. We do not pursue this approach here, but the general solution includes the possibility of energy transfer from the electron to the photon, a situation that is known as *inverse Compton scattering* and is dealt with in greater detail in Sect. 8.6. For an initially *thermal* electron (i.e. the electron velocity distribution is Maxwellian), the condition for determining which direction energy is transferred is given by (Ref. [101]),

$$\begin{aligned}
&h\,\nu > 4\,k\,{\rm T_e} \qquad \text{Compton Effect} \\
&\text{(energy transferred to electrons)} \\
&h\,\nu = 4\,k\,{\rm T_e} \quad \text{(no energy exchange)} \\
&h\,\nu < 4\,k\,{\rm T_e} \quad \text{Inverse Compton Effect} \\
&\text{(energy transferred to photons)}
\end{aligned} \tag{D.29}$$

where $E_{\rm ph} = h\,\nu$ is the initial energy of the photon, $\rm T_e$ is the electron temperature, and k is Boltzmann's constant.

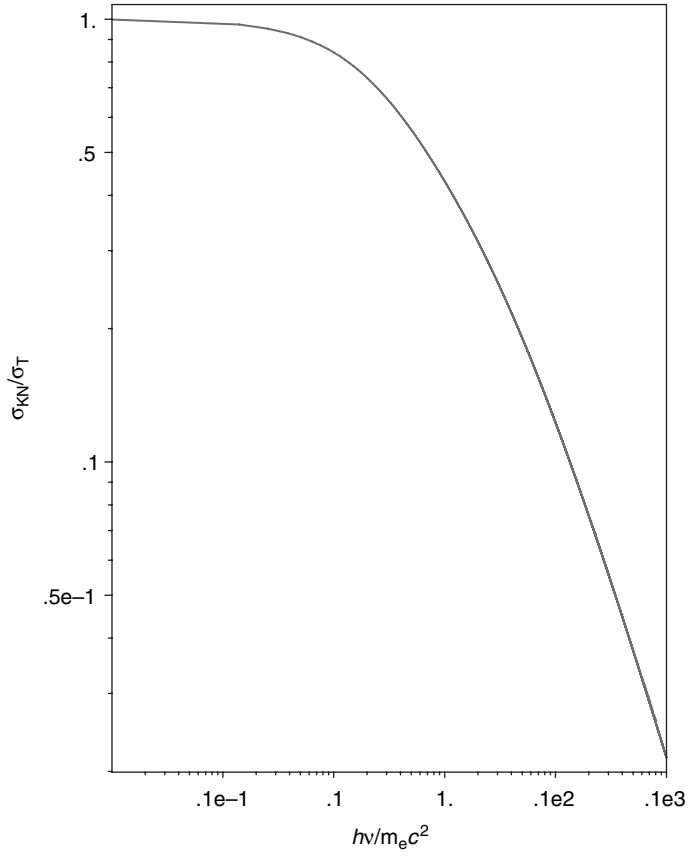


Figure D.4. Ratio of the Klein–Nishina cross-section to the Thomson cross-section, σ_{KN}/σ_T as a function of the parameter, $h\,\nu/m_e\,c^2$, showing the decrease in cross-section with increasing photon energy.

D.3 Scattering by dust

The interaction of light with dust can be complicated, but it is possible to gain some insight into the interaction by starting with some simplifying assumptions, such as grains that are spherical and of uniform composition. Dust scattering is not normally considered in isolation, however, but rather is treated together with absorption, the total being referred to as *extinction* (Chapter 5). The scattering and absorption properties resulting from different grain compositions are taken into account via a *complex* index of refraction (see below). A theory that takes this approach is called *Mie Theory.*

The extinction is expressed in terms of the unitless *extinction efficiency factor* for a grain of radius, a,

$$Q_{\mathrm{ext}\lambda} = \frac{C_{\mathrm{ext}\lambda}}{\pi\, a^2} = Q_{\mathrm{abs}\lambda} + Q_{\mathrm{scat}\lambda} = \frac{C_{\mathrm{abs}\lambda}}{\pi\, a^2} + \frac{C_{\mathrm{scat}\lambda}}{\pi\, a^2} \tag{D.30}$$

where $C_{\mathrm{ext}\lambda}$, $C_{\mathrm{abs}\lambda}$ and $C_{\mathrm{scat}\lambda}$ are the effective extinction, absorption and scattering cross-sections, respectively (equivalent to σ_{eff}, e.g. Sect. 3.4.3, Sect. 5.5) and $Q_{\mathrm{abs}\lambda}$ and $Q_{\mathrm{scat}\lambda}$ are the absorption and scattering efficiency factors, respectively. Thus, values of Q measure the effective grain cross-section in comparison to its physical cross-section. The *albedo*, that is, the fraction of incoming light that is scattered, first introduced in Sect. 4.2, can also be expressed in terms of these quantities,

$$A_\lambda = \frac{Q_{\mathrm{scat}\lambda}}{Q_{\mathrm{ext}\lambda}} \tag{D.31}$$

The response of a grain to incoming radiation depends on its composition which can be expressed in terms of a *complex* index of refraction, m, that has a real part, n, and an imaginary part, k,

$$\mathrm{m} = \mathrm{n} - \mathrm{k}\,i \tag{D.32}$$

where $i \equiv \sqrt{-1}$. This expression is a mathematical convenience that allows one to consider, together, the scattered portion of the wave, taken into account by n, and the absorbed portion, taken into account by k. The scattered portion includes all radiation that is not absorbed. This could include radiation that travels through the grain and emerges out the other side, if this is occurring, radiation that is diffracted about the grain, as well as radiation that is scattered in a more classical sense, like a photon hitting a large ball. A *pure dielectric* material (a substance with negligible conductivity) has $\mathrm{k} = 0$. Examples are ices and silicates (Sect. 3.5.2). On the other hand, metals can have values of k and n that are comparable in magnitude.

With these properties in mind, it is then possible to calculate the efficiency factors, Q as a function of the size of the grain, a, and wavelength of incident light, λ. Both a and λ are usually folded into a single variable, x,

$$x \equiv \frac{2\,\pi\,a}{\lambda} \tag{D.33}$$

and then plotted. Results for a weakly absorbing grain are shown in Figure D.5. Other compositions show different curves, but the general trend of a strong increase, a peak, and an approximately constant portion is quite typical.

The variable, x, is a useful parameter to plot because it indicates how efficiently the grain interacts with incoming radiation in terms of the size-to-wavelength ratio. It is clear that the strongest interaction (peak of the curve) corresponds to a situation in which λ is of order the grain size. For the example given in Figure D.5, the peak occurs at $x \approx 4$, or a wavelength about 80 per cent of the grain diameter.

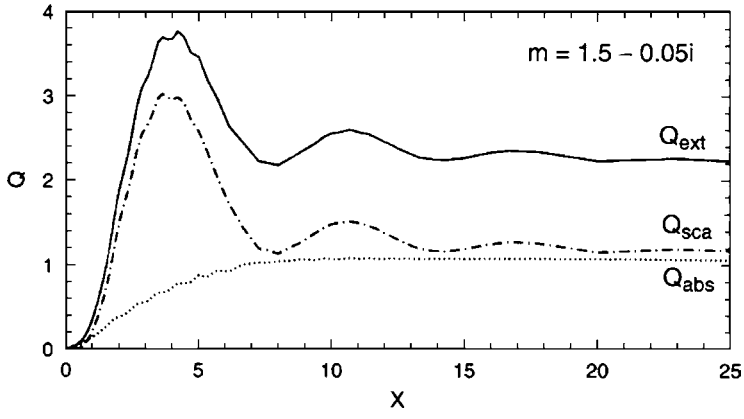


Figure D.5. Absorption, scattering, and extinction efficiency factors as a function of the parameter, $x = 2\pi a/\lambda$, for a weakly absorbing grain with index of refraction as indicated. When Q is multiplied by the geometric cross-section, the result is the effective interaction cross-section of the dust particle to the incident light. (Whittet, D.C.B, *Dust in the Galactic Environment, 2nd Ed.*, Institute of Physics Publishing)

If the grain size is very small in comparison to the wavelength of light (left on the plot), then the extinction is a strong function of wavelength. In this case, the following useful approximations result (Ref. [180]),

$$Q_{\mathrm{scat}\,\lambda} \approx \frac{8}{3}\left(\frac{2\pi a}{\lambda}\right)^4 \left|\frac{\mathrm{m}^2-1}{\mathrm{m}^2+2}\right|^2 \tag{D.34}$$

$$Q_{\mathrm{abs}\,\lambda} \approx \left(\frac{8\pi a}{\lambda}\right) \mathrm{Im}\left\{\frac{\mathrm{m}^2-1}{\mathrm{m}^2+2}\right\} \tag{D.35}$$

where Im refers to the imaginary part of the term within the braces. As long as the grain is not strongly absorbing (k is small), the term, $(\mathrm{m}^2-1)/(\mathrm{m}^2+2)$ is roughly constant with wavelength or only weakly dependent on wavelength. Thus, for a given grain size which is small in comparison to λ, $Q_{\mathrm{scat}} \propto 1/\lambda^4$ (dash-dotted curve at left) which is just Rayleigh scattering (Sect. D.1.5). For the absorption portion, $Q_{\mathrm{abs}} \propto 1/\lambda$ (dotted curve at left). In both cases, if the wavelength is very large in comparison to the grain, then the interaction of the grain with light (Q_{ext}) is negligible, such as when radio waves traverse the ISM.

On the other hand, if the grain size is large in comparison to the wavelength of light (right on the plot), then the scattering and absorption of radiation achieve a roughly constant value, similar to light hitting a large solid surface. In this case, the scattering and absorption cross-sections are just the geometric cross-sections ($Q_{\mathrm{scat}} \approx Q_{\mathrm{abs}} \approx 1$). In this limit, the extinction efficiency factor $Q_{\mathrm{ext}} \approx 2$. This means that the effective cross-sectional area of the grain is twice the geometrical cross-section, or twice as much energy is removed from the incident beam by the particle than its geometrical size would suggest. The reason for this is that light diffracts around the grain and this bending also removes light from the line of sight.

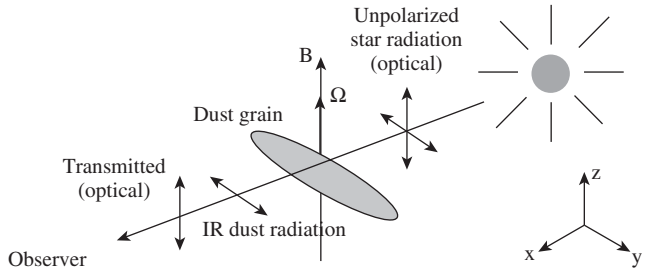


Figure D.6. An unpolarized signal from a background star becomes polarized as it interacts with an elongated dust grain that rotates about a magnetic field line. In this example, the magnetic field is parallel to the *z* axis and light propagates in the *x* direction. Incident light that has $\vec{E}$ in the *y* direction is more likely to be absorbed than waves that have $\vec{E}$ in the *z* direction, letting the light that has a *z* orientation pass through more easily. Thus, the transmitted light is polarized in a plane aligned with the magnetic field. A charged dust grain that absorbs some fraction of incident light would also re-emit it in the infrared. The emitted light would be polarized in a plane perpendicular to the magnetic field

The term *Mie scattering* is sometimes used to apply only to scattering for which a particle is *not* small with respect to the wavelength (to right in Figure D.5), in order to differentiate it from Rayleigh scattering (particles small with respect to the wavelength, far left in figure). Since $Q_{\rm scat}$ is about constant with respect to x at the right of Figure D.5 then $Q_{\rm scat} \propto 1/\lambda$ for a fixed particle size. This is a 'greyer' (flatter) response with wavelength than Rayleigh scattering ($\propto 1/\lambda^4$) and is why larger particles such as water droplets in clouds result in a greyer colour in comparison to the blue sky (Sect. 5.1.1.3).

Although Mie theory is a simple approximation to the reality of complex grains, more sophisticated theory indicates that, provided no grain alignment is occurring, large modifications of the theory are not required (Ref. [160]). In reality, however, we do know that some fraction of the grains must be aligned because scattering of starlight by dust results in polarized light. In the interstellar medium, the degree of polarization, $D_p \approx 2$ per cent (Sect. 1.7). Polarization is generally not seen unless there is a magnetic field, indicating that the magnetic field is playing an important role in aligning weakly charged grains. The grain alignment mechanism is not yet entirely understood. It is believed, though, that stellar radiation pressure, which exerts strong torques on the grains, plays a part (Ref. [52]). The degree of polarization depends on the properties of the grains and their orientation, as well as the wavelength of the incident light, so studying the polarization of optical light as a function of wavelength is also a way of obtaining information about the grains. Figure D.6 illustrates a possible configuration that would result in polarized light.

Appendix E
Plasmas, the Plasma Frequency, and Plasma Waves

As indicated in Sect. 3.4, a *plasma* is an ionized gas. Under most astrophysical conditions, a plasma is globally neutral[1] since the negative charges from free electrons will balance the positive charges of the ions. However, since the electrons are free, they can be easily perturbed by an applied electric field. Figure E.1 illustrates such a situation in which a small perturbation has caused some fraction of the charge to separate in the x direction. Since there is a net positive charge on one side and a net negative charge on the other, the plasma approximates a parallel plate capacitor with a region of width, x, on either side corresponding to the two 'plates' of the capacitor, and globally neutral ionized material is in between. The magnitude of the electric field in this capacitor will be (Table I.1),

$$\mathrm{E} = \frac{4\,\pi\,e\,N_\mathrm{e}}{A} \tag{E.1}$$

where e is the charge on the electron, N_e is the number of electrons in a plate and A is the area of a plate. The total number of electrons in a plate is,

$$N_\mathrm{e} = n_\mathrm{e}\,x\,A \tag{E.2}$$

where n_e is the number density of electrons, taken to be roughly the same in all regions since the perturbation is small. Therefore,

$$\mathrm{E} = 4\,\pi\,e\,n_\mathrm{e}\,x \tag{E.3}$$

[1]There are some exceptions, such as near sunspots or other regions in which there are ordered magnetic fields. There may be current flow in such circumstances.

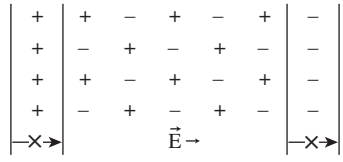


Figure E.1. Perturbation of the electric field in an astrophysical plasma induces a slight charge separation in it. The plasma then approximates a parallel plate capacitor filled with ionized, neutral material within the 'plates', and having an electric field, $\vec{E}$ between them.

