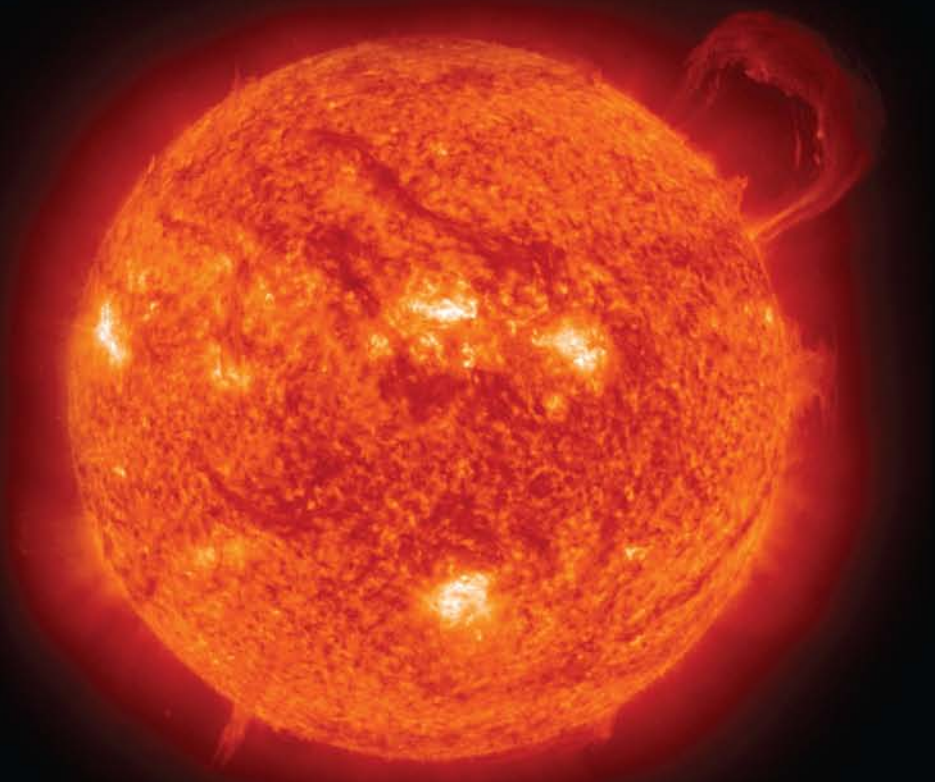
Series in PURE and APPLIED PHYSICS

Physics of the Sun

A First Course



Dermott J. Mullan

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Series in PURE and APPLIED PHYSICS

Physics of the Sun

CRC SERIES *in* PURE *and* APPLIED PHYSICS

Dipak Basu
Editor-in-Chief

PUBLISHED TITLES

Handbook of Particle Physics
M. K. Sundaresan

High-Field Electrodynamics
Frederic V. Hartemann

Fundamentals and Applications of Ultrasonic Waves
J. David N. Cheeke

Introduction to Molecular Biophysics
Jack A. Tuszynski
Michal Kurzynski

Practical Quantum Electrodynamics
Douglas M. Gingrich

Molecular and Cellular Biophysics
Jack A. Tuszynski

Concepts in Quantum Mechanics
Vishu Swarup Mathur
Surendra Singh

Physics of the Sun: *A First Course*
Dermott J. Mullan

Series in PURE and APPLIED PHYSICS

Physics of the Sun

A First Course

Dermott J. Mullan
University of Delaware
Newark, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20131120

International Standard Book Number-13: 978-1-4200-8308-8 (eBook - PDF)

Contents

Preface

The goal of this course is to undertake a quantitative examination of the physical structure of the Sun, the massive object which dominates the solar system, and which helps to support life as we know it on Earth. The text is aimed at upper-level physics undergraduates, and at graduate students in their first or second year of graduate study.

In the broadest context, the human race has one principal question that it requires solar physicists to answer: how stable is the Sun's output of energy? Specifically, is its power output steady enough that the Earth will not be subjected to fluctuations in heating which could lead to serious negative effects on life?

In order to arrive at reliable answers to these questions, we must begin with what we know about the Sun. And the place where reliable knowledge starts is with observations. A variety of instruments is at hand for observing the Sun. Using the unaided eye, we know that the Sun appears as a circular object (a "disk") with a sharp edge (the "limb"). Apart from those pieces of information, however, the eye is not especially useful. The Sun's light overwhelms the rods and cones in the eye, and the data handling system cannot deal with the flood of photons which have to be processed. Since the time of Galileo in the early 1600s, optical telescopes have been used to observe the Sun. These observations have led to the discovery of certain features on the visible surface, i.e., features which exist in the region known as the "photosphere." The best known among these features are dark regions called "sunspots."

Access to spectroscopy in the 1800s led to the discovery of a region of hotter gas above the photosphere called the "chromosphere."

During an eclipse, the human eye is an excellent detector for a faint outer extension of the Sun's atmosphere known as the "corona." Although some information about the corona can be learned from observations of certain lines in the visible spectrum, detailed knowledge about the physics of the corona had to wait for the application of the techniques of radio astronomy and X-ray astronomy. When the Sun is "viewed" in radio at microwave frequencies, the emission is patchy, with certain regions much brighter than other. When the Sun is "viewed" in X-rays, different images emerge depending on the energy of the photon. At low energies, there is extended diffuse emission covering a large fraction of the solar area, and also there are some dark areas which seem to be empty (coronal "holes"). When viewed in higher energy photons, the Sun is dominated by bright more or less localized emission regions: these

are referred to generically as "active regions." Within each active region, it is sometimes possible to trace out bright discrete structures ("loops"). The brightest of all features in the corona, and also in the chromosphere, are short-lived brightenings ("flares") which occur unpredictably from time to time in certain active regions.

Observations of the photosphere, the chromosphere, and the corona show beyond any doubt that the surface of the Sun and its outer atmosphere are subject to variability of different kinds. Sometimes dark spots appear on the surface for a few days or weeks. The total number of spots which are visible on the surface at any given time waxes and wanes every 11 years or so. At times, the Sun ejects massive quantities of material from the corona, and there are also, at times, sudden outbursts of high energy photons. Before the modern era of solar observations, these occasional ejections and outbursts would have affected life on Earth no more than providing a more or less brilliant show of "northern lights," when the northern sky would light up at night. But in our day and age, the eruptions and outbursts from the Sun can have more serious effects, including damage to expensive communication satellites or the equipment that electric power companies rely on to distribute electricity to their customers.

A good opportunity to examine graphical illustrations, examples, and videos of the broad variety of highly dynamic phenomena as they occur on the Sun can be found on the worldwide web at a site (maintained by NASA, the National Aeronautics and Space Administration) which is devoted to a particular spacecraft: the Solar and Heliospheric Observatory. This spacecraft, known by its abbreviated title, SOHO, is located in orbit around the Sun at a special point in space. The point lies between Earth and Sun at the location where the gravitational pull of the Earth is comparable to the gravitational pull of the Sun. SOHO has a continuous and uninterrupted view of the Sun since there is no day-night cycle and/or clouds to interfere with observations. The website is http://sohowww.nascom.nasa.gov/, an excellent resource for "viewing" the Sun in a broad variety of wavelengths.

Among the data sets that can be examined at the website, there is one in particular that holds the key to understanding the fundamental nature and origin of the most dynamic solar phenomena. This is an instrument known as Michelson Doppler imager (MDI) which searches for magnetic fields on the surface of the Sun. The data show clearly that magnetic fields are present in the Sun, although the total amount of magnetic flux varies with time. At certain times, there are many magnetic areas at various locations on the disk, while at other times, the Sun appears almost devoid of fields. The fields wax and wane in strength as the years go by: the waxing and waning is almost periodic, with a period of about 11 years. When we compare the magnetic field data that MDI measures with images of the photosphere, chromosphere, and corona, it becomes apparent that the phenomena of sunspots, the patchy radio corona, coronal loops, flares and mass ejections, are all related in different ways to regions where stronger than average magnetic fields are present.

The question that is important to address is the following. Do the solar eruptions and outbursts that provide such a spectacular component to the SOHO observational database represent perturbations that we should be worried about in the context of life on Earth? Or do they constitute relatively minor disturbances against the backdrop of a much larger, and much steadier, output of energy which the Sun generates continuously as the days, years, and eons go by? In order to address these questions, we need on the one hand to determine the properties of the Sun as a whole, and we need on the other hand to determine the properties of the magnetically driven phenomena in the atmosphere. Specifically we must determine how hot and how dense the material is in the deep interior, and how much inertia this can provide to offset the dynamic phenomena which attract our attention from time to time in the surface layers.

The laws of physics indicate that the power output of the Sun depends on how the physical parameters temperature T, pressure p, and density ρ vary as a function of radial location r between the center of the Sun ($r = 0$) and a region that we will refer to as "the visible surface" ($r = R_\odot$). Here, and throughout this book, the subscript $\odot$ denotes a parameter of the Sun as a whole. An important goal in our study of the Sun is to use various laws of physics to determine the radial profiles $T(r), p(r)$, and $\rho(r)$ between $r = 0$ and $r = R_\odot$.

The starting point for these profiles is provided by observations of certain parameters at the "visible surface" of the Sun. Photons from those visible layers reach us on Earth and carry information on local conditions at $r = R_\odot$. Once the global parameters of the Sun are determined, they serve as boundary conditions to help us get started on our calculation of $T(r)$, $p(r)$, and $\rho(r)$ *beneath* the surface, i.e., at $r < R_\odot$. By applying a variety of physical laws, and also by a judicious choice of computational techniques, our goal is to calculate $T(r)$, $p(r)$, and $\rho(r)$ for all values of r down to $r = 0$. This will tell us about the *internal* structure of the Sun.

An important aspect of modern solar physics is that once we have calculated the internal structure, we can check our calculations by examining the properties of the vast array of oscillations which the Sun supports. This area of research, known as helioseismology, has opened up windows on the solar interior which were totally absent prior to the 1970s. Not only can the oscillations help us to check the structural calculations, but they can also help us determine how the Sun rotates at depths far below the surface. By 1990, the oscillation data had become good enough that it became possible to detect variations in the oscillation frequencies as the Sun varies during the 11-year sunspot cycle. Even more recently, in 2007, claims have been reported for a type of oscillation (g-modes) which propagate mainly in the Sun's deep interior.

In view of the great advances in solar physics which the oscillations have made possible, it is important that, even in a first course in solar physics, attention be paid to understanding how to calculate the basic properties of solar oscillations.

However, as well as helping us get started on learning about the internal structure, conditions at the visible surface also serve as boundary condition for another problem: how do the physical parameters vary as a function of radial location *above the surface?* A rich variety of physical phenomena occur in these locations, which may be referred to as the "outer atmosphere of the Sun." Many of these physical phenomena, especially those which exhibit pronounced variability with time, are covered by the umbrella term "solar activity." This term includes sunspots, flares, and coronal mass ejections. Physical parameters of the gas which lies between $r = R_\odot$ and the Earth's orbit can also be determined by applying the laws of physics to the increasingly rarefied environment which exists at greater and greater radial distances from the center of the Sun.

In summary, our approach to studying the physics of the Sun consists of starting at a certain location ($r = R_\odot$) where the local physical parameters can be reliably measured, and then proceeding in two distinct directions: first inward, and subsequently outward.

As regards the long-term stability of the Sun, two aspects of the physics are key. The first has to do with the pressure p_c at the center of the Sun: if p_c can support the weight of the overlying material (i.e., the weight of the entire Sun), then a condition known as hydrostatic equilibrium is ensured. In this case, the Sun will be in a structural condition where, in a global sense, all the mechanical forces are in balance.

The second aspect is that energy must be generated at such a rate that the power output remains steady on time-scales of several billion years. The only source of energy that will satisfy this is nuclear fusion. This requires that the central temperature T_c be high enough that nuclear reactions can occur at a suitably rapid rate to keep the Earth warm.

Our goal in this book is to determine enough information about the physical conditions inside the Sun, and in its extended atmosphere, so that we may appreciate, from a global perspective, the amazing entity that enables life to survive on Earth for eons of time.

In this study, a particular emphasis is placed on numerical modeling. In five of the chapters (Chapters 5, 7, 9, 10, and 14), the reader is given step-by-step instructions for calculating, in a simplified manner, the numerical values of various physical quantities as a function of radial distance inside the Sun. I believe that a student can gain a lot of insight into the conditions in the Sun by watching, step by step, how the pressure, or the temperature, or the amplitude of an oscillation, vary as one moves from one radial position inside the Sun to another. When the student, in a later more advanced course, eventually encounters the complete equations of stellar structure, including detailed expressions for the equation of state, the opacity, and the energy generation rate, the codes can become so complicated that it is not easy to understand why the solutions behave the way they do. In the present course, I would like the student to obtain a firm grasp of how the pressure (in "real" units, i.e., dyn cm^{-2}) varies as a function of radial location (also in "real" units, i.e., cm)

from the center of the Sun, to the photosphere, to the chromosphere, to the corona, and eventually into the distant wind. Likewise, I want the student to obtain a firm grasp on the radial profiles of density and temperature. A feel for the actual physical length-scales and pressures can help a student to appreciate the immensity of the Sun. And as the student will learn in Chapter 11, this immensity is fundamentally what gives rise to the nuclear reactions which make life on Earth possible.

Dermott J. Mullan
University of Delaware

Author

Dermott J. Mullan has been a professor in the Bartol Research Center, Department of Physics and Astronomy, University of Delaware, since 1982. His research interests are in stellar structure, especially as the structure is altered by the presence of magnetic fields. His research, supported by several dozen grants from NASA, is based on observations with a number of satellites which detect photons from the Sun and stars over a range of wavelengths, from infrared to X-rays. His research results have been published in more than 200 papers in the refereed literature. At the University of Delaware, he has taught graduate-level courses in introduction to astrophysics, stellar structure, plasma astrophysics and solar physics, and undergraduate courses in concepts of the universe and bioastronomy. He is currently director of the NASA Space Grant Program, as well as director of the NASA EPSCoR Program in Delaware.

Chapter 1

Global Parameters of the Sun

In order to determine the physical processes which occur in the Sun, we need to know certain properties of the Sun, including mass, radius, and other quantities. In this chapter, we summarize the relevant information, with emphasis on describing *how* the information is obtained, and *how precise* the current measurements actually are.

When it comes to astrophysical measurements, the quantity which can be measured with greatest accuracy is TIME. As a result, we begin our discussion of the determination of solar parameters by referring to measurements of certain intervals of time.

1.1 Orbital Motion of the Earth

The single most important property which determines the evolutionary behavior of a star is its mass; it is the mass which determines whether a star will eventually ends its life quietly or explosively.

In order to determine the mass of the Sun, $M_\odot$, we begin by referring to the time that is required for the Earth to orbit the Sun. Determination of this time-scale is achieved by observing the interval of time required for the Sun to return to a given location relative to the "fixed" stars as seen by an observer on Earth. Actually the "fixed stars" used for this determination are a class of galaxies known as quasi-stellar radio sources ("quasars"). Relative to the quasar frame, this defines the unit of one sidereal year: $P = 365.25636$ days = 31,558,150 seconds. For this, and other precise estimates of various parameters of interest to solar system dynamics, refer to the website maintained by NASA's Jet Propulsion Laboratory: http://ssd.jpl.nasa.gov/?constants.

Now that we know the orbital period of the Earth, we now turn to the equation of motion of the Earth in its orbit. This equation can be written in terms of position vectors of Sun and Earth. Relative to a zero point which can be arbitrarily chosen, the position vector of the Sun is $r(S)$ and the position vector of the Earth is $r(E)$. The position vector of Earth relative to the Sun is $r = r(E) - r(S)$, and the unit vector associated with the relative position vector, $\hat{r}$, is directed from the Sun toward the Earth.

The forces which act on the Sun (with mass $M_\odot$) and on the Earth (with mass $m_\oplus$) are given by Newton's law of gravitation. The gravitational force causes the Earth to accelerate according to the equation

$$m_\oplus \frac{d^2 r(E)}{dt^2} = -\frac{GM_\odot m_\oplus}{r^2}\hat{r} \tag{1.1}$$

where G is Newton's gravitational constant, and the negative sign indicates that the force is toward the Sun, i.e., in the negative $\hat{r}$ direction.

The gravitational force causes the Sun to accelerate according to the equation

$$M_\odot \frac{d^2 r(S)}{dt^2} = +\frac{GM_\odot m_\oplus}{r^2}\hat{r} \tag{1.2}$$

where the positive sign indicates that the force is toward the Earth, i.e., in the positive $\hat{r}$ direction.

In terms of the relative position vector r, the above equations can be combined to yield

$$\frac{d^2 r}{dt^2} = -\frac{G(M_\odot + m_\oplus)}{r^2}\hat{r} \tag{1.3}$$

The solution of this equation is an ellipse with the center of mass at one focus. With a semimajor axis D for the ellipse, the period P of orbital motion is given by

$$P^2 = \frac{4\pi^2 D^3}{G(M_\odot + m_\oplus)} \tag{1.4}$$

This leads to an expression for the mass of the Sun:

$$\frac{GM_\odot}{D^3} = \frac{4\pi^2}{P^2[1 + m_\oplus/M_\odot]} \tag{1.5}$$

The ratio of $m_\oplus$ to $M_\odot$ is very small (we will evaluate it shortly). If we were to neglect the ratio $m_\oplus/M_\odot$ compared to unity, then we would get a fairly precise first approximation to $GM_\odot/D^3$. According to this approximation, for each planet in the solar system, the square of the period P^2 is proportional to the cube of the mean distance D^3. This property was first identified empirically by Kepler as his third law of planetary motion.

However, with or without the correction for the Earth's mass, we cannot determine the value of $M_\odot$ unless we first determine the value of D (= 1 astronomical unit).

1.2 Astronomical Unit (AU)

Once the orbital periods of the various planets are known, the application of Kepler's third law provides a scale model of the solar system. The scale

model provides knowledge, at any given instant of time, of the distances of planets and other solar system objects in terms of AU, the semimajor axis of the Earth's orbit. As a result, at any given instant of time, we know how far away any solar system object is from Earth in terms of AU. In favorable conditions, radar reflection can be used to determine the linear distance to the object at that instant. This has the advantage that a distance measurement is performed in terms of a measurement of a time interval, which can be done with high precision.

Reliable radar reflection measurements were first made around 1960 using the planet Venus. It was not just the intensity of the signal which was measured, but also the Doppler shift. This means that the orbital motion of Venus can be allowed for in the course of an extended period of observations. When Venus is closest to Earth, the round-trip time for radar reflections during the experiment is close to 5 minutes, and this interval can be measured with a precision of many significant figures.

The International Astronomical Union currently defines the AU as follows:

$$1 \text{ AU} \equiv D = 149{,}597{,}870.691 \text{ km}$$

For future reference, we note that at a distance equal to D, the linear diameter of any object which has an angular diameter of 1 arc sec is 728.8 km.

By inserting the value of D in Equation 1.5, a first approximation to the quantity $GM_\odot$ can be determined. Moreover, in this approximation, Kepler's 3$^{\text{rd}}$ law becomes an equality $P^2 = D^3$ provided that P is expressed in years and D is expressed in AU.

To obtain a more precise estimate of $M_\odot$, we need to evaluate the ratio $m_\oplus/M_\odot$. We do that by comparing the motions of two objects, one in orbit around the Sun, the other in orbit around the Earth. For both objects, we need to determine two quantities: a period and a distance.

For the object (the Earth) that is in orbit around the Sun, with period $P(S)$ and semimajor axis $D(S)$, we know that

$$P(S)^2 \sim \frac{D(S)^3}{M_\odot + m_\oplus} \tag{1.6}$$

For the object (an artificial satellite) that is in orbit around the Earth, with period $P(E)$ and semimajor axis $D(E)$, we have that

$$P(E)^2 \sim \frac{D(E)^3}{m_\oplus} \tag{1.7}$$

where we have made the reasonable assumption that the mass of the artificial satellite is entirely negligible (by 20 orders of magnitude or more) compared to the mass of the Earth.

Combining the above equations, we have that

$$\frac{M_\odot}{m_\oplus} + 1 = \left(\frac{D(S)}{D(E)}\right)^3 \left(\frac{P(E)}{P(S)}\right)^2 \tag{1.8}$$

There are a large number or choices which we can make for an artificial satellite in orbit around the Earth. Any one will suit our purpose. To obtain information about any particular artificial satellite, it is convenient to examine the website http://www.heavens-above.com/, where information is available about many satellites in orbit around the Earth. There, by way of example, we find that a satellite called RADCAT has the following orbital details: the perigee lies at an altitude of 491 km above the Earth's surface, the apogee at 495 km. This suggests that the mean altitude of RADCAT above the Earth's surface is $h = 493$ km. Note that only three significant digits are provided for these distances: this will limit the precision of our evaluation of $m_\oplus/M_\odot$.

With an mean altitude of $h = 493$ km, the semimajor axis of the orbit is $D(E) = R_\oplus + h$, where $R_\oplus$ is the radius of the Earth. The equatorial radius of the earth has been accurately measured, by the International Union of Geodesy and Geophysics, to be $R_\oplus = 6378.137$ km. This leads to $D(E) = 6871.137$ km for RADCAT. Comparing this with the value of $D(S) = 1$ AU, we see that $D(S)/D(E) = 21{,}771.924$ for RADCAT. Thus, the first factor on the right-hand side of Equation 1.8 is 1.0320255×10^{13}.

Turning now to the period, information on heavens-above.com indicates that the RADCAT satellite orbits 15.24243084 times per day, corresponding to a period $P(E) = 5668.387$ sec. Compared to the value of $P(S)$ (= 1 sidereal year), we find that the second factor on the right-hand side of Equation 1.8 has the value 3.2262344×10^{-8}.

Combining the terms in Equation 1.8, we find that $m_\oplus/M_\odot = 1/332{,}955$. This is the mass ratio which we obtain when we use the orbital data for a single satellite (RADCAT), for which we know the altitude to only three significant digits. When multiple satellites are used, the currently accepted value of $m_\oplus/M_\odot$ is found to be 1/332,946. Thus, our use of RADCAT data alone leads to an error in the mass ratio of about 1 part in 30,000. Our calculations would have led to the currently accepted value of $m_\oplus/M_\odot$ if we were to use a value of 493.08 km (rather than 493 km) for the mean altitude of RADCAT. It should be kept in mind that even small errors in a measurement may translate to significant uncertainties in some of the quantities which are of interest to us when we study the Sun.

1.3 $GM_\odot$ and the Mass of the Sun

We now have enough information to evaluate the product of the gravitational constant and the mass of the Sun:

$$GM_\odot = 1.327124 \times 10^{26} \text{ cm}^3 \text{ sec}^{-2} \tag{1.9}$$

The precision with which the product $GM_\odot$ is known has increased over the course of the space age, as more and more spacecraft have traveled throughout

the solar system, always subject to a sunward acceleration which is proportional to the above product. Currently, the numerical value of $GM_\odot$ is known to 11–12 significant digits, but we do not need all those digits here, because G is not that well known.

To extract a value for the mass of the Sun, we need to divide the above product by G. The numerical value of G is among the most poorly measured constants of nature, known only to 1 part in 10^4: $G = 6.67428(\pm 0.00067) \times 10^{-8}$ dyn cm^2 gm^{-2} (see the list of physical constants maintained by the National Institute of Standards and Technology: http://physics.nist.gov/cuu/Constants/).

Using this, we obtain the following estimate of the mass of the Sun, reliable to 1 part in 10^4:

$$M_\odot = 1.9884 \times 10^{33} \text{ gm} \tag{1.10}$$

1.4 Power Output of the Sun: The Solar Luminosity

Spacecraft which are equipped with radiometers can measure the flux of radiant energy coming from the Sun: this flux, known as the solar irradiance, I_s, is reported on the SOHO website to have a mean value of about 1366 W m^{-2}, i.e., 1.366×10^6 ergs cm^{-2} sec^{-1}. The magnitude of I_s is observed to vary slightly in the course of a sunspot cycle (see Figure 1.1): the variations are at the 0.1% level, i.e., about one part in 1000.

Given the distance from Earth to Sun (D), the mean I_s transforms to an output power from the Sun of $L_\odot = 4\pi D^2 I_s$, i.e.,

$$L_\odot = 3.8416 \times 10^{33} \text{ ergs sec}^{-1} \tag{1.11}$$

This power output from the Sun (also referred to as the "solar luminosity") varies by roughly ± 1 part in 1000 on 10–12 year cycles. During each cycle, the surface of the Sun is occupied by a greater or smaller number of dark patches called "sunspots." We will discuss sunspots in Chapter 16.

For future reference, comparing Equations 1.11 and 1.10, we note that the ratio of $L_\odot/M_\odot$ has a numerical value close to 2 ergs gm^{-1} sec^{-1}. We will find it useful to use this ratio when we calculate the internal structure of the radiative interior of the Sun (Chapter 8).

Also for future reference, we note that the power output from the Sun relies on the conversion of (nuclear) mass into energy in the deep inner core of the Sun. Using the conversion formula $E = mc^2$, we note that the value of the Sun's power output requires the conversion of mass to energy at a rate $(dM/dt)_{\rm nucl} = 4.274 \times 10^{12}$ gm sec^{-1}. In the course of the Sun's lifetime, which is estimated to be about 4.6 Gy, the mass of the Sun has been reduced by nuclear processing by a few parts in 10^4.

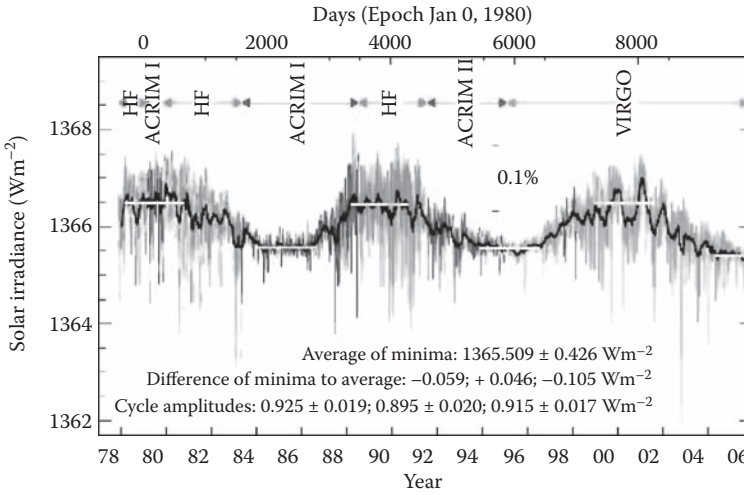


FIGURE 1.1: The solar irradiance, normalized to a solar distance of 1 AU, measured over almost 30 years. (Courtesy of SOHO/VIRGO consortium. SOHO is a project of international cooperation between ESA and NASA.)

1.5 Radius of the Sun: $R_\odot$

Now that the mean distance to the Sun is known, it would seem to be a simple matter to obtain the linear radius (diameter) of the Sun by measuring the angular radius (diameter). However, measuring the angular radius of the Sun from the ground is difficult to do with precision on account of the phenomenon of "seeing." Turbulent eddies in the Earth's atmosphere make the image of the Sun unsteady, smearing out the edge of the solar disc on angular scales of order 1 arc sec. This leads to an uncertainty in the solar radius of order 700 km derived from ground-based measurements.

Empirically, the existence of two distinct classes of eclipses of the Sun (total and annular) indicates that the Sun has an angular diameter which is comparable to the Moon's: the latter is close to 32 arc min, i.e., 1920 arc sec. So the Sun's angular radius is of order 960 arc sec. But from the ground, this cannot be measured to better than about 1 arc sec, i.e., to one part in 1000.

In order to obtain more reliable measurements, observations from space are required. The SOHO spacecraft, launched in 1996, has made the most careful measurements in this regard. In order to provide the best possible calibration of the CCD pixels in the SOHO/MDI detector, a transit of Mercury was observed on May 7 2003. In the course of a 5–6 hour period, Mercury moved along a track which was known from planetary dynamics to a precision of ± 0.025 arc sec. Kuhn et al. (2004) report that the angular radius of the Sun (when observed at $D = 1$ AU) is 959.28 ± 0.15 arc sec.

When this is converted to linear measure, it corresponds to a (linear) solar radius of $R_\odot = 6.9574(\pm 0.0011) \times 10^5$ km.

An alternate method of determining the (linear) radius of Sun is provided by helioseismic data. In Chapters 13 and 14, we will study oscillatory modes in the Sun, and we shall see that the periods of various modes can be measured to better than 1 part in 30,000. By referring to models of the solar structure, the frequencies of identifiable modes allow the solar radius to be determined with a precision which is comparable to the precision of the frequency determinations (Schou et al., 1997). The best determination of solar radius from helioseismic data is

$$R_\odot = 6.9568 \pm 0.0003 \times 10^5 \text{ km} \tag{1.12}$$

This result overlaps with the above estimate of angular diameter from SOHO, but is about three times more precise. The improvement in precision can be attributed to the fact that the analysis depends on measurements of time (frequency) rather than angle.

For future reference, when we come to discuss the solar wind (Chapter 18), it will be helpful to know how far the Earth is from the Sun in units of the solar radius. Combining D with $R_\odot$, we see that 1 AU is equivalent to 215.04 $R_\odot$.

1.6 Surface Gravity of the Sun

Now that we know the mass and radius of the Sun, we can calculate the acceleration due to gravity g_s at the solar surface.

$$g_s = \frac{GM_\odot}{R_\odot^2} = \frac{1.327124 \times 10^{26}}{(6.9568 \times 10^{10})^2} = 27{,}421.6 \text{ cm sec}^{-2} \tag{1.13}$$

For future reference, we note that a convenient way to remember this value is to recall the logarithmic value: $\log g_s = 4.44$.

1.7 Escape Speed from the Solar Surface

The escape speed from the surface of the Sun is given by

$$V_{\text{esc}} = \sqrt{\frac{2GM_\odot}{R_\odot}} = 617.7 \text{ km sec}^{-1} \tag{1.14}$$

This escape speed is a measure of the depth of the gravitational potential well due to the mass of the material in the entire Sun. It is a measure of how

strongly the Sun's weight crushes the gas in the core of the Sun. It is a law of physics that, if the Sun is to remain in hydrostatic equilibrium, the crushing effects of the weight of the overlying material on the core have to be balanced by outward-directed pressure.

Now, pressure is determined by the momentum flux of the individual gas particles. As a result, the thermal pressure in the core is related to the mean square velocity of the thermal particles there (e.g., Sears, 1959). Thermal particles, each with mass m, and in a medium with temperature T, have a root-mean-square (rms) velocity $V_{\rm rms} = \sqrt{(3kT/m)}$ where $k = 1.3806504 \times 10^{-16}$ ergs deg^{-1} is Boltzmann's constant.

The existence of the two velocities, $V_{\rm rms}$ and $V_{\rm esc}$, which are both characteristic of the Sun, suggests that in a model of the Sun which is in hydrostatic (i.e., mechanical) equilibrium, $V_{\rm rms}$ and $V_{\rm esc}$ should have comparable magnitudes. We shall check on this expectation when we complete our calculation of a mechanical model of the Sun (Chapter 9).

For future reference, we note that for a gas consisting of hydrogen atoms, $1/m = 1/m_{\rm H}$ (where $m_{\rm H}$ is the mass of a hydrogen atom), and this equals Avogadro's number N_a, which is the number of molecules in one mole. The combination kN_a is referred to as the gas constant $R_g = 8.314472 \times 10^7$ ergs deg^{-1} mole^{-1}. For a gas consisting of particles with atomic mass μ, the rms velocity $V_{\rm rms} = \sqrt{(3R_gT/\mu)}$.

1.8 Effective Temperature of the Sun

Now that we know the output power of the Sun as well as the radius, we can calculate the effective temperature. This is the temperature of the equivalent black-body which would radiate a flux equal to that emitted by the Sun:

$$L_\odot = 4\pi R_\odot^2 \sigma_B T_{\rm eff}^4$$

where the Stefan–Boltzman coefficient $\sigma_B = 5.67040 \times 10^{-5}$ ergs cm^{-2} sec^{-1} deg^{-4}. The surface flux of energy at the Sun, $F_\odot = L_\odot/4\pi R_\odot^2$ has the numerical value 6.3155×10^{10} ergs cm^{-2} sec^{-1}. Using this, we find that the effective temperature of the Sun is

$$T_{\rm eff} = 5777 \text{ K} \tag{1.15}$$

1.9 Shape of the Sun

To the unaided eye, the Sun appears to be essentially circular in shape. But careful measurements reveal a slight departure from circularity. The difference

between the solar radius at the equator and the solar radius at the pole is expressed in terms of the oblateness $\varepsilon = (R_{\rm eq} - R_{\rm pole})/R_{\rm eq}$.

First attempts to measure ε were made using ground-based observations. But the effects of seeing make this very difficult to do. Early results in the 1960s claimed that the Sun was oblate with $\varepsilon = 4.2 \times 10^{-5}$. Such an oblateness would correspond to a linear difference of 30 km between the equatorial radius and the polar radius. It would mean that the difference in angular radii would be of order 0.04 arc sec. This is much smaller than the effects of seeing (typical amplitude $\approx$ 1 arc sec), and so it is not surprising that it is difficult to make the measurements reliably from the ground.

A balloon-borne instrument, the Solar Disk Sextant (SDS), flown in 1992 and 1994, made observations at altitudes which were above most of the atmosphere. The reported oblateness was $\varepsilon = 9 \pm 1 \times 10^{-6}$ (Lydon and Sofia, 1996), considerably smaller than had been suggested by the earlier ground-based data.

Measurements from space were made by SOHO: the spacecraft was rolled through 360 degrees in small angular increments, each 0.7 degrees in extent, corresponding to $360/0.7 = 514$ individual "pie slices" of data around the entire circumference. Each "pie slice" was fitted with a radial profile: taking a numerical radial derivative of each profile, and squaring the derivative, the location of the peak of squared derivative was defined to be the location of the limb. With more than 500 samples, the rules of statistics suggest that the noise in individual "pie slices" can be reduced from 0.15 arc sec (see Section 1.5) to $0.15/\sqrt{514} = 0.007$ arc sec = 5 km. By making multiple observations over several months, the authors claimed that they could achieve a precision of 0.5 km in the solar radius (Kuhn et al., 1998).

Observations obtained in 1996–1997 indicated a solar oblateness of

$$\varepsilon = (7.77 \pm 0.66) \times 10^{-6} \tag{1.16}$$

This oblateness overlaps with the 1992/1994 results from SDS, although with somewhat improved error bars.

The existence of a finite oblateness in the solar figure is expected because the material in the Sun is subject to forces arising from rotation. If rotation were absent, the Sun's figure would settle into an equi-potential surface, for which the potential would be spherically symmetric: $\varphi = -GM_\odot/r$. With such a potential, the surface acceleration due to gravity $g = -d\varphi/dr$ is also symmetric. In the presence of rotation, however, the (inward) force due to gravity is counter-acted to some extent by the (outward) centrifugal force. With a solar angular velocity Ω, the net gravitational acceleration at colatitude θ becomes

$$g(\text{rot}) = g - r\Omega^2 \sin^2\theta \tag{1.17}$$

corresponding to a potential

$$\varphi = -\frac{GM_\odot}{r} - 0.5 r^2 \Omega^2 \sin^2\theta \tag{1.18}$$

This leads to an equi-potential surface which, in the presence of an equatorial rotational velocity $V(\mathrm{eq}) = r\Omega$, has an oblateness of $\varepsilon = 0.5V(\mathrm{eq})^2/gr$.

What is the rotational velocity of the Sun? We can answer this question as regards the *surface* of the Sun by means of direct observations. Rotational periods of material on the surface of the Sun can be measured from the Doppler shifts of spectral lines at east and west limbs. An important finding is that the rotational period is *not* constant at all latitudes. Instead, the period is found to be *shortest* at the equator, and the period becomes *longer* as we observe closer to the poles. This behavior is called "latitudinal differential rotation." An empirical fit to the rotation can be achieved by the following expression for angular velocity as a function of latitude λ:

$$\Omega(\lambda) = \Omega(0)\left[1 - b\sin^2\lambda - c\sin^4\lambda\right] \tag{1.19}$$

In a study involving Doppler shift data from many points on the surface, obtained in the course of 14 years, Howard et al. (1983) reported average values for the parameters in this fit: $\Omega(0) = 2.867 \times 10^{-6}$ rad sec^{-1}, $b = 0.121$, and $c = 0.166$. At the equator, the measured angular velocity $\Omega(0)$ corresponds to a rotational period of $P(\mathrm{rot,eq}) = 2\pi/\Omega(0) = 25.4$ days. The equatorial rotational velocity $V(\mathrm{eq}) = \Omega(0)R_\odot$ has a numerical value of 1.99 km sec^{-1}. At latitudes of 60°, the rotational period $P(\mathrm{rot},60) = 31.3$ days. At the north and south poles, Equation 1.19 indicates that $\Omega(90) = 0.713\Omega(0) = 2.044 \times 10^{-6}$ rad sec^{-1}, corresponding to a polar rotational period $P(\mathrm{rot,poles}) = 2\pi/\Omega(90) = 35.6$ days. Remarkably, the gas in the polar regions of the Sun rotates almost 30% more slowly than the gas near the equator. If we needed any reminder that the Sun is *not* a solid body (but is composed entirely of gas), differential rotation would provide the evidence.

If the entire Sun were to rotate at a period of 25.4 days, then the oblateness due to rotation would have the numerical value $\varepsilon(\mathrm{rot}) = 10.4 \times 10^{-6}$. This is several standard deviations *larger* than the oblateness reported by SOHO. It seems that the entire Sun cannot be rotating with a period that is as short as 25.4 days: some regions must be rotating more slowly than that.

The observed oblateness values ($\varepsilon = (8\text{–}9) \times 10^{-6}$) would be consistent with rotational effects if the entire Sun were to rotate with a period which is *longer* than 25.4 days by a factor of $\sqrt{(10.4/(8\text{–}9))} = 1.07 - 1.14$. The observed oblateness *could* be due entirely to rotation if the Sun were to rotate as a solid body with angular velocity $\Omega(\mathrm{obl}) = (2.5\text{–}2.7) \times 10^{-6}$ radians sec^{-1}.

Converting the angular velocities to (temporal) frequency, $\nu = \Omega/2\pi$, we note that the equatorial rotation $\Omega(0)$ corresponds to $\nu(0) = 456$ nanoHertz (nHz), while $\Omega(\mathrm{obl})$ corresponds to $\nu(\mathrm{obl}) = 398\text{–}430$ nHz.

As it turns out, the analysis of helioseismological data has revealed that the inner regions of the Sun do *not* rotate as a solid body. Different regions in the Sun rotate with different periods, depending on latitude and radial location. The fastest rotation, at equatorial latitudes, and at radial locations close to the surface, is about $\nu = 470$ nHz, while the slowest (at polar latitudes, and

also close to the surface) is about $\nu = 320$ nHz. Thus, the material inside the Sun spans a rather broad range of rotational frequencies: 320–470 nHz.

The rotational frequencies which are derived when the entire observed oblateness is attributed to rotational effects *viz.* $\nu(\text{obl}) = 398$–430 nHz, are entirely consistent with the range of rotational frequencies which exist inside the Sun. It appears therefore that most (or all) of the observed oblateness of the Sun can be ascribed without serious contradiction to rotational effects.

1.10 Critical Frequency for Solar Oscillations

Now that we know the radius and mass of the Sun, there is a critical frequency which can be constructed from $R_\odot$, $M_\odot$, and G which will be relevant when we come to discuss the various modes of oscillations inside the Sun. By analogy with a pendulum, for which the period is given by $P_g = 2\pi\sqrt{(d/g)}$ if the length of the pendulum is d and the local acceleration due to gravity is g, a critical period in the gravity field of the Sun in a global sense can be written down by considering a pendulum with a length that is equal to the natural length of the system: the solar radius. This leads to $P_g = 2\pi\sqrt{(R_\odot/g_s)}$. This can be written as

$$P_g = 2\pi\sqrt{\frac{R_\odot^3}{GM_\odot}} \tag{1.20}$$

Substituting $GM_\odot = 1.327124 \times 10^{26}$ c.g.s. and $R_\odot = 6.9568 \times 10^{10}$ cm, we find $P_g = 10{,}008$ sec. The associated frequency $\nu_g = 1/P_g$ has the numerical value 99.92 microHertz (μHz). Thus, a fundamental frequency ν_g which is very close to 100 μHz is expected to provide a significant marker among the oscillation frequencies of the modes in which the Sun oscillates on a global scale.

1.11 Mean Density of the Sun

Another quantity which can be calculated once the mass and radius of the Sun are known is the mean density:

$$\bar{\rho} = \frac{M_\odot}{(4/3)\pi R_\odot^3} \tag{1.21}$$

Inserting the values of $M_\odot$ and $R_\odot$ from Equations 1.10 and 1.12, we find $\bar{\rho} = 1.410$ gm cm^{-3}. That is, the mean density of the (gaseous) Sun is somewhat greater than the mean density of (liquid) water. Once we calculate a model for the interior of the Sun (Chapter 9), it will be a matter of interest

to compare the density at the *center* of the Sun to the *mean* density. We shall find that the central density in the Sun is much larger than the density of liquid water: the central density is actually about ten times larger than that of solid lead.

Despite these large densities, the material of which the Sun is composed does not behave as a liquid or a solid: instead, we shall find that it obeys the laws which govern the behavior of a gas (Chapter 9).

We note that the critical period in Equation 1.20 scales as $1/\sqrt{G\bar{\rho}}$.

Now that we have information on the relevant physical parameters on a global scale, we can turn to a study of the internal structure of the Sun.

Exercises

1.1 Consult a table of orbital periods for the planets Mercury, Mars, Jupiter, and Neptune. Using Kepler's third law, determine the mean distance of each planet from the Sun in AU and in cm.

1.2 We shall see (Chapter 18) that the Sun's influence over the surrounding space extends out as far as about 100 AU. Determine the period (in years) that a planet would have if it were in an orbit with that value of D. Assuming the orbit is circular, determine the speed of the planet in its orbit.

1.3 The Sun is currently estimated to be some 4.6 Gy old. When the Sun was younger than 1 Gy, theory suggests that its luminosity was only 70% of what it is today. Assuming that the Sun's radius has not changed, calculate the effective temperature of the young Sun.

1.4 Using information provided above, determine the mean density of the Earth. Given the scaling $P_g \sim 1/\sqrt{\bar{\rho}}$, calculate P_g for the Earth. How does P_g compare with the orbital period of the RADCAT satellite in its near-Earth orbit?

1.5 Stars belonging to a feature called "the main sequence" have radii R_* and masses M_* which scale roughly as $R_* = R_\odot (M_*/M_\odot)^{0.7}$. For stars with masses 0.1, 0.3, 1, 3, and 10 $M_\odot$ on the main sequence, calculate R_*, and evaluate the surface gravity (Equation 1.13) and the escape speed $V_{\rm esc}$ from the surface (Equation 1.14).

1.6 The masses and luminosities of main sequence stars can be approximated by $L_* \sim M_*^{3.8}$. Using the formula for luminosity in Section 1.8, and the $R_* - M_*$ formula in Exercise 5, show that $T_{\rm eff}$ for main sequence stars scales as $M_*^{0.6}$. Using this scaling along with Equation 1.15, calculate $T_{\rm eff}$ for main sequence stars with masses of 0.1, 0.3, 1, 3, and 10 $M_\odot$.

References

Howard, R. et al. 1983. "Solar rotation results at Mount Wilson," *Solar Phys.*, 83, 321.

Kuhn, J. R., Bush, R. I., Scherrer, P., and Schieck, X. 1998. "The sun's shape and brightness," *Nature*, 392, 155.

Kuhn, J. R., Bush, R. I., Emilio, M., and Scherrer, P. H. 2004. "On the Constancy of the Solar Diameter. II.," *Astrophys. J.*, 613, 1241.

Lydon, T. J. and Sofia, S. 1996. "A Measurement of the Shape of the Solar Disk: The Solar Quadrupole Moment, the Solar Octopole Moment, and the Advance of Perihelion of the Planet Mercury," *Phys Rev. Lett.*, 76, 177.

Schou, J., Kosovichev, A. G., Goode, P. R., and Dziembowski, W. A. 1997. "Determination of the Sun's Seismic Radius from the SOHO Michelson Doppler Imager," *Astrophys. J. Lett.*, 489, L197.

Sears, F. W. 1959. "The distribution of molecular velocities," *An Introduction to Thermodynamics.* Addison-Wesley, Reading, MA, pp. 223–255.

Chapter 2

Radiation Flow through the Solar Atmosphere

Now that we have knowledge of the global parameters of the Sun, we are in a position to turn to an interpretive study of the photons which are the principal means by which information comes to us from the Sun. If we can make certain measurements on the photons from the Sun, such as their distribution in wavelength, and the integrated flux of radiant energy, we may hope to extract information about the temperature and other physical quantities in the region from which the photons originated. The photons originate mainly in a region which can be considered roughly as "the (visible) surface of the Sun": a more precise definition of this region will emerge subsequently from a discussion of radiative transfer.

Our goal is to use the information carried by the solar photons to undertake a task of physical interpretation which will take us in two opposite directions away from the "surface": (i) into the deep inner regions of the Sun, and (ii) outward toward the rarefied material which lies above the visible surface.

We aim to use certain laws of physics to help us determine a "model of the Sun," i.e., to determine the radial profile of physical parameters such as temperature, density, and pressure.

2.1 Radiation Field in the Solar Atmosphere

The goal of radiative transfer in the solar atmosphere is to determine how radiation interacts with the medium as it passes through material with a particular set of physical properties. The interaction is mutual: on the one hand, the medium imprints certain properties on the radiation, and on the other hand, the material in the medium is affected (as far as its temperature and density are concerned) by the photons which stream outward from deep inside the star.

An important way to characterize the radiant energy is the intensity I_λ: this is the amount of radiant energy which flows through unit area per unit time per unit wavelength and per unit solid angle. The units of I_λ which are

appropriate for the visible spectrum from the Sun are ergs cm^{-2} sec^{-1} cm^{-1} steradian^{-1}.

An alternative approach to quantifying the radiant power is to specify the intensity *per unit frequency*: this is given the symbol I_ν in units of ergs cm^{-2} sec^{-1} Hz^{-1} steradian^{-1}. Conservation of energy requires that $I_\lambda d\lambda = I_\nu d\nu$. Since $\lambda\nu = c$ (where c is the speed of light), this means that $I_\lambda = I_\nu(c/\lambda^2)$.

The numerical value of I_λ (or I_ν) at any point inside a medium (whether it is inside the Sun or in a star, or inside an oven on Earth) depends on the local temperature: other things being equal, the higher the temperature, the larger the value of I_λ (or I_ν). The value of I_λ (or I_ν) also depends on the wavelength λ at which observations are made.

The simplest example of I_λ (or I_ν) which is useful for astrophysical studies refers to the radiant energy field which is in thermal equilibrium inside a closed cavity. This leads to the so-called black-body radiation. In thermal equilibrium, the radiation is in equilibrium with the walls, and equal numbers of radiant modes are being absorbed and emitted by the walls per unit time. Inside a cavity, the only radiant modes which are present with significant amplitudes have wavelengths such that an integral number of half-wavelengths fit into the cavity. The existence of a discrete number ($\sim 1/\lambda^3 \sim \nu^3/c^3$) of such wavelengths means that it is possible, using classical physics, to enumerate the numbers of radiant modes which are permitted to exist per unit volume within a certain range of wavelengths (or within a certain range of frequencies) inside the cavity. The number of such modes per unit frequency can be shown to be equal to $8\pi\nu^2/c^3$ cm^{-3} Hz^{-1}.

The close coupling of radiation and walls in these circumstances suggests that, since the thermal energy of a single particle corresponding to temperature T is kT, where k is Boltzmann's constant, it might also be appropriate to assign an energy of kT to each radiant mode. With such an assignment, the radiant energy density per unit frequency would be $E_\nu = 8\pi\nu^2 kT/c^3$ ergs cm^{-3} Hz^{-1}.

Evaluation of the energy density of the radiation field is the first step toward deriving an expression for the intensity: the latter is associated specifically with the *flow of energy* across an element of surface area and into unit solid angle in a particular direction. To transform from energy density to intensity, the energy density must be multiplied by the speed of propagation (c) and also by the factor 1/4: the latter includes a factor of 1/2 to allow for inward and outward propagation, and a factor of 1/2 for geometric averaging over spherical angles. This leads to

$$I_\nu = 2\pi\nu^2 kT/c^2 \text{ ergs cm}^{-2} \text{ sec}^{-1} \text{ Hz}^{-1} \text{ steradian}^{-1} \tag{2.1}$$

Equation 2.1 for the radiant intensity is referred to as the Rayleigh–Jeans law: it provides a good fit to the radiant flux which emerges from a black-body at long wavelengths. However, as we apply the formula to progressively shorter (ultraviolet) wavelengths, i.e., as $\nu \to \infty$, the above expression for I_ν diverges, a phenomenon known historically as the "ultraviolet catastrophe."

In order to avoid this catastrophe, Max Planck in 1900 suggested that, despite the arguments of classical physics, it is *not* correct to assign the same energy (namely, the mean thermal energy of a particle kT) to each and every mode of the radiation field. Instead, Planck postulated that for modes of a given frequency v, only certain discrete energies are allowed: $E(i) = 0,\ h\nu, 2h\nu, 3h\nu, \ldots$. In this situation, the total energy which is available from the thermal energy of the cavity is distributed among a large number of modes. In this distribution, modes with energy $E(i)$ are present in numbers which are proportional to their Boltzmann factor $\exp(-E(i)/kT)$: this takes into account the fact that very few photons in the cavity are expected to have energies which are greatly in excess of kT. Adding up the occupation numbers to determine the overall partition function, it is possible to calculate the mean energy per mode. This mean energy is found to be no longer $\langle E\rangle = kT$ (as in the classical case): instead, Planck found

$$\langle E\rangle = \frac{h\nu}{e^{h\nu/kT} - 1} \tag{2.2}$$

In the limiting case of low energy photons, $h\nu \ll kT$, Planck's formula has the desirable property that it reduces to the classical result: $\langle E\rangle = kT$. However, in the opposite limit, for photons with energies which are greatly in excess of kT, the mean energy per photon falls well below the classical value: in the limit $hv \gg kT$, the mean photon energy $\langle E\rangle$ tends toward zero.

Using the revised estimate of mean energy in the radiation modes, Planck found that the classical energy density $E_v = 8\pi(v^2/c^3)kT$ is replaced by $E_v = 8\pi(\nu^2/c^3)h\nu/[\exp(h\nu/kT) - 1]$. Multiplying this energy density by the factor $c/4$ (as in the classical treatment), the radiant intensity which is radiated *per unit frequency* from a surface at temperature T is found to be:

$$I_\nu = \frac{2\pi h\nu^3}{c^2}\frac{1}{e^{h\nu/kT} - 1} \text{ ergs cm}^{-2} \text{ sec}^{-1} \text{ Hz}^{-1} \text{ steradian}^{-1} \tag{2.3}$$

If we wish instead to express the radiant intensity in terms of intensity *per unit wavelength*, the corresponding result is

$$I_\lambda = \frac{2\pi hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda T} - 1} \text{ ergs cm}^{-2} \text{ sec}^{-1} \text{ cm}^{-1} \text{ steradian}^{-1} \tag{2.4}$$

These are the expressions (the "planck functions") for the intensity of radiation associated with a "black-body." Deep in the solar atmosphere, where local thermodynamic equilibrium holds, we shall find that the mean free path for photons is so short (typically a few km) that the photons within a "small" volume can be considered to zeroth order to be essentially contained in a cavity where the temperature changes only insignificantly across the cavity. In such conditions, where a unique temperature is not a bad fit to local conditions, the "black-body" expressions provide a useful approximation to the properties of the local radiation field.

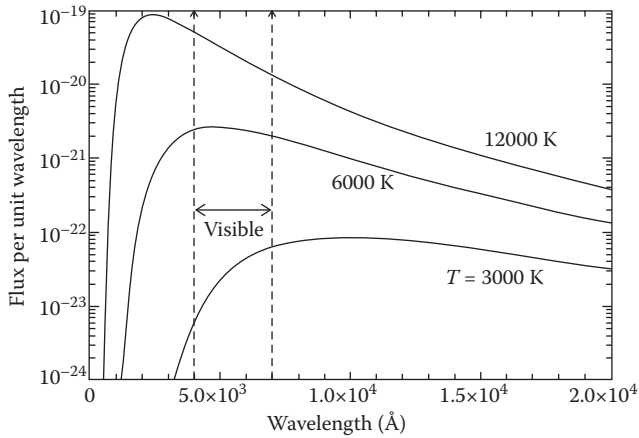


FIGURE 2.1: Black-body radiation: radiant energy flux per unit wavelength as a function of wavelength for black-bodies of different temperature. The wavelengths are expressed in units of Ångstrom (1 Å = 10^{-8} cm). The units of the ordinate are arbitrary.

Two important characteristic properties of black-body radiation are noteworthy (see Figure 2.1). First, the curve I_λ peaks at a certain wavelength $\lambda_{\max}$ which decreases as the temperature increases according to Wien's law: $\lambda_{\max}(\text{cm}) = 0.288/T$. In the case of the Sun, where the temperature in the vicinity of the "visible surface" is about 6000 K, $\lambda_{\max}$ occurs at about 5000 Å. Empirically, the solar spectrum, when plotted in the form of I_λ, is indeed found to exhibit a peak at wavelengths near 5000 Å, suggesting that the black-body provides a reasonable zeroth order fit to the radiation which emerges from the Sun. Second, integrating over all wavelengths, the total energy density of the photons in a cavity with temperature T is $u(T) = a_R T^4$ ergs cm^{-3}, where the radiation density constant a_R is equal to 7.5658×10^{-15} ergs cm^{-3} deg^{-4}. Converting from total energy density to a total flux of radiation in a certain direction, the integral over all frequencies and over all solid angles leads to a total flux of $\sigma_{\rm B} T^4$ where $\sigma_{\rm B} = a_R c/4$ is the Stefan–Boltzmann constant. The numerical value of $\sigma_{\rm B}$ is $5.67040(\pm 0.00004) \times 10^{-5}$ ergs cm^{-2} sec^{-1} deg^{-4}.

2.2 Empirical Properties of the Radiant Energy from the Sun

The human eye is not adapted for direct observations of the Sun: under no conditions should one ever point binoculars or a telescope at the Sun.

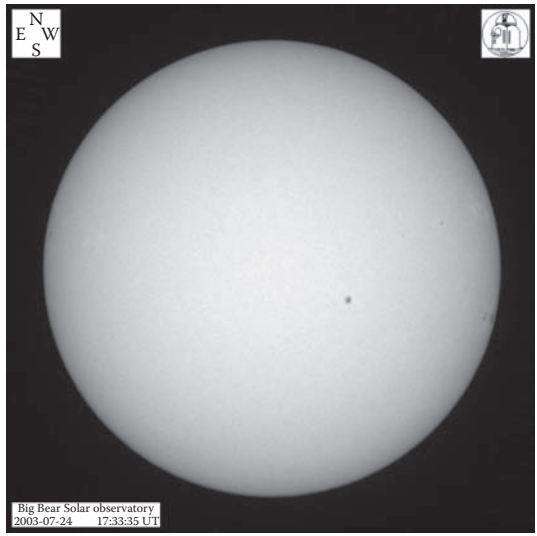


FIGURE 2.2: An image of the Sun obtained from an artificial island in a lake: in these conditions, the atmosphere around the telescope is less turbulent, and a clearer image of the Sun can be obtained from the Earth's surface. (From Big Bear Solar Observatory/New Jersey Institute of Technology. With permission.)

However, images of the Sun can be obtained with instruments which are designed for that purpose. An example of an image of the full disk of the Sun is presented in Figure 2.2.

In the image, one's eyes are usually drawn to the localized dark spots ("sunspots"), but that is not the point in the present context. (Chapter 16 deals with sunspots.) Here, we note that the solar disk is amenable to measurements of radiant intensity at all positions across the disk, from center to limb. Apart from sunspots, the intensity of the disk in visible light is azimuthally symmetric. Inspection of Figure 2.2 shows that as we move from disk center toward the limb, the intensity varies in a systematic way. The sense of the variation is such that the *limb* of the Sun is observed to be *fainter* than the center of the disk when observations are made in visible light: this gives rise to the term "limb darkening."

From our observing platform (P) on Earth, it is convenient (see Figure 2.3) to describe the location of a point S on the surface of the Sun in terms of the azimuthally symmetric angle ψ between our line of sight and the local normal to the Sun's surface at point S.

When S lies at the center of the disk (from the observer's vantage point), the line of sight from P to S enters the solar atmosphere along a line which is parallel to the local normal to the solar surface. As a result, $\psi = 0$ at the center of the disk. On the other hand, if the point S lies close to the limb,

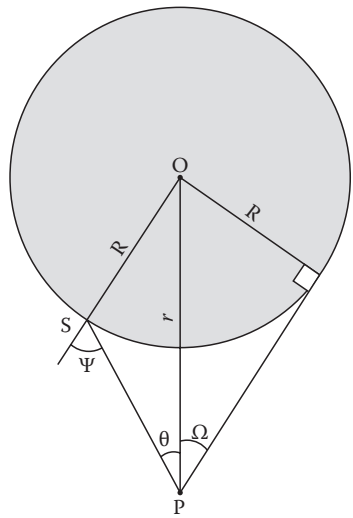


FIGURE 2.3: Defining the angle ψ which is relevant to limb darkening. The observer is at point P (e.g., the Earth), S is a point on the Sun's surface, O is the center of the Sun, Ω is the angular radius of the Sun as seen by the observer, R is the linear radius of the Sun. The angle ψ is the angle between the normal to the Sun at point S and the line of sight to the observer P. An imaginary observer standing on the Sun at point S would see the earth (P) at an angle ψ away from the local zenith. (Image downloaded from http://upload.wikimedia.org/wikipedia/commons/9/94/Limb_darkening_geometry.png. With permission.)

the line of sight from P to S intersects the solar surface at S at a large angle to the local normal. As a result, as we attempt to make measurements of the intensity at positions which lie progressively closer to the limb, our observing direction corresponds to the limit $\psi \rightarrow 90$ degrees.

In terms of the variable $\mu = \cos\psi$, observations at disk center correspond to $\mu = 1$, while observations at the limb correspond to $\mu \rightarrow 0$.

Measurements of limb darkening have been made from many sites on Earth, and at many different wavelengths. These measurements indicate that, for any particular wavelength, I_λ varies across the solar disk in a way that can be described to first order as a *first order polynomial in* μ:

$$I_\lambda(\mu) \approx a_\lambda + b_\lambda \mu \tag{2.5}$$

It may also be possible to fit higher order polynomials to the limb darkening, but the linear fit Equation 2.5 often works well. For wavelengths in the vicinity of 5000 Å, if we normalize I_λ to its value at disk center, $I_\lambda(\mu = 1)$, the empirical values of the coefficients in the linear fit are given roughly by $a_\lambda = 0.4$ and $b_\lambda = +0.6$. That is to say, when we observe the Sun in visible

light, the intensity at the limb is only 40% as large as at the center of the disk. The limb is fully 60% fainter than the center of the disk when observed at 5000 Å.

A priori, it is difficult to say whether the empirical values of the coefficients in Equation 2.5 are "reasonable." As regards a_λ (the intensity at the limb relative to the central intensity), it is limited to nonnegative values: the intensity is a physical quantity which cannot fall below zero. The actual value of a_λ is related to the mechanism of energy transport in the atmosphere: in an atmosphere where adiabatic equilibrium existed, such as would be the case if efficient convection were transporting the energy, it can be shown that the numerical value of a_λ would approach zero. Thus, the fact that the empirical value of a_λ is definitely nonzero conveys physical information: energy transport in the photosphere does *not* occur primarily by means of efficient convection. We shall see below (Section 2.8) that radiative transport yields an excellent fit to the empirical value of a_λ.

As regards b_λ, there is no obvious *a priori* reason why b_λ should necessarily be restricted to a particular algebraic sign. The empirical fact that b_λ is a positive number (for wavelengths around 5000 Å) indicates that the intensity at the limb, when observed at 5000 Å, is *less* than the intensity at disk center (hence the term: "limb darkening"). Limb darkening is a readily discernable feature of images of the Sun which are obtained in visible light (e.g., Figure 2.2).

At near infrared wavelengths ($\lambda \approx 1$ μm), b_λ it is observed to be less than 0.6: limb darkening is less severe in the infrared than at visible wavelengths. At wavelengths as long as 5 μm, limb darkening is no more than about 10%: the intensity at the limb is roughly 90% of the center intensity. At wavelengths in the near ultraviolet ($\lambda \approx 0.3$ μm), b_λ is observed to be greater than 0.6, indicating that the limb darkening is more severe in the near ultraviolet than at visible wavelengths.

The fact that the Sun is limb darkened when observed at visible wavelengths does not exclude the possibility that at other wavelengths, the limb may be observed to be *brighter* than the disk center. In fact, at long radio wavelengths, limb brightening is observed. In such a case, b_λ takes on a negative value: there is no mathematical difficulty with this as long as the sum $a_\lambda + b_\lambda$ remains nonnegative.

The numerical values of the empirical coefficients a_λ and b_λ, and the algebraic sign of b_λ, contain important information as to how the temperature $T(z)$ in the Sun's atmosphere varies as a function of the linear depth z. (Note that the depth z increases as we go *downward* into the interior of the Sun. We will also have occasion to use a linear height variable h which increases as we go *upward* in the solar atmosphere.) Obtaining this depth dependence of $T(z)$ is the first step toward determining the radial profile of temperature inside the Sun. In order to extract $T(z)$, it is first necessary to derive the radiative transfer equation (RTE), which describes how I_λ varies as a function of a related coordinate known as the "optical depth" τ.

2.3 RTE

In a radiant medium, such as the solar atmosphere, the gas emits radiant energy at a rate ε_λ ergs cm^{-3} sec^{-1} Å^{-1} ster^{-1}. This quantity is a rate of energy generation in each unit of volume in a certain region of the spectrum, across a width of spectrum equal to 1 Å, and into unit solid angle. The origin of this radiant emission can be traced ultimately to the pool of thermal energy which resides in the particles of the gas (at temperature T): as particles in this thermal pool collide with each other, the mutual accelerations of charges give rise to emission of photons with energies which are related to the thermal energies, of order kT. Because of this, the emissivity ε_λ is a function of temperature.

The gas in the solar atmosphere also absorbs radiant energy at a rate which is described by a (*linear*) *absorption coefficient* k_λ cm^{-1}. The subscript indicates that the absorption coefficient depends on the wavelength. There is also a temperature-dependence. The wavelength-dependence of k_λ is sometimes extremely rapid, e.g., in the vicinity of a strong spectral line, or near an "ionization edge." However, in other cases, k_λ varies only slowly with wavelength. This is the case in the spectrum of the Sun, for wavelengths which span much of the visible spectrum, from about 0.4 μm to about 1 μm.

In the presence of absorption, when a beam of radiation with intensity $I_\lambda(0)$ enters a uniform slab of linear thickness x, the emergent intensity is given by

$$I_\lambda(x) = I_\lambda(0)e^{-\tau} \tag{2.6}$$

The quantity $\tau = k_\lambda x$ is a dimensionless number called the optical depth of the slab at wavelength λ. The quantity $1/k_\lambda$ is a linear distance such that a slab of this thickness reduces the intensity of a beam by a factor of $1/e$. If light passes through material with optical depth $\tau = 10$, the emergent intensity is attenuated below the initial value by a factor of order 2×10^4. This means that when we view a medium where radiation is coming from a variety of depths, it becomes progressively harder to detect a significant fraction of the radiation which originated in layers of gas which have optical depths that are much larger than $\tau = 1$.

What numerical value is typical of the quantity k_λ? The answer depends on the medium. In the surface layers of the Sun, the value of k_λ at $\lambda \approx 5000$ Å is found to be of order 10^{-6} cm^{-1}. This means that two points in those layers which are separated by a distance of 10 km are separated by a medium with $\tau \approx 1$. Radiation which is emitted by gas that is, say, 100 km deeper than at a reference point P′ arrives at P′ with an intensity of only 1/20,000 times its original value. If P′ lies on the visible surface of the Sun, this means that radiation emitted from 100 km below the Sun's surface is essentially all absorbed before we have a chance to see it.

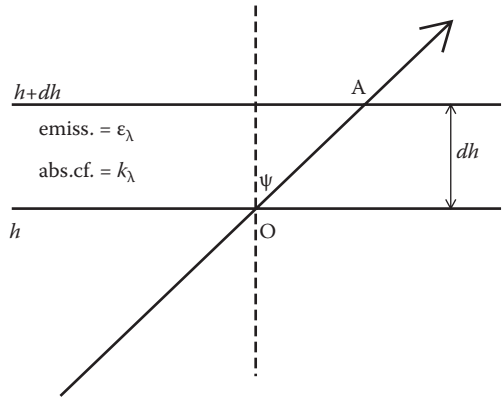


FIGURE 2.4: The arrowed line represents the path of a ray of light propagating along a slanted path from below to above, passing through a slab of material with vertical thickness *dh*. Inside the slab, the material has absorption coefficient k_λ and emissivity ε_λ. Dashed line indicates the local vertical direction. The ray enters the slab at point O, and exits the slab at point A. The ray path makes an angle ψ relative to the local normal. The path length OA of the ray as it passes through the slab has a length $dh \sec\psi$.

If k_λ is nonuniform in the slab, then the optical depth is given by $\tau = \int k_\lambda dx$. In the limit of a thin slab, with thickness dx, we see that

$$I_\lambda(x) = I_\lambda(0) - k_\lambda dx I_\lambda(0) \tag{2.7}$$

Thus, the magnitude of the reduction in I_λ associated with passing through a path length dx is proportional to the path length and also to the intensity of the radiation.

Combining the concepts of emission and absorption, we can now derive the RTE using Figure 2.4.

Consider the line OA along which radiation is propagating in a stellar atmosphere. Let OA extend across a slab which is located between heights h and $h + dh$: in our notation, the numerical value of h increases in the *upward* direction. Let the line OA lie along a direction which makes an angle ψ relative to the local normal. The intensity at height h in the direction of the line is $I_\lambda(h, \psi)$: this is a measure of the energy flux which enters an element of unit area per unit time at height h. The element of area is perpendicular to the line OA. After traversing the atmosphere and arriving at the higher level, the intensity which emerges from the element of unit area per unit time $I_\lambda(h + dh, \psi)$ differs from the value $I_\lambda(h, \psi)$ because of two processes: (i) *reduction* in intensity due to absorption along OA, and (ii) *enhancement* in intensity due to emission along OA. Let us imagine that the elements of unit area at h and at $h + dh$ are connected by a rectangular prism of unit area: the length dl of such a prism is equal to the slant length $dl = dh\sec\psi$

along the axis of the prism. The *reduction* in energy along this path is equal to $-I_\lambda(h,\psi)k_\lambda\, dh\sec\psi$. The volume of the prism is $dh\sec\psi$. In such a volume, the increase in energy flux due to local emission is $\varepsilon_\lambda(h)\,dh\sec\psi$.

Combining the enhancement and the reduction, we find that

$$I_\lambda(h+dh,\psi) - I_\lambda(h,\psi) = \varepsilon_\lambda(h)\,dh\sec\psi - I_\lambda(h,\psi)k_\lambda\, dh\ \sec\psi \tag{2.8}$$

In the limit $dh \to 0$, and setting $\mu = \cos\psi$, this leads to

$$\mu\frac{dI_\lambda}{dh} = \varepsilon_\lambda - k_\lambda I_\lambda \tag{2.9}$$

2.4 Optical Depth and the Concept of "the Photosphere"

At this point, we introduce the concept of the *optical depth which characterizes a particular layer of the atmosphere* τ_λ. The increment of this optical depth $d\tau_\lambda$ is defined by $d\tau_\lambda = -k_\lambda\, dh$, where the negative sign indicates that the numerical value of τ_λ *increases* as the height coordinate h becomes more negative, i.e., as we move *deeper* into the star. At any height h' in the atmosphere, the local optical depth is computed by integrating from a height of infinity in to the height h':

$$\tau_\lambda(h') = \int_\infty^{h'} k_\lambda\, dh \tag{2.10}$$

The zero point of the optical depth scale lies far above the visible surface of the Sun. For an observer P at a remote point (such as on Earth) (Figure 2.3), the optical depth along the line of sight between P and a point S′ (which is closer to the Sun than P) is determined by the location of S′. If S′ lies close to P, the point S′ lies in gas of extremely low density. As a result, the optical depth of point S′ (as viewed by observer P) remains small. However, as point S′ is moved progressively closer to the Sun, there is a monotonic increase in the optical depth of S′ as viewed by P. Eventually, point S′ is immersed in atmospheric gas where the density has a value that is large enough to make k_λ appreciable: now the increments $d\tau$ start to build up appreciably in the integral of $\tau_\lambda(h')$. Eventually, the gas surrounding S′ is dense enough, and deep enough in the solar atmosphere, that the integral $\tau_\lambda(h')$ approaches a value of order unity. We call that location the photosphere (from the Greek: photo = having to do with light). The numerical value of h' at the photosphere depends on the wavelength: at certain wavelengths (e.g., $\lambda \approx 1.6\ \mu$m), the absorption coefficient is smaller than at other wavelengths, and we can see deeper into the atmosphere.

Why is the photosphere significant as far as our study of solar radiation is concerned? Because as we go deeper into the Sun, below the photosphere, the optical depth of the deeper layers rapidly becomes so large that any radiation

emitted from the deep layers is significantly reduced before it can reach our observational instruments. The photosphere can be regarded as more or less the deepest lying gas from which we still have a good chance of seeing most of what is emitted.

Casting the above equation in terms of optical depth, we can re-write it as

$$\mu \frac{dI_\lambda}{d\tau_\lambda} = I_\lambda - S_\lambda \tag{2.11}$$

where $S_\lambda = \varepsilon_\lambda / k_\lambda$ is referred to as the source function at wavelength λ. The units of S_λ are the same as those of I_λ, namely, ergs cm^{-2} sec^{-1} Å^{-1} ster^{-1}.

Equation 2.11 is referred to as the radiative transfer equation (RTE).

2.5 Special Solutions of the RTE

In the following illustrative solutions, we shall for simplicity omit the wavelength subscript on all variables, but it is implied.

Because of the structure of Equation 2.11, a general aspect of the solution can readily be identified: the integrating factor is $e^{-\tau/\mu}$. Multiplying both sides by the integrating factor, we can re-write the RTE as

$$\frac{d}{d\tau}(I(\tau)e^{-\tau/\mu}) = -\frac{S(\tau)e^{-\tau/\mu}}{\mu} \tag{2.12}$$

The solution of this equation yields the intensity which an observer would "detect" if an instrument were located in a layer with optical depth τ, and if the instrument were pointed in such a way as to be observing only the radiation propagating along a direction which makes an angle $\psi = \cos^{-1}\mu$ relative to the local normal.

The formal solution $I(\tau, \mu)$ of Equation 2.12 can be considered in the limit of two distinct regimes of the μ parameter, one for the radiant intensity which flows *into the upper hemisphere* (relative to the point where the optical depth is τ), and the other for the radiation which flows *into the lower hemisphere*.

For radiation which is flowing into the *upper* hemisphere, ψ takes on values which range from 0 to 90 degrees, (i.e., μ takes on positive values between 1 and 0). In this case, the local intensity $I(\tau, \mu)$ is due to radiation which emerges from deeper layers (inside the Sun) and which is flowing outward toward free space. As a result, the local value of intensity at depth τ in the upper hemisphere (denoted by $\mu+$) involves an integration over all gas which lies deeper than τ, from $\tau \to \infty$ up to the level where the optical depth equals τ:

$$I(\tau, \mu+) = -e^{\tau/\mu} \int_\infty^\tau \frac{S(t)}{\mu} e^{-t/\mu} dt \tag{2.13}$$

For radiation which is flowing into the *lower* hemisphere, ψ takes on values from 90 to 180 degrees, (i.e., μ takes on negative values between -1 and 0). In this case, the local intensity $I(\tau, \mu)$ is due to radiation which emerges from shallower layers (higher in the atmosphere) and which is flowing inward toward the interior of the Sun. As a result, the local value of intensity at depth τ in the lower hemisphere (denoted by $\mu-$) involves an integration over all gas which lies shallower than τ, from $\tau \rightarrow 0$ down to the level where the optical depth equals τ:

$$I(\tau, \mu-) = -e^{\tau/\mu} \int_0^{\tau} \frac{S(t)}{\mu} e^{-t/\mu} dt \tag{2.14}$$

Let us limit our considerations now to the outermost layers of the Sun. That is, let us move our radiation instrument to the upper atmosphere where $\tau \rightarrow 0$. This could include moving the instrument all the way to the Earth's orbit. In this way, we would be recording what is truly the "emergent intensity" from the Sun. In this case, since there is essentially zero source of radiation coming in from free space, the integral into the lower hemisphere $I(\tau, \mu-)$ vanishes. Only the intensity $I(\tau, \mu+)$ entering into the upper hemisphere retains a nonzero value. And this component, in the limit $\tau \rightarrow 0$, becomes

$$I(0, \mu+) = \int_0^{\infty} \frac{S(t)}{\mu} e^{-t/\mu} dt \tag{2.15}$$

Let us consider some simple cases.

2.5.1 $S =$ Constant at all optical depths

If $S(\tau) = S$, independent of τ, the integral in Equation 2.15 is straightforward: we find $I(0, \mu+) = S$. Thus, the emergent intensity of radiation is just equal to S itself. Moreover, $I(0)$ is independent of μ: there is neither limb darkening nor limb brightening.

2.5.2 Constant S in a slab of finite thickness

In the case of a slab with finite optical depth τ', in which S is constant, the emergent intensity is

$$I(0, \mu+) = S \int_0^{\tau'} e^{-t/\mu} dt/\mu = S(1 - e^{-\tau'/\mu}) \tag{2.16}$$

Thus, the emergent intensity is not as large as S, but is reduced by an optical depth term. In the special case where we observe perpendicular to the slab, we can set $\mu = 1$, and then find that

$$I(0,1) = S(1 - e^{-\tau'}) \tag{2.17}$$

In the limit of infinite thickness, $\tau' \to \infty$, we recover the solution in Section 2.5.1: $I(0,1) \to S$.

In the opposite limit, when the slab is optically thin, we find

$$I(0,1) \to \tau' S \tag{2.18}$$

Thus, the emergent intensity from a very thin slab can take on values which are much less than the source function. The reduction factor is just the optical depth of the slab.

In general, in the case of constant S, the emergent intensity cannot be greater than S, but it may be much smaller than S if the optical depth is small. This is an important result in helping to interpret certain properties of the upper solar atmosphere.

2.5.3 Depth-dependent S: polynomial form

We now revert to the case of an infinite atmosphere, and consider a case where the source function depends on the optical depth. Specifically, we consider the polynomial form $S(\tau) = a + b\tau + c\tau^2$. (We shall see in Section 2.8, that there is some basis for such a choice.) To obtain the emergent intensity from such an atmosphere, we insert this function into Equation 2.15.

The first term in $S(\tau)$ corresponds to the case in Section 2.5.1 (i.e., constant S): this term results in a contribution of a to $I(0, \mu+)$. The term $b\tau$, when inserted in Equation 2.15 leads to an integral which can be integrated by parts: it contributes a term $b\mu$ to $I(0, \mu+)$. Finally, the term $c\tau^2$, when inserted in Equation 2.15, requires two integrations by parts: this leads to a term $2c\mu^2$ to $I(0, \mu+)$. Combining terms, we find that

$$I(0, \mu+) = a + b\mu + 2c\mu^2 \tag{2.19}$$

Clearly, this solution is of particular interest for the Sun's atmosphere since the empirical limb darkening of the Sun (Equation 2.5) is of precisely this form (in the special case $c = 0$, although empirical fits *can* be extended to include a term in μ^2). It therefore appears that the source function at optical depth τ in the Sun (at visible wavelengths) can be described by the function $S(\tau) = a + b\tau$. In view of the fact that the empirical value of the coefficient b is positive, the source function in the solar atmosphere at visible wavelengths *increases with increasing* τ.

We shall see below that the source function can be related to the local temperature. In view of this, the empirical observation of limb *darkening* (i.e., $b > 0$) provides us with a significant piece of information: in the visible layers of the solar atmosphere, the temperature *increases* as we penetrate *deeper* into the atmosphere. This is the start of a radial temperature gradient which will carry us from temperatures of order $T_{\rm eff}(\approx 6000$ K) near the surface to much larger temperatures in the deep interior of the Sun.

2.5.4 Depth-dependent S: exponential form

Suppose the source function has the form $S(\tau) = e^{\alpha\tau}$ where $\alpha < 1/\mu$. Inserting this into Equation 2.20, we find that

$$I(0, \mu+) = \frac{1}{1 - \alpha\mu} \tag{2.20}$$

2.6 Eddington–Barbier Relationship

The fact that a linear source function $S(\tau) = a + b\tau$ yields a limb darkening function $I(0, \mu+) = a + b\mu$ which is exactly the same linear function of μ leads to the Eddington–Barbier relationship: the intensity which is observed at any value of μ equals the source function at the level where the local optical depth τ has the value $\tau = \mu$. In other words, at any particular location on the disk of the Sun, i.e., at a given value of μ, the radiation that is observed comes effectively from gas where the optical depth is equal to μ.

In effect, when one observes the Sun at disk center ($\mu = 1$), one's line of sight penetrates essentially down to the gas in the solar atmosphere where $\tau = 1$. On the other hand, when observing near the limb, say at $\mu = 0.1$, one's line of sight penetrates only down to the gas where $\tau = 0.1$. In terms of a height scale in the solar atmosphere, the gas at $\tau = 1$ lies 100–200 kilometers deeper than the gas at $\tau = 0.1$. The deeper gas is hotter.

2.7 Is Limb Brightening Possible?

Although limb darkening is certainly the feature which is most relevant to observations of the Sun in the visible continuum (see Figure 2.2), this does *not* exclude the possibility of limb brightening when the Sun is observed at other wavelengths.

The existence of limb *brightening* requires that $I(0, \mu = 0)$ exceed $I(0, \mu = 1)$.

In the case of a polynomial source function, this possibility is formally excluded if all coefficients (a, b, c) are nonnegative. Limb brightening is possible only in cases where either b or c is sufficiently negative to ensure that $b + 2c$ is negative.

In the case of an exponential source function, the ratio $I(0, \mu = 0)/I(0, \mu = 1)$ is equal to $1 - \alpha$. To avoid nonphysical (negative) intensities, this requires that α have a value that is no greater than 1. This is stricter than the limit (already noted above) $\alpha < 1/\mu$. Thus, limb brightening is possible if $\alpha < 0$, i.e., if the source function *decreases* exponentially as the

optical depth *increases.* Although this result is of little relevance to the solar photosphere, it will be useful when we consider the solar chromosphere and corona.

2.8 Is $S(\tau) = a + b\tau$ Realistic? The Gray Atmosphere

We have seen that the observed limb darkening in the Sun, which can be described by $I(\mu) = a + b\mu$, agrees with the limb darkening which is predicted to be observed if the source function has a particular form: $S(\tau) = a + b\tau$. Now we ask, is there any physical reason why $S(\tau) = a + b\tau$ should be an acceptable description of the depth-dependent source function in the Sun?

The answer is Yes, provided we consider a limiting case known as the *gray atmosphere.* In this case, the opacity is independent of wavelength, allowing immediate integration of Equation 2.11 (RTE) over frequency.

$$\mu \frac{dI(\tau)}{d\tau} = I(\tau) - S(\tau) \tag{2.21}$$

The unsubscripted τ-dependent variables $I(\tau)$ and $S(\tau)$ in this section refer to quantities which, at any given optical depth, have been integrated over the frequency spectrum.

At any optical depth in the atmosphere, the flux of radiation $F(\tau)$ flowing towards an outside observer, integrated over all frequencies, is given by the flux integral:

$$F(\tau) = \int \mu I(\tau) d\omega \tag{2.22}$$

where the integration over $d\omega$ represents all solid angles, and the direction $\psi = 0$ points toward the observer. In conditions of radiative equilibrium, there are no new sources of energy within the atmosphere: the energy flux $F(\tau)$ is provided by processes which occur deep inside the star. As far as the atmosphere is concerned, $F(\tau)$ is to be regarded as a boundary condition: a certain quantity of energy flux "arrives" from the deep interior at the base of the atmosphere, and must be transported through the atmosphere and released into the darkness of space. Therefore, $F(\tau)$ is a constant at all optical depths in the atmosphere. We use the symbol F_o to denote this constant flux.

At optical depth τ in the atmosphere, we define the mean intensity of radiation $J(\tau)$ as

$$J(\tau) = \frac{1}{4\pi} \int I(\tau) d\omega \tag{2.23}$$

With these definitions of $F(\tau)$ and $J(\tau)$, we integrate both sides of Equation 2.11 over $d\omega$ and find that at any given location in the atmosphere,

in the vicinity of optical depth τ,

$$\frac{dF(\tau)}{d\tau} = 4\pi J(\tau) - 4\pi S(\tau) \tag{2.24}$$

In performing the integrations over $d\omega$, we have assumed that the source function $S(\tau)$, which is determined by atomic processes in the immediate neighborhood of τ, is spherically symmetric. Inserting $F(\tau) = F_o$ in Equation 2.24, we find that

$$J(\tau) = S(\tau) \tag{2.25}$$

Recall that both $J(\tau)$ and $S(\tau)$ have been integrated over all frequencies.

The quantities $J(\tau)$ and $F(\tau)$ represent zeroth and first moments of the radiation intensity at optical depth τ. Now we introduce the second moment of the radiation intensity at depth τ (again integrated over all frequencies):

$$K(\tau) = \frac{1}{4\pi} \int \mu^2 I(\tau) d\omega \tag{2.26}$$

The quantity $K(\tau)$ is proportional to the radiation pressure p_r at optical depth τ. When the spectrum of a black-body is integrated over all wavelengths, the radiation pressure p_r is related to the energy density u (see Section 2.1) by $p_r = u/3 = a_R T^4/3$.

With these definitions, let us now multiply both sides of Equation 2.11 by $(\mu/4\pi)$ and integrate over $d\omega$. This operation leads to $dK(\tau)/d\tau$ on the left-hand side of RTE. On the right-hand side, the first term reduces to the constant $F_o/4\pi$. The second term, involving integration of μS over all solid angles, reduces to zero due to the spherical symmetry of S. Thus we find

$$\frac{dK(\tau)}{d\tau} = \frac{F_o}{4\pi} \tag{2.27}$$

Since F_o is a constant, this equation can be integrated immediately to obtain:

$$K(\tau) = \tau \frac{F_o}{4\pi} + \text{constant} \tag{2.28}$$

How are we to evaluate the "constant" in Equation 2.28? We introduce the "two-stream approximation": the angular distribution of the radiant intensity is replaced by two streams, one with $\mu = +1$ going into the outer (upper) hemisphere $I_o(\tau)$, the other with $\mu = -1$ going into the inner (lower) hemisphere $I_i(\tau)$. In this approximation, and noting that the element of solid angle $d\omega$ can be written as $2\pi d\mu$, we find the following expression for the three moments of the radiation field:

$$J(\tau) = \frac{1}{2}(I_o(\tau) + I_i(\tau)) \tag{2.29}$$

$$F(\tau) \equiv F_o = \pi(I_o(\tau) - I_i(\tau)) \tag{2.30}$$

$$K(\tau) = \frac{1}{6}(I_o(\tau) + I_i(\tau)) \tag{2.31}$$

Comparing $J(\tau)$ and $K(\tau)$, we see that in this approximation, referred to as the Eddington approximation, $K(\tau) = J(\tau)/3$ at all depths in the atmosphere.

In particular, let us consider the top of the atmosphere, where $\tau = 0$. At that location, the incoming flux of radiant energy $I_i(\tau = 0)$ is zero. As a result, $K(0) = I_o(0)/6$ and $F_o = \pi I_o(0)$. Reverting to Equation 2.28, we now have enough information to evaluate the "constant": it equals $F_o/6\pi$. Replacing $K(\tau)$ by $J(\tau)/3$, and multiplying both sides by three, we then find

$$J(\tau) = \frac{F_o}{2\pi} + \tau\frac{3F_o}{4\pi} \tag{2.32}$$

Since $J(\tau) = S(\tau)$ (see Equation 2.25 above), we finally have

$$S(\tau) = \left(\frac{3F_o}{4\pi}\right)\left(\frac{2}{3} + \tau\right) \tag{2.33}$$

This is the Eddington solution for the gray atmosphere.

Now we can answer the question: is there any physical basis for considering the function $S(\tau) = a + b\tau$?

Indeed there is: the Eddington solution yields just such a solution, with a specific value of 2/3 for the ratio of a/b. The function in Equation 2.33 therefore leads to a limb darkening of the form $I(\mu) \sim (2/3) + \mu$. Thus, we see that the intensity at the limb (where $\mu = 0$), i.e., $I(0) \sim (2/3)$, is only 40% of the intensity at the center of the disk (where $\mu = 1$), i.e., $I(1) \sim (5/3)$. In this regard, we recall that the limb darkening of the Sun (Section 2.2) is in fact observed to be close to a linear function of μ, with a limb intensity of about 40% of center intensity in visible wavelengths, just as the model here predicts. Apparently, the assumptions of the Eddington solution provide a valuable approach to replicating, in a quantitative manner, the observed limb darkening of the Sun at visible wavelengths. We shall return to why this might be so when we discuss possible sources of opacity in the solar photosphere at visible wavelengths.

A more detailed solution of the RTE in a gray atmosphere (Chandrasekhar, 1944) shows that, rather than the solution in Equation 2.33, where $S(\tau) \sim \tau + (2/3)$, a more exact solution is $S(\tau) \sim \tau + q(\tau)$ where $q(\tau)$ is a slowly-varying function of τ, ranging from 0.58 as $\tau \to 0$ to 0.71 as $\tau \to \infty$. The two-stream approximation, which replaces $q(\tau)$ with the constant 2/3 is entirely consistent with the more detailed solution.

2.9 How Does Temperature Vary as a Function of τ?

Now that we derived how the source function behaves as a function of τ, our aim here is to derive the τ-dependence of the temperature. To do this,

we use the equality established above between $S(\tau)$ and $J(\tau)$. Recalling the frequency-dependent definition of $J_\nu(\tau) = (1/4\pi)\int I_\nu(\tau)d\omega$, we notice that $J_\nu(\tau)$ at any particular frequency is related to the energy density of the radiation u_ν at that frequency. In order to evaluate u_ν (with units of ergs cm^{-3} Hz^{-1}) we must integrate $I_\nu(\tau)$ (with units of ergs cm^{-2} sec^{-1} ster^{-1} Hz^{-1}) over solid angle, and divide by the speed of light:

$$u_\nu(\tau) = \frac{1}{c}\int I_\nu(\tau)d\omega \tag{2.34}$$

Comparing Equations 2.23 and 2.34, we see that

$$J_\nu(\tau) = \frac{c}{4\pi}u_\nu(\tau) \tag{2.35}$$

We have already noted (Section 2.1) that when the energy density u_ν of black-body radiation is integrated over all frequencies for an object of temperature T, the result is $u(T) = a_R T^4$, where a_R is the radiation density constant. Therefore, if we integrate $J_\nu(\tau)$ over all frequencies, we find $J(\tau) = (a_R c/4\pi)T(\tau)^4$. Noting also that $a_R c/4 = \sigma_{\rm B}$ (the Stefan–Boltzmann constant), we find that

$$J(\tau) = \sigma_{\rm B}T(\tau)^4/\pi \tag{2.36}$$

Since $S(\tau) = J(\tau)$ (see Equation 2.25), Equation 2.36 can be written as

$$S(\tau) = \sigma_{\rm B}T(\tau)^4/\pi \tag{2.37}$$

Inserting this into Equation 2.33 we find

$$\frac{\sigma_{\rm B}T(\tau)^4}{\pi} = \left(\frac{3F_o}{4\pi}\right)\left(\frac{2}{3}+\tau\right) \tag{2.38}$$

The constant flux F_o which propagates through the atmosphere can be expressed in terms of an effective temperature $T_{\rm eff}$ by means of the definition

$$F_o \equiv \sigma_{\rm B}T_{\rm eff}^4 \tag{2.39}$$

Combining Equations 2.38 and 2.39, we finally arrive at an expression for the profile of temperature as a function of optical depth in an Eddington atmosphere:

$$T(\tau)^4 = \frac{T_{\rm eff}^4}{4}(2+3\tau) \tag{2.40}$$

We shall refer to Equation 2.40 as the "Eddington relation." It is this relation which will eventually start us on the way to deriving profiles of density and pressure at various heights in the atmosphere of the Sun.

Note that the source function, while at first sight (Equation 2.37) appears similar to the expression for the flux F_o (Equation 2.39), differs from the flux in two ways: (i) $S(\tau)$ varies with τ, but F_o is independent of τ; (ii) $S(\tau)$ includes an extra factor of π in the denominator.

2.10 Properties of the Eddington Relation

In the upper layers of the atmosphere, as $\tau \to 0$, the Eddington relation predicts that the temperature does *not* by any means fall to zero. Instead, it approaches an asymptotic limit, the "boundary temperature,"

$$T_{\text{boundary}} \equiv T(\tau = 0) = \frac{T_{\text{eff}}}{2^{1/4}} \tag{2.41}$$

In the case of the Sun, with $T_{\text{eff}} = 5777$ K, we find a boundary temperature of 4858 K. As a result, the Planck function in the upper atmosphere does *not* go to zero, but tends to a constant value in the uppermost levels.

The optical depth at which the local temperature has a value which is equal to T_{eff} is $\tau = 2/3$. This is consistent with the observation that the photons we see coming from the Sun, emerging on the whole from layers where the optical depth cannot be much greater than $\tau \approx 1$, appear to emerge from a gas with a temperature that is about 6000 K.

Reference

Chandrasekhar, S. 1944. "On the Radiative Equilibrium of a Stellar Atmosphere. II.," *Astrophys. J.*, 100, 76.

Chapter 3

Toward a Model of the Sun: Opacity

Now that we have limb-darkening information as to how the temperature in the vicinity of the solar photosphere behaves as a function of optical depth, we have taken the first step in achieving one of the principal goals of solar physics: to calculate how the physical quantities in the Sun behave as a function of radial location. We refer to such a radial profile as a "solar model."

When the only information that we have access to is limb darkening, the range of radial locations in the Sun which can be modeled reliably is quite restricted: we can extract information only for a range of heights in the vicinity of the photosphere. For present purposes, the "vicinity of the photosphere" refers to locations in the solar atmosphere which lie *above* the convection zone, and *below* the chromosphere. For the sake of brevity, we refer to these limits as the "lower" photosphere and the "upper" photosphere, respectively. In what follows, given the physical conditions which exist in the solar atmosphere, we shall find that the "lower" and "upper" photosphere differ in height by $\Delta h =$ several hundred kilometers.

In subsequent chapters, when we wish to progress our modeling efforts deeper into the interior of the Sun, or upward into the chromosphere, we shall need access to data over and above what limb darkening can provide. But in all cases, we need to know the answer to the following question: what is the opacity? That is the subject of this chapter.

3.1 Relationship between Optical Depth and Linear Absorption Coefficient

Up to this point, the RTE has been discussed in terms of k_λ cm^{-1}, which is a *linear* absorption coefficient. More customary in astrophysics is the opacity, $\kappa_\lambda = k_\lambda/\rho$: this is a measure of how opaque a medium of density ρ is to light of wavelength λ. The units of κ_λ are cm^2 gm^{-1}. The units of κ_λ indicate that the opacity is associated with a cross-sectional area which impedes the free passage of radiation as the radiation propagates through 1 gm of material.

Since opacity includes a cross-sectional area which is responsible for scattering and/or absorbing light, it is worthwhile to mention a fundamental cross-section associated with a free electron. A free electron scatters a photon with

a cross-section which is given by the Thomson formula $\sigma_{\rm T} = (8/3)\pi r_{\rm e}^2$ where $r_{\rm e} = e^2/mc^2$ is the classical radius of the electron. In discussions about the interactions between photons and matter, the Thomson cross-section is an important quantity:

$$\sigma_{\rm T} = 6.6245873 \times 10^{-25} \ {\rm cm}^2 \tag{3.1}$$

In an ionized gas, there will always be free electrons which contribute to opacity with the above cross-section. Note that $\sigma_{\rm T}$ is independent of wavelength (at least for photons with energies less than $m_{\rm e}c^2 = 0.5$ MeV).

However, in the gas which exists at various locations in the Sun, as well as free electrons, there are also atoms and/or ions in which some electrons are still held in bound orbits. Opacity is a measure of how strongly photons interact with the atoms/ions of the medium through which the photons are passing. We shall find that quantitative aspects of the opacity, and especially its sensitivity to temperature, play a key role in modeling three regions of the Sun: the photosphere, the deep interior, and the chromosphere. Because of this, the more we understand the properties of opacity, its numerical values and its variations with temperature, the more insight we will have into the structure of the Sun.

3.2 Two Approaches to Opacity: Atomic and Astrophysical

There are two different approaches to opacity, depending on one's interest: atomic physics or astrophysics. From the *atomic* point of view, the main goal is to understand the following: given a photon with a specified wavelength (but of unspecified origin), what is the cross-section for photon absorption by a particular atom? Quantum mechanics can be used to derive quantitatively the numerical value of the cross-section at any particular wavelength. We shall take this viewpoint in Sections 3.4, 3.5, and 3.6.

From the point of view of *astrophysics*, the main goal is to understand how photons with a wide range of wavelengths interact with the medium through which they are passing. In the solar atmosphere, opacity involves a process of interaction between photons and atoms: it is not merely the atomic physics which is relevant, but also the "spectrum of radiation" in the atmosphere, and how this spectrum overlaps with regions of large and small opacity. The key question is: how many photons are present at wavelengths where the opacity is large, and how many photons are present at wavelengths where the opacity is small? It is all very well for the atomic physicist to report that hydrogen absorbs most strongly at wavelengths 912–1216 Å, but if researcher A is studying the flow of radiation through an atmosphere where there are essentially no photons at 912–1216 Å, then the peak in absorption at 912–1216 Å

is of no great relevance to researcher A: the radiation in the atmosphere, with most of its photons at (say) long wavelengths, may encounter very little effective opacity. However, if researcher B is considering an atmosphere where the photon spectrum peaks at 912–1216 Å, then the effective opacity may indeed be very large.

In astrophysics, the ease with which photons propagate through an atmosphere depends on a convolution of the photon spectrum with the wavelength-dependent opacity. Thus, we need to have a way of calculating a mean opacity of some sort, such that the relevant atomic physics as well as the relevant spectral information can be merged in a meaningful way. This will lead us to define a "mean opacity" in Section 3.6 after we have considered certain details of atomic physics.

3.3 Atomic Physics: (i) Opacity due to Hydrogen Atoms

To see how the opacity is related to the properties of individual atoms, consider Figure 3.1. This shows what an observer sees when looking through the "endface" of a column of length l. The area of each "endface" of the column is 1 cm^2, and the observer's line of sight passes through a medium (inside the column) which has density $\rho = n_{\rm a} m_a$. Here, $n_{\rm a}$ cm^{-3} is the number density of absorbers, each of mean mass m_a (gm).

Each absorber in the column has a cross-sectional area σ for the absorption of light. The total number of absorbers in the column is the *column number density*: $N = n_{\rm a} l$, with units of cm^{-2}. (An equivalent quantity is the *mass*

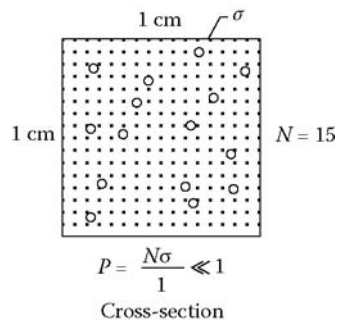


FIGURE 3.1: Indicating how to estimate the blockage of light passing through a column of material. Each endface of the column is a square with dimensions 1×1 cm. In the Figure, we are looking in at one endface, and the column extends a length l behind the page. (From the website http://mysite.du.edu/∼jcalvert/phys/scat2.gif, used with permission from J. Calvert.)

column density $d_c = Nm_a$ with units of gm cm^{-2}.) In the limit $N\sigma \ll 1$, light which enters the column at one "endface" and emerges through the other "endface," encounters a total absorption area of $N\sigma$ cm^2. Comparing this to the 1 cm^2 area of the "endface," this means that the light emerging from the column is reduced by the fractional amount $N\sigma$. Recalling that, in the limit of small optical depth, $e^{-\tau} = 1-\tau$, we see that the fractional reduction in the light can be set equal to τ. By definition, in the limit of small τ, we have that $\tau = lk_\lambda$. This indicates that k_λ can be considered as equal to $n_{\rm a}\sigma$. Note that σ depends on wavelength, so we can write, more formally, that $k_\lambda = n_{\rm a}\sigma_\lambda$.

Converting to opacity, this leads to $\kappa_\lambda = k_\lambda/\rho = \sigma_\lambda/m_a$. This expression helps us to see an important aspect of opacity. Since $1/m_a$ is the number of absorbers in 1 gm of the material, the opacity can be written as the product of two factors: (i) the *cross-section of an individual absorber*, and (ii) the *number of absorbers per gram.*

For gas of a given composition, m_a is a constant, independent of wavelength. So if we examine a plot of κ_λ versus λ, then we will be able to trace how the cross-section σ_λ of individual absorbers behaves at each wavelength. When the atoms (or ions) in the medium through which light is propagating contain electrons in bound energy levels, σ_λ may vary by many orders of magnitude as a function of wavelength. To see why this is so, let us consider the simplest case: hydrogen atoms.

3.3.1 Absorption from the ground state: dependence on λ

If hydrogen atoms are in a medium with low enough temperature, only the ground state has a significant population. Since the energy of the ground state lies at an energy of $E_1 = -13.6$ eV below the continuum, and since the lowest excited state lies at an energy $E_2 = -3.4$ eV, photons with energies less than 10.2 eV (i.e., with wavelength $\lambda > 1216$ Å) do not have sufficient energy to excite the electron from the ground state into any other energy level. As a result, apart from some weak (Rayleigh) scattering off the atom as a whole, κ_λ is small at $\lambda > 1216$ Å (see Figure 3.2).

Notice the units of opacity in Figure 3.2. First of all, they are expressed per hydrogen atom. Also, they are expressed in the unit which has already been mentioned above (Equation 3.1): the Thomson cross-section $\sigma_{\rm T}$ for scattering of photons off a free electron. The plotted numbers are at many wavelengths much larger than unity: the largest number in the plot is 10^{10}. Thus, the plot gives a clear indication that an electron which is *bound inside a hydrogen atom* can be much better (by several orders of magnitude) at scattering photons than a *free electron.* This is a key concept for understanding opacity.

For a narrow range of wavelengths in the vicinity of $\lambda = 1216$ Å, the value of σ_λ increases to large values: this interaction is associated with the bound-bound transition which gives rise to the Lyman-α spectral line. Shortward of Lyman-α, over a range of about 200 Å in wavelength, there is another range of wavelengths in which σ_λ is again small. In the vicinity of $\lambda = 1026$ Å

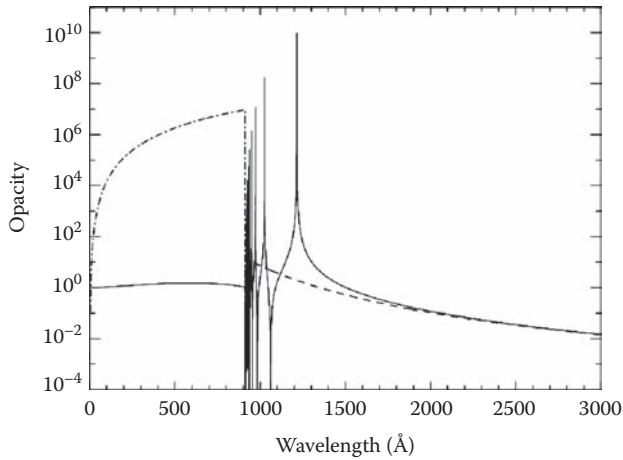


FIGURE 3.2: Lyman scattering opacity per hydrogen atom, in units of the Thomson scattering cross-section per electron. At wavelengths longer than 912 Å, Lyman lines are plotted as solid "spiky" features. At wavelengths shorter than 912 Å, opacity due to the Lyman continuum is plotted as a dot-dashed line. (From Stenflo, J. O. 2005. *Astron. Astrophys.,* 429, 713. With permission from Springer Science and Business Media.)

(Lyman-β), excitation from the ground state to the energy level E_3 at -1.51 eV gives rise to a locally significant value of σ_λ. A series of line absorptions (the Lyman series of lines), separated by regions of continuum where σ_λ is small continues until the wavelength becomes as short as $\lambda_1 = 912$ Å. At that point, σ_λ increases by several orders of magnitude, and stays large over a wavelength interval of hundreds of Å. Why? Because photons with $h\nu_1 \equiv hc/\lambda_1 = 13.6$ eV have an energy which is large enough to cause a bound-free transition, i.e., to ionize the atom. At wavelengths shorter than 912 Å, σ_λ decreases systematically ($\sim\lambda^3$) toward shorter wavelengths.

When σ_λ is plotted as a function of λ (see Figure 3.2), the values are small for all λ greater than (roughly) 1500 Å. At 912 Å, the value of σ_λ increases discontinuously to a peak value. The discontinuity is referred to as the Lyman "edge."

How large does the cross-section for photon absorption become at $\lambda =$ 912 Å? Quantum mechanical calculations indicate that the peak value is $\sigma_{912} = 6 \times 10^{-18}$ cm^2. Should this be considered as a "large" or a "small" number in terms of the ability to interact with a photon? To make a meaningful comparison, we note that the Thomson cross-section is some *seven orders of magnitude less* than σ_{912}.

By occupying the ground state of hydrogen, an electron enhances its ability to interact with a photon of wavelength 912 Å by a factor of 10 million. The conclusion is that an electron which is in a bound orbit can be (if the

wavelength is right) a powerful and effective absorber of photons, up to 10 million times more effective at absorbing a photon than a free electron.

Another way to look at the "largeness" of σ_{912} is to consider the area of the ground state orbit. According to the Bohr model, the orbital radius of the ground state is 0.528×10^{-8} cm. The corresponding orbital area is 0.88×10^{-16} cm^2. The value of σ_{912} amounts to almost 10% of this area. As far as a passing 912 Å photon is concerned, it is as if the ground state electron in the H atom has "spread itself out" over a significant fraction of the orbital area.

The large absorbing efficiency of a bound electron is an important result for understanding why the numerical value of opacity varies strongly in different regions of the Sun. In any region where there are still an appreciable fraction of electrons bound to abundant nuclei, the opacity may be much larger than elsewhere in the star. This fact plays a key role in determining the internal structure of the Sun: in certain regions where hydrogen retains its electron, or where helium retains at least one of its electrons, the passage of photons may be rendered so difficult that radiation can no longer serve as an effective method of transporting energy.

Now that we know the cross-section at the Lyman edge, we can calculate the opacity. In a gas consisting only of hydrogen, the number of H atoms per gram is $1/m_{\rm H} \approx 6 \times 10^{23}$. As a result, the opacity at the Lyman edge has the numerical value $\kappa_\lambda \approx 3.6 \times 10^6$ cm^2 gm^{-1}. We shall return to this number later.

We have referred to ways in which photons can lose/gain energy from bound-bound transitions and from bound-free transitions. There is a third class of interactions which also allow loss or gain of photon energy: these are "free-free" transitions. A free electron in a plasma can be thought of as being in a orbit (admittedly unbound) around a distant proton. If a passing photon can cause that electron to move farther away from, or nearer to, the distant proton, then the photon has caused the electron essentially to make a transition into a different (again unbound) orbit. If the total energy of the new orbit is greater than before, then the electron has gained energy from the photon: the photon experiences this interaction as "free-free" opacity.

3.3.2 Absorption from excited states: dependence on λ and T

So far, we have considered hydrogen atoms at "low" temperatures. In such a case, there is essentially no interaction with photons at wavelengths longward of (roughly) 1500 Å. However, in a medium which is sufficiently hot, absorption at longer wavelengths becomes possible. Thermal excitation causes some electrons to populate excited levels at a more or less significant rate. As far as photon absorption is concerned, each excited level displays at a certain wavelength its own "edge" where a free-bound transition can occur. Each of these "edges" has similar characteristics to those of the Lyman "edge." That is, photons on the longward side of the "edge" pass essentially

freely through the gas, but photons which lie shortward of the "edge" can be effectively absorbed. The "edges" corresponding to the energy levels with principal quantum numbers n lie at wavelengths $\lambda_n(\mathring{A}) = 912\,n^2$. Two of these "edges" are of interest for the Sun because they lie in a part of the spectrum where the Sun emits much of its power: $\lambda_2 = 3648$ Å, and $\lambda_3 = 8208$ Å. These are the Balmer and Paschen edges, respectively.

The peak opacity of hydrogen gas *at the Lyman edge* is essentially independent of temperature, at least as long as hydrogen is not significantly ionized. But this is not the case for the other "edges": in those cases, the magnitude of the peak opacity (in cm^2 *per gram*) depends on the fraction of the atoms which have an electron in the corresponding excited state. In a medium where the total number density of hydrogen is $n_{\rm H}$, the number density n_i of H atoms with electrons in the $n = i$ level is related to the number density in the ground state (n_1) by

$$n_i = \frac{g_i}{g_1} n_1 \exp\left(-\frac{\Delta E_{i1}}{kT}\right) \tag{3.2}$$

The term g_i is the statistical weight of level i: in a hydrogen atom, $g_i = 2i^2$. The quantity ΔE_{i1} is the energy difference between levels 1 and i. When the energy difference ΔE is expressed in units of eV, the exponential term is more conveniently written in the form $10^{-\theta \Delta E}$, where $\theta = 5040/T$.

Consider, for example, the $n = 2$ level of hydrogen which can be ionized by photons at the Balmer edge. For this level, $\Delta E = 10.2$ eV. If we consider by way of example a medium where the temperature T=10^4 K, we find $n_2/n_1 = 2.89 \times 10^{-5}$. Thus, in a parcel of gas which contains 1 gm of hydrogen at $T = 10^4$ K, only one atom in 35000 is capable of absorbing photons at the Balmer edge. Now, in terms of quantum mechanics, the cross-section for a single H atom to undergo photoionization from $n = 2$ by means of a photon of wavelength λ_2 is not greatly different from that for photoionization from $n = 1$ by means of a photon of wavelength λ_1. As a result, when we convert to the absorption cross-section *per gram of material* (=opacity), the peak opacity at the Balmer edge at $T = 10^4$ K is no more than $\zeta_{\rm BL} = 3 \times 10^{-5}$ times the peak opacity at the Lyman edge.

Because of the exponential factor in Equation 3.2, the magnitude of the reduction factor $\zeta_{\rm BL}$ becomes rapidly smaller at low temperatures. For example, in the lower photosphere, where $T = 6000$ K, $\zeta_{\rm BL}$ is of order 10^{-8}, while in the upper photosphere, where $T = 4900$ K, $\zeta_{\rm BL}$ falls to 10^{-10}.

Since the opacity κ_λ for a medium consisting of pure hydrogen has the value $\approx 3.6 \times 10^6$ cm^2 gm^{-1} at the Lyman peak, the numerical value of opacity in the solar photosphere at the Balmer peak $\kappa_{\rm B}$ does not exceed 0.036 cm^2 gm^{-1}. On the redward side of the peak, i.e., at wavelengths longer than 3648 Å, $\kappa_{\rm B}$ is zero.

Another bound level of hydrogen which is relevant in a discussion of photons at visible wavelengths is the $n = 3$ level, which can be ionized by photons at the Paschen edge. For this level, $\Delta E = 12.1$ eV. As a result, in the lower

photosphere, where $T = 6000$ K, $n_3/n_1 = 6 \times 10^{-10}$. As a result, the Paschen peak opacity is 10 to 20 times smaller than the Balmer peak opacity in the lower solar photosphere. Thus, κ_P does not exceed 0.0036 cm^2 gm^{-1} at wavelengths near 8200 Å. Applying the λ^3 law, we see that at wavelengths close to 4000 Å, the Paschen continuum opacity in the photosphere does not exceed 0.0004 cm^2 gm^{-1}.

As a result, when we consider the "visible spectrum" of the Sun, usually considered to extend between wavelengths of $\lambda \approx$ 4000–7000 Å, the opacity due to atomic hydrogen κ_v ranges from about 0.0004 to at most 0.002 cm^2 gm^{-1}. Later, we shall see (Chapter 5, Section 5.1) that the mass column density in the photosphere is roughly $d_c \approx 4$ gm cm^{-2}. Multiplying κ_v and d_c, we find that absorption by atomic hydrogen contributes an optical depth $\tau_v = \kappa_v d_c$ of no more than 0.008 in the photosphere. Since by definition, the photosphere is the region where the optical depth is of order unity, it appears that atomic hydrogen is *not* a significant contributor to the optical depth at visible wavelengths in the solar photosphere.

So what is providing most of the absorption in the Sun's photosphere? It turns out to be an unusual "atom."

3.4 Atomic Physics: (ii) Opacity due to Negative Hydrogen Ions

The principal absorber in the solar photosphere at visible wavelengths is the *negative hydrogen ion*, i.e., a hydrogen atom with an extra electron attached. The standard hydrogen atom consists of one electron and one proton bound (by means of an attractive central force) in a stable arrangement with an infinity of bound energy levels. But there also exists the possibility that, if free electrons are available in the surrounding medium, an extra electron can be added without the system necessarily being unstable. In essence, the two electrons in an H^- ion arrange themselves (because of Coulomb repulsion) to remain on opposite sides of the proton, as far away from each other as possible. In this situation, the force acting on one of the electrons is no longer central, and no longer purely attractive. As a result, it turns out that there is no longer an infinite set of bound levels. But a bound state does exist. Just one. The bound level lies at an energy of $E(H^-) = -0.754$ eV.

Photons which have the capacity to excite a free-bound transition (removing one of the electrons and leaving a neutral hydrogen atom), have a wavelength $\lambda < \lambda(H^-) = 16{,}450$ Å. In contrast to the sharp edge which occurs in the case of free-bound transitions in the H atom, photoionization of the H^- ion shows a much more gradual wavelength dependence: the cross-section σ_λ rises from zero at 16,450 Å to a maximum at wavelengths around $\lambda_{\max} \approx 8500$ Å, i.e., at energies of about $2E(H^-)$. Detailed calculations of

atomic structure indicate that the maximum cross-section has the numerical value $\sigma_\lambda(\max) = 4.5 \times 10^{-17}$ cm^2. This value is almost an order of magnitude larger than the maximum Lyman continuum cross-section for atomic hydrogen. As a result, the absorption due to H^- is by no means a negligible process. On the redward side of $\lambda_{\max}$, the value of σ_λ falls to one-half of its maximum value at $\lambda \approx 1.3$ μm. On the blueward side, σ_λ falls to $0.5\sigma_\lambda(\max)$ at $\lambda \approx 0.4$ μm.

Another source of continuous opacity due to H^- arises when free electrons pass by a "free" hydrogen atom. This "free-free" process (see Section 3.3.1) contributes opacity which is relatively small at visible wavelengths, but which increases monotonically toward longer wavelengths as λ^2.

The total opacity due to H^- is the sum of the bound-free and free-free processes. The minimum opacity due to H^- occurs at $\lambda \approx 1.6$ μm, where the free-bound process has its "edge."

Thus, across the visible portion of the solar spectrum, from about 4000 Å to about 7000 Å, where most of the solar energy flux emerges, the free-bound opacity due to H^- does not vary by more than a factor of two. There are no large discontinuities in opacity throughout the visible spectrum. As a result, for most of the photons passing through the solar photosphere (i.e., those in the visible spectrum), the opacity is nearly independent of wavelength. Because of this, the *gray* approximation is not too bad when we consider the solar photosphere. This could explain why the gray atmosphere solution ($S(\tau) = a + b\tau$) fits the solar limb darkening ($I(\mu) = a + b\mu$) quite well at visible wavelengths.

We shall find (Section 5.1) that the column density N above the photosphere is about 2×10^{24} cm^{-2}. Most of this column is composed of hydrogen *atoms.* Let the ratio of the abundance of H^- ions to the abundance of H atoms in the photosphere be φ. Then with a cross-section which does not depart significantly from $\sigma_\lambda(\max) = 4.5 \times 10^{-17}$ cm^2, the optical depth above the photosphere due to H^- is $\tau \approx 9 \times 10^7 \varphi$. This can attain values of order unity if the ratio φ is close to 10^{-8}.

Is it plausible that the fractional number of H atoms which will capture a second electron in the solar atmosphere is of order 10^{-8}? To answer that question, we need to know how ionization equilibrium depends on temperature and pressure. For this, we need to consider the Saha equation, which will be the topic of Chapter 4. For now, we note that in the solar photosphere, we shall find (Chapter 4, Section 4.6) that it is entirely consistent with local temperature and pressure that φ should be of order 10^{-8}. As a result, there are indeed enough H^- ions in the solar atmosphere to cause the optical depth at visible wavelengths to be of order unity in the photosphere.

It is essentially the formation of H^- ions which limits how deeply we can observe into the Sun at visible wavelengths. The favorable formation of H^- in the solar photosphere requires the presence of both hydrogen atoms and free electrons. Where do the free electrons in the photosphere come from? Not from the ionization of hydrogen: after all, we are seeking to have H atoms *add*

an electron, not to *lose* an electron. In the photosphere, free electrons are provided mainly by the most abundant "metals" which have low ionization potentials (Mg, Si, S, and Fe). Together, these provide electrons with an abundance of a few times 10^{-4} relative to the abundance of H.

In the lower photosphere, where the gas pressure p_g is found to be of order 10^5 dyn cm^{-2} (see Chapter 5, Section 5.1), the electron pressure is given roughly by the product of p_g and a few times 10^{-4}. This leads to an electron pressure of roughly $\log p_{\rm e} \approx 1.6$. In the upper photosphere, where hydrogen is becoming progressively ionized, the relative abundance of electrons is larger than in the photosphere. As a result, the electron pressure falls off less rapidly with height than the gas pressure. At a height where p_g has fallen off by a factor of (say) 100 relative to the photosphere, $\log p_{\rm e} \approx 0$. We shall use these below when we apply the Saha equation to estimating the fractional abundance of H^- in the photosphere.

3.5 Atomic Physics: (iii) Opacity due to Helium Atoms and Ions

The spectrum of HeII is analogous to that of HI, except that all wavelengths are reduced by a factor of four. Thus, the HeII-Lyman edge lies at $\lambda = 228$ Å.

For HeI, with an ionization potential of 24.6 eV, the edge at which absorption is maximum lies at $\lambda = 504$ Å. The maximum absorption cross-section at the HeI edge is 8×10^{-18} cm^2, comparable to the maximum cross-section at the Lyman edge of H.

3.6 Astrophysics: The Rosseland Mean Opacity

Once we know how opacity varies as a function of wavelength (κ_λ), or as a function of frequency (κ_ν), a mean opacity can in principle be obtained simply by taking an arithmetic average over all wavelengths or frequencies. But that would not be especially useful in our attempts to determine how photons make their way through an atmosphere. The derivation of a relevant "mean opacity" should also incorporate somehow the shape of the spectrum of radiation.

The most common method for calculating a mean opacity which is relevant to the passage of radiation in the deep interior of a star is called the Rosseland mean $\kappa_{\rm R}$, defined by:

$$\frac{1}{\kappa_{\rm R}} \int \frac{dB_\nu}{dT} d\nu = \int \frac{1}{\kappa_\nu} \frac{dB_\nu}{dT} d\nu \tag{3.3}$$

In deriving Equation 3.3, it is assumed that the shape of the spectrum is related to the local Planck function B_ν. The appearance in Equation 3.3 of (a) $1/\kappa_\nu$ and (b) the first derivative of B_ν with respect to temperature can be traced ultimately to the fact that the RTE provides an expression for $1/\kappa_\nu$ times the first spatial derivative of the intensity (dI_ν/dx): in a given atmosphere, the spatial gradient can be converted to the derivative with respect to temperature.

The Rosseland mean as defined in Equation 3.3 is a "transparency mean": as far as opacity is concerned, Equation 3.3 gives maximal weight to regions in the spectrum where the opacity is smallest. In the atmosphere of the Sun, photons will tend to "leak out" through such regions. Also, Equation 3.3 weighs more heavily those parts of the spectrum where dB_ν/dT is larger, i.e., at wavelengths somewhat shorter than the Wien maximum.

The units of $\kappa_{\rm R}$ are cm^2 gm^{-1}.

For a medium which contains a certain mixture of elements, the Rosseland mean is calculated by first determining the frequency-dependence of the opacity due to each atomic species in the mixture. At each frequency, the total opacity is obtained by summing the contributions before performing the integral in Equation 3.3. For absorption from the respective ground-states of the various types of atoms in the mixture, there is no dependence on temperature. But absorption due to excited states introduces significant temperature dependence at longer wavelengths. Further temperature dependence enters as a result of the Planck function which enters in Equation 3.3. As a result, it is not surprising that $\kappa_{\rm R}$ varies significantly with temperature.

To illustrate how $\kappa_{\rm R}$ depends on temperature, we show in Figure 3.3 the results for a gas consisting of H and He (with mass fractions $X = 0.7$ for H and $Y = 0.28$ for He) plus "metals" (with mass fraction $z = 0.02$). The range of temperatures along the abscissa extends from $T = 1000$ K ($\log T = 3$) to temperatures ($\log T = 8$) in excess of those at the center of the Sun. Each curve is labeled with a value of $\log(R)$ where the parameter $R = \rho/T_6^3$ is a combination of density ρ and T_6, the temperature in units of 10^6 K.

How can we understand the behavior of $\kappa_{\rm R}$ which appears in Figure 3.3? Let us examine certain limiting behaviors. Two aspects of the figure stand out: at the highest temperatures, all curves converge to a single finite value, and at the lowest temperatures, all curves plunge steeply toward zero opacity.

3.6.1 Limit of low ρ and/or high T: electron scattering

In the limit of very low density, e.g., when the value of $\log(R)$ has its smallest value ($= -6$ in Figure 3.3), and also in the limit of the highest temperatures, H and He are essentially completely ionized. There are essentially no bound states available for electrons to occupy: all electrons are free particles, and electron scattering is the sole source of photon interaction. With $\sigma_{\rm e} = 6.6 \times 10^{-25}$ cm^2 per electron, and a mean molecular weight of 1.3

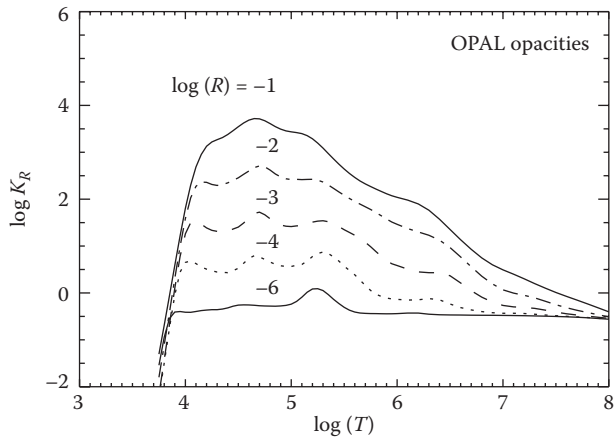


FIGURE 3.3: Rosseland mean opacity ($\kappa_{\rm R}$, in units of cm^2 gm^{-1}) as a function of temperature T (deg K) for a number of densities. The parameter on each curve is the quantity $R = \rho/T_6^3$, where ρ is density in units of gm cm^{-3}, and T_6 is temperature in millions of deg K. Data are taken from a table of OPAL opacities for near-solar composition (hydrogen mass fraction $X = 0.7$, helium mass fraction $Y = 0.28$, metals mass fraction $Z = 0.02$). (More extensive tables of OPAL opacities are publicly available at the Lawrence Livermore website: ftp://www-phys.llnl.gov/pub/opal/type1data/GN93/ascii/GN93hz)

(corresponding to some 5×10^{23} electrons per gram), we expect that in conditions of complete ionization, $\kappa_{\rm R} \approx 0.3$ cm^2 gm^{-1}. Thus, the ordinate in Figure 3.3, $\log(\kappa_{\rm R}) = -0.5$ when ionization is complete. Indeed, the curve labeled -6 in Figure 3.3 lies close to this value at all temperatures except for the very coolest (less than 6000 K).

3.6.2 Low T limit

As T falls below 10^4 K, all curves in Figure 3.3 fall steeply to very small values. The reason for this behavior is readily understood: as $T \to 0$, all electrons collect in the ground states of HI and HeI. The only photons which can be absorbed effectively require wavelengths shorter than 912 Å and 228 Å, respectively. But at temperatures below 10^4 K, the Wien maximum lies at $\lambda_{\rm max} > 2880$ Å, and the flux of photons at $\lambda \leq 912$ Å is exponentially small. The lower the temperature, the more drastic is the exponential reduction in photon flux at $\lambda \leq 912$ Å. In conditions which apply to the upper photosphere, i.e., $T = 4900$ K, the Rosseland mean opacity falls to values of order 0.001 cm^2 gm^{-1}.

3.6.3 Higher ρ: free-bound absorptions

At a fixed temperature, increasing density in Figure 3.3 corresponds to an increase in $\log(R)$. At the highest temperatures ($\log(T) > 7.5$), density effects are minimal: at such high temperatures ionization is almost complete at all densities which are relevant to the Sun. As a result, electron scattering dominates the opacity. However, at intermediate temperatures, increasing the density leads to increased recombination rates in the gas. In such conditions, the number of atoms with populated bound states grows, and this causes significant increases in opacity, especially at the various bound-free "edges."

For the lowest density curve in Figure 3.3 (labeled -6), the most significant departure from electron scattering (i.e., departure from $\log(\kappa_{\rm R}) = -0.5$) is the "bump" at temperatures $\log(T)$ in the range 5.0–5.5. In this range, Wien's law indicates that the peak of the Planck function lies at wavelengths of 100–300 Å. It is noteworthy that this range of wavelengths overlaps with the HeII–Lyman edge at 228 Å. The overlap of the peak in the spectrum with the largest peak in atomic opacity at the Lyman edge is conducive to creating enhanced opacity. Of course, the ionization of HeII is almost complete due to the low densities, so there are relatively few HeII ions available to contribute their Lyman-edge absorptions: this explains why the "bump" reaches maximum opacities ($\log(\kappa_{\rm R}) = 0$) which are larger than electron scattering opacity, but not by orders of magnitude. However, when we examine higher densities, e.g., on the curves labeled -4 and -3 in Figure 3.3, larger numbers of HeII ions survive at $\log(T) = 5.0$–5.5, and the "bumps" in opacity grow to larger amplitudes.

There is a second "bump" on the curve labeled -6 in the range $\log(T) =$ 4.5–4.75: at such temperatures, Wien's law predicts a peak in the Planck function around $\lambda \approx 500$–900 Å. Such photons have energies which approach the energies required to cause bound-free absorptions from HeI. This peak also becomes more prominent at higher densities, where the number of HeI atoms per gram (at a given temperature) increases significantly.

A third "bump" in $\log(\kappa_{\rm R})$, most prominent on the curves in Figure 3.3 labeled -4 and -3, occurs at $\log(T) \approx 4.0$. At such temperatures, excited states in HI are rapidly being populated: the populations grow exponentially with increasing T. Each of these states contributes absorption due to its free-bound "edges." In particular the Balmer "edge" overlaps with the peak of the black-body function at $\log(T) \approx 4.0$. The exponential growth in bound populations causes $\kappa_{\rm R}$ to increase rapidly with increasing T. However, as $\log(T) \rightarrow 4.0$, hydrogen is also ionizing rapidly. The competition between excitation and ionization leads to a rather narrow peak in the opacity curves in Figure 3.3.

3.6.4 Magnitude of the largest opacity

As density increases, the ranges of temperatures at which significant ionizations of HI, HeI, and HeII occur begin to overlap more and more. The various

relatively narrow individual "bumps" in Figure 3.3 which are apparent at relatively low densities tend to merge into a single broad "bump," although "shoulders" are still apparent on both sides of the broad "bump." The broad "bump" lies at temperatures of $\log(T) = 4.5 - 5$.

For the density range which enters into the data shown in Figure 3.3, the largest values of opacity have numerical values in the range $\log(\kappa_{\rm R}) = 3 - 4$.

What is the maximum density for the results in Figure 3.3? At a temperature of $\log(T) = 4.75$ (i.e., $T_6 = 0.06$), the maximum density, corresponding to the curve $\log(R) = -1$, is of order 2×10^{-5} gm cm^{-3}.

Other investigations (e.g., Ezer and Cameron, 1963) have extended the calculation of Rosseland mean opacity to higher densities. With $\rho = 10^{-3}$ gm cm^{-3}, the peak in opacity at $\log(T_{\rm max}) = 4.5 - 4.75$ has a value $\log(\kappa_{\rm R}) \approx 5$. If we attempt a rough extrapolation of the results of Ezer and Cameron, it appears that the peak numerical value of $\kappa_{\rm R}$ might be as large as 10^6 cm^2 gm^{-1}. Recalling that the (wavelength-dependent) opacity at the Lyman edge has the numerical value $\kappa_\lambda(\max) \approx 3.6 \times 10^6$ cm^2 gm^{-1}, it seems unlikely that the maximum permissible value of $\kappa_{\rm R}$, could be much larger than κ_λ (max). If a black-body curve were to be matched optimally in wavelength such that its peak overlapped with the Lyman edge, then we might expect to see $\kappa_{\rm R}$ values approaching κ_λ (max): the temperature required for such matching would be such that $\lambda_{\rm max} = 0.288({\rm cm})/T_{\rm max} = 912$ Å. This leads to $\log(T_{\rm max}) = 4.5$. This is consistent with the peak in Figure 3.3 above.

The results in Figure 3.3 refer to a gas where H and He are the principal constituents. There are a few "metals" included in the calculations, and these alter the H/He opacity curves slightly, giving rise to some "bumps" appearing at $\log(T) = 5 - 6$. The reason that the changes are only slight has to do with the relatively small abundances of the "metals" compared to H and He: even the most abundant "metals" have fractional abundances of no more than 0.001 times H (by number).

3.7 Power-Law Approximations to the Rosseland Mean Opacity

For future reference, we note that it is at times convenient to fit the Rosseland mean opacity to power laws of temperature and density. This allows analytic solutions to be extracted for certain problems. Different power laws apply in different parameter regimes.

In the earliest attempts to model the deep interior of the Sun, at temperatures in excess of roughly 10^6 K, the Kramers opacity "law" was developed. In this temperature range, the dominant constituents of the Sun are almost completely ionized. The fractional abundances of incompletely ionized atoms are becoming rapidly smaller as the temperature increases. As a result, the

strongest contributors to opacity (bound electrons) are becoming progressively scarcer in the gas. This leads to a rapid *decrease* in opacity as the temperature *increases*. Valid for both bound-free and free-free transitions, certain approximations lead to the functional form $\kappa = \kappa_o \rho / T^{3.5}$. We shall find this useful below in Chapter 8 (Section 8.3) when we model the radiative interior of the Sun.

In a different limit, at temperatures below 10^4 K, the temperature sensitivity is very different. Starting at the lowest temperatures, at say a few thousand degrees, essentially all H and He atoms are in their ground states, while most photons in the local black-body function are at long wavelengths. As a result, the opacity falls off to very low levels. However, increasing the temperature has two effects: (i) it populates excited states of H and He, creating opportunities for longer wavelength photons to be absorbed; (ii) the peak in the black-body spectrum moves toward shorter wavelengths. Both effects combine to cause opacity to *increase* rapidly as the temperature *increases*. The opacity also increases as the density increases, although the sensitivity to density is much less pronounced than the temperature-sensitivity. A power law fit to opacities in this temperature regime suggests that $\kappa = \kappa_1 \rho^a T^b$ could provide a reasonable zeroth order fit. In order to determine what the power law indices a and b are, we refer to a specific table of opacities which has been calculated by Kurucz (1992). Some of the opacities from that table will be used later (in Chapter 5) to calculate a model of the photosphere. For present purposes, we note that Kurucz lists log (Rosseland mean opacity) for a series of log(temp) and log(press). (A portion of this table appears as Table 5.1.). In Kurucz's complete table, at each value of $\log(T)$ and $\log(p)$, the local density is also listed. Using the values of $\log(\kappa)$ tabulated by Kurucz at all values of $\log(T) \leq 4.0$, we have obtained least squares fits for the coefficients in the relationship $\log(\kappa) = c + a \log(\rho) + b \log(T)$. We find $a = 0.343$, $b = 9.0583$, and $c = -31.97$. The steep dependence on temperature is noteworthy. We shall find this formulation of the opacity useful in Chapter 15 when we model the rise in temperature between photosphere and chromosphere.

3.8 Narrow Band Opacity: Absorption Lines in the Spectrum

So far, the opacity we have discussed occurs mainly over broad regions of the spectrum. For example, the negative hydrogen ion contributes significant opacity at wavelengths from as short as 4000 Å to as long as 1–2 μm. And bound-free absorption from the $n = 2$ level of hydrogen contributes opacity at all wavelengths shortward of 3648 Å. These are truly "broad-band" sources of opacity, and they help to determine, in conjunction with the atmospheric

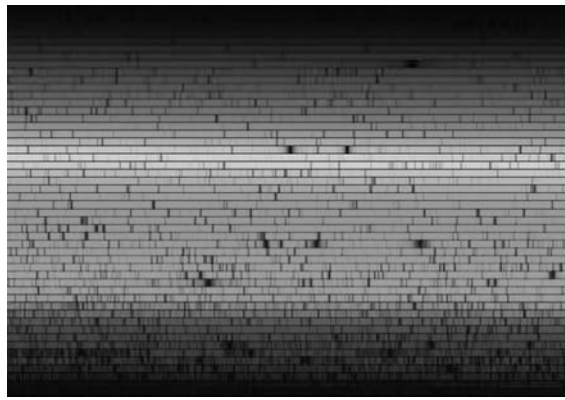


FIGURE 3.4: **(See color insert following page 202.)** Fraunhofer lines: dark narrow absorption features in the solar spectrum. (Downloaded from http://www.noao.edu/image_gallery/html/im0600.html Credit: National Optical Astronomy Observatory/Association of Universities for Research in Astronomy/National Science Foundation. With permission.)

temperature profile, the overall "rainbow" of the solar spectrum, extending from peak intensity in the yellow-green region toward the red and toward the violet, where the intensity gradually fades from sight.

However, when one is presented with a spectrum of the Sun (see Figure 3.4), the first thing that attracts one's attention is the presence of a multitude of more-or-less narrow dark "lines" distributed across the entire range of visible wavelengths. These are the features that Fraunhofer first discovered in 1814 when he fed sunlight into the entrance slit of a spectroscope. Fraunhofer drew up a list of the strongest lines that he could discern, and labeled them with a series of upper case letters from A to K (in order of decreasing wavelength). A list of weaker lines was subsequently labeled with lower case letters, starting with a in the red and moving toward shorter wavelengths. Although Fraunhofer did not identify the origin of his lines, many of his labels persist in common use to this day: e.g., Fraunhofer's D line near 5900 Å is now known to be a close doublet, and is referred to as the D_1 and D_2 lines of neutral sodium. The K line at 3934 Å, the strongest line in the solar spectrum, is now known to arise from ionized calcium: observation of the Sun in the Ca K line is a common practice in modern solar physics. It is now known that the features labeled A, B, and a by Fraunhofer (at wavelengths near 7590, 6870, and 6280 Å, respectively) have nothing to do with the Sun: they are caused by oxygen molecules in the Earth's atmosphere.

The occurrence of the solar "Fraunhofer lines," so striking in their darkness against the backdrop of the rainbow, indicates the presence in the solar atmosphere of atoms and ions of particular elements in particular stages of excitation.

3.8.1 Characterizing the properties of absorption lines

It is helpful to consider quantitative measures of absorption lines. To do this, we examine the "line profile," i.e., how does the radiant intensity (or flux) vary as a function of wavelength?

In each line, when one plots the radiation flux F_λ as a function of wavelength (see Figure 3.5), one starts far from line center on (say) the blueward side, with an intensity essentially equal to the continuum F_c. For all lines, if one chooses a wavelength that is far enough from line center, the ratio F_λ/F_c approaches unity (apart from some "noise" associated with other spectral lines). As wavelength increases, and one enters into the line, the intensity decreases more or less rapidly: this decrease gives rise to what are called the "wings" of the line (Figure 3.5). At a certain wavelength, the radiation flux reaches a minimum value. This location, where the depth of the line is maximum, is defined to be line center. The depth of line center varies from line to line: some lines are so weak that they dip to no more than a percent or two below the continuum: such weak lines, with central intensities of 0.98–0.99 of the local continuum, can be difficult to identify against the continuum. At the other extreme, the strongest lines have depths in excess of 90% . In the center of such deep lines, the residual intensity may amount to only a few percent of the continuum.

The strength of each line can be characterized in an empirical sense by the area that the line subtracts from the continuum radiation which would pass

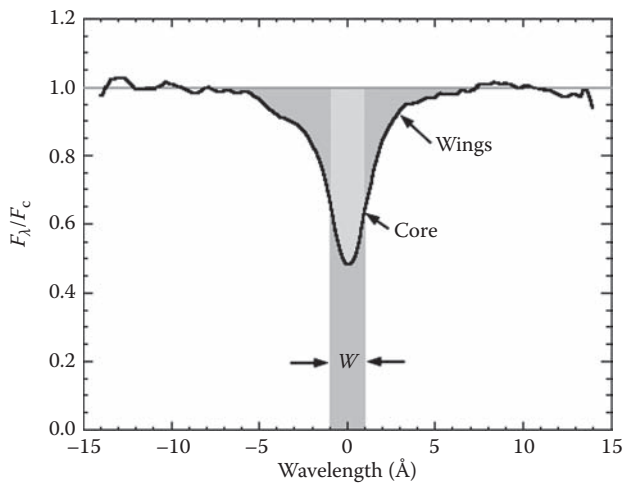


FIGURE 3.5: **(See color insert following page 202.)** The shape of an absorption line in the spectrum of the Sun. The rectangle with width W has an area equal to the total area absorbed by the line. W is referred to as the equivalent width. (From http://web.njit.edu/∼gary/321/equiv_width.gif. With permission.)

through the local wavelength region if the line were absent. This area of absorption is known as the "equivalent width" W. The equivalent width of a line, when expressed in wavelength units, is the width of an equivalent rectangular shape, extending from the continuum to an intensity level of zero, which would have the same area as the area absorbed by the line (see Figure 3.5). The weakest lines that the eye can detect in a solar spectrum of good quality have EW of a few milliangstrom. Lines which attract the most attention when one takes a quick look at the spectrum, such as a closely spaced pair of lines due to sodium in the yellow part of the spectrum (at wavelengths of 5890 and 5896 Å), have EW of order 1 Å. The strongest lines in the visible spectrum of the Sun are the H and K resonance lines of Ca II: the EW of the K line is some 20 Å.

The EW of a weak spectral line increases in proportion to the total number of atoms/ions along the line of sight which contain electrons in the lower state of the transition. This gives rise to a linear slope in what is called the "curve of growth": this curve illustrates how the EW increases with increasing column density.

The EW is a measure of the *area* of a line. But there is also physically valuable information contained in the actual *shape* of the line: this shape is determined by a number of physical factors. For the weakest lines, random motions cause the absorption to be spread out in wavelength as a result of the Doppler effect. In such lines, where the width is determined by "velocity broadening," the line shape is Gaussian if the velocities of the atoms obey a Maxwellian law. As a result, the absorption is maximum at line center, where the wavelength has the value λ_o, and the absorption falls off along a Gaussian curve to smaller values as one examines wavelengths that are farther and farther from line center, on both sides of line center. This gives rise to a characteristic shape of the "core" of the line (see Figure 3.5), with a characteristic half-width of $\Delta\lambda_o$. Thermal motions, with mean speeds V_t of about $\sqrt{(2kT/m_a)}$, are certainly a contributor to $\Delta\lambda_o$. If thermal motions are the only random motions which are present in the absorbing gas, the value of $\Delta\lambda_o(\text{th})$ is determined entirely by the mean thermal speed of the atoms/ions which create the line: according to the Doppler formula:

$$\frac{\Delta\lambda_o(\text{th})}{\lambda_o} = \frac{V_t}{c} \tag{3.4}$$

where c is the speed of light. In order to fully resolve the profile of an absorption line which is broadened only by thermal broadening, one needs to use a spectrometer with a resolving power R_p which is at least as large as $\lambda_o/\Delta\lambda$ (th).

3.8.2 Line-broadening: more than merely thermal motions

However, thermal motions are *not* the only contributor to the widths of lines in the solar spectrum. To see that this is the case, note that if thermal

motions were the only contributors, then lines from massive atoms/ions would be much narrower than the lines from light atoms/ions. For example, lines due to nitrogen ($m_a = 14$) should have $\Delta\lambda_o$ which is twice as broad as lines due to iron ($m_a = 56$). And lines from even more massive atoms should become increasingly narrow as the atomic weight increases. But observations of lines in the solar spectrum do not behave in this way: instead, there is a minimum width which is observed in lines from even the heaviest atoms. That minimum width, when interpreted in terms of velocity by means of the Doppler formula, is found to correspond to a speed of 1–2 km sec^{-1}.

This is a very important piece of information about the gas in the solar atmosphere. Apparently, the gas in the solar photosphere is subject to motions of 1–2 km sec^{-1} over and above the purely thermal motions. The generic term for such nonthermal motions is "microturbulence."

What could the microturbulence in solar spectral lines be due to? In part, it may arise because of the effects of convection in the lower photosphere of the Sun (see Chapters 5 through 7), with upflows overshooting to some extent into the layers of gas where the absorption lines are formed. Also, the microturbulence may be caused by waves of some kind propagating upward and downward through the photosphere, or even standing waves which arise when waves are trapped in a cavity of some kind. (We shall see in Chapters 13 and 14 that sound waves of both kinds are present in the Sun as a whole.)

For stronger lines, there may be so many atoms along the line of sight that some atoms cast a shadow on others, preventing the latter from absorbing as much light as one might have expected. In such lines, the optical depth for photons at line center τ_o exceeds unity by factors which may be large. In order for photons to escape from such shadows, the photons undergo a scattering process which leads to a slow increase in line width. The process by which this increase in line width occurs (called "opacity broadening") is quite different from velocity broadening. In the presence of opacity broadening, a Gaussian profile may still be present, but the half-width is larger than the thermal value $\Delta\lambda_o(\text{th})$ by an amount which scales as $\sqrt{(\ln\tau_o)}$. For a line in which τ_o is (say) equal to 10^4, the line profile has a width of roughly $3\Delta\lambda_o(\text{th})$. In the presence of such opacity broadening, the EW no longer increases in proportion to the number of available atoms along the line of sight: instead, the EW increases more slowly than a linear proportionality. This slow increase is referred to as "the flat part of the curve of growth." In such lines, the presence of microturbulence helps to broaden the line in such a way that the shadowing effect is less severe: as a result, larger microturbulence leads to larger EW.

The strongest lines in the solar spectrum have broad wings: these are not at all Gaussian in shape, but are much more widely spread out. Various physical processes can give rise to such broad wings. In the case of hydrogen lines, free electrons in the solar atmosphere create fluctuating electric fields as they pass by a particular hydrogen atom. As a result of the Stark effect, the energy levels of the hydrogen atom are perturbed, and the energy of the transition is shifted

to either higher or lower energies. This gives rise to absorption at wavelengths which may be relatively far removed from λ_o. In the case of strong lines such as the Ca K line, the probability of spontaneous transition within the ion is so large that an electron in the upper energy level survives for only a very short interval of time. This leads to "radiation broadening" of the transition.

One final cause of line broadening in the Sun must be mentioned: magnetic fields. In the presence of such fields, the energy levels within an atom or ion become split into several distinct sublevels. Only certain transitions between the sublevels are permitted. As a result, a line which originally was a single narrow feature in the spectrum can break up into a group of distinct, but closely spaced, components. The amount of spacing is a measure of the strength of the field (see Chapter 16, Section 16.4.1). The polarization of the various components contains information as to the direction of the field lines relative to the line of sight.

In summary, absorption lines in the spectrum of the Sun provide important information about abundances of elements, about velocity fields, and about magnetic fields in the solar atmosphere.

Exercise

3.1 Lines due to hydrogen, sodium, calcium, and iron occur in the solar spectrum at wavelengths of 6563 Å, 5890 Å, 3933 Å, and 5250 Å, respectively. Assuming that each line arises in a gas with $T = 6000$ K, calculate the thermal width in Å for each line. What resolving power is required to resolve the thermally broadened line profile in each case?

References

Ezer, D. and Cameron, A. G. W. 1963. "The early evolution of the sun," *Icarus*, 1, 422.

Kurucz, R. 1992. Opacity tables made available to the author upon request.

Stenflo, J. O. 2005. "Polarization of the Sun's continuous spectrum," *Astron. Astrophys.*, 429, 713.

Chapter 4

Toward a Model of the Sun: Ionization

The properties of opacity at high temperature (and low pressure) are controlled by the physical process of ionization. Other physical properties of the gas, including thermodynamic quantities such as the specific heats (which are important for transport of energy), are also significantly affected by the ionization process. In order to have a clear understanding of the physics of certain regions in the Sun, it is important to have a quantitative model of ionization. This leads us to consider an equation originally derived by Saha (1921) for ionization equilibrium.

We have already mentioned how populations of atoms are distributed among energy levels when the latter are *bound*: the number density n_i of atoms with electrons in the i^{th} energy level is related to the number density of atoms in the ground state n_1 according to a Boltzmann distribution:

$$\frac{n_i}{n_1} = \left(\frac{g_i}{g_1}\right) \exp\left(\frac{-\Delta E_{i1}}{kT}\right) \tag{4.1}$$

where ΔE_{i1} is the difference in energy between level i and the ground state. The statistical weight g_i has the value $2i^2$ for the ith level in a hydrogenic atom. This value of g_i is derived by noting that for principal quantum number i, the angular momentum sublevels take on integer quantum numbers L from zero up to $i-1$: each of these sublevels has orbital multiplicity $2L+1$, as well as multiplicity 2 for electron spin. Summing the combined multiplicity $4L+2$ per sublevel over L sublevels from $L=0$ to $L=i-1$ yields a total of $2i^2$.

4.1 Statistical Weights of Free Electrons

Now, when we wish to consider the case of ionization, we need to write down an expression for the population of *unbound ions and electrons.* Analogously to the Boltzmann formula for bound states, the Boltzmann distribution for ions and electrons is:

$$\frac{n_{\text{i}}}{n_{\text{a}}} = \frac{g_{i+\text{e}}}{g_{\text{a}}} \exp\left(-\frac{I}{kT}\right) \tag{4.2}$$

Here, $n_{\rm i}$ and $n_{\rm a}$ are number densities of ions and atoms respectively, and I is the "ionization potential," i.e., the energy required to ionize an atom from the ground state. The statistical weight of the atom $g_{\rm a}$ is essentially that of the ground state, i.e., two for a hydrogenic atom.

The principal difference is that the statistical weight of the ionized system, consisting of ion plus electron, must include not only the statistical weight of the ion g_i (also mainly in its ground state), but must also include the statistical weight $g_{\rm e}$ of the free electron. The overall statistical weight for ion and electron is the product of two terms: $g_{i+{\rm e}} = g_i g_{\rm e}$. In contrast to an electron which occupies a bound energy level, where the number of available sublevels is small (leading therefore to small statistical weight), a free electron has access to an enormous number of states. This has the effect that, in Equation 4.2, the right-hand side of Equation 4.2 may grow to values of order unity even when the exponential term is small, i.e., even when kT is much smaller than I.

For example, consider the case of hydrogen, where $I = 13.6$ eV. What do we have to do in order to achieve significant ionization, i.e., in order to cause $n_i \approx n_a$? The answer is; we need to make the right-hand side of Equation 4.2 of order unity. At first, let us suppose that the statistical weight factors are equal to unity, i.e., suppose that the only relevant term on the right-hand side of Equation 4.2 is the exponential factor. In such a case, if we are to achieve $n_i \approx n_a$, the temperature would have to satisfy the equation $kT \approx 13.6$ eV. This corresponds to $T \approx 158000$ K. On the other hand, when statistical weights are included (as of course they must be), it is possible to achieve $n_i \approx n_a$ even when T is less than 10000 K (at low pressure).

In order to evaluate $g_{\rm e}$, we need to know the number of states available to a free electron. Such an electron moves in a 6-dimensional (6-D) $p - r$ phase space, where p is 3-D momentum, and r is 3-D coordinate space. The uncertainty principle restricts the uncertainties in 1-D momentum and 1-D position such that $dp_x dr_x$ cannot be less than a quantity of order h, the Planck constant. This leads to the concept of minimally occupiable "cells" in phase space, each with volume $d^3p d^3x \approx h^3$. Allowing for electron spin, each "cell" in phase space cannot be occupied by more than two electrons. As a result, in a gas which occupies a volume $V(r)$ of coordinate space and a volume $V(p)$ in momentum space, the total number of states which are available to the free electrons is $g_{\rm e} = 2V(p)V(r)/h^3$.

What are we to use for $V(r)$ and $V(p)$? In a medium where the number density of electrons is $n_{\rm e}$, the mean volume of coordinate space $V(r)$ which is occupied by a single electron is $V(r) = 1/n_{\rm e}$. Each such electron has access to a large volume of momentum space. In momentum space, few electrons have momenta faster than the mean thermal speed: $V_{th} = \sqrt{(8kT/\pi m)}$. As a result, momentum space is filled up out to a radius of order $p_{th} = mV_{th}$. The associated volume $V(p)$ in momentum space is $(4/3)\pi p_{\rm th}^3$.

Combining $V(r)$ and $V(p)$ we find

$$g_{\rm e} = \frac{8\pi}{3}\left(\frac{8km}{\pi}\right)^{3/2}\frac{1}{h^3}\frac{T^{3/2}}{n_{\rm e}} \tag{4.3}$$

It is usual to rewrite this relation in terms of electron *pressure* rather than in terms of electron *number density*. Replacing n_e with p_e/kT in Equation 4.3, Equation 4.2 becomes

$$\frac{n_\mathrm{i} p_\mathrm{e}}{n_\mathrm{a}} = \frac{g_i}{g_\mathrm{a}} C_i T^{2.5} \exp\left(\frac{-I}{kT}\right) \tag{4.4}$$

where C_i is a combination of numerical and physical constants. The physical constants in C_i appear in the combination $m_\mathrm{e}^{1.5} k^{2.5}/h^3$: in c.g.s. units, the numerical value is 0.021. Including the numerical constants, the value of C_i turns out to be (coincidentally) close to unity (0.72).

4.2 Saha Equation

In logarithmic form, we can write Equation 4.4 as

$$\log\left(\frac{n_\mathrm{i}}{n_\mathrm{a}}\right) = -\log p_\mathrm{e} + 2.5 \log T - \theta I + \log\left(\frac{g_i}{g_\mathrm{a}}\right) - 0.14 \tag{4.5}$$

where I is the ionization potential in units of eV, $\theta = 5040/T$, and the logarithms are to base 10. As an aid to memory, we note that in cases where $g_i \approx g_\mathrm{a}$, no significant error is made if we make the approximation of retaining only the first three terms on the right-hand side of Equation 4.2. In what follows, we adopt this approximation.

Equation 4.5 is the Saha equation: it allows us to evaluate the degree of ionization in a medium of given T and p_e.

It is worth reiterating why considerations of ionization are important as far as opacity is concerned. If a gas is highly ionized, the lack of bound states for electrons leads to low values of opacity. Because a bound electron may be up to seven orders of magnitude more efficient than a free electron at interacting with photons, the presence of even a few bound electrons can cause the opacity to be enhanced significantly.

4.3 Application of the Saha Equation to Hydrogen in the Sun

There are two distinct locations in the Sun where hydrogen makes a transition from mostly neutral to significant ionization. Since the photosphere is the location in the Sun where neutral hydrogen is most abundant, it is no

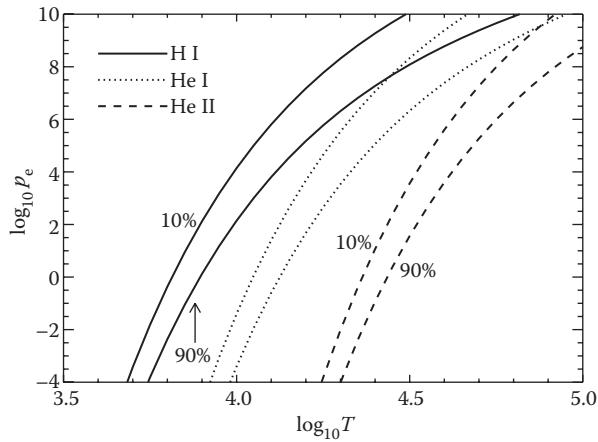


FIGURE 4.1: Ionization strips of neutral hydrogen HI, neutral helium HeI, and singly ionized helium HeII in the ($\log p_e$–$\log T$) plane. Units of T are K, and units of p_e are dyn cm^{-2}. The solid curves labeled 10 and 90% indicate the loci along which hydrogen is 10% ionized and 90% ionized, respectively. The dashed curves labeled 10 and 90% indicate the loci along which HeII is 10% ionized and 90% ionized, respectively. The dotted lines refer to HeI, but the 10 and 90% labels are omitted from the upper and lower lines (respectively) for the sake of clarity.

surprise to find that the two distinct locations lie on either side of the photosphere. One lies above the photosphere, in the low-density gas of the upper chromosphere. The second lies well below the surface, in the denser gas of the convection zone. Let us use the Saha equation to determine the temperatures of these locations.

In order to discuss the Saha equation, it is convenient to plot the Saha equation in the ($p_e - T$) plane, and to introduce the concept of an "ionization strip," as shown in Figure 4.1. Along the curves labeled P% in Figure 4.1, H and He are P percent ionized. To avoid crowding in the figure, we plot only two curves for each element: $P = 10\%$ (essentially the onset of ionization), and $P = 90\%$ (essentially the near completion of ionization). The area between the 10 and 90% ionization levels can be considered an "ionization strip," where either H or He is in the process of transitioning from mostly neutral to mostly ionized. Within each ionization strip, where heat input tends to increase the ionization rather than increase the temperature, the specific heat of the element in question becomes much larger (by an order of magnitude or more) than the standard value from kinetic theory (see Chapters 6 and 7).

To be specific, let us quantify the onset of "significant" ionization as the location where $n_i \approx n_a$, corresponding to 50% ionization. According to the

simplified version of Equation 4.5, this occurs for hydrogen when the temperature satisfies the equation

$$\theta I - 2.5 \log_{10} T \approx - \log_{10} p_e \tag{4.6}$$

where we must use $I = 13.6$ for hydrogen.