The force on an electron produced by this field is (Table I.1),

$$F = \mathrm{m_e}\,\ddot{x} = -e\,E \tag{E.4}$$

where $\mathrm{m_e}$ is the electron mass. Eqs. (E.3) and (E.4) lead to an equation for simple harmonic motion, like a mass on a spring, i.e.,

$$\ddot{x} = -\omega_\mathrm{p}^2\,x \tag{E.5}$$

where

$$\omega_\mathrm{p} \equiv \left(\frac{4\,\pi\,e^2\,n_\mathrm{e}}{\mathrm{m_e}}\right)^{1/2} = 5.6 \times 10^4\,n_\mathrm{e}^{1/2} \quad \text{rads s}^{-1} \tag{E.6}$$

is the angular *plasma frequency*. Thus, when electrons in a plasma are perturbed they oscillate at a natural frequency, ω_p, similar to the way in which a bell rings at a natural frequency when hit[2]. These oscillations are called *plasma waves*. One could consider the response of the ions to such a perturbation as well, but the ions are very sluggish in comparison to the electrons and can be considered at rest (Prob. 5.11).

Since accelerating electrons radiate, electrons in an ionized gas that oscillate at the plasma frequency will themselves emit radiation at the same frequency. This is the explanation for Type III solar radio bursts, as shown in Figure E.2. The declining electron density between the Sun and the Earth[3] corresponds to a declining plasma frequency.

Since the response of the gas is as an harmonic oscillator, this development is similar to that of the oscillator model for bound electrons described in Appendix D.1.2. That is, we now consider the perturber to be an electromagnetic wave whose electric field[4]

[2]Harmonics of this frequency may also be excited.
[3]The electron density, n_e, ranges from $\sim 10^8$ cm^{-3} in the Solar corona (see Figure 6.5) to ~ 5 cm^{-3} near the Earth (with some variation, Ref. [155a]).
[4]As indicated in Appendix D.1, the oscillating magnetic field has a small effect in comparison to the oscillating electric field and can be neglected.

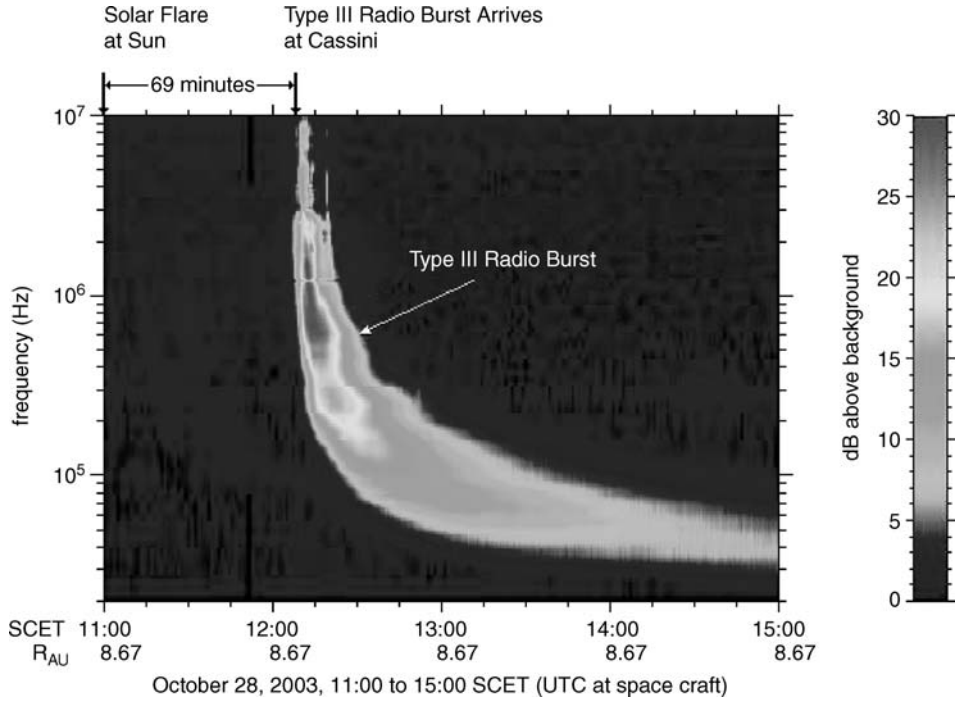


Figure E.2. This very low frequency radio emission resulting from a Solar flare was detected by the Cassini spacecraft in Oct, 2003. A solar flare has produced a fast stream of highly energetic (1−100 keV) electrons that encounters ionized gas between the Sun and the Earth, in this case, starting at a distance of about 0.5 AU from the Sun. The electrons excite waves in the gas at the plasma frequency corresponding to the electron density at that position ($\nu_p = 8.9\, n_e^{1/2}$ kHz, Eq. E.9). The radio emission can only escape if its frequency is greater than the plasma frequency of gas that it encounters, a situation that is possible for a density distribution that declines with distance from the Sun. The lower envelope of this plot denotes the plasma frequency cutoff for a declining density. The initial time delay of 69 min is the time it takes for the radio emission to first reach the spacecraft which was a distance of 8.67 AU from the Sun on that date. The maximum frequencies detected at the earliest time are somewhat higher than shown on this plot (of order 100 MHz). (Credit: D. Gurnett, NASA/JPL, & the University of Iowa)

oscillates at an angular frequency of ω, and consider what the result will be for different regimes of ω in comparison to ω_p (see below).

As the wave travels through a plasma, it will have a phase velocity (the rate at which the peaks and troughs of the wave, or the phase, advances through the medium),

$$v_{ph} = c\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right]^{-1/2} \tag{E.7}$$

and a group velocity (the rate at which energy, or any wave modulation or 'information' advances through the medium),

$$v_g = c\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right]^{1/2} \tag{E.8}$$

(Ref. [96]). Since the group velocity is the rate of energy transfer, it must always be less than or equal to c, consistent with Special Relativity. The phase velocity, however, is greater than or equal to c. This does not violate Special Relativity because no energy is transfered at this rate. Note that, from Eqs. (E.7) and (E.8), $v_{ph}\, v_g = c^2$.

Now, if $\omega \gg \omega_p$, the natural plasma oscillations are very slow in comparison to the perturbing wave. The medium then has very little effect on the electromagnetic wave and the wave propagates through it, as if it were transparent, with $v_{ph} \approx v_g \approx c$. If $\omega \sim \omega_p$ (but $\omega \not< \omega_p$), there is a strong interaction between the wave and the electrons in the gas. In this case, v_{ph} can be much greater than c and v_g can be much less. The result is a *resonance*, a situation in which matter is perturbed at a rate similar to the rate of its natural oscillation (recall resonance scattering from Appendix D.1.2). However, if $\omega < \omega_p$, then there is no solution to Eqs. (E.7) and (E.8). Such a situation corresponds to *no* propagation through the medium since the natural oscillations are so fast in comparison to the disturbance produced by the wave that they essentially cancel the disturbance. In this case, the wave reflects from the gaseous surface in much the same way that optical light reflects from a solid[5].

Expressed in cycles per second, the plasma frequency is,

$$\nu_p = 8.9\, n_e^{1/2} \quad \text{kHz} \tag{E.9}$$

For a typical ISM electron density of $n_e = 10^{-3}$ cm^{-3}, $\nu_p = 280$ Hz. This is a very low frequency in comparison to those at which astronomical measurements are made so the ionized component of the ISM is transparent to light (note, however, that scattering and absorption may still be occurring). However, the plasma frequency can become important for higher density gases. For example, the reason that low frequency radio waves do not reach the ground, as shown in the atmospheric transparency curve of Figure 2.2, is because of reflection of these signals from the Earth's ionosphere (Prob. 5.1.2). Reflection of low frequency waves from plasmas is also important in other contexts. For example, the reflection of radar signals from ionized meteor trails is used to understand the properties of incoming meteoroids, and the reflection of man-made Earth-based signals from the underside of the ionosphere allows us to hear radio stations that are located over the physical horizon.

[5]It is interesting to note that, for similar reasons, metals reflect optical light but can become transparent at UV wavelengths.

Appendix F
The Hubble Relation and the Expanding Universe

F.1 Kinematics of the Universe

As indicated in Sect. 7.2.2, the expansion of the Universe (see Figure F.1) produces the observed Hubble Relation (Figure F.2). For galaxies that are not very distant (closest to the origin in Figure F.2), the Hubble Relation can be written simply as the linear function,

$$c\,z = H_0\,d \quad \mathrm{km\,s^{-1}} \tag{F.1}$$

where H_0, the slope of the curve, is called the *Hubble constant*. With distance, d, in Mpc, H_0 is expressed in units of km s^{-1} Mpc^{-1} (cgs units of s^{-1}). The inverse of this is called the *Hubble time*,

$$t_{\mathrm{H}} = \frac{1}{H_0} \tag{F.2}$$

which gives an approximation to the age of the Universe. For $H_0 = 71\,\mathrm{km\,s^{-1}}$ $\mathrm{Mpc^{-1}}$ (Table G.3), $t_{\mathrm{H}} = 13.8$ Gyr. A more exact age requires consideration of curvature in the plot, although the current best value is not significantly different from t_{H} (cf. 13.7 Gyr determined from the fluctuations in the cosmic microwave background, see Sect. 3.1). Although the term, $c\,z$, is often referred to as the galaxy's *recession velocity*, it is more accurately a description of the expansion of space itself. The same relation would result even if the galaxies were fixed motionless in space (Figure F.1). Superimposed on the recessional velocity is the *peculiar velocity* of each galaxy as it moves through space

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

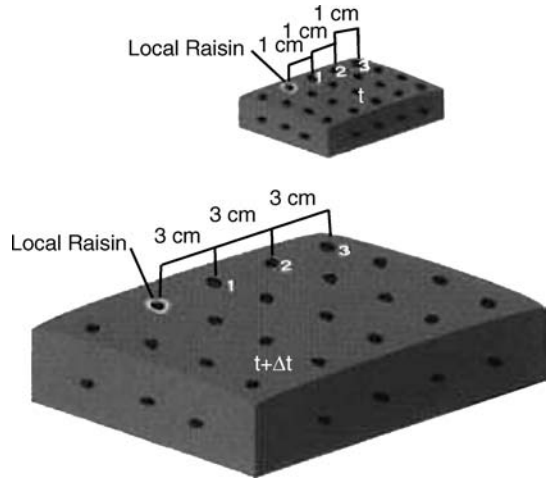


Figure F.1. This 'raisin cake model' helps to conceptualize the expansion of the Universe. As the dough (space) expands, it carries the raisins (galaxies) with it. As measured from the local raisin (the Milky Way), in time Δt, the nearest raisin has moved a distance of two units ($v_1 = 2/\Delta t$) and the farthest raisin has moved six units ($v_3 = 6/\Delta t$). Thus, the most distant raisins recede at higher speeds than the nearby raisins, as the Hubble relation indicates. The same conclusion would be reached from any raisin so, even though the origin of the Hubble plot is at the observer, the centre of the Universe is not.

under the influence of the gravitational fields of other systems near it. The peculiar velocities result in Doppler shifts which (along with random errors) produce scatter about the curve of Figure F.2. Peculiar velocities are small in comparison to the expansion redshift, except for the nearest galaxies (see Example 7.3), and are always $< c$.

Because of the finite speed of light, galaxies that are more distant (to the right) in the Hubble plot are seen as they were at earlier times in the history of the Universe. The origin, with $z = 0$, $d = 0$, $t = t_0$ is 'here and now'. Over cosmological scales, the simple linear relation of Eq. (F.1) no longer applies (Prob. 7.7) and we need to allow for curvature, meaning that we must allow for an expansion rate that changes with time. A more general form for the Hubble Relation, then, requires a *Hubble parameter*, $H(t)$, of which the Hubble constant is simply the value now ($H_0 = H(t_0)$), as well as a parameter describing the curvature called, for historical reasons, the *deceleration parameter*[1], $q(t)$, for which the *deceleration constant* is the value now ($q_0 = q(t_0)$). These are defined by,

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \qquad q(t) \equiv -\frac{\ddot{a}(t)\,a(t)}{\dot{a}(t)^2} \equiv -\frac{\ddot{a}(t)}{\dot{a}(t)[H(t)]^2} \tag{F.3}$$

[1]This terminology should not be taken to imply that the universe is currently decelerating.

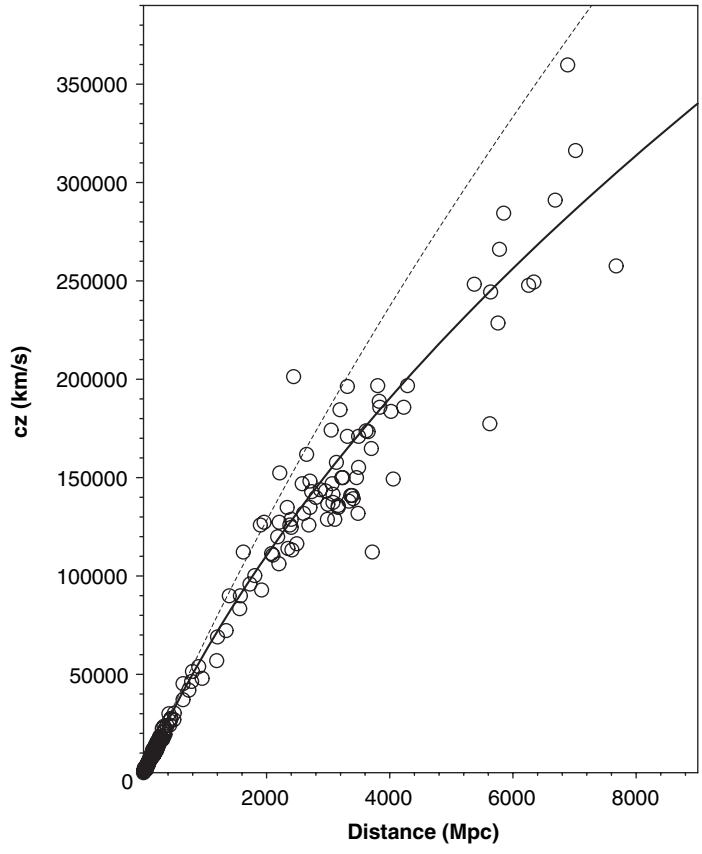


Figure F.2. The Hubble Relation, shown here to $z \approx 1$, using the Type Ia Supernova data of Ref. [171]. The curves plot Eq. (F.16) using $H_0 = 71\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Table G.3), with the dark solid curve corresponding to $q_0 = -0.55$ ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$) and the light solid curve showing $q_0 = 0.5$ ($\Omega_m = 1$, $\Omega_\Lambda = 0$). The observer (us) is at the origin. Scatter in the figure is due to the peculiar motions of galaxies as well as random error in the data. At larger distances, the errors are larger. Note that the distance is the luminosity distance, defined in Eq. (F.11) and cz should only be thought of as the 'velocity' of expansion for $z \ll 1$.

where $a(t)$ is the dimensionless scale factor introduced in Sect. 7.2.2 and we use the dot notation to indicate differentiation with time. From the right hand side of the expression for $q(t)$, we can see that the denominator is always positive. Therefore, if $q(t) < 0$, $\ddot{a}(t)$ is positive and the Universe is *accelerating* at time t, whereas if $q(t) > 0$ then the Universe is *decelerating* at t.