Let us consider the upper chromosphere, where the pressure is relatively low: $\log p_e \approx 0$ (see Chapter 15). In this case, Equation 4.6 reduces to $\theta = (2.5/13.6)\ \log\ T$, i.e., $T\ \log\ T = 27400$. To solve this, we can guess at the slowly varying logarithm term and then iterate. For example, if we guess $\log\ T = 4$ (i.e., $T = 10^4$ K), then the first iteration at the solution would be $T = 27{,}400/4 = 6850$, too low to be consistent with $\log\ T = 4$. If we guess $\log\ T = 3.7$ (i.e., $T = 5000$ K), then the solution would be $T = 7400$ K, too high to be consistent with the initial guess. An iterative solution to this equation yields $T = 7100$–7200 K. Thus, when the electron pressure is as low as it is in the upper chromosphere, hydrogen begins to ionize significantly when the temperature rises above 7100–7200 K.

Now let us consider the gas below the photosphere. Photospheric models (see Chapter 5) indicate that the gas pressure in the photosphere is $\log p_g \approx 5$, and it rises as we move deeper below the photosphere. In the photosphere, electron number densities are less than atom number densities by about 2000. (Reasons for this number will be discussed below.) As a result, $\log\ p_e \approx 1.7$ in the photosphere. But with increasing depth, and increasing temperature, increasing ionization of hydrogen causes the value of p_e to approach closer to the value of p_g. At a depth where $\log p_g \approx 7$, the electron pressure is within an order of magnitude of p_g. In this case, the 50% ionization point occurs when

$$\theta \approx 0.18 \log_{10} T - 0.53 \tag{4.7}$$

Iterative solution of this equation yields $T = 20{,}000$–$21{,}000$ K.

The contrast between the ease of ionization in the chromosphere and in the subphotosphere is apparent. In the low pressure conditions of the chromosphere, it is relatively easy to ionize hydrogen: a temperature of just above 7000 K will suffice. But in the high pressure gas below the surface, the ionization of hydrogen is "postponed" to higher temperatures: 50% ionization of hydrogen does not occur until the temperature has risen above 20,000 K.

As a final application of the Saha equation to hydrogen in the Sun, let us check the fractional ionization of hydrogen in the photosphere. In the lower photosphere, inserting $\log p_e = 1.7$, $T = 6000$ K (i.e., $\theta = 0.84$), we find $\log(n_i/n_a) = -3.7$. In the upper photosphere, inserting $\log p_e = 0$, $T = 4900$ K (i.e., $\theta = 1.03$), we find $\log(n_i/n_a) = -4.8$. Thus, the average degree of hydrogen ionization in the solar photosphere is about $10^{-4.25}$: only one H in some 20,000 is ionized. In contrast, all elements in the solar gas which have ionization potentials of about 9 eV and less are essentially completely ionized. This includes (in order of decreasing abundances) Si, Mg, Fe, Al, Ca, and Na. Using the standard abundances of each of these elements, and summing over

them, we find that their ionization provides about 10^{-4} electrons for every hydrogen atom. This clearly exceeds the average degree of ionization of hydrogen in the upper photosphere: therefore, the "metals" are the primary source of free electrons in the upper photosphere. In the lower photosphere, the supply of free electrons comes from metals and hydrogen in comparable amounts.

4.4 Application of the Saha Equation to Helium in the Sun

The double ionization of helium requires $I = 54$ eV. This means that the equation for 50% ionization is

$$54\theta - 2.5 \log_{10} T \approx -\log_{10} p_{\rm e} \tag{4.8}$$

Solutions to this equation occur at higher temperatures than those for hydrogen. In the chromosphere ($\log p_{\rm e} \approx 0$), the solution satisfies $T \log T = 1.1 \times 10^5$: this corresponds to $T \approx 25{,}000$ K.

In the case of the subphotosphere, we must go to deeper layers than those where hydrogen ionization reaches the 50% level. In the deeper layers, $\log p_{\rm e}$ may be as high as 10–12. This leads to

$$\theta \approx 0.046 \log_{10} T - 0.2 \tag{4.9}$$

The solution of this is $T \approx 140{,}000$ K.

Again, the contrast between the ease of ionization in the chromosphere and in the subphotosphere is apparent. In the low pressure conditions of the chromosphere, it is relatively easy to ionize helium: a temperature of just above 25,000 K will suffice. But in the high pressure gas below the surface, the ionization of helium is "postponed" to higher temperatures: 50% ionization of helium does not occur until the temperature has risen above 140,000 K.

These are quantitative illustrations of a point we made already in discussing Figure 3.3: at a given temperature, increasing pressure leads to lower degrees of ionization, and therefore more bound states to cause high opacity.

4.5 Contours of Constant Ionization: The Two Limits

Another way to look at Equation 4.5 is in terms of contours in the $(p_{\rm e} - T)$ plane along which the degree of ionization is constant. Such contours are shown in Figure 4.1 for 10 and 90% degrees of ionization. There are two principal segments of each contour, with different dependences on temperature. At low T, the term θI dominates on the left-hand side of Equation 4.6. As a

result, each ionization contour is described essentially by $\theta I = -\log p_{\rm e}$, i.e., $p_{\rm e} \sim \exp(-1/T)$. This is a curve which falls off steeply in the $(p_e - T)$ plane at low T. In the opposite limit, where temperatures are high, $\theta I \to 0$, and each contour is described essentially by $p_{\rm e} \sim T^{2.5}$. This is a gently sloping line. The transition from one segment to the other occurs at the location where the terms θI and $2.5 \log T$ are comparable in magnitude.

4.6 Application of the Saha Equation to the Negative Hydrogen Ion

What about H^-? Can we apply the Saha equation to this ion? Yes, except that in this case, we start with a charged particle and end up with one neutral particle plus one electron. But the principle remains the same, as long as we replace $n_{\rm a}$ in Equation 4.2 by $n(H^-)$, and replace n_i in Equation 4.2 by $n(H)$. (Here we use the symbol $n(H)$ to emphasis that the resulting "ionized particle" is actually a neutral hydrogen atom.) Inserting the known ionization potential (0.754 eV), we find

$$\log_{10}\left(\frac{n(H)}{n(H^-)}\right) = -\log_{10} p_{\rm e} + 2.5 \log_{10} T - 0.754\theta \qquad (4.10)$$

It is more common to be interested in the (small) ratio $\varphi = n(H^-)/n(H)$. So we rewrite the equation as

$$\log_{10} \varphi = \log_{10} p_{\rm e} - 2.5 \log_{10} T + 0.754\theta \qquad (4.11)$$

In the lower solar photosphere, where $T \approx 6000$ K ($\theta \approx 0.84$), we have already seen (Chapter 3, Section 3.4) that the electron pressure is given by $\log p_{\rm e} \approx 1.6$. Inserting numerical values in Equation 4.6, we find $\log \varphi \approx -7.22$. In the upper photosphere, where $T \approx 4900$ K ($\theta \approx 1.03$), and $\log p_{\rm e} \approx 0$ (Chapter 3, Section 3.4), $\log \varphi \approx -8.45$. In view of these estimates for upper and lower photosphere, we conclude that the average value of $\log \varphi$ in the solar photosphere is close to -8. It was a value of precisely this order, $\varphi \approx 10^{-8}$, which we found in Chapter 3 (Section 3.4) to be necessary to have H^- contribute the dominant opacity in the solar photosphere.

Exercises

4.1 The ionization strips in Figure 4.1 are defined by the somewhat arbitrary percentages of 10 and 90% ionization. Determine where the ionization strips lie in the $\log p_{\rm e}$–$\log T$ plane for the cases where the ionization percentages are 0.1 and 99.9%. Do this for HI, HeI, and HeII.

4.2 In Section 4.5, there is a definition of a transition point between the two segments of the ionization contours. Use the appropriate ionization potentials to evaluate the temperature of the transition point for HI, HeI, and HeII.

References

Saha, M. N. 1921. "On a physical theory of stellar spectra," *Proc. Roy. Soc London A,* 99, 135.

Chapter 5

Computing a Model of the Sun: The Photosphere

Now that we have information about opacity, we are almost ready to undertake the calculation of a model of the first segment of the Sun which is accessible to modeling: the photosphere. In calculating this model, we do not inquire into the *origin of the energy* which flows through the solar atmosphere. Instead, we simply take the luminosity (or flux) as a given, and calculate how the physical parameters of the medium arrange themselves so as to "handle" the energy which is passing through.

But first we require the answer to two questions: (i) how is the pressure related to the temperature? (ii) how does the pressure vary with height? As regards (i), the gas in the solar photosphere is of sufficiently low density ρ and of sufficiently high temperature T that the gas can be taken to behave as a perfect gas, with pressure given by the formula $p = R_g T\rho/\mu_a$. Here μ_a is the mean molecular weight, and $R_g = 8.31448 \times 10^7$ ergs deg^{-1} mole^{-1} is the gas constant. The chemical composition of the solar photosphere, consisting of some 90% hydrogen (by number), about 9% of helium, and about 1% of heavier elements ("metals"), leads to $\mu_a \approx 1.3$. Moreover, between the upper and lower photosphere (as defined earlier), the temperature ranges from the boundary value $T_o \approx 4900$ K to a temperature of order 6000 K. Over such a range of temperature, variations in the degree of ionization of hydrogen and helium are very small. This allows us to assume, without significant error, that μ_a remains constant throughout the photosphere.

5.1 Hydrostatic Equilibrium: The Scale Height

As for question (ii), the variation of gas pressure with height in any medium may be determined readily if the medium satisfies the condition of hydrostatic equilibrium (HSE). The HSE condition is applicable if the pressure $p(h)$ of the atmospheric material at any location (at height h) supports the weight of all of the atmospheric material which lies at heights above h. The equation which describes HSE is

$$\frac{dp}{dh} = -g\rho \tag{5.1}$$

where g is the local acceleration due to gravity, acting downward, in the direction of decreasing h. In the photosphere of the Sun, we have already mentioned that $g = 27{,}421.6$ cm sec^{-2}. The value of g at any particular height h decreases as h increases in proportion to $1/(R_\odot + h)^2$. When we consider a photospheric model in which the height h varies over a range of (say) $\Delta h = 500$ km, the relative change in gravity from lower to upper photosphere is given by

$$\frac{\Delta g}{g} \approx \frac{2\Delta h}{R_\odot} \approx 10^{-3} \tag{5.2}$$

The ratio $\Delta g/g$ is so small that, in a model of the photosphere, g may safely be taken to be a constant.

In a medium where g is constant, a particular solution of HSE provides a useful length-scale. To see this, consider a medium which is an isothermal perfect gas, with $p = R_g T\rho/\mu$. Then the solution of Equation 5.1 for the pressure as a function of height is as follows:

$$p(h) = p(0) \exp\left(-\frac{h}{H_p}\right) \tag{5.3}$$

And for the density as a function of height, the solution is analogous:

$$\rho(h) = \rho(0) \exp\left(-\frac{h}{H_p}\right) \tag{5.4}$$

That is, the pressure and the density both decrease exponentially with increasing height. At some arbitrary height, which is chosen as the zero-point of h, the local pressure and density are $p(0)$ and $\rho(0)$, respectively. The characteristic length scale H_p is referred to as the "scale height" or the "pressure scale height" or the "density scale height" of the isothermal atmosphere. The formula for H_p is $R_g T/g\mu_a$. Inserting numerical values, we find that in the photosphere, H_p varies over the range 115–140 km, i.e.,

$$H_p = (1.15 - 1.4) \times 10^7 \text{cm} \tag{5.5}$$

The atmosphere which lies above the level where $h = 0$ presses down on the gas at $h = 0$ due to the gravitational pull. The weight of the overlying material exerts a pressure on the gas at $h = 0$. In order to evaluate the pressure, let us consider a 1 cm^2 horizontal element of area at $h = 0$, and let us imagine a column with cross-sectional area 1 cm^2 extending upward to infinity from that element. The total amount of mass in that column can be obtained by integrating Equation 5.4 from $h = 0$ to $h = \infty$. The result is a mass column density d_c equal to $\rho(0)H_p$ gm cm^{-2}. An alternative way to state this information is to note that the column density, i.e., the number N_c of atoms in a sq.cm column above $h = 0$ equals $n(0)H_p$ cm^{-2}, where $n(0)$ is the number density of atoms at $h = 0$.

We shall find that in the solar photosphere, $\rho(0) \approx (2 - 3) \times 10^{-7}$ gm cm^{-3}, i.e., $n(0) \approx (1 - 2) \times 10^{17}$ cm^{-3}. Combining $\rho(0)$ and $n(0)$ with a

mean $H_p \approx$ 130 km, we find $d_c \approx (3-4)$ gm cm^{-2}, and $N_c \approx (1-3)\times 10^{24}$ cm^{-2}. In HSE, the pressure which occurs due to the weight of this column is $p(0) = d_c g \approx 10^5$ dyn cm^{-2}. (We shall want to check, when we compute a model of the photosphere, that our model yields pressures of this order: see Section 5.6 below.)

For comparison, we note that the atmospheric pressure on the surface of the Earth is about ten times larger than the photospheric $p(0)$. Of course, the processes which determine the atmospheric density and pressure at the surface of the Earth are very different from those which determine $\rho(0)$ in the Sun: the latter is determined by the requirement that the optical depth $\tau(0)$ be of order unity. There is no such requirement for the Earth: the fact that an observer standing on the surface of the Earth can see the Sun and stars clearly indicates that the optical depth τ_E of cloud-free atmosphere (in visible light) is actually considerably less than unity.

5.2 Sharp Edge of the Sun's Disk

Before moving on to the photospheric model, we make a short diversion here to address a problem which we now have enough information to solve. Combining the scale height in the photosphere with the results of Chapter 2 (Section 2.5.2) helps us to understand why the Sun, although a gaseous body, has a sharp edge.

It is a fact of life on Earth that the atmosphere we breathe is in turbulent motion. When we observe a distant object through the atmosphere, the turbulence causes smearing of the object. This is referred to as "the effects of seeing." As a result of "seeing," it is typically true that an observer on the Earth cannot distinguish two objects which are closer together than about 1 arc sec.

Suppose an observer wishes to make two measurements of solar intensity, I_1 and I_2, near the limb of the Sun. The first measurement I_1 is along a line of sight which is as nearly as possible "on the limb." This line of sight, at its closest approach to the Sun, passes through gas at a certain height h_1 in the upper photosphere. This is the measurement which, in visible light, yields an intensity $I_1 = a_\lambda = 0.4$ (relative to disk center) (see Chapter 2, Equation 2.5). According to the results of Chapter 2 (Section 2.5.2), the value of I_1 is determined by the product of the local source function S_λ times the optical depth τ_1 along the line of sight: $I_1 = \tau_1 S_\lambda$. To make the second measurement of intensity, I_2, the observer chooses a line of sight which is displaced off the limb by the smallest possible amount permitted by "seeing." The second line of sight, shifted by 1 arc sec relative to the first, at its closest approach to the Sun passes through gas which lies at a height $h_2 = h_1 + 730$ km (see Chapter 1, Section 1.2). The gas which lies at height h_2 has a density which is reduced below that at height h_1 by a factor $\exp[(h_2 - h_1)/H_p] \approx e^{5.6} = 10^{2.44} \approx 270$.

The reduction in density has the effect that the optical depth through such gas τ_2 is less than τ_1 by a factor of 270. Therefore the intensity $I_2 = \tau_1 S_\lambda/270$. We have seen (Chapter 2, Section 2.10) that the temperature in the upper photosphere approaches a constant value as height increases. Thus, T does not change significantly between h_2 and h_1. To the extent that the source function can be identified with the Planck function, this means that S_λ is essentially the same along both lines of sight. Therefore, $I_2 = I_1/270 = 0.0015$.

That is, by shifting my line of sight by a mere 1 arc sec away from the limb, I measure that the observed intensity falls off by a factor of almost 300. This is in contrast to what happens on the solar disk: as the line of sight is moved from disk center to the limb, i.e., as the line of sight traverses some 960 arc sec, the intensity decreases gradually from 1.0 to 0.4. But with a further shift of only 1 arc sec in the line of sight, the intensity falls by a further factor of almost 300. It is this rapid transition in behavior between observations on the disk and off the disk which gives the Sun its sharp edge.

5.3 Preparing to Compute a Model of the Solar Photosphere

The aim of this exercise is to combine HSE and the temperature structure of the gray atmosphere to calculate a table of values of various physical parameters as a function of the vertical height coordinate (increasing upward). The model begins by tabulating temperature as a function of optical depth τ (increasing downward). Transformations between τ and h require knowledge of the opacity as a function of relevant physical parameters.

The HSE equation (Equation 5.1) can be converted to an optical depth scale by noting the definition $d\tau = -\kappa\rho dh$ where κ is the (gray) opacity: we shall use the Rosseland mean opacity for this quantity. This leads to the central equation for the present chapter:

$$\frac{dp}{d\tau} = \frac{g}{\kappa} \tag{5.6}$$

In order to solve this equation, we need to have access to values of Rosseland mean opacities as a function of temperature and pressure. A table of such values was kindly made available by Dr. R. L. Kurucz of the Harvard-Smithsonian Center for Astrophysics. For the convenience of the reader, these are presented in Tables 5.1 and 5.2. (The reader may also be able to find results obtained by other researchers on the web.) In the tables below, the (log) opacities are tabulated as functions of temperature and pressure for a mixture of elemental abundances which is a "standard" solar mixture. Bound-bound, bound-free, and free-free transitions are included for many stages of ionization of all elements in the mixture. Negative hydrogen ions and hydrogen molecules are also included. For bound-bound

TABLE 5.1: $\log_{10}(\kappa_R)$ in units of $\mathrm{cm}^2\ \mathrm{gm}^{-1}$ for pressures (dyn cm^{-2}) in the lower subrange $LP \equiv \log_{10}(p) = 3.00 - 5.20$ and for temperatures in the range $LT \equiv \log_{10}(T) = 3.32 - 4.90$

LP→	3.00	3.20	3.40	3.60	3.80	4.00	4.20	4.40	4.60	4.80	5.00	5.20
LT												
3.32	−5.16	−5.13	−5.08	−5.04	−4.99	−4.93	−4.87	−4.80	−4.74	−4.66	−4.58	−4.49
3.34	−4.88	−4.85	−4.82	−4.78	−4.73	−4.68	−4.63	−4.58	−4.52	−4.47	−4.40	−4.34
3.36	4.56	−4.53	−4.49	−4.46	−4.42	−4.37	−4.33	−4.28	−4.23	−4.18	−4.12	−4.06
3.38	−4.23	−4.19	−4.15	−4.11	−4.07	−4.03	−3.99	−3.95	−3.91	−3.86	−3.81	−3.76
3.40	−3.94	−3.88	−3.82	−3.77	−3.72	−3.68	−3.64	−3.61	−3.57	−3.53	−3.49	−3.44
3.42	−3.70	−3.61	−3.54	−3.46	−3.40	−3.35	−3.30	−3.26	−3.22	−3.18	−3.14	−3.11
3.44	−3.50	−3.40	−3.30	−3.21	−3.12	−3.05	−2.99	−2.93	−2.88	−2.84	−2.80	−2.76
3.46	−3.35	−3.22	−3.10	−2.99	−2.89	−2.80	−2.72	−2.64	−2.58	−2.53	−2.48	−2.43
3.48	−3.21	−3.08	−2.95	−2.82	−2.70	−2.59	−2.48	−2.39	−2.32	−2.24	−2.18	−2.13
3.50	−3.07	−2.94	−2.81	−2.67	−2.54	−2.41	−2.30	−2.19	−2.09	−2.00	−1.92	−1.85
3.52	−2.91	−2.79	−2.66	−2.53	−2.40	−2.27	−2.14	−2.02	−1.90	−1.80	−1.70	−1.62
3.54	−2.72	−2.61	−2.49	−2.37	−2.24	−2.12	−1.99	−1.87	−1.75	−1.63	−1.52	−1.42
3.56	−2.54	−2.43	−2.31	−2.19	−2.07	−1.96	−1.84	−1.71	−1.59	−1.48	−1.36	−1.25
3.58	−2.39	−2.27	−2.14	−2.02	−1.90	−1.79	−1.67	−1.55	−1.43	−1.32	−1.20	−1.09
3.60	−2.28	−2.14	−2.01	−1.88	−1.75	−1.63	−1.51	−1.39	−1.27	−1.15	−1.04	−0.92
3.62	−2.22	−2.07	−1.92	−1.77	−1.64	−1.50	−1.37	−1.25	−1.12	−1.00	−0.88	−0.77
3.64	−2.20	−2.03	−1.87	−1.71	−1.56	−1.41	−1.27	−1.13	−1.00	−0.87	−0.75	−0.62
3.66	−2.21	−2.03	−1.86	−1.69	−1.52	−1.36	−1.21	−1.05	−0.91	−0.77	−0.63	−0.50
3.68	−2.19	−2.03	−1.86	−1.69	−1.52	−1.35	−1.18	−1.01	−0.86	−0.70	−0.56	−0.41
3.70	−2.11	−1.97	−1.83	−1.67	−1.51	−1.34	−1.17	−1.00	−0.83	−0.67	−0.51	−0.35
3.73	−1.86	−1.75	−1.63	−1.51	−1.39	−1.25	−1.11	−0.96	−0.80	−0.64	−0.48	−0.31
3.76	−1.53	−1.44	−1.34	−1.24	−1.13	−1.02	−0.91	−0.79	−0.66	−0.53	−0.39	−0.25
3.79	−1.18	−1.10	−1.01	−0.92	−0.82	−0.73	−0.63	−0.53	−0.42	−0.31	−0.20	−0.08

(Continued)

TABLE 5.1: (Continued)

3.82	−0.81	−0.74	−0.66	−0.58	−0.50	−0.41	−0.32	−0.23	−0.14	−0.04	0.06	0.16
3.85	−0.43	−0.37	−0.30	−0.23	−0.16	−0.08	0.00	0.08	0.16	0.25	0.34	0.43
3.88	−0.03	0.02	0.07	0.13	0.19	0.26	0.33	0.40	0.47	0.55	0.63	0.71
3.91	0.37	0.41	0.46	0.51	0.56	0.61	0.67	0.73	0.79	0.86	0.93	1.00
3.94	0.76	0.80	0.84	0.88	0.93	0.97	1.02	1.07	1.12	1.18	1.23	1.30
3.97	1.08	1.13	1.18	1.23	1.27	1.32	1.36	1.40	1.45	1.49	1.54	1.59
4.00	1.28	1.37	1.45	1.51	1.57	1.62	1.66	1.71	1.75	1.79	1.83	1.88
4.05	1.30	1.46	1.60	1.73	1.85	1.95	2.03	2.10	2.16	2.21	2.26	2.30
4.10	1.07	1.26	1.44	1.62	1.79	1.95	2.10	2.23	2.35	2.45	2.53	2.60
4.15	0.78	0.96	1.16	1.35	1.54	1.73	1.92	2.10	2.27	2.43	2.58	2.71
4.20	0.54	0.71	0.88	1.06	1.25	1.45	1.64	1.84	2.04	2.23	2.42	2.60
4.25	0.36	0.51	0.67	0.83	1.01	1.20	1.38	1.58	1.77	1.97	2.17	2.38
4.30	0.21	0.34	0.48	0.63	0.80	0.98	1.16	1.35	1.55	1.74	1.94	2.14
4.35	0.09	0.20	0.32	0.47	0.62	0.78	0.96	1.15	1.34	1.54	1.74	1.94
4.40	0.00	0.09	0.19	0.32	0.46	0.61	0.78	0.96	1.15	1.35	1.54	1.75
4.45	−0.04	0.04	0.12	0.23	0.35	0.48	0.64	0.81	0.99	1.18	1.37	1.58
4.50	−0.04	0.02	0.10	0.19	0.30	0.41	0.55	0.70	0.87	1.04	1.23	1.43
4.55	−0.08	0.00	0.08	0.17	0.27	0.38	0.51	0.64	0.80	0.96	1.13	1.32
4.60	−0.15	−0.08	0.00	0.10	0.21	0.33	0.46	0.60	0.75	0.90	1.07	1.24
4.65	−0.22	−0.17	−0.10	−0.02	0.08	0.20	0.33	0.47	0.63	0.80	0.97	1.15
4.70	−0.29	−0.25	−0.20	−0.13	−0.05	0.05	0.16	0.29	0.44	0.60	0.77	0.95
4.75	−0.33	−0.30	−0.27	−0.22	−0.16	−0.09	0.00	0.11	0.24	0.38	0.54	0.71
4.80	−0.35	−0.33	−0.31	−0.28	−0.24	−0.19	−0.13	−0.05	0.05	0.17	0.31	0.46
4.85	−0.35	−0.34	−0.32	−0.31	−0.28	−0.25	−0.21	−0.15	−0.08	0.01	0.12	0.24
4.90	−0.36	−0.35	−0.34	−0.32	−0.30	−0.27	−0.24	−0.20	−0.15	−0.09	−0.01	0.08

Source: Used with the permission of R. L. Kurucz (Harvard-Smithsonian C of A).

TABLE 5.2: $\log_{10}(\kappa_R)$ in units of $\mathrm{cm}^2\ \mathrm{gm}^{-1}$ for pressures (dyn cm^{-2}) in the upper subrange $LP \equiv \log_{10}(p) = 5.40 - 8.00$ and for temperatures in the range $LT \equiv \log_{10}(T) = 3.32 - 4.90$

LP→	5.40	5.60	5.80	6.00	6.25	6.50	6.75	7.00	7.25	7.50	7.75	8.00
LT												
3.32	−4.39	−4.27	−4.14	−3.99	−3.78	−3.51	−3.19	−2.98	−2.88	−2.77	−2.67	−2.57
3.34	−4.28	−4.21	−4.14	−4.07	−3.99	−3.90	−3.81	−3.72	−3.63	−3.54	−3.44	−3.36
3.36	−4.00	−3.93	−3.87	−3.80	−3.71	−3.62	−3.53	−3.44	−3.35	−3.26	−3.17	−3.08
3.38	−3.70	−3.64	−3.58	−3.51	−3.43	−3.34	−3.26	−3.17	−3.08	−2.99	−2.90	−2.81
3.40	−3.39	−3.34	−3.28	−3.22	−3.14	−3.06	−2.98	−2.89	−2.80	−2.72	−2.63	−2.55
3.42	−3.06	−3.02	−2.97	−2.92	−2.85	−2.77	−2.69	−2.61	−2.53	−2.45	−2.36	−2.28
3.44	−2.73	−2.69	−2.64	−2.60	−2.54	−2.47	−2.40	−2.33	−2.25	−2.17	−2.10	−2.02
3.46	−2.39	−2.35	−2.32	−2.28	−2.22	−2.17	−2.11	−2.04	−1.97	−1.90	−1.83	−1.75
3.48	−2.08	2.03	−1.99	−1.96	−1.91	−1.86	−1.81	−1.75	−1.69	−1.63	−1.56	−1.49
3.50	−1.79	−1.74	−1.69	−1.65	−1.60	−1.55	−1.50	−1.45	−1.40	−1.35	−1.29	−1.23
3.52	−1.54	−1.48	−1.42	−1.37	−1.31	−1.26	−1.21	−1.16	−1.12	−1.07	−1.02	−0.96
3.54	−1.33	−1.25	−1.18	−1.11	−1.04	−0.98	−0.93	−0.88	−0.84	−0.79	−0.75	−0.70
3.56	−1.15	−1.05	−0.97	−0.89	−0.81	−0.74	−0.67	−0.62	−0.57	−0.53	−0.49	−0.44
3.58	−0.98	−0.88	−0.78	−0.69	−0.60	−0.51	−0.44	−0.38	−0.32	−0.28	−0.23	−0.19
3.60	−0.81	−0.71	−0.61	−0.51	−0.41	−0.31	−0.23	−0.16	−0.09	−0.04	0.00	0.05
3.62	−0.65	−0.55	−0.44	−0.34	−0.23	−0.13	−0.04	0.04	0.11	0.17	0.22	0.27
3.64	−0.51	−0.39	−0.28	−0.18	−0.06	0.05	0.14	0.23	0.30	0.37	0.43	0.48
3.66	−0.38	−0.26	−0.14	−0.04	0.09	0.21	0.31	0.40	0.48	0.56	0.62	0.67
3.68	−0.28	−0.15	−0.03	0.09	0.23	0.35	0.47	0.57	0.65	0.73	0.79	0.85
3.70	−0.21	−0.07	0.07	0.19	0.34	0.48	0.60	0.71	0.81	0.89	0.96	1.02
3.73	−0.15	0.00	0.15	0.30	0.46	0.62	0.76	0.89	1.01	1.11	1.19	1.25
3.76	−0.10	0.06	0.21	0.36	0.54	0.71	0.87	1.02	1.16	1.28	1.38	1.46
3.79	0.05	0.18	0.31	0.45	0.62	0.79	0.96	1.12	1.27	1.41	1.53	1.63
3.82	0.27	0.38	0.50	0.62	0.76	0.91	1.06	1.22	1.37	1.52	1.65	1.77

(Continued)

TABLE 5.2: (Continued)

3.85	0.53	0.63	0.73	0.83	0.96	1.09	1.23	1.37	1.50	1.64	1.78	1.90
3.88	0.80	0.89	0.98	1.07	1.19	1.31	1.43	1.55	1.68	1.81	1.93	2.06
3.91	1.08	1.15	1.24	1.32	1.43	1.54	1.65	1.76	1.88	1.99	2.11	2.23
3.94	1.36	1.43	1.50	1.57	1.67	1.77	1.87	1.97	2.08	2.19	2.30	2.41
3.97	1.65	1.71	1.76	1.83	1.91	2.00	2.09	2.18	2.28	2.38	2.49	2.59
4.00	1.92	1.97	2.02	2.08	2.15	2.23	2.31	2.39	2.48	2.57	2.67	2.77
4.05	2.35	2.39	2.43	2.48	2.54	2.60	2.66	2.73	2.81	2.89	2.97	3.07
4.10	2.67	2.72	2.78	2.82	2.88	2.94	3.00	3.06	3.13	3.20	3.28	3.36
4.15	2.82	2.92	3.00	3.08	3.16	3.23	3.30	3.36	3.43	3.50	3.57	3.65
4.20	2.77	2.92	3.06	3.19	3.32	3.43	3.52	3.61	3.69	3.76	3.84	3.92
4.25	2.57	2.76	2.95	3.13	3.32	3.50	3.64	3.77	3.88	3.98	4.07	4.16
4.30	2.35	2.55	2.76	2.95	3.20	3.43	3.63	3.82	3.98	4.12	4.24	4.35
4.35	2.14	2.35	2.56	2.76	3.03	3.28	3.53	3.76	3.97	4.16	4.33	4.48
4.40	1.96	2.16	2.38	2.59	2.85	3.12	3.38	3.64	3.88	4.12	4.33	4.52
4.45	1.78	1.99	2.20	2.41	2.68	2.95	3.23	3.50	3.77	4.03	4.26	4.49
4.50	1.63	1.83	2.05	2.26	2.53	2.79	3.07	3.35	3.63	3.90	4.17	4.41
4.55	1.51	1.70	1.91	2.11	2.38	2.64	2.92	3.19	3.47	3.75	4.03	4.29
4.60	1.41	1.60	1.79	1.99	2.24	2.50	2.76	3.03	3.30	3.58	3.85	4.12
4.65	1.33	1.50	1.69	1.87	2.11	2.36	2.61	2.86	3.12	3.38	3.65	3.91
4.70	1.14	1.33	1.53	1.72	1.96	2.20	2.44	2.67	2.91	3.16	3.42	3.67
4.75	0.89	1.08	1.27	1.47	1.72	1.96	2.21	2.45	2.69	2.93	3.17	3.41
4.80	0.63	0.81	0.99	1.19	1.43	1.67	1.92	2.17	2.41	2.66	2.90	3.14
4.85	0.39	0.55	0.72	0.90	1.14	1.38	1.62	1.87	2.11	2.36	2.60	2.84
4.90	0.20	0.33	0.48	0.64	0.87	1.10	1.34	1.58	1.82	2.07	2.31	2.55

Source: Used with the permission of R. L. Kurucz (Harvard-Smithsonian C of A).

transitions, the lines are assumed to be broadened with a microturbulent velocity of 2 km $\sec^{-1}$. Each *row* of Tables 5.1 and 5.2 is labeled with LT, which is equal to the logarithm (to base 10) of the temperature (in degrees K): the temperatures range from close to 2000 K to almost 100000 K. Each *column* of Tables 5.1 and 5.2 is labeled with LP, which is equal to the logarithm (to base 10) of the pressure (in dyn cm^{-2}): the pressures range over five orders of magnitude. In order to fit the results into a standard page width, the opacities are presented in the form of two tables, corresponding to a low subrange of pressure and a high subrange of pressure. The tabulated values of opacity exhibit the overall behavior described earlier in Chapter 3 (Figure 3.3) (where the results are presented in a different format): (i) in the limit of high temperature (and low pressure), $\log(\kappa) \rightarrow -0.5$; (ii) in the limit of low temperature, $\log(\kappa)$ tends to very small values; (iii) numerical values of opacity reach maximum values at $\log(T) = 4.0$–4.5; (iv) maximum values of opacity are 10^{4-5} cm^2 gm^{-1}.

The goal of the present chapter is to calculate a tabulated model of the solar photosphere. This means that we wish to obtain a table of values where each row of the table refers to a particular optical depth in the atmosphere. On that row, we wish to provide numerical values for the temperature, pressure, density, and height in the solar atmosphere.

5.4 Computing a Model of the Photosphere: Step by Step

The calculation proceeds by way of the following steps.

1. Choose a value of τ for the first row in the tabulated model: e.g., $\tau(1) = 10^{-4}$.

2. For row 1, choose the vertical depth coordinate to be $z(1) = 0$. (This is an arbitrary choice, and is done merely for convenience. Afterward, you may change the zero-point of height if you choose.) We will use h (increasing upward) and $-z$ (where z increases downward) interchangeably in the calculation.

3. Calculate the temperature in row 1, $T(1)$, from the Eddington solution (Chapter 2, Equation 2.40), using $T_{\mathrm{eff}} = 5777$ K (see Chapter 1, Equation 1.15).

4. To obtain the pressure $p(1)$ in row 1, one could guess any finite starting value and then iterate. To avoid complications, I suggest that you simply choose the following guess: $\log p(1) = 3.0$. Note that in all cases, the log function refers to logarithms to base 10, and the physical quantities are in c.g.s. units.

5. Now that you know $T(1)$ and $p(1)$, calculate the density in row 1, $\rho(1)$, from the perfect gas expression $\rho(1) = p(1)\mu_a/(R_g T(1))$. Here, R_g is the gas constant (see Chapter 1, Section 1.7), and μ_a, the mean molecular weight, can be set equal to a constant value 1.3 for the material of the solar photosphere. (The value 1.3 arises because solar material is roughly 90% H, 10% He by numbers, plus less than 1% heavier elements.)

6. Now that you know log $T(1)$ and log $p(1)$, interpolate in Tables 5.1 and/or 5.2 to find a local value for the opacity $\kappa(1)$.

7. Step forward to the second row of the table, i.e., to the next value of τ. In order to reduce numerical errors, I suggest that you keep the step size small. For example, consider using $\tau(2) = 2 \times 10^{-4}$. This means that the interval in optical depth between rows 1 and 2 is $\Delta\tau = 10^{-4}$.

8. With the new value of τ, calculate the new value of T from the Eddington solution. Call this $T(2)$.

9. Calculate the increase in pressure between rows 1 and 2 using an approximation to Equation 5.6: $\Delta p = g\Delta\tau/\kappa(1)$, where $g = 2.7 \times 10^4$ cm sec^{-2}. This then gives $p(2) = p(1) + \Delta p$.

10. Knowing $T(2)$ and $p(2)$, interpolate in the opacity table for $\kappa(2)$.

11. Calculate the density $\rho(2)$ from $T(2)$ and $p(2)$.

12. Convert the step in optical depth to a step in linear depth: $\Delta z = +\Delta\tau/(\rho(2)\kappa(2))$. If you want to be more precise, replace the denominator by the mean value of $\rho\kappa$ between row 1 and row 2: $\rho\kappa \approx 0.5(\rho(1)\kappa(1)+\rho(2)\kappa(2))$. Once Δz is calculated, you can calculate the depth $z(2) = z(1) + \Delta z$ which is appropriate for row 2 of the tabulated model.

13. Calculate the local temperature gradient $dT/dz = (T(2) - T(1))/\Delta z$. (This gradient will be used later when we wish to calculate a model for the convection zone.)

    At this point, there should be seven entries in row 2: τ, T, p, ρ, z, κ, and (dT/dz).

    Use those values to step forward to row 3. For generality, we refer to the quantities in the row we have just calculated as row i.

14. Step forward to row $i + 1$. To start this step, choose a new value of $\tau(i+1) = \tau(i) + \Delta\tau$. What step size should be used? Plausible choices might be $\Delta\tau = 10^{-4}$ until the optical depth τ reaches a value of 10^{-3}. Then use a step size of $\Delta\tau = 10^{-3}$ until $\tau = 10^{-2}$. Then use a step size of $\Delta\tau = 10^{-2}$ until $\tau = 10^{-1}$. Then use a step size of $\Delta\tau = 10^{-1}$ until $\tau = 1$. Finally use a step size $\Delta\tau = 1$ until $\tau = 10$.

15. Repeat steps 8–14 multiple times until τ reaches a value of about 10. In each iteration, replace $T(1)$ in the above instructions by $T(i)$ and $T(2)$ by $T(i+1)$. Do the same replacements for the other variables.

5.5 The Outcome of the Calculation

The outcome of the exercise is a table of physical quantities as a function of height: this is called a "model of the photosphere." An example of such an exercise, using the above tables of opacity, is given in Table 5.3. At each

TABLE 5.3: A model of the solar atmosphere

τ	Temperature	Pressure	Density	z	$\log(\kappa)$	grad T
2.00E-04	4.86E+03	1.77E+03	5.69E-09	1.91E+06	$-2.04E+00$	9.57E-08
3.00E-04	4.86E+03	2.02E+03	6.52E-09	3.35E+06	$-1.97E+00$	1.26E-07
4.00E-04	4.86E+03	2.26E+03	7.26E-09	4.51E+06	$-1.93E+00$	1.57E-07
5.00E-04	4.86E+03	2.47E+03	7.94E-09	5.48E+06	$-1.89E+00$	1.87E-07
6.00E-04	4.86E+03	2.66E+03	8.57E-09	6.32E+06	$-1.86E+00$	2.17E-07
7.00E-04	4.86E+03	2.85E+03	9.17E-09	7.06E+06	$-1.83E+00$	2.48E-07
8.00E-04	4.86E+03	3.02E+03	9.73E-09	7.71E+06	$-1.80E+00$	2.78E-07
9.00E-04	4.86E+03	3.19E+03	1.03E-08	8.30E+06	$-1.78E+00$	3.07E-07
1.00E-03	4.86E+03	3.35E+03	1.08E-08	8.84E+06	$-1.76E+00$	3.38E-07
1.10E-03	4.86E+03	3.50E+03	1.13E-08	9.33E+06	$-1.75E+00$	3.68E-07
2.10E-03	4.86E+03	4.97E+03	1.60E-08	1.27E+07	$-1.73E+00$	5.42E-07
3.10E-03	4.86E+03	6.07E+03	1.95E-08	1.47E+07	$-1.60E+00$	8.85E-07
4.10E-03	4.87E+03	7.00E+03	2.25E-08	1.62E+07	$-1.53E+00$	1.20E-06
5.10E-03	4.87E+03	7.83E+03	2.51E-08	1.74E+07	$-1.48E+00$	1.52E-06
6.10E-03	4.87E+03	8.58E+03	2.75E-08	1.84E+07	$-1.44E+00$	1.82E-06
7.10E-03	4.87E+03	9.27E+03	2.98E-08	1.93E+07	$-1.40E+00$	2.13E-06
8.10E-03	4.87E+03	9.92E+03	3.18E-08	2.00E+07	$-1.37E+00$	2.43E-06
9.10E-03	4.87E+03	1.05E+04	3.38E-08	2.07E+07	$-1.35E+00$	2.73E-06
1.01E-02	4.88E+03	1.11E+04	3.56E-08	2.13E+07	$-1.33E+00$	3.02E-06
2.01E-02	4.89E+03	1.67E+04	5.32E-08	2.51E+07	$-1.31E+00$	4.71E-06
3.01E-02	4.91E+03	2.06E+04	6.55E-08	2.73E+07	$-1.16E+00$	8.11E-06
4.01E-02	4.93E+03	2.39E+04	7.57E-08	2.89E+07	$-1.08E+00$	1.11E-05
5.01E-02	4.95E+03	2.67E+04	8.45E-08	3.01E+07	$-1.02E+00$	1.39E-05
6.01E-02	4.96E+03	2.94E+04	9.25E-08	3.11E+07	$-9.79E-01$	1.66E-05
7.01E-02	4.98E+03	3.18E+04	9.97E-08	3.20E+07	$-9.45E-01$	1.92E-05
8.01E-02	5.00E+03	3.40E+04	1.06E-07	3.28E+07	$-9.14E-01$	2.18E-05
9.01E-02	5.01E+03	3.61E+04	1.13E-07	3.35E+07	$-8.88E-01$	2.43E-05
1.00E-01	5.03E+03	3.82E+04	1.19E-07	3.41E+07	$-8.64E-01$	2.67E-05
2.00E-01	5.19E+03	5.67E+04	1.71E-07	3.80E+07	$-8.30E-01$	3.95E-05
3.00E-01	5.33E+03	6.98E+04	2.05E-07	4.04E+07	$-6.80E-01$	6.14E-05
4.00E-01	5.46E+03	8.02E+04	2.30E-07	4.20E+07	$-5.78E-01$	8.05E-05
5.00E-01	5.59E+03	8.89E+04	2.49E-07	4.33E+07	$-4.99E-01$	9.75E-05
6.00E-01	5.70E+03	9.64E+04	2.64E-07	4.43E+07	$-4.38E-01$	1.12E-04
7.00E-01	5.81E+03	1.03E+05	2.77E-07	4.52E+07	$-3.73E-01$	1.28E-04
8.00E-01	5.92E+03	1.08E+05	2.87E-07	4.59E+07	$-3.06E-01$	1.47E-04
9.00E-01	6.01E+03	1.13E+05	2.94E-07	4.65E+07	$-2.46E-01$	1.64E-04

tabulated value of optical depth (column headed "τ"), we list the temperature (in units of K), pressure (units=dyn cm^{-2}), density (in gm cm^{-3}), depth (in cm, relative to an initial $h = 0$ at the top), the logarithm of the Rosseland mean opacity (in units of cm^2 gm^{-1}). In the seventh column, we give the temperature gradient, dT/dz (in units of deg K cm^{-1}). The reason for including this quantity will be explained when we discuss Convection in Chapter 6.

5.6 Overview of the Model of the Solar Photosphere

Now we have a table which lists certain physical properties of the gas over a range of heights in the solar photosphere. It is worthwhile to take a look at the properties we have obtained.

First, the gas temperature in the photosphere ($\tau = 0.667$) is 5777 K. This is of course a natural consequence of the Eddington solution (Chapter 2, Equation 2.40). But it is useful to remember that the effective temperature of the *radiation* (i.e., the photons) which comes to us from the Sun (Chapter 1, Equation 1.15) has a direct connection with the local thermodynamic temperature of the atoms *in the photosphere*. The photons we see on Earth have energies which are characteristic of the thermal energies of the atoms back in the photosphere. The reason for this is that the radiation and the gas in the photosphere are close to local thermodynamic *equilibrium*. (This is very different from the condition on the surface of the Earth, where the dominant photons [sunlight] have energies of a few electron volts [eV], while the gases in Earth's atmosphere have thermal energies of only 0.03 eV: this is far from thermodynamic equilibrium.)

Second, the density in the photosphere $\rho(\mathrm{ph})$ is $(2-3)\times10^{-7}$ gm cm^{-3}. If the gas were purely hydrogen, the corresponding number density of atoms would be $n(\mathrm{ph}) = (1.2-1.8)\times10^{17}$ cm^{-3}. (Given the presence of He and other heavier elements, the true value of $n(\mathrm{ph})$ is somewhat smaller than this.) The number column density above the photosphere $N(\mathrm{ph})$ is $n(\mathrm{ph})H_p$. Inserting $H_p = (1.15-1.4)\times10^7$ cm (Equation 5.5), we find $N(\mathrm{ph}) = (1.4-2.5)\times10^{24}$ cm^{-2}. This is the number of atoms which lie above each sq cm of the solar photosphere. Of these, roughly 1 in 10^8 is an H^- ion. Thus, there are some $N(H^-) = (1.4-2.5)\times10^{16} H^-$ ions lying above each sq cm of the photosphere.

Third, the pressure in the photosphere $p(\mathrm{ph})$ is of order 10^5 dyn cm^{-2}. This is the pressure which is necessary to support the weight of the overlying gas. Since each of the $N(\mathrm{ph})$ atoms overlying the photosphere has a mass that is close to the mass of a hydrogen atom (1.67×10^{-24} gm), we see that the exponents cancel out and the mass which lies above each sq cm of the photosphere is about 4 gm. In the presence of gravity with $g = 27422$ cm sec^{-2}, the corresponding force, mg, pressing down on each sq cm is therefore close to 10^5 dyn cm^{-2}. The fact that the pressure $p(\mathrm{ph})$ is equal to mg is not an accident; it indicates that vertical forces are in balance in the photosphere.

This is another way to say that in deriving the model atmosphere, we have assumed HSE.

Fourth, the range of depths Δz between the photosphere and the "top" of the model photosphere (which we have chosen to be at $\tau = 10^{-4}$) is 400–500 km. The actual height range depends on the choice of opacity. If we had used a different opacity table from Tables 5.1 and 5.2, then the height range could have been somewhat different. Other parameters would also have changed somewhat. But the values just cited give a reliable zeroth order overview of the physical parameters in the solar photosphere.

Fifth, we recall that the principal contributor to opacity in the photosphere is the negative hydrogen ion. The numerical values of p(ph) and ρ(ph) cited above take on the numerical values they do mainly because of the particular cross-section ($\sigma = 4.5 \times 10^{-17}$ cm^2) which is presented by an H^- ion to the photons that are most abundant in the solar spectrum. Given that our model has $N(H^-)H^-$ ions lying above each sq cm of the photosphere, the column density and the cross-section combine to yield an optical depth $N(H^-)\sigma$ of order unity in the solar photosphere.

An important question arises concerning the last line of Table 5.3. Why does Table 5.3 stop at a depth where $\tau = 0.9$? This seems like an odd place to stop the computation of a model photosphere. Shouldn't we keep going deeper? The answer is No, and the reason for this answer has to do with the numerical value of the temperature gradient dT/dz. We shall see in Chapter 6 that when dT/dz increases above a critical value g_{ad}, convection sets in and radiative transfer is no longer the dominant mode of energy transport in the atmosphere. Now, the computation which led to the results in Table 5.3 is based on solution (Equation 2.40) of the equation of radiative transfer, i.e., radiation carries the entire energy flux through the atmosphere. There is little meaning in applying such a computation to gas where convection is occurring. In Chapter 6, we shall show that the critical value g_{ad} is about 1.7×10^{-4} deg cm^{-1}. Inspection of Table 5.3 above shows that dT/dz increases with increasing depth, and is approaching this critical value as we approach the bottom of Table 5.3. In fact, if we were to continue the calculation of Table 5.3 to greater depths, we would find that at optical depth τ=1.0, the local value of dT/dz would exceed g_{ad}. Thus, our results suggest that convection sets in at optical depths between 0.9 and 1.0. This corresponds to a depth of only a few tens of kilometers below the photosphere.

Exercises

5.1 Evaluate the pressure scale height in regions of the solar atmosphere where the gas has temperature of 10^4, 10^6, and 10^7 K. (Such temperatures exist in the chromosphere, in the corona, and in flares: see Chapters 15 and 17.) Use molecular weight $\mu \approx 0.5$.

5.2 Perform the step-by-step calculation described in Section 5.4, using the opacities given in Tables 5.1 and 5.2. Compare your results with those in Table 5.3.

5.3 Repeat the calculations for different choices of various parameters. For example, use a starting pressure $\log p(1) = 2$ or 4. What differences do you find compared to the results in Table 5.3? Are some parameters more sensitive than others to the alteration in starting pressure? As a further example, use $\Delta\tau$ values which are twice as large as those suggested in Section 5.4. Then repeat the calculations using $\Delta\tau$ values which are one-half of the values suggested in Section 5.4. What differences do you find in the various cases?

5.4 The opacity tables given in Tables 5.1 and 5.2 were computed (by Dr R. Kurucz) using a number of choices of parameters (chemical mixture, microturbulence, etc.). Other tables of Rosseland mean opacities, using different choices of some parameters, exist in the literature (e.g., Iglesias and Rogers, 1996 and references therein). Use one of those tables to repeat the calculations in Section 5.4. Which parameters are altered most compared to the results in Table 5.3?

Note: in Tables 5.1 and 5.2, opacities are listed as functions of T and of *gas pressure* p. Opacity tables in the literature *may* list the opacity as functions of the R parameter (Chapter 3, Section 3.6), or as a function of T and *electron pressure* p_e. In order to use such tables in the procedure described in Section 5.4, you will need to convert from density to pressure (assuming a perfect gas), or you will need to find auxiliary tables which first convert from p_e to gas pressure p.

References

Iglesias, C. A. and Rogers, F. J. 1996. "Updated OPAL opacities," *Astrophys. J.*, 464, 943.

Kurucz, R. L. 1992. *Table of Rosseland Mean Opacities,* personal communication.

Chapter 6

Convection in the Sun: Empirical Properties

So far, we have been restricting attention to the upper parts of the *photosphere* in the Sun, where energy is transported through the gas almost entirely by means of radiation. The gas in the photosphere (at least in all parts except the very lowest regions) does not move, but simply "processes" the photons, absorbing, emitting, and scattering them in such a way that there is a net transport of energy in the outward (radial) direction. Because of the existence of radiative equilibrium, the equation of radiative transfer allows us to extract reliable physical properties of the gas in the photosphere, where there are no systematic gas motions.

Now we turn our attention to a region of the Sun where radiative equilibrium becomes progressively less important. In this new region, as we go deeper into the Sun, photons play a progressively minor role in transporting energy. In the deeper layers which now draw our attention, energy eventually is transported essentially completely by means of convection. Convection occurs when the material itself experiences bulk flows in order to carry the heat radially outward.

In order to obtain a complete physical model of the Sun, we shall eventually have to develop a theory which will allow us to model convective heat transport in solar conditions. The details of one such theory will be the subject of Chapter 7. Before embarking on the task of developing such a theory, however, we need to learn as much as we can from the empirical properties of convection as they present themselves to us in the visible layers of the solar atmosphere: these empirical properties will guide us in developing a convective model.

6.1 Nonuniform Brightness

Evidence for the presence of convection in the Sun can be seen when an image of the solar surface is obtained with sufficiently high angular resolution. An example is shown in Figure 6.1.

The Figure shows clearly that the surface of the Sun is not uniformly bright. We see that there are brighter areas (called granules) and darker areas

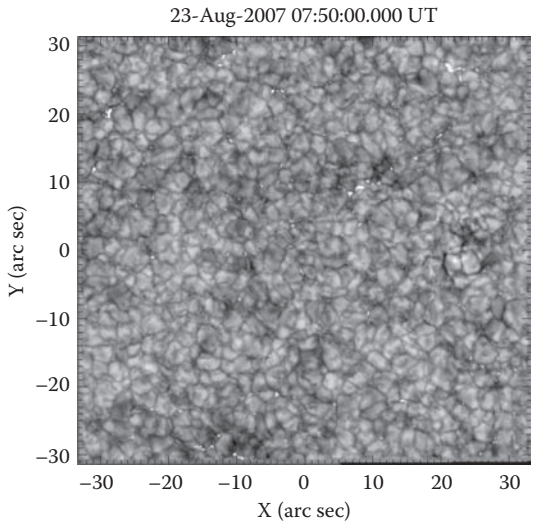


FIGURE 6.1: A close-up view of the solar surface. The field of view is about 48×48 Mm (where 1 Mm = 1000 km). The essential aspect is the "granular" structure of lighter patches surrounded by darker boundaries. (From Dr. M. Mathioudakis, Queen's University, Belfast, and the Swedish Solar Telescope. With permission.)

(called intergranular lanes). Our knowledge of radiative transfer suggests that the bright granules contain hotter gas than the dark intergranular lanes. The difference in temperature between the bright and the dark gas will be discussed quantitatively below.

We note the differing topology between brighter and darker features. It is possible to start in one location in an intergranular lane and move to other dark lanes without traversing a bright granule. But bright granules are for the most part isolated from one another: one cannot go from one bright granule to another without traversing an intergranular lane.

6.2 Granule Shapes

The individual granules in the image above have irregular shapes. On close inspection, one notices that some of the boundaries of many granules consist of lines which are nearly straight. One seems to be looking at a collection of shapes which hint at the geometrical structures known as polygons. Is there

any physical reason why polygons might be of interest when we speak of convection? Indeed there is.

From an empirical standpoint, polygons have been found to be the preferred spatial pattern of cells which occur in laboratory convection under certain conditions. When a layer of liquid is heated from below, and the temperature gradient between bottom and top is not too large, heat can be carried up through the liquid by conduction. In this phase, each molecule is pursuing its own purposes, carrying heat upward as an individual.

But at some point, when the temperature gradient becomes larger than a critical value, the liquid begins to move in a macroscopic way. In conditions which were studied by Benard (1900), the motions organize themselves into a geometrical pattern consisting of polygons. Each element of the pattern is called a cell (a "Benard" cell), involving the organized motion of trillions of atoms or molecules. No longer do the molecules behave as individuals in order to carry heat: instead, a lower energy condition can be reached if molecules cooperate in a macroscopic pattern of motion, with hot liquid rising and cool liquid sinking. This method of transporting heat is called convection. Benard cells are long-lived structures. Thus, Benard cells provide heat transfer by means of steady state convection.

In the Sun, the hint of polygonal structures among some granules is an intriguing reminder of the cells which Benard found in his experiments. However, the polygons in the Sun are by no means in steady state. Quite the contrary: individual granules are observed to live for only a finite time. When one performs correlation studies on images of granules over a large area of the Sun, one finds that on average, individual granules can no longer be clearly identified after a time-scale of 5–10 minutes (Title et al., 1989). This time-scale can be regarded as a sort of average "life-time" of a granule. We shall return to the significance of this time-scale after we have discussed spatial scales and velocities.

The contrast between Benard cells and solar granules contains important information about fluid dynamics. In Benard cells, the material moves in a steady pattern known as "laminar flow": this is appropriate in conditions where a fluid of relatively high viscosity flows at relatively low speeds. But in solar granules, the gas flows in a highly nonsteady pattern: this is appropriate in "turbulent" conditions where a low-viscosity medium moves faster than a critical speed. The turbulent flows in solar convection are reminiscent of eddies in fast-flowing water: each eddy lives for only a finite time.

6.3 Upflow and Downflow Velocities

For now, we need to note an important correlation which exists in the bright and dark gas. When spectra of individual granules and intergranular

lanes are obtained, it is found that the *brighter* regions are systematically associated with *upflows*, while the *darker lanes* are associated with *downflows*. As far as physics is concerned, the correlation of upflow with excess temperature, and downflow with reduced temperature, is the essential characteristic which indicates that convection is at work, transporting heat *upward* through the solar material. There is a real difference between up and down in the solar atmosphere when it comes to heat: heat flows in the *upward* direction.

As far as the quantitative flux of energy is concerned, an important physical parameter is the algebraic difference in *vertical velocity* between upflows and downflows. Spectroscopic data are required to evaluate these velocities. The average amplitude of this velocity difference is reported to be about 2 km $\sec^{-1}$ (Bray et al., 1976). In some case, amplitudes as large as 6 km $\sec^{-1}$ have also been reported (Beckers, 1968). Although not directly associated with upward energy transport, it is of interest to mention that *horizontal* velocities associated with granules can be measured by tracking algorithms: the rms values are observed to be as large as 1.5 km $\sec^{-1}$ (Title et al.,1989).

We conclude that gas in the solar granulation moves with speeds V of order a few km $\sec^{-1}$.

6.4 Linear Sizes of Granules

It is not a trivial matter to observe granules on the surface of the Sun: they are small features which were not at all apparent to the early telescopic observers of the Sun. The first clear images of granules were not reported for almost three centuries after Galileo turned his small telescope to the Sun. The granules do not become detectable until the resolution of the telescope becomes large enough, and the disturbing effects of Earth's atmosphere are reduced to a minimum.

How high does the angular resolution have to be in order to distinguish clearly the bright and dark areas? The empirical answer is: the observing instrument must be able to resolve angles of about 1 arc sec or better. Observations from favorable locations on the Earth's surface may on occasion satisfy this criterion. Observations from space routinely satisfy the criterion.

What are the horizontal spatial scales associated with granules and intergranular lanes? Some granules have diameters as large as 2–3 arc sec, while others have diameters as small as the limiting resolution of the telescope. Title et al. (1989) conclude: "it is fair to say that there is a characteristic granule size in the vicinity of 1.2–1.4 arc sec." The corresponding characteristic linear dimensions are 900–1000 km. The fact that the size distribution extends to 2–3 arc sec indicates that the largest granules are up to 2000 km in diameter. Observational limitations prevent us from determining the dimensions of the smallest granules.

6.5 Circulation Time around a Granule

Now that we know (1) the horizontal diameter D associated with the top of a granule, and (2) a velocity of the gas flow in the cell, it is of some interest to estimate how much time it takes the gas to circulate around a cell.

To make the estimate, we need to know also the vertical depth H of the granule. Let us estimate the distance which an element of gas travels as it starts at the bottom of the cell, rises to the top (*distance* H), spreads out horizontally to the edge (*distance* $D/2$), sinks to the bottom (*distance* H), and then returns to the center of the cell (*distance* $D/2$). The total distance traversed by the element of gas is $C = D + 2H$. The time required for the gas to complete one circulation is $t(\text{circ}) = C/v$. With $D \approx 1000$ km, and $V \approx 2$ km sec^{-1}, we find $t(\text{circ}) \approx 500 + H$ sec, if H is expressed in km.

What vertical depth should we consider for a convection cell in the Sun? We have already identified a natural length scale which exists in the stratified solar atmosphere: the scale height H_p. We have already seen (Chapter 5, Section 5.1) that $H_p \approx 100$ km in photosphere. It may be that H is related to H_p by a number which is of order unity: $H = \alpha H_p$. With such a choice, we find $t(\text{circ}) \approx 500 + 100\alpha$ sec.

We have already seen that granules live for 300–600 sec. This range of lifetimes overlaps with $t(\text{circ})$ as long as α does not exceed unity by a significant amount. Detailed modeling of the Sun suggests $\alpha \approx 1.5$ (e.g. Mullan et al., 2007). This leads to $t(\text{circ}) \approx 650$ sec.

Thus, it appears that solar granules do not live long enough to complete *even a single complete circulation of the cell.*

This is in significant contrast with Benard cells: under the carefully controlled conditions of a laboratory experiment, convection cells can survive for arbitrarily large numbers of circulation times.

6.6 Temperature Differences between Bright and Dark Gas

The efficiency of heat transport by convection depends on the temperature difference between rising and falling gas. This temperature difference gives rise to the observed differences in intensity ΔI between hot gas (where the intensity is I_h) and cold gas (where the intensity is I_c). Using the highest quality ground-based observations, measurements of the emergent intensity in different locations across the solar surface show that relative to a mean intensity, there are excursions in $\Delta I/I$ of up to $\pm 20\%$ in visible light (Stein and Nordlund, 1998). The rms excursions are about $\pm 10\%$. Measurements from space (Title et al., 1989) suggest that the rms values of $\Delta I/I$ are up to $\pm 16\%$ in quiet Sun. However, when the effects of acoustic oscillations are

allowed for, the rms values of $\Delta I/I$ are found to be $\pm 10\%$ at wavelengths around 6000 Å (Title et al., 1989).

Thus, the bright granules have an rms intensity which is some 10% in excess of the average , while the dark intergranular lanes have an rms intensity which is some 10% smaller than the average.

For purposes of calculating how much energy is transported by convection, we need to convert the intensity difference to a temperature difference between the temperatures in the hot and cold gases, T_h and T_c. When we observe bright granules, we are seeing down into the solar atmosphere essentially to an optical depth of $\tau_h \approx 1$ in the hot rising gas. When we observe dark intergranular lanes, we are seeing into the solar atmosphere to an optical depth of $\tau_c \approx 1$ in the cold sinking gas. Although the optical depths are the same in hot and cold gas, we are not observing gas at the same vertical height: we see more deeply into the cold gas (where the opacity is lower).

Since we see to equal optical depths in both bright and dark gas, the emergent intensities I_h and I_c are proportional to the respective source functions S_h and S_c at optical depth unity. Thus, $\Delta I/I = \Delta S/S$ to a rough approximation. To the extent that the continuum source function at $\tau \approx 1$ can be equated with the Planck function, we expect that

$$\frac{I_h}{I_c} \approx \frac{\exp(c_2/\lambda T_c) - 1}{\exp(c_2/\lambda T_h) - 1} \tag{6.1}$$

where $c_2 = hc/k$ is the second radiation constant: $c_2 = 1.44$ cm deg.

In view of 10% fluctuations in rms intensity, we have that $I_h \approx 1.1$ while $I_c \approx 0.9$. For purposes of this calculation, we assume that the dark intergranular gas has a temperature at $\tau \approx 1$ equal to the effective temperature, i.e., $T_c = 5777$ K. Then we find that, according to Equation 6.1 (assuming an observing wavelength $\lambda = 5500$ Å), the "rms" temperature for the hot gases is $T_h \approx 6050$ K. Thus, in the black-body approximation, the rms temperature difference ΔT between bright granules and dark intergranular material at $\tau \approx 1$ in both materials is roughly 300 K.

More careful treatment of radiative transfer, using a fully 3-D radiative-hydrodynamic code (Stein and Nordlund, 1998), indicates that the temperatures at $\tau \approx 1$ in the coldest and hottest elements of gas can reach extreme values which range from 5800 to 7000 K. Thus the extreme temperature differences at $\tau \approx 1$ is 1200 K. The rms temperature differences are smaller than this extreme value: from the results of Stein and Nordlund we estimate $\Delta T_{\rm rms} \approx 500$–$600$ K. Thus, our rough estimates of $\Delta T = 300$ K at $\tau \approx 1$ using the black-body approximation to interpret the observed intensity fluctuations are too small by a factor of about two.

As we move upward and downward from the level $\tau \approx 1$, the results of 3-D modeling (Stein and Nordlund, 1998) indicate that the range of temperature differences ΔT increases. In the upper atmosphere, the extreme range is 1500 K at $\tau \approx 0.001$, while in the deeper layers, at $\tau \approx 1000$, the extreme ranges of ΔT may be as large as 2500 K.

If we were to compare the rising and sinking gas at equal geometric depths (rather than at equal optical depths), the temperature differences would be larger than the above estimates. The modeling of Stein and Nordlund (1998) indicates that at equal depths close to the photosphere, the extremes of temperature difference may be as large as 4000 K. However, the extremes of temperature differences are confined to relatively small spatial regions: they arise mainly in certain narrow downdraft columns of gas which plunge downward after losing a lot of energy at the surface.

As a large-scale spatial average of temperature differences between rising and falling gas, it is more appropriate, for our approximate estimates, to adopt the values of $\Delta T_{\rm rms} \approx 500$–$600$ K.

6.7 Energy Flux Carried by Convection

Now that we know how fast the gas is moving and how much temperature difference exists between the hot rising material and the cold sinking material, we can turn to a consideration of the key question which is relevant for solar physics: how much heat energy is being carried upward by the gas motions?

In a parcel of gas, a temperature difference of ΔT corresponds to a difference in the heat content of $C_p \Delta T$ ergs gm^{-1}. Here, C_p is the specific heat at constant pressure, in ergs gm^{-1} K^{-1}. This is the excess amount of internal heat that the hot rising gas contains compared to the cool sinking gas.

Now, to calculate the upward flux of energy, in units of ergs cm^{-2} sec^{-1}, we need to multiply the heat content (in ergs gm^{-1}) by the mass flux, $F_m = \rho V$(in gm cm^{-2} sec^{-1}) . This leads us to the upward heat flux due to convection:

$$F(\text{conv}) = \rho V C_p (\Delta T) \tag{6.2}$$

This formula is applicable in any region of the Sun where bulk gas motions are present.

6.7.1 Convective energy flux in the photosphere

To start off a discussion of convective energy transport, we consider the particular case of a region in the Sun where we already know the magnitudes of all the parameters which enter into the above formula: this particular region is the photosphere. Subsequently, we shall consider deeper lying layers.

Let us see what is the magnitude of this energy flux in the photosphere. The model of the solar atmosphere which we derived in Chapter 5 indicates that in the photosphere, $\rho \approx (2\text{–}3) \times 10^{-7}$ gm cm^{-3}. The discussion given in the earlier sections of the present chapter suggests that velocity differences between hot and cold gas are of order a few km sec^{-1}, say $V \approx 3{\times}10^5$ cm sec^{-1}, and that $\Delta T \approx 500$ K. The final quantity that we need to evaluate in order to estimate the convective flux is C_p.

To evaluate the specific heat, we note that in a perfect gas, where the particles are monatomic, and are not undergoing ionization, the internal energy of an atom consists of a single term, due to thermal motion, i.e., $(3/2)kT$ per atom, where k is Boltzmann's constant. In a gas composed of hydrogen atoms only, the internal energy per gram is $U = (3/2)kT/m_H = (3/2)R_gT$ ergs per mole, where $R_g = k/m_H$ is the gas constant. For a gas mixture with mean molecular weight μ, we find $U = (3/2)R_gT/\mu$ per gram. This leads to a specific heat per gram at constant volume

$$C_v = \frac{dU}{dT} = \frac{3R_g}{2\mu} \tag{6.3}$$

The specific heat at constant pressure, $dU/dT + p(dV/dT)_p$ (where subscript p denotes constant pressure) is given by

$$C_p = C_v + \frac{R_g}{\mu} = \frac{5R_g}{2\mu} \tag{6.4}$$

For the gas in the solar photosphere, consisting of a non-ionizing mixture of H (90%) and He (10%), we have $\mu \approx 1.3$. This leads to $C_p \approx 1.6 \times 10^8$ ergs gm^{-1} deg^{-1}.

In a monatomic, nonionizing gas, the ratio of specific heats $\gamma = C_p/C_v$ has the numerical value of 5/3. In terms of γ, the value of C_p can be written as

$$C_p = \frac{\gamma}{\gamma - 1}\frac{R_g}{\mu} \tag{6.5}$$

Combining Equation 6.2 with the various parameters, we find that in the photosphere, the numerical value of the flux of energy which is being transported by convection is

$$F(\mathrm{conv, ph}) \approx 7 \times 10^9 \text{ ergs cm}^{-2} \text{ sec}^{-1} \tag{6.6}$$

Is this a large quantity or a small one? Well, large and small are relative terms. What should we compare the convective flux to in order to decide whether it is "large" or "small"? The relevant flux which passes thru the Sun's atmosphere is the flux of energy $F_\odot$ which eventually leaves the Sun and travels out into space. In radiative terms, we have already specified (Chapter 1, Section 1.8) what this is: $F_\odot = \sigma T_{\mathrm{eff}}^4 = 6.3155 \times 10^{10}$ ergs cm^{-2} sec^{-1}.

Comparing $F(\mathrm{conv}, ph)$ with $F_\odot$, we can see that in the photosphere, convection is carrying some 10–15% of the total energy flux which passes upward through the solar atmosphere. This shows us that although radiation dominates in the process of transporting the energy flux up though the photosphere, radiation is not the *only* process which contributes to transporting energy though the photosphere. Gas motions are *also* of material assistance in the photosphere.

6.7.2 Convective energy above the photosphere?

Once we move away from the particular case of the photosphere, and consider gas which lies either shallower or deeper in the Sun, we can expect that $F(\mathrm{conv})$ will take on different values from that at the photosphere.

In what sense will the numerical values of $F(\text{conv})$ change as we move upward into shallower layers? Well, we have already seen (Chapter 5) that the density of the gas falls off exponentially with increasing height. Moreover, as the density falls off, it becomes increasingly difficult for the temperature difference between rising and sinking gas to be maintained: leakage of photons in the increasingly rarefied gas has the effect that the rising and sinking gases tend toward the same temperatures. As a result, ΔT decreases in the upper photosphere. In view of Equation 6.2, this, combined with the decrease in density, causes $F(\text{conv})$ to *decrease* as we examine gas which lies above the photosphere.

At a height of, say, 300 km above the photosphere, the density has fallen to about 0.1 times the photospheric density. Due to this factor alone, the local value of $F(\text{conv})$ would fall to a value of only 1–2% of the radiative flux. Allowing for the radiative leakage, $F(\text{conv})/F_{\odot}$ in fact falls below 1% at heights of a few hundred km above the photosphere.

6.7.3 Convective energy flux below the photosphere

What about the deeper gas? How large is the convective flux there? In these layers, densities increase exponentially rapidly as the depth increases. Furthermore, the temperature differentials between rising and sinking gas ΔT are larger than in the photosphere (see Section 6.6). Both of these factors contribute to an *increase* in the convective flux as we go deeper beneath the photosphere.

Moreover, there is a further factor, associated with ionization, which helps the gas to transport convective flux more easily. As the temperature rises in the deeper gas, the atoms begin to experience an increasing amount of ionization; the gas enters into one of the "ionization strips" in Figure 4.1. When ionization is in process, the internal energy of an atom of the gas U is no longer due solely to thermal motions: instead, there is an extra term associated with the ionization potential energy I. For hydrogen, $I = 13.6$ eV. This is larger by more than an order of magnitude than the thermal energy: at temperatures corresponding to gas just below the photosphere (where $T = 6000$–$10{,}000$ K), the thermal energy per atom is only of order 0.6–1 eV. The occurrence of ionization energy of 13.6 eV represents the addition of such a large temperature-sensitive contribution to internal energy U that the numerical value of specific heat $C_v (= dU/dT)$ increases significantly compared to the values cited above for a monatomic (and nonionizing) gas. The specific heat reaches a maximum when the gas is roughly 50% ionized: as we noted in Chapter 4, this occurs for hydrogen at depths where the temperature is about 20,000 K. In the vicinity of 50% hydrogen ionization, the value of C_v rises to ≈ 36 times its "normal" value $[(3/2)R_g/\mu]$, while the value of C_p rises to ≈ 27 times its "normal" value $[(5/2)R_g/\mu]$ cited above for a monatomic nonionizing gas.

When the gas is undergoing ionization, and C_p and C_v both increase in value by significant amounts, the ratio of specific heats $\gamma = C_p/C_v$ is no longer as large as 5/3 (the value for a nonionizing monatomic gas). Moreover, in

adiabatic conditions, the pressure and density are no longer related by the simple relationship $p \sim \rho^{\gamma}$. Instead, the pressure and density are related by $p \sim \rho^{\Gamma}$, where the numerical value of the generalized exponent Γ is no longer strictly equal to C_p/C_v. When the degree of ionization is 50%, Γ for a gas composed of hydrogen alone falls to minimum values of ≈ 1.135 (Clayton, 1968). In the Sun, where hydrogen is not the sole constituent, but helium contributes almost 10% by number, the minimum value of Γ is not so small: when hydrogen is 50% ionized, helium is still essentially neutral, with $\Gamma(\mathrm{He}) = 5/3$. The combination of (roughly) 90% of the atoms with $\Gamma(\mathrm{H}) = 1.135$ and (roughly) 10% of the atoms with $\Gamma(\mathrm{He}) = 5/3$ lead to an overall minimum value of $\Gamma(\mathrm{H} + \mathrm{He})$ of 1.19. We will return to this minimum value of Γ when we consider the calculation of a model of the convection zone in Chapter 7.

In even deeper layers, where hydrogen ionization is approaching completion, the ionization energy becomes less important in the internal energy. Thermal energy once again dominates. As a result, the specific heat reverts (almost) to the value cited above. But now there is a difference: for every hydrogen atom in the photosphere, there are now two particles at great depth (a proton and an electron), each with its own equal share of thermal energy. As a result, the internal energy per gram is *twice as large* as in the photosphere.

Combining the increased factors of ρ, C_p, and ΔT, it is clear that $F(\mathrm{conv})$ must increase rapidly as we go below the surface. The only possibility for off-setting this rapid increase in convective flux would be for the velocity V to undergo a dramatic reduction as depth increases. However, this does not seem likely: convective models indicate that V remains of order 1 km sec^{-1} even at depths as great as 1000 km below the photosphere. The models indicate that at depths of ≥ 100 km below the photosphere, convection is able to carry more than 90% of the total energy flux.

The bulk motion of gas provides a very efficient method for the Sun to transport energy in the layers of gas which lie not far beneath the photosphere.

Now that we have seen how effective convection is for energy transport below the photosphere, this raises the question: does convection dominate the transport of energy within the entire interior of the Sun? In order to arrive at an answer to this question, we need to understand the physical causes of convection. We now turn to those causes.

6.8 Onset of Convection in the Sun: The Critical Gradient g_{ad}

Let us perform the following thought experiment. Consider an atmosphere in which T is increasing as the depth increases. In the context of optical depth, we have already found such a case when we discussed radiative equilibrium in a grey atmosphere: $T^4 \sim (\tau + 2/3)$. Once we know the opacity, we can

convert this into the functional form which indicates how T varies with the linear depth z. This was part of the solution of the solar atmosphere which we obtained in Chapter 5.

In order to set the stage for a discussion of convection, it is helpful to consider how T varies as a function of z. At each depth, the slope of T versus z has a certain numerical value: we refer to that slope as the local temperature gradient $g_T = dT/dz$ in units of degrees per cm. Because we are considering an atmosphere in which T increases as z increases, the sign of g_T is positive. This is important for understanding the possibility for convection to occur.

Consider a parcel of gas which lies initially at a depth z. The gas has a well-defined temperature T. The question we ask is: is this parcel of gas stable or unstable when we displace it vertically? To answer this question, we consider the change in energy which occurs as a result of the displacement. The algebraic sign of this energy change plays a key role in what follows.

Suppose we displace the parcel upward along a vertical path which has length dz. The local ambient temperature at the new depth $z - dz$ is $T - dT$ where $dT = |g_T\, dz|$. Let the upward displacement of the parcel to the final depth $z - dz$ be performed in a time that is so short that the parcel has no time to lose any of its internal energy by leakage to the ambient gas. When the parcel arrives at $z - dz$, it will still have its initial temperature T: therefore the parcel finds itself hotter than its surroundings by dT. That is, the parcel will contain internal thermal energy which is in excess of that in the ambient gas. Now let enough time elapse that the parcel releases all of its excess thermal energy into the local gas under conditions where pressure is maintained constant: the amount of thermal energy which it will release is given by $C_p\ dT$ ergs gm^{-1}.

Is this energy release significant as far as the displacement of the parcel is concerned? To answer that, we must compare the amount of thermal energy which has been released with another energy term which arises as a result of the displacement: gravitational potential energy. In order to displace the parcel upward by the linear amount $-dz$, work has to be done against gravity. The amount of that work is $-gdz$ per gram of material.

Now we ask: what has happened to the total energy of the parcel in the course of its displacement to its final position? On the one hand, the gas has released $C_p dT$ ergs gm^{-1}. On the other hand, energy had to be found (from somewhere) to increase the potential energy by $-gdz$ ergs gm^{-1}. The total amount of energy ΔW associated with the displacement of 1 gm of material is therefore given by

$$\Delta W = -gdz + C_p dT \tag{6.7}$$

The key point to note in Equation 6.7 is the possibility that ΔW takes on numerical values which can be either positive or negative.

For example, suppose that the magnitude of gdz exceeds the magnitude of $C_p dT$. In this case, ΔW is negative: in order to displace the parcel upward

to its final position, we have had to supply *more* work than is released by the thermal excess. There is not enough thermal energy released to compensate for the work that was done in order to lift the parcel. As a result, the parcel will tend to sink back down to its initial position. There is no incentive for the parcel to move upward. This is referred to as "convective stability."

But in the opposite limit, suppose that the magnitude of $C_p dT$ exceeds the magnitude of gdz. In this case, release of thermal excess at the end of the displacement is *more than enough* to compensate for the work of lifting the parcel. All of that lifting work can be provided for by releasing the excess internal energy of the parcel. In fact, after the work of lifting has been performed, there is even some internal energy left over to make sure that the parcel remains hotter than the ambient gas. In other words, the gas itself contains more than enough internal energy to do the work of lifting the parcel against gravity. This is an unstable situation: the parcel, once displaced, keeps on moving upward. We refer to this as "convective instability."

The boundary between stability and instability as far as convection is concerned occurs when ΔW has a value which is neither positive nor negative, i.e., when the total energy exchange between the parcel and its surroundings is zero. In such a situation, the parcel undergoes a change which is adiabatic. This particular case occurs when $gdz = C_p dT$, i.e., when the temperature gradient $g_T \approx dT/dz$ takes on the particular value known as the "adiabatic temperature gradient":

$$g_T = g_{ad} \equiv g/C_p \tag{6.8}$$

Gas in which the local temperature gradient has a particular value g_T is convectively *stable* if $g_T < g_{ad}$. In gas where the opposite holds, i.e., where $g_T > g_{ad}$, the gas is convectively *unstable*.

This reminds us that the algebraic sign of g_T is important: on the right-hand side of Equation 6.8, the quantities g and C_p are both positive definite. If g_T is a negative quantity, i.e., if the temperature *decreases* as the linear depth *increases*, it is *impossible* to satisfy the condition for convective *instability:* $g_T > g_{ad}$. As a result, such gas is always convectively *stable*.

The conclusion of the present section is the following. The onset of convection is determined by the answer to the question: is the local temperature gradient dT/dz larger or smaller than the value of the adiabatic temperature gradient g_{ad}?

6.9 Numerical Value of g_{ad}

In the solar photosphere, we have seen (Chapter 1, Section 1.13) that $g = 27422$ cm sec^{-2}. And in Section 6.7.1, we have seen that $C_p \approx 1.6 \times 10^8$

ergs gm^{-1} in the photosphere. Combining these values, we find the important result

$$g_{ad} \approx 1.7 \times 10^{-4} \mathrm{deg\ cm}^{-1} \tag{6.9}$$

as the numerical value of the critical temperature gradient which must be exceeded if convection is to occur in the solar photosphere.

Now it is remarkable that, when we were calculating a model solar atmosphere in Chapter 5, we eventually did reach a layer of gas (with optical depth $\tau = 0.9$–1.0, i.e., near the photosphere) where the local temperature gradient dT/dz increased to a numerical value in excess of the above critical gradient g_{ad}. This means that, in the model presented in Table 5.3 in Chapter 5, the gas is convectively stable for all levels of the photosphere which are listed in the table. According to the model, convection in the Sun sets in at optical depth $\tau \geq 0.9$–1.0, i.e., just below the photosphere.

Why is this remarkable? It means that we Earth-based observers are lucky enough to see down into the Sun deep enough to catch a glimpse of at least the uppermost layers of convection (see Figure 6.1). There is nothing to say *a priori* that this *must* happen: it is certainly possible that the onset of convection might have occurred so deep below the photosphere that Earth-based observers would be able to see nothing whatsoever of the convective motions. (For example, if we lived near a hot star of spectral class O or B, we would see no evidence for convection, which is confined to the innermost core of the star.)

As it is, we Earthlings *are* able to see the Sun's convection, with its up-and-down gas motions of hot and cold gas. Without this privilege, we might know a lot less about convection not just in the Sun, but in other stars also.

Below the photosphere, where hydrogen begins to undergo appreciable ionization, the rapid increase of C_p (by factors which may be as large as ten or more), has the effect that g_{ad} takes on values which are numerically much smaller (by a factor of order 10) than in the photosphere. Because of this, it is much easier for the local temperature gradient to exceed the local value of g_{ad} when the ambient gas is undergoing ionization. As a result, it is much easier to satisfy the convective instability condition $g_T > g_{ad}$. Therefore, in a region where gas is undergoing ionization, we are likely to find convection.

6.10 Alternative Expression for g_{ad}

As mentioned above, an alternative expression for C_p is $[\gamma/(\gamma-1)]R_g/\mu$. Using this, we can rewrite g_{ad} as

$$g_{ad} \equiv \left(\frac{dT}{dz}\right)_{ad} = \frac{g\mu(\gamma-1)}{\gamma R_g} \tag{6.10}$$

Now, for a medium which is in HSE, we also know that the pressure gradient is given in Chapter 5 by Equation 5.1. In the present context, where we are using the independent variable z (i.e., the depth), rather than (as in Chapter 5, Equation 5.1) the height h, the HSE equation is written as

$$\frac{dp}{dz} = +g\rho \tag{6.11}$$

Dividing Equation 6.10 by Equation 6.11, we find that

$$\left(\frac{dT}{dp}\right)_{ad} = \frac{1}{\rho C_p} = \frac{(\gamma - 1)\mu}{\gamma \rho R_g} \tag{6.12}$$

Among the terms on the right-hand side, we note that for a perfect gas, $\mu/\rho R_g$ equals T/p. Carrying this over to the left-hand side of the equation, we find

$$\left(\frac{d\log T}{d\log p}\right)_{ad} = \frac{\gamma - 1}{\gamma} \tag{6.13}$$

We note that if the local conditions in a gas in any region of the Sun are in fact adiabatic, then the local temperature and pressure will vary in such a way that the local gradient of temperature relative to pressure, $d\log T/d\log p$, will take on the value $(d\log T/d\log p)_{ad}$ as given by Equation 6.13. This has the effect that the pressure p in that region of the Sun will vary as a power law of T. In the presence of ionization, we need to replace γ in Equation 6.13 with the more generalized exponent Γ:

$$p \sim T^{\Gamma/(\Gamma-1)} \tag{6.14}$$

In a monatomic nonionizing gas, where $\Gamma = \gamma = 5/3$, the right-hand side of Equation 6.13 has the numerical value 0.4. In such a gas, adiabatic processes lead to a pressure-temperature relationship of the form $p \sim T^{2.5}$.

But if ionization is at work, the power law relationship becomes steeper. For example, in a gas composed of pure hydrogen where the degree of ionization is 50%, $\gamma \approx 1.135$. In such a case, Equation 6.13 indicates that $p \sim T^8$; in such conditions, small increases in temperature would be associated with much larger increases in pressure than in the nonionizing limit $p \sim T^{2.5}$. These results will be applied to the solar convection zone in the next chapter.

Exercise

6.1 Consider flows at the surface of the Sun $h = 0$ with vertical speeds of 1, 3, 6, and 10 km sec^{-1}. Given the gravity at the surface of the Sun (Chapter 1, Equation 1.13), calculate the maximum heights to which these flows can reach above the surface.

References

Beckers, J. M. 1968. "High-resolution measurements of photospheric and sunspot velocity and magnetic fields using a narrow-band birefringent filter," *Solar Phys.*, 3, 258.

Benard, H., 1900. "Les tourbillons cellulaires dans une nappe liquide," *Rev. Gen. des Sciences Pures Appl.*, 11, 1261 and 1309.

Bray, R. J., Loughhead, R. E., and Tappere, E. J. 1976. "Convective velocities derived from granule contrast profiles in FeI at 6569.2 Å," *Solar Phys.,* 49, 3.

Clayton, D. D. 1968. "Thermodynamic state of the stellar interior," *Principles of Stellar Evolution and Nucleosynthesis.* McGraw-Hill, New York, pp. 77–165.

Mullan, D. J., MacDonald, J., and Townsend, R. D. H. 2007. "Magnetic cycles in the Sun: Modeling the changes in radius, luminosity, and *p*-mode frequencies," *Astrophys. J.*, 670, 1420.

Stein R. F. and Nordlund, A. 1998. "Simulations of solar granulation. I. General properties," *Astrophys. J.*, 499, 914.

Title, A. M., Tarbell, T. D., Topka, K. P., Ferguson, S. H., Shine, R. A., and the SOUP Team. 1989. "Statistical properties of solar granulation derived from the SOUP instrument on Spacelab 2," *Astrophys. J.*, 336, 475.

Chapter 7

Computing a Model of the Sun: The Convection Zone

In this chapter, we wish to calculate the structure of the region in the Sun where convection dominates the transport of energy. As in Chapter 5, we will not discuss the *origin of the energy which is flowing through the convection zone.* Instead, we again accept the total luminosity (or flux) of the Sun as a boundary condition, and seek to determine how the material arranges itself so as to "handle" the energy which is passing through. We will examine the forces which act on the medium, and determine how the medium responds. In this sense, the model we will derive is better referred to as a mechanical model rather than a complete model.

Based on empirical evidence, the gas in the photosphere of the Sun is moving (up and down) with speeds of a few km $\sec^{-1}$. In order to determine the equations which will allow us to describe solar convection in plausible physical terms, we need first to understand why the convective motions in the surface layers of the Sun have speeds of this order of magnitude. Why are the motions not of order a few cm $\sec^{-1}$? Or hundreds of km $\sec^{-1}$? What is the determining factor which sets the scale of the speeds?

7.1 Quantifying the Physics of Convection: Vertical Acceleration

We have seen that certain parcels of gas in the Sun are observed to be rising, while others are sinking. The rising parcels are hotter than the sinking ones, and the rms temperature differences are of order $\Delta T \approx 500$–600 K in the photosphere.

From a physics perspective, it is important to note that the speeds of convective motion are less than the local (adiabatic) speed of sound, $c_s = \sqrt{(\gamma R_g T/\mu)}$, where γ is the ratio of specific heats and μ is the mean molecular weight. (In the solar photosphere, $c_s \approx 9$ km $\sec^{-1}$.) This has the effect that sound waves can propagate quickly between hot and cold gas and equalize the pressures. Thus, the differences in pressure between hot and cold gas at any height are not significant. Now, for material which obeys the equation of

state of a perfect gas, the pressure difference Δp is related to the temperature difference by $\Delta p/p = \Delta T/T + \Delta\rho/\rho - \Delta\mu/\mu$. In the photosphere, there is no significant difference in the degree of ionization between hot and cold gas: therefore there is no significant difference in the molecular weights, and we may set $\Delta\mu/\mu = 0$.

Using this, and setting $\Delta p/p = 0$, we see that the observed temperature difference ΔT between rising and sinking gas in the photosphere corresponds to a density difference $\Delta\rho/\rho = -\Delta T/T$. The negative sign indicates that the hotter (rising) gas has lower density than the cooler (sinking) gas. With a fractional temperature difference in the photosphere observed to be $\Delta T/T \approx (500\text{–}600)/5800 \approx 0.1$, we see that hotter gas has a density which is about 10% smaller than the density of the cooler gas.

Now the photosphere of the Sun is for the most part in hydrostatic equilibrium: this means that there are no net forces acting on the gas in the photosphere. This is not to say that there are no forces whatsoever acting on the gas: it means only that whatever forces *are* at work, they are in general balanced in the photosphere. On the one hand, there is an upward force (per unit volume) due to the vertical pressure gradient dp/dz. On the other hand, there is a downward force (per unit volume) due to the weight of the gas, ρg. When these forces are in balance, there is no net acceleration, and the gas remains at rest. This is the situation throughout the model of the photosphere which was presented in Chapter 5. That is, given a photospheric model where, at depth z, the density is ρ_o and the pressure is p_o, then dp_o/dz has a numerical value which is precisely equal to $\rho_o g$ at all heights in the photospheric model:

$$\frac{dp_o}{dz} = g\rho_o \quad (HSE)$$

This is the equation of HSE (cf. Chapter 5, Equation 5.1), rewritten in terms of the depth z (which increases *downward*) rather than the height parameter h (which increases *upward*).

But in a gas where convection is possible, the forces are no longer balanced. Let us consider the imbalanced forces and the accelerations which they cause.

Suppose a certain parcel of gas is hotter than the ambient medium. The density ρ' in the parcel will be *lower* than the ambient density ρ_o. As a result, the downward force on unit volume of gas in the parcel due to its weight $\rho' g$ is now *less* than the local upward force due to pressure dp/dz. The unbalanced force $dp/dz - \rho' g$ acting on a parcel of gas with unit volume leads to an acceleration of that parcel in the upward direction. Since unit volume of the gas has a mass of ρ', Newton's second law of motion (force=mass times acceleration) tells us that the unit volume will be subject to an upward acceleration dV/dt such that

$$\rho' \frac{dV}{dt} = \frac{dp}{dz} - g\rho' \tag{7.1}$$

This equation expresses the conservation of momentum. Notice that in the absence of flows ($V = 0$), conservation of momentum reduces to HSE.

As we have seen, the motions which are observed in the solar photosphere are such that pressure remains equalized between hot and cold gas. That is, the pressure of the gas remains relatively unchanged in hot or cold gas compared to the ambient medium. This means that we can, without serious error, replace dp/dz by dp_o/dz. But we already know that $dp_o/dz = \rho_o g$. Therefore, the upward acceleration dV/dt experienced by the low-density gas parcel is given by $(dV/dt)_u = g(\rho_o - \rho')/\rho'$. The fact that ρ' is less than ρ_o has the effect that the sign of the right-hand side is positive. Therefore, the acceleration is in the *upward* direction. Buoyancy forces create this upward acceleration.

If a parcel of gas is locally cooler (and denser) than ambient, with density $\rho'' > \rho_o$, then the acceleration will be downward, with a magnitude $(dV/dt)_d = g(\rho_o - \rho'')/\rho''$. Again, the effect is due to buoyancy.

Since the differences in density between the ambient medium and the hot (upgoing) and cold (downgoing) gas are not large, we can write the relative acceleration a_{hc} between hot and cold gas as

$$a_{hc} \equiv \left(\frac{dV}{dt}\right)_u - \left(\frac{dV}{dt}\right)_d = \frac{g(\rho'' - \rho')}{\rho_o} \approx \frac{g\Delta\rho}{\rho} = -g\frac{\Delta T}{T} \qquad (7.2)$$

Inserting the empirical result $\Delta T/T \approx 0.1$, we find that the magnitude of the relative acceleration a_{hc} between hot and cold gas in the photosphere is given by $a_{hc} \approx 0.1g \approx 2.7 \times 10^3$ cm sec^{-2}.

7.2 Velocities and Vertical Length Scales

Now that we have an estimate for the relative vertical acceleration between hot and cold gas, we can ask: over what vertical length scale s_v must the acceleration a_{hc} be allowed to operate in order to build up a vertical velocity difference V which is comparable to the observed values, i.e., a few km sec^{-1}?

The relevant formula is $V^2 = 2a_{hc}\, s_v = 2s_v g\Delta T/T$. Setting $V = (2-3) \times 10^5$ cm sec^{-1}, we find $s_v = 74$–167 km. Thus, if the buoyancy forces due to the density differences between hot and cold gas in the solar photosphere are allowed to operate over distances of 74–167 km, the vertical velocities which can be produced are comparable to the observed values.

Is there any physical significance to lengths scales of 74–167 km? Well, we have seen (Chapter 5, Section 5.1) that the pressure (and density) scale height H_p in the photosphere is 115–140 km. We note that this range of H_p values is completely contained within the range of values for the vertical distance s_v. This suggests that the dynamics of convection in the solar photosphere are constrained in such a way that the vertical acceleration due to buoyancy is allowed to operate over vertical length scales which are comparable to H_p. Specifically, with the above numbers, it appears that $s_v \approx (0.5\text{–}1.5)H_p$.

This is an empirical conclusion. It is based on the *observed* temperature differences between hot and cold gas, and on the *observed* relative differences in velocity between rising and sinking gas. If we were not able to resolve the granulation in the Sun, thereby measuring differences in temperature and velocity between rising and sinking gas, we would have to rely on indirect arguments in order to decide what might be the best choice for s_v.

7.3 Mixing Length Theory (MLT) of Convection

It was mentioned earlier (Chapter 6, Section 6.2) that granules in the Sun have properties which are similar to eddies in a fast-flowing river: such eddies survive for a finite time and then dissolve into the ambient water. During their life-time, they travel a finite distance, before they "mix" their contents back into the river. This finite distance is called the "mixing length."

In solar convection, by analogy, a parcel of convective flow can preserve its identity for a finite time only. During that time, the material travels a finite distance L (the "mixing length") in the vertical direction (buoyancy forces determine that the motion is preferentially vertical), and then mixes in with the ambient gas. Based on the discussion in the previous section, it seems plausible to equate L with the vertical distance s_v, which is (as we have seen) of order the local H_p. In solar convection, a "mixing length parameter" α is defined as the ratio between the mixing length and the local scale height: $\alpha = L/H_p$. Based on the discussion in Section 7.2, an appropriate choice for α in the solar photosphere is 0.5–1.5. Again, this is an empirical conclusion, based on measured velocities and temperature differences.

Is there any theoretical reason why the mixing length might be expected to be of order H_p? Well, when a parcel of gas starts its upward "lifetime" at depth z, it has a density which is only slightly smaller than that of the ambient gas. Once the parcel has risen to a new depth $z - L$, it finds itself in lower density gas. When the parcel has time to adjust its density to the ambient value at depth $z-L$, the parcel must expand in volume. If the vertical distance L were as large as, say, $2.1H_p$, the parcel would find itself at the top of its path (at depth $z - L$) in a medium with an ambient density which is $e^{2.1} \approx 8$ times smaller than the initial. This would lead to an eight-fold increase in the parcel's volume, along with a four-fold increase in surface area. As a result, if at depth z at any instant of time, a snapshot of the gas at that depth showed the aggregate of all rising parcels occupying, say, 10–25% of the available surface area, then at depth $z - L$, the parcels would have expanded to occupy 40–100% of the available surface area. There would be no more room for further expansion.

This leads us to suspect that the existence of the empirical limit $\alpha = 0.5$–1.5 may be related to a self-regulating process: there is simply not enough

room for parcels which would expand, in the course of their lifetime, to ten or more times their initial volume. Moreover, we have already mentioned (Chapter 6, Section 6.5) that $\alpha = 1.5$ seems to be the best choice to describe the Sun.

7.4 Temperature Excesses Associated with MLT Convection

As a check on the plausibility of choosing $\alpha \approx 1$, let us estimate how large the temperature excess is expected to become between rising gas and the ambient medium. In other words, what is the temperature excess relative to ambient after a parcel of gas has traveled a length L?

To answer this question, consider a parcel of gas which rises from an initial depth z to a new (upper) depth $z - L$, and rises so fast that it preserves its initial temperature along the way. At the upper depth, if the gas in the parcel were not called upon to perform any work, the parcel would have a temperature in excess of the ambient temperature by an amount $\Delta T_o = Lg_o$. (Here, $g_o = dT/dz$ is the local temperature gradient in the ambient medium.)

Now, our discussion of the adiabatic gradient in Chapter 6 shows that, in a convective region, some of the internal energy of the parcel of gas is used to do the work of raising the parcel a distance L against gravity. Specifically, the work against gravity, i.e., gL per gram, can be performed by extracting the amount $C_p \Delta T_{ad} = Lg$ from the internal energy per gram of the gas. (Note, $g_{ad} = g/C_p$ is the adiabatic temperature gradient.)

As a result, when the parcel reaches its upper position, $z - L$, it finds itself with a temperature that exceeds the ambient by an amount ΔT which is not as large as the ΔT_o mentioned above. Instead, the temperature excess ΔT is given by the reduced quantity $\Delta T = \Delta T_o - \Delta T_{ad}$. Expressing ΔT_o and ΔT_{ad} in terms of the temperature gradients, we can write

$$\Delta T = L(g_o - g_{ad}) \equiv L(\Delta g_T) \tag{7.3}$$

Here we define the Δg_T as the superadiabatic gradient, i.e., the amount by which the ambient temperature gradient g_o exceeds the adiabatic gradient g_{ad}.

Let us estimate the numerical value of the temperature excess ΔT. We have already seen that in the solar photosphere, g_{ad} has a numerical value of about 1.7×10^{-4} deg cm^{-1}. In regions of vigorous convection near the photosphere, the local temperature gradient g_o may exceed g_{ad} by an amount which is not necessarily small. There is no reason to exclude the possibility that that g_o could exceed g_{ad} by an amount which is comparable to g_{ad} itself. This suggests that Δg_T could have a value of order 10^{-4} deg cm^{-1}. In such a case, and setting $L \approx H_p \approx 10^7$ cm, we find that the mixing length theory predicts $\Delta T \approx 10^3$ K in the photosphere.

How does this compare with the temperature differences which exist in the solar granulation? We have seen, from rough analysis of the empirical brightness fluctuations in the granulation, that the rms temperature differences are estimated to be in the range 500–600 K. These are consistent, within factors of two, with the above estimate.

It seems that estimates of temperature excesses based on MLT are not inconsistent with empirical data by significant amounts.

The fact that the velocities of solar granulation, as well as the temperature differences between hot and cold gas, can be replicated, at least roughly, in the context of MLT suggests that the theory can be of service when we attempt to model the complexities of turbulent solar convection.

7.5 MLT Convective Flux in the Photosphere

In the context of MLT, the convective heat flux $F(\text{conv}) \approx \rho V C_p \Delta T$ (see Equation 6.2) can be expressed as follows. We replace V by the expression $\sqrt{(2Lg\Delta T/T)}$, and we replace ΔT by the expression $L\Delta g_T$. This leads to

$$F(\text{conv}) \approx \rho C_p \sqrt{\frac{2g}{T}} L^2 (\Delta g_T)^{3/2} \tag{7.4}$$

Near the photosphere, substitution of appropriate quantities ($\rho \approx (2-3) \times 10^{-7}$gm cm^{-3}, $T \approx 6000$ K, $L \approx 10^7$ cm, $\Delta g_T \approx 10^{-4}$ deg cm^{-1}) leads to $F(\text{conv}) \approx (1-2) \times 10^{10}$ ergs cm^{-2} sec^{-1}. This result is consistent, within a factor of two, with the estimate given in Chapter 6 (Section 6.7.1).

7.6 MLT Convective Flux below the Photosphere

It is of great interest to determine what happens to $F(\text{conv})$ as we examine material that lies deeper inside the Sun. As we go beneath the surface, temperatures increase greatly, and the mean molecular weight decreases (by a factor of about two). As a result, $L = \alpha H_p \sim T/\mu$ increases to values which are much greater than those near the surface. In regions of the solar interior where the gas has $T \approx 10^6$ K, H_p approaches values as large as 10^{10} cm. In such gas, the density is also much larger than the photospheric value: ρ approaches values of order 1 gm cm^{-3}. Moreover, C_p increases above the surface values by a factor of at least two. Let us see how these values affect the expression for $F(\text{conv})$.

The value of $F(\text{conv})$ cannot exceed the overall flux of energy which emerges from the deep interior of the Sun. At the surface, this flux $F_\odot$ equals

6.3155×10^{10} ergs cm^{-2} sec^{-1} (Chapter 1, Section 1.8). As we go inward into the Sun, the surface area decreases, but the total power from the Sun (the "luminosity") remains constant. By the time we reach a radial location of $0.7R_\odot$, the energy flux increases to about 10^{11} ergs cm^{-2} sec^{-1}.

Returning to Equation 7.4, this leads us to ask: how large must the superadiabaticity Δg_T be in order to transport a flux of 10^{11} ergs cm^{-2} sec^{-1} deep inside the Sun? Substituting the above numerical values for the subsurface gas, we can evaluate the Δg_T which is needed: we find $\Delta g_T \approx 10^{-11}$deg cm^{-1}.

What is the significance of this result? The answer depends on what we compare Δg_T to. Since Δg_T has the dimensions of a temperature gradient, it is natural to ask: is there another temperature gradient which is relevant to convection in the Sun? Indeed there is (see Chapter 6, Section 6.8): it is the adiabatic gradient g_{ad}. Deep in the Sun, $g_{ad} = g/C_p$ still has a numerical value of order 10^{-4} deg cm^{-1}: the subsurface increase in C_p is offset by the subsurface increase in g. Compared to g_{ad}, we see that the superadiabaticity Δg_T is seven orders of magnitude smaller.

For all practical purposes, the superadiabaticity is zero. That is, the temperature gradient in the deeper layers of the convection zone is *equal to the adiabatic gradient* g_{ad}. Since we already have a simple expression for g_{ad}, this provides an enormous simplification in our task of obtaining a model of the convection zone. We do not have to be concerned with how the opacity or density or pressure behave as a function of depth: instead, we simply accept that (to a high degree of precision) $dT/dz = g/C_p$. In regions where g and C_p are constant, this allows us to perform an immediate integral:

$$T(z) = T(z_o) + (z - z_o)\frac{g}{C_p} \tag{7.5}$$

The fact that the temperature gradient in the deep convection zone equals the adiabatic gradient means that the processes which occur in the solar convection zone are essentially adiabatic in nature. This will help us determine how pressure and density vary with depth.

7.7 Adiabatic and Nonadiabatic Processes

Once it has been determined that the temperature profile in the Sun's deep convection zone is essentially the adiabatic profile, we can in principle apply the laws of adiabatic processes to the variations of density and pressure. Thus, if the density varies as a function of depth according to $\rho(z)$, then the pressure at depth z is related to $\rho(z)$ according to $p(z) \sim [\rho(z)]^\Gamma$ (see Chapter 6, Section 6.7.3.) The index $\Gamma \equiv d(\log p)/d(\log \rho)_{ad}$ is the adiabatic exponent for pressure-density variations. For monatomic gases, under conditions where ionization is not occurring (or is essentially complete), the numerical value of Γ is 5/3.

For a perfect gas, $p \sim T\rho$, and so the density $\rho(z)$ in an adiabatic region is related to $T(z)$ by $\rho(z) \sim [T(z)]^{1/(\Gamma-1)}$. Also, as we have seen already (Equation 6.14), the pressure $p(z)$ in the adiabatic region is related to $T(z)$ by $p(z) \sim [T(z)]^{\Gamma/(\Gamma-1)}$.

Thus, if we were considering an adiabatic medium where $\Gamma = 5/3$ at all depths, then, given a temperature, density, and pressure at a reference depth z_o, the quantities at depth z would be given by Equation 7.5 plus the following two equations:

$$\rho(z) = \rho(z_o)\left[\frac{T(z)}{T(z_o)}\right]^{1.5} \equiv K_d T(z)^{1.5} \tag{7.6}$$

$$p(z) = p(z_o)\left[\frac{T(z)}{T(z_o)}\right]^{2.5} \equiv K_p T(z)^{2.5} \tag{7.7}$$

In Equations 7.6 and 7.7, we have introduced proportionality constants $K_d \equiv \rho(z_o)/T(z_o)^{1.5}$ and $K_p \equiv p(z_o)/T(z_o)^{2.5}$ for density and pressure, respectively. The constants K_d and K_p are related to the specific entropy of the gas at the top of the solar convection zone.

If Equations 7.5 through 7.7 were all that we needed to describe solar convection, then the computation of a model of the convection zone would be simple. We would start with our model of the photosphere (Chapter 5), evaluate the constants K_d and K_p using the conditions at the base of the photosphere (where convection sets in), and then proceed to deeper layers by increasing the depth z.

Unfortunately, things are not so simple in the Sun.

Two effects are particularly important in seriously modifying the properties of the convection zone near its upper boundary. First, radiative losses near the solar surface from convective elements (granules and intergranular regions) are severe. As a result, processes in the granulation are highly *non*adiabatic within the uppermost 1–2 megameters (Mm) of the convection zone. Nonadiabaticity has the effect that the local temperature gradient in the uppermost 1–2 Mm rises to values which are well in excess of the adiabatic gradient. (We have already used this information in Section 7.4 above, when we estimated temperature differences between rising and falling material.) We simply cannot assume that, as soon as convection sets in, the processes all become adiabatic.

The second important effect is that the value of the exponent Γ departs significantly from the monatomic value of 5/3 because of the effects of ionization. To be sure, Γ *is* close to 5/3 in the photosphere, and Γ again reverts to values within a few percent of 5/3 at depths in excess of $20-30$ Mm below the photosphere. In such regions, the exponents which appear in Equations 7.6 and 7.7 are entirely appropriate. However, at depths of a few megameters, where the degree of ionization of hydrogen is greater than (say) 10% , and less than (say) 90% (i.e., the gas lies in an "ionization strip" in Chapter 4,

Figure 4.1), the numerical values of Γ falls well below 5/3. As was mentioned earlier (Chapter 6, Section 6.7.3), Γ may fall as low as ≈ 1.19 in certain regions in the Sun. Now, in a medium where $\Gamma = 1.19$, the exponents in Equations 7.6 and 7.7 would take on values of 5.3 and 6.3, respectively. In such a medium, given an increase in temperature from depth z_o to depth z, the accompanying increases in density and pressure under adiabatic conditions would be significantly *larger* than Equations 7.6 and 7.7 would predict. The reason for this behavior has to do with the increase in entropy associated with ionization. Because the ionization energy is large compared to the thermal energy, a large input of energy dQ is required to cause ionization in unit mass of material, without any significant increase in temperature. This leads to a significant increase in the specific entropy $dS = dQ/T$.

Fully consistent modeling of solar convection requires inclusion of 3-D radiative transfer as well as a detailed treatment of the ionization of hydrogen (e.g., Stein and Nordlund, 1998). The results of such calculations indicate that if we use the conditions at the top of the convection zone to calculate $K_d \equiv \rho(z_o)/T(z_o)^{1.5}$ and $K_p \equiv p(z_o)/T(z_o)^{2.5}$, we will make large numerical errors. The errors are in the following sense: if we were to use the above numerical values of K_d and K_p in Equations 7.6 and 7.7, the pressures and densities we would calculate in the deep convection zone would be *too small* by two to three orders of magnitude.

In a complete model of the solar convection zone, we should include the full effects of radiative losses and include ionization effects at all depths. Such a model would demonstrate a behavior where $p \sim T^{2.5}$ at the shallowest depths near the surface ($z < 1$ Mm), then a narrow region of intermediate depths (a few Mm) where the exponent would be significantly larger than 2.5, followed by a deeper region where the exponent would decrease to approach 2.5 once more. At depths $z \geq 20 - 30$ Mm, conditions would revert to $p \sim T^{2.5}$. We shall see that the convection zone has a depth of order 200 Mm. Thus, the functional form $p \sim T^{2.5}$, as in Equation 7.7, applies throughout some 90% of the depth of the convection zone, although K_p takes on different values in the upper and lower portions of the convection zone.

7.8 Computing a Model of the Convection Zone: Step by Step

How can we make allowance for the above properties of the solar material? In this first course in solar physics, rather than following in detail the complicated calculations of radiative transfer and of the ionization of hydrogen at each depth, we make the following simplification: we use a single effective value for the exponents in Equations 7.6 and 7.7 throughout the convection

zone. To select the effective values, we use an effective value of Γ which is given by the arithmetic mean of the minimum and maximum values cited above, i.e., $\Gamma(\text{eff}) = 0.5(1.19 + 1.67) = 1.43$. With this choice, the exponents in Equations 7.6 and 7.7 become 2.3 and 3.3. We therefore use the following depth-dependences of density and pressure:

$$\rho(z) = \rho(z_o) \left[\frac{T(z)}{T(z_o)} \right]^{2.3} \tag{7.8}$$

$$p(z) = p(z_o) \left[\frac{T(z)}{T(z_o)} \right]^{3.3} \tag{7.9}$$

These equations, together with Equation 7.5, are the equations which we use to compute a model of the solar convection zone. We proceed as follows.

Start at the deepest layer in the photospheric model (Chapter 5), which also corresponds to the top of the convection zone. There, the temperature, depth, pressure and density are already known: since these are the first (topmost) values in our model of the convection zone, we refer to these as $T(1) = 6010$ K, $z(1) = 465$ km, $p(1) = 1.13 \times 10^5$ dyn cm^{-2}, and $\rho(1) = 2.94 \times 10^{-7}$gm cm^{-3}.

Step down below the photosphere by taking a step of say $\Delta z = 1000$ km. Assuming adiabatic conditions, the increase in temperature across the step Δz is

$$\Delta T = \Delta z \frac{g}{C_p} \tag{7.10}$$

What g should we use in estimating ΔT? The convection zone occupies a spherical shell which extends inward to great depths in the Sun, as deep as 20–30% of the solar radius. Within the convective shell, the total amount of mass is small compared to the total mass of the Sun. As a result, most of the mass of the Sun lies interior to the convection zone. Because of this, the value of g varies as $1/r^2$. Thus, at depth z, the local acceleration due to gravity can be calculated from

$$g(z) = 27{,}422 \left[\frac{R_\odot}{R_\odot - z} \right]^2 \text{ cm sec}^{-2} \tag{7.11}$$

At the base of the convection zone, $g(z)$ is about twice as large as the surface value.

What value of C_p should be used? Equation 6.5 provides a starting point. In the photosphere of the Sun, where $\gamma = 5/3$ and the mean molecular weight $\mu \approx 1.3$, we find $C_p \approx 1.6 \times 10^8$ ergs gm^{-1} K^{-1}. Both quantities γ and μ vary with depth. Let us consider μ first. Deep inside the Sun, where H and He are completely ionized, there are two particles for each H nucleus, and three particles for each He nucleus. As a result, the mean molecular weight per particle is 1/2 for H, and 4/3 for He. In a mixture of 90% H and 10% He, $\mu \approx (0.5^*0.9) + (1.33^*0.1) \approx 0.58$. This is the value of μ which we shall use in the deep interior of the Sun (Chapter 9). In the convection zone, where

ionization is underway, causing μ to vary from 1.3 (at the top) to 0.58 (at the bottom), we shall approximate the value of μ by the average of these limits, i.e. $\mu_{\rm conv} = 0.94$. Let us now consider γ. As in Section 6.7.3, we replace γ in the convection zone with the generalized Γ. Specifically, Equations 7.8 and 7.9 are based on the effective value $\Gamma = 1.43$. In order to preserve consistency, in Equation 6.5 we replace $\gamma/(\gamma - 1)$ by $\Gamma/(\Gamma - 1) = 3.3$. For simplicity, we assign a constant value to C_p throughout the convection zone, namely $C_p = 3.3\ R_g/\mu_{\rm conv}$. Of course, this does not take into account the largest values which C_p takes on at certain depths in the solar convection zone. We shall therefore not be surprised if our simplified model of the solar convection zone will be defective in certain ways.

Now that g and C_p can be evaluated at depth $z(1)$, we can calculate ΔT using Equation 7.10. Therefore, at the new depth $z(2) = z(1) + \Delta z$, the temperature $T(2)$ has the value $T(1) + \Delta T$.

Knowing the temperature $T(2)$ at $z(2)$, we calculate the local pressure and density using $p(2) = p(1)[T(2)/T(1)]^{3.3}$ and $\rho(2) = \rho(1)[T(2)/T(1)]^{2.3}$.

Repeating the calculation at a greater depth, $z(3) = z(2) + \Delta z$, we step inwards into the Sun, evaluating temperature, pressure and density at each step according to Equations 7.5 and Equations 7.8 and 7.9.

We continue increasing the depth until the temperature rises to a certain value, $T_b \approx 2 \times 10^6$ K. At that point, we stop the calculation. Why? Because we shall find in Chapter 8 that the base of the convection zone lies at a well-defined temperature T_b which is close to 2 million K.

This step-by-step procedure leads to a table of values of z, T, p, and ρ down to the base of the convection zone. An example of values selected from such a table are shown in Table 7.1.

7.9 Overview of Our Model of the Convection Zone

We see that at the base of the convection zone, i.e., at the location where the temperature T_b equals 2×10^6 K, our simplified model yields a pressure p_b of order 3×10^{13} dyn cm^{-2} and a density ρ_b of order 0.2 gm cm^{-3}. These values compare favorably with results from a recent detailed model of the Sun (Bahcall et al., 2006): $T_b = 2.01 \times 10^6$K, $p_b = 4.3 \times 10^{13}$ dyn cm^{-2}, and $\rho_b = 0.16$ gm cm^{-3}.

As far as the depth z_b of the convection zone is concerned, our model indicates a depth of 163 Mm. In terms of the solar radius, this is a depth of 23–24% of $R_\odot$. That is, the convection zone occupies about one-quarter of the distance from the solar surface to the center. This indicates clearly that convection in the Sun is by no means confined to a thin shell. Instead, we can properly refer to a thick "convective envelope" which penetrates inward by some 25% of the solar radius in the outermost layers of the Sun.

TABLE 7.1: A simplified model of the solar convection zone

Depth z (cm)	T**(K)**	p**(dyn cm^{-2})**	ρ**(gm cm^{-3})**
9.6492E+07	1.0819E+04	7.8964E+05	1.1412E-06
1.9649E+08	2.0260E+04	6.3889E+06	4.9305E-06
3.9649E+08	3.9224E+04	5.7757E+07	2.3023E-05
6.9649E+08	6.7877E+04	3.5920E+08	8.2740E-05
9.9649E+08	9.6780E+04	1.1716E+09	1.8928E-04
1.4965E+09	1.4552E+05	4.5614E+09	4.9010E-04
2.0465E+09	1.9997E+05	1.3156E+10	1.0287E-03
3.0465E+09	3.0127E+05	5.1560E+10	2.6759E-03
4.0465E+09	4.0566E+05	1.3897E+11	5.3565E-03
5.0465E+09	5.1329E+05	3.0446E+11	9.2742E-03
6.0465E+09	6.2432E+05	5.8471E+11	1.4643E-02
7.0465E+09	7.3889E+05	1.0252E+12	2.1694E-02
8.4965E+09	9.1169E+05	2.0652E+12	3.5418E-02
1.0046E+10	1.1057E+06	3.9286E+12	5.5552E-02
1.1546E+10	1.3034E+06	6.7961E+12	8.1527E-02
1.3046E+10	1.5115E+06	1.1135E+13	1.1518E-01
1.4546E+10	1.7310E+06	1.7496E+13	1.5804E-01
1.6046E+10	1.9628E+06	2.6598E+13	2.1187E-01
1.6296E+10	2.0027E+06	2.8443E+13	2.2205E-01

Actually, according to inversions of helioseismological data (Chapter 13), the Sun's convective envelope is somewhat thicker than 25% of the radius. The base of the convection zone is found to lie at a depth z_b which amounts to 28.7$\pm$ 0.3% of $R_\odot$ (Christensen-Dalsgaard et al., 1991). The simplified approach which we have used in calculating Table 7.1 yields a shallower convection zone than the helioseismological result by some 5% of $R_\odot$, i.e., by ≈ 35 Mm. How can we understand such a discrepancy? It is due in large part to our neglect of the large increases in C_p which occur in regions where hydrogen is undergoing ionization. In such ionization regions, the true values of C_p are *up to* 10 *or more times larger* than the value of 3.3 $R_g/\mu_{\rm conv}$ which we have adopted in the upper convection zone. As a result, for a given step in depth Δz, our computed $\Delta T = (g/C_p)\Delta z$ in the ionization zone is some ten times too large. Conversely, for a given temperature interval, our estimated value of the corresponding Δz is too small in the ionization zone by a factor of order ten. Thus, in an ionization region which spans a depth range of 1–3 Mm in the "real" Sun, our method has the effect that as much as 10–30 Mm of depth are "missing" by the time the integrated value of temperature reaches the limit T_b.

When solar models are computed with state-of-the-art computing techniques (e.g., Bahcall et al., 2006), the models yield estimates of the convection zone thickness which depend on the chemical composition which one assumes for the model. With two different choices of the solar composition, Bahcall et al. compute that the convective envelope has a thickness of 28.7 and 27.2% of $R_\odot$.

Exercises

7.1 Perform the step-by-step calculation of the convection zone described in Section 7.8, using values of $T(1)$, $z(1)$, $p(1)$, and $\rho(1)$ which you obtained in one of your models of the photosphere (Chapter 5). What differences do you find from the results in Table 7.1?

7.2 Repeat the calculation of Section 7.8 using a different value of the step size, Δz, e.g., 500 km, 2000 km. How much do the various parameters differ from those in Table 7.1?

7.3 Repeat the calculations of Section 7.8 using different values of $\Gamma(\text{eff})$. Instead of using $\Gamma(\text{eff}) = 1.43$, consider $\Gamma(\text{eff}) = 1.3$ *and* 1.6. Each of these will lead to changes in the exponents in Equations 7.8 and 7.9. Proceed in each case to the depth z_b where $T = T_b = 2$ MK. In each case, how do your values of z_b compare with the value obtained from helioseismology (200 Mm)?

References

Bahcall, J. N., Serenelli, A. M., and Basu, S. 2006. "10,000 standard solar models: a Monte Carlo simulation," *Astrophys. J. Suppl.*, 165, 400.

Christensen-Dalsgaard, J., Gough, D. O., and Thompson, M. J. 1991. "The depth of the solar convection zone," *Astrophys. J.*, 378, 413.

Stein, R. F., and Nordlund, A. 1998. "Simulations of solar granulation. I. General properties," *Astrophys. J.*, 499, 914.

Chapter 8

Radiative Transfer in the Deep Interior of the Sun

Continuing inward to the deep interior of the Sun, we note that, below the Sun's convection zone, energy is transported once again by means of radiation. We refer to this region as the radiative interior of the Sun. The aim of this chapter is to derive the equations which determine the radial profiles of temperature, pressure, and density in the radiative interior.

In this region of the Sun, hydrogen and helium are essentially completely ionized. As a result, there are no longer many bound electrons available. At temperatures in excess of 2 million K, the only remaining bound electrons belong to some of the metals, and their relative abundances are small. Therefore, photons are not as strongly absorbed in the radiative interior as they are in the cooler gas in the convection zone. As a result, the opacity decreases rapidly in the radiative interior.

With reduced opacity, radiation can more readily carry the energy flux outward through the Sun without requiring the temperature gradient to become large. That is, radiation once again takes over as the preferred means of energy transport. Thus, we can consider energy transport through the Sun in an overall sense in terms of a "sandwich": there are two regions in which radiation transports the energy (the photosphere, but the radiative interior), separated by a region where convection transports the energy.

8.1 Thermal Conductivity for Photons

When we come to consider how radiation travels deep inside the Sun, we find that it is easier to describe the flow of radiative energy there than was the case in the surface layers. As was described in Chapter 2 (especially Equations 2.29 through 2.31), when we considered radiative transfer *in the surface layers*, we had to give careful consideration to the large relative difference between outgoing I_o and incoming I_i intensities: in the extreme case of $\tau \rightarrow 0$, the incoming intensity can be set to zero, while the outgoing intensity is proportional to the full outward flux of energy generated by the Sun. But deep in the interior, the situation is different: I_o and I_i at any given point are both enormous compared to their values in the photosphere, but the

difference between outgoing and incoming intensities is very small compared to the magnitude of either. In this situation, the photons flow down the temperature gradient in a manner which can be well described in *diffusive* terms.

This means that the flux F of radiant energy can be described in the form of a generalized Fick's law: the flux F is linearly proportional to the local gradient of temperature. That is

$$F(r) = -k_{th}\frac{dT}{dr} \tag{8.1}$$

where k_{th} is the thermal conductivity in units of ergs $\mathrm{cm}^{-1}\mathrm{sec}^{-1}\mathrm{deg}^{-1}$.

Referring to the kinetic theory of gases, we find (e.g., Roberts and Miller, 1960) that in a medium where *particles* are responsible for the transport of heat, the general formula for thermal conductivity is

$$k_{th} = \frac{1}{3}\lambda V_t \rho C_v \tag{8.2}$$

Here, V_t is the mean thermal speed of the particles which are transporting the heat, λ is the mean free path of the particle (i.e., the mean distance a particle travels between collisions with the background medium), ρ is the mass density of the medium, and C_v is the specific heat per gram at constant volume of the medium which is doing the heat transport. Inserting the appropriate units, it is readily seen that the units of k_{th} are ergs $\mathrm{cm}^{-1}\mathrm{sec}^{-1}\mathrm{deg}^{-1}$, as required by Equation 8.1.

In the radiative interior of the Sun, we are dealing with a medium consisting of two distinct components: photons and material particles. The two components are closely coupled by means of emission and absorption of radiation. Energy transport is performed by the photons, while mass density is provided by the material particles. Let us apply Equation 8.2 to this case, where photons are the "particles" which transport energy. For photons, the mean speed of the "particles" is the speed of light: therefore, we replace V_t in Equation 8.2 with $c = 3 \times 10^{10}$ cm sec^{-1}. The mean free path for a photon is determined by the length scale λ corresponding to optical depth of order unity. Using the definitions in Chapter 3, Section 3.1, we see that this length scale is given by the condition $\lambda\kappa\rho = 1$ where κ is the local opacity. Thus, in Equation 8.2, we replace λ by the quantity $1/\kappa\rho$. The density ρ is the local mass density.

How do we estimate the term C_v for the case of a photon-material mixture in which the photons are the carriers of energy? We start by recalling (Chapter 2, Section 2.1) that the energy contained in radiation *per unit volume* is $u = a_R T^4$ ergs cm^{-3} where $a_R = 7.5658 \times 10^{-15}$ erg cm^{-3} deg^{-4} is the radiation density constant. In terms of units, we note that $C_v = (dU/dT)_v$ refers to an energy content U *per gram* of the medium. To convert from energy per unit volume to energy per gram, we divide u by the local mass density. This leads to $U = a_R T^4/\rho$ ergs gm^{-1}. This yields $C_v = 4a_R T^3/\rho$ ergs $\mathrm{gm}^{-1}\mathrm{deg}^{-1}$.

Combining the various terms, we finally have

$$k_{th} = \frac{4a_R cT^3}{3\kappa\rho} \tag{8.3}$$

8.2 Flux of Radiant Energy at Radius r

Noting that the Stefan–Boltzmann constant σ_B is related to the radiation constant a_R by the formula $\sigma_B = a_R c/4$, we finally have that

$$k_{th} = \frac{16\sigma_B T^3}{3\kappa\rho} \tag{8.4}$$

This is the thermal conductivity of a medium in which photons are transporting energy. The larger the opacity, the smaller the thermal conductivity.

Now that we know the thermal conductivity, we can write down the local flux of radiant energy at radial location r in terms of the local temperature gradient:

$$F(r) = -\frac{16\sigma_B T(r)^3}{3\kappa(r)\rho(r)}\frac{dT}{dr} \tag{8.5}$$

8.3 Base of the Convection Zone

At this point, we can determine a quantity to which we have already referred in Chapter 7: the temperature at the base of the convection zone. This temperature, T_b, is determined by the location where the temperature gradient due to radiation (given by Equation 8.5) becomes as large as the adiabatic gradient g/C_p. As was mentioned in Chapter 7, the numerical values of g and C_p at the base of the convection zone (at a radial location of about $0.7R_\odot$) are both larger than their respective surface values by a factor of about two. Thus, the surface value of $g/C_p (\approx 1.7 \times 10^{-4}$ deg cm^{-1}: see Chapter 6, Equation 6.9) can be inserted for dT/dr in Equation 8.5 to determine the base of the convection zone.

Also at the base of the convection zone, where $r \approx 0.713R_\odot$, $F(r)$ is larger than the surface flux $F_\odot (= 6.3155 \times 10^{10}$ ergs cm^{-2} sec^{-1}: Chapter 1, Section 1.8) by a factor of $(R_\odot/r)^2 \approx 2$. Thus, we can set $F(r_b) \approx 1.3 \times 10^{11}$ ergs cm^{-2}sec^{-1}.

As regards the opacity, we have already noted (Chapter 3, Section 3.7) that at temperatures in excess of about 10^6 K, a reasonable fit to the opacities can be obtained by the Kramers "law": $\kappa = \kappa_o \rho/T^{3.5}$ cm^2 gm^{-1}. By fitting to

tabulated values of opacity in conditions which are relevant to the solar interior (e.g., Harwit, 1973), we have determined that a plausible numerical value of κ_o is roughly 10^{24} when ρ is in units of gm cm^{-3} and T is in units of K. (This choice leads to a value of $\kappa \approx 10^3$ cm^2 gm^{-1} in gas where $\rho \approx 1$ gm cm^{-3} and $T \approx 10^6$ K.)

Inserting these factors in Equation 8.5, we find that the temperature and density at the base of the convection zone are related by

$$\frac{T_b^{6.5}}{\rho_b^2} \approx 3 \times 10^{42} \tag{8.6}$$

The uncertainties in estimating the numerical values of the various parameters entering into Equation 8.5 are such that we retain only one significant digit in Equation 8.6.

To proceed further, we need to know the relationship between T_b and ρ_b. Such a relationship is already available: in the deep convection zone, we have seen (Chapter 7) that the density and temperature are related by an adiabatic function. This function, Equation 7.8 (Chapter 7), when applied to the base of the convection zone, indicates that $\rho_b = K_d T_b^{2.3}$, where K_d is related to the parameters at the top of the convection zone by $K_d = \rho(z_o)/T(z_o)^{2.3}$. Inserting values of $\rho(z_o) = 2.9 \times 10^{-7}$ gm cm^{-3} and $T(z_o) = 6010$ K (from Chapter 5, Table 5.3), we find that $K_d \approx 6 \times 10^{-16}$, where we again retain only one significant digit in view of the simplification which enters into the choice of the exponent 2.3 (see Chapter 7).

Inserting these values into Equation 8.5, we find $T_b^{1.9} = 1 \times 10^{12}$. This leads finally to $T_b \approx 2 \times 10^6$ K. This is the origin of our choice of temperature at the base of the convection zone when we computed a model of the convection zone in Chapter 7.

8.4 Temperature Gradient in Terms of Luminosity

It is useful to convert from units of flux to units of power (i.e., luminosity). At any radial location inside the Sun, the luminosity $L(r)$ (in units of ergs sec^{-1}) has a value which is determined by the summation of energy sources which lie interior to radial location r. The value of $L(r)$ is zero near the center of the Sun, and it increases rapidly in magnitude in the energy-generating core. Detailed models indicate that $L(r)$ rises to $> 90\%$ of its asymptotic value at a radial location of about 0.2 $R_\odot$. The local flux of radiant energy $F(r)$ (in units of ergs cm^{-2} sec^{-1}) is related to $L(r)$ by $F(r) = L(r)/4\pi r^2$.

Combining this with Equation 8.5, we see that we can write

$$\frac{T(r)^3}{\kappa(r)} \frac{dT}{dr} = -\frac{3L(r)}{64\pi\sigma_B} \frac{\rho(r)}{r^2} \tag{8.7}$$

8.5 Temperature Gradient in Terms of Pressure

The usefulness of Equation 8.7 can be seen by comparing it with the equation of HSE:

$$\frac{dp}{dr} = -g(r)\rho(r) = -GM(r)\frac{\rho(r)}{r^2} \tag{8.8}$$

where $M(r)$ is the mass interior to radial location r. Does HSE apply to the radiative interior of the Sun? Yes: there are no bulk flows of gas in that part of the Sun.

Notice that on the right-hand sides, both Equations 8.7 and 8.8 contain the factor $\rho(r)/r^2$. Therefore, if we take the ratio of Equations 8.8 and 8.7, we find an equation which relates T and p at any radial location in the star:

$$\frac{T(r)^3}{\kappa(r)}\frac{dT}{dp} = \frac{3}{64\pi\sigma_B G}\frac{L(r)}{M(r)} \tag{8.9}$$

As already mentioned, detailed solar models indicate that $L(r)$ builds up rapidly to its asymptotic value as r increases from $r = 0$, reaching 90% of $L_\odot$ at $r = 0.2R_\odot$. For the mass function, $M(r)$ also rises from zero at $r = 0$, and tends to the asymptotic value $M_\odot$ as r increases. The rate of rise in $M(r)$ is not as great as for $L(r)$: $M(r)$ reaches 90% of its asymptotic value $M_\odot$ around $r = 0.5R_\odot$. In the outer parts of the radiative interior, where both $M(r)$ and $L(r)$ are within 10% of their asymptotic values, the ratio *L(r)/M(r)* can be well approximated with the asymptotic value $(L/M)_a \equiv L_\odot/M_\odot \approx 2$ ergs $\mathrm{sec}^{-1}\mathrm{gm}^{-1}$ (see Chapter 1, Section 1.4). Closer to the center of the Sun, where $L(r)$ remains large while $M(r)$ decreases, the ratio $L(r)/M(r)$ becomes larger than $(L/M)_a$. Examination of detailed models suggests that $L(r)/M(r)$ exceeds $(L/M)_a$ by factors of 2, 4, and 6 at $r \approx 0.25R_\odot$, $r \approx 0.15R_\odot$, and $r \leq 0.1R_\odot$. Thus, throughout 98% of the volume of the Sun, the right-hand side of Equation 8.9 retains a constant value, within a factor of 2.

For purposes of the simplified solar model we are considering here, we shall set the right-hand side of Equation 8.9 equal to a constant, $C_1 = 8 \times 10^9$ c.g.s. This is the appropriate value for regions of the Sun where $L(r)/M(r) \approx (L/M)_a$.

8.6 Integrating the Temperature Equation

In order to proceed with the integration of Equation 8.9, we use the Kramer's opacity law, as described above: $\kappa = 10^{24}\ \rho/T^{3.5}\ \mathrm{cm}^2\mathrm{gm}^{-1}$. Substituting this in Equation 8.9, we find

$$T^{6.5}\ dT = 10^{24} C_1 \rho dp \tag{8.10}$$

Using the perfect gas law, $\rho = p\mu/R_g T$, Equation 8.10 can be written as

$$T^{7.5}\ dT = C_2 p dp \tag{8.11}$$

where $C_2 = 10^{24}\ C_1\ \mu/R_g \approx 10^{26}\mu$. In the radiative interior of the Sun, $\mu \approx 0.5$. Integration yields an expression for pressure in terms of the temperature:

$$p^2 = C_3 T^{8.5} + \text{const} \tag{8.12}$$

where $C_3 = 1/(4.25\ C_2)$. In order to avoid the use of large numbers, it is convenient to express the temperature as $T_6 = T/10^6$ K. In these units, we find

$$p^2 = 5 \times 10^{24}\ T_6^{8.5} + \text{const} \tag{8.13}$$

To evaluate the constant, we use the conditions which have been computed for the base of the convection zone: according to Table 7.1 (Chapter 7), we see that at that location, $T_6 = 2.0027$, and $p = 2.84 \times 10^{13}$ dyn cm^{-2}. Inserting these in Equation 8.13, we find that the constant has the value -1.02×10^{27} c.g.s. We shall use this in the next Chapter.

Exercise

8.1 The Kramers "law" is not a perfect fit to the opacities in the solar interior. Other possible fits to the opacities include the cases $\kappa = \kappa'\rho/T^3$ cm^2 gm^{-1} and $\kappa = \kappa''\rho/T^4$ cm^{2}gm^{-1}. For both these cases, evaluate κ' and κ'' by fitting (in both cases) $\kappa = 10^3$ cm^{2}gm^{-1} at $\rho = 1$ and $T = 10^6$ K. Starting at Equation 8.10, and keeping $C_1 = 8 \times 10^9$ c.g.s., obtain revised versions of Equation 8.13, including revised values for the constant of integration.

References

Harwit, M. 1973. "Stars," *Astrophysical Concepts.* J. Wiley and Sons, New York, pp. 303–373.

Roberts, J. K. and Miller, A. R. 1960. "The transfer of heat by conduction and convection," *Heat and Thermodynamics.* Blackie and Son, London, UK, pp. 280–315.

Chapter 9

Computing a Mechanical Model of the Sun: The Radiative Interior

Following the spirit of Chapters 5 and 7, we now proceed deeper into the Sun and calculate a radial profile for the physical variables. As in the earlier chapters, we still refrain from considering the *origin of the energy* which is passing through. Our aim is purely mechanical: given a total luminosity, how does the medium arrange itself so as to "handle" the energy which is passing through? A complete model of the Sun would of course include a description of the processes whereby the energy is generated: we will discuss that in Chapter 11. But in this chapter, we do not attempt to calculate a complete model. Our goal is as follows: given the solar luminosity as a boundary condition, what can we deduce about the structure of the Sun?

In Chapter 8 (Section 8.6), we derived the following relationship between pressure and temperature:

$$p^2 = 5 \times 10^{24} T_6^{8.5} - 1.02 \times 10^{27} \tag{9.1}$$

This equation applies (within our simplification of constant L/M ratio) to the radiative interior. We use Equation 9.1 to continue our computation of a solar model. The model will consist of a table in which each line refers to a particular depth (i.e., radial location), at which we calculate the local temperature, pressure, and density.

9.1 Computational Procedure: Step by Step

We start at the base of the convection zone, where we already (see Chapter 7, Table 7.1) have numerical values for the quantities z_b, $T_b (\approx 2 \times 10^6$ K), p_b, and ρ_b. It is convenient to convert now from depth z to radial distance from the center of the Sun: $r = R_\odot - z$. Thus, the starting values for the four parameters in the table we wish to compute for the radiative interior are: $r(1) = R_\odot - z_b$, $T_6(1) = 2.0027$, $p(1) = p_b$, and $\rho(1) = \rho_b$. These are the parameters we enter into the first line of our table of the solar radiative interior.

The computation proceeds by means of the following steps:

1. Choose an increase in temperature of (say) $\Delta T_6 = 0.01$. Thus, $T_6(2)$, the temperature (in units of 10^6 K) of the second row in the table, is given by $T_6(2) = T_6(1) + \Delta T_6$.

2. Using $T_6(2)$ in Equation 9.1, the pressure $p(2)$ on the second line of the table can be calculated.

3. The pressure increment between lines 1 and 2 is $\Delta p = p(2) - p(1)$.

4. The density on the second line of the table is calculated from the perfect gas law: $\rho(2) = p(2)\mu/(R_g T(2))$. Here, μ can be set equal to 0.5 (see Chapter 7, Section 7.8). The mean density between lines 1 and 2 is $\rho(a) = 0.5(\rho(1) + \rho(2))$.

5. The linear distance between lines 1 and 2 can be derived from the equation of hydrostatic equilibrium:

$$\Delta r = -\frac{\Delta p}{g(r)\rho(a)} \tag{9.2}$$

   However, before this step can be taken, we need to discuss what value we should use for the acceleration $g(r)$.

6. In order to calculate $g(r)$, the physics of the solar interior tells us that there are two different zones in the radial coordinate which we need to distinguish. In the first zone, in the outermost parts of the Sun, at radial location r, the local gas density is small enough that the mass $M(r)$ enclosed within radius r is essentially constant. As a result, $g(r) = GM(r)/r^2$ in the outermost parts of the Sun can be written essentially as $g(r) = GM_\odot/r^2$. As a result, in this outer zone of the radial coordinate, $g(r)$ *decreases* as the radial location *increases*, according to the inverse square law, $1/r^2$. In the second zone, near the center of the Sun, the density does not change rapidly: within the inner 10% of the solar radius, the density changes by a factor of only about 2. In the limit of constant density ρ_c near the center, the local acceleration due to gravity tends toward

$$g(r) = \frac{GM(r)}{r^2} \approx \frac{4\pi G\rho_c}{3} r \tag{9.3}$$

   Thus, near the center of the Sun, the acceleration due to gravity at radius r *increases* linearly with *increasing* r.

   The existence of the two distinct zones means that the radial profile of $g(r)$ inside the Sun is *not* monotonic. Instead, there exists, inside the Sun, a radial location r_m where $g(r)$ has a maximum value.

   In order to include this feature, and in the spirit of simplicity which informs our approach to modeling the interior of the Sun, we assume that the behavior of $g(r)$ inside the Sun can be captured adequately by a composite of two functions, depending on the radial location.

At radial locations *outside* a certain critical radius r_m, we use the inverse square law:

$$g(r) = g_s \left(\frac{R_\odot}{r} \right)^2 \quad (r \geq r_m) \tag{9.4}$$

where $g_s = 27{,}422$ cm sec^{-2} is the acceleration at the surface of the Sun (see Chapter 1, Equation 1.13).

At radial locations *inside* r_m, we use the linear law:

$$g(r) = g(r_m) \frac{r}{r_m} \quad (r \leq r_m) \tag{9.5}$$

where $g(r_m)$ is chosen so that $g(r)$ is continuous at $r = r_m$.

What value is appropriate for r_m? Various values can be chosen in order to determine what effect the choice would have on the solar model. One possibility that we have used for the tabulated model to be reported below is to identify r_m with the radial location where the mass interior to $r = r_m$ is 50% of the solar mass, i.e., $M(r_m) \approx 0.5 M_\odot$. Using information from detailed solar models, it turns out that this condition corresponds to $r_m \approx 0.25 R_\odot$. With this choice, as we move inward into the solar interior, Equation 9.4 indicates that the value of g at first *increases* from its surface value $g_s = 27{,}422$ cm sec^{-2} to a sharp peak of $g(r_m) = 16 g_s$. Then in the inner zone, between r_m and the center of the Sun, g *decreases* along a linear ramp toward a value of zero at the center. In detailed solar models (see Exercise 5 at the end of this chapter), the radial profile is not sharply peaked, and has a smaller maximum value 8–9 g_s.

7. Now that $g(r)$ can be evaluated at any radial location, Δr can be computed using Equation 9.2, and then $r(2) = r(1) + \Delta r$.

8. Repeat steps $1 - 7$ n times until the computed radial location $r(n+1) = r(n) - \Delta r$ reaches the value zero. At this point, the model has reached the center of the Sun, and the tabulated parameters refer to conditions at the center.

An example of an abbreviated table computed according to the above prescription is given in Table 9.1.

9.2 Overview of Our Model of the Sun's Radiative Interior

What do the results of the model in Table 9.1 tell us about conditions in the deep interior of the Sun?

According to this model, the gas at the center of the Sun has a temperature of roughly $T_c = 16.5$ million K, a density of $\rho_c = 141$ gm cm^{-3}, and a pressure

TABLE 9.1: A mechanical model for the radiative interior of the Sun: $r_m = 0.25 R_\odot$

r(cm)	T(K)	p(dyn cm^{-2})	ρ(gm cm^{-3})	g(cm sec^{-2})
5.3093E+10	2.0127E+06	2.9803E+13	1.0329E-01	4.6766E+04
5.1046E+10	2.1027E+06	4.1817E+13	1.3873E-01	5.0548E+04
4.9373E+10	2.2027E+06	5.5592E+13	1.7605E-01	5.4111E+04
4.7997E+10	2.3027E+06	7.0541E+13	2.1369E-01	5.7301E+04
4.5697E+10	2.5027E+06	1.0561E+14	2.9436E-01	6.3262E+04
4.3737E+10	2.7027E+06	1.4960E+14	3.8613E-01	6.9084E+04
4.1179E+10	3.0027E+06	2.3715E+14	5.5092E-01	7.7965E+04
3.8249E+10	3.4027E+06	4.0589E+14	8.3209E-01	9.0393E+04
3.5726E+10	3.8027E+06	6.5218E+14	1.1964E+00	1.0364E+05
3.3520E+10	4.2027E+06	9.9832E+14	1.6570E+00	1.1775E+05
3.1572E+10	4.6027E+06	1.4696E+15	2.2273E+00	1.3275E+05
2.9839E+10	5.0027E+06	2.0944E+15	2.9204E+00	1.4865E+05
2.7924E+10	5.5027E+06	3.1399E+15	3.9805E+00	1.6977E+05
2.6240E+10	6.0027E+06	4.5442E+15	5.2808E+00	1.9229E+05
2.4747E+10	6.5027E+06	6.3846E+15	6.8491E+00	2.1621E+05
2.3416E+10	7.0027E+06	8.7472E+15	8.7135E+00	2.4153E+05
2.2220E+10	7.5027E+06	1.1726E+16	1.0903E+01	2.6825E+05
2.1141E+10	8.0027E+06	1.5426E+16	1.3446E+01	2.9636E+05
2.0162E+10	8.5027E+06	1.9958E+16	1.6373E+01	3.2588E+05
1.9269E+10	9.0027E+06	2.5443E+16	1.9715E+01	3.5680E+05
1.8452E+10	9.5027E+06	3.2014E+16	2.3501E+01	3.8912E+05
1.7702E+10	1.0003E+07	3.9810E+16	2.7763E+01	4.2284E+05
1.6998E+10	1.0503E+07	4.8980E+16	3.2532E+01	4.2916E+05
1.6272E+10	1.1003E+07	5.9684E+16	3.7840E+01	4.1087E+05
1.5513E+10	1.1503E+07	7.2092E+16	4.3720E+01	3.9173E+05
1.4714E+10	1.2003E+07	8.6382E+16	5.0204E+01	3.7161E+05
1.3870E+10	1.2503E+07	1.0274E+17	5.7325E+01	3.5033E+05
1.2970E+10	1.3003E+07	1.2138E+17	6.5116E+01	3.2768E+05
1.2004E+10	1.3503E+07	1.4249E+17	7.3612E+01	3.0334E+05
1.0953E+10	1.4003E+07	1.6630E+17	8.2845E+01	2.7687E+05
9.7901E+09	1.4503E+07	1.9304E+17	9.2851E+01	2.4760E+05
8.4694E+09	1.5003E+07	2.2295E+17	1.0366E+02	2.1438E+05
6.9014E+09	1.5503E+07	2.5628E+17	1.1532E+02	1.7498E+05
4.8543E+09	1.6003E+07	2.9330E+17	1.2785E+02	1.2370E+05
2.0805E+08	1.6503E+07	3.3427E+17	1.4130E+02	1.5297E+04

of $p_c = 3.34 \times 10^{17}$ dyn cm^{-2}. Protons at a temperature of T_c have an rms speed $V_{\rm rms} = \sqrt{(3 R_g T_c)} \approx 640$ km sec^{-1}.

It is important to compare this proton speed at the center of the Sun with another characteristic speed associated with the Sun as a whole: the escape speed $V_{\rm esc}$ from the surface. In Chapter 1, Section 1.7, we saw that $V_{\rm esc} = 617.7$ km sec^{-1}. The latter is a measure of the strength of the inward pull of gravity which holds the hot gas at the center of the Sun together by means of the crushing weight of the overlying gas. We see that in our simplified

mechanical model of the Sun, we have found that the conditions at the center are such that $V_{\rm rms}$ agrees with $V_{\rm esc}$ within about 4%. This agreement indicates that in our model, the inward pull of the gravitational forces and the outward force of pressure are close to achieving a balance.

How does our mechanical model in Table 9.1 compare with models which have been computed by including many more details of the physics (including energy generation by nuclear fusion)? To answer that, we note that in a series of 10,000 "standard solar models" reported by Bahcall et al. (2006), the best estimates of temperature, density and pressure at the center are found to be 15.48 million K, 150.4 gm cm^{-3}, and 2.34×10^{17}dyn cm^{-2}. Compared with these, our mechanical model yields a central temperature which is too high by about 7%, a density which is too low by about 6%, and a pressure which is too high by about 40%. The main reason for the error in pressure has to with the mean molecular weight: we have assumed $\mu = 0.58$ throughout the radiative interior, whereas in the "real Sun," nuclear reactions build up more and more helium in the core as time goes on. As a result, in the Bahcall et al. model, the central value of μ equals 0.83, i.e., some 40% larger than our assumed value. In the best estimate models of Bahcall et al., protons at the center of the Sun have $V_{\rm rms} = 621$ km sec^{-1}, agreeing with $V_{\rm esc}$ to better than 1%. With a central density of 150 gm cm^{-3}, and a mean nuclear weight of about 2 (He is building up in abundance in the core because of the reactions), the corresponding number density of nuclei is of order 0.5×10^{26} cm^{-3}.

The central density in the Sun (150 gm cm^{-3}) exceeds the mean density of the Sun (1.41 gm cm^{-3}: see Chapter 1, Equation 1.21) by a factor of slightly more than 100. This is a measure of the "central condensation" of the Sun to which we shall return in Chapter 10 (Section 10.9).

Another aspect of our model which deserves attention concerns the ratio of radiation pressure p_r to gas pressure. The value of p_r at the center of the Sun can be determined from the result (see Chapter 2, Section 2.8) $p_r = a_R T^4/3$: our model yields $p_r \approx 2 \times 10^{14}$ dyn cm^{-2}. The gas pressure at the center of the Sun exceeds p_r by a factor of more than 1000. Elsewhere in the Sun, the gas pressure exceeds the radiation pressure by even greater factors. For example, in the photosphere, the gas pressure ($\approx 10^5$ dyn cm^{-2}) exceeds the radiation pressure by a factor of more than 3×10^4. These numerical values indicate that we are justified in neglecting radiation pressure compared to the gas pressure when we calculate a first model of the Sun: when we wrote down the equation of hydrostatic equilibrium (Chapter 5, Equation 5.1), the quantity p in Equation 5.1 includes only the gas pressure. In certain stars other than the Sun, this might not be an acceptable approximation: but in the case of the Sun, radiation pressure does not contribute significantly to supporting the Sun against gravity. Finally, it should not be too surprising that our model is not perfect. Because of various simplifications, and because we use different modeling techniques and different μ values in different regions of the Sun, our model includes an artificial "step" in density between the bottom of the convection zone (see Table 7.1, bottom line) and the top of the radiative interior (see Table 9.1, top line).

9.3 Photons in the Sun: How Long before They Escape?

Now that we have obtained a model of the interior of the Sun where photons transport the energy, it is worthwhile to ask: how long does it take for a photon to propagate from the center of the Sun to the surface? Subsequently we shall compare the photon time-scale with the time-scale for the escape of a very different type of elementary particle (the neutrino: Chapter 12) which is also generated in the core of the Sun.

To estimate the photon time-scale, we note that photons which are generated at the center of the Sun make their way outward by diffusing through the material of the solar interior. In this process, the photons make their way outward in radius by means of a random walk: if the length of each step in the random walk is on average l_p, then after N steps, the photon will have moved outward in the radial direction by a distance $r_N \approx l_p\sqrt{N}$.

As a first step toward estimating the diffusion time-scale, we consider the mean free path l_p which a photon travels between interactions with the solar material. In view of the definition of opacity, it is clear that $1/\kappa\rho$ is an appropriate length-scale (see Chapter 8, Section 8.1). Therefore, we may take $l_p \approx 1/\kappa\rho$.

What is a typical numerical value for this length scale? Near the center of the Sun, the model in Table 9.1 indicates $\rho \approx 140$ gm cm^{-3}. Moreover, with opacity in the solar interior given by the Kramer's law, $\kappa = \kappa_o \rho T^{-3.5}$, and using the value $\kappa_o = 10^{24}$ (see Chapter 8), we find that near the center of the Sun, where $T \approx 1.6 \times 10^7$ K, the numerical value of the opacity κ is of order 10 cm^{2}gm^{-1}. This leads to $l_p \approx 0.7 \times 10^{-3}$ cm. In the center of the Sun, photons are restricted to very short mean free paths.

How is l_p expected to vary with increasing radial distance from the center of the Sun? In the radiative interior, we have seen that the pressure (at least near the center) varies as $T^{4.25}$. In such conditions, the perfect gas law indicates that $\rho \sim T^{3.25}$. Moreover, assuming Kramers opacity, $\kappa \sim \rho/T^{3.5}$, we find $l_p \approx 1/\kappa\rho \sim T^{3.5}/\rho^2 \sim 1/T^3$. As a result, the value of l_p increases as we move away from the center of the Sun. However, even when we reach the base of the convection zone, at a radial location of $r_b \approx 0.7 R_\odot$, where the temperature has fallen to 2 million K, l_p has increased by no more than about 450, i.e., $l_p \approx 0.3$ cm. Throughout the radiative interior the mean free time $t_p = l_p/c$ between photon collisions is no more than 10^{-11} sec.

In terms of the random walk argument given above, the number of "steps" N_b that the photon must take in order to move outward from the center of the Sun to the base of the convection zone is given by $N_b \approx (r_b/l_p)^2$. The time required for this number of steps is $t_b \approx N_b t_p \approx r_b^2/l_p c$. Most of the time required by a photon to random walk to the base of the convection zone is spent in the core of the Sun, where l_p is smallest. Inserting the values $r_b \approx 5 \times 10^{10}$ cm, $l_p \approx 0.001$ cm we find $t_b \approx 8 \times 10^{13}$ sec, i.e., 2–3 $\times 10^6$ years. Thus, photons require on average a few million years to propagate from the core of the Sun out to the base of the convection zone. From there, their

energies are transported to the surface by fluid flow on much shorter time-scales.

This indicates that when we observe the Sun today, the energy entering our eyes was actually generated several million years ago. It will be a matter of interest in a subsequent chapter to compare this photon time-scale with the corresponding value for neutrinos.

9.4 Global Property of the Solar Model

We can obtain a complete model of the Sun, based on our simplified approach, by combining Tables 5.3, 7.1, and 9.1.

In view of the large ranges of physical parameters between surface and center, is there some way that we can check our calculations in some global sense? There is one test we can do.

In Chapters 13 and 14, we shall be interested in the topic of helioseismology, i.e., the study of eigenmodes of oscillation within the Sun. One class of eigenmodes, relying on pressure as the restoring force, are referred to as p-modes. Each eigenmode has an eigenfunction which, when plotted as a function of radius from the center of the Sun to the surface, exhibits a definite number of "nodes" (where the eigenfunction passes through 0). The number of such nodes in the radial direction, n_r, helps to define each mode.

We shall find that at high frequencies, the p-modes display a well-defined asymptotic behavior: for modes of a given angular degree (l), the frequencies of modes which differ by unity in the value of n_r differ from each other by a characteristic frequency spacing $\Delta\nu$. Theory indicates that the frequency spacing $\Delta\nu$ is related to the time t_s required for sound to propagate from the center of the star to a reflection point at radial location $R(r)$ near the photosphere:

$$t_s = \int_0^{R(r)} \frac{dr}{c_s(r)} \tag{9.6}$$

where c_s is the sound speed. Specifically, $\Delta\nu$ can be shown to be equal to $1/(2t_s)$.

Now that we have obtained a model of the Sun, albeit only a simplified model, it is of interest to inquire: what is the sound travel time from center to photosphere according to our model? The integration in Equation 9.6 can be performed using the combined information in Tables 5.3, 7.1, and 9.1. When we do this, we find that the sound crossing time from the center to the photosphere of our combined model is $t_s = 3804$ sec, a few minutes longer than one hour. This leads to $\Delta\nu = 131.5$ μHz.

Empirically, p-modes in the Sun with low values of $l(= 0, 1, 2)$, are found to have asymptotic spacings of $\Delta\nu = 134.8 - 135.1$ μHz (Appourchaux et al.,

1998). Thus, our mechanical model of the Sun replicates the solar asymptotic spacings within 2–3%.

9.5 Does the Material in the Sun Obey the Perfect Gas Law?

In computing the model of the three regions of the Sun (Chapters 5, 7, and 9) we have used the equation of state for a perfect gas. Now that we have calculated the conditions in the interior of the Sun, we need to perform a consistency check, and ask: does the gas in the Sun really obey the perfect gas law? After all, we have found that the density at the center of the Sun exceeds 140 gm cm^{-3}: this is denser than solid gold or solid lead, and the latter materials certainly do not obey the perfect gas law.