To better understand the scale factor, consider the Universe to consist of concentric shells, each of which has a radial coordinate, r, originating at the observer and ending on a shell. As the Universe expands, r, called the *comoving coordinate*, is 'attached' to the shell and expands with it. Since the coordinate system itself is expanding, r is a constant with time for a given shell. The scale factor then describes the form of the

expansion. We are now helped by the fact that our Universe appears to be flat on the largest scales (e.g. Ref [142], Ref. [159]). This means that if two parallel beams of light are transmitted into space, over large scales they will remain parallel (space is *Euclidean*) rather than converging or diverging as they might if space had some intrinsic curvature to it[2]. The *proper distance*, i.e. the commonly understood distance one would obtain by laying down rulers, end to end, can then be expressed in terms of the scale factor and comoving coordinate as,

$$d(t) = a(t)\, r \tag{F.4}$$

We now derive the Hubble Relation for $z < 1$ which allows us to drop terms higher than second order in any expansions following. Expanding the scale factor, $a(t)$ in a Taylor series (Appendix A.1) about t_0 we have,

$$a(t) = a_0 + \dot{a}_0(t - t_0) + \frac{\ddot{a}_0}{2}(t - t_0)^2 + \ldots \tag{F.5}$$

where the subscript, 0, refers to values at $t = t_0$, and for simplicity, we have dropped the functional dependence on t on the right hand side only. For example, we now write[3] a_0 rather than $a(t_0)$ and the values in parentheses on the right hand side now imply multiplication rather than a functional dependence. Using the relations of Eqs. (F.3) evaluated now, the above becomes,

$$a(t) = a_0\,[1 - H_0(t_0 - t) - \frac{q_0}{2}\, H_0^2(t_0 - t)^2] \tag{F.6}$$

where $(t_0 - t)$ is called the *look back time*, a positive quantity since $t_0 > t$. If we use the definition for expansion redshift (Eq. 7.11) and a binomial expansion (Appendix A.2) the redshift becomes,

$$z(t) = H_0(t_0 - t) + (1 + \frac{q_0}{2})\, H_0^2(t_0 - t)^2 \tag{F.7}$$

We then invert this equation to solve for $t(z)$,

$$(t_0 - t) = \frac{1}{H_0}\,[z - (1 + \frac{q_0}{2})\, z^2] \tag{F.8}$$

Thus, for any measured redshift less than 1, if we know the *currently observed* values of H_0 and q_0, the look back time can be evaluated from Eq. (F.8).

[2]On small scales, the path of a photon passing by a mass will be altered (see Sect. 7.3.1), but this can be neglected on large scales.

[3]Since a_0 is a dimensionless scale factor that refers to 'now', it is normally set to the value, 1. We have retained it in this development, but $a_0 = 1$ could be used throughout.

A photon, travelling radially towards us through expanding space, moves from a larger comoving coordinate through smaller ones with increasing time. The infinitesimal distance travelled by such a photon is determined by the Robertson–Walker metric, an expression describing the separation between events in space–time in an expanding homogeneous and isotropic universe. For a flat universe this distance is $c\,dt = -a(t)\,dr$ (cf. Eq. F.4). Then between the time the photon is emitted from the source, t, and the time it is observed, t_0, it travels from the comoving coordinate of the source, r, to us at $r = 0$,

$$\int_r^0 dr = -c \int_t^{t_0} \frac{1}{a(t)}\, dt \tag{F.9}$$

If we insert Eq. (F.6) into the right hand side of Eq. (F.9), again use a binomial expansion, integrate, and use Eq. (F.8), we find (Prob. 7.8) the comoving coordinate of the source as a function of its redshift,

$$r = \frac{c\,z}{a_0\,H_0}\left[1 - \frac{1}{2}(1 + q_0)\,z\right] \qquad (z < 1) \tag{F.10}$$

Although it is tempting to use Eq. (F.4) to convert the comoving coordinate, r, to a proper distance, d, in reality, we cannot measure the proper distance. Rather we measure the flux of an object, f, for which the intrinsic luminosity, L, is believed to be known (see Sects. 1.1 and 1.2 for a discussion of flux and luminosity). The distance that is inferred is the *luminosity distance* defined as,

$$d_{\rm L}{}^2 \equiv \frac{L}{4\,\pi f} \tag{F.11}$$

(compare Eq. F.11 to Eq. 1.9) where the observed flux is,

$$f = \frac{L}{4\,\pi (a_0\,r)^2 (1 + z)^2} \tag{F.12}$$

The term $4\,\pi (a_0\,r)^2$ is the surface area of a sphere when and where the flux is measured (at $t = t_0$ or 'now'; recall that r is a constant with time for a given object). The term $(1 + z)^2$ results from two effects. Firstly, since the wavelength is stretched by the expansion, the received energy of a photon, $E_{\rm obs}$, is related to the energy emitted, $E_{\rm em}$ by (see Table I.1, Eq. 7.1),

$$E_{\rm obs} = \frac{h\,c}{\lambda} = \frac{h\,c}{\lambda_0 (1 + z)} = \frac{E_{\rm em}}{(1 + z)} \tag{F.13}$$

Thus the received flux ($\propto E_{\rm obs}$) is diminished by the same factor. Secondly, a time interval as measured by a distant observer, $\Delta t_{\rm obs}$, is larger than a time interval in the

source rest frame, $\Delta t_{\rm em}$, (Prob 7.9) according to,

$$\Delta t_{\rm obs} = \Delta t_{\rm em}(1+z) \tag{F.14}$$

Consequently, the flux ($\propto$ the number of photons received per unit time) is less again by the same factor. It is important to note that, since the measurement of f applies to the light when it was emitted in the past, $d_{\rm L}$ as plotted in Figure F.2 therefore is the distance to the object at the time its light was emitted and is not the distance to the object now.

Combining Eq. (F.11) and Eq. (F.12), the luminosity distance can be written[4],

$$d_{\rm L} = a_0\, r(1+z) \tag{F.15}$$

For very small redshifts ($z \ll 1$), the luminosity distance is just the proper distance of Eq. (F.4) (measured at $t = t_0$) and is the value used in Eq. (F.1). We now need only to substitute our expression for r from Eq. (F.10) into Eq. (F.15) (and eliminate the diminishingly small term in z^3) to arrive at the Hubble Relation for $z < 1$,

$$H_0\, d_{\rm L} = c\, z\, [1 + \frac{1}{2}(1-q_0)\, z] \qquad z < 1 \tag{F.16}$$

where we have expressed the result in a slightly rearranged form. Again, for very small redshifts, this reduces to Eq. (F.1). For $z > 1$, Eq. (F.16) cannot be used since higher order terms in the expansions become important (see Sect. F.3). Note that even though this equation describes the relation between z and $d_{\rm L}$ over a range of redshift, it is parameterized in terms of the *current* values of the Hubble and deceleration parameters.

To determine (calibrate) this relation, it is straightforward to measure z from the wavelength shift of spectral lines in distant sources. However, measurements of distance are more challenging. Those objects for which the absolute luminosity is believed to be known are called *standard candles*, as mentioned in Sect. 5.4.2. The flux is measured and Eq. (F.11) is then used to find $d_{\rm L}$. A variety of standard candles are used, but for the larger scales, these candles must be quite bright in order to be seen over cosmological scales. The Cepheid variables discussed in Sect. 5.4.2 are only bright enough to measure distances to the closer galaxies. Fortunately, good data come from Type Ia supernovae (SNe, as plotted in Figure F.2) which are believed to result from the thermonuclear explosion of a *white dwarf star*[5] when it accretes more gas from a nearby companion than its internal structure can support. As the explosion proceeds, we observe a *light curve* which is a plot of increasing, then decreasing brightness with time. From measurements of Type Ia SNe of known distance, it has

[4]Another common measure of distance in astronomy is called the *angular size distance* defined as $d_{\rm as} = x/\theta$, where x is the physical linear diameter of the source in the plane of the sky and θ is its apparent angular diameter as seen from the Earth. It is used when angular sizes are measured rather than luminosities and is related to the luminosity distance via $d_{\rm L} = d_{\rm as}(1+z)^2$.

[5]White dwarfs are very compact stars having an upper mass limit of 1.4 $M_\odot$ and a size approximately that of the Earth. They are the end point of evolution of a low mass star like the Sun once the outer layers have been dispersed (Sect. 3.3.2). This type of supernova therefore has a different origin than the Type II supernovae discussed in Sect. 3.3.3.

been found that the peak luminosity of this curve is almost constant for all Type Ia SNe. Moreover, those SNe with higher peak luminosities have a slower rate of decline. Therefore, a comparison of the rate of decline of a distant supernova with supernovae of known luminosity allows one to determine the intrinsic peak luminosity of the distant supernova to high accuracy. Once the intrinsic luminosity is known, the distance can be determined as described above. The Hubble relation can then be plotted from the redshifts and distances of a sufficient number of these standard candles. H_0 and q_0 may then be found (Prob. 7.10) via fits to Eq. (F.16). With H_0 and q_0 known, the Hubble relation has then been calibrated.

Supernovae have been observed in only a small fraction of all galaxies and other standard candles are faint and difficult to observe in the more distant systems. This is why the calibrated Hubble relation is a very important tool for determining distances to galaxies. One need only measure a redshift and a distance is found through this relation (e.g. Eq. F.16 for $z < 1$). Because calibration is difficult, however, the value of the Hubble constant has been the subject of much debate over the years, and the value of the deceleration constant even more so since it characterizes a change in H with time. It is currently believed that H_0 is known to within about 10 per cent (Ref. [39]). Since q_0 measures curvature, its value is revealed only in the most distant objects which are also subject to the most error. Figure F.2 shows a clear trend which suggests that our Universe is currently *accelerating* with time since $q_0 < 0$ (see also next section). There has been some suggestion that systematic errors may be creating this trend. For example, if the supernovae were intrinsically fainter in the distant past, then they would appear farther away, causing the curve to turn over as observed. Another possibility is that some intervening 'grey dust' (dust whose effects on the signal are independent of wavelength and therefore would be difficult to detect) would make them appear fainter. However, independent data from spatial fluctuations in the cosmic background radiation (see Figure 3.2, Sect. 3.1) are consistent with the requirement of an acceleration term (Ref. [159]). This has been one of the greatest surprises of modern cosmology.

F.2 Dynamics of the Universe

What is producing this large scale behaviour of the Universe? We have dealt, so far, only with the *kinematics* of the expanding Universe. We have not yet considered the *dynamics* which would relate this expansion to the forces that drive it. For this, Einstein's Field Equations of General Relativity are used together with the Robertson–Walker metric, resulting in an expression called the *Friedmann Equation*. In the version used here, we adopt a flat universe, neglect radiation pressure which is important only in the radiation-dominated era of the early universe at $z \geq 1000$, and modify the result to include an acceleration term expressed as the *cosmological constant*, Λ (cgs units of s^{-2}). The result is,

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\pi G\rho}{3} - \frac{\Lambda}{3} = 0 \tag{F.17}$$

where ρ (g cm^{-3}) is the total mass density of the Universe including all matter, both light and dark[6]. It is more general to express ρ and Λ in terms of energy densities, ϵ (erg cm^{-3}). The mass density can be expressed in terms of the energy density of matter and the cosmological constant can be thought of in terms of the energy density of a vacuum. For matter, we use Einstein's famous equation for the equivalence of mass and energy, expressed as a density,

$$\epsilon_{\rm M} = \rho\, c^2 \tag{F.18}$$

and for the accleration term, we define an energy density,

$$\epsilon_\Lambda \equiv \frac{\Lambda\, c^2}{8\,\pi\, G} \tag{F.19}$$

With these relations, the Friedmann Equation is rewritten using Eq. (F.3),

$$H^2 - \frac{8\,\pi\, G}{3\, c^2}(\epsilon_{\rm M} + \epsilon_\Lambda) = 0 \tag{F.20}$$

We now define a *critical energy density*,

$$\epsilon_c \equiv \frac{3\, H^2\, c^2}{8\,\pi\, G} \tag{F.21}$$

as well as *density parameters*,

$$\Omega_{\rm M} \equiv \frac{\epsilon_{\rm M}}{\epsilon_{\rm c}} = \frac{8\,\pi\, G\,\rho}{3\, H^2} \qquad \Omega_\Lambda \equiv \frac{\epsilon}{\epsilon_{\rm c}} = \frac{\Lambda}{3\, H^2} \tag{F.22}$$

where the right hand side of each equation makes use of Eq. (F.19) and Eq. (F.21). So the Friedmann Equation reduces, for a flat universe, to the simple expression,

$$\Omega_{\rm M} + \Omega_\Lambda = 1 \tag{F.23}$$

Note that both $\Omega_{\rm M}$ and Ω_Λ change with time, but their sum should always equal 1 in a flat universe with an acceleration term. If there were no cosmological acceleration term, then $\Omega_{\rm M} = 1 \Rightarrow \epsilon_{\rm M} = \epsilon_{\rm c}$. Thus, the critical energy density is the energy density associated with matter in a flat universe, in the absence of a value of Ω_Λ. In a universe with an acceleration term, both $\Omega_{\rm M}$ and Ω_Λ must be < 1. Because $\Omega_{\rm M}$ includes all of the contents of the universe, one could also break it down into its various components. For example, dark matter, $\Omega_{\rm D} = \epsilon_{\rm D}/\epsilon_{\rm c}$, and matter that radiates, $\Omega_{\rm L} = \epsilon_{\rm L}/\epsilon_{\rm c}$, so that $\Omega_{\rm D} + \Omega_{\rm L} = \Omega_{\rm M}$. Dark matter could be expressed as baryonic and non-baryonic ($\Omega_{\rm B} + \Omega_{\rm N} = \Omega_{\rm D}$), and so on. Determining these quantities is an important goal of

[6]It also includes the radiation content, although this is negligible for $z < 1000$ and so will be neglected throughout.

modern cosmology since they tell us about the kind of universe we live in. Thus we now need to relate the density parameters of Eq. (F.22) to observable quantities, such as H_0 and q_0, which can be measured from the Hubble Relation.