What criterion can we use in order to test whether the perfect gas law, which follows from the classical kinetic theory of gases, is actually obeyed inside the Sun? The answer is: classical theories of matter are acceptable as long as quantum mechanical effects are negligible.

According to quantum mechanics, particles in certain circumstances behave with wave-like properties. The wavelength λ_p associated with a particle of mass m, moving with speed V, is given by de Broglie's formula: $\lambda_p = h/(mV)$, where $h = 6.62606896 \times 10^{-27}$ gm cm^2 sec^{-1} is Planck's constant.

Classical physics provides a reliable description of the behavior of matter as long as the de Broglie waves of individual particles do not overlap one another significantly. But the laws of classical physics break down if the de Broglie wave of one electron is so large that it overlaps significantly with the de Broglie waves of a number of the neighboring electrons. How large should the number of overlaps be? It must be at least two, because the existence of electron spin allows two electrons to occupy the same element in phase space without contradicting Pauli's exclusion principle. Once the wave of one particle overlaps the waves of (say) ten or more neighboring particles, then the electrons with overlapping de Broglie waves begin to "feel the pressure" of the Pauli exclusion principle. In a very real sense, the electrons are subject to a physical pressure which "drives them away" from their neighbors in phase space. In such a situation, quantum effects must be taken into account, and the electrons are said to be "degenerate," and the pressure which they exert is called "electron degeneracy pressure."

In conditions where quantum effects are important, the pressure associated with electron degeneracy can be strong enough to support the overlying weight of an object with a mass of order the Sun's mass. In such an object, the pressure of the *thermal gas* is no longer the physical agent which supports the star against its own weight. As a result, even if HSE were to hold, there would no longer be any reason why the rms *thermal speed* of the particles at the center of the star should be equal to the escape speed from the surface.

Let us see what happens to de Broglie waves at the center of the Sun. In order to make the quantum effects as large as possible, we consider electrons: their small mass ensures that their de Broglie wavelengths are larger than those of other constituents. Therefore, electrons have the best chance of having their de Broglie waves overlap with their neighbors. For electrons at the center of the Sun, where the rms velocity is $\sqrt{(3kT/m_e)}$, the de Broglie wavelength has a mean value of $\lambda_e = h/\sqrt{(3kTm_e)}$. Inserting the temperature at the center of the Sun, $T = 1.6 \times 10^7$ K, we find $\lambda_e = 2.7 \times 10^{-9}$ cm.

Now that we know the value of the de Broglie wavelength of an electron at the center of the Sun, we can determine the conditions which are required for degeneracy to be an important contributor to the pressure. In order that the de Broglie wave of one particular electron would extend throughout a volume in which there are ten or more other electrons, the mean distance between the electrons at the center of the Sun must be less than λ_e by a factor of at least $10^{1/3}$. Thus if electron degeneracy is to be important in the center of the Sun, the mean distance between electrons in the center of the Sun must be no larger than $d_e \leq 1.3 \times 10^{-9}$ cm.

In a medium where the electron density is n_e cm^{-3}, the mean distance d_e between an electron and its neighbors is roughly $1/n_e^{1/3}$ cm. Thus, for electron degeneracy to be important at the center of the Sun, the number density of electrons would have to be at least as large as 5×10^{26} cm^{-3}. At the center of the Sun, where hydrogen burning has been going on for several billion years, helium has increased to an abundance which is comparable to H. Corresponding to each electron from He, there are two nucleons to contribute mass, each with a mass of 1.67×10^{-24} gm. Therefore, the mass density in a helium-dominated region with $n_e \geq 5 \times 10^{26}$ cm^{-3} would exceed 1700 gm cm^{-3}. This is more than ten times larger than the best estimates for the density at the center of the Sun (150.4 gm cm^{-3} in the model of Bahcall et al., 2006).

Thus, despite the high gas density at the center of the Sun, the central temperature is so large that the average de Broglie wavelength of each electron is relatively short. As a result, there is no significant overlap of the electron de Broglie waves at the center of the Sun.

This means that the laws of classical physics are adequate to describe the gas at the center of the Sun. In particular, we are justified in assuming that the material in the Sun, even at the very center where densities are highest, obeys the equation of state of a perfect gas.

9.6 Summary of Our Solar Model

We have found that, even without considering the generation of energy in detail, it is nevertheless possible, using the equation of hydrostatic equilibrium, to calculate the radial profiles of various physical parameters from center to surface. These are contained in our Tables 5.3, 7.1, and 9.1.

Our results allow us to appreciate the great range which is spanned by the various physical parameters in the Sun. Of the three principal parameters ($T, p,$ and ρ), we note that the smallest range is exhibited by the temperature: the central temperature exceeds the photospheric temperature by only about 3.5 orders of magnitude. Pressure exhibits the widest range: the central pressure exceeds the photospheric pressure by at least 12 orders of magnitude. Density presents an intermediate case: the central density exceeds the photospheric density by 8–9 orders of magnitude.

It is encouraging to find that, despite these great ranges, we have arrived at a model of the Sun which is consistent with a variety of observational data, despite the fact that the model is purely mechanical, i.e., our model has included the equation of momentum conservation but we have not paid much attention to the energy equation.

Of course it is precisely the generation of energy which sets the Sun (and stars) apart from other celestial bodies. Now that we have derived estimates of certain physical parameters inside the Sun, we need to examine the processes by which energy (which we have treated so far as a boundary condition) is actually generated.

However, before we enter into the details of nuclear reactions, we make a digression into a topic which at first sight seems to be rather idealized and far-afield from a study of the Sun. The topic has to do with mathematical entities called polytropes. However idealized these may seem, we shall find that we have already been working with models (in the convection zone, and in the radiative interior) which are actually not far removed from polytropes. Moreover, the discussion of polytropes will stand us in good stead in a subsequent chapter when we consider oscillations in the Sun.

Exercises

9.1 Perform the computation described in Section 9.1, using whatever computational technique that you prefer for the numerical work. How do your numbers at the center of the Sun compare with those in Table 9.1?

9.2 In step 6 of the procedure in Section 9.1, there is a recommendation for an approximation to the nonmonotonic radial profile of gravity inside the Sun. Experiment with different choices for the parameters of the approximation, e.g., choose a smaller value of r_m e.g., $0.20R_\odot$ or $0.15R_\odot$, and recalculate the model according to Section 9.1. How do the parameters at the center change? Then choose a larger value of r_m e.g., $0.30R_\odot$ or $0.35R_\odot$, and see how the central parameters change.

9.3 The formula in Equation 9.1 is based on the Kramers' opacity law. Use the two revised versions of Equation 9.1 which you obtained in

Exercise 1 of Chapter 8. For each case of the revised opacity "law," repeat the calculation of Section 9.1. How do your revised results for the parameters at the center of the Sun agree with the results of Bahcall et al. (2006)?

9.4 For each of your models, calculate the sound crossing time (Equation 9.6) and the associated frequency interval $\Delta\nu$ (express the frequency in μHz). How well does your value of $\Delta\nu$ agree with the observed solar value of (about) 135 μHz?

9.5 An alternative approach to modeling the radial profile of the gravity is to examine the profile of the mass parameter $M(r)$ in a detailed solar model, such as that on the website of J. Christensen-Dalsgaard. The tabulated values of the model are contained in the file http://www.phys.au.dk/~jcd/solar_models/fgong.l5bi.d.15c. A description of the different columns and rows can be found at: http://www.phys.au.dk/~jcd/solar_models/file-format.pdf. Extract $M(r)$ at a number of points in the tabulated model, and plot $g(r) = GM(r)/r^2$ as a function of $\log(r)$. A peak ($\approx$ 8.4 times the surface gravity) occurs at a certain value of $\log(r/R_\odot)(\approx -0.8)$. Is there a simple functional form (such as a parabola) that you can find to fit the peak? Use that functional form to obtain a better estimate of the local value of g in step 6 of the procedure in Section 9.1. How do the central parameters of the Sun change as a result?

References

Appourchaux, T. et al. 1998. "The art of fitting p-mode spectra: II. Leakage and noise covariance matrices," *Astron. Astrophys. Suppl.*, 132, 121.

Bahcall, J. N., Serenelli, A. M., and Basu, S. 2006. "10,000 standard solar models: a Monte Carlo simulation," *Astrophys. J. Suppl.*, 165, 400.

Chapter 10

Polytropes

Now that we have computed a model of the solar interior, albeit a simplified one, it is worthwhile to pay a certain amount of attention to a particular aspect of our solutions. This digression will be valuable in a later chapter when we come to consider how to compute the properties of oscillations in a solar model.

10.1 Power-Law Behavior

Inspection of Equation 9.1 indicates that, when we consider regions in the radiative interior of the Sun where the temperature is sufficiently large, the constant term in Equation 9.1 can be neglected. In this limit, we see that the pressure p varies as a power law of T : $p \sim T^{\beta}$. In the particular case we are considering, the exponent β has the numerical value of 4.25. The origin of this particular numerical value for the exponent can be traced to two simplifications: (i) $L(r)/M(r)$ remains constant as a function of radial location, and (ii) the opacity depends on density and temperature according to "Kramers' law": $\kappa \sim \rho T^{-3.5}$. These are simplifications which deal with matters of (i) energy generation, and (ii) energy transport. The fact that the radial profiles of pressure and temperature are related to each other by a power law therefore depends on certain assumptions we have made concerning the *energy equation*.

In a very different context, the power-law relation $p \sim T^{\beta}$ also emerged when we were modeling the deep convection zone: see Equations 6.14 and 7.9. Detailed models of the solar convection zone indicate that throughout some 90% of the depth of the convection zone, the power law in the $p \sim T$ relationship is close to 2.5. In the case of the convection zone, the reason for the power law behavior can be traced to the physics of adiabatic processes, i.e., on processes which enter into the transport of energy. Once again, certain assumptions which we make concerning the *energy equation* lead to a power law behavior between p and T.

It is noteworthy that we have encountered, in two quite different regions of the Sun, a functional dependence of pressure on temperature which can be written as a power law. In both cases, the emergence of the power law has to do

with *certain assumptions about the energy equation*, although the assumptions are quite different in the two different regions of the Sun.

In this chapter, we discuss the properties of equilibrium gas spheres ("polytropes") in which no explicit attention whatsoever is given to the energy equation. Instead of attending to the details of the energy equation, a certain functional form is *assumed to exist* for the relationship between pressure and density. Specifically, pressure and density are assumed to be related by means of a *power law*. For a perfect gas, this means that pressure and temperature are also related by a *power law*. A power law relationship between pressure and density is referred to as a "polytropic" equation of state.

In our search for deriving radial profiles of physical parameters inside the Sun (or for that matter inside any star), the only reason why polytropes have any claim on our attention is that in "real stars," the generation and transport of energy occurs in fact in such a way that, in certain cases, the pressure and density *do* turn out to obey the polytropic functional form, at least over certain ranges of the radial coordinate.

10.2 Polytropic Gas Spheres

A polytrope is defined to be a medium in which the pressure and density are related by the following relationship:

$$p = K\rho^{(n+1)/n} \tag{10.1}$$

where K and n are constants. The constant n is referred to as the "polytropic index." (In previous chapters, lower case n has been used to denote number density of atoms/electrons in a gas. In the present chapter, and also in Chapter 14, there are historical reasons for using n as the polytropic index, a dimensionless number, which has nothing to do with density.) The properties of polytropic spheres of gas have been discussed by a number of authors, including Lane, Ritter, Kelvin, Emden, and Fowler. A detailed study can be found in a book by Chandrasekhar (1958).

Why are polytropes relevant to our study of the Sun? Because, in a non-ionizing medium which obeys the perfect gas equation of state, $p \sim \rho T$, the polytropic relationship can be written in the form of a power law relationship between pressure and temperature: $p \sim T^{\beta}$, where β takes on a specific value: $\beta(\text{polytrope}) = n + 1$.

This leads us to consider an application of polytropic concepts to the two portions of the Sun in which we have already identified a power-law relationship between pressure and temperature.

First, in the adiabatic portions of the convection zone, the fact that p varies as $T^{2.5}$ suggests that the radial profile of the physical properties of those portions of the Sun is related to the radial profile of a polytrope with

index $n = 1.5$. We may say that the structure of the Sun in the convection zone corresponds to an "effective polytropic index" of 1.5.

Second, in the radiative interior, the fact that p varies as $T^{4.25}$ suggests that a polytrope with index $n = 3.25$ might provide useful information on the radial profile of physical parameters. In this case, we may say that the structure of the Sun in the radiative interior corresponds to an "effective polytropic index" of 3.25.

To make this more quantitative, it is instructive to calculate models of polytropes. As was mentioned above, in this chapter, we continue the practice of not referring to the energy equation explicitly: all details of energy generation and transport are subsumed into Equation 10.1. Then a polytrope model is obtained by solving the equation for conservation of mass, and the (static version of the) equation for conservation of momentum. In this way, we are using information about the *mechanical* properties of the star without including the *thermal* properties explicitly. Despite this limitation, since the sound speed at any radial location r depends on the ratio of $p(r)$ to $\rho(r)$ (both of which we shall calculate), our polytrope models will still provide sufficient information to allow us to study quantitatively the propagation of acoustic waves through the star.

The equation of mass conservation is:

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \tag{10.2}$$

As we have already seen (Chapter 7, Equation 7.1), the equation of momentum conservation (in the limit of zero velocity) is simply the equation of HSE:

$$\frac{dp(r)}{dr} = -\frac{GM(r)\rho(r)}{r^2} \tag{10.3}$$

These can be combined into a single second order equation:

$$\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2}{\rho(r)}\frac{dp(r)}{dr}\right) = -4\pi G\rho(r) \tag{10.4}$$

This is Poisson's equation for a self-gravitating sphere. It is a second order equation which includes two unknown functions of the radial coordinate: $p(r)$ and $\rho(r)$. The polytropic assumption, i.e., that p is related to ρ at each and every value of r by the relation $p(r) = K\rho(r)^{(n+1)/n}$, allows Equation 10.4 to be reduced to an equation for a single function of r.

Let the density and pressure at the center of the polytrope by ρ_c and p_c. Then at all values of radial location r, the local pressure and density satisfy the relation

$$\frac{p(r)}{\rho(r)^{(n+1)/n}} = \frac{p_c}{\rho_c^{(n+1)/n}} \tag{10.5}$$

10.3 Lane–Emden Equation: Dimensional Form

Let us introduce a dimensionless function y of the radial coordinate according to the definition

$$y^n = \frac{\rho(r)}{\rho_c} \tag{10.6}$$

The function y is referred to as the Lane–Emden function. The goal of the polytropic exercise is to derive, or compute, the function y as a function of the radial location. In view of the definition of y, it is clear that the boundary condition on y at the center of the polytrope is $y(r = 0) = 1$.

Inserting Equation 10.6 into Equation 10.5, we find that

$$\frac{p(r)}{p_c} = y^{n+1} \tag{10.7}$$

From Equations 10.6 and 10.7, we see that at any radial location, $y \sim p(r)/\rho(r)$. From the equations of thermodynamics, we already know that the ratio of pressure to density in a medium is related to the (square of the) sound speed. Thus, the numerical value of y at any radial location is proportional to the local sound speed. Because of this, in our subsequent study of helioseismology, we shall be able to use polytropes to determine realistic global properties of oscillation modes which involve the propagation of acoustic waves in a sphere.

If the material of which the polytrope is composed happens to obey the perfect gas equation of state, then at any given radial location, $y \sim p(r)/\rho(r)$ is also proportional to $T(r)$, the local temperature. In fact, $y(r) = T(r)/T_c$, where T_c is the central temperature. In this case, Equations 10.6 and 10.7 indicate that $\rho(r)$ scales as $T(r)^n$, while $p(r)$ scales as one higher power $T(r)^{n+1}$. We have seen scalings of this kind earlier: see Chapter 7, Equations 7.8 and 7.9.

Using Equations 10.6 and 10.7 to replace $\rho(r)$ and $p(r)$ in Equation 10.4, and collecting all the constants on the left-hand side of the equation, we find

$$\frac{(n+1)p_c}{4\pi G\rho_c^2}\left(\frac{1}{r^2}\right)\frac{d}{dr}\left[r^2\frac{dy}{dr}\right] = -y^n \tag{10.8}$$

Given a value of the polytropic index n, the radial profile of the function y can be obtained by solving Equation 10.8.

10.4 Lane–Emden Equation: Dimensionless Form

In order to convert the Lane–Emden equation to dimensionless form, we introduce a new unit r_o, the Emden unit of length, which is defined by

$$r_o^2 = \frac{(n+1)p_c}{4\pi G\rho_c^2} \tag{10.9}$$

How can we be sure that r_o has the dimensions of length? To answer this, we note that on the right-hand side of Equation 10.9, we can first isolate the ratio of p_c to ρ_c: this is related to the square of a sound-speed. Thus, the ratio has dimensions of $[\text{length}]^2/[\text{time}]^2$. The remaining dimensional units are those belonging to $1/G\rho_c$. In Chapter 1, Section 1.10, we mentioned a characteristic period P_g associated with the gravitational field of the Sun. The value of P_g is proportional to $\sqrt{(R_\odot^3/GM_\odot)}$. The dimensions of M/R^3 are those of density, indicating that the combination $1/\sqrt{(G\rho)}$ has the dimensions of [time]. Therefore the factor $1/G\rho_c$ in Equation 10.9 has the dimensions of $[\text{time}]^2$. Combining the dimensions, we see that the dimensions of the right-hand side of Equation 10.9 are indeed $[\text{length}]^2$. Therefore, r_o has the dimensions of length.

Is the Emden unit of length related to a length scale which might be relevant in the context of the structure of "real stars"? In particular, would it be useful to consider the dimensions of a star such as the Sun in terms of r_o? Or does the radius of a "real star" differ from r_o by many orders of magnitude? To answer these questions, we evaluate r_o using the values of central density and pressure which we have already obtained in our simplified solar model. Substituting $p_c = 3.34\times10^{17}$ dyn cm^{-2} and $\rho_c = 141$ gm cm^{-3} from Table 9.1, we find that $r_o = 4.5\times10^9\sqrt{(n+1)}$ cm. We shall see below that the radius R_p of a polytropic star is larger than r_o by a factor x_1 where the numerical value of x_1 depends on the n value. For example, with $n = 3.25$, the numerical value of x_1 is about 8 (see Table 10.1 below). Moreover, for $n = 3.25$, we see that the Emden unit of length r_o has a numerical value of about 9.3×10^9 cm. Multiplying r_o by x_1, we find that R_p is about 7.4×10^{10} cm. This is within a few percent of the actual solar radius. So it appears that the linear dimensions of polytropes in which central pressures and densities overlap with those of the "real Sun" provide a realistic and useful unit of length for characterizing a star such as the Sun. Applicability of polytropes to other stars will be discussed below.

In terms of the Emden unit of length, we introduce a new dimensionless variable x for the radial coordinate: $x = r/r_o$. This allows us to rewrite Equation 10.8 in the form

$$\frac{1}{x^2}\frac{d}{dx}\left[x^2\frac{dy}{dx}\right] = -y^n \tag{10.10}$$

This is the dimensionless form of the *Lane–Emden equation.* It is an ordinary differential equation of second order containing one unknown, $y(x)$.

In certain cases, it is convenient to rewrite Equation 10.10 in terms of an auxiliary function z defined by $z \equiv xy$. Inserting this into Equation 10.10, we find that the Lane–Emden equation can also be written in the form

$$\frac{d^2z}{dx^2} = -\frac{z^n}{x^{n-1}} \tag{10.11}$$

10.5 Boundary Conditions for the Lane–Emden Equation

When we set out to calculate a polytropic model, the aim of the exercise is to determine how physical parameters vary between the center of the sphere and the surface. This involves solving Equation 10.10 for y as a function of the radial coordinate x. Once we have such a solution, a plot of y as a function of x will show, for a perfect gas, a curve which is proportional to the radial profile of temperature from center to surface. According to Equation 10.6, the radial profile of the density will be obtained by raising the local value of y to the power n (the polytropic index). According to Equation 10.7, the radial profile of the pressure will be obtained if the local value of y is raised to the power $n + 1$.

Since Equation 10.10 is second order, we need two boundary conditions (BC's) in order to obtain a unique solution for any given value of n. One BC is readily available from the definition in Equation 10.6: $y = 1$ at $x = 0$.

To obtain a second BC, it is helpful to consider how the acceleration due to gravity $g(r)$ is related to the Lane–Emden function $y(r)$. To derive such a relation, we recall that the value of $g(r)$ is related to the gravitational potential Φ by the formula $g = -d\Phi/dr$. This allows us to rewrite the equation of hydrostatic equilibrium in the form $dp/dr = \rho d\Phi/dr$, leading to $dp = \rho d\Phi$. In view of the definition of a polytrope (Equation 10.1), we can write $dp = [(n + 1)/n]K\rho^{1/n}d\rho$. This leads to the following differential equation relating ρ and Φ in hydrostatic equilibrium: $d\Phi \sim \rho^{\delta}d\rho$ where the exponent $\delta = -1 + (1/n)$. Integrating the equation, we find $\Phi \sim \rho^{1/n} + constant$. Typically, the gravitational potential is set to zero at infinity, where $\rho \to 0$. This choice leads to $\Phi \sim \rho^{1/n}$. Recalling the definition of y in Equation 10.6, we see that $\Phi \sim y$.

Now at the center of the Sun, where density approaches a constant value, we have already seen (Chapter 9, Section 9.1 (step 6, Equation 9.5)) that $g \to 0$ as $r \to 0$. In other words, $d\Phi/dr \to 0$ as $r \to 0$. Converting to the dimensionless length parameter x, this is equivalent to $dy/dx \to 0$ as $x \to 0$. This provides us with the second BC which we need in order to obtain a unique solution for Equation 10.10 for any specified value of the polytropic index n.

In order to satisfy the two BC's, a series expansion can be obtained near the origin. For a polytrope with index n, the result is found to be (e.g., Chandrasekhar, 1958, p. 95)

$$y = 1 - \frac{x^2}{6} + \left(\frac{n}{120}\right)x^4 - \cdots \tag{10.12}$$

10.6 Analytic Solutions of the Lane–Emden Equation

Since the boundary conditions both apply at $x = 0$, we obtain a solution for $y(x)$ (for any given value of n) by starting at the center of the polytrope and integrating outward.

We note that since $dy/dx \sim d\Phi/dr$, and $d\Phi/dr = -g$ (a negative number), the slope dy/dx is negative. Therefore, although y starts with the value $y = 1$ at x=0, the value of y decreases as x increases, for all values of n. Since y decreases as we move outward from the center, there exists a certain radial location, $x = x_1$ (which is different for different values of n), at which the value of y passes through zero for the first time. At that radial location, pressure and density are both equal to zero. The ratio of p/ρ (i.e., the temperature, if the medium obeys the perfect gas law) is also zero at $x = x_1$. Compared to the values of unity at the center of the polytrope, it is natural to consider that the first zero point of y corresponds to the "surface" of the polytrope.

Analytic solutions are known for the Lane–Emden equation for three particular values of n.

10.6.1 Polytrope $n = 0$

In this case, Equation 10.10 becomes

$$\frac{d}{dx}\left(x^2 \frac{dy}{dx}\right) = -x^2 \tag{10.13}$$

Integrating once, we find

$$x^2 \frac{dy}{dx} = -\frac{x^3}{3} + \text{const} \tag{10.14}$$

In order to satisfy the boundary condition $dy/dx = 0$ at $x = 0$, the constant must be zero. This leads to

$$\frac{dy}{dx} = -\frac{x}{3} \tag{10.15}$$

Integrating again, and applying the condition $y = 1$ at $x = 0$, we find

$$y(n = 0) = 1 - \frac{x^2}{6} \tag{10.16}$$

The first zero of $y(n = 0)$ occurs at $x_1 = \sqrt{6}$.

10.6.2 Polytrope $n = 1$

In this case, it is convenient to use Equation 10.11, which reduces, in the case $n = 1$, to the simple form

$$\frac{d^2 z}{dx^2} = -z \tag{10.17}$$

The solution of this equation, consistent with both boundary conditions at $x = 0$ is $z = \sin(x)$. Reverting to the solution for y, we have

$$y(n = 1) = \frac{\sin(x)}{x} \tag{10.18}$$

The first zero of $y(n = 1)$ occurs at $x_1 = \pi$.

10.6.3 Polytrope $n = 5$

Derivation of the solution in this case is more complicated than the two cases above. (See Chandrasekhar's 1958 book, pp. 93–94, for a derivation.) Here we simply state the result:

$$y(n = 5) = \frac{1}{(1 + x^2/3)^{1/2}} \tag{10.19}$$

The first zero of $y(n = 5)$ occurs at $x_1 \to \infty$. Thus, for the case $n = 5$, the equilibrium configuration of the polytrope is infinitely extended.

Numerical solutions of the Lane–Emden equation for arbitrary values of n will be discussed in Section 10.8.

10.7 Are Polytropes Relevant for "Real Stars"?

Note that in all three polytropes for which analytic solutions exist, inspection of the solutions indicates that y is a monotonically decreasing function of x for all values of x between 0 and x_1. This property also emerges from numerical solutions of the other polytropes, where only nonanalytic solutions exist. Recalling that y is proportional to temperature (in a perfect gas) (see Section 10.3), the fact that temperature decreases monotonically from center to surface indicates that if energy is generated at the center (by an unspecified mechanism), that energy will find itself in a medium which has a negative temperature gradient: this facilitates the transport of energy toward the surface. Here again, we come across a feature which makes it attractive to consider polytropes as structures which, although highly idealized, nevertheless have properties which are physically relevant in the context of modeling "real stars."

The most successful application of polytropes to stellar structure is found when one is modeling a star in which the equation of state in fact obeys the polytropic relation (Equation 10.1). Do such stars exist? Yes. We can summarize four examples.

First, the polytrope $n = 1.5$ is relevant to low-mass stars. We recall (Chapter 7) that the Sun has a convection zone which occupies a spherical shell with a finite thickness: the shell ends at a well-defined radial location ($r \approx 0.7 R_\odot$), so that only an outer envelope of the Sun is convective. It turns out that when models are computed for stars with masses less than the Sun's mass, the convective envelope becomes deeper, reaching ever farther into the star as we consider stars with lower and lower masses. Eventually, a mass M_c is reached where the convective "envelope" extends all the way to the center of the star. For stars with masses less than M_c, the entire star is convective, and the adiabatic limit of convection applies throughout essentially the entire star. Such stars can be represented quite well by the $n = 1.5$ polytrope.

Second, for quite different reasons, the polytrope $n = 1.5$ also turns out to be relevant to old stars called "white dwarfs." These are objects where nuclear burning is no longer happening: in these stars, electron degeneracy pressure (see Chapter 9, Section 9.5) supports the star against gravity. In such cases, if the electrons are nonrelativistic, Equation 10.1 applies with $n = 1.5$ and a value of K which depends only on certain physical constants. In this case, it can be shown that white dwarfs should obey a mass-radius relationship $R \sim M^{-1/3}$. This is a very different relationship from that which applies to solar-like stars: for the latter, the radius R increases as the mass increases, roughly as $R \sim M$. But white dwarfs are predicted to have radii which decrease as their mass increases. There is observational evidence to support this prediction (e.g., Provencal et al., 1998).

Third, if the electrons supporting a star are relativistic, it turns out that the equation of state is given by Equation 10.1 with $n = 3.0$ and another value of K, again determined by a (different) combination of physical constants. For the particular value $n = 3.0$, it turns out that a unique mass M_3 exists, determined by physical constants. For typical stellar compositions, M_3 is found to be close to $1.4M(\text{sun})$: this is the most massive object which can exist in hydrostatic equilibrium with support from degenerate electrons. It is remarkable that a polytrope which corresponds to an object being supported against gravity by the pressure of relativistic electrons has a unique mass of the same order as a "real star" such as the Sun. Nevertheless, this conclusion has emerged as of fundamental importance in observational attempts to probe the evolution of the stellar universe in its earliest stages. The stars which can be observed farthest away in space (and therefore farthest back in time), are exploding stars called supernovae. One class of supernova occurs when a white dwarf accumulates so much mass that it exceeds M_3: when that happens, the star cannot exist in equilibrium, but collapses and releases gravitational energy in an explosion which is so large that it can be seen all the way across the universe. The fact that each member of this class of supernova relies on the same physical principles allows cosmologists to assume that each member of the class is a "standard candle," with a unique output power. This allows a distance to be assigned to each such event.

Fourth, we have already seen (Chapter 9) that the Sun consists of distinct regions in which a polytropic equation is "not too bad": the convective envelope has $n = 1.5$ and the radiative core has $n = 3.25$. The Sun can therefore not be regarded as a "true polytrope" in the strict sense of the word. But how about considering the possibility of approximating the Sun as having a single "effective polytropic index" from surface to center? Might this help us to understand some of the global properties of the Sun? Let us see. From the results which have emerged from our model of the Sun (see Chapter 9, Section 9.6), we have seen that from surface to center, the temperatures, densities, and pressures increase by (roughly) 3, 9, and 12 orders of magnitude respectively. Now, if a single "effective polytropic index" n_e could be considered as applying to the Sun as a whole, let us recall that in a polytrope, ρ scales as T to

the power of n_e, while p scales as T to the power of $n_e + 1$ (see discussion between Equations 10.7 and 10.8). Therefore, an increase in T by 10^3 would be accompanied by increases in ρ and in p by 10^9 and 10^{12} respectively if $n_e \approx 3$. This value of n_e has a value that is intermediate between the values of 1.5 and 3.25 which are applicable to the Sun's envelope and core, respectively. As a result, even in the case of a composite object such as the Sun, the concept of a polytrope helps us (roughly) to understand why some of the global physical properties of the Sun behave in the way that they do.

In summary, the study of polytropes is not at all irrelevant as far as "real stars" are concerned. To be sure, the treatment is not complete: it tells us nothing about the sources of opacity, or the sources of energy. Nevertheless, there is useful information to be gained in this "first course" by considering the mechanical properties that polytropes allow us to describe.

10.8 Calculating a Polytropic Model: Step by Step

For arbitrary values of the polytropic index n, numerical solutions can be obtained for the Lane–Emden equation. These numerical solutions (e.g., Chandrasekhar, 1958) indicate that the first zeroes of polytropes with $n = 1.5$, $3.0, 3.25$, and 4.0 occur at $x_1 \approx 3.65, 6.90, 8.02$, and 15.0, respectively. In dimensional units, the radius of the corresponding polytrope is $R(n) = x_1 r_o$ where r_o is the Emden unit of length corresponding to the particular polytropic index.

By way of illustration, and because we shall use this particular case in discussing certain oscillations in the Sun (Chapter 14, Section 14.5) let us consider the polytrope $n = 3.25$. In this case, we have already pointed out (Section 10.4) that the value of r_o is 9.3×10^9 cm. Combining this with the appropriate value of x_1, we have seen that the radius of a complete $n = 3.25$ polytrope with a central pressure and density equal to that of our simplified solar model would be $R(3.25) = 7.4 \times 10^{10}$ cm. Of course, the Sun is not a complete polytrope, with a constant n value all the way from center to surface. Nevertheless, the dimensional radius that we determine for such a polytrope, is within 6% of the radius of the "real Sun." It is amazing that a structure which is as simple as a polytrope (and in which the energy equation is replaced by a gross simplification) can have macroscopic properties which are not far removed from those of an actual star.

To calculate the structure of a polytrope for arbitrary n, the aim is to compute the value of y at each of a tabulated list of values of x. Also, at each value of x, we wish to calculate the quantity $y'(= dy/dx)$. For numerical purposes, it is convenient to start with the version of the Lane–Emden equation given in Equation 10.11, where the function z is defined by $z = xy$.

Then we can rewrite Equation 10.11 in the form of two coupled first order differential equations for the functions $f_1 = z$ and $f_2 = dz/dx$. In terms of these functions, the Lane–Emden equation can be replaced by two equations for two unknowns:

$$f_2 = \frac{df_1}{dx} \tag{10.20}$$

$$\frac{df_2}{dx} = -\frac{f_1^n}{x^{n-1}} \tag{10.21}$$

We start to integrate these equations at $x = 0$ using the BC's $f_1 = 0$ and $f_2 = 1$. In order to start off the numerical integration correctly, we use the series expansion for the Lane–Emden equation near the origin (Equation 10.12). Then we find

$$f_1 = x - \frac{x^3}{6} + \left(\frac{n}{120}\right)x^5 \tag{10.22}$$

and

$$f_2 = 1 - \frac{x^2}{2} + \left(\frac{n}{24}\right)x^4 \tag{10.23}$$

The step-by-step procedure for calculating a polytrope, especially one that will be useful when we come to determining the oscillation properties (see Chapter 14), proceeds as follows. The goal is to obtain a table of values of three quantities (x, y, and y'), extending from $x = 0$ (the center of the "star") to $x = x_1$ (the surface of the "star"). The process is as follows.

1. Choose a value for the polytropic index n.

2. The first entries in the table refer to the center of the star. They are $x(1) = 0$, $y(1) = 1$, and $y'(1) = 0$.

3. Choose a step size Δx which may be as small as you like. A value $\Delta x = 0.01$ will eventually lead to a table of values which, for $n = 3.25$, will contain about 800 rows.

4. Advance the x value to its value for the second row in the table: $x(2) = \Delta x$. Use $x(2)$ in Equations 10.22 and 10.23 to calculate the corresponding values of $f_1(2)$ and $f_2(2)$. Then the value of $y(2)$ is given by $f_1(2)/x(2)$. And the value of $y'(2)$ is given by $y'(2) = (f_2(2) - y(2))/x(2)$.

5. For the third row of the table, we advance to $x(3) = x(2) + \Delta x$. Now we have enough information to start to use an integrator (such as a Runge-Kutta routine) to step forward the solution of Equations 10.20 and 10.21. This leads to values of $f_1(3)$ and $f_2(3)$ which we then convert to $y(3)$ and $y'(3)$ using the expressions in step 4.

TABLE 10.1: Solution of Lane–Emden equation for the polytrope $n = 3.25$ (Notation: $x \cdot x$ D$_{yy} = x \cdot x$ times 10^{yy})

x	y	y'
0.00	1.0	0.0
0.02	0.99993D+00	−0.66662D-02
0.10	0.99833D+00	−0.31573D-01
0.20	0.99337D+00	−0.64203D-01
0.30	0.98521D+00	−0.95610D-01
0.40	0.97400D+00	−0.12524D+00
0.50	0.95995D+00	−0.15262D+00
0.70	0.92434D+00	−0.19925D+00
1.00	0.85655D+00	−0.24651D+00
1.50	0.72480D+00	−0.27000D+00
2.00	0.59385D+00	−0.24945D+00
2.50	0.47832D+00	−0.21231D+00
3.00	0.38202D+00	−0.17418D+00
3.50	0.30362D+00	−0.14111D+00
4.00	0.24015D+00	−0.11434D+00
4.50	0.18856D+00	−0.93307D-01
5.00	0.14626D+00	−0.76917D-01
5.50	0.11119D+00	−0.64141D-01
6.00	0.81774D-01	−0.54120D-01
6.50	0.56812D-01	−0.46189D-01
7.00	0.35392D-01	−0.39844D-01
7.50	0.16823D-01	−0.34709D-01
8.00	0.63652D-03	−0.30492D-01
8.02	0.28430D-04	−0.30332D-01

6. For each new row of the table, increase the x value by Δx, and compute the updated values of y and y'.

7. It is easy to see when the integration must be stopped: y cannot take on negative values. Therefore, the last row of the table should contain a y value which is close to zero, say, $y \leq 0.001$. The last row in the table should contain an x value that is close to x_1 for the polytrope you have chosen, e.g., for $n = 3.25$, the value of x_1 is known to be 8.01894 (Chandrasekhar, 1958). All values of y' in the table will be negative numbers.

An example of an abbreviated table for the polytrope $n = 3.25$ is given in Table 10.1. We will have occasion to use the results in (an expanded version of) Table 10.1 in Chapter 14, when we calculate the periods of a certain class of oscillations known as g-modes in a polytrope. It will be instructive to compare the periods to the values which are observed for certain oscillations in the Sun. We shall find that once again, the use of a polytrope, however idealized, to

TABLE 10.2: Central condensation in various polytropes

$n = 1.0$	1.5	2.0	3.0	3.25	3.5
$C_c = 3.29$	5.99	11.40	54.18	88.15	152.9

describe the structure of a star (e.g., the Sun) provides information which is quite useful in interpreting data from the "real Sun."

10.9 Central Condensation of a Polytrope

A polytrope has the property that, when one evaluates the gradient y' at the surface of the "star," one can then calculate (Chandrasekhar, 1958) the ratio C_c of the central density ρ_c to the mean density $\rho_m = M/(4/3)\pi R^3$. The quantity C_c is referred to as the "central condensation." Values of C_c for some polytropes (taken from Chandrasekhar, 1958, his Table 4) are given in Table 10.2.

We have already noted (Chapter 9, Section 9.2) that the "real Sun" has $C_c \approx 100$. Therefore, in terms of central condensation, the Sun behaves like a polytrope with an index n slightly larger than 3.25. Recall (Chapter 10, Section 10.2) that for the radiative interior of the Sun, there are physical reasons (related to opacity) why the polytrope $n = 3.25$ is relevant to the relationship between pressure and temperature.

Exercises

10.1 Use the step-by-step procedure in Section 10.8 to calculate a table of values x_i $(i = 1, 2, 3, \ldots)$ of $y_i = y(x_i)$ and $y' = dy/dx$ from center to surface for the polytropes $n = 1.0, 1.5$, and 3.25.

10.2 For the case $n = 1$, also evaluate the analytic solution $y_a(x) = \sin(x)/x$ for each x_i. For each entry in the table, x_i, calculate the fractional difference $\delta y/y$ between your numerical $y(x_i)$ and the analytic solution $y_a(x_i)$. Repeat the calculation with a smaller and a larger choice of step size Δx. How do the fractional differences $\delta y/y$ change?

10.3 At the "surface" of each polytrope in Exercise 1, your table will give you the local values of x and y'. For each polytrope, use those surface values to evaluate the quantity $-x/3y'$ at the surface. Compare the results with the central condensations C_c listed in Table 10.2.

References

Chandrasekhar, S. 1958. "Polytropic and isothermal gas spheres," *An Introduction to the Study of Stellar Structure.* Dover Publications, New York, pp. 84–182.

Provencal, J. L., Shipman, H. L., Hog, E., and Thejll, P. 1998. "Testing the white dwarf mass-radius relation with HIPPARCOS," *Astrophys. J.*, 494, 759.

Chapter 11

Energy Generation in the Sun

Historically, the source of energy generation on the Sun has been attributed to a number of causes, including gravitational collapse and radioactive decay. The possibility that nuclear fusion might be the source of solar energy could not be evaluated quantitatively until certain key pieces of information were in place. In particular, the masses of the relevant isotopes had to be measured to three or four significant digits before it became evident that atomic masses, although close to integer values, actually deviated from integers by small, but systematic, amounts. The deviations amount to only a few parts per thousand, but those small deviations are at the very heart of nuclear energy generation.

The characteristic which sets the Sun (and stars in general) apart from other structures in the universe is precisely the fact that *the Sun is able to generate its own supply of energy by means of nuclear reactions.* On Earth, nuclear reactions can be made to happen by accelerating particles to energies of millions of electron volts (MeV), and then "slamming" the fast particles into a target nucleus. But there are no MeV accelerators in the Sun. Instead, the only available particles are those belonging to a *thermal population* in which the mean energies are much smaller than 1 MeV: mean thermal energies in the Sun's core (where $T = 15$–16 MK) are of order $kT \approx 1$ keV only. Despite energies of merely keV, the fact remains that the solar particles *can* (and do) participate in nuclear reactions. The fact that the reacting particles are *thermal* gives rise to the term "thermo-nuclear reactions" to describe the process whereby light nuclei in the Sun fuse into heavier ones. The (slight) loss of mass which occurs in the fusion reactions emerges in the form of kinetic energy and energetic photons. It is this emergent energy which makes the Sun a power generator.

Two distinct cycles of reactions were identified, almost simultaneously, by Bethe in the years 1938 and 1939 as possibly contributing to nuclear energy generation in the Sun. The cycles are referred to as the pp-cycle (Bethe and Critchfield, 1938) and the CNO-cycle (Bethe, 1939). Recent models of the Sun indicate that the CNO cycle contributes only 0.5% to the Sun's energy output (Bahcall et al., 2005). In this first course on solar physics, we focus on the predominant pp-cycle.

The important questions in the context of solar energy generation are: (i) Which reactions occur? (ii) How much energy is liberated in each reaction? (iii) How many reactions occur per second? Now that we know certain physical parameters in the Sun, we can address these questions in turn.

TABLE 11.1: Isotope nuclear masses in atomic mass units (a.m.u.)

Proton (H^1)	1.0072764
Neutron (n)	1.0086649
Deuteron ($D = H^2$)	2.0135532
Helium-3 (He^3)	3.0149321
Helium-4 (He^4)	4.0015061

In the following discussion, masses of the relevant nuclei (see Table 11.1) will be cited in terms of atomic mass units (a.m.u.). The data in Table 11.1 are obtained by starting with *atomic* masses (Audi and Wapstra, 1993) and then subtracting electron masses to obtain the *nuclear* masses. In c.g.s. units, according to the NIST Reference list (http://physics.nist.gov/cuu/Constants/), 1 a.m.u.$= 1.6605388 \times 10^{-24}$ gm. The rest-mass energy equivalent of 1 a.m.u. is $E(1) = 1.4924178 \times 10^{-3}$ ergs. Expressed in units of electron volts (1 eV $= 1.6021765 \times 10^{-12}$ ergs), we find $E(1) = 931.494$ MeV.

11.1 The pp-I Cycle of Nuclear Reactions

In the Sun, the most common set of reactions which occur are referred to as the pp-I cycle. There are also less common cycles referred to as pp-II and pp-III, but all have the same overall end-result, namely, four protons are fused into one helium nucleus. We shall return to the pp-II and pp-III cycles in the next chapter when we discuss neutrinos. In the present section, where energy generation is the principal focus of our discussion, we confine our attention to the pp-I cycle.

There are three reactions to be taken into account in the pp-I cycle.

$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{D} + \mathrm{e}^+ + \nu \, (E = 1.442 \text{ MeV}) \qquad \text{(a)}$$

Here, p + p denotes the reaction of two protons, both of which belong to the thermal distribution which exists at any given radial location with local temperature $T(r)$. The reaction products include a deuteron (a nucleus consisting of one proton and one neutron), a positron (e^+), and a particle known as a neutrino (ν: see Chapter 12).

In order to determine the amount of energy which is released in reaction (a), we use the masses of the various particles in Table 11.1. Using these, we find that the combined mass on the left-hand side of reaction (a) is 2.0145528 a.m.u. This exceeds the deuteron mass by $\Delta m = 0.0009996$ a.m.u. The fractional excess in mass is small, only 0.1%, but the existence of an excess (however small) ensures that the reaction is exothermic. In energy units, the corresponding energy is $c^2 \Delta m = 0.931$ MeV. The positron is an antiparticle which requires an equivalent rest-mass energy of 0.511 MeV. The net energy which is available for the neutrino and D from reaction (a) is the remainder $0.931 - 0.511 = 0.420$ MeV: this is the maximum ("end-point") energy that

the neutrino can carry away. Actually, because the reaction energy is shared by three particles, the neutrino energy has a continuous spectrum between zero and 0.420 MeV: the average energy carried off by the neutrino is about one-half of the end-point energy, i.e., about 0.2 MeV. The positron quickly annihilates on an ambient electron, releasing an energy of 1.022 MeV. Thus, the total amount of energy released into the core of the Sun by reaction (a) is 1.442 MeV. Allowing for the average energy carried off by the neutrino, the amount of energy which is available (on average) to be deposited in the core of the Sun, thereby contributing to the thermal energy pool, is about 1.2 MeV.

The second reaction in the pp-I cycle is

$$\mathrm{p} + \mathrm{D} \rightarrow \mathrm{He}^3 + \gamma(E = 5.493 \text{ MeV}) \qquad \text{(b)}$$

In this second step of the pp-I cycle, a proton from the thermal population reacts with a deuteron produced in reaction (a), and forms a nucleus of He^3 plus an energetic photon. The combined masses of p and D on the left-hand side (= 3.0208296 a.m.u.) exceeds the mass of the He^3 nucleus (see Table 11.1) by $\Delta m = 0.0058975$ a.m.u. Once again, the fractional mass excess is small, about 0.2%, but it is finite. Therefore, the reaction is exothermic, with an energy release $c^2\Delta m$ of 5.493 MeV. This energy is divided between the He^3 nucleus and the photon.

The final reaction in the pp-I cycle is

$$\mathrm{He}^3 + \mathrm{He}^3 \rightarrow \mathrm{He}^4 + 2\mathrm{p}\,(E = 12.860 \text{ MeV}) \qquad \text{(c)}$$

In this third step of the pp-I cycle, after reactions (a) and (b) have occurred twice, the two He^3 nuclei fuse to create one nucleus of He^4, releasing two protons. The sum of the rest masses of two He^3 on the left-hand side (6.0298642 a.m.u.) exceeds the sum of the rest masses of the three particles on the right-hand side (6.0160589 a.m.u.) by $\Delta m = 0.0138053$ a.m.u. The corresponding energy release $c^2\Delta m$ is 12.860 MeV.

For the Sun to produce energy by hydrogen fusion, it is essential that in each reaction of the above cycle, the combined mass of the products is *less* than the combined mass of the reactants. In the early days of measuring atomic weights, when the masses of the isotopes were known with a precision of only two significant digits, the mass of the reactants would equal the mass of the products. It is only when the atomic weight measurements reach a precision of at least three significant digits that the mass difference $2m(\mathrm{p}) - m(\mathrm{D})$ becomes a nonzero number. And in order to derive the energy release in the reaction with a precision of N significant digits, the isotopic masses have to be measured with precisions of $N + 2$ significant digits.

11.2 Reaction Rates in the Sun

Altogether, in a complete pp-I cycle, consisting of two reactions each of (a) and (b), plus one reaction (c), the total energy released is $2(1.442 + 5.493) +$

12.860 = 26.73 MeV. However, some of this is carried off by neutrinos. The amount of energy which is deposited into the thermal pool of the Sun's core, and which can therefore, contribute to the radiant output power of the Sun, is roughly $2(1.2 + 5.493) + 12.860 = 26.25$ MeV.

This is the amount of thermal energy which is released into the thermal pool in the Sun's core when four protons fuse into one helium nucleus. Converting to c.g.s. units, each pp-I cycle generates $\Delta E(\text{pp-I}) = 4.206 \times 10^{-5}$ ergs.

Now, we already know the total output power of the Sun (Chapter 1, Section 1.4): $L_\odot = 3.8416 \times 10^{33}$ ergs sec^{-1}. Therefore, since pp-I cycles are by far the largest source of energy generation in the Sun, the number of these reactions which occur in the Sun every second (i.e., the frequency of the reactions) is

$$F_r = \frac{L_\odot}{\Delta E(\text{pp-I})} = 0.91 \times 10^{38} \text{ reactions sec}^{-1} \tag{11.1}$$

The Sun also relies, in a small percentage of cases, on pp-II and pp-III cycles: however, both of those cycles also begin with reactions (a) and (b) above, and their rates are controlled primarily by the (slowness of) reaction (a). Moreover, some ($< 1\%$) of the solar energy output comes from the CNO-cycle: in this cycle, carbon acts as a catalyst to bring about the same overall effect as in the pp-cycle, namely, fusing four protons into one helium nucleus.

In summary, we will not make a gross error if we take $F_r \approx 10^{38}$ per second as the number of nuclear reactions which occur in the Sun every second.

11.3 Proton Collision Rates in the Sun

In order to set the reaction rate F_r in context, let us compare F_r to the overall rate F_c at which collisions between protons occur in the nuclear-generating core of the Sun. By the word "collision," we mean an event in which the momenta of the individual particles are altered, in a manner analogous to the collision of two billiard balls. Two protons which at first happen to approach each other, feel an increasingly strong Coulomb repulsion, which eventually causes the two to move apart, changing directions and speeds in such a way as to conserve energy and momentum.

With a mean velocity of V, and a number density of n_p protons cm^{-3}, the rate at which a single proton in the Sun experiences momentum-altering collisions with other protons is $f_c = n_p V \sigma$ per second, where σ is the momentum collision cross-section. (Lower case f denotes the collision rate for a *single* proton.) Between two protons, the value of σ is determined by the Coulomb force. To calculate the Coulomb cross-section σ_c, we note that in a gas with temperature T, the mean kinetic energy of thermal motion, of order

kT, allows two protons to approach one another within a minimum distance r_m such that $e^2/r_m \approx kT$. Such close collisions result in large deflections of the protons in their motion. The cross-sectional area associated with r_m (i.e., πr_m^2) would be a reasonable estimate for σ_c if large deflections were the only contributors to deflecting protons in their motion. But because the Coulomb force is a long-range force, protons are also subject to a multitude of small deflections as a result of distant collisions. The net effect of these is to yield a cross-section which is larger than the estimate based on r_m by a multiplying factor called the Coulomb logarithm. It is conventional to write $\sigma = \pi r_m^2 \Lambda$, where Λ is a logarithmic term which includes the effects of distant collisions. Thus, $\sigma_c \approx \pi e^4 \Lambda/(kT)^2$. (Note that, for thermal particles, with $V_{th} \sim \sqrt{T}$, the Coulomb cross-section σ_c scales as $1/V_{\rm th}^4$. i.e., fast particles have many fewer collisions than slow particles.) In the core of the Sun, where $n_p \approx 10^{26}$ cm^{-3} (see Chapter 9, Section 9.2) and $T = (1.5\text{–}1.6) \times 10^7$ K, the value of $\Lambda \approx 5 - 8$. This leads to $\sigma_c \approx (2\text{–}3) \times 10^{-19}$ cm^2. Since the mean thermal velocity of a proton in the core of the Sun is $V_{\rm th} \approx 6 \times 10^7$ cm sec^{-1} (Chapter 9, Section 9.2), we see that each proton undergoes $f_c \approx (1\text{–}2) \times 10^{15}$ momentum-changing collisions per second.

The overall rate F_c of Coulomb collisions in the core of the Sun is given by f_c times the total number of protons $N_p(c)$ in the core. The core of the Sun, in which nuclear reactions occur, is confined, according to detailed models, within the innermost 20% (or so) of $R_\odot$. Although the volume of this core is a small fraction of the total solar volume, the high densities in the core have the effect that the mass of the core is of order $0.1 M_\odot$, i.e., about 2×10^{32} gm. Dividing this by the mass of a proton, we find, $N_p(c) \approx 10^{56}$. This leads to $F_c = f_c N_p(c) \approx 10^{71}$ momentum-changing collisions occur every second in the nuclear-generating core of the Sun.

Comparing the reaction rate $F_r (\approx 10^{38}$ sec$^{-1})$ in the solar core with the collision rate F_c in the core, we see that a proton undergoes (on average) a huge number of collisions N_c(react), where

$$N_c(\text{react}) \approx \frac{F_c}{F_r} \approx 10^{33} \text{ collisions} \tag{11.2}$$

before that proton ever participates in a pp-I cycle in the Sun's core. Since an individual proton experiences on average $f_c \approx (1\text{–}2) \times 10^{15}$ collisions per sec, each proton in the Sun will participate in a pp-cycle only after an average time-span of about $N_c(\text{react})/f_c \approx (5\text{–}10) \times 10^{17}$ sec, i.e., once in 15–30 billion years. This explains why the life-time of the Sun in its hydrogen-burning phase will be of order 10 billion years.

This discussion indicates that the occurrence of the nuclear reactions in the pp-cycle is a very rare event indeed in the conditions of the Sun's core: *only one collision in* 10^{33} *results in a nuclear reaction.* It is worth while examining why this is so: part of the answer will lead us to understand why thermonuclear reactions are very sensitive to temperature.

11.4 Conditions Required for Nuclear Reactions in the Sun

Now that we know that nuclear reactions are occurring in the Sun, it is worthwhile to consider the physical requirements that must be satisfied before a reaction can occur at all.

In order to have a nuclear reaction occur, whether in the Sun or in the laboratory, certain conditions have to occur. First, two nuclei must undergo a collision with each other. The collision must be of a particular kind. We are not interested merely in momentum-changing collisions where the particles stay far apart and experience only a "glancing" blow off each other. Such "distant" collisions are certainly important in a plasma when we wish to evaluate certain transport coefficients in the plasma: because the Coulomb force is long-range, the overall effect of many distant collisions can dominate over the rare large-angle collisions. (This is in fact the origin of the factor Λ in the Coulomb cross-section mentioned above.) However, distant collisions of this kind contribute nothing to nuclear reactions.

11.4.1 Nuclear forces: short-range

In order for a nuclear reaction to occur, it is essential that two nuclei must approach one another so closely that the strong force, which binds nucleons (protons, neutrons) together inside a nucleus, can come into play. How close do such collisions have to be?

The answer depends on the range of the strong force. From measurements of nuclear size, it is known that nuclei have radii which are a few times 10^{-13} cm. This indicates that the strong force operates only within a finite length-scale, of order $r_N \approx 10^{-13}$ cm ($\equiv 1$ "fermi" [fm]). Inside 1 fm, the force between two nucleons is very strongly attractive. As a result, nuclear reactions occur if, and only if, two nuclei can be brought as close together as r_N.

The same can be said about the way that the weak force operates in the Sun: in order for the pp chain of reactions to occur, the weak force must cause a proton to "decay" into a neutron. This is only possible if the proton is inside the deep potential well of the nuclear force. In effect, the weak force in the Sun does its work only when particles are within a distance of r_N.

As a result, if two nuclei approach each other no closer than, say, 10^{-11} cm (or more), neither the nuclear force nor the weak force has a chance to come into play. Such nuclei simply have a momentum-changing collision, bouncing off each other and continuing on their way, completely unchanged as far as their nuclear properties are concerned.

How strong is the attractive force which holds two nucleons together? Well, it certainly has to be strong enough to overcome the Coulomb repulsion. The Coulomb repulsion between two protons separated by only 1 fm has a potential energy $e^2/r \approx 1.5$ MeV. Moreover, the nucleons inside a nucleus must not

allow their de Broglie waves to "leak out of the nucleus": this requires that each nucleon must be moving with a speed V_n which is so fast that h/mV_n is no larger than a few fermi. At such speeds, a proton has a kinetic energy of about 10 MeV. Such high speeds tend to disrupt the nucleus, and so the strong force has to have an attractive energy of at least 10 MeV. Detailed calculations suggest that the strong force has an attractive energy of 30–40 MeV.

11.4.2 Classical physics: the "Coulomb gap"

According to classical physics, when two positive point charges $+Z_1e$ and $+Z_2e$ are separated from each other by a distance r, they experience a Coulomb repulsive force. The potential energy of the repulsion is $Z_1Z_2e^2/r$. The closer the two particles approach each other, the stronger the repulsion becomes.

How closely can such particles be made to approach each other? In classical terms, the answer is straightforward: the two can come no closer than a distance r_c where their relative kinetic energy (KE) is equal to the repulsive potential energy. Let the masses by $A_1\ m_p$ and $A_2\ m_p$ where m_p is the proton mass. In terms of the reduced mass Am_p of the two nuclei ($A = A_1\ A_2/(A_1 + A_2)$), the average KE is given by $0.5\ Am_pV^2$.

This leads to

$$r_c = \frac{2Z_1Z_2e^2}{Am_pV^2} \tag{11.3}$$

For the collision of two protons, this reduces to

$$r_c = \frac{4e^2}{m_pV^2} \tag{11.4}$$

In order to appreciate how nuclear reactions occur in the Sun, and in order to appreciate that something beyond classical physics is at work, we need to ask a specific question: what is the magnitude of r_c for two protons near the center of the Sun? With a mean thermal speed of ≈ 600 km sec^{-1} (Chapter 9, Section 9.2), and inserting the values of e and m_p, we find that $r_c \approx 1.5 \times 10^{-10}$ cm ≈ 1500 fm.

The critical point of this result is that r_c *greatly exceeds* the range of the nuclear force r_N. Specifically, with the values we use above, the classical distance of closest approach of two protons in the center of the Sun is roughly 1000 times larger than the nuclear force range.

To be sure, not all of the protons have velocities equal to the rms speed. There are some faster ones. For example, in a thermal distribution, one proton in e^{10} (i.e., one proton in 20,000) has a speed which exceeds the mean by a factor of 3.2. If two such protons collide, then their distance of closest approach, based on classical physics, would be reduced to ≈ 150 fm. Even so, this is still some 100 times larger than r_N, much too far apart for the strong force to operate.

Because of the Coulomb repulsive force, classical physics indicates that protons in the core of the Sun are *forbidden* from approaching each other closely enough to allow the nuclear force to come into play. In classical terms, two protons are always separated by a distance which is at least as large as r_c: we refer to r_c as the "Coulomb gap."

If the Sun were governed by classical physics alone, the Coulomb gap would be an insuperable barrier which would prevent any nuclear reactions from occurring in the Sun in its present condition. In order to understand why the Sun shines, we are forced to the following important conclusion: we need to go beyond classical physics. We must admit that the Sun is an object in which quantum physics plays an essential role.

11.4.3 Quantum physics: bridging the "Coulomb gap"

So, let us enter the world of quantum mechanics. In this world, particles in certain circumstances no longer behave as points: a particle of mass m moving with speed V has a finite probability of occupying an extended region of space. This nonpoint-like behavior is modeled by saying that a quantum particle can be represented by an associated wave. The wavelength is given by a formula first derived by de Broglie (1924):

$$\lambda_p = \frac{h}{mV} \tag{11.5}$$

where h is Planck's constant. According to de Broglie, a quantum particle can be considered as being "spread out" over a finite distance of order λ_p.

Now we come to the heart of the matter of nuclear fusion in the Sun: the fact that any individual particle is actually "spread out" *over a finite length scale* is precisely the property which allows for the possibility of "bridging the Coulomb gap." When classical physics has reached its limit, and two particles can come no closer than the Coulomb gap, we appear to be faced with two "point particles" separated by r_c. But now quantum mechanics steps in and replaces each particle by a structure which is no longer point-like: instead, each "particle" has a finite size, of order λ_p. When the two protons approach each other to a critical separation of $2\lambda_p$, the wave of one proton extends far enough to "touch" the wave of the other proton. Since the reduced mass of two protons is $0.5m_p$, the critical separation equals the de Broglie wavelength $\lambda_p(Am_p)$ for a single particle with a mass equal to the reduced mass, Am_p.

We now have two key length-scales in the problem: $\lambda_p(Am_p)$ and $r_c(Am_p)$. The two scales depend on different physical constants, and on different powers of the particle speed. As regards numerical value, there is no *a priori* reason why, in any particular environment, they might not differ from each other by orders of magnitude: the ratio r_c/λ_p in general might be much greater than unity or much less than unity.

But let us consider a particular location where physical parameters have the values necessary to make λ_p comparable to r_c. What happens then? Each particle "spreads out" and in effect, the particles "touch" each other across the Coulomb gap, i.e., they in effect come so close together that the distance between them is essentially zero. In particular, the two particles effectively approach each other within r_N, the range of the nuclear force. This sets the stage for nuclear reactions to occur.

The conclusion is that quantum effects allow the "Coulomb gap" to be "bridged" if r_c becomes small enough to be comparable to λ_p. Since λ_p and r_c both depend on the particle speed V (although to different powers), the "bridging" condition reduces to a condition on V. For collisions between two protons, the critical speed V_c is the speed for which $2e^2/Am_pV^2 \approx h/Am_pV$, i.e.,

$$V_c \approx \frac{2e^2}{h} \approx 697 \text{ km sec}^{-1} \tag{11.6}$$

It is noteworthy that the critical speed which allows for "bridging the Coulomb gap" between two nuclei is determined by two of the fundamental constants of nature.

Even more interesting is the numerical value of the critical speed. Inserting constants into Equation 11.6, we find $V_c \approx 690 - 700$ km sec^{-1}. This is a very significant number for the Sun.

11.4.4 Center of the Sun: thermal protons bridge the Coulomb gap

We note that the critical speed V_c is close to V_{th}, the rms speed of *protons* at the center of the Sun ($\approx$ 620 km sec^{-1}). Specifically, the ratio $r_c/\lambda_p = V_c/V_{th}$ in the core of the Sun has a numerical value of about 1.1, i.e., close to unity.

In any gaseous object which is in hydrostatic equilibrium, gravitational effects ensure that the central temperature is such that the rms speed of the dominant constituent in the core is comparable to the escape speed from the surface of the object. In order for the object to further qualify for the special title of "star," this rms speed in the core must be large enough *to allow the Coulomb gap to be bridged by quantum effects* (Mullan, 2006). Once this condition is satisfied, nuclear reactions *between thermal protons* can occur in the core. The Sun satisfies this condition. Therefore, the Sun can have access to nuclear reactions, and the energy which emerges from such reactions. *It is this which makes the Sun a star.*

In a thermal population, the particle speeds are distributed over a finite range of values. Thus, not all protons in the core of the Sun have the same speed. However, the possibility that thermo-nuclear reactions will set in is quite sensitive to the proton speed. On the one hand, if the proton speed is

a factor of (say) two *less* than V_c, then the Coulomb gap $r_c \sim 1/V^2$ opens up to a value which is four times *wider* than estimated above. At the same time, the wavelength $\lambda_p \sim 1/V$ increases by a factor of only two. Thus, the Coulomb gap is now *too wide to be bridged by the de Broglie wave*. On the other hand, if the proton speed is two times *larger* than V_c, then the wavelength λ_p decreases by a factor of two, but the Coulomb gap is now four times *smaller*. Therefore, the gap can still be bridged. This indicates that once the temperature reaches a value that is high enough to ensure that the rms speed is of order V_c, nuclear reactions will occur. But if the temperature is too small to allow the rms speed to equal V_c, then nuclear reactions will not occur.

11.4.5 Other stars: bridging the Coulomb gap

In a global sense, the Sun's mass M and radius R have values which have the effect that the crushing effects of gravity [as measured by $V_{\rm esc} \approx \sqrt{(2GM/R)}$] provide enough "thermo" at the center of the Sun to create a certain temperature. At that temperature, thermal protons have mean speeds V_{th} of order $V_{\rm esc}$. When conditions are such that V_{th} is comparable to V_c, then quantum mechanics bridges the Coulomb gap between two protons, and pp-nuclear reactions can set in.

Since V_c is determined by physical constants only, *any star* which has the same M/R ratio as the Sun will also have pp-reactions in its core. Now, astronomers discovered in the 1920s that if the stars we see in the night sky are plotted in a diagram of luminosity *versus* effective temperature, 90% of the stars lie close to a band known as the "main sequence." After decades of study, astronomers also determined masses M and radii R for many of the stars. A striking result emerged from these data: although the masses and radii vary by factors of 100–1000 along the main sequence, the *ratio* M/R is almost constant from one end of the main sequence to the other. This means that the main sequence is occupied by objects (stars) in which the mean thermal velocity in the core $V_{th} \approx V_{\rm esc} \sim \sqrt{M/R}$ is essentially equal to V_{th} in the Sun. But the latter is , as we have seen, close to V_c: therefore, along the main sequence, all stars have $V_{th} \approx V_c$. In such objects, pp-reactions can occur in the core. Therefore, *the main sequence is the locus of stars which are fusing hydrogen in their core.*

11.4.6 Inside the nuclear radius

Once two particles approach each other closer than r_N, nuclear reactions become possible. If the strong force between nucleons is at work, then nuclear reactions occur on a short time-scale. For example, in the Sun, once a deuterium nucleus is formed by reaction (a) in the pp-cycle (see Section 11.1 above), an ambient proton will interact with the deuteron *via* reaction (b): the latter reaction involves the strong force, and it occurs within time-scales of a few seconds.

11.5 Rates of Thermo-Nuclear Reactions: Two Contributing Factors

The overall rate of any particular thermonuclear reaction in thermal plasma depends on two factors. One has to do with bridging the Coulomb gap: this factor is sensitive to the temperature. The second has to do with the operation of forces within the nuclear radius: this is essentially independent of temperature.

11.5.1 Bridging the Coulomb gap: quantum tunneling

We have described the process of bridging the Coulomb gap in terms of the comparative equality of the two lengths r_c and λ_p. More formally, quantum mechanics treats the process in terms of tunneling through a potential barrier.

To quantify the tunneling, we first note that in quantum mechanics, the dynamics of particles are described by an equation called the Schrodinger wave equation. According to this equation, a free particle has a propagating wave-like character which is described, in 1-D motion, by a sinusoidal relation in space and time, i.e., an exponential with an imaginary argument:

$$\psi(x,t) \sim \exp 2\pi i \left(\frac{x}{\lambda_p} - ft\right) \tag{11.7}$$

In Equation 11.6, the spatial wavelength is λ_p, and $f = E/h$ is the frequency associated with a particle with energy E. When such a wave encounters a vertical wall (or "mountain") which is too high for a particle of energy E to surmount, the sinusoidal solution of the Schrodinger equation is replaced by a damped (nonpropagating) exponential:

$$\psi(x) \sim \exp\left[-2\pi\left(\frac{x}{\lambda_p}\right)\right] \tag{11.8}$$

In the Sun, *the very heart of energy generation depends on applying Equation* 11.8 *to the "mountain" of the Coulomb gap*, i.e., to the (huge) obstacle which prevents two thermal protons from approaching each other any closer than r_c.

According to quantum mechanics, the probability $P(V)$ that a particle with speed V (and associated de Broglie wavelength λ_p) can penetrate a 1-D barrier of width r_c is proportional to

$$P(V) \approx |\psi(r_c)|^2 \approx \exp\left[-4\pi\left(\frac{r_c}{\lambda_p}\right)\right] \tag{11.9}$$

The fact that $|\psi(r_c)|^2$ is nonzero as long as r_c/λ_p is finite, means that, in the quantum world, there is a *finite chance* that a particle can penetrate through a wall (or pass over a "mountain") which would be completely

insurmountable in the classical world. This process is known as "quantum tunneling."

When the tunneling calculation is done rigorously, in 3-D and in the presence of a "mountain" which has the particular shape of the Coulomb barrier, it is found that the numerical coefficient $4\pi(=12.6)$ in Equation 11.8 must be replaced by the somewhat larger number $2\pi^2(=19.7)$. That is, $P(V) \approx \exp(-2\pi^2 r_c/\lambda_p)$. Inserting the expressions given above for r_c and λ_p, we find that the probability $P(V)$ for Coulomb barrier penetration is given by

$$P_G(V) = \exp\left[\frac{-4\pi^2 Z_1 Z_2 e^2}{hV}\right] \tag{11.10}$$

This expression for the probability is known as the Gamow factor, in honor of the physicist who first performed the tunneling integral (Gamow, 1928). In recognition of Gamow's role, we use subscript G in Equation 11.10.

What is the numerical value of the tunneling probability in the core of the Sun? We have seen (Section 11.4.4) that in the core, $r_c/\lambda_p \approx 1.1$. In that case, Equation 11.10 tells us that $P_G(V) \approx \exp(-2.2\pi^2) \approx 4 \times 10^{-10}$.

Note that $P_G(V)$ is quite sensitive to the particle speed, e.g., suppose that, instead of considering particles moving with speed V_{th}, we were to consider the collisions of two particles each of which moves with speed $2V_{th}$. In such a case, the tunneling probability $P_G(2V_{th}, \text{Sun})$ would be $\approx \exp(-1.1\pi^2) = 2 \times 10^{-5}$. Thus, by doubling the speed, we have increased the pp-tunneling probability by a large amount (5×10^4). At first sight, this suggests that we may have made an error of many orders of magnitude by evaluating the tunneling probability at the particular speed V_{th}. But upon further consideration, we can see that the error is much less serious.

To see why this is so, we note that in a thermal velocity distribution, where $f(V) \sim V^2 \exp(-V^2/V_{\text{th}}^2)$, there are *fewer* particles moving at faster speeds. For example, for every particle which moves with speed V_{th}, there are only $4e^{-4} \approx 0.07$ particles in a Maxwellian distribution moving with $2V_{th}$. For this reason alone, the number of possible interactions which might occur every second between particles each of which moves with speed $2V_{th}$ is smaller by $0.07^2 \approx 1/200$ than the collision rate between two particles moving with speed V_{th}. Furthermore, the cross-section for Coulomb collisions is smaller for faster particles: $\sigma_c \sim 1/V^4$ (see Section 11.3). This further reduces the collision rate by a factor of 16 when we compare particles with speed $2V_{th}$, to particles with speed V_{th}. Combining the Coulomb and Maxwellian factors, we see that the increase in pp-tunneling probability by 5×10^4 is offset by $200 \times 16 \approx 3 \times 10^3$. Therefore, as far as the actual rate of tunneling, particles with speed $2V_{th}$, are indeed more effective than particles with speed V_{th}, but not by many orders of magnitude. The increase in effectiveness is a factor of ~ 17.

If we were to repeat this exercise for particles moving even faster, say $4V_{th}$, we would find that the increase in tunneling probability (by a factor of $\approx 10^7$) is more than offset by the combined Maxwellian and Coulomb factors. The relative number of Maxwellian particles is $16\text{e}^{-16} \approx 2 \times 10^{-6}$, and

Coulomb collisions occur 256 times less frequently. Thus, despite the increased Gamow factor, particles with speed $4V_{th}$ are about ten times *less* effective than particles with speed V_{th}. Overall, the peak in pp-tunneling probability in a thermal distribution of protons occurs for particles with speeds of $2 - 3V_{th}$, and closer to $2V_{th}$ than to $3V_{th}$.

This suggests that our estimate of tunneling probability obtained above for particles moving with speed $V_{th} (\approx 4 \times 10^{-10})$ is a lower limit on the actual probability in the Sun. The lower limit should be increased by a factor of perhaps 20 in order to obtain a more realistic pp-tunneling probability for a Maxwellian distribution in the Sun: $P_G(\text{Sun}) \approx 8 \times 10^{-9}$.

We can now see an important conclusion of this discussion. Even in the "favorable" conditions which exists in the core of the Sun, only one collision in (roughly) 125 million results in one proton tunneling close enough to another to "feel" the nuclear force. On the other hand, as we have seen (Section 11.3), each proton in the core undergoes some 2×10^{15} collisions per second. Therefore, each proton in the Sun's core experiences roughly 10^7 tunneling events every second. When combined with the relevant post-tunneling processes (see Section 11.5.2), this suffices to provide the Sun with its mighty output power.

We shall return below to examine how the functional form of $P_G(V)$ has the effect that the rates of thermonuclear reactions increase rapidly with increasing temperature. But for now, we turn to what happens inside the nucleus once the tunneling has occurred.

11.5.2 Post-tunneling processes

Once tunneling has occurred, the two particles are close enough together that they can be regarded as being together inside a nucleus. The processes which then occur depend on which forces come into play.

We have already mentioned (Section 11.4.6) that the strong force is at work in reaction (b) of the pp-cycle. The strong force is also at work in reaction (c) of the pp-cycle. However, even though reaction (b) occurs on a time-scale of a few seconds in the Sun, reaction (c) requires on average several million years to occur. The principal reason that reaction (c) is so much slower than reaction (b) in the Sun has to do with the tunneling factor: referring to Equation 11.10, we see that the product $Z_1 Z_2$ is four times larger for reaction (c) than for (b). (We will return to this in Section 11.7.)

But reaction (a) in the pp-cycle is different. The strong force is *not* the predominant factor which controls this reaction. When two protons interact via the strong force, they in effect attempt to form a nucleus consisting of two protons and nothing else. Such a nucleus would be a "di-proton." However, calculations of nuclear structure indicate that such a nucleus is not stable: the combination of kinetic energy, Coulomb repulsion, and exchange forces overwhelms the attractive nuclear energy. As a result the di-proton is unbound. The strong force is simply not strong enough to bind the two protons in reaction (a) together.

So how does reaction (a) proceed? We note that the product of the reaction, the deuteron, *is* a stable (bound) nucleus consisting of one proton and one neutron. To form such a nucleus, one of the protons which enters into reaction (a) must become a neutron. During the (very) brief interval of "collision-time" when the two protons are within a distance of r_N of each other, one of the protons must become transformed into a neutron.

How long does the "collision time" last? The duration of a collision is $t_c \approx r_N/V$ where $V \approx 6 \times 10^7$ cm sec^{-1} is the mean thermal speed of protons (and therefore neutrons) in the core of the Sun. Setting $r_N \approx 10^{-13}$ cm, we find $t_c \approx 2 \times 10^{-21}$ sec.

What is the chance that a proton-to-neutron transformation will happen during an interval of duration t_c? If we were considering a free proton in the Sun, the answer would be straightforward: the chance would be zero. It is impossible for a free proton in the Sun to decay into a neutron because the proton would have to *gain* a mass of 0.0014 a.m.u. (see Table 11.1). This is equivalent to an energy gain of 1.29 MeV, about 1000 times larger than the thermal energies in the Sun. However, inside a nucleus, in the presence of the strong force, with an attractive energy of 30–40 MeV, the transformation of a proton into a neutron becomes possible: in such an environment, in a potential well some 30–40 MeV deep, the possibility of "picking up" 1.29 MeV is no longer out of the question. In fact, the transformation (or "decay") of a proton into a neutron inside the nucleus occurs because of the operation of the weak force.

This requires that the weak force must work its transforming effects precisely during the "collision-time." Now, a measure of the strength of the weak force is provided by the empirical result that free neutrons decay with a half-life $t_{1/2}$ of about 650 sec.

What is the probability $P_d(p)$ that a proton will decay into a neutron during the "collision-time"? The correct answer to this question requires a theory of beta-decay: Fermi's theory was used by Bethe and Critchfield (1938) in their calculation of the rate of the pp-cycle in the Sun.

Without going into the details of beta-decay theory, we can estimate an upper limit to the probability by considering a hypothetical analog to proton-proton collisions. Suppose two free *neutrons* were available in the thermal population in the Sun's core, and suppose they were to undergo a collision in which the distance of closest approach happened to be r_N. A free neutron always has the option of decaying into a proton. So, what is the probability $P_c(n)$ that one of the neutrons would decay into a proton *during the collision time* t_c? The answer is: $P_c(n)$ can be estimated roughly by the ratio of t_c to the neutron half-life, $t_{1/2}$. This leads to $t_c/t_{1/2}$ of order 3×10^{-24}. Thus, the probability that a (free) neutron in the Sun's core would decay into a proton during the collision is $P_c(n) \approx 3 \times 10^{-24}$. Even with the advantage of free neutron decay, this is still a very small probability.

Returning now to the case of proton-proton collisions, we recall that the proton and the neutron are both nucleons with similar properties. (In the technical language of nuclear physics, protons and neutrons are members of

the same "isospin doublet.") As a result, they are expected to behave to a certain extent in similar ways when they are within a distance of r_N of each other. However, there is a difference in the energy ΔE which is released in the reaction: whereas reaction (a) above releases an energy of 1.44 MeV, the excess mass energy of 1.29 MeV of the neutron relative to the proton would have the effect that the energy released in the reaction $\mathrm{n}+\mathrm{n} \rightarrow \mathrm{D}$ would equal $1.44 + 1.29 = 2.73$ MeV. Now according to a general rule in particle decays (known as the "Sargent rule"), the rate of beta decay scales as $(\Delta E)^{\alpha}$, where $\alpha = 5$ in the limit that the decay products are relativistic. As a result, the reaction $\mathrm{p} + \mathrm{p} \rightarrow \mathrm{D}$ is predicted to be less probable than $\mathrm{n} + \mathrm{n} \rightarrow \mathrm{D}$ by a factor of order $(2.73/1.44)^5 \approx 24$ in the relativistic limit. Even in the non-relativistic limit, the probability $P_c(p)$ that a proton will decay into a neutron during the collision is expected to be smaller than $P_c(n)$. Defining the ratio of $P_c(n)/P_c(p)$ as $\xi(>1)$, we write $P_c(p) \approx (3/\xi) \times 10^{-24}$.