To do this, we make use of conservation of energy, i.e. the total energy content of the Universe associated with matter (which includes radiation), ε, within any comoving radius, r, must not change with time. Expressing this as $\varepsilon = \rho(r\,a)^3\,c^2 = \text{constant}$ (having used Eq. F.4) and differentiating with respect to time leads to $(\dot{\rho}\,a^3 + 3\,\rho\,a^2\,\dot{a}) = 0$. Solving for $\dot{\rho}$ and using Eq. (F.3),

$$\dot{\rho} = -3\,\rho\,H \tag{F.24}$$

Now differentiating Eq. (F.17) with respect to time, and substituting Eq. (F.17) and Eq. (F.24) into the result yields,

$$\frac{\ddot{a}}{a} + \frac{4\,\pi\,G\,\rho}{3} - \frac{\Lambda}{3} = 0 \tag{F.25}$$

This is the acceleration equation, sometimes called the *Second Friedmann Equation*. Using Eq. (F.3), Eq. (F.22), and Eq. (F.21), this becomes,

$$q = \frac{\Omega_{\mathrm{M}}}{2} - \Omega_{\Lambda} \tag{F.26}$$

If there were no cosmological constant ($\Omega_{\Lambda} = 0$), then a flat universe would require $\Omega_{\mathrm{M}} = 1$ and $q = 0.5$.

The above equations apply to any time, t, and therefore they also apply now, so any quantity that has a dependence on t can be expressed with its present day value. Thus, Eq. (F.22), Eq. (F.23) and Eq. (F.26) become,

$$\Omega_{\mathrm{M0}} = \frac{8\,\pi\,G\,\rho_0}{3\,{H_0}^2}, \quad \Omega_{\Lambda 0} = \frac{\Lambda}{3\,{H_0}^2}, \quad \Omega_{\mathrm{M0}} + \Omega_{\Lambda 0} = 1, \quad q_0 = \frac{\Omega_{\mathrm{M0}}}{2} - \Omega_{\Lambda 0} \tag{F.27}$$

Often, the subscript, 0, is dropped and the values are taken to be understood as the respective values now. These are extremely useful relations since they relate the observed values, q_0, and H_0, obtainable from the Hubble Relation, to quantities that describe the fundamental nature of our Universe. Notice that Eqs. (F.27) give us four equations in the four unknowns, Ω_{M0}, $\Omega_{\Lambda 0}$, ρ_0, and Λ. Thus, all quantities can, in principle, be determined. By considering how these quantities vary with time, moreover, the history and nature of the expansion may be revealed. For example, dividing the Friedmann equation by ${H_0}^2$, using $\rho = \rho_0(a_0/a)^3$ (since total mass is conserved) and applying our previous definitions evaluated today, we find,

$$\frac{H^2}{{H_0}^2} = \Omega_{\mathrm{M0}}\left(\frac{a_0}{a}\right)^3 + \Omega_{\Lambda 0} \tag{F.28}$$

The only time-variable quantities in this equation are H and a. This equation makes it quite clear that the expansion of the universe will become more and more dominated by the acceleration term in later times as a increases (Prob. 7.11), if this is indeed the kind of universe we live in.

The Hubble Relation shown in Figure F.2 shows two curves, one with $\Omega_{\Lambda 0} = 0$, $\Omega_{\mathrm{M}0} = 1$, $q_0 = 0.5$ (light curve) and one with $\Omega_{\Lambda 0} = 0.7$, $\Omega_{\mathrm{M}0} = 0.3$, $q_0 = -0.55$ (dark curve). The latter set of values are a better match to the data. As pointed out in Sect. F.1, $q_0 < 0$ implies that the Universe is currently accelerating. From Eq. (F.22) and H_0 from Table G.3, the latter set of values gives $\rho_0 = 2.8 \times 10^{-30}$ g cm^{-3} and $\Lambda = 1.1 \times 10^{-35}$ s^{-2}. It is these quantities which have been presented as percentages in Figure 3.4, but the Figure 3.4 values are taken from a broader range of measurements than outlined in this Appendix. In that figure, 'dark energy' (not to be confused with dark matter) refers to whatever energy source is producing $\Omega_{\Lambda 0}$ and 'matter' refers to all matter, both light and dark, that is represented by $\Omega_{\mathrm{M}0}$ (see also the discussion in Sect. 3.2). It is clear that, provided there are no systematic errors in Figure F.2, a matter dominated universe is unable to describe the universe we live in. Since an acceleration term is needed, there must be some kind of dark energy that is producing it. The form that this dark energy might take is currently unknown. Theoretical cosmologists are actively pursuing a variety of possibilities including the concept of *quintessence* which allows the acceleration term to vary with time, rather than being expressed in the form of a cosmological constant. It may be that advances in the realm of fundamental physics will be required before its nature will be fully revealed.

It is astonishing that more than half of the energy of the universe appears to be in a mysterious form that accelerates it and, of the remainder, most of the matter is dark (Sect. 3.2). Thus, our knowledge of the universe is gleaned from only a tiny fraction of its contents. Remarkably, it indeed appears to be possible to discern the large-scale structure and behaviour of the Universe from the visible tracers within it.

F.3 Kinematics, dynamics and high redshifts

Since we now have a clearer view as to the kind of Universe we live in, it is becoming more common to write the kinematic equations of Sect. F.1 in terms of the dynamical expressions of Sect. F.2. For example, substituting the expression for q_0 as well as $\Omega_{\Lambda 0} + \Omega_{\mathrm{M}0} = 1$ from Eqs. (F.27) into Eq. (F.10), gives a dynamical expression for the comoving distance when $z < 1$,

$$r = \frac{c\,z}{a_0\,H_0}\left(1 - \frac{3}{4}\,\Omega_{\mathrm{M}0}\,z\right) \qquad (z < 1) \tag{F.29}$$

Using Eq. (F.15) and eliminating terms in z^3 which are small for $z < 1$, this can be converted to a luminosity distance,

$$d_{\rm L} = \frac{c\,z}{H_0}\left[1 + \left(1 - \frac{3}{4}\,\Omega_{\rm M0}\right) z\right] \qquad (z < 1) \tag{F.30}$$

Eq. (F.30) is another way, besides the kinematical result of Eq. (F.16), of relating the redshift, z, to the distance, $d_{\rm L}$, of the Hubble Relation (Figure F.2).

In Sect. F.1, we did not consider any redshifts higher than 1. Currently, the highest measured redshifts are of order $z \sim 7$ (Ref. [89]) and there is every reason to believe that higher redshifts will be routinely measured as more powerful telescopes are trained towards the faintest objects in the deepest reaches of the Universe. For $z > 1$, however, the higher order terms of earlier expansions cannot be neglected and Eqs. (F.16) and (F.30) can no longer be used to describe the Hubble Relation or to determine the distance. For the sake of completeness, we provide here the full expression for comoving distance for a flat Universe with a cosmological constant in which $\Omega_{\rm M} + \Omega_\Lambda = 1$,

$$r = \frac{c}{a_0\,H_0}\int_0^z \frac{dz}{\left(\Omega_{\rm M0}\,z^3 + 3\,\Omega_{\rm M0}\,z^2 + 3\,\Omega_{\rm M0}\,z + 1\right)^{1/2}} \tag{F.31}$$

As before, this expression can be used with Eq. (F.15) to determine a luminosity distance, or any other distance (see Footnote 4 of this Appendix) that may be desired.

Appendix G
Tables and Figures

Table G.1. The Greek alphabet

Capital	Small	Name
A	α	alpha
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ϵ	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	mu
N	ν	nu
Ξ	ξ	xi
O	o	omicron
Π	π	pi
P	ρ	rho
Σ	σ	sigma
T	τ	tau
Υ	υ	upsilon
Φ	ϕ	phi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Astrophysics: Decoding the Cosmos Judith A. Irwin
 ISBN: 978 0-470-01305-2(HB) 978-0-470-01306-9(PB)

Table G.2. Fundamental physical data[a]

Symbol	Meaning	Value
Physical constants		
c	Speed of light in vacuum	$2.997\,924\,58 \times 10^{10}$ cm s^{-1}
G	Universal gravitational constant	$6.6742(10) \times 10^{-8}$ cm^3 g^{-1} s^{-2}
g	Standard gravitational acceleration (Earth)	$9.806\,65 \times 10^{2}$ cm s^{-2}
k	Boltzmann constant	$1.380\,6505(24) \times 10^{-16}$ erg K^{-1}
		$8.617\,343(15) \times 10^{-5}$ eV K^{-1}
h	Planck constant	$6.626\,0693(11) \times 10^{-27}$ erg s
N_A[b]	Avogadro's constant	$6.022\,1415(10) \times 10^{23}$ mol^{-1}
$\mathcal{R}$	Molar gas constant	$8.314\,472(15) \times 10^{7}$ erg mol^{-1} K^{-1}
R_∞	Rydberg constant	109 737.315 68 525(73) cm^{-1}
Atomic and nuclear data		
e[c]	Atomic unit of charge	$4.803\,250(21) \times 10^{-10}$ esu
eV	Electron volt	$1.602\,176\,53(14) \times 10^{-12}$ erg
m_e	Electron mass	$9.109\,3826(16) \times 10^{-28}$ g
m_n	Neutron mass	$1.674\,927\,28(29) \times 10^{-24}$ g
m_p	Proton mass	$1.672\,621\,71(29) \times 10^{-24}$ g
u_{amu}	Unified atomic mass unit	$1.660\,538\,86(28) \times 10^{-24}$ g
r_e[d]	Classical electron radius	$2.817\,940\,325(28) \times 10^{-13}$ cm
λ_C	Compton wavelength	$2.426\,310\,238(16) \times 10^{-10}$ cm
λ_{C_n}	Neutron Compton wavelength	$1.319\,590\,9067(88) \times 10^{-13}$ cm
λ_{C_p}	Proton Compton wavelength	$1.321\,409\,8555(88) \times 10^{-13}$ cm
a_0	Bohr radius	$0.529\,177\,2108(18) \times 10^{-8}$ cm
σ_T	Thomson cross section	$0.665\,245\,873(13) \times 10^{-24}$ cm^2
Radiation constants		
σ	Stefan–Boltzmann constant	$5.670\,400(40) \times 10^{-5}$ erg s^{-1} cm^{-2} K^{-4}
b	Wien's Displacement Law constant	$2.897\,7685(51) \times 10^{-1}$ cm K
a[e]	Radiation constant	$7.565\,91(25) \times 10^{-15}$ erg cm^{-3} K^{-4}

[a] Data have been taken from Ref. [168], converted to cgs units, unless otherwise indicated. The values in parentheses, where provided, represent the standard error of the last digits which are not in parentheses.
[b] This is the number of particles in one *mole* (mol) of material.
[c] From Ref. [96]. An esu is an 'electrostatic unit'. The corresponding SI unit of charge is 1.602×10^{-19} C. Note that the force equation in cgs units, $F = (q_1\, q_2)/r^2$, does not have a constant of proportionality for F in dynes, q_1, q_2 in esu, and r in cm.
[d] $r_e = e^2/(m_e\, c^2)$.
[e] Ref. [44].

Table G.3. Astronomical data[a]

Symbol	Meaning	Value
Distance		
AU	Astronomical unit	$1.495\ 978\ 706(2)\times10^{13}$ cm
pc	Parsec	$3.085\ 677\ 582\ \times10^{18}$ cm
ly	Light year	$9.460\ 730\ 47\ \times10^{17}$ cm
Time		
yr	Tropical year[b]	$3.155\ 692\ 58\ \times10^{7}$ s
yr	Sidereal year[c]	$3.155\ 814\ 98\ \times10^{7}$ s
day	Mean sidereal day[c]	$23^{h}\ 56^{m}\ 04.090\ 524^{s}$
Earth		
$M_\oplus$	Earth mass	$5.973\ 70(76)\ \times10^{27}$ g
$R_\oplus$	Mean Earth radius[d]	$6.371\ 000\ \times10^{8}$ cm
R_{equ}	Earth equatorial radius[d]	$6.378\ 136\ \times10^{8}$ cm
R_{pol}	Earth polar radius[d]	$6.356\ 753\ \times10^{8}$ cm
$\rho_\oplus$	Mean Earth density[d]	5.515 g cm^{-3}
Moon		
M_m	Moon mass[d]	$7.353\ \times10^{25}$ g
R_m	Mean Moon radius[d]	$1.738\ 2\ \times10^{8}$ cm
ρ_m	Mean Moon density[d]	3.341 g cm^{-3}
r_m	Mean Earth – Moon distance[d]	$3.844\ 01(1)\ \times10^{10}$ cm
Sun		
$M_\odot$	Solar mass	$1.989\ 1(4)\ \times10^{33}$ g
$R_\odot$	Solar radius	$6.955\ \times10^{10}$ cm
$\rho_\odot$	Mean solar density	1.41 g cm^{-3}
$L_\odot$	Solar luminosity[d]	$3.845(8)\ \times10^{33}$ erg s^{-1}
$T_{eff\,\odot}$	Solar effective temperature[e]	5781 K
$S_\odot$	Solar constant[f]	$1.367\ \times10^{6}$ erg s^{-1} cm^{-2}
$m_{V\,\odot}$	Solar apparent visual magnitude[e]	−26.76
$M_{V\,\odot}$	Solar absolute visual magnitude[e]	4.81
$M_{b\,\odot}$	Solar absolute bolometric magnitude[e]	4.74
$\theta_\odot$	Angular diameter of Sun from Earth	32.0′
Galaxy		
$\mathcal{R}_\odot$	Sun Galactocentric distance[g]	8.5(5) kpc
V_{LSR}	Velocity of the Local Standard of Rest[h]	220(20) km s^{-1}
Extragalactic		
H_0	Hubble constant[i]	71^{+4}_{-3} km s^{-1} Mpc^{-1}
T_{CMB}	Cosmic Microwave Background temperature	$2.726\ \pm\ 0.005$ K

[a]Ref. [96] unless otherwise indicated. Values in parentheses indicate the one standard deviation uncertainties in the last digits.
[b]Equinox to equinox.
[c]Fixed star to fixed star.
[d]Ref. [44].
[e]Ref. [18].
[f]Flux of the Sun at distance of the Earth.
[g]Distance of Sun from the centre of the Galaxy. A value of 8 kpc is often used (e.g. Sect. 10.2.1).
[h]Velocity of an object at a Galactocentric distance of $\mathcal{R}_\odot$ in circular orbit about the centre of the Galaxy. The Sun's velocity about the Galactic centre agrees with this value to within the quoted error.
[i]Rate of expansion of the Universe. The value is from Ref. [159]; errors are likely closer to 10 per cent (Ref. [39]).