11.5.3 Probability of pp-cycle in the solar core: reactions (a) and (b)

Combining the probability factors for quantum tunneling and for the post-tunneling process of proton transformation, we see that in the center of the Sun, the overall probability $P(pp)$ of a pp-nuclear reaction (i.e., reaction (a) in the pp-cycle) in a collision in the solar core is given by the product of $P_G(\mathrm{Sun})$ and $P_c(p)$. Using the estimates given above, we find $P(pp) \approx (24/\xi) \times 10^{-33}$.

We recall that the observed properties of the Sun indicate that a pp-cycle occurs on average only once in every $N_c(react) \approx 10^{33}$ collisions in the Sun's nuclear-burning core (Equation 11.2). That is, the empirical probability of a nuclear reaction is of order 10^{-33} per collision. Compared with our estimates of $P(pp)$, we see that we can replicate the empirical probability of nuclear reaction in the Sun as long as $P_c(n)$ does not exceed $P_c(p)$ by more than ≈24. This is precisely what is available based on the Sargent rule.

Thus, of the 33 orders of magnitude which occur in the empirical reaction probability $1/N_c$ (react), the process of tunneling through the Coulomb barrier provides about eight orders of magnitude, while the weak interaction which occurs in the post-tunneling process contributes the remaining 25 orders of magnitude. The weak interaction truly dominates (by ≈17 orders of magnitude) in regulating the slowness of the thermonuclear processes in the Sun.

It is the low value of the probability associated with the weak interaction that causes reaction (a) of the pp-cycle to be so much slower than reactions (b) or (c). We recall (Section 11.3) that on average, a proton participates in reaction (a) once in 15–30 billion years. In reaction (b), since the Coulomb barrier is similar to that in reaction (a), the tunneling probability is comparable to that for reaction (a). However, the post-tunneling process in reaction (b) involves the interaction between two nuclei so as to form a third stable nucleus. The interaction in reaction (b) therefore operates by way of the *strong* force, in sharp contrast to reaction (a), where the *weak* force is at work. In the

nature of things, we expect that the strong force operates on much shorter time scales than the weak force. In fact, measurements of the cross-section for reaction (b) indicate that the post-tunneling process in (b) operates almost 18 orders of magnitude more rapidly than in reaction (a). As a result, instead of a time-scale of almost 10^{18} sec between occurrences of reaction (a), reaction (b) occurs on time-scales of order seconds. We shall return to discuss the time-scale for reaction (c) in Section 11.7 below, after we quantify how the tunneling probability depends on charge and mass.

11.6 Temperature Dependence of Thermonuclear Reaction Rates

A significant characteristic of the Gamow tunneling probability $P_G(v)$ (Equation 11.8) is the occurrence of the particle speed in the *denominator* of the exponential argument. This has the effect that $P_G(V)$ falls off exponentially rapidly to zero as the speed decreases below a value that is related to V_c (see Equation 11.6). In the opposite limit, for speeds $V \gg V_c$, $P_G(V)$ eventually saturates at a value of unity.

In contrast to this behavior of the tunneling factor, the significant property of a thermal velocity distribution $f(V) \sim V^2 \exp(-V^2/V_{\rm th}^2)$ is that as the speed increases, the exponential term eventually dominates over the V^2 term. As a result, the number of available particles falls off exponentially rapidly at high speed.

The overall rate of thermonuclear reactions involves an integral of the product $\Pi(V) = P_G(V)f(V)$ over all velocities. Because of the contrasting behavior of each of the terms as a function of velocity, the integral receives essentially zero contribution from particles with low speeds or from particles with high speeds. The integrand peaks at an intermediate velocity V_o, corresponding to energy E_o. The particles which contribute most to the rate of thermonuclear reactions are those which lie within a range of velocities ΔV in the neighborhood of V_o. As a result, when we integrate over all velocities, the thermonuclear reaction rate r_{tn} is proportional to $f(V_o)$ (the number of particles in the thermal distribution at $V = V_o$) times ΔV.

Converting from velocity to energy, we note that the exponential term in $f(V)$ converts to $f(E) \sim \exp(-E/kT)$, while $P_G(V)$ converts to $P_G(E) = \exp(-\beta/\sqrt{E})$. In the expression for $P_G(E)$, $\beta = C' Z_1 Z_2 \sqrt{A}$ and $C' = 2\pi^2 e^2 \sqrt{(2m_p)}/h = 1.23 \times 10^{-3}$ c.g.s. units. Since the mean thermal energy kT ($\approx 1.9 \times 10^{-9}$ ergs) in the core of the Sun is of order 1 keV ($= 1.6 \times 10^{-9}$ ergs), it is convenient (Clayton, 1968) to express energy in units of keV: $E_k = E/(1\ \text{keV})$. In these units, C' is replaced by $C'_k = 31\ \text{keV}^{0.5}$.

In terms of energy, the product $\Pi(E) = P_G(E)f(E)$ has a maximum value at an energy E_o where the sum of the two terms $\beta/\sqrt{E} + E/kT$ in the exponent

is a minimum. Taking the derivative with respect to energy, we find that this minimum occurs when

$$\frac{-\beta}{2E_o^{3/2}} + \frac{1}{kT} = 0 \tag{11.11}$$

This leads to $E_o = (\beta kT/2)^{2/3}$: this is the energy at which the particles in the thermal distribution participate with maximum effectiveness in quantum tunneling, and therefore, also in thermonuclear reactions. For example, in the case of reaction (a) in the pp-cycle in the core of the Sun, we have $Z_1 = Z_2 = 1$ and $A = 0.5$. These lead to $\beta = 22$ keV$^{0.5}$. Since $kT \approx 1.2$ keV in the core of the Sun, we find $E_o \approx 5.6$ keV, i.e., ≈ 4.7 times larger than the mean thermal energy. The velocity of particles with energy E_o is therefore $\approx \sqrt{4.7}$ times the mean thermal speed, i.e., $\approx 2.2 V_{th}$. This confirms our discussion in Section 11.5.1.

Using the estimate of E_o, we find that the rate of thermonuclear reactions f_r is proportional to $\Pi(E_o)$, i.e., $f_r \sim \exp(-3E_o/kT)$. Because of the exponential factor, the rate f_r is quite sensitive to temperature. To quantify this, let us insert the expression derived above for E_o, and take the natural logarithm. We find $\ln(f_r) = -3(\beta/2)^{2/3}/(kT)^{1/3}$. It is often convenient to write the reaction rate in terms of a power law of the temperature, $f_r \sim T^{\delta}$. This leads to

$$\delta \equiv \frac{d\ln f_r}{d\ln T} = +\left(\frac{\beta^2}{4k}\right)^{1/3} \frac{1}{T^{1/3}} \tag{11.12}$$

Inserting c.g.s. values for β (for the pp-reaction) and k, and expressing the temperature in units of 10^6 K (i.e., $T_6 \equiv T/10^6$ K) we find

$$\delta = +\frac{11.1}{T_6^{1/3}} \tag{11.13}$$

In the core of the Sun, where $T_6 \approx 15 - 16$, Equation 11.13 indicates that $\delta \approx 4 - 5$. Thus, the rate of pp-reaction increases rather rapidly as temperature increases.

11.7 Rate of Reaction (c) in the pp-cycle

Reaction (c) (Section 11.1) involves a larger Coulomb barrier than reactions (a) or (b). It is interesting to see quantitatively how sensitive the tunneling barrier is to the reacting nuclei.

In calculating the quantity β for reaction (c), using $Z_1 = Z_2 = 2$ and $A = 1.5$, we find $\beta = 152$ keV$^{0.5}$. Setting $kT = 1.2$ keV, this leads to $E_o =$

20.3 keV, which is much larger than the 5.6 keV value for reaction (a). As a result, the reaction rate, which is proportional to $\exp(-3E_o/kT)$, is reduced in reaction (c) compared to reaction (a) by $\exp(-3[20.3 - 5.6]/1.2) \approx 10^{-16}$ in the core of the Sun. (Note, $kT = 1.2$ keV in the solar core.) However, the post-tunneling process in reaction (c) depends on the strong force: the rate of this process therefore greatly exceeds that for reaction (a). Empirically, the excess in rates is found to be of order 10^{25} (Clayton, 1968, p. 380).

Combining the factors 10^{-16} and 10^{25}, we see that each He^3 nucleus reacts at a rate which is 10^9 times more frequent than each proton. However, the equilibrium abundance of He^3 nuclei is much smaller than the proton abundance: specifically, in equilibrium, for every proton, there are only 10^{-5} He^3 nuclei. This has the effect that the mean free time interval that a particular He^3 nucleus must wait between collisions with another He^3 nucleus is 10^5 times longer than the mean free time for collisions with a proton.

The combination of the enhancement factor of 10^9 (due to tunneling plus post-tunneling processes) and the decrease of 10^5 (due to abundances) has the net effect that an individual He^3nucleus has a collision leading to reaction (c) in a time-scale which is some 10^4 times shorter than the time-scale for a proton to undergo reaction (a). As a result, whereas a time-scale of order 10^{10} years is characteristic of reaction (a), the time-scale for reaction (c) is of order 10^6 years.

For reactions other than reaction (a), the numerical coefficient 11.1 in the expression for δ must be replaced by $11.1(Z_1Z_2)^{2/3}(A/0.5)^{1/3}$, where the 0.5 refers to the reduced mass which enters into reaction (a). This has the effect that reaction (c) in the core of the Sun has a rate which increases as T^{16}. The great sensitivity to temperature arises from the sensitivity of tunneling to the strength of the Coulomb barrier.

Exercises

11.1 From Exercise 5 in Chapter 1, you already know the values of $V_{\rm esc}$ for main sequence stars with masses of $0.1, 0.3, 1, 3$, and $10\ M_\odot$. Assuming thermal speeds in the core $V_{th} \approx V_{\rm esc}$, evaluate the ratio $r_c/\lambda_p = V_c/V_{th}$ in the core of each star (where V_c is given by Equation 11.6). Show that on the main sequence, the ratio r_c/λ_p does not vary by more than a factor of roughly 2.

11.2 Using the tunneling probability formula $P(V) \approx \exp(-2\pi^2 r_c/\lambda_p)$, calculate $P(V)$ for the five stars in Exercise 1. Show that $P(V)$ for the $10M_\odot$ star is two to three orders of magnitude larger than for the $1M_\odot$ star, while $P(V)$ for the $1M_\odot$ star is three to four orders of magnitude larger than for the $0.1M_\odot$ star. Show how these results help us to ex-

plain the empirical results that the luminosity of a $10M_\odot$ star exceeds $L_\odot$ by about 1000, while $L_\odot$ exceeds the luminosity of a $0.1M_\odot$ star by about 1000.

References

Audi, G. and Wapstra, A. H. 1993. "The 1993 atomic mass evaluation: (I) Atomic mass table," *Nuclear Phys., A*, 565, 22.

Bahcall, J. N., Serenelli, A. N., and Basu, S. 2005. "New solar opacities, abundances, helioseismology, and neutrino fluxes," *Astrophys. J. Lett.*, 621, L85.

Bethe, H. A. 1939. "Energy production in stars," *Phys. Rev.*, 55, 434.

Bethe, H. A. and Critchfield, C. L. 1938. "The formation of deuterons by proton combination," *Phys. Rev.*, 54, 248.

Clayton, D. D. 1968. "Thermonuclear reaction rates," *Principles of Stellar Evolution and Nucleosynthesis.* McGraw-Hill, New York, pp. 283–361.

De Broglie, L. 1924. *Recherches sur la théorie des quanta*. PhD Thesis, Sorbonne University, Paris.

Gamow, G. 1928. "Zur quantentheorie des atomkernes," *Zeitschrift fur Physik*, 51, 204.

Mullan, D. J. 2006. "Why is the Sun so large?," *Amer. J. Phys.*, 74, 10.

Chapter 12

Neutrinos from the Sun

As a result of the calculations in Chapters 5, 7, and 9, we have obtained a model for the interior of the Sun. Is there any way to check our model, to see if it is correct?

In the 1960s, only one answer existed for this question: we need to detect a certain kind of energetic particle (neutrinos) which emerges from nuclear reactions in the core of the Sun. The goal of such experiments would be to check that the numbers of neutrinos which reach the Earth, as well as their energies, are consistent with the properties we calculated for nuclear reactions in the solar core.

The existence of neutrinos was first postulated by Pauli in 1930 in order to preserve the laws of conservation of momentum and energy in certain radioactive decays. Pauli's approach was a bold one: he had to postulate that in these decays, a hitherto unseen particle with zero electric charge must emerge with a finite energy and momentum, but with a mass that must be so small as to be almost zero, certainly much smaller than the mass of an electron. The term "neutrino" was subsequently coined by Fermi for Pauli's unseen particle. The absence of electric charge means that the neutrino does *not* interact with its surroundings by means of electromagnetic processes. The fact that the neutrino is associated with radioactive decay (a process which is driven by the weak interaction) means that the neutrino interacts with other particles via the weak force.

We have already seen (Chapter 11, Section 11.5.3) how the weak force in the Sun makes for very long time-scales in certain reactions, whereas the strong force makes reactions occur much more rapidly. There is also a large difference in strength between the weak force and the electromagnetic force. Because of this difference, photons (signatures of the electromagnetic force) and neutrinos (signatures of the weak force) behave very differently as they propagate inside the Sun. In Chapter 9, Section 9.3, we determined that photons which originate in the core of the Sun take millions of years to escape from the Sun. In contrast to the photons, we shall see that neutrinos from nuclear reactions in the solar core can reach the surface of the Sun in a matter of a few seconds.

12.1 Generation and Propagation of Solar Neutrinos

Every time the pp cycle occurs, a neutrino emerges from the first step of the cycle (Chapter 11, Section 11.1, reaction (a)). A complete pp-cycle requires this step to occur twice. As a result, since the pp-cycle occurs some 10^{38} times per second in the Sun (see Chapter 11, Section 11.2), there are some 2×10^{38} neutrinos generated per second in the Sun's core.

Are the neutrinos likely to be absorbed as they pass through the Sun? Or can they escape more or less freely? To answer this, we return to the same sort of calculation we did in Chapter 9, Section 9.3 when we were considering how *photons* propagate inside the Sun. The relevant physical quantity is the mean free path λ_m that a neutrino can travel between collisions inside the Sun.

In general, when a projectile moves through a medium containing n "target objects" per cc, each with a cross-section of σ, the mean free path is given by $\lambda_m = 1/n\sigma$. In the case of photons, where the opacity κ is conventionally expressed in units of $\mathrm{cm}^2\ \mathrm{gm}^{-1}$, the product $n\sigma$ can be replaced by the product $\kappa\rho$ (Chapter 3, Section 3.3). As a result, as we have already seen (Chapter 9, Section 9.3), the mean free path of a photon $1/\kappa\rho$ is very short (≈ 0.001 cm) in the Sun's core.

Turning now to neutrinos, we revert to the general formula $\lambda_m = 1/n\sigma$. The "target objects" that a neutrino from the core of the Sun encounters on its way to the surface are mainly protons. In the core of the Sun, where the mass density ρ is ≈ 140 gm cm^{-3} (see Chapter 9, Table 9.1), the number density of nuclei is roughly ρ/m_{H} (where $m_{\mathrm{H}} = 1.67 \times 10^{-24}$ gm). Thus, $n \approx 10^{26}\mathrm{cm}^{-3}$, mainly hydrogens, but including He and a few "metal" ions.

Now we come to the key difference between photons and neutrinos. Photons which try to propagate past the ions in the core of the Sun "see" the ions as having, on average, effective areas of order $10^{-23}\mathrm{cm}^2$. The reason that the cross-section has a value larger than the Thomson cross-section (see Chapter 3, Equation 3.1) is that the photon interacts via electromagnetism with some bound electrons which remain in heavy nuclei. But for neutrinos, electromagnetism is not important: neutrinos interact with the nuclei in the Sun by means of the weak interaction. For this, the cross-section is *much* smaller than the Thomson value. Cowan et al. (1956) detected neutrinos from a fission reactor by measuring their interaction with a specially designed detector. Depending on the power level of the reactor, Cowan et al. found a mean neutrino reaction rate in the detector between 0.6 and 2.9 per hour. Running for almost 1400 hours, Cowan et al. determined that for the neutrinos associated with a fission reactor (with energies of a few MeV, comparable to the energies of solar neutrinos: see Figure 12.1), the scattering cross-section $\sigma \approx 6 \times 10^{-44}\mathrm{cm}^2$. In subsequent more refined experiments, the cross-section was found to be larger by a factor of about 2. To a good approximation, we may take $\sigma \approx 10^{-43}\mathrm{cm}^2$ for neutrinos with energies of a few MeV.

The contrast between the photon cross-section and the neutrino cross-section is striking: the difference amounts to some 20 orders of magnitude.

"Weak" is indeed an appropriate adjective to describe the interaction that neutrinos have with matter.

For neutrinos with energies of a few MeV, the cross-section is insensitive to energy, and so we can, without serious error, apply the cross-section determined from fission reactor neutrinos to the conditions in the core of the Sun. Combining the neutrino σ with the value of n in the core of the Sun, we see that even in the densest region of the Sun, the neutrino mean free path $\lambda_m = 1/n\sigma \approx 10^{17}$ cm. In terms of a unit of length that is more familiar to astronomers, this equals one-tenth of a light-year! As the neutrinos move outward from the core of the Sun, and pass through gas of lower density, the value of λ_m becomes even larger. Even at the center of the Sun, the value of λ_m exceeds the solar radius (7×10^{10} cm) by more than six orders of magnitude.

As a result, the neutrinos from the pp reaction in the core of the Sun barely "feel" the material of the solar interior at all. Less than one neutrino in a million will undergo a scattering between the core of the Sun and its surface. For the rest, the Sun is essentially "transparent," and the neutrinos simply stream freely out of the Sun. With essentially zero rest mass, a neutrino travels at the speed of light: once a neutrino is generated in the core, it reaches the surface in a time of $R_\odot/c = 2.3$ sec. Some 500 sec later, the neutrino passes the Earth's orbit.

The lack of scattering in physical space does not mean that the neutrinos feel *no* effects whatsoever from passing through the Sun: another type of effect occurs, one which causes the neutrino to change into another type of neutrino. We will return to this below, after we describe the experiments which have been built to detect solar neutrinos.

12.2 Fluxes of Solar Neutrinos at the Earth's Orbit

The Sun generates $N_n \approx 2 \times 10^{38}$ pp-neutrinos per second. When the neutrinos pass by the Earth, at a distance of $D = 1$ AU, the flux is $F_n = N_n/4\pi D^2 \approx 6 \times 10^{10}$ cm^{-2} sec^{-1}.

The neutrinos which emerge from the pp reaction have a range of energies, all less than 0.42 MeV. Other channels of the pp-chain, as well as contributions from the CNO cycle, ensure that the Sun generates other neutrinos with a range of energies. The spectrum of solar neutrinos, calculated from a detailed model of the Sun, and evaluated at the Earth's orbit, is shown in Figure 12.1. Neutrinos which emerge from reactions involving only two outgoing particles are emitted at unique energies: these appear as "lines" in the spectrum. Reactions in which more than two outgoing particles are present (including pp $\rightarrow$ De$^+\nu$) give rise to a "continuum" of energies for the neutrinos, up to a well-defined maximum "cut-off" energy, which is determined by the difference in energy between initial and final state.

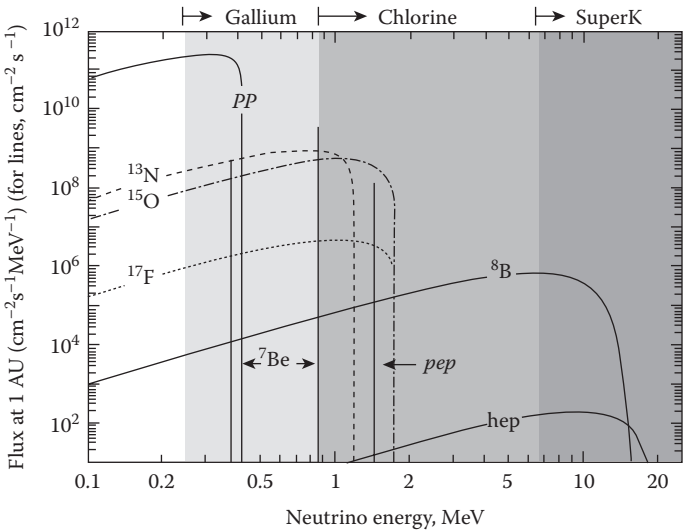


FIGURE 12.1: Fluxes of solar neutrinos as a function of energy at a distance of 1 AU from the Sun. Different ordinates are used depending on whether one is dealing with "lines" or with "continua." (From Nakamura K. 2000. *Euro. Phys. J. C,* 15, 366. With permission.)

For the "lines" in the solar neutrino spectrum, the ordinate in Figure 12.1 refers to the flux in units of particles $\mathrm{cm}^{-2}\,\mathrm{sec}^{-1}$ at Earth. For the "continua," the ordinate in Figure 12.1 refers to a differential energy flux, in units of particles $\mathrm{cm}^{-2}\ \mathrm{sec}^{-1} MeV^{-1}$ at Earth.

In terms of overall flux, the neutrinos from the Sun are predominantly those which emerge from reaction (a) of the pp-I chain. As Figure 12.1 illustrates, the maximum differential flux of pp neutrinos is $(2-3)\times 10^{11}\mathrm{cm}^{-2}\ \mathrm{sec}^{-1}\mathrm{MeV}^{-1}$, while the cut-off energy is 0.42 MeV. Thus, the area under the curve, i.e., the total flux of pp neutrinos at Earth orbit is no more than $(8-12)\times 10^{10}$ neutrinos $\mathrm{cm}^{-2}\mathrm{sec}^{-1}$. This is consistent with the rough estimate of 6×10^{10} neutrinos $\mathrm{cm}^{-2}\ \mathrm{sec}^{-1}$ given above.

12.3 Neutrinos from Reactions other than pp-I

In the Sun, most of the energy generation occurs via the pp-I chain of reactions which were discussed in Chapter 11. Now, in order to provide a more complete discussion of the neutrinos which come from the Sun, we need to look at certain less frequent reactions which also occur in the solar core, in particular the pp-II and pp-III chains. These reactions do not contribute much to the

energy output of the Sun, but they do contribute significantly to the neutrino fluxes which have been detected on Earth. In fact, for the first 25 years of solar neutrino experiments, the only neutrinos which could be detected were those from the pp-III chain. The reason has to do with the fact that neutrinos must have a certain minimum energy before a detector can respond.

12.3.1 pp-II and pp-III chains

Both of these chains at first rely on the reactions pp-I (a) and (b) (see Chapter 11, Section 11.1) to produce He^3. Then, instead of interacting with another He^3 nucleus, the following reaction occurs:

$$He^3 + He^4 \rightarrow Be^7 + \gamma \qquad \text{(d)}$$

How fast does reaction (d) go compared with reaction (c) (see Chapter 11) of the pp-I chain? To address this, we can first address the difference in Coulomb barrier tunneling by proceeding analogously to the discussion in Chapter 11, Section 11.7. For reaction (d), we use $Z_1 = Z_2 = 2$ and $A = 1.71$: this leads to $\beta = 162$ keV$^{0.5}$. Using $kT = 1.2$ keV in the solar core, we find $E_o = (\beta kT/2)^{2/3} = 21.1$ keV. This is somewhat larger than the 20.3 keV value for reaction (c) of the pp-I chain. As a result, the tunneling rate, which is proportional to $\exp(-3E_o/kT)$, is reduced in reaction (d) compared to reaction (c) by almost ten in the core of the Sun. Thus, quantum tunneling reduces the reaction rate of (d) compared to (c) by about one order of magnitude.

On the other hand, a significant factor which strongly favors the occurrence of reaction (d) over reaction (c) has to do with the fact that He^4 is much more abundant in the Sun than He^3: the excess is some four orders of magnitude. As a result, any nucleus of He^3 finds itself likely to collide, in a given time interval, with 10^4 times more He^4 nuclei than with He^3 nuclei.

The final determination of how rapidly reaction (d) occurs compared to reaction (c) has to do with what happens in the post-tunneling process, when the strong force comes into operation. There is no easy way to see what differences should be expected when the strong force comes into play: one must rely on detailed quantum mechanical calculations. These indicate (Clayton, 1968) that for (d), the reaction rate has a numerical value which is 10^4 times smaller than for reaction (c).

Combining the reduction in tunneling rate (10^{-1}) with the increase in abundance (10^4) and the decrease (10^{-4}) in the post-tunneling rate, we find that the net effect is that reaction (d) occurs about 10 times less frequently in the solar core than reaction (c). As a result, the pp-I chain occurs about 90% of the time in the solar core, while the pp-II and pp-III chains, in combination, occur about 10% of the time.

Following reaction (d), the pp-II chain proceeds according to the reactions:

$$Be^7 + e^- \rightarrow Li^7 + \nu \qquad \text{(e)}$$

$$Li^7 + p \rightarrow 2He^4 \qquad \text{(f)}$$

Reaction (e), which involves electron capture, leads to two (and only two) particles in the exit channel. As a result, the neutrinos have a unique energy. The energy difference between the ground states of the nuclei is 0.86 MeV. A neutrino "line" indeed appears in Figure 12.1 at 0.86 MeV.

As it happens, the Li^7 nucleus has an excited state at an energy of 0.48 MeV above ground: this lies low enough that it is also below the energy of the ground state of Be^7. The transition from the ground state of Be^7 to this excited state produces a neutrino with an energy of $0.86 - 0.48 = 0.38$ MeV. A neutrino "line" also appears in Figure 12.1 at this energy. Laboratory measurements indicate that the 0.86 MeV transition occurs 90% of the time. This accounts for the fact that in Figure 12.1, the flux of 0.86 MeV neutrinos is about ten times larger than the flux of 0.38 MeV neutrinos.

Following reaction (d), the pp-III chain proceeds as follows:

$$Be^7 + p \rightarrow B^8 + \gamma \tag{g}$$

$$B^8 \rightarrow Be^8 + e^+ + \nu \tag{h}$$

$$Be^8 \rightarrow 2He^4 \tag{i}$$

Reaction (g) differs from reaction (e) in the qualitative sense that in (g), the repulsive force between two positively charged particles has to be penetrated, whereas in (e), there is an attractive force between the Be^7 nucleus and the electron. For these reasons, the pp-II chain gets off to a faster start than the pp-III chain. Detailed calculations show that the pp-II chain occurs about 100 times more frequently in the Sun than the pp-III chain.

Nevertheless, reaction (h), in the pp-III chain is the reaction which first allowed solar neutrinos to be detected. Three particles emerge from the decay, and as a result, the neutrino energies are spread across a continuum. Significantly, the cut-off energy of the continuum is quite large, some 14 MeV. This large value arises from the large difference in mass between the parent and the daughter nuclei (see Audi and Wapstra 1993) 8.021864 a.m.u. (B^8), and 8.003111 a.m.u. (Be^8). If the decay in reaction (h) were to occur between the ground states of parent and daughter, the cut-off energy would be the energy corresponding to the total mass difference (0.0188 a.m.u.), i.e., 17.5 MeV. However, although the decay starts in the ground state of B^8, it is forbidden to go to the ground state of Be^8: instead, the decay goes to an excited state of Be^8 which lies 2.9 MeV above ground. The positron also requires 0.51 MeV. As a result, the cut-off energy for reaction (h) is 14.1 MeV.

12.3.2 Other reactions

The pep reaction involves an electron capture by a proton and a subsequent collision with another proton:

$$p + e^- + p \rightarrow D + \nu \tag{j}$$

The energetics are the same as reaction (a) of the pp-I chain, except that the electron appears on the left-hand side. Moreover, only two particles emerge, and as a result, the emergent neutrino has a unique energy: 1.44 MeV. Reaction (j) in the Sun occurs once for every 400 pp reactions.

The Hep reaction leads to a neutrino continuum with a cut-off at 18.8 eV:

$$\mathrm{He}^3 + \mathrm{p} \rightarrow \mathrm{He}^4 + \mathrm{e}^+ + \nu \qquad \text{(k)}$$

These are certainly the most energetic neutrinos generated by solar nuclear reactions. However, models indicate that the Hep reaction is very rare, occurring less than once for every million pp reactions. There are no claims in the literature that any experiment has confidently identified a neutrino from the Hep reaction.

In stars hotter than the Sun, energy is generated preferentially by a "bi-cycle" of reactions in which carbon nuclei act as catalysts for fusing four protons into one helium. In this "bi-cycle," three decays occur (N^{13}, O^{15}, and F^{17}) with the emission of a neutrino with a continuous energy spectrum extending up to $1 - 2$ MeV. These are shown in Figure 12.1. In the Sun, combinations of models and data suggest that the CNO cycle contributes only 0.5% to the Sun's energy output (Bahcall et al., 2005). We will not consider CNO neutrinos any further in this first course.

12.4 Detecting Solar Neutrinos on Earth

The very smallness of the interaction cross-section which allows neutrinos to escape from the center of the Sun has the inevitable corollary that detection of neutrinos on Earth requires efforts which are nothing short of Herculean.

There are two general classes of experiment for the detection of neutrinos. In one class, we rely on the properties of certain nuclei to absorb a neutrino, thereby transforming the initial nucleus into the nucleus of a new element: the goal is then to identify the amount of the new element which is produced in a given time interval. In the second class of experiments, we do not use nuclear physics at all: instead, we detect events in which a fast neutrino "smashes into" an electron in a certain medium (e.g., water), giving the electron a speed which exceeds that of light in the medium. When that happens, a burst of Cherenkov radiation is emitted and is detected by light-sensitive photo-tubes.

12.4.1 Chlorine detector

The first neutrino detector, built by Raymond Davis Jr in the 1960s, used a large tank of cleaning fluid ($\mathrm{C_2Cl_4}$) containing 520 tons of Cl^{37}. The goal was to have solar neutrinos interact with Cl^{37} nuclei to produce nuclei of Ar^{37}, and then count how many Argons were in the tank after a certain length of

running time. In order to avoid contamination from backgrounds, the detector was buried deep, almost one mile, underground in a mine in South Dakota (Davis et al., 1968).

The (forward) decay reaction

$$Ar^{37} \rightarrow Cl^{37} + e^{+} + \nu$$

is driven by the mass difference between the ground states of Ar^{37} and Cl^{37}, corresponding to an energy of 0.814 MeV (Audi and Wapstra 1993).

As a result, the Davis' detector (which records events driving the above reaction backward), responds only to neutrinos with energies in excess of 0.814 MeV. In principle, this means that if neutrino capture were to occur mainly via a transition from the Cl^{37} ground state to the Ar^{37} ground state, then Davis should be able to detect the line neutrinos from the pep reaction, and from the higher energy Be^{7} decay, as well as continuum neutrinos from B^{8} decay, the Hep reaction, as well as three decays in the CNO bi-cycle.

However, the nuclear physics is such that transitions from the Cl^{37} ground state to excited states of the Ar^{37} nucleus are preferred, especially to a level at an energy of about 5 MeV above the ground state. As a result, neutrinos in the B^{8} and Hep continua dominate the signal in the Davis detector. Of these, the B neutrinos are dominant by far.

Now that we know which continua will be dominant, we turn to the experimental results. In principle, we are seeking a measurement of the neutrino flux. A convenient unit can be devised which incorporates the likelihood of a neutrino being detected. The common unit for discussing solar neutrino experiments is the solar neutrino unit (SNU): this is defined to be one neutrino capture per second in a detector which contains 10^{36} target nuclei.

Why is this unit useful? Because the interaction cross-section between a neutrino and one of the target atoms is expected to be of order 10^{-43} cm^{2}, while the input flux from the Sun in the B^{8} continuum is expected to be a few times 10^{7} neutrinos cm^{-2} sec^{-1}. The product of these numbers yields an expected capture rate in a detector of a few times 10^{-36} per second. In view of this, a detector containing 10^{36} targets should yield a detection rate of a few per second. By definition, a detection of one per second per 10^{36} targets equals 1 SNU.

The Davis' detector contained roughly 10^{31} chlorine target nuclei. The standard solar model predicted that in one day of running (i.e., about 10^{5} sec), the Davis' detector should record about two neutrino captures. When the exact calibration was done, it was found that two captures per day would correspond to a solar neutrino rate of 8.1 SNU's. (See Figure 12.2). Even allowing the experiment to run for several months at a time, the total yield of argon atoms in the tank at the end of the run was expected to be no more than a few hundred, out of a tank which contained some 10^{31} chlorine atoms: the chemical expertise required to flush out those few argon atoms from an "ocean of chlorine" was truly impressive.

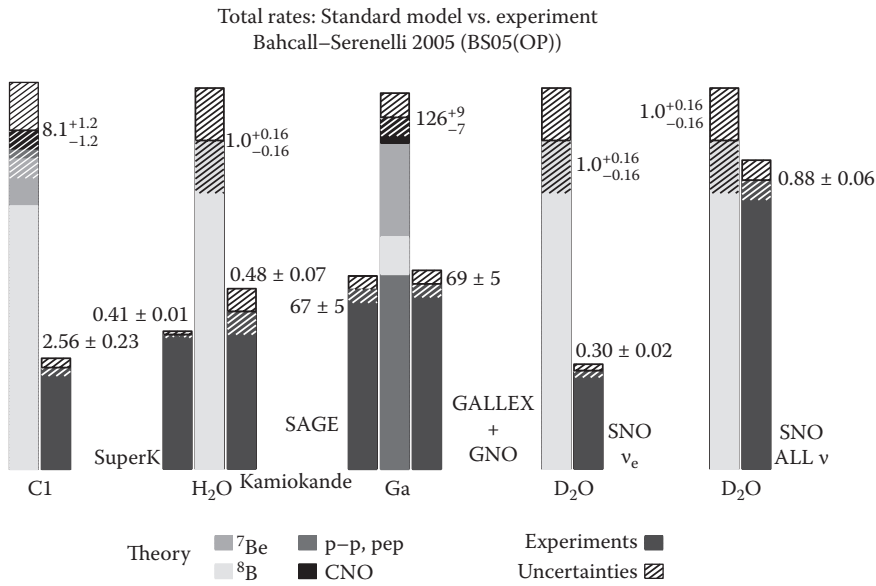


FIGURE 12.2: **(See color insert following page 202.)** Neutrino counting rates: comparison between theory and experiments. The rates are given in units of SNU's (see text). (From the late John Bahcall's website: www .sns.ias.edu/∼jnb, under the heading "Solar Neutrinos: viewgraphs." With permission.)

The experimental results were a surprise. The observed count rates, when averaged over 20 years and more, yielded a rate of only about $0.6-0.7$ captures per day. The corresponding average solar neutrino rate is 2.6 ± 0.2 SNU.

Davis' experiment led to the startling conclusion that the experimental capture rates of solar neutrinos were smaller than predicted by a factor of about three. This shortfall became known as the "solar neutrino problem." We shall return to this below.

12.4.2 Cherenkov emission

If a neutrino collides with an electron in a medium, the electron, called the "knock-on electron," picks up some of the neutrino energy. Solar neutrinos, with energies of up to 14 MeV, can create knock-on electrons which also have energies which are measured in MeV. Such electrons travel with speeds that are close to the speed of light *in vacuo*. If such an electron travels through a medium where the speed of light is reduced to (say) $0.7c$ (such as water), the knock-on electron will be moving faster than light in the medium. This causes emission of light in a Cherenkov cone, with an opening angle determined by the electron's energy. To make the electron fast enough for the Cherenkov process to be possible, the initial neutrino must have a minimum energy.

The Kamiokande detector in Japan, containing some 2000 tons of water, was instrumented with a spherical shell of phototubes to track the Cherenkov cones from solar neutrinos: this detector came on line in 1983. Subsequently, the super-Kamiokande detector, with 50,000 tons of water, and with more than 10,000 phototubes, came on line in 1996. In both cases, the minimum energy required to create knock-on electrons with significant Cherenkov emission is ≥ 5 MeV (e.g., Rothstein, 1992; Takeuchi, 2005). As a result, neither detector could record the main (pp) neutrinos from the Sun: the detectors could respond only to the upper end of the spectrum of B^8 neutrinos.

The standard solar model predicted that the neutrino detection rates should be 1 ± 0.2 SNU. But the experimental results yielded no more than $0.4 - 0.5$ SNU.

A major advantage of the Kamiokande detectors is that they provide information as to the *direction* of the incoming neutrino. The data confirmed that the neutrinos are indeed coming from the Sun.

12.4.3 Gallium detectors

In order to detect the most abundant neutrinos from the Sun (i.e., those from reaction (a) of the pp-I chain), it is necessary to devise a detector in which the threshold energy lies well below the cut-off energy (0.42 MeV) of the pp neutrinos. An isotope of gallium satisfies this criterion.

The relevant decay scheme on which this neutrino detector is based is

$$Ge^{71} \rightarrow Ga^{71} + e^{+} + \nu$$

The mass difference between Ge^{71} and Ga^{71} corresponds to an energy difference of only 0.23 MeV. Consequently, neutrinos can be captured by Ga^{71} if the neutrino energy exceeds 0.23 MeV. Most of the pp neutrinos satisfy this criterion.

With a detector which is sensitive to the most abundant solar neutrinos, the count rate is predicted to be much larger than in the chlorine detector or in the Cherenkov detectors: the gallium detectors were predicted to respond at the rate of 126 SNU's.

Two experiments were built, one in Russia (SAGE, using 50 tons of liquid gallium, with operations starting in 1990: see Abdurashitov et al., 1999), and one in Italy (GALLEX, using a solution containing 30 tons of gallium, with operations starting in 1991: see Kirsten, 2008).

The detection results from both experiments were in agreement with each other, some $67 - 69$ SNU's, but both detection rates were definitely lower than the predictions.

12.4.4 Heavy water detector

A detector containing 1000 tonnes of heavy water (D_2O), surrounded by an even larger volume of clean "ordinary water" (H_2O), was buried 2 km be-

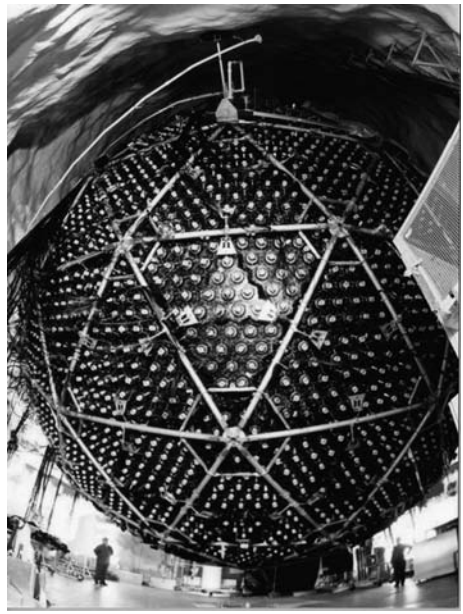


FIGURE 12.3: The inner part of the SNO detector. Notice the scale of this detector: the scale can be estimated from the size of the human beings near the bottom of the image. (Image courtesy of SNO.)

low ground at the Sudbury Neutrino Observatory (SNO) in Ontario, Canada (see Figure 12.3). The container of heavy water plus "ordinary water" was viewed by almost 10000 photomultiplier tubes, arranged on a geodesic dome framework, in order to detect the faint flashes of radiation emitted by particle interactions inside the heavy water. The size of the instrument can be estimated by comparison with the two workers near the bottom of the image. The entire container plus geodesic dome was immersed in a 30-meter barrel of ordinary water (H_2O): the barrel is as tall as a ten-storey building.

The presence of deuterium allowed three distinct classes of reactions to occur involving neutrinos (Ahmad et al., 2002).

a. Knock-on electrons are created as in Kamiokande; these gave results similar to those in Section 12.4.2.

b. Neutrinos associated with neutron/proton decays (called electron neutrinos) interact with D to cause the neutron to decay into a proton. The nucleus then becomes a "di-proton," which is unstable (see Chapter 11, Section 11.5.2). There is a rapid decay into two free protons plus an electron. If the electron is fast enough, a Cherenkov pulse can be detected. The standard solar model predicted 30 of these events per day: the experiment actually recorded only about ten per day.

c. Neutrinos associated with decays *other* than neutron/proton decays belong to distinct families: they are referred to as μ-neutrinos and τ-neutrinos, to indicate the decays with which they are associated. All three neutrino families can interact with deuterium by a process known as the neutral current reaction. This splits the deuterium nucleus, and a free neutron emerges. In the presence of a suitable contaminant nucleus (such as Cl^{35}, added to the water tank in the form of table salt), neutron capture can occur, and gamma rays are emitted.

The standard solar model predicted about 30 neutrinos per day: the experiment recorded essentially that rate. For the first time, a neutrino detector responded in the way that was predicted by the standard solar model.

12.5 Solution of the Solar Neutrino Problem

For a decade or more after Davis announced his first results, the commonest explanation for the solar neutrino problem was that there must be something wrong with the solar model. Attempts were made by solar modelers to add extra effects in the Sun (strong magnetic fields, fast rotation, atypical metal abundances), but these were mostly *ad hoc.* However, as helioseismology (see Chapters 13 and 14) came into its own in the 1980s and 1990s, it became clear that there was very little wrong with the solar model. The solution of the solar neutrino problem must lie in the physics of the elementary particles.

According to the standard model of particle physics, the fundamental constituents of matter consist of six "flavors" of quarks (two of which exist in protons and neutrons) and six leptons. The latter consist of electrons, μ-mesons, and τ-mesons, plus the "flavors" of corresponding neutrinos (electron neutrinos, μ-neutrinos, and τ-neutrinos). Leptons interact only through the weak force, and also (if they are electrically charged) through the electromagnetic force. Although in the standard model, all neutrinos have zero mass, experimental evidence from cosmic rays indicates that this is not exactly true. It turns out that neutrinos have nonzero masses, although the masses are orders of magnitude less than the next lightest lepton (the electron). The existence of finite mass has the effect that neutrinos in different flavors can "mix" among themselves.

The Sun generates electron neutrinos only: all of the decays which generate neutrinos in the Sun involve electrons only (see Chapter 11, Section 11.1, reaction (a), and Chapter 12, Section 12.3.1, reactions (e) and (h), and Section 12.3.2, reactions (j) and (k)). The Sun does not generate either μ- or τ-neutrinos directly. However, as the electron neutrinos propagate through

the Sun, and between the Sun and Earth, the electron neutrinos undergo a mixing process, thereby producing neutrinos in the other two flavors. If enough mixing occurs so as to populate equally all three flavors, then roughly equal numbers of neutrinos are produced in all three flavors.

As a result, only about one-third of the (electron) neutrinos which are generated at Sun survive to reach the Earth as electron neutrinos. The remaining two-thirds reach the Earth as roughly equal mixtures of μ-neutrinos and τ-neutrinos.

The chlorine and gallium detectors are sensitive only to electron neutrinos. Their count rates are smaller than expected because the detectors do not "see" the μ- or τ-neutrinos. The Cherenkov pure-water experiments are in principle sensitive to all three neutrinos, but in practice the cross-section for scattering off electrons in the water favors the electron neutrinos. Therefore, the pure-water detectors respond best to the one-third electron neutrinos, with a weaker response to the other two-thirds. But when SNO experiment (c) was performed, with salty water, all three flavors of neutrinos could participate in the reactions, and were therefore detectable.

The history of the solar neutrino "problem" reads like an exciting detective story. It took 35 years of "big science" in multiple countries to identify the "culprit." The case was solved in 2002 by the SNO experiment (see Section 12.4.4c).

Two significant results emerged from the neutrino detective story. First, as regards the physics of the internal structure of the Sun, the solar models survived a stringent test. Second, in the field of particle physics, a new window "beyond the standard model" was opened up. Both areas of research, solar physics and particle physics, benefited from the long process of solving the solar neutrino problem.

Exercises

12.1 Use the isotope masses in Table 11.1 to show that the cut-off energy in reaction (k) (Section 12.3) is 18.8 MeV.

12.2 For reaction (k) (Section 12.3), show that the energy E_o at which quantum tunneling has maximum effectiveness (see Chapter 11, Section 11.6) is equal to 10.1 keV. Using this value of E_o, show that, due to Coulomb effects alone, reaction (k) occurs almost 10^5 times less frequently than reaction (a) (Chapter 11, Section 11.1).

12.3 Using tabulated values of atomic weights for C and Cl, how that a detector which contains 520 tons of C_2Cl_4 contains close to 10^{31} atoms of chlorine.

References

Abdurashitov, J. N. et al. 1999. "Measurement of the solar neutrino capture rate by SAGE and implications for neutrino oscillations in vacuum," *Phys Rev. Lett.*, 83, 4686.

Ahmad, Q. et al., 2002. "Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory," *Phys. Rev. Lett.*, 89, 011301.

Audi, G. and Wapstra, A. H. 1993. "The 1993 atomic mass evaluation: (I) Atomic mass table," *Nuclear Phys., A*, 565, 22.

Bahcall, J. N., Serenelli, A. M., and Basu, S. 2005. "New solar opacities, abundances, helioseismology, and neutrino fluxes," *Astrophys. J. Lett.*, 621, L85.

Clayton, D. D. 1968. "Major nuclear burning stages in stellar evolution," *Principles of Stellar Structure and Nucleosynthesis.* McGraw-Hill, New York, pp. 362–435.

Cowan, C. L., Reines, F., Harrison, F. B., Kruse, H. W., and McGuire, A. D. 1956. "Detection of the free neutrino: a confirmation," *Science*, 124, p. 103, 1956.

Davis, R., Harmer, D. S., and Hoffman, K. C. 1968. "Search for neutrinos from the Sun," *Phys. Rev. Lett.*, 20, 1205.

Kirsten, T. 2008. "Retrospect of GALLEX/GNO," *Journal of Physics: Conference Series* 120, 052013 (IOP Publishing).

Nakamura, K. 2000. "Solar neutrinos," *Euro. Phys. J. C*, 15, 366.

Rothstein, I. Z. 1992. "Solar ν_e production does not enhance event rates in the Kamiokande detector," *Phys. Rev. D*, 45, R2583.

Takeuchi, Y. 2005. "Solar neutrino measurements in Super-Kamiokande," *Nucl. Phys. B (Proc. Suppl.)*, 149, 125.

Chapter 13

Oscillations in the Sun: The Observations

We have already (in Chapter 12) raised the important question: how can we possibly check on our models of the internal structure of the Sun? After all, the interior of the Sun is surely one of the most inaccessible parts of the world we live in. So it is natural to raise the question: how do we know we are on the right track? Could it be that our calculations are far from reality, and are just plain wrong? Is it possible to check on these calculations?

One way to address this issue is by studying neutrinos, which come from the hottest parts of the solar interior, where nuclear reactions occur. The neutrinos allow us to check on our calculations in the very core of the Sun (see Chapter 12).

But in the 1970s, and especially following a landmark experiment in 1980 at the South Pole (where the Sun was observed continuously for more than five days), a second method of testing the solar models became available. In terms of physics, the radial profile of the properties of the entire solar interior can be checked by studying the properties of certain waves which propagate back and forth inside the Sun. These waves provide us with a "window into the Sun." In this chapter and in Chapter 14, we turn to a study of these waves, and how they can help us to check our calculations of the internal structure of the Sun.

The Sun, although appearing to the unaided eye as being constant in its output, nevertheless is not absolutely unchanging. The most obvious forms of solar variability are sunspots: dark regions on the surface which appear and disappear on semiregular time-scales of days to years (see Chapter 16).

However, when one observes the Sun with sufficiently high resolution, one finds that there are some highly regular variations which occur on time-scales of *minutes*. In this case, the periodicities of the variations are not at all semiregular: on the contrary, they occur on highly precise time scales, which are reproducible every time one observes the Sun. These extremely periodic variations provide a means for us to obtain valuable information about the solar interior akin to the information that geologists obtain about the Earth's interior by studying earthquakes.

The purpose of this chapter is to describe the observations which allow us to determine the properties of the Sun's variations on time-scales of minutes.

The variations can be studied from the point of view of *temporal* variations alone (with no regard for spatial resolution). They can also be studied in data

which are *spatially* resolved across the disk of the Sun. We turn first to the purely temporal variations.

13.1 Variability in *Time* Only

When the Sun is observed "as a star," data are gathered without regard to spatial location on the surface of the Sun. The detector integrates over the entire disk of the Sun. Using high *spectral* resolution, small variations in *velocity* can be detected. When these are analyzed as a time series, a power spectrum is obtained, showing how much power occurs (in velocity) as a function of frequency. An example is shown in Figure 13.1 (Fossat et al., 1981). The data were obtained by observing the Sun continuously for a time interval $T_o \approx 0.5 \times 10^6$ sec. The abscissa show the frequency in units of milliHertz (mHz), while the ordinate shows the power in velocity ($\sim V^2$ per unit frequency). The frequency resolution, of order $1/T_o$, is a few microHertz (μHz).

The striking result in Figure 13.1 is that large amounts of power are observed at certain frequencies, while at other frequencies, there is so little power that it can hardly be distinguished from noise.

Inspection of Figure 13.1 illustrates that the Sun generates significant quantities of power at certain well-defined frequencies ν which extend between (roughly) 2.5 mHz and (roughly) 4.5 mHz. The corresponding periods ($= 1/\nu$) range from about 400 sec down to about 220 sec. Earlier observations of this type, at lower resolution, had detected only a broad peak of power centered

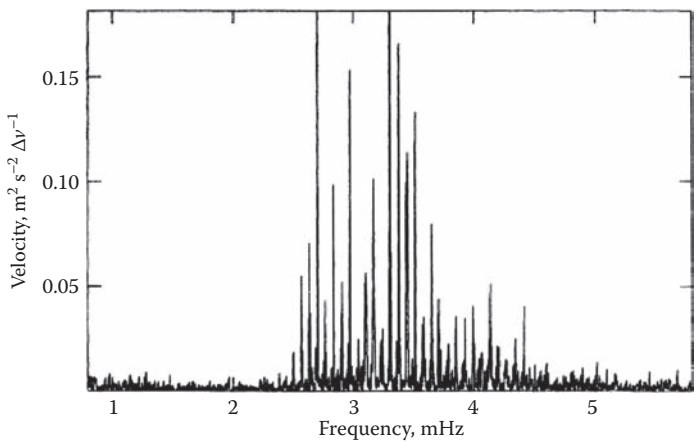


FIGURE 13.1: Power spectrum of solar oscillations in velocity. (From Fossat, E., Grec, G., and Pomerantz, M. A. 1981. *Solar Phys.*, 74, 59. Used with permission of Springer Science and Business Media.)

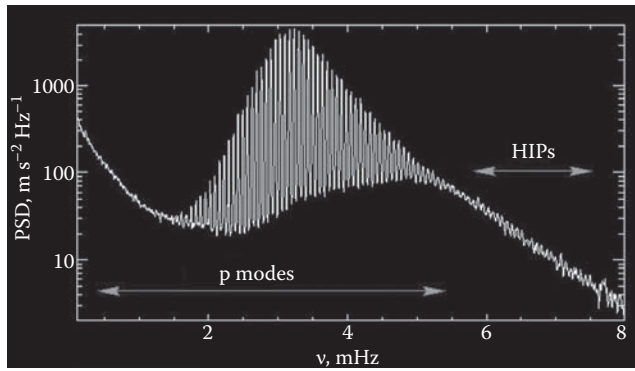


FIGURE 13.2: Logarithmic plot of p-mode power in the Sun. Ordinate: power spectral density (PSD) of velocity oscillations in the Sun derived from 800 days of measurements. The p-modes are the dominant signal at low frequencies ($\leq$ 5 mHz). At high frequencies, a different phenomenon appears: high interference peaks (HIPs), arising from partial wave reflection in the solar atmosphere. (Courtesy of SOHO/GOLF consortium. SOHO is a project of international cooperation between ESA and NASA.)

at frequencies of about 3.3 mHz, i.e., periods of about 300 sec: for this reason, the early observers referred to the oscillations as "five-minute oscillations."

The plot in Figure 13.1 uses a *linear* axis for the power scale. This allows us to separate readily the largest peaks in the power spectrum. But a linear plot makes it difficult to identify the smallest oscillations in the spectrum. In order to enhance our ability to see the smaller oscillations, a logarithmic plot (taken from the website of the Solar and Heliospheric Observatory [SOHO]) is presented in Figure 13.2. The range in frequency ν extends from less than 0.5 mHz to 8 mHz. The oscillations with the largest power levels (having amplitudes of $4000 - 5000$ m^2 sec^{-2} Hz^{-1}), exist at $\nu = 3 - 3.5$ mHz, i.e., in the 5-minute range, as expected. But now, the plot allows us to identify some oscillatory power from the "noise" out to $\nu \approx 5$ mHz (i.e., periods as low as 200 sec). The last identifiable modes at large ν have amplitudes of less than 100 m^2 sec^{-2} Hz^{-1} (after subtracting the background), i.e., almost 100 times weaker than the peak power.

On the low-frequency side of the peak, oscillatory power can be identified down to $\nu \approx 1.5$–1.6 mHz (i.e., periods of order 10 minutes). The last identifiable modes at low ν contain power which, after subtracting the background, has a numerical value of perhaps 10 m^2 s^{-2} Hz^{-1}, i.e., three orders of magnitude smaller than the peak power in the 5-minute range.

The data in Figures 13.1 and 13.2 illustrate that the Sun produces power at a multitude of remarkably "spiky" peaks, i.e., the Sun is oscillating ("ringing") in many *very specific* tones. The tones are referred to as "p-modes": they are caused by acoustic modes (i.e., pressure waves) which are trapped inside the

Sun (see Chapter 14). The narrowness of the "spiky" peaks in Figure 13.1 is striking. The ratio of line frequency ($\nu \approx 3$ mHz) to line width (≈ 1 μHz) is a few thousand, indicating that when the Sun "rings," the "quality factor" of the resonant cavity is very high.

It is also apparent from Figures 13.1 and 13.2 that the spikes are not distributed at random in frequency: even the unaided eye can see that there is a preferred spacing (≈ 0.07 mHz) between adjacent peaks. Actually, a more fundamental spacing turns out to have about twice this value: many modes are found to be separated by $\Delta\nu = 135$–136 μHz (Appourchaux et al., 1998). This interval contains important information about the interior of the Sun (see Chapter 14).

The peak *power* in any given p-mode does not remain invariant, but fluctuates with time. An example of the variability in a single mode ($l = 0$, $n_r = 21$: these labels will be defined below) is shown in Figure 13.3. The mode amplitude is plotted in a perspective 3-D diagram with time along one axis, frequency of the mode along another axis, and power along the vertical axis. The mode in question has $\nu \approx 3034$ μHz. In the course of an observing run of 4–5 months, the *power* of the mode varies by a factor of ≈ 10, forming a "mountain range" in the plot. The peak in power does *not* shift significantly in *frequency* as time goes on. The reason why peaks come and go in the "mountain range" has to do with how the mode is generated: modes are generated by convective flows at certain depths (see Chapter 14, Section 14.8), and these flows are highly variable.

13.2 Variability in *Space* and *Time*

The data in Figures 13.1 and 13.2 refer to variability in *time* only: such data are obtained when the Sun is observed *as a star*, with no attempt to resolve the Sun's disk spatially. However, valuable information about the Sun can also be extracted from the *spatial* properties of the variations. To do that, the Sun must be observed with data which are not only well-resolved in time, but also *spatially* resolved. The higher the angular resolution which is used to obtain the data, the smaller the patches on the Sun's surface which can be examined for oscillation.

From a mathematical perspective, when one analyzes the properties of spatial variations on a spherical surface, it is natural to use "spherical harmonic functions" $Y_{lm} = P_l^m(\cos\theta)e^{im\phi}$ to describe the surface structure. Here, θ is the colatitude, and ϕ is the longitude. The index l refers to structure in the latitudinal direction, between the North pole ($\theta = 0$) and the South pole ($\theta = \pi$). The index m refers to structure in longitude. Initially, we neglect longitudinal variations, and consider $m = 0$. This allows us to reduce Y_{lm} to the Legendre functions $P_l(\cos\theta)$. For $l = 0, 1, 2,$ and 3, $P_l(x) = 1$, x, $(3x^2 - 1)/2$, and $(5x^3 - 3x)/2$.

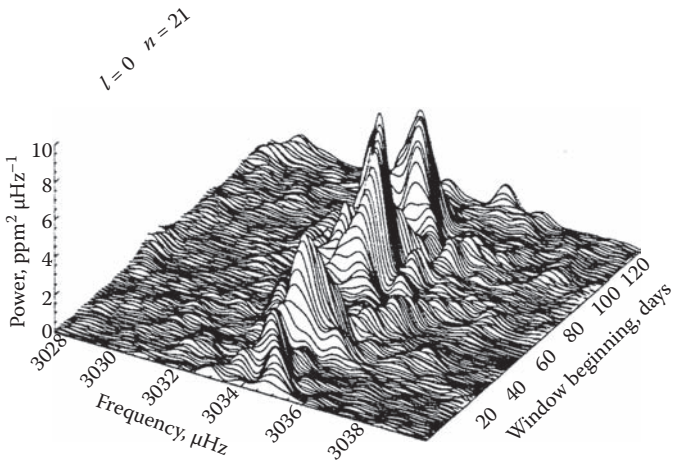


FIGURE 13.3: A 3-D plot showing, as a function of time, the variations which occur in the amplitude and in the frequency of one particular mode of oscillation in the Sun. The mode in this figure has degree $l = 0$, and radial order $n_r = 21$. (The label $n = 21$ in the upper left corner of the figure corresponds to $n_r = 21$ in our notation.) Of the two axes which lie in the "horizontal plane" in the plot, the one on the right-hand side of the figure indicates the passage of time (in units of days) since the beginning of the observing window. This particular observing run lasted more than four months. The second "horizontal axis" indicates the frequency of the oscillation (in units of μHertz). The particular mode in the figure has a frequency which lies between 3028 and 3038 μHz. Rising above the "horizontal plane" in the figure, the "vertical axis" shows the power in the oscillation mode at each instant of time and at each frequency: units of power are ppm^2 μHz^{-1} (where ppm = parts per million). (From Gavryusev, V. G. and Gavryuseva, E. A.1997. *Solar Phys.*, 172, 27. Used with permission from Springer Science and Business Media.)

The parameter l is the "angular degree" of the mode, the number of nodes (i.e., regions of zero amplitude) which exist in the oscillatory structure between North and South poles. For modes with $l = 0$ (no nodes in latitude), gas motions are synchronized over the entire surface: at a given instant, the gas is moving outward (at all points of the surface), and then one half-cycle later, the gas is moving inward (at all points of the surface). For modes with $l = 1$, there is one node in latitude, at $cos(\theta) = 0$, i.e., the equator: at a given instant, when the gas in the Northern hemisphere is moving outward, the gas in the Southern hemisphere is moving inward. One half-cycle later, the Northern gas moves in, while the Southern gas moves out. For modes with $l = 2$, at a given instant, gas moves outward between the North pole and latitude $\sin^{-1}(1/\sqrt{3}) = 35°$N, gas moves inward in the equatorial regions (at latitudes between 35° N and 35°S), and gas moves outward from 35°S to the South pole. One half-cycle

later, the outward motion is confined to the equatorial regions, while the polar "caps" move inward. For modes with $l = 3$, nodes occur at latitudes 51°N, 0 (the equator), and 51°S.

The larger the l value, the more nodes one must "squeeze" into the range of latitudes from $+90°$ to $-90°$, and the closer the nodes approach each other. When l is large, the linear distance between adjacent nodes along a great semicircle from the North to South pole is roughly equal to the length of that semicircle divided by l. The distance between adjacent nodes is equivalent to one-half of one wavelength. Formally, a linear distance which can be regarded as the "horizontal wavelength" λ_h of a mode is given by

$$\lambda_h = \frac{2\pi R_\odot}{\sqrt{l(l+1)}} \tag{13.1}$$

Thus, a mode with degree $l = 250$ has $\lambda_h \approx 17500$ km. The angular scale of this on the Sun's surface, as observed from Earth, is about 20–25 arc sec: therefore, in order to obtain meaningful information about modes with $l \geq 250$, we need to make observations of the Sun with angular resolutions which are at least as good as 5–10 arc sec: this will allow us to have a few "pixels" across one wavelength of the $l = 250$ modes.

So, given an observing scheme which allows us to measure *velocities* across the disk of the Sun with resolutions that are as good as *a few arc seconds*, we can analyze the data in terms of its *spatial* properties as well as its *temporal* properties. To extract a power spectrum corresponding to a given l value, a data set (which has been averaged over longitude) is convolved with the particular spherical harmonic Y_l. For each l, the resulting series is subjected to time-series analysis, and a power spectrum is obtained for that l value: the power spectrum will consist of "spikes" at a number of discrete frequencies (reminiscent of Figure 13.1). Repeating the analysis for many l values, the resulting power spectrum can conveniently be plotted in 2-D, with spatial information (the degree of the mode, l) along one axis, and temporal information (the frequency of the mode) along the other axis. (See Figure 13.4). The power in any particular mode is indicated by the intensity of the "dot" which is used to plot that mode.

What do the results in Figure 13.4 tell us about the Sun? They tell us that oscillations in the Sun occur preferentially at certain frequencies: this is easiest to see near the left-hand edge of the figure, where individual white dots can be distinguished clearly from one another. If, e.g., we fix attention on modes with $l = 10$, i.e., with $\lambda_h \approx 420$ Mm, we can determine the distribution of power in the Sun by running our eye vertically along the left-hand side of the figure. We see that power is present at detectable levels only at a particular set of certain discrete frequencies.

Moreover, the white spots in Figure 13.4 do not exist near the top and bottom edges of the plot, i.e., at high and low ν. In this data set, obtained at the Earth's South Pole, the noise in solar observing is such that power cannot

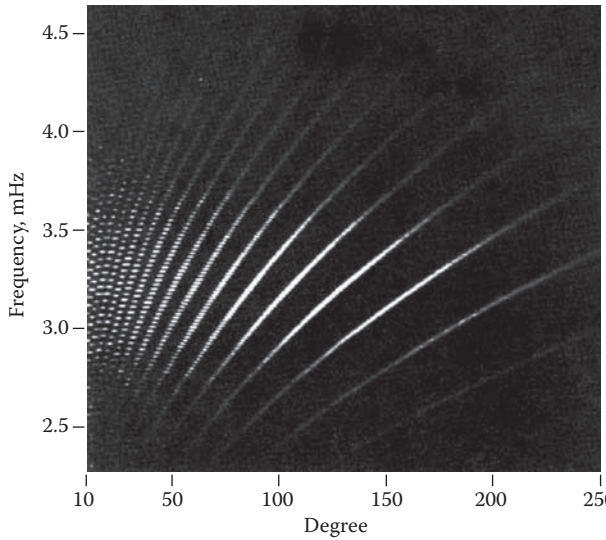


FIGURE 13.4: Spatial-temporal power spectrum of solar oscillations. Abscissa: angular degree of the mode. Ordinate: frequency of the mode. Power in each mode is indicated by the intensity of the white "dot." Observations obtained during a 50-hour continuous run at the South Pole in December 1981. (From Duvall, T. L., Harvey, J. W., Libbrecht, K. G., Popp, B. D., and Pomerantz, M. A., *Astrophys. J.*, 324, 1158, 1988. With permission.)

be detected reliably at $\nu \leq 2.5$ mHz, or at $\nu \geq 4$ mHz. So far, this does not seem to go much beyond what we learned from Figure 13.1.

However, in Figure 13.4, there is actually a great deal more information than in Figure 13.1. Specifically, we can also choose $l =$ (say) 11 or 30 or 100, and run our eye vertically through Figure 13.4 to identify where "white spots" lie. Again, each white spot lies at a particular frequency, although not the same set of frequencies as we found for $l = 10$. As we examine modes with higher ν for *a fixed value of* l, a striking feature of solar oscillations emerges: the interval in frequency between adjacent modes approaches a constant asymptotic value $\Delta\nu$. The value of $\Delta\nu$ varies only slowly with l: for l in the range 0–10, $\Delta\nu$ is found to lie in the range 135–138 μHz (Appourchaux et al., 1998).

Frequencies can be derived for modes with all values of l down to $l = 0$ (although results in Figure 13.4 only go down to $l = 10$). For each value of l a list of frequency "spikes" can be prepared. Each such list corresponds to a vertical "cut" through Figure 13.4. By way of example, we note that, for $l = 0$, some 20 peaks were listed by Duvall et al. (1988), with ν ranging from 1824 to 4669 μHz, e.g., the list contains entries at $\nu = 2899$, 3034, and 3169 μHz. The mode at $\nu = 3034$ μHz is the one which appears in Figure 13.3 above.

Notice that the intervals between these three modes are 135 and 135 μHz. The presence of a well-defined separation in frequency between adjacent modes is an example of "asymptotic behavior" (see Chapter 14).

13.3 Radial Order of a Mode

The question is: for a fixed value of l, what do the different "spikes" in frequency correspond to? Why are there only certain frequencies in the list of "spikes"? Empirically, the answer is not immediately obvious. We need to turn to theory (Chapter 14). Theory tells us that each mode of the Sun follows a certain pattern *on the surface of the Sun* (with nodes at well-defined latitudes, determined by l), but as well, each mode has a certain functional form *in the radial coordinate between the center of the Sun and the surface.* The functional form in radius is called the "radial eigenfunction," and it contains a series of "ups and downs" (see e.g., Chapter 14, Figures 14.2 and 14.3). Between each "up" and the next "down," the eigenfunction passes through zero at a specific radial location. Each such zero is a node of the radial eigenfunction. The number of nodes n_r between the center and the surface is called the "*radial order*" of the mode.

Now we are in a position to interpret the "spikes" in frequency for a given l: each "spike" in ν corresponds to a particular integer n_r. For a fixed l, the value of n_r increases as the ν increases.

In contrast to the angular *degree* l, which can be determined *empirically* by examining how oscillations with a particular period are distributed across the *surface* of the Sun, there is no *purely empirical* way to determine the value of n_r. The radial order can be determined only by comparing the observed frequency with calculations of the interior structure of the Sun, and seeing which frequency fits best. Such an exercise leads to the conclusion that the three $l = 0$ modes mentioned above with $\nu = 2899$, 3034, and 3169 μHz correspond to $n_r = 20$, 21, and 22. In terms of this notation, the mode in Figure 13.3 corresponds to $l = 0$ and $n_r = 21$. (Note that the authors of Figure 13.3 use the slightly different notation $l = 0$, $n = 21$: the reason we prefer to use the label n_r for the radial order rather than n is to avoid confusion with n as the polytropic index.)

Now that we have introduced the radial order n_r, we are in a position to draw attention to another feature of Figure 13.4. Inspection of Figure 13.4 shows that, for modes with the largest l values in the plot, the individual white "dots" merge together and form narrow "ridges" of whiteness (i.e., power) slanting from lower-left to upper-right. Model fitting indicates that along each of these ridges, *the radial order* n_r *retains a unique value.* The value of n_r is smallest for the ridges which lie closest to the right-hand side of Figure 13.4. Model fitting suggests that the right-most (barely) visible ridge in Figure 13.4 has $n_r = 4$. More sensitive instruments (e.g., the long running MDI on the

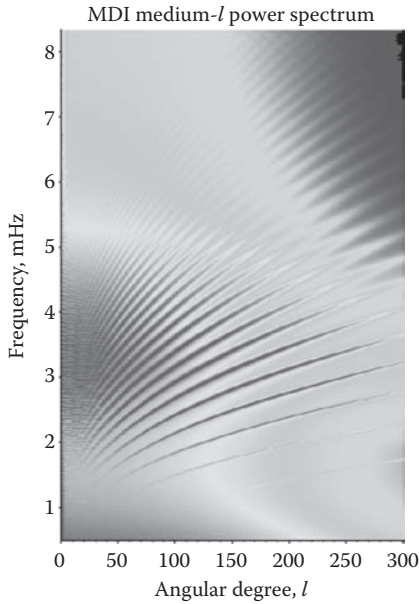


FIGURE 13.5: **(See color insert following page 202.)** Power spectrum of solar oscillations obtained from long observing runs in space. The lowest lying "ridge" at $\nu = 1.3$–1.7 mHz and $l = 200 - 300$ is due to fundamental modes, with $n_r = 0$. (Courtesy of SOHO/MDI consortium. SOHO is a project of international cooperation between ESA and NASA.)

SOHO spacecraft) can identify ridges of power even farther to the right, with $n_r = 3, 2, 1$, and even $n_r = 0$ (the fundamental mode) (Figure 13.5).

Figure 13.5 shows that not only does the Sun emit acoustic power at frequencies between 2.5 and 4 mHz (as in Figure 13.4), but there are also ridges of power at frequencies down to (about) 1.3 mHz and up to (about) 5 mHz. Because the observing conditions at the South Pole are noisier than at SOHO, the weaker ridges in the SOHO data (Figure 13.5) cannot be reliably identified in the South Pole data (Figure 13.4). Nevertheless, the conclusion we drew from Figure 13.4 remains valid: the amount of oscillatory power which is observed in each ridge (i.e., the "height" of the ridge) becomes progressively smaller as n_r decreases to the smallest values.

13.4 Which *p*-Modes Have the Largest Amplitudes?

Inspection of Figure 13.4 shows that the Sun appears to pump power into modes with lower l and/or larger n_r values. For example, in tables of modes

which were reliably identified in the Sun in the earliest data sets (e.g., Duvall et al., 1988), modes with $l = 0$ and those with $l = 1$ have been identified with n_r values ranging from $n_r = 12$ to $n_r = 33$. Modes with $l = 2$ and $l = 3$ have been reliably identified with n_r values ranging from $n_r = 11$ to $n_r = 32$. Outside these ranges, e.g., at $n_r < 10$, it is difficult to detect modes reliably in most data sets. In terms of frequency, the highest and lowest ν which are reliably detected are close to 5 and 1.5 mHz, respectively (Figure 13.2).

It is natural to ask: why does the Sun produce acoustic modes with just this range of frequencies? Why do we not detect modes with ν as high as 10 or 20 mHz, or as low as 0.5–1 mHz? As regards the low ν behavior, we postpone the discussion until we have a mathematical description of the eigenfunctions (Chapter 14, Section 14.8). But we do not need to understand eignfunctions in order to determine why high ν modes are absent. So let us turn to that piece of the puzzle.

13.5 Trapped and Untrapped Modes

The reason why p-modes are detectable in the Sun has to do with the fact that certain acoustic waves are able to build up to large amplitudes. And the reason for the large amplitude is that certain waves are *not allowed* to propagate freely out of the Sun from the place where they are generated. Instead, once those waves are generated, there is something about the Sun's internal structure that causes the waves to be *trapped:* given favorable conditions, trapped waves can build up to large amplitudes. It is precisely such waves which give rise to the high levels of power in the discrete peaks in Figures 13.1 and 13.2, and to the high levels of power in the discrete sets of "white dots" in Figure 13.4.

On the other hand, certain other acoustic waves *can* propagate freely from their place of origin, reaching into the upper atmosphere of the Sun. Those propagating waves, which are not subject to trapping, are of great interest in the context of the heating of the solar chromosphere (see Chapter 15, Section 15.9).

In this section, we ask: can we identify a dividing line between trapped and untrapped acoustic waves in the Sun? We shall find that the dividing line occurs at a certain critical wave period P_c: waves which have periods shorter than P_c are untrapped, while waves with periods longer than P_c can be trapped.

In this regard, we recall the empirical result that the p-modes which have significant amplitudes are observed to have $\nu = 2.5$–4.5 mHertz: such frequencies correspond to wave periods of 220–400 sec. This tells us that there is something about the solar atmosphere which causes waves with such periods to be trapped: the "something" has to do with the fact that the atmosphere is stratified. We now consider the physical reasons why this is the case.

13.5.1 Wave propagation in a stratified atmosphere

Consider the propagation of a sound wave vertically in an atmosphere which is in HSE. As a result of HSE, the pressure in the unperturbed atmosphere has a vertical distribution $p_o(z)$ which obeys the equation

$$\frac{dp_o}{dh} = -\rho_o g \tag{13.2}$$

In this equation, the vertical coordinate h increases in the *upward* direction. The right-hand side represents the weight of 1 cm^3 of gas acting as a downward force on each square cm of the atmosphere at height h. This downward force is balanced by the gradient of the pressure acting upward on that square cm.

We now superpose a small vertical displacement ξ on a parcel of gas caused by a local perturbation in pressure. We assume that the displacement raises or lowers the parcel in such a way that initially there is no change in density. Denoting the new pressure by $p(h) = p_o(h) + \Delta p$, we see that the vertical forces acting on 1 cm^3 of gas,

$$\Delta F = \frac{dp}{dh} + g\rho_o \tag{13.3}$$

no longer sum to zero: HSE is not satisfied. In view of the conservation of momentum (i.e., Newton's second law of motion), the imbalance in vertical forces ΔF acting on 1 cm^3 of gas (with mass ρ_o) leads to a vertical acceleration $d^2\xi/dt^2$ such that

$$\rho_o \ddot{\xi} = -\frac{dp}{dh} - \rho_o g = -\frac{d}{dh}(\Delta p) \tag{13.4}$$

In Equation 13.4, double dots denote the second *time* derivative. This is the equation of motion for the parcel of gas as it responds to the change in ambient pressure gradient. (In Chapter 7, Section 7.1, we also applied Newton's law of motion to gas in the convection zone, where the breakdown of HSE also leads to an imbalance of forces, with consequent vertical acceleration.)

As the parcel of gas moves, its density does not remain constant: the internal density changes by a finite amount $\Delta\rho$ if the magnitude of the displacement ξ varies with h. To quantify this, we consider the conservation of mass associated with the displacement. Suppose that the initial parcel of gas spanned an interval of height Δh, but when it is displaced, that parcel spreads out over a different interval of height $\Delta\xi$. There are two possibilities. On the one hand, if $\Delta\xi$ *exceeds* Δh, the parcel has been "stretched," and the internal density of the gas *decreases* as a result of the displacement. On the other hand, if $\Delta\xi$ is *less than* Δh, the parcel has been compressed, and the internal density of the gas *increases* as a result of the displacement. In both cases, the fractional change in density is related to the ratio of $\Delta\xi$ to Δh. In the limit, the fractional change in density which occurs in the parcel of gas can be written as

$$\frac{\Delta\rho}{\rho} = -\frac{d\xi}{dh} \tag{13.5}$$

Note that if ξ is the same at all h, i.e., if the atmosphere is displaced as a whole by a constant amount, the derivative $d\xi/dh$ equals zero, and there is no change in density.

Now that we have taken into account the conservation of momentum and the conservation of mass, it remains to address the conservation of energy. In the present case, we will incorporate this by assuming that the sound waves propagate in an adiabatic manner, i.e., $p \sim \rho^\gamma$. Thus, $\Delta p/p = \gamma \Delta\rho/\rho = -\gamma d\xi/dh$. To first order, we replace Δp in Equation 13.4 by $-\gamma p_o d\xi/dh$.

Using this in Equation 13.4, we find

$$\rho_o \ddot{\xi} = \frac{d}{dh}\left(\gamma p_o \frac{d\xi}{dh}\right) = \gamma p_o \xi'' + \gamma \xi' \frac{dp_o}{dh} \tag{13.6}$$

Here, primes denote *spatial* derivatives with respect to h. Dividing through by ρ_o, and noting that the adiabatic sound speed c_s is given by $c_s^2 = \gamma p_o/\rho_o$, we rewrite Equation 13.6 in the form

$$\ddot{\xi} = c_s^2 \xi'' - \gamma g \xi' \tag{13.7}$$

Equation 13.7 is the equation which describes how *a sound wave propagates in a stratified atmosphere*, i.e., in an atmosphere where finite gravity is present.