Table G.4. Planetary data[a]

Planet	Equatorial diameter (km)	Oblateness	Mass (Earth = 1)	Density (g cm^{-3})	Rotation period[b] (days)	Incl.[c]	Albedo[d] (Bond)	a^e (AU)	e^f
Mercury	4 879	0	0.055 274	5.43	58.646	0.0	0.119	0.3871	0.2056
Venus	12 104	0	0.815 005	5.24	243.019	2.6	0.750	0.7233	0.0068
Earth	12 756	1/298	1.000 000	5.52	0.9973	23.4	0.306	1.0000	0.0167
Mars	6 792	1/148	0.107 447	3.94	1.0260	25.2	0.250	1.5237	0.0935
Jupiter	142 980[g]	1/15.4	317.833	1.33	0.4101[h]	3.1	0.343	5.2020	0.0490
Saturn	120 540[g]	1/10.2	95.163	0.69	0.4440	26.7	0.342	9.5752	0.0568
Uranus	51 120[g]	1/43.6	14.536	1.27	0.7183	82.2	0.300	19.1315	0.0501
Neptune	49 530[g]	1/58.5	17.149	1.64	0.6712	28.3	0.290	29.9681	0.0086
Pluto[i]	2 390	0?	0.002 2	1.8	6.3872	57.4	0.4–0.6	39.5463	0.2509

[a]Ref. [71] unless otherwise indicated.
[b]Sidereal.
[c]Inclination of equator to orbital plane.
[d]Albedo is the fraction of incident light that is reflected. This table gives the *Bond albedo* (from http://nssdc.gsfc.nasa.gov), defined as the total fraction of reflected Sunlight over all wavebands.
[e]Mean distance to the Sun (equal to the semi-major axis of the ellipse).
[f]Mean orbital eccentricy. The perihelion and aphelion distances are given by $r_{\min} = a(1 - e)$ and $r_{\max} = a(1 + e)$, respectively.
[g]At 1 atm (101.325 kPa).
[h]For the most rapidly rotating equatorial region.
[i]At the XXVIth General Assembly of the International Astronomical Union held in Prague in August, 2006, 'Resolution 6' was adopted, indicating that 'Pluto is a ''dwarf planet'' ... and is recognized as the prototype of a new category of Trans-Neptunian Objects.'

Table G.5. Data for human vision[a]

Quantity	Value
Pupil diameter (dark-adapted)	0.8 cm
(daylight)	0.2 cm
Approx. number of photoreceptor cells[b]	
Rods	10^8
Cones	6.5×10^6
Retina effective area	10 cm^2
Focal length	2.5 cm
Resolution of human eye	$1'$[c]
Wavelength response of human eye:	
Bright-adapted (photopic) visible range	400–700 nm
Dark-adapted (scotopic) visible range	400–620 nm
Wavelength of peak sensitivity of eye:	
Bright-adapted	555 nm
Dark-adapted	507 nm
Representative wavelengths of colours:	
Violet	420 nm
Blue	470 nm
Green	530 nm
Yellow	580 nm
Orange	610 nm
Red	660 nm

[a]Wavelengths are from Ref. [71]. Note that values for the eye are typical, but vary with individual.
[b]In order to register as a separate source, each photoreceptor cell must be separated by at least one other unexcited photoreceptor cell. Dark-adapted vision response is primarily due to rods.
[c]This value varies depending on the individual and lighting conditions.

Table G.6. The electromagnetic spectrum[a]

Waveband	Wavelength range (cm)	Frequency range (Hz)	Energy range[b] (eV)
Radio[c]	≥ 1	$\leq 3\times10^{10}$	$\leq 1.2\times10^{-4}$
(Microwave)[c]	$(100 \rightarrow 0.1)$	$(3\times10^{8} \rightarrow 3\times10^{11})$	$(1.2\times10^{-6} \rightarrow 1.2\times10^{-3})$
Millimetre – Submillimetre[c]	$1 \rightarrow 0.01$	$3\times10^{10} \rightarrow 3\times10^{12}$	$1.2\times10^{-4} \rightarrow 1.2\times10^{-2}$
Infrared	$0.01 \rightarrow 10^{-4}$	$3\times10^{12} \rightarrow 3\times10^{14}$	$1.2\times10^{-2} \rightarrow 1.2$
Optical	$10^{-4} \rightarrow 3\times10^{-5}$	$3\times10^{14} \rightarrow 10^{15}$	$1.2 \rightarrow 4.1$
Ultraviolet	$3\times10^{-5} \rightarrow 10^{-6}$	$10^{15} \rightarrow 3\times10^{16}$	$4.1 \rightarrow 124$
X-ray	$10^{-6} \rightarrow 10^{-9}$	$3\times10^{16} \rightarrow 3\times10^{19}$	$124 \rightarrow 1.2\times10^{5}$
Gamma-ray	$\leq 10^{-9}$	$\geq 3\times10^{19}$	$\geq 1.2\times10^{5}$

[a]There is some variation as to where the 'boundaries' of the various wavebands lie.
[b]The energy, in electron volts, can be computed from the definition of an electron volt (Table G.2) and the relation between energy and wavelength or frquency (Table 1.1), hence 1 eV => $\lambda = 1.24 \times 10^{-4}$ cm and $\nu = 2.42 \times 10^{14}$ Hz.
[c]In astrophysics, the radio band is taken to include microwave frequencies and occasionally to include the millimetre – submillimetre bands as well.

Table G.7. Stellar data[a], Luminosity Class V (Dwarfs)

Spectral type	B–V	V–R	M_V	BC	T_{eff}(K)	$R/R_\odot$	$M/M_\odot$
O3				−4.3?	48000		50?
O5			−5?	−4.3?	44000		30?
O6	−0.32	−0.15	−4.8	−4.25	43000	12?	25?
O8	−0.31	−0.15	−4.1	−3.93	37000	10	20?
B0	−0.29	−0.13	−3.3	−3.34	31000	7.2	17
B1	−0.26	−0.11	−2.9	−2.6	24100	5.3	10.7
B2	−0.24	−0.1	−2.5	−2.2	21080	4.7	8.3
B3	−0.21	−0.08	−2	−1.69	18000	3.5	6.3
B4	−0.18	−0.07	−1.5	−1.29	15870	3	5
B5	−0.16	−0.06	−1.1	−1.08	14720	2.9	4.3
B8	−0.1	−0.02	0	−0.6	11950	2.3	3
A0	0	0.02	0.7	−0.14	9572	1.8	2.34
A2	0.06	0.08	1.3	0	8985	1.75	2.21
A5	0.14	0.16	1.9	0.02	8306	1.69	2.04
A7	0.19	0.19	2.3	0.02	7935	1.68	1.93
F0	0.31	0.3	2.7	0.04	7178	1.62	1.66
F2	0.36	0.35	3	0.04	6909	1.48	1.56
F5	0.44	0.4	3.5	0.02	6528	1.4	1.41
F8	0.53	0.47	4	−0.01	6160	1.2	1.25
G0	0.59	0.5	4.4	−0.05	5943	1.12	1.16
G2	0.63	0.53	4.7	−0.08	5811	1.08	1.11
G5	0.68	0.54	5.1	−0.11	5657	0.95	1.05
G8	0.74	0.58	5.6	−0.16	5486	0.91	0.97
K0	0.82	0.64	6	−0.23	5282	0.83	0.9
K2	0.92	0.74	6.5	−0.3	5055	0.75	0.81
K3	0.96	0.81	6.8	−0.33	4973	0.73	0.79
K5	1.15	0.99	7.5	−0.43	4623	0.64	0.65
K7	1.3	1.15	8	−0.54	4380	0.54	0.54
M0	1.41	1.28	8.8	−0.72	4212	0.48	0.46
M2	1.5	1.5	9.8	−0.99	4076	0.43	0.4
M5	1.6	1.8	12	−1.52	3923	0.38	0.34

[a]Ref. [68]. These values should be considered typical, but individual stars may differ from the values quoted here. Blanks imply that the data are not sufficiently well known to be quoted here.

Table G.8. Stellar data[a], Luminosity Class III (Giants)

Spectral type	B–V	V–R	M_V	BC	T_{eff} (K)	$R/R_\odot$
F0	0.31		1	0.04	7178	7
F2	0.36		0.9	0.04	6909	
F5	0.44		0.8	0.02	6528	
F8	0.54		0.7	−0.02	6160	8
G0	0.64		0.6	−0.09	5943	9
G2	0.76		0.5	−0.17	5811	10
G5	0.9	0.69	0.4	−0.29	5657	11
G8	0.96	0.7	0.3	−0.33	5486	12
K0	1.03	0.77	0.2	−0.37	5282	14
K2	1.18	0.84	0.1	−0.45	5055	17
K3	1.29	0.96	0.1	−0.53	4973	21
K5	1.44	1.2	0	−0.81	4623	40
K7	1.53		−0.1	−1.15	4380	60
M0	1.57	1.23	−0.2	−1.36	4212	100
M2	1.6	1.34	−0.2	−1.52	4076	130
M5	1.58	2.18	−0.2		3923	

[a]Ref. [68]. See notes to Table G.7. Masses of Type III giants are not well known but are likely $\approx$ 2–2.5 $M_\odot$.

Table G.9. Stellar data[a], Luminosity Class Ib (Supergiants)

Spectral type	B–V	V–R	M_V	BC	T_{eff} (K)	$R/R_\odot$
A0	0.01	0.03	−5	−0.12	9550	70
A2	0.05	0.07	−5	−0.02	9000	
A5	0.1	0.12	−5	0.01	8500	
A7	0.13		−4.9	0.02	8300	
F0	0.16	0.21	−4.8	0.02	8030	80
F2	0.21	0.26	−4.8	0.02	7780	
F5	0.33	0.35	−4.7	0.04	7020	
F8	0.55	0.45	−4.6	−0.02	6080	
G0	0.76	0.51	−4.6	−0.18	5450	100
G2	0.87	0.58	−4.6	−0.26	5080	
G5	1	0.67	−4.5	−0.35	4850	
G8	1.13	0.69	−4.5	−0.41	4700	
K0	1.2	0.76	−4.5	−0.47	4500	110
K2	1.29	0.85	−4.5	−0.53	4400	
K3	1.38	0.94	−4.5	−0.68	4230	130
K5	1.6	1.2	−4.5	−1.52	3900	200
K7	1.62		−4.5		3870	
M0	1.65	1.23			3850	230
M2	1.65	1.34			3800	

[a]Ref. [68]. See notes to Table G.7. Masses of Type Ib supergiants are not well known but are likely $\approx 10\,M_\odot$.

Table G.10. Solar interior model[a]

$R/R_{\odot}^{b}$	$M/M_{\odot}{}^{c}$	L^{d} (erg s^{-1})	T^{e} (K)	ρ^{f} (g cm^{-3})
1.14E-03	1.64E-07	5.66E+27	1.58E+07	1.56E+02
2.48E-02	1.64E-03	5.50E+31	1.56E+07	1.49E+02
1.03E-01	8.29E-02	1.84E+33	1.30E+07	8.57E+01
1.87E-01	3.00E-01	3.52E+33	9.82E+06	3.95E+01
2.73E-01	5.44E-01	3.83E+33	7.40E+06	1.62E+01
3.55E-01	7.21E-01	3.85E+33	5.80E+06	6.50E+00
4.39E-01	8.36E-01	3.85E+33	4.64E+06	2.56E+00
5.29E-01	9.09E-01	3.85E+33	3.72E+06	1.01E+00
6.25E-01	9.53E-01	3.85E+33	2.94E+06	4.02E-01
7.22E-01	9.77E-01	3.85E+33	2.09E+06	1.76E-01
8.04E-01	9.90E-01	3.85E+33	1.33E+06	8.80E-02
8.66E-01	9.96E-01	3.85E+33	8.41E+05	4.41E-02
9.09E-01	9.99E-01	3.85E+33	5.34E+05	2.21E-02
9.39E-01	1.00E+00	3.85E+33	3.41E+05	1.12E-02
9.59E-01	1.00E+00	3.85E+33	2.19E+05	5.58E-03
9.72E-01	1.00E+00	3.85E+33	1.43E+05	2.76E-03
9.81E-01	1.00E+00	3.85E+33	9.57E+04	1.34E-03
9.86E-01	1.00E+00	3.85E+33	6.44E+04	6.52E-04
9.90E-01	1.00E+00	3.85E+33	4.48E+04	3.10E-04
9.93E-01	1.00E+00	3.85E+33	3.35E+04	1.39E-04
9.95E-01	1.00E+00	3.85E+33	2.65E+04	5.93E-05
9.96E-01	1.00E+00	3.85E+33	2.18E+04	2.43E-05
9.97E-01	1.00E+00	3.85E+33	1.84E+04	9.69E-06
9.98E-01	1.00E+00	3.85E+33	1.58E+04	3.81E-06
9.988E-01	1.00E+00	3.85E+33	1.36E+04	1.50E-06
9.993E-01	1.00E+00	3.85E+33	1.16E+04	6.04E-07
9.998E-01	1.00E+00	3.85E+33	8.72E+03	2.76E-07
1.0000E+00	1.00E+00	3.85E+33	5.67E+03	1.55E-07
1.0001E+00	1.00E+00	3.85E+33	5.11E+03	9.61E-08
1.0002E+00	1.00E+00	3.85E+33	4.82E+03	5.62E-08
1.0003E+00	1.00E+00	3.85E+33	4.70E+03	3.10E-08
1.0004E+00	1.00E+00	3.85E+33	4.65E+03	1.68E-08
1.0005E+00	1.00E+00	3.85E+33	4.64E+03	9.07E-09
1.0006E+00	1.00E+00	3.85E+33	4.63E+03	4.87E-09
1.0007E+00	1.00E+00	3.85E+33	4.63E+03	2.58E-09
1.0008E+00	1.00E+00	3.85E+33	4.63E+03	1.34E-09
1.0009E+00	1.00E+00	3.85E+33	4.63E+03	6.67E-10
1.0011E+00	1.00E+00	3.85E+33	4.63E+03	3.03E-10

[a]Modelled conditions for the interior of the Sun adopting $T_{eff} = 5779.6$ K, $L_{\odot} = 3.851 \times 10^{33}$ erg s^{-1}, X= 0.708, Z= $0.02 R_{\odot} = 6.9598 \times 10^{10}$ cm and $M_{\odot} = 1.98910 \times 10^{33}$g. (Ref. [70]). (Reproduced by permission of D. Guerther)
[b]Radius as a fraction of the Solar radius.
[c]Mass as a fraction of the Solar mass.
[d]Luminosity.
[e]Temperature.
[f]Mass density.

Table G.11. Solar photosphere model[a]

$\log(\tau_0)$[b]	T[c] (K)	$\log(P_g)$[d] [log(dyn cm^{-2})]	$\log(P_e)$[e] [log(dyn cm^{-2})]	$\log(\kappa_0/P_e)$[f] [log(cm^3 s^2 g^{-2})]	x[g] (km)
−4.0	4310	3.13	−1.42	−1.22	−476
−3.8	4325	3.29	−1.28	−1.23	−443
−3.6	4345	3.42	−1.15	−1.24	−415
−3.4	4370	3.54	−1.03	−1.25	−389
−3.2	4405	3.66	−0.92	−1.27	−365
−3.0	4445	3.78	−0.80	−1.28	−340
−2.8	4488	3.89	−0.69	−1.30	−317
−2.6	4524	4.00	−0.58	−1.32	−293
−2.4	4561	4.11	−0.48	−1.33	−269
−2.2	4608	4.22	−0.37	−1.35	−245
−2.0	4660	4.33	−0.26	−1.37	−221
−1.8	4720	4.44	−0.14	−1.40	−196
−1.6	4800	4.54	−0.01	−1.42	−172
−1.4	4878	4.65	0.11	−1.45	−147
−1.2	4995	4.76	0.26	−1.49	−122
−1.0	5132	4.86	0.42	−1.54	−96
−0.8	5294	4.96	0.60	−1.58	−73
−0.6	5490	5.04	0.83	−1.64	−51
−0.4	5733	5.12	1.10	−1.71	−31
−0.2	6043	5.18	1.43	−1.79	−14
0.0	6429	5.22	1.80	−1.88	0
0.2	6904	5.26	2.21	−1.99	11
0.4	7467	5.28	2.64	−2.10	19
0.6	7962	5.30	2.96	−2.17	26
0.8	8358	5.32	3.20	−2.22	33
1.0	8630	5.43	3.34	−2.25	41
1.2	8811	5.36	3.46	−2.26	50

[a]Modelled conditions for the surface region of the Sun (Ref. [68]). The model assumes that the solar effective temperature is $T_{eff} = 5770$ K and that $\log(g) = 4.44$, where g (cm s^{-2}) is the surface gravity.
[b]Log of the optical depth (Sect. 5.4.1) at the reference wavelength of 500 nm. $\text{Log}(\tau_0) = 0$ represents the 'surface' of the Sun's photosphere.
[c]Temperature at the corresponding optical depth.
[d]Log of the gas pressure.
[e]Log of the electron pressure.
[f]Log of the ratio of the mass absorption coefficient (Sect. 5.4.1), κ_0 (cm^2 g^{-1}) at the reference wavelength of 500 nm, to P_e.
[g]Geometrical depth, where 0 corresponds to the photosphere surface and positive values are below the surface.

Table G.12. Acronym key to bibliography

A&A	Astronomy & Astrophysics
A&ARv	Astronomy & Astrophysics Review
A&AS	Astronomy and Astrophysics Supplement Series
AIP Conf. Proc	American Institute of Physics Conference Proceedings
AJ	The Astronomical Journal
AO	Applied Optics
ApJ	The Astrophysical Journal
ApJS	Astrophysical Journal Supplement Series
ARAA	Annual Review of Astronomy and Astrophysics
AREPS	Annual Review of Earth and Planetary Sciences
ASP Conf. Ser.	Astronomical Society of the Pacific Conference Series
Astron. Nachr.	Astronomische Nachrichten
GeoRL	Geophysical Research Letters
JApA	Journal of Astrophysics and Astronomy
JGR	Journal of Geophysical Research
JKAS	Journal of the Korean Astronomical Society
JOSA	Journal of the Optical Society of America
JQSRT	Journal of Quantitative Spectroscopy & Radiative Transfer
MNRAS	Monthly Notices of the Royal Astronomical Society
Natur	Nature
NewA	New Astronomy
NewAR	New Astronomy Reviews
OE	Optical Engineering
PASJ	Publications of the Astronomical Society of Japan
PASP	Publications of the Astronomical Society of the Pacific
Phys. Rev.	Physical Review
Rev. Mod. Phys	Reviews of Modern Physics
Sci	Science
SIAMR	Society of Industrial and Applied Mathematics Review

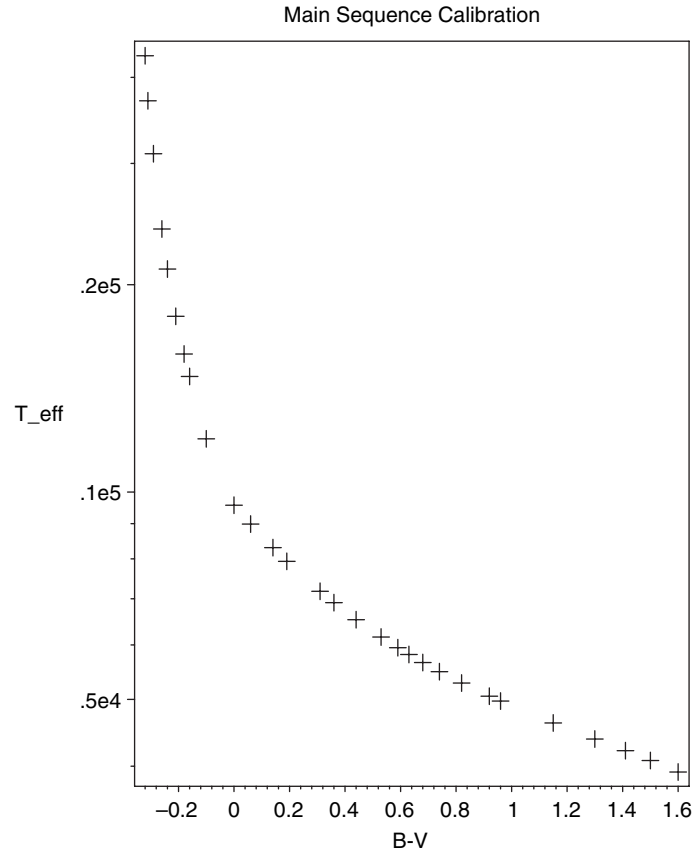


Figure G.1. B–V colour – temperature calibration for the Main Sequence, from the data of Table G.7

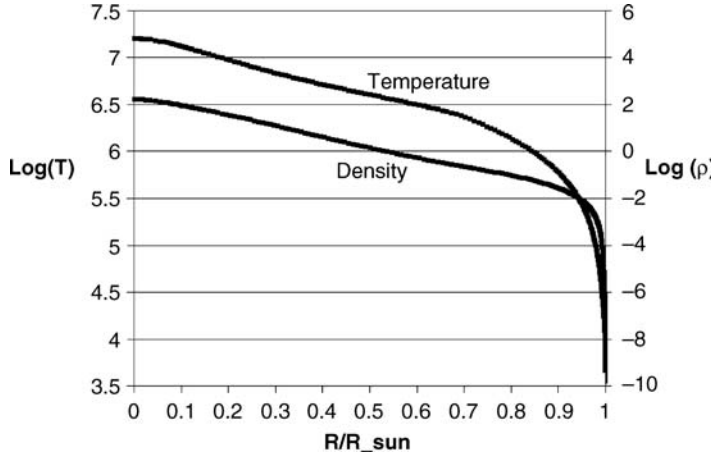


Figure G.2. Solar interior model from the data of Table G.10. The temperature is in kelvins and density in g cm^{-3}.

References

1. AbdelSalam, H. M., Saha, P., & Williams, L. L. R., 1998, AJ, **116**, 1541
2. Afflerbach, A., *et al.,* 1996, ApJS, **106**, 423
3. Afonso, C., *et al*., 2003, A&A, **400**, 951
4. Alcock, C., *et al*., 2001, ApJ, **550**, L169
5. Alcock, C., *et al*., 1998, ApJ, **499**, L9
6. Allen, G. E., *et al*., 1997, ApJ, **487**, L97
7. Anderson, M., Rudnick, L., Leppik, P., Perley, R., & Braun, R., 1991, ApJ, **373**, 146
8. Anderson, M., & Rudnick, L., 1996, ApJ, **456**, 234
9. Alves, J. F., Lada, C. J., & Lada, E. A., 2001, Natur, **409**, 159
10. Auriere, M., Lauzeral, C., & Ortolani, S., 1990, Natur, **344**, 638
11. Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., & Witzel, A., 1977, A&A, **61**, 99
12. Bagenal, F. 1992, AREPS, **20**, 289
13. Beaulieu, J.-P., *et al*., 2006, Natur, **439**, 437
14. Bekefi, B., *Radiation Processes in Plasmas*, John Wiley & Sons, Inc., New York, 1966
15. Beck, R., & Krause, M., 2005, Astron. Nachr., **326**, 414
16. Bennett, C. L., *et al*., 2003, ApJS, **148**, 1
17. Bessell, M. S., 1990, PASP, **102**, 1181
18. Bessell, M. S., Castelli, F., & Plez, B., 1998, A&A, **333**, 231
19. Bessell, M. S., & Brett, J. M., 1988, PASP, **100**, 1134
20. Bhattacharjee, P., & Sigl, G., 2000, Physics Reports, **327**, 109
21. Bignami, G. F., Caraveo, P. A., de Luca, A., & Mereghetti, S., 2003, Natur, **423**, 725
22. Bonamente, M., Lieu, R., Joy, M. K., & Nevalainen, J. H., 2002, ApJ, **576**, 688
23. Born, M., *Atomic Physics*, Blackie & Sons Ltd., London, 1946
24. Boselli, A., Lequeux, J., & Gavazzi, G., 2002, A&A, **384**, 33
25. Brogan, C. L., 2006, ASP Conf. Ser., **345**, 187
26. Brown, D., Hupp, E., & Martinez, C., 2006, NASA News Release: 2006–097, http://saturn.jpl.nasa.gov/news/

Astrophysics: Decoding the Cosmos Judith A. Irwin

ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

27. Brown, R. L., & Mathews, W. G., 1970, ApJ, **160**, 939
28. Brussaard, P. J., & van de Hulst, H. C., 1962, Rev. Mod. Phys. **34**, 507
29. Buckley, L., 2005, Earth Policy Institute, http://www.earth-policy.org/Indicators/Temp/2004.htm
30. Cairns, I., McCusker, C. B. A., Peak, L. S., & Woolcott, R. L. S., 1969, Phys. Rev. **186**, 1394
31. Canada France Hawaii Telescope Observatory Manual, Version 1.0, January 2003, http://www.cfht.hawaii.edu/Instruments/ObservatoryManual/
32. Carlstrom, J. E., Holder, G. P., & Reese, E. D., 2002, ARAA, **40**, 643
33. Carroll, B. W., & Ostlie, D. A., *An Introduction to Modern Astrophysics*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1996
34. Casadei, D., & Bindi, V., 2004, ApJ, **612**: 262–267
35. Chabrier, G., 2003, PASP, **115**, 763
36. Charbonneau, D. *et al*., 2005, ApJ, **626**, 523
37. Chandra Supernova Remnant Catalog at http://snrcat.cfa.harvard.edu/index.html
38. Chauvin, G. *et al*., 2004, A&A, **425**, 29
39. Chen, G., Gott III, J. R., & Ratra, B., 2003, PASP, **115**, 1269
40. Church, M. J., *et al*., 2005, MNRAS, **359**, 1336
41. Ciddor, P. E., 1996, AO, **35**, 1566
42. Churchwell, E., 2002, ARAA, **40**, 27
43. Comito, C., *et al*., 2005, ApJS, **156**, 127
44. A. N. Cox (Ed.), *Allen's Astrophysical Quantities*, Springer-Verlag, New York, 1999
45. Deming, D., Seager, S., Richardson, L. J., & Harrington, J., 2005, Natur, **434**, 740
46. Dickey, J. M., & Lockman, F. J., 1990, ARAA, **28**, 215
47. Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J., 2003, ApJ, **585**, 801
48. Diehl, R., Prantzos, N., & von Ballmoos, P., 2005, astro-ph/0502324
49. Diehl, R., *et al*., 2006, Natur, **439**, 45
50. Dopita, M. A., & Sutherland, R. S., *Astrophysics of the Diffuse Universe*, Springer, New York, 2003
51. Darbon, S., Perrin, J.-M., & Sivan, J.-P., 1998, A&A, **333**, 264
52. Draine, B. T., 2003, ARAA, **41**, 241
53. Dravins, D., Lindegren, L., Mezey, E., & Young, A. T., 1997, PASP, **109**, 725
54. Dunne, L., *et al*., 2003, Natur, **424**, 285
55. Dupree, A. K., & Goldberg, L., 1970, ARAA, **8**, 231
56. Ebeling, H., *et al*., 1998, MNRAS, **301**, 881
57. Ellison, D. C., Drury, L. O'C., Meyer, J.-P., 1997, ApJ, **487**, 197
58. Elvis, M., 2000, ApJ, **545**, 63
59. Freeman, K., & Bland-Hawthorn, J., 2002, ARAA, **40**,487
60. Frei, Z., Guhathakurta, P., Gunn, J. E., & Tyson, J. A., 1996, AJ, **111**, 174
61. Fukui, Y., et. al., 2006, Sci, **314**, 106
63. Ghez, A. M., *et al*., 2005, ApJ, **620**, 744
64. Gingerich, O., *The Book Nobody Read*, Penguin Books, London, 2004
65. Goncharov, A. V., Owner-Petersen, M., Andersen, T., & Beckers, J. M., 2002, Optical Engineering **41**, 1065–1072, Donald C. O'Shea, Ed
66. Gordon, M. A., & Sorochenko, R. L., 2002, *Radio Recombination Lines*, Kluwer Academic Publishers, Dordrecht
67. Gratton, R. G., *et al*., 2003, A&A, **408**, 529
68. Gray, D. F., 2005, *The Observation and Analysis of Stellar Photospheres, 3rd Ed.*, Cambridge University Press, Cambridge
69. Grosdidier, Y., Moffat, A. F. J., Joncas, G., & Acker, A., 1998, ApJ, **506**, L127