In the special case where gravity is absent, i.e., $g = 0$, when the background medium is unstratified, Equation 13.7 reduces to the standard wave equation $\ddot{\xi} = c_s^2 \xi''$: this describes waves propagating with speed c_s. In this case, there are no limitations on the direction of travel, or on the frequency of the waves which may propagate.

However, in the presence of gravity, the second term on the right-hand side of Equation 13.7 comes into play. This is the term which makes a significant difference to the properties of sound wave propagation in a stratified atmosphere.

13.5.2 Simplest case: the isothermal atmosphere

Let us consider the case of an isothermal atmosphere with temperature T. In this case, we have already (Chapter 5, Section 5.1) seen that the density is stratified as a function of height according to an exponential law: $\rho(h) = \rho_o \exp(-h/H)$ where the scale height H is given by the expression $H = R_g T/g\mu_a$ where μ_a is the mean atomic weight of the atmospheric gas. For a perfect gas, the sound speed c_s can be written as $c_s^2 = \gamma R_g T/\mu_a = \gamma g H$.

In order to proceed with the solution in this case, it is convenient to transform to dimensionless variables. We introduce a new dimensionless length coordinate: $h' = h/2H$. Note that in order to convert the dimensional length h to dimensionless form, we normalize to a length which is *not* the scale height, but *twice* the scale height.

We also introduce a new dimensionless time scale: $t' = t/(2H/c_s)$. In order to arrive at a dimensionless time, we normalize to the time required for sound to traverse the length scale $(2H)$ which appears in the definition of h'.

Converting now the temporal and spatial derivatives to the new dimensionless variables, and making use of the relation $c_s^2 = \gamma g H$, we find that Equation 13.7 takes on the form

$$\ddot{\xi} = \xi'' - 2\xi' \tag{13.8}$$

Here, dots denote differentiation with respect to the dimensionless time t', while primes denote differentiation with respect to the dimensionless length h'. Equation 13.8 describes, in terms of dimensionless length and time variables, how a stratified atmosphere responds to a sound wave propagating vertically.

To solve Equation 13.8, one further change of variables is helpful: we replace ξ with the auxiliary variable $u = \xi \exp(-h')$. This leads to the equation

$$\ddot{u} = u'' - u \tag{13.9}$$

In order to describe wave motion, we seek a periodic solution to this equation: $u = u_o e^{i\omega t'}$ where u_o is a function of the spatial coordinate h' only, and ω is the frequency associated with the dimensionless time t'. Substituting this in Equation 13.9, we find

$$u_o'' = -(\omega^2 - 1)u_o \equiv -Au_o \tag{13.10}$$

where A is defined as $\omega^2 - 1$.

Mathematically, Equation 13.10 has two well-known classes of solutions, depending on the algebraic sign of A.

Class (i) $A > 0$. The parcel of gas undergoes a displacement u which is simple harmonic motion in the (dimensionless) height coordinate h':

$$u_o = \exp(\pm ih'\sqrt{A}) \tag{13.11}$$

When combined with the sinusoidal time factor $e^{i\omega t'}$, Equation 13.11 represents a *sound wave freely propagating* in a vertical direction through the atmosphere.

The condition that A must exceed zero for the solution of Equation 13.11 to be valid means that ω^2 must exceed unity. That is, the wave *frequency must exceed a certain value in order that the wave may propagate freely in the vertical direction.*

Class (ii) $A < 0$. In this case, the solution is an exponential with a real argument:

$$u_o = \exp(\pm h'\sqrt{A}) \tag{13.12}$$

In order to avoid divergence at infinity, only the *damped* solution in Equation 13.12 is physically meaningful. This damped solution indicates that waves *do not propagate* in this case. The nature of the solution in Class (ii) is very different from the solution in Class (i)

13.5.3 Critical frequency and the critical period

The transition between the *freely propagating* solutions in Class (i) and the *damped (nonpropagating)* solutions in Class (ii) occurs at $A = 0$, i.e., at $\omega^2 = 1$. Reverting to dimensional variables, the corresponding transition occurs at the critical frequency

$$\omega = \omega_{\rm ac} \equiv \frac{c_s}{2H} \tag{13.13}$$

This critical frequency, identified by subscript "ac" is referred to as *the acoustic cut-off frequency.*

The cut-off *period* $P_{\rm ac}$ associated with the cut-off *frequency* is given by

$$P_{\rm ac} = \frac{2\pi}{\omega_{\rm ac}} = \frac{4\pi H}{c_s} = \frac{4\pi}{g}\sqrt{\frac{R_g T}{\gamma \mu_a}} \tag{13.14}$$

13.5.4 Physical basis for a cut-off period

Why, in physical terms, does a cut-off period exist in the Sun's atmosphere? Why is it that waves with periods longer than $P_{\rm ac}$ cannot propagate vertically? To answer this, consider what happens if one tries to launch a wave with a certain period into the atmosphere. If the wave period is longer than $P_{\rm ac}$, the spatial extent of one wavelength ($\lambda = c_s P_{\rm ac}$) of such a wave extends over many (4π, i.e., > 10) scale heights of the atmosphere. With such a large length-scale, during the time that the wave is propagating across one of its own wavelengths, the atmosphere has time to "adjust itself" to the perturbation: the effects of the adjustment are to cancel out the wave. The stratified atmosphere in effect can "short out" the wave.

On the other hand, a wave with a short period can propagate across one of its own wavelengths before the atmosphere has time to adjust and "cancel out" the wave.

13.5.5 Numerical value of the cut-off period

Now we come to a key question: what is the numerical value of the critical period in the Sun? In the upper photosphere of the Sun, where $T = 4860$ K and $\mu_a \approx 1.3$, we find

$$P_{\rm ac} \approx 195 - 200 \text{ sec} \tag{13.15}$$

The corresponding cut-off frequencies are $\nu_{\rm ac} \approx 5.0$–5.1 mHz.

Recall that the p-modes in the Sun have detectable amplitudes for frequencies which are no greater than (about) 5 mHz (see Figure 13.2). Now that we have derived the concept of the acoustic cut-off of a stratified atmosphere, we can understand why p-modes are not detectable with periods shorter than (roughly) 200 sec: such waves, with frequencies $\nu > 5$ mHz, have $\nu > \nu_{\rm ac}$.

Such waves are free to propagate vertically through the solar atmosphere. Such waves are therefore *not* trapped: they escape easily from the location where they are generated. They do not "stick around" long enough to build up their amplitude.

On the other hand, acoustic waves with periods longer than 200 sec *cannot* propagate vertically through the upper solar photosphere. When such waves encounter the upper photosphere, they are not permitted to propagate further in a vertical direction: *instead, they are reflected back down into the Sun.* This sets up the possibility of those waves becoming trapped. And if they are trapped, then they can "stick around" long enough to have energy pumped into them. The more energy is pumped in, the larger their amplitudes becomes, and the easier it is for us to detect them.

The principal conclusion of this section is the following: the effects of atmospheric stratification explain why p-modes are *not* detectable at ν *larger* than (roughly) 5 mHz.

In order for waves to be trapped, the Sun needs to provide a "cavity" of some kind to contain the waves. We have now identified wave reflection as a reason why there exists an *upper* boundary to such a cavity in the Sun. In Chapter 14, we shall discuss a mechanism which causes a *lower boundary* to the cavity.

13.6 Long-Period Oscillations in the Sun

The presence of oscillations in the Sun at periods of a few minutes has been known for several decades. These are the well-studied p-modes.

But we can also ask: are there oscillations in the Sun with much longer periods (e.g., hours)? The answer to this question was controversial for many years. However, in 2007, it was reported that the Sun exhibits a large number of modes with periods as long as 15 hours (Garcia et al., 2007).

As mentioned in Chapter 1, Section 1.10, there exists a critical (global) frequency $\nu_g \approx 100$ μHz which is expected to be relevant to solar oscillations. The physical origin of ν_g is very different from the physical origin of the acoustic cut-off frequency (Equation 13.13). The latter arises from acoustic processes which are *localized* in the stratified atmosphere, within lengths scales of order one scale height (≈ 100 km), whereas the critical (global) frequency is determined by length scales of order the solar radius ($\approx 700{,}000$ km), ν_g is controlled by gravity, rather than by acoustic processes.

The critical period corresponding to ν_g is 2–3 hours. We note that the p-modes in the Sun all have frequencies which are *larger* than ν_g. It now appears that the long-period modes reported by Garcia et al. (2007) have frequencies which are *smaller* than ν_g.

The theory which will be presented in Chapter 14 predicts that two distinct classes of oscillation should exist, one at high frequency (the p-modes), the

other at low frequency (the g-modes). The critical frequency ν_g provides a natural dividing point between these two classes. In view of this, it seems likely that Garcia et al. have identified g-modes in the Sun.

Exercises

13.1 In Chapter 1, Exercise 5, you have already calculated surface gravities for five "main sequence" stars with masses of 0.1–10$M_\odot$. Calculate the cut-off period $P_{\rm ac}$ for each star, assuming $\mu_a = 1.3$, and setting $T = T_{\rm eff}$ as calculated in Chapter 1, Exercise 6.

13.2 Using the properties of the same five "main sequence" stars as in Exercise 1, calculate the critical gravity period P_g (Chapter 1, Equation 1.20) for each star.

13.3 How large must the degree l of a mode in the Sun be in order to have a horizontal wavelength equal to (a) a supergranule diameter ($\approx$ 30 thousand km; see Chapter 15), (b) a granule diameter?

References

Appourchaux, T., Rabello-Soares, M.-C., and Gizon, L. 1998. "The art of fitting p-mode spectra. II. leakage and noise-covariance matrices," *Astron. Astrophys. Suppl.*, 132, 131.

Duvall, T. L., Harvey, J. W., Libbrecht, K. G., Popp, B. D., and Pomerantz, M. A. 1988. "Frequencies of solar p-mode oscillations," *Astrophys. J.*, 324, 1158.

Fossat, E., Grec, G., and Pomerantz, M. A. 1981. "Solar pulsations observed from the geographic south pole: latest results," *Solar Phys.*, 74, 59.

Garcia, R. A. et al., 2007. "Tracking solar gravity modes; the dynamics of the solar core," *Science*, 316, 1591.

Gavryusev, V. G. and Gavryuseva, E. A. 1997. "Statistical properties of the pulses in solar p-mode power," *Solar Phys.*, 172, 27.

Chapter 14

Oscillations in the Sun: Theory

In order to understand in physical terms why the Sun exhibits oscillations at precisely defined frequencies, we consider in this chapter the oscillations in an idealized "star." Specifically, we revert to the topic of polytropes (see Chapter 10), since these provide in certain cases an analytic form for the radial profile of pressure and density in which oscillations can occur. Of course, if we were undertaking a detailed examination of the Sun, we would have to make use of the full numerical radial profiles of pressure and density: but those numerical solutions make it more complicated to derive the properties of the oscillations. So in this first course in solar physics, we simplify the problem by considering the oscillation modes of a polytrope. Results from the polytropic case contain many of the important characteristics of oscillations in the "real Sun."

In this chapter, we derive a pair of first order differential equations (Equations 14.17 and 14.18) which describe the properties of oscillations in a polytrope. The pair of equations which we shall derive represent a simplification of the full oscillation problem (which requires four equations to specify completely). Nevertheless, many of the properties which are observed to occur in solar oscillations can be modeled with the simpler system. Moreover, students will have an easier time exploring the computational properties of the simpler system.

From Chapter 10, Equation 10.10, we recall the Lane–Emden equation for the polytrope of order n:

$$\frac{1}{x^2}\frac{d}{dx}\left[x^2\frac{dy}{dx}\right] = -y^n \tag{14.1}$$

Here, y is related to the density at any given radial location by the expression $y^n = \rho/\rho_c$ (where subscript c denotes the central value), and the (dimensionless) spatial coordinate x is related to the (dimensional) radial coordinate r by $x = r/r_o$, where the Emden unit of length r_o is defined by

$$r_o^2 = \frac{(1+n)p_c}{4\pi G\rho_c^2} \tag{14.2}$$

14.1 Small Oscillations: Deriving the Equations

In order to derive the equations which govern oscillations in a polytrope, we follow the discussion first given (in the midst of World War II) by Cowling (1941). Let the material at any point in the polytrope undergo a vector displacement **h**. This vector displacement is in a general direction, but the component of **h** along the radial direction is of particular interest: we refer to this radial component as R. The displacement is accompanied by localized perturbations in density, pressure, and gravitational acceleration: we refer to these as $\delta\rho$, δp, and δg.

Our goal in this section is to derive equations which will allow us to solve for the quantities which describe the oscillations: R, δp, and $\delta\rho$.

The perturbations induced in the star by the displacement **h** have the effect that the equation of HSE is no longer satisfied. The imbalance of pressure and gravity forces leads to an acceleration which is given by the equation for *conservation of momentum*

$$\rho\frac{d^2\mathbf{h}}{dt^2} = -\nabla p - \rho g \tag{14.3}$$

where the spatial gradient operator ∇ includes a component along the radial (outward) direction. Equation 14.3 is a more general (3-dimensional) form of the 1-D Equation 7.1, where only *vertical* motions were being considered. Instead of containing only the vertical velocity V, Equation 14.3 includes the 3-D velocity $\mathbf{v} = d\mathbf{h}/dt$. Moreover, in Equation 7.1, the pressure gradient involves d/dz, where z increases *inward*, whereas in Equation 14.3, the operator ∇ involves d/dr, where r increases *outward*: this accounts for the difference in sign in the first term on the right-hand side. Why do we need to consider more than 1-D (radial) motions in the present chapter? Because the oscillations in the Sun are *not* confined to the radial direction: most of the oscillations are actually *non-radial* in nature. This is made explicit in the title of Cowling's (1941) article.

Suppose that the displacement **h** is periodic, with a time dependence $e^{i\omega t}$, where $i = \sqrt{-1}$ and the angular frequency ω is related to the frequency ν (used in Chapter 13) by $\omega = 2\pi\nu$. Then the left-hand side of Equation 14.3 becomes $-\rho\omega^2\mathbf{h}$. Retaining only terms which are of first order in the perturbation amplitude, the right-hand side of Equation 14.3 becomes

$$-\nabla\delta p - g\delta\rho - \rho\delta g$$

In order to keep the discussion as simple as possible, but retain the essential physics of oscillation, we now invoke what is called the "Cowling approximation": we neglect changes in the gravity, i.e., we set $\delta g = 0$. Why is it plausible to neglect changes in the gravitational acceleration? Because the mass in a star is concentrated toward the center: the central density is much larger than the density in the outer regions. Now, the oscillations we consider here consist of motions which have maximum amplitudes in the outer regions of the star,

where the eignfunctions reach their maximum amplitudes (see Figures 14.2 and 14.3). As a result, the mass interior to a certain point remains almost unchanged by the slight changes associated with oscillations. This allows us to assume $\delta g = 0$ as a reasonable simplifying approximation.

The small oscillations occur as perturbations in a medium (the polytrope) which is in HSE. This allows us to write $g = -\nabla p/\rho$, and so we can rewrite Equation 14.3 in the form

$$-\rho\omega^2\mathbf{h} = -\nabla\delta p + (\delta\rho/\rho)\nabla p \tag{14.4}$$

Turning now to the *conservation of mass*, the equation of continuity $\partial\rho/\partial t + \nabla\cdot(\rho\mathbf{v}) = 0$ can be written, to first order in the perturbations, as $\delta\rho = -\nabla\cdot(\rho\mathbf{h})$

We can now eliminate $\mathbf{h}$ from Equation 14.4 by taking the divergence of both sides:

$$\omega^2\delta\rho = -\nabla^2(\delta p) + \nabla\cdot\left[\frac{\delta\rho}{\rho}\nabla p\right] \tag{14.5}$$

In a spherical object, it is natural to separate the oscillations into two components: one depends only on the angular coordinates, and the second is a function of the radial coordinate r. The Laplacian operator in Equation 14.5 contains a radial component and an angular component. The latter can be taken to be a spherical harmonic, Y_{lm}. Here, we ignore the m (longitudinal) subscript, and consider only the latitudinal variations, which are characterized by l, the degree of the mode. In this case, the angular part of the Laplacian in Equation 14.5 is replaced by $-l(l+1)\delta p/r^2$ where l is the degree of the mode.

In spherical coordinates, the expressions for the radial components of Laplacian and divergence lead to the following form for Equation 14.5:

$$\omega^2\delta\rho = \frac{l(l+1)}{r^2}\delta p - \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\delta p}{\partial r} - r^2\frac{\delta\rho}{\rho}\frac{\partial p}{\partial r}\right) \tag{14.6}$$

So far we have considered the generalized displacement $\mathbf{h}$. Now let us confine our attention to the radial component of the displacement R. The radial component of Equation 14.4 gives us the following expression for R:

$$\rho\omega^2 R = \frac{\partial}{\partial r}(\delta p) - \frac{\delta\rho}{\rho}\frac{\partial p}{\partial r} \tag{14.7}$$

This gives us the first equation of three which relate the three unknowns R, δp, and $\delta\rho$.

Substituting Equation 14.7 in Equation 14.6, we find a second equation which relates the three unknowns:

$$\omega^2\delta\rho = \frac{l(l+1)}{r^2}\delta p - \frac{1}{r^2}\frac{\partial}{\partial r}\left(\rho\omega^2 r^2 R\right) \tag{14.8}$$

Now that we have eliminated the angular dependences, there is only one remaining independent variable: the radial coordinate. As a result, we can safely replace the partial derivative $(\partial/\partial r)$ by the total derivative (d/dr).

Equations 14.7 and 14.8 are two equations which relate the displacement of the fluid element in the radial direction to the perturbations in pressure and density. A third equation is needed if we are to solve for the three unknowns. Having already used the equations which describe conservation of *mass* and of *momentum*, we now turn to the equation for *energy* conservation.

The particular form of the energy equation which we use in this first course in solar physics is the following: the oscillations are assumed to be *adiabatic*. That is, when the oscillations occur, the total pressure variation Δp is related to the total density variation $\Delta\rho$ by an adiabatic relation: $\Delta p/p = \gamma\Delta\rho/\rho$. Here, γ is the adiabatic exponent. In a notation which is analogous to the definition of the polytropic index (see Chapter 10, Equation 10.1), Cowling (1941) writes γ in the form $\gamma = 1 + (1/N)$. For a typical adiabatic index $\gamma = 5/3$, the value of N is 1.5.

What are the total changes in pressure which occur as a result of the oscillation? First, there is δp itself. However, since the oscillating element of fluid has also moved a radial distance R, the fluid element finds itself at a radial location where the ambient pressure is *different* from the value it had in the unperturbed location. Thus, the total change in pressure Δp associated with the oscillation is the sum of two terms: $\Delta p = \delta p + R(dp/dr)$. An equivalent sum of terms applies also to the total change in density. Then the adiabatic version of the energy equation can be written:

$$\frac{\delta p + R(dp/dr)}{p} = \left(1 + \frac{1}{N}\right)\left(\frac{\delta\rho + R(d\rho/dr)}{\rho}\right) \tag{14.9}$$

We now have three equations for three unknowns: Equations 14.7 through 14.9.

14.2 Conversion to Dimensionless Variables

To help cast the equations into more convenient form, we introduce some dimensionless variables. In this process, we are guided by the choices which were made in Chapter 10 in connection with dimensionless variables in a polytrope of order n.

First, we change from the dimensional frequency ω to the dimensionless quantity α according to the definition:

$$\alpha = \frac{\omega^2(1+n)}{4\pi G\rho_c} \tag{14.10}$$

The fact that α is dimensionless can be verified by recalling the definition of the Emden unit of length r_o (see Equation 10.9). Note that the (dimensional) frequency ω scales as $\sqrt{\alpha}$.

Second, we reduce the radial displacement R to dimensionless form by normalizing to the Emden unit of length: $X = R/r_o$. Analogously, we express the radial coordinate r as a new dimensionless variable $x = r/r_o$. We reduce the pressure and density perturbations to dimensionless forms by normalizing to their central values: $\theta = \delta p/p_c$ and $\eta = \delta\rho/\rho_c$.

In terms of these dimensionless variables, we can convert Equations 14.7 through 14.9 into dimensionless form. So far, the derivation is quite general, and could be applied to any particular star in order to solve for the three unknowns.

But now, following Cowling (1941), and in the spirit of Chapter 10 above, we restrict our attention to the case of a "polytropic star." Specifically, we now apply our three equations (in dimensionless form) to a polytrope where the density profile is given by $\rho = \rho_c y^n$.

Then we find that Equation 14.8 becomes

$$\alpha\eta = \frac{l(l+1)}{x^2}\theta - \frac{1}{x^2}\frac{d}{dx}\left(\alpha x^2 y^n X\right) \tag{14.11}$$

Equation 14.11 shows, in dimensionless form, how the density perturbation η is related to the pressure perturbation θ and to the radial gradient of the radial displacement X.

In dimensionless form, Equation 14.7 becomes

$$\alpha y^n X = \frac{d\theta}{dx} - (1+n)\eta y' \tag{14.12}$$

where y' denotes the spatial gradient dy/dx.

In dimensionless form, Equation 14.9 can be written

$$y\eta(1+N) = N\theta - (n-N)Xy^n y' \tag{14.13}$$

Using Equation 14.13 to obtain an expression for η, we can eliminate η from Equations 14.11 and 14.12 and obtain two equations for two unknowns, the radial displacement X and the pressure perturbation θ. These two equations describe the oscillations which occur inside a polytrope.

Substituting the expression for η into Equation 14.12, and gathering terms in θ on the left-hand side, we find

$$\frac{d\theta}{dx} - \frac{(1+n)Ny'}{(1+N)y}\theta = X\left(y^n\alpha - \frac{(1+n)(n-N)}{1+N}y^{n-1}y'^2\right) \tag{14.14}$$

Substituting the expression for η into Equation 14.11, and gathering terms in X on the left-hand side, we find

$$\frac{1}{x^2}\frac{d}{dx}\left(x^2y^nX\right) - \frac{(n-N)}{1+N}y^{n-1}y'X = \theta\left[\frac{l(l+1)}{\alpha x^2} - \frac{N}{(1+N)y}\right] \tag{14.15}$$

The Equations 14.14 and 14.15 are coupled: the radial gradient of one variable is expressed in terms of the value of the other variable.

It is convenient to define auxiliary variables, one for the pressure fluctuation θ, the other for the radial displacement X. In order to see which auxiliary variable is most helpful, we note that the left-hand side of Equation 14.14 can be written in the form $\theta' - (Ey'/y)\theta$ where $E = N(1+n)/(1+N)$ is a numerical constant. This form of a differential equation suggests an integrating factor y^{-E}. This leads us to convert the pressure perturbation variable θ and the radial displacement variable X to new auxiliary variables:

$$w = \theta y^{-E}, \quad z = X y^{E} \tag{14.16}$$

The two variables $w = w(x)$ and $z = z(x)$ describe how the pressure perturbation and radial displacement vary as a function of radial location in the polytrope. In terms of these two variables, and also introducing the constant $Q = 2E - n$, we finally arrive at two ordinary differential equations for w and z as functions of the radial coordinate:

$$\frac{dw}{dx} = z y^{-Q} \left(\alpha - \frac{(1+n)(n-N)}{1+N} \frac{y'^2}{y} \right) \tag{14.17}$$

$$\frac{d}{dx}\left(x^2 z \right) = w y^{Q} \left(\frac{l(l+1)}{\alpha} - \frac{N x^2}{(1+N)y} \right) \tag{14.18}$$

14.3 Overview of the Equations

Let us summarize what we have done up to this point. Equations 14.17 and 14.18 describe the profiles of radial displacements ($\sim z(x)$) and fluctuations in pressure ($\sim w(x)$) which occur when a "star" oscillates with a particular frequency ($\sim\sqrt{\alpha}$). The "star" extends in radial coordinates from $x = 0$ (the center) to $x = x_1$ (the surface). Inside the star, the function $y(\sim T)$ varies from $y = 1$ at the center to $y = 0$ at the surface. The radial gradient y' (which appears on the right-hand side of Equation 14.17 as a squared quantity) is negative throughout the star. The star obeys a polytrope equation of state: $p \sim \rho^{1+1/n}$ and the oscillations are adiabatic: $\Delta p/p = (1 + 1/N)\Delta\rho/\rho$. The oscillations vary in latitude such that l nodes exist between the North and the South poles of the star.

For any chosen polytrope (with index n), the Lane-Emden equation can be integrated (either analytically or numerically) to obtain a table of values of y and y' as a function of x between 0 and x_1. (The latter is set by the choice of the polytropic index n). A value is assumed for N (typically 1.5), and this then fixes the value of Q. Then for each assigned value of $l(= 0, 1, 2, 3, \ldots)$, the quantities inside the large brackets on the right-hand side of Equations 14.17 and 14.18, as well as the y^Q terms, can be evaluated at all tabulated values

of x between 0 and x_1. This provides the information required to undertake a numerical integration for the two unknowns $w(x)$ and $z(x)$.

The properties of oscillations in a polytrope can be determined by integrating Equations 14.17 and 14.18 numerically for $w(x)$ and $z(x)$ with appropriate boundary conditions at the center and at the surface. Numerical integration can be performed either by programming a Runge-Kutta subroutine, or by using one of the widely available software packages such as MATLAB® or Mathematica.

In order to begin the integration at the center of the star, asymptotic functional forms for w and z must be specified: see step 6 in Section 14.4.1 below. The boundary conditions at the surface of the "star" for an eigenmode are that $w \to 0$ and $z \to 0$ as $x \to x_1$.

By experimenting with different choices of polytropic index n, one can learn a great deal about the properties of oscillations in stars.

The properties of Equations 14.17 and 14.18 are such that, in asymptotic terms, there are two distinct classes of eigenmodes: one is relevant in the limit $\alpha \to \infty$, the second is relevant in the limit $\alpha \to 0$ (see Chapter 14, Section 14.6). In the limit of high frequency ($\alpha \to \infty$), pressure dominates as the restoring force: such modes are referred to as p-modes. The p-modes exist in all polytropes. In the limit of long period ($\alpha \to 0$), gravity dominates as the restoring force: these are g-modes. Unlike the p-modes, g-modes do *not* exist in all polytropes: because of the presence of the term $(n - N)$ in Equation 14.17, g-modes with finite periods do not exist if the polytropic index n is less than the value that is assumed for N (typically $N = 1.5$).

14.4 The Simplest Exercise: Solutions for the Polytrope $n = 1$

In order to get a feel for how the oscillation equations work, and how they lead to eigenfrequencies, it is instructive to integrate Equations 14.17 and 14.18 numerically for the case of a particularly simple polytrope, namely, $n = 1$. Although that polytrope makes no claim to describe any actual star, it still retains the overall structure of high pressure and density at the center, and much lower pressure and density at the surface. For present purposes, the outstanding advantage of the polytrope $n = 1$ is that the functions $y(x)$ and $y'(x)$ are known analytically: $y(x) = \sin(x)/x$, $y'(x) = \cos(x)/x - \sin(x)/x^2$. As a result, we do not need to prepare a table of values of y and y': the local values can be calculated analytically. The surface of the star occurs at $x_1 = 3.14159$. We assume $N = 1.5$. This leads to $E = 1.2$ and $Q = 1.4$.

To perform the calculation, one must first specify a certain value of α: this remains fixed throughout the "star." Start at the center of the star ($x = 0$), and increase the value of x by some chosen increment Δx. Integrate the

coupled Equations 14.17 and 14.18 for the two unknowns w and z at each step. Because the two equations are coupled, one integrates outward first in (say) w using the current value of z: given the current value of z, the right-hand side of Equation 14.17 can be evaluated, and this therefore allows one to evaluate dw/dx. Knowing this, one can take a step Δx and use one's numerical scheme (e.g., Runge-Kutta) to calculate an updated value of w. This is then inserted in the right-hand side of Equation 14.18 in order to evaluate dz/dx. This then allows the numerical scheme to calculate an updated value of z across the step Δx. This process is repeated for all tabulated values of x between 0 and 3.14159. The result is a table of values of w and z as a function of x.

Of special interest is the value of $w(x)$ at the star's surface ($x = x_1 = 3.14159$). In most cases, for arbitrary values of α, the value of $w(x_1)$ will be found to be nonzero. But in certain special cases, when α takes on certain values, one finds that when the calculation reaches the surface of the polytrope, the computed value of $w(x_1)$ equals zero. Those special cases are the eigenmodes of the polytrope.

14.4.1 Procedure for computation

1. Pick a value of l among the set 0, 1, 2, and 3.

2. Pick a starting guess for α, the (dimensionless) frequency. Because of the choice of normalizations, the starting guess for α should not be too far from unity. A recommended starting guess is $\alpha = 0.1$. The reason for this choice is as follows: for $l = 0, 1, 2$, and 3, the lowest eigenfrequency in the $n = 1$ polytrope is found to have the numerical value $\alpha \approx 1, 0.3, 0.55$, and 0.7, respectively.

3. Start the integration near the center of the polytrope by setting $x = 0.01$ (or $x = 0.1$ if you are confident about your equation solver routine).

4. At that value of x, evaluate the local values of $y(x) = \sin(x)/x$ and $y'(x) = \cos(x)/x - \sin(x)/x^2$.

5. With choices now made for $l, \alpha, n(=1)$, and $N(=1.5)$, you have all the information you need to compute the local numerical values of the expressions in large brackets on the right-hand side of Equations 14.17 and 14.18.

6. You will have to choose initial values for w and z. What initial values should you use for w and z? When one examines the asymptotic properties of Equations 14.17 and 14.18, it turns out that in the limit $x \to 0$, the functional form of z is as follows:

$$z(x) = x^{l-1} \tag{14.19}$$

for all values of l. Also in the limit $x \to 0$, the functional form of w is

$$w(x) = \frac{\alpha x^l}{l} \tag{14.20}$$

for $l > 0$, and $w = 1$ for $l = 0$.

Using these, evaluate w and z at whatever value of x you have chosen as the starting point.

7. Increase x by $\Delta x = 0.01$ (if that is your choice of step size). Using the values of w and z from step (6), and recalculating the local values of y and y' at the new value of x, evaluate the right-hand side of Equations 14.17 and 14.18. With the new numerical values for the derivatives, step forward to calculate the new values of w and z.

8. Repeat step (7) until x has a value that is slightly smaller than $x_1 = 3.14159$. Where should you stop the integration in x? That is, how close should you approach the limiting value $x_1 = 3.14159$? You cannot go too close, because then the terms in $1/y$ on the right-hand side of both equations will become infinitely large. One possible approach is to stop the integration in x when the value of $y(x)$ has decreased to a "small" value, such as 0.001. (Recall that starting at the center, y has a value of 1.0.) For purposes of the oscillation calculation, this stopping point may be considered to be "the surface" of the polytrope.

9. Once you reach this "surface," your code will give you a certain value for the pressure fluctuation variable w. Call this w(surf), and enter this value into a table alongside the value you specified for α (in step 2).

10. Pick a new, *larger* value for α. How large should the new value of α be? Recommended increases are 0.1 up to $\alpha = 5$. That is $\alpha = 0.2, 0.3, 0.4, \ldots, 4.9, 5.0$. Then increase the increment to 0.5 for values of α between 5 and 20. Then use increments of three for α up to (about) 300. Alternatively, one could choose constant steps in $\log(\alpha)$.

11. For each value of α, repeat steps (3)–(9). For each α, tabulate the value you compute for w(surf). Since α is proportional to the square of the frequency (see Equation 14.10), it is more convenient to convert from α to a dimensional frequency using the unit $\nu_g \approx 100$ μHz (Chapter 1, Section 1.10). For a polytrope with index n, the conversion factor for a "star" with mass and radius equal to the solar values is (Mullan and Ulrich, 1988) $\nu = \nu_g\sqrt{(3C_c\alpha/(n+1))}$ where C_c is the central condensation of the polytrope (Chapter 10, Section 10.9). In the present case, $n = 1$, this leads to $\nu = 222.0\sqrt{\alpha}$ μHz.

12. Once you have computed results for all values of α from 0.1 to about 300, plot w(surf) versus α, or (more conveniently) w(surf) versus ν. Two such plots are shown in Figure 14.1, one for $l = 2$ (solid curve) and the

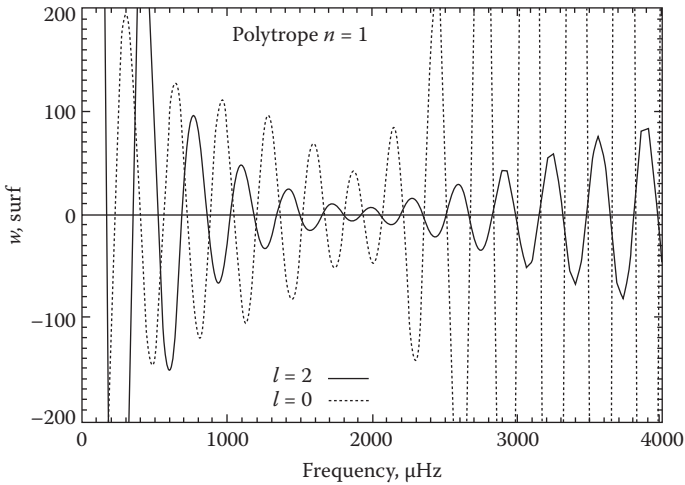


FIGURE 14.1: Surface value of oscillation variable w (the pressure fluctuation) as a function of frequency for polytrope $n=1$ and for two values of the degree l. Crossings of the horizontal line $w(\text{surf})=0$ identify the eigenfrequencies.

other for $l = 0$ (dotted curve). One sees that, as the frequency increases, w(surf) swings back and forth between positive and negative values. For each value of l, there exists a discrete set of values of ν (i.e., ν_1, ν_2, $\nu_3 \ldots$) at which the computed curves cross the horizontal axis, i.e., w(surf) passes through a value of zero. At such values of ν, the pressure fluctuation in the oscillation goes to zero at the surface of the polytrope. Such an oscillation is an eigenmode of the polytrope for the particular value of l that was chosen for the calculation. The sequence of values ν_i are eigenfrequencies of the polytrope.

For your chosen value of l, you will now have a series of eigenfrequencies for the $n = 1$ polytrope.

13. Pick a new value of l and repeat steps (2)–(12). For each value of l, you will find a different series of eigenfrequencies.

14.4.2 Comments on the results: patterns in the eigenfrequencies

Our results for the Cowling approximation in a polytrope, although greatly simplified, nevertheless allow us to draw conclusions which would remain valid if we had applied the (more complicated) full oscillation equations to a detailed solar model. Because of this, it is instructive to examine certain properties of the output of the polytrope oscillation program.

To begin with, we note that at the highest frequencies plotted in Figure 14.1, the last four eigenfrequencies for $l = 0$ are found to be at $\nu = 3490$, $3653, 3816$, and 3980 μHz. For $l = 2$, the highest four eigenfrequencies are found to be at $\nu = 3474, 3637, 3801$, and 3965 μHz. Two patterns are striking here.

First, for modes with a given degree l, the intervals between adjacent eigenfrequencies are $\Delta\nu = 163, 163$, and 164 μHz for $l = 0$, and $\Delta\nu = 163$, 164, and 164 μHz for $l = 2$. That is, both sequences of modes show a striking asymptotic behavior: there is a (roughly) constant *frequency* interval between adjacent eigenmodes. The asymptotic frequency separation we have found here is slightly larger than $\Delta\nu (= 153$ μHz) obtained for the $n = 1$ polytrope from a precise calculation (Mullan and Ulrich, 1988): this difference can be ascribed to inadequacies in treating the surface boundary conditions in the solutions of the equations presented in the above figure.

Although the present results pertain to the $n = 1$ polytrope, it will be shown (Chapter 14, Section 14.6) that a constant *frequency* interval between adjacent p-modes is predicted to be a *general property* of the oscillation equations *in the limit of high frequencies*. For reasons which will soon become clear, we refer to $\Delta\nu$ as the "large separation" between adjacent modes. In the Sun, observations indicate that the "large separation" $\Delta\nu$ has values of 135–136 μHz for modes with $l = 0-3$ (Appourchaux et al., 1998). Clearly, this observed separation is smaller than the 153–163 μHz that we have found for the $n = 1$ polytrope. But this is not a matter for any great concern: we have never claimed that the $n = 1$ polytrope is supposed to be an accurate representation of *the Sun*. The point is, there *is* a "large separation" for the modes in the "real Sun," just as we have discovered in our model for the $n = 1$ modes.

Second, we note that the sequences of eigenfrequencies for $l = 0$ and $l = 2$ pair up with each other such that the two curves in Figure 14.1 cross the horizontal axis at *almost* the same frequencies. This indicates that the eigenfrequencies for certain modes with $l = 0$ and $l = 2$ differ from each other by an amount that is small compared to the "large separation" of either sequence. The frequency separations between corresponding $l = 0$ and $l = 2$ modes are $\delta\nu(0-2) = 16, 16, 15$, and 15 μHz, i.e., about one order of magnitude smaller than the numerical values of $\Delta\nu$. The frequency differences $\delta\nu(0-2)$ are referred to as "small separations," to distinguish them from the "large separations" ($\Delta\nu$) between adjacent modes at constant l.

Although we do not present the results graphically here, we note that when the analog of Figure 14.1 is plotted for $l = 1$ and $l = 3$, results similar to those in Figure 14.1 emerge. There is again an asymptotic "large separation" $\Delta\nu$ of about 163 μHz in frequency between adjacent modes with the same l value, and a "small separation" between corresponding $l = 1$ and $l = 3$ modes. In this case, the "small separations" turn out to be $\delta\nu(1-3) = 10 - 12$ μHz.

In the Sun, empirically it is found that the "large separations" are 135–136 μHz, while the "small separations" for modes with frequencies of 3500–4000 μHz are $\delta\nu(0-2) \approx 10$ μHz and $\delta\nu(1-3) \approx 12$ μHz (Appourchaux

et al., 1998). We notice that the numerical values of the "small separations" are about one order of magnitude smaller than the numerical values of the "large separations." Of course there is no reason to expect the $n = 1$ polytrope to reproduce the structure of the Sun in detail. Nevertheless, it is encouraging that even with the simplification of using the $n = 1$ polytrope, and also using the simplification of the Cowling approximation, we recover some salient features of the Sun's eigenfrequencies.

Finally, there is one further piece of information that we need in order to interpret Figures 13.1 and 13.2. It is this: given the frequencies of two adjacent modes with $l = 0$ (say the modes observed at $\nu = 3034$ and 3169 μHz: see Chapter 13, Section 13.2), it is observed that there exists a mode with $l = 1$ with a frequency which is *almost exactly half-way* between the two adjacent $l = 0$ modes. Thus, the ($l = 1$, $n_r = 21$) mode is observed at $\nu = 3099$ μHz, only 0.1% away from the mid-point frequency of the two surrounding $l = 0$ modes. Now, modes with $l = 0$ are excited in the Sun with almost equal power to those with $l = 1$. Therefore, rather than seeing in Figures 13.1 and 13.2 separations between peaks of 135 μHz (the "large separation"), one sees separations of only about one-half that value, i.e., about 0.07 mHz.

14.4.3 Eigenfunctions

Now that frequencies of the eigenmodes have been identified, it is also important to consider the structure of the radial eigenfunctions. Two examples are shown in Figure 14.2, where we plot the radial profile of the function w (corresponding to the pressure perturbation) in Equations 14.17 and 14.18. The abscissa in Figure 14.2 is the radial coordinate x in Equations 14.17 and 14.18: it runs from 0 at the center of the "star" to the boundary value $x = x_1 = \pi$ which is appropriate for the $n = 1$ polytrope.

Two features are noticeable about the eigenfunctions in Figure 14.2. First, the numerical values have excursions on both sides of the $w = 0$ (horizontal) axis. As a result, there exist a finite number of "nodes" where the eigenfunction passes through the value of zero. The number n_r of times that an eigenmode crosses the $w = 0$ axis between center and surface is used to label the mode as being of "radial order n_r."

Second, as we approach the surface, the excursions of the eigenfunction increase to larger (absolute) values. The peaks in the eigenfunction can be considered as "antinodes" where the pressure fluctuation has a local maximum. The antinode which occurs nearest to the surface has a larger amplitude than those which lie somewhat deeper. (This is not always true for some of the very lowest order l modes, such as $l = 0$, but for moderate and high l values, the antinode nearest the surface has the largest amplitude.)

A question that is of particular interest in solar physics is the following: at what radial location is the largest antinode of any given eigenfunction situated? The answer to this question has a bearing on the basic question: why are only certain p-modes excited in the Sun? For the two examples in Figure 14.2,

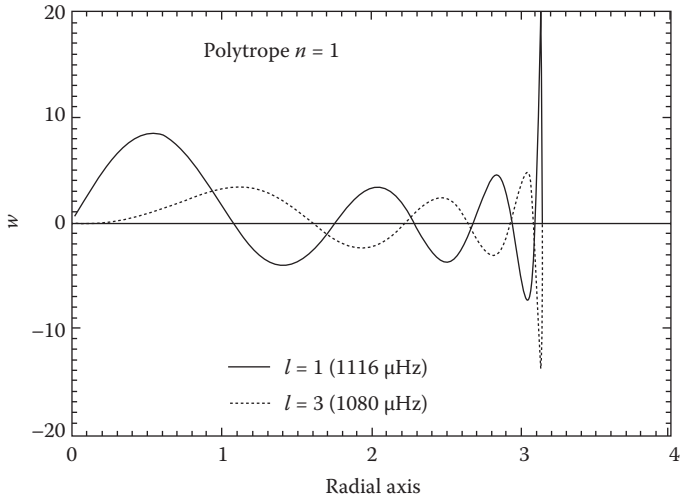


FIGURE 14.2: Pressure eigenfunctions of two modes with similar frequencies, differing in degree l by $+2$, and differing in radial order by -1. The radial coordinate in the abscissa extends from center to surface.

the radial locations at which the largest antinode occurs are located at about 0.99 times the radius of the $n = 1$ polytrope. That is, the largest antinodes occur at depths $z_{\rm an}$ which are close to the surface, no more than 1% of the stellar radius below the surface. The larger the numerical value of n_r, the closer the last (and largest) antinode lies to the surface (for a given l value). In order to demonstrate this result in more detail, we show in Figure 14.3 some details of eigenfunctions which do *not* refer to a polytropic model, but which instead were obtained from *a realistic solar model*. For the three modes shown, with $n_r = 10, 15$, and 25 (and $\nu = 1610, 2290$, and 3650 μHz, respectively), we see that the largest antinodes lie at fractional depths of 0.5%, 0.2%, and $< 0.1\%$ of $R_\odot$, respectively. These depths will be important subsequently when we consider why certain p-modes are excited in the Sun more effectively than others.

The eigenfunctions in Figure 14.3 were computed theoretically using an oscillation code: they show the scaled amplitudes of the eigenfunctions. What is *not* shown in the figure is how much power the Sun actually pumps into the various modes. In fact, there are striking differences in the levels of power which are observed to occur in the modes whose eigenfunctions are plotted in Figure 14.3. The frequency of the $n_r = 25$ mode in Figure 14.3 ($\nu = 3650$ μHz) is such that the mode lies near the peak of power for solar p-modes (see Figure 13.2): the power level is observed to be almost 5000 $\mathrm{m^2\ s^{-2}\ Hz^{-1}}$. On the other hand, the frequency of the $n_r = 10$ mode in Figure 14.3 ($\nu = 1610$ μHz) is such that the mode lies in the barely detectable regime in Figure 13.2: the observed power level is perhaps 10 $\mathrm{m^2\ s^{-2}\ Hz^{-1}}$ above background. Thus, the observations indicate that the ($l = 1$, $n_r = 10$) mode is present in the Sun

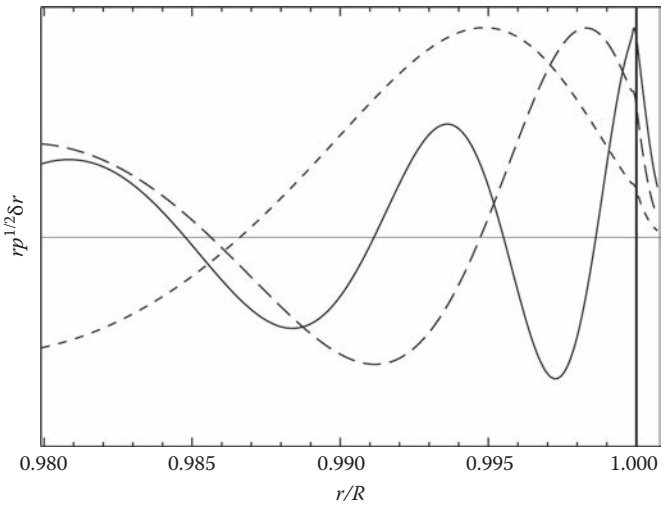


FIGURE 14.3: Eigenfunctions of (scaled) radial displacement δr (analogous to z in Equation 14.18) for $l = 1$ modes in a realistic model of the Sun. The abscissa is radial location expressed in terms of the solar radius R. Notice that the horizontal scale is greatly expanded: only the outermost 2% of the radial coordinate is plotted. Dotted, dashed, and solid curves refer to modes with radial orders $n_r = 10, 15$, and 25, respectively. (From Christensen-Dalsgaard, J. 2003. p. 88 of http://www.phys.au.dk/∼jcd/ oscilnotes/chap-5.pdf. With permission.)

at a power level which is almost three orders of magnitude *smaller* than the power in the ($l = 1$, $n_r = 25$) mode. The intermediate mode ($l = 1$, $n_r = 15$), at $\nu = 2290$ μHz, is present in Figure 13.2 at a power level which is about 10 times smaller than the peak power. We will return to these power levels when we consider the excitation of the modes in Section 14.8 below.

The eigenfunctions in Figure 14.2 belong to two modes which are separated by the "small separation" $\delta\nu(1-3)$. As can be seen from Figure 14.2, each of the two eigenfunctions crosses the $w = 0$ axis several times between the center and the surface. Inspection shows that, of the two eigenfunctions in Figure 14.2, the $l = 1$ curve has *one more node in the radial direction* than the $l = 3$ curve. It is a general relation that modes which differ by *two* units in l are separated by the "small separation" $\delta\nu$ if they also differ by *one* unit in n_r. That is, $\nu(l+2, n_r) \approx \nu(l, n_r + 1)$.

Although not plotted in Figure 14.2, when we plot the radial profile of the *radial displacement z of the oscillations*, the maximum excursions of z are found to be smaller by an order of magnitude or more than the excursions of the pressure fluctuation variable w. The fact that pressure fluctuations are dominant confirms that we are dealing with p-modes.

14.5 What about g-Modes?

Our choice of polytrope $n = 1$ (chosen for the simplicity of its analytic solution) prevents us from discussing g-modes: because of our choice $N = 1.5$, no g-modes with finite periods exist in any polytrope with $n \leq 1.5$. If we wished to numerically model g-modes in a polytrope, we must use $n > 1.5$ (assuming $N = 1.5$). As we have seen (Chapter 10), for polytropes with $n > 1.5$, no analytic formulas exist for the polytrope structure (apart from the uninteresting case of $n = 5$ for an infinitely distended star). Therefore, a study of g-modes in polytropes requires us first to determine numerically a table of values of y and y' as a function of x, and then interpolate in this Table to obtain local values of y and y' at each value of x in the right-hand side of Equations 14.17 and 14.18. Such a study has been reported for the polytropes $n = 2, 2.5, 3, 3.5$, and 4 by Mullan (1989).

Since the case $n = 3.25$ is relevant for the radiative interior of the Sun (see Chapter 10, Section 10.2), we focus on that case here. An abbreviated table of y and y' values as a function of x in the $n = 3.25$ polytrope has already been given (see Table 10.1). A more extended version of that table, including more than 800 rows, was prepared so that it could be used for interpolation in the right-hand side of Equations 14.17 and 14.18. Those equations were then numerically integrated from center to surface, using a series of frequencies appropriate for g-modes, using the steps outlined in Section 14.4.1 above.

There are four alterations to the steps outlined in Section 14.4.1 when we discuss g-modes. First, the range of permissible l does not include $l = 0$: so step (1) should read: "Pick a value of l among the set $1, 2, 3$." Second, in step (4), the local values of y and y' for each value of x cannot be obtained analytically: instead, they must be obtained by interpolating into Table 10.1 (or an extended version thereof). Third, in step (10), one must choose a new, *smaller* value of α: the study of g-modes requires that one goes to smaller and smaller frequencies, i.e., long and longer *periods*. Fourth, in step (11), the conversion factor from α to (dimensional) frequency for $n = 3.25$ is $\nu = 789.0\sqrt{\alpha}$ μHz.

Apart from these alterations, the numerical integration proceeds step-by-step at each frequency as described in Section 14.4.1. The result is again that, at each frequency, the code provides a value for the oscillation variable, w(surf), at the surface of the polytrope.

Analogous to the plot in Figure 14.1 for the p-modes, we present in Figure 14.4 the surface values of w in the polytrope $n = 3.25$ for g-modes with $l = 1$ and $l = 2$. In contrast to Figure 14.1, where we plotted w(surf) as a function of *frequency*, in Figure 14.4, it makes more sense to plot w(surf) for the g-modes as a function of *the period*. The modes we plot have periods between about 8 and 16 hours: this range overlaps with the range that is relevant to observations of oscillations which have been reported in the Sun (Garcia et al., 2007). Once again, the zero points of w(surf) define the locations of eigenmodes. Inspection of Figure 14.4 shows that, for the g-modes,

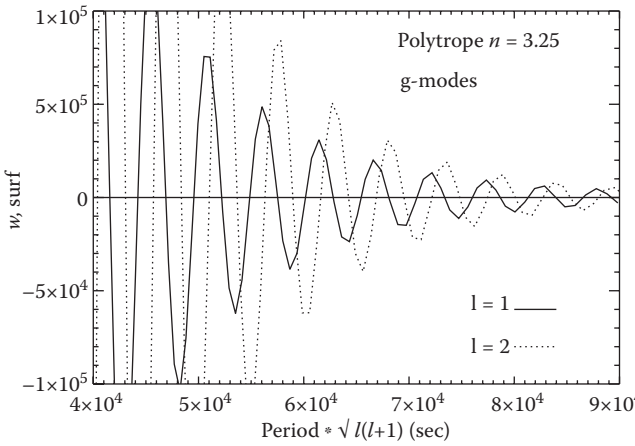


FIGURE 14.4: Calculation of g-modes in the polytrope $n = 3.25$. Plotted is the surface value of w, one of the two oscillation variables, as a function of the mode *period* times $\sqrt{l(l+1)}$ for $l = 1$ and 2. Eigenmodes occur where $w(\text{surf}) = 0$. Note that the separation between adjacent eigenmodes approaches a constant value in *period*. The asymptotic separation in the quantity $P\sqrt{l(l+1)}$ is about 2500 sec. Contrast this plot with Figure 14.1, where the abscissa was in terms of *frequency*.

the separation between adjacent eigenmodes approaches an asymptotic limit *which is constant in period*. In the case of the particular examples that are plotted in Figure 14.4, the period separation between adjacent eigenmodes times $\sqrt{l(l+1)}$ is found to be about 2500 sec. For $l = 1$, this corresponds to a period separation of 1770 sec, i.e., about 30 minutes between adjacent modes. For $l = 2$, this corresponds to a period separation of 1020 sec, i.e., about 17 minutes between adjacent modes.

Thus, whereas p-modes exhibit constant asymptotic separation in *frequency* (with a "large separation" $\Delta\nu$ between the *frequencies* of adjacent modes at high frequency), g-modes exhibit constant asymptotic separation ΔP in *period*. This dramatic distinction in asymptotic behavior between p-modes and g-modes is a striking feature of the oscillation equations.

We turn now to a discussion as to why the p-modes and the g-modes display these distinctly different asymptotic behaviors.

14.6 Asymptotic Behavior of the Oscillation Equations

We can use the mathematical properties of Equations 14.17 and 14.18 to see why p-modes have asymptotically equal spacing in *frequency* between

adjacent modes, whereas g-modes have asymptotically equal spacing in *period*.

14.6.1 p-modes

As regards p-modes, asymptotic behavior emerges in the limit of high frequencies, $\alpha \to \infty$. In this limit, Equation 14.17 reduces to $dw/dx = \alpha z y^{-Q}$ while Equation 14.18 reduces to $d(x^2 z)/dx = -w\varphi(x, y)$ where $\varphi = N x^2 y^{Q-1}/(N+1)$. Let us concentrate on the values of the quantities w and z close to the "surface" of the star, i.e., at a fixed value of x (close to x_1). In essence, this is a rough method of concentrating on the value of w which we referred to as w(surf) when we plotted Figure 14.1. Although this is not a mathematically rigorous procedure, it helps us to see (roughly) that we can write $dz/dx \approx -Aw$ where $A = y^{Q-1}/\Gamma$ is a constant at a fixed value of x. (We use the quantity Γ to replace $(N+1)/N$). The numerical value of A depends on the (almost zero) value of y close to x_1. Again treating y^{-Q} as essentially constant (because we are treating a point at a fixed value of x close to the surface), we take the second derivative of w and find $d^2w/dx^2 \approx \alpha y^{-Q} dz/dx$. Substituting the above expression for dz/dx, we find

$$\frac{d^2w}{dx^2} \approx -w\alpha A y^{-Q} \approx -w\left(\frac{\alpha}{\Gamma y}\right) \tag{14.21}$$

Recall that in a polytrope composed of a perfect gas, the Lane-Emden function y is related to the local temperature by $y = T/T_c$ (Chapter 10, Section 10.3). The combination Γy (which occurs on the right-hand side of Equation 14.21) is therefore proportional to the square of the local sound speed: $\Gamma y = A' c_s^2$.

Substituting this in Equation 14.21, and also substituting the expression $\nu = A''\sqrt{\alpha}$ (see Equation 14.10) relating α to the (dimensional) frequency, we can rewrite Equation 14.21 as

$$\frac{d^2w}{dx^2} \approx -w\left(\frac{B\nu}{c_s}\right)^2 \tag{14.22}$$

where $B^2 = 1/A'A''^2$.

The solution of Equation 14.22 is a sinusoidal function: $w \sim sin(xB\nu/c_s)$. Here, w is to be interpreted as w(surf), (i.e., the ordinate in Figure 14.1), while ν is the frequency (i.e., the abscissa in Figure 14.1). The quantity x can be regarded as fixed at the value $x = x_1$. As a result, w(surf) is expected to vary sinusoidally as ν increases. The essence of a sinusoid is that it passes through successive zeroes as the argument $x_1 B\nu/c_s$ passes through the set of discrete values $j\pi$ where j is an integer. The interval between successive zeroes of the sinusoid corresponds to increments of π between successive eigenfrequencies multiplied by $x_1 B/c_s$. Thus, in this limit, adjacent modes differ in *frequency* by a constant amount $\Delta\nu = \pi c_s/Bx_1$.

Note that $\Delta\nu \sim c_s/x_1$. Now, the combination x_1/c_s is related to t_s, the time for sound to propagate from the Sun's center to the surface (Chapter 9, Equation 9.6). Thus, $\Delta\nu$ is proportional to $1/t_s$. Detailed mathematical work shows that in fact, the asymptotic frequency separation $\Delta\nu$ should equal $1/(2t_s)$.

14.6.2 g-modes

As regards g-modes, the treatment of asymptotic behavior follows the above discussion for p-modes, except that now we consider the limit $\alpha \to 0$. In this limit, the dominant term in dz/dx is $l(l+1)/\alpha$, while the dominant term in dw/dx is proportional to $(n-N)$. Repeating the above steps, we again find a sinusoidal solution, with zeroes separated by a constant interval π in the argument. In this case, adjacent modes differ by a constant value of the argument $A_g = (n-N)\sqrt{l(l+1)}/\nu$.

There are two features to be noted about A_g. First, the presence of $n-N$ has the effect that sinusoidal solutions exist only for $n > N$: in a polytrope where $n \leq N$, the solutions are no longer propagating waves, but are instead damped exponentials.

Second, A_g includes the frequency in the *denominator*. Thus, A_g is proportional to the *period* of the mode. As a result, the constant interval between neighboring g-modes is proportional to the *period*. Thus adjacent g-modes (with fixed l) are separated by a constant interval $P_o/\sqrt{l(l+1)}$ in the *period*. The quantity P_o is predicted to have a well-defined value for a given value of the polytropic index: e.g., for $n=3$, $P_o=3497$ sec, while for $n=3.5$, $P_o=1927$ sec (see Mullan, 1989). Thus, for modes with degree $l=1$, the asymptotic separation in period between adjacent modes is predicted to be 41 minutes for $n=3$, and 23 minutes for $n=3.5$. As can be seen from Figure 14.4, for the intermediate case $n=3.25$, the asymptotic separation for $l=1$ modes is about 30 minutes.

The feature of a constant separation in *period* between adjacent g-modes may aid observers in identifying such modes in the Sun. In fact, precisely this technique was used by Garcia et al. (2007) to identify modes in the Sun with periods between 1 and 15 hours. These modes exhibit asymptotic separations of about 9, 13, and 24 minutes with $l = 3, 2$, and 1, respectively. The discovery of 24-minute asymptotic separations between adjacent g-modes with $l = 1$ lies between the predicted separations for the $l = 1$ g-modes in the polytropes $n = 3.5$ and 3.25. We note that in Chapter 10, Section 10.2, for independent reasons, "the structure of the Sun in the radiative interior corresponds to an effective polytropic index of 3.25." Now the g-modes can propagate only in the radiative interior of the Sun (where $n > 1.5$): they are evanescent in the convection zone (where $n = 1.5$). This explains why their amplitudes at the surface are so small as to make them difficult to identify. Therefore, observational properties of the g-modes should be determined by the physical structure in the radiative interior. The asymptotic separations in period reported by Garcia et al. can be fitted with the results from a polytrope $n = 3.25 - 3.5$.

It is remarkable that a structure as simple as a polytrope, and one with an index close to what is expected for completely independent reasons in the radiative interior of the Sun, can yield asymptotic g-mode separations which are consistent with the observed separations.

In contrast to the asymptotic behavior of p-modes (where the asymptotic interval in *frequency* is determined by the radial profile of the *sound speed*), it is not surprising that for g-modes, the asymptotic interval in *period* between adjacent g-modes is *not* determined by the sound speed: instead, it is determined by the radial profile of a very different physical quantity known as the Brunt–Vaisala frequency ($\nu_{BV} \sim n - N$).

In the solar convection zone, where $n = 1.5$, and therefore $n = N$, g-modes are exponentially damped.

14.7 Depth of Penetration of p-modes beneath the Surface of the Sun

Now that we know that p-modes are associated with the propagation of *sound* waves, we can obtain a valuable piece of information about the following question: how deeply into the Sun do p-modes with a particular l penetrate? To see this, we note that when sound waves propagate in a medium where the local sound speed is c_s, the wave frequency ω is related to the wavelength λ by a dispersion relation: $\omega^2 = k^2 c_s^2$. Here, the wavenumber k is defined by $k = 2\pi/\lambda$. Let us examine k^2 in more detail.

We have already (see Chapter 13, Section 13.2) introduced the concept of "horizontal wavelength" $\lambda_h = 2\pi R_\odot/\sqrt{[l(l+1)]}$ in connection with modes of angular degree l: this wavelength is a measure of how many nodes exist along the meridian from north to south pole. Associated with λ_h, we introduce the "horizontal wave number" k_h defined by $k_h = 2\pi/\lambda_h = \sqrt{[l(l+1)]}/R_\odot$.

A second component of the wave number which enters into the solar p-modes is associated with the wavelength in the radial direction λ_r. The "radial wave number" is defined by $k_r = 2\pi/\lambda_r$, such that $k^2 = k_h^2 + k_r^2$.

The dispersion relation for sound waves can be written in the form

$$k_h^2 + k_r^2 = \frac{\omega^2}{c_s^2} \tag{14.23}$$

Let us consider what happens to a sound wave as it penetrates deeper and deeper below the surface of the Sun. At the surface, the degree of the mode is identified in terms of the angular degree l: this fixes the value of k_h, and the sound-wave associated with the mode of degree l retains that value of k_h at all depths.

With increasing depth below the surface, the ambient temperature T increases, leading to an *increase* in $c_s^2 = \gamma R_g T/\mu_a$. As a result, the right-hand

side of Equation 14.23 *decreases* as we go deeper into the Sun. At a certain depth z_r, the right-hand side falls to a value which is equal to k_h^2. At that depth, the only way to satisfy Equation 14.23 is for k_r to become zero. Propagation in the radial direction is no longer permitted. The wave number becomes entirely horizontal. When the wave is near the surface, it has a finite value for both k_h and k_r: such a wave propagates at a certain angle relative to the radial direction. But at depth z_r, the wave propagates horizontally, and can penetrate no deeper into the Sun. Thus, although the wave starts off its journey into the Sun by propagating away from the surface at a finite angle to the radial direction, the wavefront gradually becomes more and more refracted away from the radial direction, until at depth z_r, the wave becomes horizontal, and then begins to refract back toward the surface.

The depth z_r is indicative of the maximum depth to which a wave with a given degree l penetrates into the Sun. We cannot expect that such a wave will contain much (or any) information about what is happening in the deep interior of the Sun, at depths in excess of z_r. If we wish to study conditions at radial locations of (say) r in the deep interior of the Sun, we must make sure to study the properties of waves which *can* propagate into depths z_r which are at least as great as $R_\odot - r$.

Setting $k_r^2 = 0$ in Equation 14.23, and setting $c_s^2 = \gamma R_g T/\mu$, we find that the depth z_r occurs when the local temperature T has the value $T_r = \omega^2\mu/\gamma R_g k_h^2$. At what depth does the temperature have such a value? Well, as long as we are considering depths which are not too far beneath the surface of the Sun, specifically, as long as we consider depths of no more than about 200,000 km (roughly 0.3 $R_\odot$), we know that the solar structure is determined by convective heat transport. In such conditions, the temperature gradient is equal to the adiabatic gradient, g/C_p. This means that the temperature as a function of depth is $T(z) = T(z_o) + g(z - z_o)/C_p$ (see Chapter 7, Equation 7.5). We set $z_o = 0$ where $T(z_o) \approx 6000$ K. Throughout most of the convection zone, $T(z) \gg 6000$ K. Therefore, the depth z_r is essentially equal to $T_r C_p/g$. Recalling that C_p can be set equal to $\gamma R_g/\mu(\gamma-1)$ (see Chapter 6, Equation 6.5), we find

$$z_r = \frac{\omega^2}{k_h^2}\frac{1}{g(\gamma-1)} \tag{14.24}$$

In the limit of large l, and for modes with (linear) frequency $\nu(= \omega/2\pi)$, the depth of penetration can be expressed as a fraction of the solar radius as follows:

$$\frac{z_r}{R_\odot} = \frac{4\pi^2\nu^2 R_\odot}{l^2}\frac{1}{g(\gamma-1)} \tag{14.25}$$

Thus, the larger the degree l, the shallower is the penetration of the mode beneath the surface of the Sun.

As an illustration of Equation 14.25, it is instructive to ask: how large must l be in order to have the depth of penetration no deeper than the con-

vection zone? In such a case, $z_r/R(\text{sun}) \leq 0.3$. Inserting this in Equation 14.25, and using $g = 27422$ cm sec^{-2} and $\gamma = 5/3$, we find that the degree l must exceed a value equal to $l_c \approx 22400\,\nu$. The modes which are most commonly excited in the Sun have $\nu \approx 0.003$ Hz. This leads to $l_c \approx 60 - 70$. This indicates that if we want to use p-modes to study the convection zone in the Sun, it will be best to concentrate on the properties of modes with degree l in excess of 60–70. Referring to Figure 13.4 in Chapter 13, we see that modes which lie on the prominent ridges toward the right-hand side of the figure are all confined within the solar convection zone.

For modes with large l, i.e., for modes which do not penetrate deeply beneath the surface, the eigenfunctions are effectively "squeezed" into a shell in the outer parts of the Sun between a depth of z_r and the surface. For example, observed modes with the largest l in Figure 13.4, with $l \approx 200$, and $\nu = 3$ mHz, are confined to a shell which penetrates beneath the surface to a distance of only about $0.03R_\odot$. And yet there still exist a series of modes, each with its own radial order n_r, which must be "squeezed" into this thin shell. It is obvious that, in such a case, even relatively small values of n_r will result in having the last antinode very close to the surface.

14.8 Why Are Certain Modes Excited More Than Others in the Sun?

In order to understand how the p-modes are excited in the Sun, and why certain p-modes are not excited at all (or at best, only at a very low level), let us recall an important feature of the observed power spectrum in Chapter 13, Figure 13.2. We have seen (Chapter 14, Section 14.4.3) that the solar mode with $l = 1$ and $n_r = 10$ is present at a power level which is 500 times smaller than the mode with $l = 1$ and $n_r = 25$. And for the mode with $l = 1$ and $n_r = 15$, the power level is some ten times smaller than the mode with $l = 1$ and $n_r = 25$.

14.8.1 Depths where p-modes are excited

These results invite comparison with the eigenfunctions which are plotted in Figure 14.3. We note that the largest antinode of the mode with $l = 1$ and $n_r = 10$ lies at a depth of 0.5% of the solar radius, i.e., at a depth of 3500 km below the photosphere. On the other hand, the mode with $l = 1$ and $n_r = 25$ has its largest antinode at depths of <700 km. The ($l = 1$, $n_r = 15$) mode has its largest antinode at an intermediate depth, 1400 km.

These results lead us to ask the question: is there some physical quantity in the Sun which can provide power to the p-modes, and is favorable for excitation of p-modes at depths of <700 km but is less favorable (by a factor

of ten) at depths of 1400 km, and is even less favorable (by factors of 500) at depths of 3500 km?

14.8.2 Properties of convection at the excitation depth

The most obvious characteristic of depths $z = 700$–3500 km is that they lie within the solar convection zone. The principal characteristic of that zone is that the gas is driven effectively to finite velocities by means of convective instability. Convective granules come and go on time-scales of a few minutes. When a compressible medium is in motion, it is inevitably a source of pressure fluctuations, i.e., sound waves. As a result, the solar convection zone is an effective generator of sound waves. This raises the possibility that such waves may serve as a source of p-modes, if conditions are favorable to allow transfer of energy into the modes.

At what location is energy likely to be transferred most effectively into a p-mode? The answer is: at the location where the mode's eigenfunction has its largest antinode. That is where the mode "likes" to have a large pressure fluctuation. This suggests that p-modes can be excited in the Sun most effectively if the largest antinode (i.e., in general, the antinode which lies closest to the photosphere) lies at a depth where convection generates sound waves effectively.

How much power do the convective flows in the Sun emit as sound waves? The maximum available power can be computed by noting that an individual granule survives for only about one turnover time, i.e., for a time interval $t_c \approx D/V$, where D is a length associated with circulation around the convection cell, and V is the convective velocity. When a granule reaches the end of its lifetime, and loses its identity by dissolving back into the ambient medium, it is as if the energy density of the convective flow ($E_d \approx \rho V^2$ ergs cm^{-3}), equivalent to a ram pressure, is made available (over a time-scale of order t_c) as a pressure pulse in the ambient medium. The maximum available power P_p emerging from each cm^3 in the pressure pulse is of order

$$P_p \approx \frac{E_d}{t_c} \approx \frac{\rho V^3}{D} \text{ergs cm}^{-3}\text{sec}^{-1} \tag{14.26}$$

By integrating over the linear extent of the granule ($\sim D$), we find that the maximum available *flux* of pressure F_p from the dissolving granule is $\sim \rho V^3$ergs cm^{-2}sec^{-1}.

Only a fraction of F_p is converted into a flux of sound waves, F_s, with periods in the five-minute range. In a medium where the sound speed is $\approx$ 10 km sec^{-1}, a five-minute sound wave has a wavelength $\lambda \approx 3000$ km, i.e., larger than the linear extent of the granule D. The dissolving granule acts in essence as a "short antenna" for radiating sound waves with wavelength λ. Antenna theory indicates that an antenna of length D is quite inefficient at emitting waves with $\lambda > D$. Specifically, by considering the details of a multipole expansion, it can be shown that the efficiency of emission from a

short antenna is proportional to $(D/\lambda)^{2m+1}$, where $m = 1$ for dipole emission and $m = 2$ for quadrupole emission. It turns out that the sound which is emitted by solar convection is generated mainly by quadrupole terms, i.e., the efficiency of sound emission scales as $(D/\lambda)^5$.

The periods $P_s (= \lambda/c_s)$ of the sound waves emerging from a cell with lifetime t_c are comparable to t_c. As a result, we can write $D/\lambda \approx V/c_s$. This leads to the following expression for the flux of sound F_s

$$F_s \sim F_p \left(\frac{V}{c_s}\right)^5 \sim \rho V^3 M^5 \tag{14.27}$$

where $M = V/c_s$ is the Mach number associated with the convective flows.

In order to proceed with a quantitative discussion, we need to know how the convective velocity V varies with depth beneath the solar surface. Unfortunately, the model of the convection which we computed in Chapter 7 does not contain this information: we made no attempt to compute V because we did not attempt to apply "mixing-length theory" in detail. Instead, we "skipped over" the superadiabatic and ionizing layer and went right to the limit of setting the temperature gradient equal to the adiabatic temperature gradient: by referring to more detailed models, we pointed out (Chapter 7, Section 7.7) that the layer which we "skipped over" has a linear extent of a few Mm. These are precisely the range of depths that we now need to know about in order to discuss excitation of p-modes. Therefore, with the approach we have adopted in this "first course in solar physics," we are not really in a position to provide a quantitative answer to the question "why are certain p-modes excited more than others?"

Rather than leave this important question unanswered, it is worthwhile to refer briefly to one particular solar model in which the depth-dependence of the convective velocity was explicitly calculated. Inspection of that model (Baker and Temesvary, 1966) indicates that at depths of 700, 1400, and 3500 km below the photosphere, the combination of parameters $\rho V^3 M^5 (\sim F_s)$ takes on numerical values of 3.4×10^4, 2.3×10^3, and 24 ergs cm^{-2} sec^{-1}, respectively. That is, at 1400 km, F_s is reduced by a factor of $\approx$10 compared to F_s at 700 km: this could explain why the power observed in the ($l = 1$, $n_r = 15$) p-mode is ten times smaller than the power observed in the ($l = 1$, $n_r = 25$) p-mode. Note also that at 3500 km, F_s is reduced by a factor of $\approx$ 1000 compared to F_s at 700 km: this could explain why the power observed in the ($l = 1$, $n_r = 10$) p-mode is almost 1000 times smaller than the power observed in the ($l = 1$, $n_r = 25$) p-mode.

Finally, note that if we consider modes with large l, where the eigenfunctions are "squeezed" into a thin shell close to the solar surface (see Equation 14.25), even rather small values of n_r may result in the last antinode lying quite close to the surface, i.e., right in the zone where acoustic generation by convection is highly efficient. This explains why, at large values of l, modes with small values of n_r (e.g., $n_r = 4$) can be excited to detectable amplitudes (see Chapter 13, Figure 13.4).

14.9 Using Helioseismology to Test a Solar Model

Now that we have computed a solar model, we have obtained tables of values of various physical parameters as a function of the radial distance from center to surface inside the Sun. The question arises: how can we test the validity of the results we have obtained? After all, they are numbers in a table, and their values are only as good as the assumptions and approximations which went into their calculation. It would be good to have an independent means of checking. This is where helioseismology comes into its own: it allows us to "peer into" the interior of the Sun and check some of the physical variables we have calculated.

14.9.1 Global sound propagation

We have already seen (Chapter 14, Section 14.6.1) that the asymptotic frequency separation $\Delta\nu$ should equal $1/(2t_s)$. In view of this, we can now see the significance of a calculation we did in Chapter 9, Section 9.4. There, we computed the value of t_s for our complete solar model, and found $t_s = 3804$ sec. Using that, we find $\Delta\nu = 1/2t_s = 131.5$ μHz. This is within 2–4% of the observational values of $\Delta\nu$: 135–136 μHz (Appourchaux et al., 1998). This tells us that our model for the Sun is doing a good job of reproducing a key global property of the "real Sun."

14.9.2 Radial profile of the sound speed

A solar model provides a radial profile of (among other things) the sound speed from center to surface. Once this is available, it is in principle possible to calculate a table of the eigenfrequencies of p-modes with various values of l and n_r.

The modifier "in principle" in the previous sentence is meant to emphasize that Equations 14.17 and 14.18 above refer only to the case of a polytrope: in the case of a realistic solar model, no single value of the polytropic index exists throughout the entire model. Therefore, new (nonpolytrope) versions of Equations 14.17 and 14.18 must be derived in which the radial profile of sound speed is incorporated explicitly. Also, for maximum precision, the Cowling approximation would have to be replaced with a more complete set of equations.

Once those changes have been made, a table of mode frequencies can be calculated. These can be checked against the measured frequencies in order to determine how good the model is. In general, the calculated frequencies will not reproduce the observed values. The discrepancies can be used to determine what numerical changes need to be made to the model sound-speeds in order to achieve better fits. An example is shown in Figure 14.5.

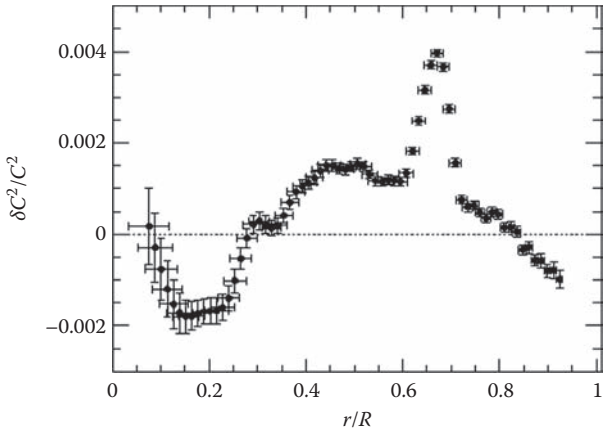


FIGURE 14.5: Radial profile of discrepancies between sound speeds in a solar model and the sound speeds required to reproduce the observed eigenfrequencies. (Courtesy of SOHO/MDI consortium. SOHO is a project of international cooperation between ESA and NASA. With permission.)

The first thing to notice about Figure 14.5 is that although discrepancies between model and data *are* present, the discrepancies are small: nowhere inside the Sun is the sound speed (squared) in error by more than 0.4%. This is a striking endorsement of the reliability of current solar models: the run of temperature inside a solar model from center to surface, ranging from > 10 million K to a few thousand K, reproduces what happens at all radial locations inside the Sun to better than a few parts per thousand.

Because modes of different degree l penetrate into the Sun by different amounts (see Equation 14.25), discrepancies between modes with the largest l values contain information about errors in the model in the outermost layers of the Sun. Modes with intermediate l values ($\approx$ 60–70) may be used to probe conditions down to the base of the convection zone (at radial locations $r \approx 0.7R_\odot$. Such regions are the site of the largest discrepancies in sound speed (see Figure 14.5): this suggests that certain physical phenomena which occur at the interface between the convection zone and the radiative core may not yet be properly incorporated in the model. Among these phenomena might be rotational shear, or magnetic fields, or overshooting of convection.

In the deepest regions of the radiative interior, information about the model is contained in p-modes with the lowest l values. But even then, the innermost part of the Sun, at radial locations within (say) $0.2R_\odot$ of the center of the Sun, cannot be probed with great reliability by p-mode data. This explains why the error bars in Figure 14.5 become considerably larger in the innermost regions of the Sun. Probing conditions near the center of the Sun will be greatly helped if the properties of g-modes can be determined with high precision.