70. Guenther, D. B., Demarque, P., Kim, Y.-C., & Pinsonneault, M. H., 1992, ApJ, **387**, 372
71. R. Gupta (Ed.), *Observer's Handbook 2005*, Royal Astronomical Society of Canada, University of Toronto Press, Toronto, 2004
72. Handa, T., *et al*., 1987, PASJ, **39**, 709
73. Hansen, B. M. S., *et al*., 2002, ApJ, **574**, L155
74. Harrison, E. R. *Cosmology, The Science of the Universe*, Cambridge University Press, Cambridge, 1981
75. Hartmann, W. K. *Moons and Planets*, Thomson Books, Belmont California, 2005
76. Harwit, M. *Astrophysical Concepts*, Springer-Verlag, New York, 1988
77. Heitler, W. *The Quantum Theory of Radiation*, *3rd Ed.*, Dover Publications Inc., New York, 1984
78. Iyudin, A. F. *et al*., 1994, A&A, **284**, L1
79. Jackson, J. D. *Classical Electrodynamics, 3rd Ed.*, John Wiley & Sons, Inc., Hoboken, New Jersey, 1998
80. Jacoby, G. H., Hunter, D. A., & Christian, C. A., 1984, ApJS, **56**, 257
81. Kanekar, N., & Briggs, F. H., 2004, NewAR, **48**, 1259
82. Kantharia, N., G., Anantharamaiah, K. R., & Payne, H. E., 1998, ApJ, **506**, 758
83. Kantharia, N., G., & Anantharamaiah, K. R., 2001, JApA, **22**, 51
84. Kaspi, V., Roberts, M. S. E., & Harding, A. K., 2006, in *Compact Stellar X-ray Sources*, eds. W. H. G. Lewin and M. van der Klis, Cambridge University Press, Cambridge p.279
85. Kassim, N. E., Perley, R. A., Dwarakanath, K. S., & Erickson, W. C., 1995, ApJ, **455**, L59
86. Kiminki, D. C., *et al*., 2006, astro-ph/0609772
87. King, L. J., *et al*., 1998, MNRAS, **295**, L41
88. Kishimoto, M.,1999, ApJ, **518**, 676
89. Kodaira, K., *et al*., 2003, PASJ, **55**, L17
90. Kollmeier, J. A. *et al*., 2006, ApJ. **648**, 128
91. Korn, A., *et al*., 2006, *astro-ph/0610077*
92. Kurucz, R. L., 1979, ApJS, **40**, 1
93. Kwok, S., 2005, JKAS, **39**, 271
94. Lada, C. J., Bergin, E. A., Alves, J. F., & Huard, T. L., 2003, ApJ, **586**, 286
95. Lallement, R., *et al*., 2003, A&A, **411**, 447
96. K. R. Lang, *Astrophysical Formulae*, Springer-Verlag, New York, 1999
97. Leibundgut, B., *et al*., 1996, ApJ, **466**, L21
98. Lerate, M. R., *et al*., 2006, MNRAS, **370**, 597
99. Lisenfeld, U., Isaak, K. G., & Hills, R., 2000, MNRAS, **312**, 433
100. Livesey, D. L. *Atomic and Nuclear Physics*, Blaisdell Publishing Company, Waltham, Massachusetts, 1966
101. Longair, M. S., *High Energy Astrophysics, Vol. 1, Second Ed.*, Cambridge University Press, Cambridge, 1992
102. Lucretius, *The Way Things Are*, translated by R. Humphries, Indiana University Press, Bloomington, 1969
103. Massey, P., & Hunter, D. A., 1998, ApJ **493**, 180
104. Mathews, J. D., Janches, D., Meisel, D. D., & Zhou, Q.-H., 2001, GeoRL, **28**, 1929
105. NASA/IPAC Extragalactic Database (NED) operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (http://nedwww.ipac.caltech.edu)
106. McDonald, A. B. *et al*., 2002, AIP Conf. Proc., **646**, 43
107. Mewe, R., Lemen, J. R., & van den Oord, G. H. J., 1986, A&AS, **65**, 511
108. Montañés-Rodriguez, P. *et al*., 2005, ApJ, **629**, 1175

109. Netterfield, C. B., *et al*., 2002, ApJ, **571**, 604
110. Neuhäuser, R. *et al*., 2005, A&A, **435**, L13
111. Newbury, P. R., & Spiteri, R. J., 2002, SIAMR, **44**, 111
112. Nicastro, F., *et al*., 2005, Natur, **433**, 495
113. Ord, S. M., van Straten, W., Hotan, A., W., & Bailes, M., 2004, MNRAS, **352**, 804
114. Olive, K. A., Steigman, G., & Skillman, E. D., 1997, ApJ, **483**, 788
115. Osterbrock, D. E. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, University Science Books, Mill Valley, California, 1989
116. Pacholczyk, A. G., *Radio Astrophysics*, W. H. Freeman and Co., San Francisco, 1970
117. Palacios, A., *et al*., 2005, A&A, **429**, 613
118. Pasquini, L. *et al*., 2004, A&A, **426**, 651
119. Pedersen, K., *et al*., 2006, NewA, **11**, 465
120. Pedlar, A., Booler, R. V., Spencer, R. E., & Stewart, O. J., 1983, MNRAS, **202**, 647
121. Penney, C. M., 1969, JOSA, **59**, 34
122. Pérez-Torres, M. A., Alberdi, A., & Marcaide, J. M., 2001, A&A, **374**, 997
123. Perley, R. A., Dreher, J. W., & Cowan, J. J., 1984, ApJ, **285**, L35
124. Pickles, A. J. 1998, PASP, **110**, 863
125. Plüschke, S., *et al*., 2001, in *Exploring the Gamma-Ray Universe* (4th INTEGRAL Workshop), Eds. B. Battrick, A. Gimenez, V. Reglero and C. Winkler, ESA SP-459, Noordwijk, 55–58
126. Polanyi, J. 2006, in a talk delived to a History class at the University of Toronto, on Oct. 2, 2006, reprinted in the *Toronto Star* newspaper, Sunday, October 22, 2006
127. Porquet, D., *et al*., 2003, A&A, **407**, L17
128. Ramaseshan, S., & Rao, C. R. (compilers), 1988, *C. V Raman, A Pictorial Biography*, Indian Academy of Sciences, Bangalore
129. Rauch, M., 1998, ARAA, **36**, 267
130. Reiprich, T. H., & Böhringer, H., 2002, ApJ, **567**, 716
131. Raymond, J. C., & Smith, B. W., 1977, ApJS, **35**, 419
132. Readhead, A. C. S., 1994, ApJ, **426**, 51
133. Regan, M. W., *et al*., 2001, ApJ, **561**, 218
134. Rengelink, R. B., *et al*., 1997, A&AS, **124**, 259
135. Richer, H. B., *et al*., 2006, Sci, **313**, 936
136. Rieke, G. H., Lebofsky, M. J., & Low, F. H., 1985, AJ, **90**, 900
137. Roelfsema, P. R., & Goss, W. M., 1992, A&ARv, **4**, 161
138. Rohlfs, K., *Tools of Radio Astronomy*, Springer-Verlag, Berlin, 1986
139. Roshi, D. A., & Anantharamaiah, K. R., 2000, ApJ, **535**, 231
140. Roshi, D. A., & Anantharamaiah, K. R., 2001, ApJ, **557**, 226
141. Roussel, H., Sadat, R., & Blanchard, A., 2000, A&A, **361**, 429
142. Ruhl, J. E. *et al*., 2003, ApJ, **599**, 786
143. Rybicki, G. B., & Lightman, A. P., *Radiative Processes in Astrophysics*, Wiley-Interscience, John Wiley & Sons, New York, 1979
144. Santos, M. R., 2004, MNRAS, **349**, 1137
145. Sauer, D., & Jedamzik, K., 2002, A&A, **381**, 361
146. Scalo, J., & Elmegreen, B. G., 2004, ARAA, **42**, 275
147. Schlickeiser, R. *Cosmic Ray Astrophysics*, Springer, Berlin, 2002
148. Schneider, N., *et al*., 2006, A&A, **458**, 855
149. Schneider, P., Ehlers, J., & Falco, E. E., *Gravitational Lenses*, Springer-Verlag, New York, 1992
150. Schödel, R., *et al*., 2002, Natur, **419**, 694
151. Schraml, J., & Mezger, P. G., 1969, ApJ, **156**, 269
152. Schwarzschild, M. *Structure and Evolution of Stars*, Dover Publications, Inc., New York, 1958
153. Shaver, P., 1980, A&A, **91**, 279

154. Shore, S. N. *The Tapestry of Modern Astrophysics*, Wiley-Interscience, a John Wiley & Sons, Inc. Publication, New Jersey, 2003
155. Shu, F. H. *The Physics of Astrophysics, Vol. I, Radiation*, University Science Books, Mill Valley California, 1991
155a. Sittler Jr., E.C., and Guhathakurta, M., 1999, ApJ, **523**, 812
156. Smoot, G. F., & Scott, D., 2002, in ''Review of Particle Properties'', K. Hagiwara, et al., Phys. Rev. D66, 010001
157. Sneep, M., & Ubachs, W., 2005, JQSRT, **92**, 293
158. Snowden, S. L., *et al*., 1995, ApJ, **454**, 643
159. Spergel, D. N. *et al*., 2003, ApJS, **148**, 175
160. Spitzer Jr., L., *Physical Processes in the Interstellar Medium*, John Wiley & Sons, New York, 1978
161. Springel, V., *et al*., 2005, Natur, **435**, 629
162. Stil, J., & Irwin, J. A., 2001, ApJ, **563**, 816
163. Stone, R. C. 1996, PASP, **108**, 1051
164. Strong, A. W., *et al*., 2005, A&A, **444**, 495
165. Strong, A. W., Moskalenko, I. V., & Reimer, O., 2004, ApJ, **613**, 962
166. Sutherland, R. S., 1998, MNRAS, **300**, 321
166a. Tammann, G. A., Sandage, A., and Reindl, B., 2003, A&A, **404**, 423
167. Taylor, A. R. *et al*., 2003, AJ, **125**, 3145
168. Taylor, B. N., & Mohr, compilers, P. J., 2002, ''The NIST Reference on Constants, Units and Uncertainty'', http://physics.nist.gov/cuu
169. Tercero, B., Cernicharo, J., & Pardo, J. R., 2005 in *IAU Symp. 231, Astrochemistry Throughout the Universe: Recent Successes and Current Challenges*, 161
170. Tercero, B., Cernicharo, J., & Pardo, J. R., 2006, in preparation
171. Tonry, *et al*., 2003, ApJ, 594, 1, *online data at* http://www.journals.uchicago.edu/ApJ/journal/issues/ApJ/v594n1/57840/57840.html, *Table 8*
172. Trotter, A. S. *et al*., 1998, ApJ, **495**, 740
173. Uchiyama, Y., Takahashi, T., Aharonian, F. A., & Mattox, J. R., 2002, ApJ, **571**, 866
174. Unsöld, A., & Baschek, B., *The New Cosmos*, *4th Ed.*, Springer-Verlag, New York, 1991
175. Vallée, J. P., 2004, NewAR, **48**, 763
176. van der Werf, S. Y., 2003, AO, **42**, 354
177. Verschuur, G. L. and Kellermann, K. I., ed. *Galactic and Extragalactic Radio Astronomy, 2nd Ed.*, Springer-Verlag, Berlin, 1988
178. Walker, G., *Astronomical Observations – An Optical Perspective*, Cambridge University Press, Cambridge, 1987
179. Westfall, R. S., *Never at Rest – A Biography of Isaac Newton*, Cambridge University Press, Cambridge, 1987
180. Whittet, D. C. B., *Dust in the Galactic Environment, 2nd Ed.*, Institute of Physics Publishing, Bristol, 2003
181. Willingale, R., *et al*., 2002, A&A, **381**, 1039
182. Willingale, R., *et al*., 2003, MNRAS, **343**,995
183. Wilson, C. D., 1995, ApJ, **448**, L97
184. Yun, M., Ho., P. T. P., & Lo, K. Y., 1994, Natur, **372**, 530
185. Yun, M. 1999, in *Proceedings of IAU Symposium 186: Galaxy Interactions at Low and High Redshift*, ed. J. E. Barnes and D. Sanders, p. 81
186. van de Hulst, H. C., 1957, IAU Symp. 4, *Radio Astronomy*, ed. H. C. van de Hulst, New York, Cambridge University Press, p.3
187. Zarka, P., 1998, JGR, **103**, 20,159
188. Zwaan, M. A., van Dokkum, P. G., & Verheijen, M. A. W., 2001, Sci, **293**, 1800

Index

Items in italic refer to figures, while items appearing in bold refer to tables.

Astrophysics: Decoding the Cosmos Judith A. Irwin

ISBN: 978-0-470-01305-2(HB) 978-0-470-01306-9(PB)

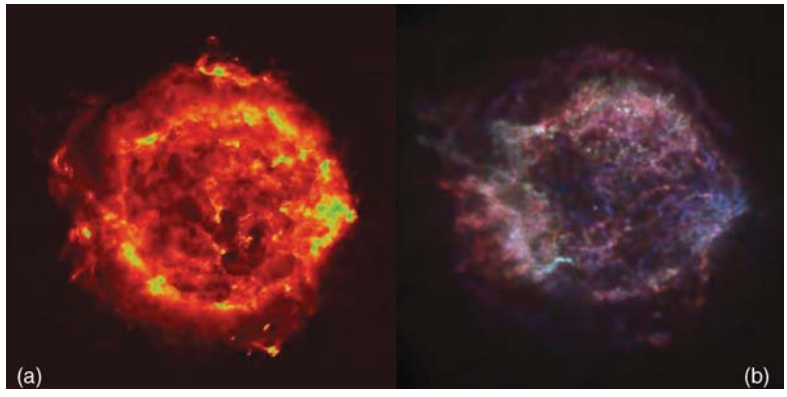


Figure 1.2

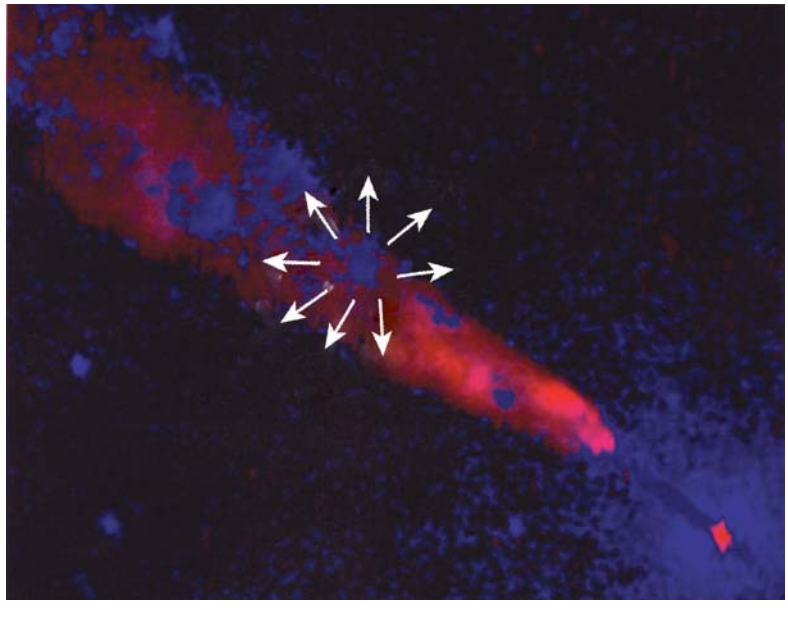


Figure 1.4

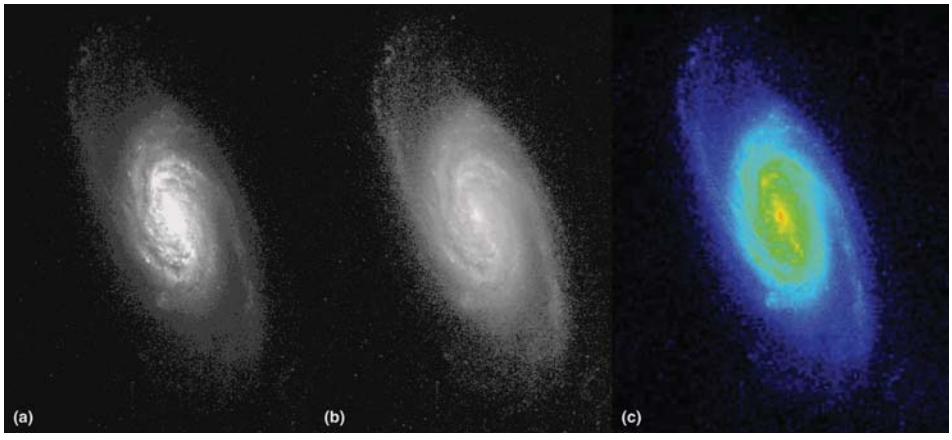


Figure 2.18

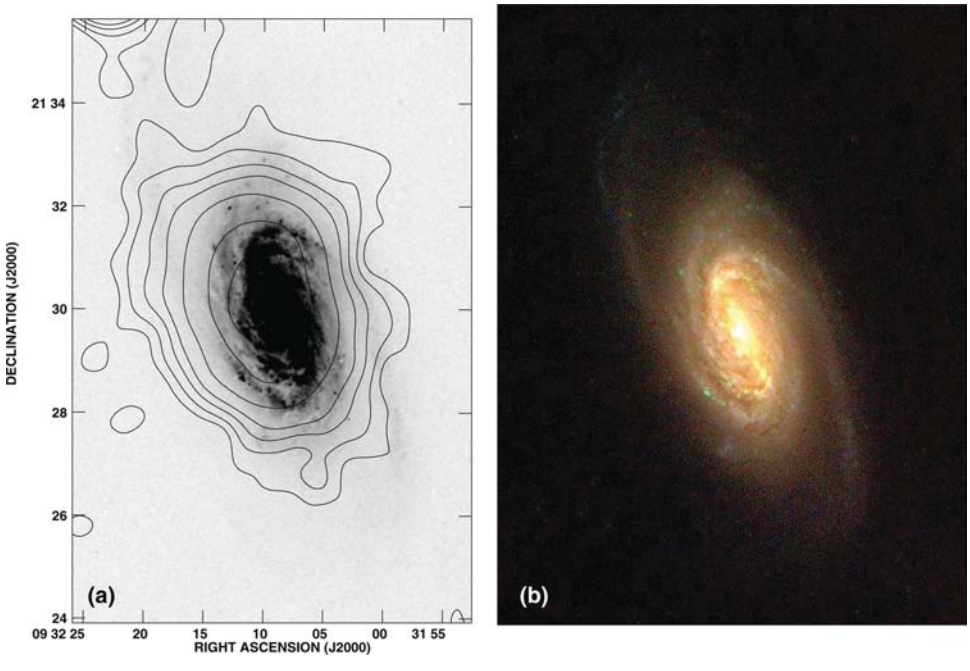


Figure 2.19

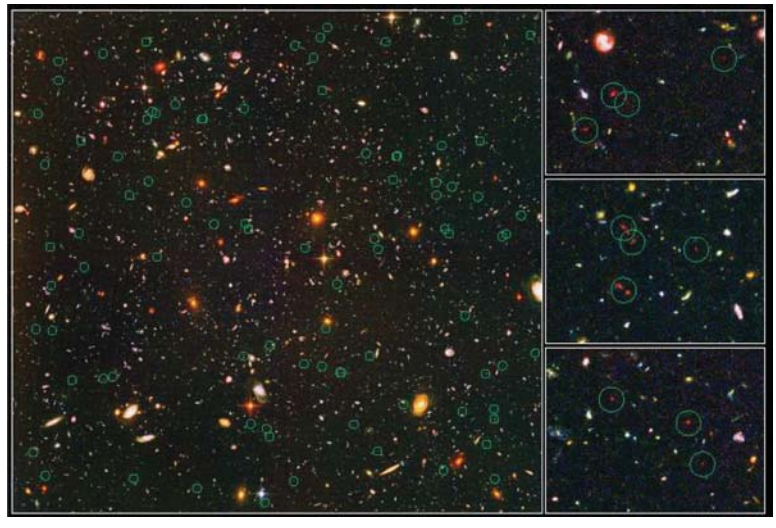


Figure 3.1

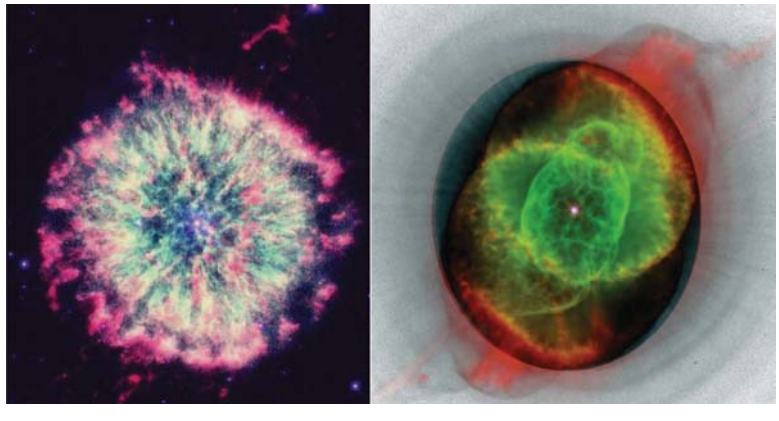


Figure 3.7

Figure 3.13

Figure 3.22

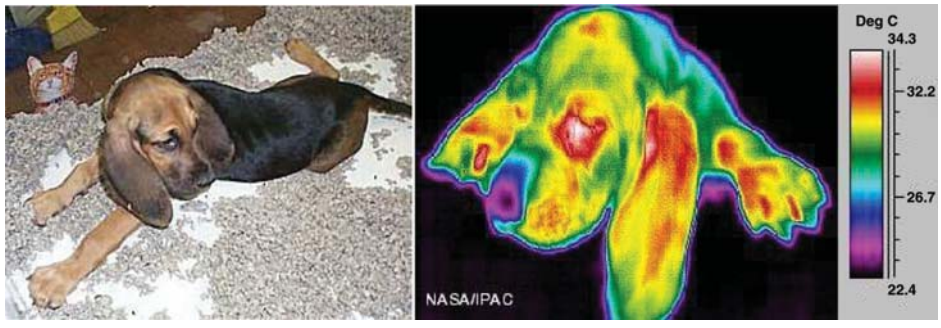


Figure 4.9

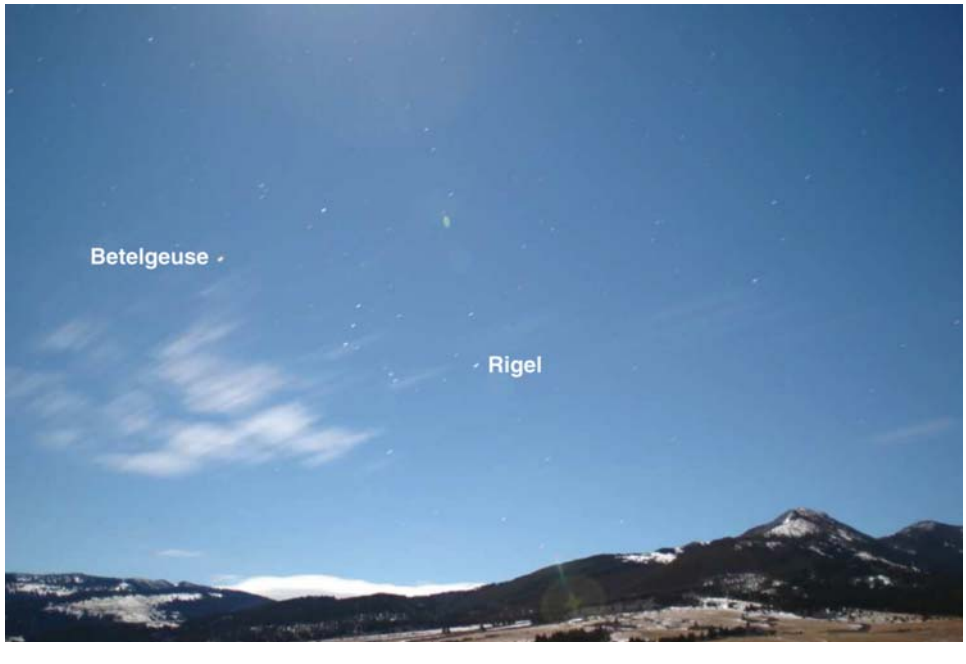


Figure 5.8

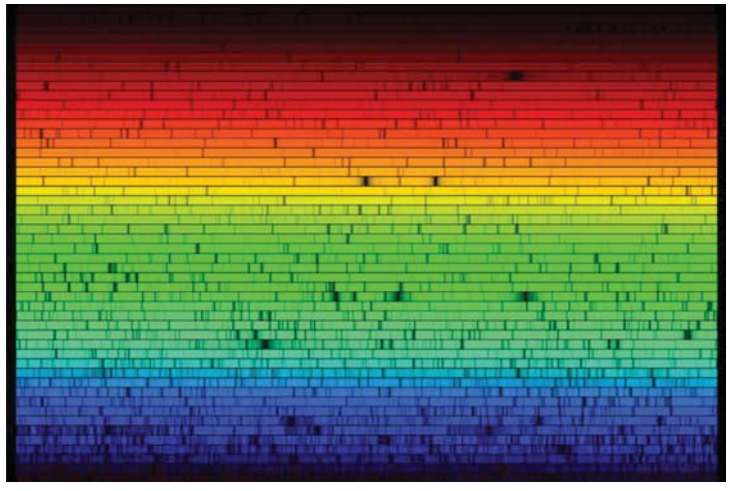


Figure 6.4

Figure 6.6

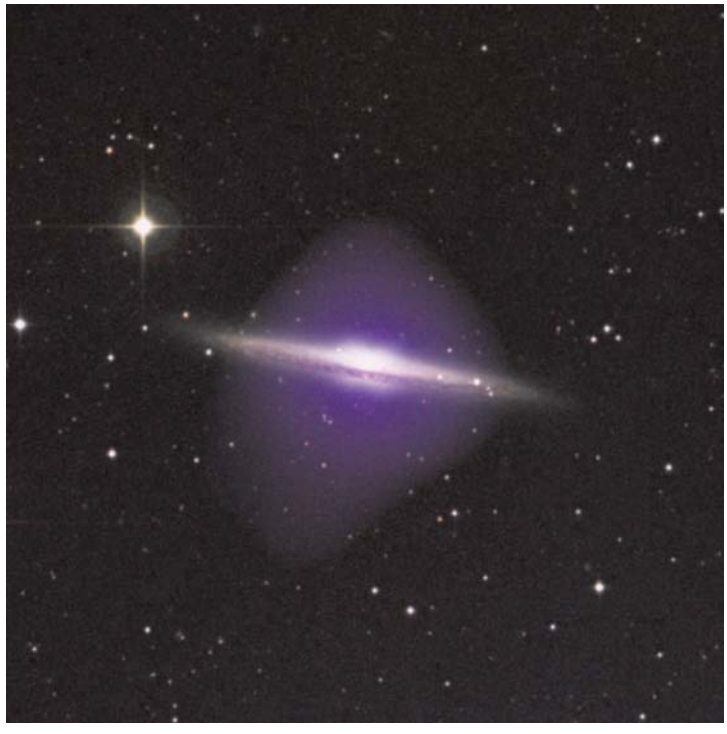


Figure 8.7

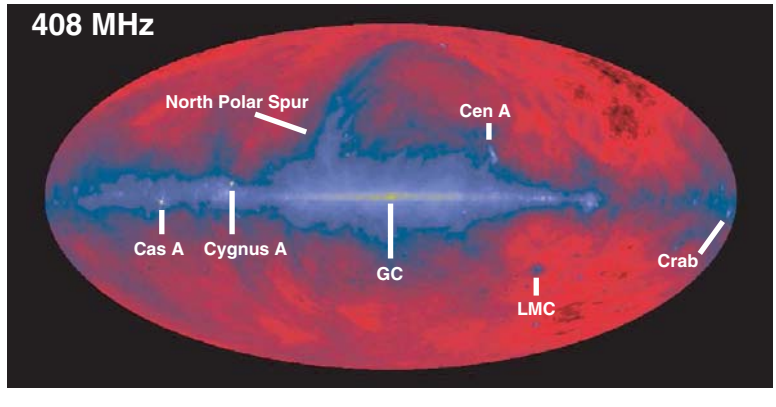


Figure 8.15

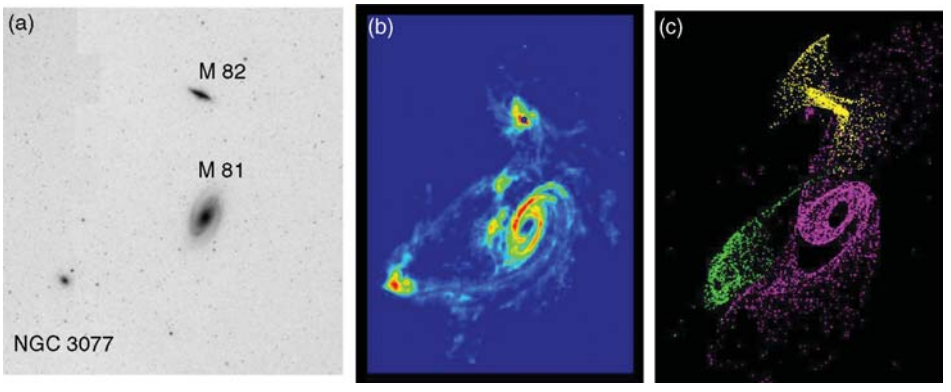


Figure 9.7

Figure 10.5