14.9.3 The Sun's rotation

We do not need helioseismology to study the rotation on the *surface* of the Sun: that rotation can be observed directly. The observations (Chapter 1, Section 1.9) show that the Sun rotates faster at the equator than at high latitudes, with a difference of almost 30% between the equator and the poles.

When it comes to studying the Sun's rotational properties *beneath* the surface, then we must rely on helioseismology. In describing the modes which exist inside the Sun, we have concentrated on only two of the integers which specify a mode: l, and n_r. These are related to properties of the modes in the *latitudinal* and *radial* directions respectively.

However, in order to study rotation, we would also need to include, in our spherical harmonic analysis, an index m to describe properties of modes in a third direction, namely, in *longitude*. For a mode with any given value of the degree l, there exist $2l+1$ submodes with m values varying from $m=-l$ to $m=+l$. The algebraic sign of m indicates the longitudinal direction in which the mode propagates. Now, the Sun is rotating with a speed V_r which, at the equator and on the surface, has a magnitude of about 2km sec^{-1}. A p-mode (sound wave) which propagates in the *same* longitudinal direction as rotation propagates with speed c_s+v_r relative to a nonrotating frame. On the other hand, a p-mode which propagates *opposite* to the direction of rotation propagates with speed c_s-v_r relative to a nonrotating frame. These differences in propagation speed lead to differences in the frequencies of eigenmodes with $\pm m$. Near the base of the convection zone, where $c_s \approx 100$ km sec^{-1}, V_r amounts to some 2% of c_s so rotational effects are not negligible. Careful measurements of the differences in frequency allow one to extract information about V_r. Again making use of the depth-dependence of mode penetration as a function of degree l, one can obtain the radial profile of rotation.

The p-modes allow us to probe rotation in the regions of the Sun which extend from the surface in to radial locations of a few tenths of $R_\odot$. In the core of the Sun, studies of g-modes will eventually allow more detailed extraction of rotational information there.

In this "first course" in solar physics, we unfortunately do not have enough information to describe how helioseismology is used to study rotation inside the Sun. The data which are now available also allow extraction of information as to how the rotation profile varies between the equator and the pole. But analysis of such a process would carry us far beyond the limits of this "first course" in solar physics.

Suffice it to say that we present in Figure 14.6 the profile of rotational angular velocity which has been determined by analyses of this kind. The results are presented as a function of two variables: (i) the fractional radial coordinate (r/R) inside the Sun, and (ii) the latitude.

Focusing first on the *surface* of the Sun, i.e., at $r/R=1$, the curves in Figure 14.6 show that at low latitudes (0 deg), the surface rotates relatively rapidly: $\Omega/2\pi \approx$ 450–460 nanoHertz (nHz), i.e., a rotation period of 25.2–25.7 days. (This range overlaps the rotation period obtained by averaging

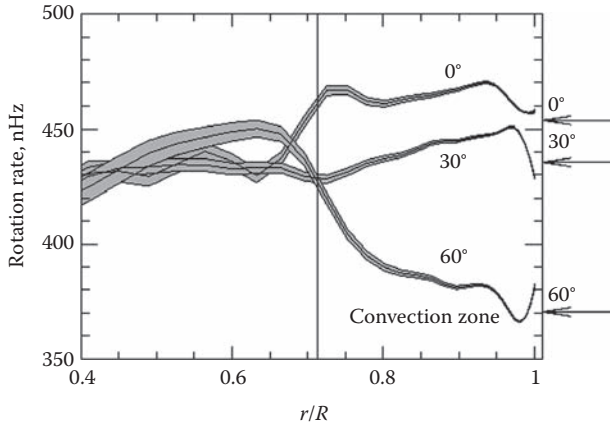


FIGURE 14.6: Angular velocity versus fractional radius inside the Sun at various latitudes. (Courtesy of SOHO/MDI consortium. SOHO is a project of international cooperation between ESA and NASA. With permission.)

surface Doppler data over 14 years: see Chapter 1, Section 1.9.) At higher latitudes, the surface rotates more slowly: at 60 deg latitudes, $\Omega/2\pi \approx 370$–380 nHz, i.e., a rotation period of 30.5–31.3 days. (This range also overlaps with the average of surface Doppler data: see Chapter 1, Section 1.9) Although the surface Doppler data allow in principle the study of rotation all the way to the poles (see Chapter 1, Equation 1.19), the helioseismological data are not sensitive enough at latitudes above (about) 60^o to allow Figure 14.6 to be extended reliably to the poles.

The fact that the Sun's *surface* does not rotate as a solid body has been known for a long time (see Chapter 1, Section 1.9). What Figure 14.6 demonstrates is that it is not merely the *surface* of the Sun which departs from solid-body rotation: this feature also exists in the *interior* of the Sun.

The long-known departure from solid body rotation at the *surface* is called latitudinal differential rotation (LDR). As we examine the *sub-surface gas* in Figure 14.6, we see that LDR persists (with values similar to the surface) down to depths of at least 0.2 solar radii, i.e., to radial locations as small as $0.8R_\odot$. At greater depths, as we approach the base of the convection zone, at radial location $r \approx 0.7R_\odot$, there is a remarkable convergence of the angular velocities to essentially a unique value. Below the convection zone, in the radiative interior, i.e., at $r \leq 0.65R_\odot$, and into a radial coordinate of order $0.5R_\odot$, the Sun rotates at essentially the same rate at all latitudes: $\Omega/2\pi \approx 430$ nHz. That is, the radiative interior of the Sun, at least in its outer regions, rotates *as a solid body* with a period of 26.9 days. Analysis of g-modes will be useful for determining the rotational properties of the central regions of the Sun, but these are not yet well developed. We do not yet know the rotation of the Sun in the inner half ($r \leq 0.5R_\odot$) with as much confidence as we do in the outer half.

The fact that the radial profiles in the convection zone in Figure 14.6 have finite slopes means that the convection zone of the Sun, at fixed latitude, has a variation in angular velocity as a function of radius. Thus, the Sun exhibits radial differential rotation (RDR), in addition to the LDR which is evident at the surface.

Why exactly the Sun shows the rotational properties shown in Figure 14.6 is not readily explainable in terms of the physics of a "first course." For example, from the simplest perspective, one might expect that the convection zone, with its turbulent stresses (which serve as a highly effective viscosity), should be able to enforce solid body rotation more easily than the radiative interior. And yet Figure 14.6 shows exactly the opposite: it is the radiative interior which exhibits solid body rotation. Given this empirical result, one might then expect (since the radiative interior contains 98% of the Sun's mass) that the gas in the convection zone (amounting to only 2% of the mass) could easily be "kept in line," and forced into solid body rotation also. But this expectation is contrary to what occurs in the real Sun. Some powerful internal dynamics must be at work to drive the convection zone into differential rotation: apparently, the forces which are at work in the convection zone (rotation, gravity, thermal buoyancy, viscous stresses) have the overall effect that the rotation of the convection zone lags behind the core (by about 10%) at high latitudes, but is ahead of the core (by about 10%) at low latitudes. Computational models are required to incorporate multiple physical effects if they are to replicate successfully the observed LDR and RDR (e.g., Kitchatinov, 2005).

We shall return to the rotational properties of the Sun when we discuss how magnetic fields are generated in the Sun (Chapter 16).

Exercises

14.1 Perform the step-by-step procedure described in Section 14.4.1 for p-modes with $l = 1$ and 3. Plot the equivalent of Figure 14.1, and obtain a table of the eigenfrequencies for $l = 1$ and $l = 3$ p-modes in the $n = 1$ polytrope for an object with solar mass and radius. For each l value, determine the "large separations" $\Delta\nu$ (in μHz) between adjacent modes. And for appropriate pairs of modes, determine the "small separations" $\delta\nu(1-3)$ between modes with $l = 1$ and $l = 3$.

14.2 You have already (Chapter 10, Exercise 1) calculated a table of values of y and y' for the polytrope $n = 3.25$. Use your tabulated values (including interpolation if necessary), and the step-by-step procedure in Section 14.5, to integrate Equations 14.17 and 14.18 in the $n = 3.25$ polytrope for g-modes with $l = 3$. Plot the results in the form shown in Figure 14.4. Compare your eigenperiods for $l = 3$ with those for $l = 1$ in Figure 14.4.

References

Appourchaux, T., Rabello-Soares, M.-C., and Gizon, L. 1998. "The art of fitting p-mode spectra: II. Leakage and noise-covariance matrices," *Astron. Astrophys. Suppl.*, 132, 131.

Baker, N. H. and Temesvary, S. 1966. "A solar model," *Table of Convective Stellar Envelope Models*, 2nd ed. NASA Institute for Space Studies, New York, pp. 18–28.

Christensen-Dalsgaard, J. 2003. "Properties of solar and stellar oscillations," p. 88 of http://www.phys.au.dk/∼jcd/ oscilnotes/chap-5.pdf

Cowling, T. G. 1941. "The non-radial oscillations of polytropic stars," *Monthly Not. Royal Astron. Soc.*, 101, 367.

Kitchatinov, L. L. 2005. "Reviews of topical problems: the differential rotation of stars," *Physics Uspekhi*, 48, 449.

Mullan, D. J. 1989. "g-mode pulsations in polytropes: high-precision eigenvalues and the approach to asymptotic behavior," *Astrophys. J.*, 337, 1017.

Mullan, D. J. and Ulrich, R. K. 1988. "Radial and non-radial pulsations of polytropes: high-precision eigenvalues and the approach of p-modes to asymptotic behavior," *Astrophys. J.*, 331, 1013.

Chapter 15

The Chromosphere

So far, when we have discussed the Sun, we have been interested in the material which extends from the visible surface *downward* into the *interior* of the Sun. The visible surface, the "photosphere" (the "light sphere"), provides the light which dominates human vision. Our model of the interior of the Sun, extending over the entire radial extent from center to photosphere, spanned a radial distance of some 700,000 km. When we computed the model, we did so in three segments, focusing on distinct laws of physics which play dominant roles in each segment. As it turned out, the three segments were found to be of unequal radial depth. The model of the deep interior (Chapter 9) extended over some 500,000 km. The model of the convection zone (Chapter 7) had a depth of some 200,000 km. And the model of the photosphere (Chapter 5) spanned no more than a few hundred kilometers in linear extent.

Now, we turn our attention in the opposite direction. Instead of starting at the photosphere and moving inward, we now start at the photosphere and move *upward* and *outward*. This brings us into the more rarefied gas that forms the outer atmosphere of the Sun. And just as we did for the interior, it will be convenient to recognize that different laws of physics are dominant in different segments of the outer atmosphere. We shall find it convenient to again discuss three segments of the outer atmosphere: the chromosphere (Chapter 15), the corona (Chapter 17), and the solar wind (Chapter 18). Of these, the linear extents are again very different: the solar wind is by far the largest, extending over vast distances of interplanetary space, with linear scales up to 10 *billion* km; the corona can be detected with optical equipment out to distances of 1–2 *million* km; and the chromosphere has a thickness of no more than 1–2 *thousand* km.

Thus, as we move outward from the surface, the chromosphere is by far the thinnest of the three segments, by analogy with the thinness of the photosphere as regards the interior of the Sun. This raises the question: why should we spend time on such a narrow region? What is there for us to learn about solar physics by paying attention to such a thin shell of gas? The answer is; the chromosphere allows us to study the effects of one particular mode of energy in the Sun: sound waves. We have already seen (Chapters 13 and 14) that sound waves provide a key mechanism for us to study the *interior* of the Sun. Now we turn to another location where sound waves play a key role, this time in terms of energy deposition.

15.1 Definition of the Chromosphere

The word "chromosphere" is derived from a Greek word meaning "color-sphere." Why is the word "color" used to describe this structure? The reason has to do with the phenomena which are visible to the human eye during an eclipse of the Sun.

There are two distinct phenomena which can be seen during an eclipse as the Moon blocks out the brilliant light of the photosphere. One of these phenomena last for a long time, while the other is over "in a flash." But both tell us something valuable about the Sun's atmosphere.

i. The long lasting phenomenon, which can be seen as long as the total phase of the eclipse lasts, is an extended white region (see Figure 15.1) that extends above the surface in a more or less uniform manner to radial distances of a few solar radii: this is the corona (Chapter 17).

ii. The short-lived phenomenon is visible only for a few (4–8) seconds at the start and end of totality (see innermost rings of Figure 15.1). A brightly colored ring is seen, confined to a narrow region very close to the solar limb. The predominant color of the narrow ring is "rose-colored." The

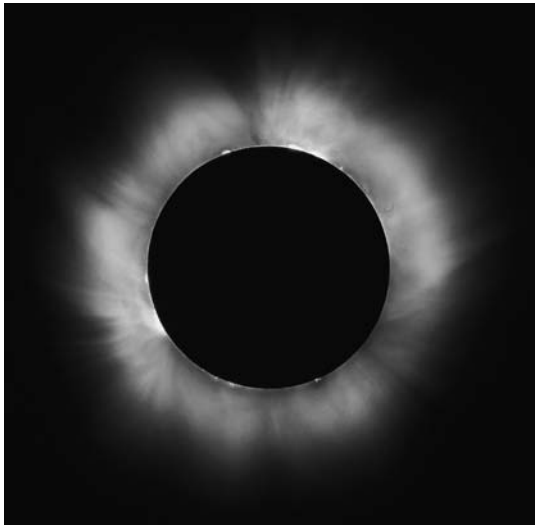


FIGURE 15.1: **(See color insert following page 202.)** The chromosphere of the Sun as seen briefly during a total eclipse of the Sun in 1999 in France. Notice that the chromosphere is confined to a narrow ring, and it has a pronounced reddish color. (Copyright Luc Viatour/www.lucnix.be. With permission.)

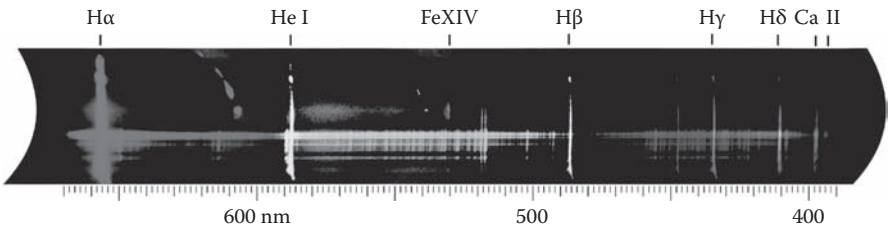


FIGURE 15.2: **(See color insert following page 202.)** Intensity of radiation from the flash spectrum: notice the strong emission lines. (From http://www.eurastro.de/pictures/sofi99/mr/linspec.jpg. Used with permission of Manfred Rudolf/Eurasto.)

> fact that the layer is obviously colored (in contrast to the white corona) gives rise to the term "chromosphere." We shall see that the fact that the dominant color is red tells us something significant about the physics of the chromosphere (see Section 15.12.3 below).

When a spectrum is obtained, the chromosphere reveals a large number of emission lines in the spectrum (Figure 15.2). The presence of emission lines in the flash spectrum provides a remarkable contrast to the photosphere, where most of the spectrum (in visible light) contains absorption lines (see Chapter 3, Figure 3.4). The fact that the chromosphere lasts for only a few seconds gives rise to the phrase "flash spectrum" for the chromospheric emission lines. The strongest lines originate in hydrogen (including the Balmer lines known as $H\alpha, H\beta, H\gamma$, and $H\delta$ at wavelengths of 6563, 4861, 4340, and 4102 Å respectively). There are also prominent lines due to helium (5876 Å), and ionized calcium (CaII H and K at 3968 and 3934 Å).

The presence of emission lines in the flash spectrum indicates that in the chromospheric gas, bound electrons in atoms and ions are cascading down from upper energy levels to lower ones: this process must start by free electrons being present in significant numbers in the gas, and then being captured by the ions. The fact that free electrons are abundant indicates that the local temperature in the chromosphere is greater than the temperature in the photosphere. Something has *heated up the gas in the chromosphere to temperatures in excess of those in the photosphere.* It will be our primary goal in this chapter to determine what causes this heating.

Historically, the presence of the line at 5876 Å in the flash spectrum is noteworthy. This line was first observed in a solar eclipse in 1868 when a spectroscope was used to view the Sun during the few seconds of the flash spectrum. The line could not be identified with any known material on Earth at the time: the name "helium" was given to the material, after the Greek work "helios" meaning the Sun. It would take 30 years before helium was discovered on Earth.

15.2 Linear Thickness of the Chromosphere

The fact that the flash spectrum lasts only for a few seconds contains information on the linear extent of the chromosphere along the radial direction. To see this, we note that the timings of the various phenomena which occur during an eclipse are determined by how fast the Moon moves across our line of sight to the Sun. Now, the Moon is in orbit around the Earth such that one orbit (360 deg) requires about 30 days. This corresponds to an angular velocity of 0.5 deg per hour, i.e., 1800 arc sec per hour, or 0.5 arc sec per second of time. As a result, in a time interval of 4–8 sec, the Moon traverses an arc having an angular extent of 2–4 arc sec. At the distance of the Sun, where the conversion factor is 728.8 km per arc sec (Chapter 1, Section 1.2), such an angular extent corresponds to a linear extent of 1500–3000 km.

The chromosphere is truly a thin shell around the Sun, extending to no more than 0.5% of the solar radius above the photosphere.

15.3 Observing the Chromosphere on the Solar Disk

Observations during a total solar eclipse allow us to see the chromosphere *at the limb* of the Sun. Such observations were the first to discover the existence of the chromosphere. But total solar eclipses do not happen very often. Is there a way to observe the chromosphere outside an eclipse? Yes: it is possible to detect the chromosphere by judicious choice of observing conditions on the disk of the Sun.

To see why this is the case, let us recall how we are best able to observe the photosphere: the strategy is to take advantage of photons which have an optical depth of unity in the photosphere. To achieve this goal, we use visible light, with a continuous spectrum in the vicinity of 5000 Å, where the continuum opacity κ is of order $0.1-0.5$ cm^2 gm^{-1} (see Chapter 5, Table 5.3). Using such light, a line of sight which starts far from the Sun (e.g., at Earth orbit) and penetrates inward, passes through enough material (with column density $H\rho_o$) to build up optical depth $\tau = \kappa H \rho_o$ of order unity in the photosphere. With scale-heights of $(1-2)\times 10^7$ cm, we recall that the level in the solar atmosphere where $\tau(5000\ \text{Å}) \approx 1$ occurs in material which has a density ρ_o of $(2-3)\times 10^{-7}$ gm cm^{-3}.

Following this line of reasoning, what must we do if we wish to probe levels of the atmosphere which lie at greater altitudes, where the gas has *lower density* than the photosphere? Clearly, we must choose an observing wavelength in some spectral line L_s where the opacity *exceeds* that in the continuum by a factor of (say) F. As a result, our line of sight coming inward

to the Sun from our observing point on Earth will reach a surface where $\tau(L_s) \approx 1$ at an altitude which is higher up than the photosphere. Now, the $\tau(L_s) \approx 1$ level occurs at a height where (provided κ and H are still about the same), ρ_o is smaller than in the photosphere by a factor of F.

Suppose $F = 10^4$: such an enhancement of opacity relative to the continuum is readily achievable in the cores of strong lines such as the first Balmer line, $H\alpha$, or the resonance lines of CaII (the so-called H and K lines). In the core of such a line, the line of sight from Earth can penetrate in only to levels where the local gas density is some 10^4 times smaller than in the photosphere. This corresponds to a density reduction of e^{9-10} compared to the photosphere. In an isothermal atmosphere, such a density reduction occurs across a vertical interval of nine to ten scale heights above the photosphere. With scale heights of 100–200 km near the photosphere, this corresponds to vertical heights of 1000–2000 km. Recalling the linear thickness of the chromosphere (1500–3000 km) which has been revealed by the flash spectrum, we see that, if we observe at a wavelength which allows our line of sight to penetrate no deeper than 1000–2000 km above the photosphere, this will put us right in the chromosphere.

The trick of observing the chromosphere without waiting for an eclipse is to tune a detector to a wavelength that is close to the center of $H\alpha$ or the CaII K line: with such a setting, the observations will reveal the "surface" of the Sun as it exists at an altitude of 1000–2000 km above the photosphere.

15.4 Appearance of the Chromosphere on the Disk

When the Sun is observed in the center of the CaK line, the chromosphere is seen to be nonuniform in brightness, especially around sunspots (see Figure 15.3).

At certain times, the things which first catch the eye are localized enhancements in brightness which lie in certain regions of the surface. These bright features lie within a range of latitudes, roughly between 10 and 30 deg, in both the Northern and the Southern hemispheres. The bright features are spatially associated with localized dark features ("sunspots") when observed in the visible continuum. The bright features in the CaK images of the chromosphere are larger in area than the sunspots: the bright features are referred to as "plages" (French for "beaches").

The rest of the surface has an appearance which is reminiscent of the skin of an orange: there are many small discrete features (pock-marks) distributed across the surface. The pock-marks look as if someone has used a broad-tipped marker to put many darker splotches on the surface, and each splotch is surrounded by a fine line of brighter material around the border. According to

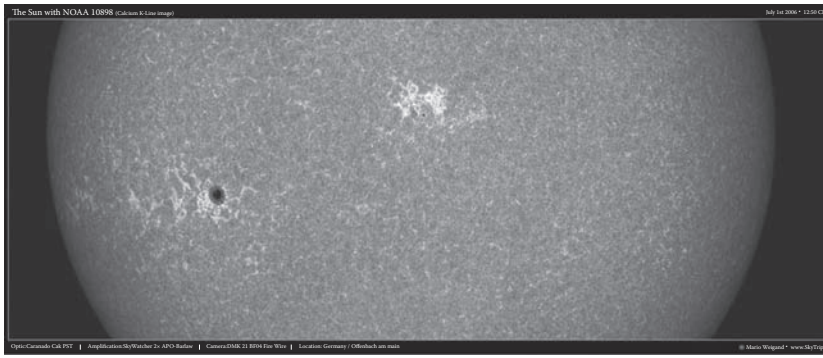


FIGURE 15.3: Part of the Sun as viewed in the K line (wavelength = 3933 Å) of singly ionized calcium. (Image downloaded from www.skytrip.de/sonne/kline2006-07-01d.jpg.)

another popular analogy, it is as if we are looking at a large number of (darker) fields, each surrounded by its own (brighter) "hedgerow" (see Figure 15.5 below).

There is well-defined topological structure: each darker region (known as a "cell") is surrounded by a brighter rim. This means that it is possible to move continuously from any element of bright rim to any other element of bright rim continuously without crossing a dark cell. But it is not possible to move from one cell to another without crossing a bright rim. The fact that the bright rims in the CaK images are connected to one another across the surface of the Sun gives rise to the descriptive term "network" for the "hedgerows" which are bright in CaK.

This topology of cells surrounded by borders is reminiscent of what we observe in the photospheric features called granules (see Chapter 6, Figure 6.1). However, whereas the photospheric granules consist of *bright* centers surrounded by *dark* rims, in the case of the chromosphere, we have the opposite structure: *dark* cell centers, surrounded by *bright* rims. The topological similarities to the granules gives rise to the term "supergranules" for the features which appear in the CaK images. A supergranule can be defined as one of the dark cells surrounded by its "hedgerow" of bright network.

15.5 Properties of Supergranules in the CaK Line

The linear dimensions of a supergranule can be determined roughly by inspection of a CaK image of the entire disk of the Sun. Starting at the center of the disk, one can draw a radial line and count the number of dark cells as

one moves out toward the limb. Estimates made from "typical" CaK images suggest that one encounters some 20–30 cells. This leads to a linear extent of order $R_\odot/(20\text{–}30) \approx 23\text{–}35{,}000$ km.

Compared with the linear scale of granules (about 1000 km), the supergranules are ≈ 30 times larger in linear extent. Each supergranule has room to contain ≈ 1000 granules within its area.

Doppler studies of supergranules indicate that the network is the site of *down-flowing* material, while the cells are regions where flows are mainly horizontal. It appears that material is flowing upward near the center of the cell, sweeping out horizontally to the network, and sinking in the network. At first sight, this is reminiscent of the convective flows in granules. However, there is a major difference: in the supergranule flows, the material which sinks is *brighter* than the rising material. This is *opposite* to the behavior in granules, where rising material is brighter (Chapter 6, Section 6.3). Therefore, it is not obvious that thermal convection, which is so helpful in interpreting the properties of granules, is at work in the supergranules.

Nevertheless, the similarity between the titles "granules" and "supergranules" raises the question: why are there preferred length scales associated with these phenomena? We have seen (Chapter 6, Section 6.4) that granules are convection cells with preferred *horizontal* scales of order 1000 km. Modeling of the convective flows in terms of mixing length theory suggests that, corresponding to the observed preferred horizontal scales, the convective flows have *vertical* extents of less than 200 km (Chapter 7, Section 7.2). The physical significance associated with such depths is that they represent 1–2 pressure scale heights: vertical motions over such distances would cause significant spreading of the material, and it is hard to imagine how gas could rise from much deeper layers and still preserve any identity. But what are supergranules? Are they really convection cells analogous to granules, except that they penetrate inward to depths that are some 30 times deeper, thereby giving rise to convection cells which have horizontal scales some 30 times larger than granules? If this is a correct interpretation, then supergranule convective flows must extend inward to depths of 6000–9000 km. Is there anything special that we could identify with material in the Sun at such depths? Maybe these are the depths where HeI $\rightarrow$ HeII ionization is increasing from less than 10% to more than 90% (see Chapter 4, Figure 4.1). However, no one has ever successfully reproduced a supergranule pattern in a model. This is in striking contrast to the remarkable success that has been achieved in replicating the observed properties of granulation in the context of thermal convection (Stein and Nordlund, 1998).

In order to make a quantitative statement about the properties of supergranules, let us ask the question: how much brighter is the network compared to the darker cell interiors? Spectra which have been obtained in plage regions (where the network is dominant) suggest that in the region of the K line, the intensity is about twice as large as the intensity in quiet Sun (where the cells are dominant) (Worden et al., 1998). We shall return later to a discussion of what causes this difference between cell and network (see Figure 15.7 below).

15.6 Supergranules Observed in the Hα Line

Another strong spectral line which is useful for observing the chromosphere is Hα. When the limb of the Sun is observed at the center of Hα, the chromosphere is observed to consist not only of an overall region of emission, but there are also multiple discrete bright linear structures (Figure 15.4).

These structures extend upward from near the photosphere to heights of a few thousand km. They appear as "spikes" of Hα emission. This spiky structure leads to the descriptive title of "spicules." Each spicule lives on average 5–10 minutes, and material moves upward along each spicule at speeds of 20–30 km sec^{-1}. Each spicule appears to be a structure along which material is guided. Magnetic fields would provide a natural conduit for ionized material to flow along.

When the disk is observed at the center of Hα (see Figure 15.5), supergranules can be identified but the topology is inverted relative to the CaK images. After one stares at the image for some time, one again sees a combination reminiscent of "field surrounded by hedgerow," except that in Hα, the "fields" are brighter and the "hedgerows" are darker. Within the hedgerow, individual short dark straight streaks (or "brush-strokes") can be identified:

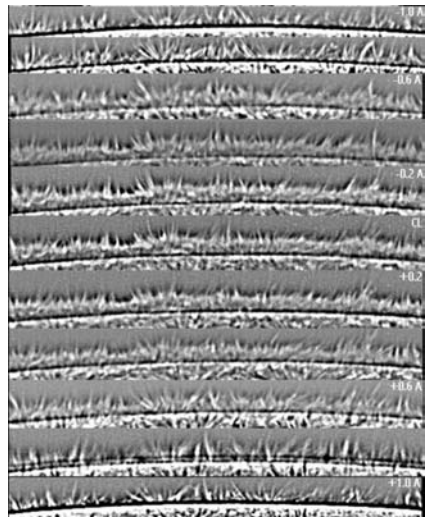


FIGURE 15.4: 11 panels showing short segments of the solar limb: each panel is taken in a wavelength near the Hα line, from the center of the line (labeled CL) to 1.0 Å on the blue side (-1.0, topmost image) to 1.0 Å on the red side ($+1.0$: image at bottom). In each panel, the Sun's disk and limb are in the lower region, and blank sky is at the top. Short linear structures reaching above the limb toward the sky are spicules. (From Big Bear Solar Observatory, New Jersey Institute of Technology. With permission.)

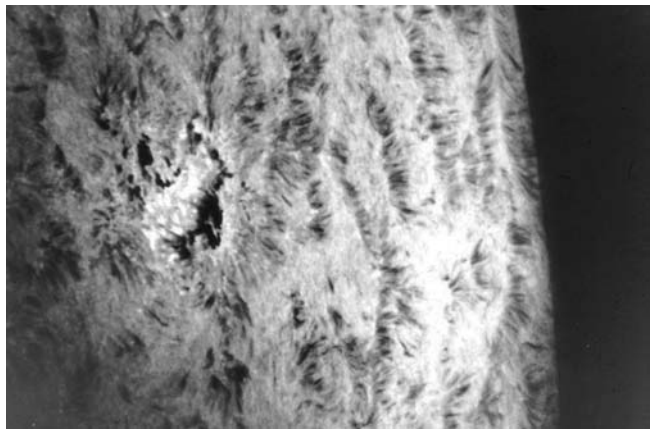


FIGURE 15.5: Image of a portion of the solar disk in $H\alpha$. The dark features that look like "brush-strokes" are composed of individual "spikes," referred to as "dark mottles." Each dark mottle is probably a spicule. Notice that the spicules are not distributed uniformly over the disk: instead, they form groups. With practice, the eye can identify brighter patches of surface which are surrounded by groups of spicules. The brighter patches correspond to supergranule cells in CaK images. The (dark) spicules correspond with the (bright) network in CaK images. (Image taken at the National Solar Observatory is operated by the Association of Universities for Research in Astronomy, under a cooperative agreement with the National Science Foundation.)

these are referred to by solar observers as "dark mottles." It seems likely that each dark mottle is actually a spicule seen in projection against the bright disk.

Why would a spicule appear bright when seen at the limb and dark when seen on the disk? Because of the background. At the limb, there is no background light to absorb, and as a result, the spicule appears as an emitting structure. But on the disk, the spicule material has plenty of H atoms in the $n = 2$ energy level which are capable of absorbing photons with wavelengths near 6563 Å coming from the photosphere. Why are there large populations of H atoms in the $n = 2$ state in a spicule? An answer is provided by Chapter 3, Section 3.3.2: because of a local increase in temperature inside the spicule, which is a region of localized energy deposition.

15.7 The Two Principal Components of the Chromosphere

Observations in both CaK and in Hα indicate that the chromosphere consists of two principal components, cell and network, each with its distinct

properties. High-resolution observations of magnetic fields indicate that the fields are concentrated in the network. It appears that the horizontal motions outward from the center of the cell "sweep up" magnetic field lines and deposit them in the network. Each spicule in the network represents a localized magnetic flux tube in which the local gas has become energized (probably by magnetic energy in some form) and has reached up to heights of several thousand km.

15.8 Temperature Increase into the Chromosphere: Empirical Results

Images of the Sun in CaK or Hα exhibit one important and noticeable difference from an image in the white light continuum. The latter shows a pronounced limb *darkening*: the intensity at the limb in visible light is a very significant 60% fainter than the center of the disk (see Chapter 2, Equation 2.5 and discussion). The observed limb darkening $I(\mu) = a + b\mu$ (where b is a positive number) can be ascribed to the source function $S(\tau) = a + b\tau$(Chapter 2, Section 2.5.3), indicating that the temperature is *increasing* as the optical depth in the continuum increases. However, when the Sun is observed in a wavelength which emphasizes the chromosphere, the eye is struck by the fact that limb darkening is not always evident (e.g., Figure 15.5). On the contrary, there are more or less extended regions where the limb actually appears to be *brighter* than the center of the disk. This is a great difference from what occurs in the visible continuum. This difference suggests that the temperature in the chromosphere is *not* increasing as the optical depth (in the chromospheric line) increases. In fact, the behavior is precisely the opposite: analysis of the chromospheric emission lines which are seen in the "flash spectrum" indicates that the temperature of the gas *increases* as the optical depth *decreases*. That is, the temperature in the chromosphere *rises* as the height above the photosphere increases.

How much does the temperature rise in the chromosphere? An empirical determination of an answer to this question requires detailed study of various lines, both absorption and emission, as well as continua, in the solar spectrum. This leads to a profile of temperature versus height. It is found that the profile of the chromospheric temperature rise in a cell differs somewhat, and in a systematic manner, from the profile of the chromospheric temperature rise in the network. In fact different pieces of the network, some of which are observed to be brighter than others, also have somewhat different profiles. And in the cell, some areas are darker than others, and these also yield somewhat different profiles. In Figure 15.6, we illustrate the profiles which have been obtained for the average cell center (B), a dark point within a cell (A), the average network (D), and a bright network element (E). (The notation is that of Vernazza et al., 1981.)

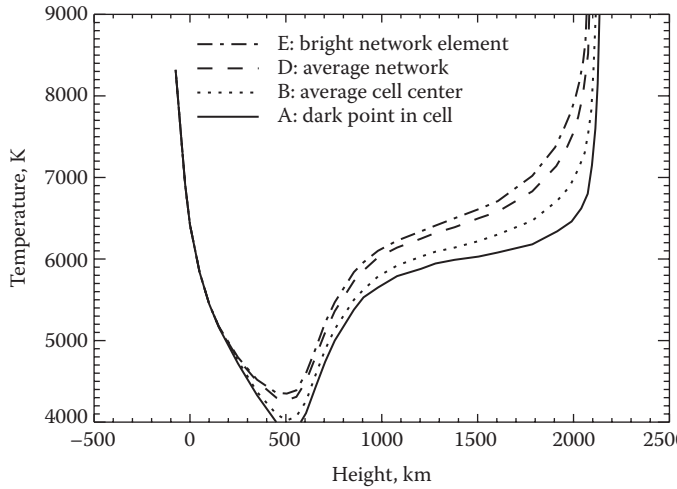


FIGURE 15.6: Temperature profiles as a function of height from the photosphere up into the chromosphere. The height scale is zeroed at the photosphere: negative heights refer to layers of the Sun which lie below the photosphere. Four distinct features on the Sun (labeled A, B, D, and E) are illustrated. Above the photosphere, notice the temperature minimum at $h \approx 500$ km, the rising temperatures between $h = 500$ and 1000 km, the temperature "plateau" between $h = 1000$ and 2000 km, and the steep increase in temperature above ≈ 2100 km. (The fair profiles were constructed by plotting data from Vernazza, J. E. Avrett, E. H., and Loeser, R. 1981. *Astrophys. J. Suppl.*, 45, 635.)

Let us consider some important features of the profiles in Figure 15.6.

First of all, all four profiles merge in and below the photosphere: it is only at heights of a few hundred kilometers above the photosphere that the curves begin to differ.

Second, in all four cases, the temperature passes through a minimum value at heights close to 500 km above the photosphere. We may regard the temperature increase at heights above 500 km as an indication that we have entered into "the lower chromosphere."

Third, above 500 km, the temperature at first begins to increase rather steeply, but at heights of about 1000 km, the temperature profile flattens out to some extent. This gives rise to a "plateau" in the temperature at heights between (roughly) 1000 and 2000 km. The temperatures in the plateau are mainly in the range 6000–6500 K. This plateau can be regarded as defining the "middle chromosphere." In the plateau, the local temperature exceeds the temperature minimum by roughly 2000 K.

Fourth, at heights of about 2100 km, there is a steep increase in temperature: this is referred to as the "upper chromosphere."

Fifth, within the plateau, the four distinct features differ from one another in temperature by several hundred degrees. E.g., at $h = 1500$ km, the average cell center has $T \approx 6200$ K, while the bright network element has $T \approx 6600$ K. Although the difference in temperature of 400 K is only some 20% of the 2000 K increase in temperature above the temperature minimum, we shall find (Section 15.12.2) that this relatively small difference in temperature actually requires a much large difference in energy deposition in the network than in the cell.

Sixth, the overall thickness of the chromosphere, from "lower" to "upper," is some 1600–1700 km. This overlaps with the range of thicknesses reported in Section 15.2 from eclipse timings.

Seventh, we have mentioned (Section 15.6) that discrete structures (spicules) exist at certain locations in the network: these may extend upward in height to a few thousand kilometers. Some spicules therefore have heights which exceed the thickness of the chromospheric profiles in Figure 15.6. The "real chromosphere" includes some spiky structure which is not well described by the results in Figure 15.6. The latter should be regarded as applicable to the "nonspiky" parts of the solar atmosphere.

15.9 Temperature Increase into the Chromosphere: Mechanical Work

The most striking result in Figure 15.6 is that the local temperature *increases* as the height increases above the photosphere. This result indicates that it is no longer useful to think in terms of the Eddington atmosphere, where radiative equilibrium was operative. In the latter conditions, we have seen (Chapter 2, Equation 2.40) that the temperature should vary as $T^4 \sim \tau + \text{const}$, i.e., T should approach a constant value as we go higher up in the atmosphere, where $\tau \to 0$. In the Eddington atmosphere, there should certainly be no tendency for the temperature to *increase* as we move toward smaller values of τ.

Clearly, something quite different from radiative equilibrium is operating in the chromosphere. What could it be?

Up to this point, in our calculations of the solar model (Chapters 5, 7, and 9), we have always been dealing with material where the temperature falls off monotonically as radial distance from the center of the Sun increases. In the presence of such a negative radial gradient of temperature, it is natural to think in terms of the heat which flows *down* the temperature gradient.

Now, as we enter into the chromosphere, T starts to *increase* as the radial distance increases. The radial gradient of temperature is now *positive*. Such a positive gradient cannot be a consequence of heat flow from the inner portions of the Sun. There is a different physical process at work in order to raise the

temperature in the chromosphere: this process is *mechanical work*. The source of the work can be identified with the thermodynamic term PdV, which occurs when a suitable pressure P compresses the volume V of 1 gm of gas. Where can we find suitable pressures to perform such work in the solar atmosphere? In sound waves: we have already seen that the Sun supports multiple p-modes, each of which is a sound wave. Sound waves are longitudinal modes in which pressure compresses and rarefies the local gas as the wave propagates past any point.

Thus, sound waves *can* do mechanical work on the gas in the Sun. However, the fact that the chromosphere extends *above* the photosphere by linear distances of a few thousand kilometers indicates that the waves which are responsible for chromospheric heating are *not* identical to the trapped p-modes: the latter are *trapped below the photosphere*, whereas now we need to have waves which are *propagating above the photosphere*. As we have seen (Equation 13.15), waves which are capable of freely propagating above the photosphere must have periods shorter than 195–200 sec, i.e., frequencies higher than 5 mHz.

The increase in chromospheric temperature is found to occur in *both* the cell and the network. However, material in the network increases in temperature faster than material in the cell (see Figure 15.6: curve E vs. curve A). Over the same range of heights, the average network gas is hotter than the average cell gas by an extra few hundred degrees. This suggests that the supply of mechanical energy is greater in the network than in the cell. Although the excess temperature in the network seems relatively small (about 20%), we shall see that it actually requires a large difference in mechanical energy deposition.

15.10 Modeling the Chromosphere: The Input Energy Flux

The aim of any attempt to model the chromosphere is to calculate how the temperature varies as a function of height. Specifically, how fast does the temperature increase above the boundary value that is predicted by the photospheric model? Does it increase by (say) 1000 K over a height interval of 10 km? or 100 km? or 1000 km?

In order to calculate the temperature rise, let us consider how sound waves could provide mechanical energy to the gas. Let us start in the photosphere with a flux of mechanical energy in sound waves, and calculate what happens as the waves propagate upward. A sound wave can be characterized by an amplitude in velocity δV. In a medium of density ρ, the energy density of the wave is $0.5\,\rho\,\delta V^2$ ergs cm^{-3}. The waves propagate at the speed of sound c_s. As a result, the acoustic energy flux is given by

$$F(\mathrm{ac}) = 0.5\,\rho\,\delta V^2 c_s \tag{15.1}$$

In the photosphere of the sun, our solar model informs us that $\rho \approx 3 \times 10^{-7}$ gm cm^{-3}. We also know that the local (adiabatic) speed of sound c_s is given by the formula $\sqrt{(1.67 R_g T/\mu)}$, where μ is the mean molecular weight. In the photosphere, the numerical value of c_s is $\approx$10 km sec^{-1}. This leads to $F(\text{ac}) \approx 0.15\ \delta V^2$.

What are we to use for δV, the amplitude of the sound waves in the photosphere? We have seen (Chapter 3, Section 3.8) that line profiles in the solar spectrum have excess widths over and above what the lines would have in the presence of purely thermal motions. The excess widths, of order 1–2 km sec^{-1}, and ascribed to "microturbulence," in all likelihood include contributions from sound waves in the photosphere. The observed amplitude of the turbulence may therefore be regarded as an upper limit on the amplitude of sound waves in the photosphere.

Of the observed microturbulence of 1–2 km sec^{-1}, let us suppose that sound waves contribute no more than 50%: i.e., we assume that the amplitude of sound waves in the photosphere is no more than δV (photo) $\approx$ 1 km sec^{-1}. This sets a limit on $F(\text{ac})$ of $\leq 1.5 \times 10^9$ ergs cm^{-2} sec^{-1}. Compared to the energy flux passing through the photosphere in the form of radiation ($F_\odot = 6.3155 \times 10^{10}$ ergs cm^{-2} sec^{-1}: see Chapter 1, Section 1.8), we see that $F(\text{ac})$ in the photosphere is less than 2% of the overall flux of energy coming up through the solar atmosphere. We are certainly not discussing a major channel for the transport of energy through the photosphere: radiation is still far and away the dominant channel for energy transport in the visible layers of the Sun.

Any acoustic energy which *is* present in the photosphere, and which contributes to microturbulent line broadening, certainly includes some p-modes with periods in excess of (about) 200 sec. However, such long-period waves cannot propagate up into the chromosphere. The only segment of the acoustic flux which is of interest as far as the heating of the chromosphere is concerned is the segment where the waves have periods which are short enough to allow vertical propagation. This segment contains only those waves with periods which are *shorter* than 200 sec. In order to estimate the flux of sound waves which can actually reach the chromosphere (thereby contributing to heating), we need to reduce the above upper limit on $F(\text{ac})$.

What fraction of $F(\text{ac})$ reaches the chromosphere? The answer depends on the spectrum of the acoustic power which is generated by the convective motions. Most of $F(\text{ac})$ is expected to be generated at periods corresponding to granule turnover times, or lifetimes (see Chapter 14, Section 14.8.2), i.e., at periods of 300–600 sec (Chapter 6, Section 6.2). Waves with periods of less than 200 sec are expected to contribute only a small fraction to the overall spectrum. According to one estimate (Musielak et al., 1994), the acoustic energy flux which reaches the chromosphere $F(\text{chr})$ is no more than 5×10^7 ergs cm^{-2}sec^{-1}. That is, only 3–4% of $F(\text{ac})$ is in the form of waves which are free to propagate vertically in the solar atmosphere: as expected, the great majority of $F(\text{ac})$ created by granules which live 300–600 sec is in the form of waves with periods that are longer than 200 sec. Musielak et al.

(1994) find that their estimate of $F(\text{chr})$ is quite sensitive to various assumptions about the properties of turbulence. It is entirely possible that $F(\text{chr})$ could be in error by a factor of two or more.

In view of the uncertainties, a conservative range of estimates of the flux of acoustic energy that is available as the input for chromospheric heating in the Sun may be

$$F(\text{chr}) = 10^{7-8} \text{ erg cm}^{-2} \text{ sec}^{-1} \tag{15.2}$$

Since the two basic components of the chromosphere are observed to differ in brightness, it seems plausible that the lower limit of the range of $F(\text{chr})$ in Equation 15.2 might apply to one component (the cell), while the upper limit in Equation 15.2 might apply to the other component (the network). We shall return to this when we discuss the amount of heating.

15.11 Modeling the Chromosphere: Depositing the Energy

What happens to the flux of acoustic energy $F(\text{chr})$ as it propagates upward in the Sun's atmosphere? At first, the amplitudes of the waves are small enough that the waves simply "ride" through the gas, dissipating no energy. In this regime, the energy flux of the waves remains constant. In the upper photosphere, where the temperature is almost constant with height (Chapter 2, Section 2.10), c_s is also almost constant with height. As a result, the constancy of $F(\text{chr}) \sim \rho \delta V^2 c_s$ requires that the wave amplitude δV varies as $1/\sqrt{\rho}$.

Now, in an atmosphere which is essentially isothermal, ρ falls off exponentially with height: $\rho(h) \sim \exp(-h/H_p)$ (Equation 5.4). As a result, δV *increases* exponentially with increasing height according to $\delta V(h) \sim \exp(+h/2H_p)$. As we have already seen (Section 5.1), in the solar photosphere $H_p =$ 115–140 km.

We have seen that the amplitude of sound waves in the photosphere layers $\delta V(\text{photo})$ is perhaps 1 km sec^{-1}. Compared to the local sound speed, the sound wave amplitudes in the photosphere are $\approx 0.1 c_s$. Applying the exponential growth formula, we see that when the waves reaches a height h_s where $\exp(h_s/2H_p) \approx 10$, then the amplitude of the sound waves will have grown to a value $\delta V(h_s)$ which approaches c_s. This occurs at a height $h_s \approx 4.6 H_p$ above the photosphere, i.e., $h_s \approx 530$–640 km above the photosphere.

What happens to an acoustic wave when its amplitude becomes comparable to the local sound speed? To see what happens, we note that a sound wave consists of a crest and a trough: the wave is moving forward as a whole at speed c_s. However, the meaning of the term "amplitude of the wave" means that the matter in the crest is moving with a speed of δV *relative to the wave.* That is, the matter in the crest of the wave is moving relative to a stationary

observer at a speed $\delta V + c_s$, while the material in the trough of the wave is moving relative to a stationary observer at speed $-\delta V + c_s$. When δV approaches c_s, the material in the crest overtakes the material in the trough. Then the wave profile becomes so steep that a vertical step in pressure develops: in this condition, the sound wave has become a shock front. This behavior is reminiscent of water waves approaching a beach: when the wave becomes vertical, the wave can no longer continue to be a sinusoidal motion. At that point, the wave "breaks," and deposits its energy in the form of a churning white-cap. Analogously, when an acoustic wave evolves to the condition of a shock front, the pressure jump across the wave "breaks," leading to local churning and compression of the gas. As a result, the PdV work appears in the form of localized heat.

This leads us to an important conclusion about a certain region in the solar atmosphere, particularly the region $h_s \approx 530$–640 km above the photosphere. At such heights, we expect that sound waves from the photosphere will begin to "break" and, as a result, acoustic energy will begin to be deposited effectively in the solar atmosphere. In this regard, it is important to note from Figure 15.6, that this height range is precisely where the empirical models of the chromosphere indicate that the temperature reaches a minimum, and starts to increase upward. In view of what we have said about acoustic waves undergoing steepening, and forming shock waves, it is natural to attribute the empirical increase in temperature at heights of 500 km or more above the solar photosphere to the onset of shock heating.

In contrast to the Eddington model, where T was predicted to *fall off* slowly as height increases, now the dissipation of acoustic power indicates that the temperature should *start to increase* above a certain height, h_s. The temperature in the solar atmosphere reaches a minimum value, $T(\text{min})$, in the vicinity of height h_s.

The region of the "temperature minimum" may be thought of as a boundary between the upper photosphere (below) and the low chromosphere (above).

Although we expect that shock heating will *set in* at heights of order h_s, we do not expect the acoustic energy to be deposited *in its entirety* at a single location. For one thing, the local heating increases the scale height, and this helps to postpone further steepening of the wave to greater heights. Instead of instantaneous local dissipation, the process is spread out in the vertical direction such that the acoustic flux falls off roughly as $\exp(-h/\lambda_d)$, where λ_d is a dissipation length scale. Since dissipation is associated with steepening of the waves, and the steepening is associated with the falling-off in density (which occurs on an e-folding scale of H_p), we expect that λ_d might be of order a few times H_p. For purposes of rough estimation, we use $\lambda_d \approx 300$ km.

This result allows us now to estimate the rate $E(\text{chr})$ at which acoustic energy is deposited into each cubic cm of the atmosphere: $E(\text{chr}) \approx F(\text{chr})/\lambda_d$. Inserting the values given above, we find that acoustic energy is deposited in the chromosphere at a rate which is, at least as to order of magnitude

$$E(\text{chr}) \approx 0.3 - 3 \text{ ergs cm}^{-3} \text{ sec}^{-1} \tag{15.3}$$

15.12 Modeling the Equilibrium Chromosphere: Radiating the Energy Away

When mechanical energy is "dumped" into a cubic cm of gas, the gas attempts to get rid of the energy by whatever means are available. One of the most efficient means available to gas at the temperature minimum is to increase the local temperature by a finite amount, and then use the increased efficiency of radiative ability at the higher temperature to radiate the energy away. If the gas is successful in finding a way to radiate energy at a rate of 0.3–3 ergs cm^{-3} sec^{-1}, then an equilibrium can be reached: the local temperature can achieve a more or less steady state.

Let us turn to a calculation of the excess temperature which would allow the solar atmosphere to reach such an equilibrium.

15.12.1 Radiative cooling time-scale

We first estimate how long it takes for gas to cool by means of radiation. Suppose a parcel of gas is heated (for whatever reason) to a temperature T which is hotter than its surroundings: the latter are at temperature T_o. How long would it take for the heated parcel to radiate away its excess heat energy? The gas (with density ρ) in a volume element dV has excess internal energy $E(\mathrm{exc}) = C_v \rho (T - T_o) dV$ ergs, where C_v is the specific heat.

How quickly can this excess energy be radiated from this volume element? It depends on what form of radiation is available to the gas. Suppose the radiation is predominantly in the continuum. Let the surface area of the volume element be dA. If the element is optically thick, then the energy will be radiated from the surface dA at a rate given by that of a black body: the emergent intensity is such that the rate at which energy is radiated out of each sq cm into a background medium with temperature T_o (integrated over 4π solid angle) is given by the difference in source functions: $S_{bb}(T) = 4\pi\sigma_B(T^4 - T_o^4)/\pi$ (see Chapter 2, Equation 2.37). In this limit, the rate at which the excess energy in the volume element is radiated away in the continuum $(dE/dt)_{\mathrm{rad}}$ would simply by $S_{bb}(T)dA$, i.e., $4\sigma_B(T^4 - T_o^4)dA$ ergs sec^{-1}.

However, as we move upward in the solar atmosphere, and encounter gas with smaller and smaller densities, the volume element will *not* always be optically thick. Instead, the line of sight through the element may have a small optical depth $\tau(< 1)$. This has the effect that the emergent intensity is reduced by the factor τ (Chapter 2, Equation 2.18): $S = \tau S_{bb}(T)$. This leads to

$$\left(\frac{dE}{dt}\right)_{\mathrm{rad}} = 4\sigma_B \tau (dA)(T^4 - T_o^4) \tag{15.4}$$

At this rate of energy loss, how long will it take for radiation to cause the parcel to cool down to $T = T_o$? This "cooling time" $t(\mathrm{cool})$ is given by

$E(\text{exc})/(dE/dt)_{\text{rad}}$. This leads to

$$t(\text{cool}) = \frac{E(\text{exc})}{(dE/dt)_{\text{rad}}} = \frac{C_v \rho (T - T_o)}{4\sigma_B (T^4 - T_o^4)} \frac{1}{\tau} \frac{dV}{dA} \tag{15.5}$$

The cooling time depends on the local conditions and also on the optical depth of the parcel. In general, the ratio of the volume of the element dV to its surface area dA is associated with the linear scale ds of the element: $dV/dA{=}ds$. However, according to the definition of optical depth, we can also write the optical depth of the element in terms of the linear scale: $\tau = \kappa\rho\ ds$. Substituting this in Equation 15.5, we obtain

$$t_{\text{cool}} = \frac{C_v (T - T_o)}{4\sigma_B \kappa (T^4 - T_o^4)} \tag{15.6}$$

This expression is valid for the regions in the solar atmosphere where continuum radiation is efficient. In higher layers, where emission lines become more efficient radiators, we do not expect to find Equation 15.6 as useful.

15.12.2 Magnitude of the temperature increase: the low chromosphere

Now that we know how rapidly energy can be radiated away from a volume element near the temperature minimum, we can estimate the local increase in temperature $\Delta T = T - T_o$ which occurs as a result of deposition of mechanical energy at a rate $E(\text{chr})$.

An increase in the local temperature by an amount ΔT causes the local thermal energy density to increase by $\Delta E = C_v \rho \Delta T$ ergs cm^{-3}. This excess energy can be radiated away at a rate that is determined by the cooling timescale $t(\text{cool})$:

$$\left(\frac{dE}{dt}\right)_{\text{rad}} \approx \frac{\Delta E}{t_{\text{cool}}} = 4\sigma_B \kappa \rho (T^4 - T_o^4) \tag{15.7}$$

The units of the right- and left-hand sides of Equation 15.7 are ergs cm^{-3} sec^{-1}. Equilibrium is possible if the rate at which energy is being deposited into a unit volume $E(\text{chr})$ (see Equation 15.3) is equal to the rate at which energy is radiated out of that unit volume $(dE/dt)_{\text{rad}}$. This leads to

$$4\sigma_B \kappa \rho (T^4 - T_o^4) = 0.3\text{–}3 \tag{15.8}$$

Inserting the value of the Stefan–Boltzmann constant σ_B, we find that the gas in the solar atmosphere *can* reach equilibrium if

$$\kappa (T^4 - T_o^4) \approx (1 - 10) \times 10^3 / \rho \tag{15.9}$$

Can we find a solution to this equation? In order to answer this, we need to know how the opacity κ depends on temperature and density in the upper

parts of the solar photosphere. We have already seen (Chapter 3, Section 3.7) that κ can be fitted in certain regimes of temperature with power laws in density and temperature. In the present case, we are interested in gas where the temperature lies below 10^4 K. In that case, as we found in Chapter 3, Section 3.7, $\kappa \approx 10^{-32}\rho^{0.3}\,T^9$. The steep dependence on temperature is noteworthy: it arises mainly because an increase in temperature (in this temperature range) leads to rapid increases in the populations of the upper levels of hydrogen atoms. Inserting this in Equation 15.9, we find

$$T^9(T^4 - T_o^4) \approx (1 - 10) \times 10^{35}/\rho^{1.3} \tag{15.10}$$

Solutions of this equation, for a given density ρ, indicate the temperature to which gas in the solar atmosphere of density ρ would be heated if (i) energy were deposited in that gas at a rate given by Equation 15.3, and if (ii) continuum opacity determines the radiative losses.

What value of density should we use in Equation 15.10? The answer depends on where exactly in the solar atmosphere the mechanical energy is being deposited. Densities in the solar atmosphere vary over a wide range. In the photosphere, our solar model suggests $\rho \approx 3\times10^{-7}$ gm cm^{-3}. Densities at the temperature minimum, i.e., some 4.6 scale heights above the photosphere, are lower than the photospheric densities by factors of $e^{-4.6} = 0.01$. Thus, local densities in the low chromosphere start at about 3×10^{-9} gm cm^{-3}. In the upper chromosphere, at heights of 2000 km, i.e., at least $14H_p$ above the photosphere, the densities are smaller than photospheric values by $e^{-14} \approx 10^{-6}$. As a result, when we consider conditions in the solar chromosphere, we are interested in the solutions of Equation 15.10 over a range of densities from (roughly) 3×10^{-9} gm cm^{-3} to 3×10^{-13} gm cm^{-3}. (For future reference, we note that the latter *mass* density corresponds to a *number* density in the upper chromosphere of order 2×10^{11} protons cm^{-3}). Using the condition of hydrostatic equilibrium, i.e., $\rho(h) = \rho_o \exp(-h/H_p)$, we can associate (roughly) each value of density with a corresponding height above the photosphere. (We use $H_p = 140$ km and $\rho_o = 3 \times 10^{-7}$ gm cm^{-3}).

Let us assume that the background atmosphere (before acoustic waves are present) has $T_o = T(\min) \approx 4000$ K (see Figure 15.6). Using this, we can obtain solutions to Equation 15.10 for any choice of density throughout the above range. For clarity, we consider two distinct components of the chromosphere: in one, the deposition of acoustic flux occurs at a low rate (we use the number 1 in brackets on the right-hand side of Equation 15.10), while in the other, the acoustic flux is deposited at a ten times higher rate (10 in brackets on the right-hand in Equation 15.10). The corresponding solutions to Equation 15.10 are presented in Figure 15.7. We see that, over a range of heights from about 500 km to about 1000 km, the temperature is predicted to rise steeply to a value that is at least 2000 K above the temperature minimum. Thus, acoustic dissipation, in combination with continuum radiative losses appears to account quite well for the initial rise in temperature in the low chromosphere.

The high-flux solution agrees best with empirical curve E, the bright network element. The low-flux solution lies closer to empirical curve A, the dark point in the supergranular cell. At heights in the low chromosphere, the high-flux solution gives rise to temperatures which, at any particular height, are hotter than on the low-flux solution by several hundred degrees. Thus, even though the empirical curves appear to be separated in temperature by a relatively small amount (a few hundred degrees), that temperature difference corresponds to input rates of mechanical energy which differ by a factor of ten.

Why are the rates of mechanical energy deposition in the bright network elements ten times larger than in the dark point in the cell? One obvious difference between such locations is the magnetic field strength: supergranule flows cause magnetic fields to be strong in the network, but weak in the cell. The strong fields in the network provide channels for spicules to exist. As a result, it seems likely that magnetic effects have something to do with ten times enhanced mechanical energy deposition rates in the network.

15.12.3 Magnitude of the temperature increase: the middle chromosphere

The results in Figure 15.7 show clearly that although we have been successful in fitting the temperatures in the low chromosphere, using Equation 15.10, the fit definitely breaks down in the middle chromosphere. The reason for the break-down is related to the choice of source function that was used for the radiative losses in the low chromosphere: we chose the black-body relation

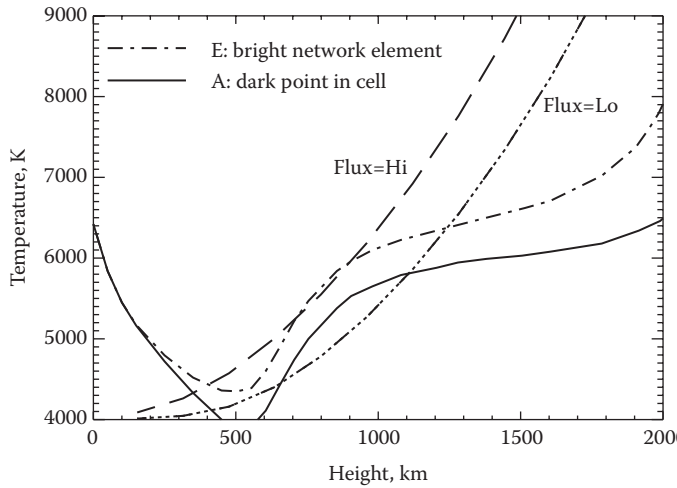


FIGURE 15.7: Fits to chromospheric temperature increases using low and high fluxes of acoustic waves. Curves A and E: portions of two of the empirical curves in Figure 15.6. Lines labeled Flux: solutions of Equation 15.10 in the limits of low and high flux.

$S_{bb} \sim T^4$, which is valid only as long as the continuum photons and the gas are tightly coupled in local thermodynamic equilibrium (LTE).

However, as we rise to greater altitudes above the solar photosphere, and the density falls off exponentially, the coupling between continuum photons and gas diminishes. The gas becomes less and less efficient as a continuum radiator.

At the same time, bound levels in the dominant atoms and ions are being increasingly well populated by the rising temperatures. Emission lines from certain bound levels become effective coolants of the gas in the middle chromosphere. Among these lines, the H and K lines of Ca II, as well as the H and K lines of Mg II are particularly effective in the lower regions of the middle chromosphere (Vernazza et al., 1981: see their Figure 49). In the upper regions of the middle chromosphere, as the temperature reaches values close to 7000 K, hydrogen lines become the predominant channels of radiative cooling. The radiative cooling rates in these various lines have temperature and density dependences which depart significantly from black-body curves.

In order to understand the plateau in temperature in the middle chromosphere, it is important to recall that the flux of acoustic power which is responsible for chromospheric heating originates in the convective turbulence below the photosphere. As a result, the acoustic flux $F(\mathrm{ac})$ is maximum near the photosphere, and it diminishes with increasing height. Moreover, the mechanical energy which is deposited in the middle chromosphere does not go simply into increasing the local temperature. Instead, the energy is diverted in increasingly large amounts to internal degrees of freedom: population of bound levels, and (ultimately) the ionization of hydrogen (and helium). In Chapter 4 (Section 4.3), we saw that at temperatures of about 7000 K in the chromosphere, H approaches 50% ionization. In the upper part of the middle chromosphere, hydrogen ionization rises above the 50% level.

The combination of reduced rates of input of mechanical energy, the onset of strong radiative cooling which occurs predominantly in emission lines, and the siphoning off of energy into bound levels and ionization, leads to a plateau in the temperature. The bound levels of hydrogen act, in effect, as a kind of thermostat for the middle chromosphere.

From the plateau in the middle chromosphere, where hydrogen is roughly 50% ionized, strong emission of the Balmer lines occurs. The first member of the Balmer series, Hα (at a wavelength of 6563 Å), is the strongest emitter from the chromosphere in the visible spectrum. The red color of this strong line accounts for the "rose-colored hue" that is a common feature of the flash spectrum during an eclipse of the Sun (see Figure 15.1).

15.12.4 Magnitude of the temperature increase: the upper chromosphere

In the upper chromosphere, where the temperature increases above 7000 K, rapidly approaching 10^4 K and higher, hydrogen approaches complete ionization. No longer are there internal degrees of freedom (bound levels,

ionization) available to absorb mechanical energy. No longer are there strong continua or lines available to radiate away the mechanical energy. Equilibrium is not possible: there is a "runaway" of the temperature to high values.

With only thermal energy available, and with the low density of the gas (approaching 10^{-13} gm cm^{-3}), the deposition of energy even at a rate $E(\mathrm{chr})$ that is much lower than in Equation 15.3, leads to rapid local heating. To see this, note that the thermal energy density e, which is comparable to the local pressure, $R_g \rho T/\mu$, obeys the equation $de/dt = E(\mathrm{chr})$ when there are no longer any effective channels to carry away the energy. Thus, even if the deposition rate is as low as (say) 0.1% of the lowest value in Equation 15.3, the rate of temperature increase in gas of density 10^{-13} gm cm^{-3} is $dT/dt = 0.001\mu E(\mathrm{chr})/R_g\rho$. With $\mu = 0.5$ in ionized hydrogen, and $R_g = 8.31 \times 10^7$ ergs gm^{-1} deg^{-1}, we find $dT/dt \approx 20$ K sec^{-1}. Within a matter of minutes, the local temperature increases by several thousand K.

This can help us to understand why the temperature rises steeply in the upper chromosphere (see Figure 15.6).

For future reference, we note that at the top of the chromosphere, where $T \approx 10^4$ K, and number densities are of order 2×10^{11} cm^{-3}, the gas pressures ($p = 2N_e kT$) are of order 0.6 dyn cm^{-2}. When we discuss the corona (Chapter 17), it will be valuable to compare this pressure near the *top of the chromosphere* with the pressure near the *base of the corona*.

Exercise

15.1 The estimates of chromospheric heating given in Equation 15.10 are obtained by picking a particular fitting formula for the opacity in Equation 15.9, $\kappa \approx 10^{-32}\rho^{0.3}T^9$. Other choices of fits to the opacities are possible, using a different coefficient and different exponents. Choose values of 5, 7, and 11 for the temperature exponent, and values of 0 and 0.5 for the density exponent. For each pair of exponents, recalculate the fitting formula such that, in all cases, $\log \kappa = 4$ when $\log \rho = 0$ and $\log T = 4$, and then recalculate the curves labelled Flux=Hi and Flux=Lo in Figure 15.7.

References

Musielak, Z., Rosner, R., Stein, R. F., and Ulmschneider, P. 1994. "On sound generation by turbulent convection: a new look at old results," *Astrophys. J.*, 423, 474.

Stein, R. F. and Nordlund, A. 1998. "Simulations of solar granulation. I. General properties," *Astrophys. J.*, 499, 914.

Vernazza, J. E. Avrett, E. H., and Loeser, R. 1981. "Structure of the solar chromsphere. III. Models of the EUV brightness components of the quiet Sun," *Astrophys. J. Suppl.*, 45, 635.

Worden, J. R., White, O. R., and Woods, T. N. 1998. "Plage and enhanced network indices derived from CaII K spectroheliograms," *Solar Phys.*, 177, 255.

Chapter 16

Magnetic Fields in the Sun

Up to this point, we have been considering the Sun in terms of material which can be described reliably by the laws of "ordinary" gas dynamics and radiative transfer. This has been sufficient to allow us to describe in some detail the overall structure of the Sun, including radial profiles of pressure, temperature and density from the center all the way to the surface. But the very concept of a radial profile incorporates the assumption that the profile is the same in all directions, i.e., the material is spherically symmetric. This is certainly an adequate assumption deep in the interior of the Sun.

However, as we approach the surface, certain features become apparent in the Sun where departures from spherical symmetry are more or less severe. One such effect is introduced by rotation: the Sun departs from a spherical shape by having a slightly oblate figure. But the effect is so small that the unaided eye cannot see the effect. In fact, reliable measurements of the oblateness are quite difficult to make (Chapter 1, Section 1.9).

16.1 Sunspots

The most dramatic departures from spherical symmetry on the Sun's surface are sunspots (see Figure 16.1). These are darker areas of the surface which are sometimes large enough to be seen by the unaided eye. Occasional reports of naked-eye sunspots by Chinese observers are on record for the past two millennia. To be sure, the advent of telescopes has greatly increased the observability of sunspots. But even during the time period 1600–1650, when telescopes first became available, there are records of as many as 33 naked-eye sunspots, including one by Galileo himself (Vaquero, 2004).

Sunspots spanning a wide range of sizes appear from time to time, in an unpredictable way, as dark spots somewhere on the surface of the Sun, usually as a pair of spots, or in groups. As solar rotation carries a pair of spots across the disk of the Sun, one spot is in the lead, and the other follows. This gives rise to the notation "leader" spot and "follower" spot. Thus, when a pair of spots first rotates onto the disk, appearing at the East limb, we see first the leader. And when the pair eventually (about two weeks later) reaches the West limb, it is the leader spot which disappears from view first.

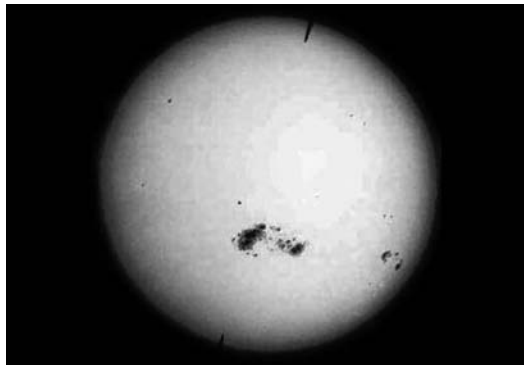


FIGURE 16.1: Sunspots on the Sun in 1947, including the largest sunspot ever photographed. Some of the small groups consist of only two spots, the "leading" (on the right) and "following" (on the left) components. The largest group contains many spots, although there is one dominant "leading" spot and one dominant "following" spot. The pointed lines at top and bottom denote the rotational North and South poles of the Sun. (Image is accessible to the public on the NASA website http://sunearthday.gsfc.nasa.gov/2006/multimedia/gal_015.php.)

In most spot pairs, the line between the centers of the leader spot and the follower spot is aligned almost East–West, although there is a slight (but definite) tilt away from the exact East–West direction. The sense of the tilt is clearly defined: the leader is situated at slightly lower latitudes (i.e., closer to the equator), while the follower is situated at slightly higher latitudes. That is, in the Northern (Southern) hemisphere, the follower is situated closer to the North (South) pole. The existence of this slight but definite tilt has a role to play in our understanding of the solar cycle (see Section 16.9).

A large spot (either leader or follower) consists of a darker central core (the "umbra"="shadow" in Latin) surrounded by a "penumbra," which is intermediate in brightness between the umbra and the photosphere (see Figure 16.2).

16.1.1 Spot temperatures

How dark is the umbra relative to the photosphere? The answer depends on the wavelength: the shorter the wavelength, the darker is the intensity of the umbra $I_\lambda(*, \mu)$ compared to the intensity of the undisturbed photosphere at the same distance from the disk center $I_\lambda(\mu)$. At wavelengths of 4000 Å, large spots may have $I_\lambda(*, \mu)$ values which are 10 to 20 times smaller than $I_\lambda(\mu)$ (Bray and Loughhead, 1979; their Table 4.1). The contrast between umbra and photosphere becomes less pronounced as we observe at longer and longer wavelengths: around 1 micron, the spot intensity is almost half as bright as the photosphere. Model atmospheric fits to umbral radiation allow one to obtain the profile of temperature versus optical depth in the umbra. Expressing tempera-

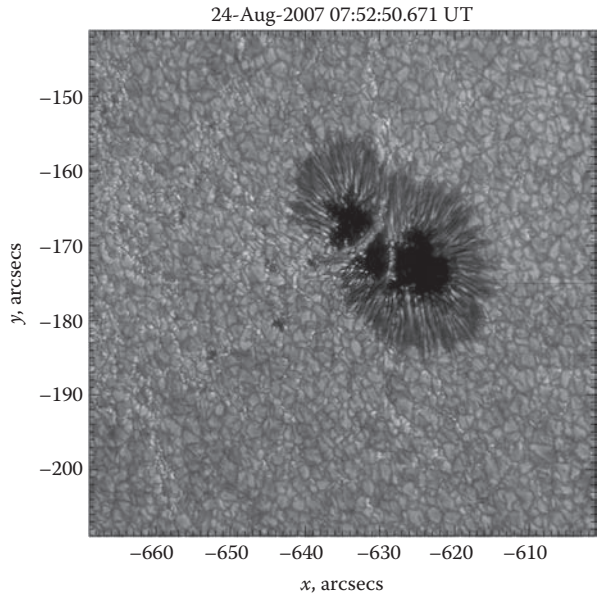


FIGURE 16.2: A sunspot, showing the dark umbra (center) plus striated penumbra, surrounded by undisturbed photosphere (containing granules). (From Dr. M. Mathioudakis, Queen's University Belfast, and the Swedish Solar telescope. With permission.)

tures in terms of the variable which appears in the Saha equation (Chapter 4, Section 4.2), $\theta = 5040/T$, the difference between $\theta(*, \tau)$ inside the spot at optical depth τ and $\theta(\tau)$ in the undisturbed photosphere at the same τ, can be as large as $\Delta\theta \approx 0.3$–0.4 at $\tau \approx 1$. Since the local temperature in the photosphere at $\tau \approx 1$ is close to 6000 K, i.e., $\theta(1) \approx 0.84$, this leads to $\theta(*, 1) \approx 1.14$–$1.24$. Thus, the temperature in the spot at $\tau \approx 1$ is 4100–4400 K. That is, the gas in the "photosphere" of the spot (i.e., around $\tau \approx 1$) is some 1600–1900 K *cooler* than the gas at equal optical depth in the photosphere. Estimates of the effective temperature of a spot are 4100–4200 K, i.e., almost 2000 K cooler than the photosphere. The fractional deficit in effective temperature in the umbra is about 30% compared to the undisturbed photosphere. Thus, the bolometric flux ($\sim T_{\rm eff}^4$) from the umbra is only about 25% of the photospheric flux. The blocking of some 75% of the photospheric energy flux indicates that there exists a severe perturbation of energy flow in the subsurface layers of the umbra. We shall see that magnetic fields hold the key to understanding this process.

16.1.2 Spot areas

How large are sunspots? The smallest ones, consisting of umbra only (with no penumbra), are called "pores," and have angular diameters of 2–5 arc sec.

The smallest pores are comparable in size to the sizes of individual granules. (We shall see in Section 16.7.2 that this is not a coincidence.) The largest pores have diameters of no more than 10 arc sec: once a pore grows to a diameter of $10''$ or more, a penumbra appears, and the feature becomes a *bona fide* sunspot. Large spots have areas which are often cited in units of "millionths of the visible hemisphere area." The largest spot ever recorded (Figure 16.1) had an area of $A \approx 6300$ millionths, i.e., it occupied about 0.6% of the visible surface. More commonly, the spots have areas of up to a few hundred millionths of the disk area: 95% of spots have $A \leq 500$ millionths (Bray and Loughhead, 1979; p. 229, their Table 6.1). Although spots are indeed striking phenomena when seen against the background of the solar surface, these numerical values of areal coverage remind us that an individual sunspot is truly a small-scale object in comparison with the Sun as a whole.

16.1.3 Spot numbers: the 11-year cycle

The number of spots on the surface of the Sun varies with time. The numbers increase and decrease in a nearly cyclical manner: sometimes there are many spots on the surface, while at other times, there are few (or even no) spots visible (see Figure 16.3).

In order to quantify this variability, observers have devised certain rules to count the number of spots on the Sun: the commonest system is referred to as the Zurich Sunspot Number, R_Z, which counts both single spots and groups of spots. When R_Z is small (or zero) for a period of time (a month or

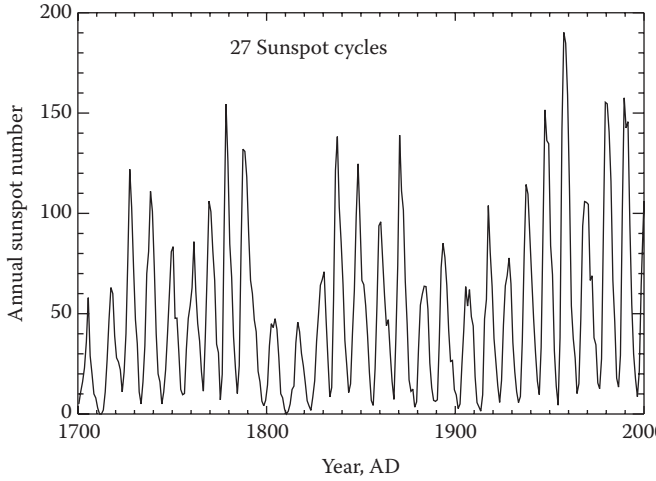


FIGURE 16.3: Number of sunspots, averaged over one year, plotted as a function of time over a span of three centuries. (The data were obtained from the Solar Influences Data Center at http://www.sidc.be/sunspot-data/)

more), the Sun is said to be in a stage of "minimum activity." When there are a great numbers of spots, and R_Z rises to values of 100–200 or more, the Sun is said to be at "maximum activity." These phases are also known as "solar minimum" and "solar maximum," respectively.

The interval of time between one "solar minimum" and the next is not constant: the interval can be as short as nine years, and as long as 12 years. The average length of the sunspot cycle (based purely on the sunspot counts) is about 11 years. In Figure 16.3, we see that recent "solar maxima" occurred around the years 1980 and 1990. The most recent "solar maximum" was observed in 2000–2001.

Something unexpected happens to the power output from the Sun in the course of the sunspot cycle (see Chapter 1, Figure 1.1). You might expect that when sunspots are *most* abundant, the Sun would emit *less* power. But this is *not* what is observed. Instead, the solar power output is observed to have *maximum* values in or around the years 1980, 1990, and 2000, when there are *most* sunspots on the surface. This surprising discovery emerged from spacecraft data during the last decades of the twentieth century: it was only in those decades that the precision of measurements of the total (bolometric) luminosity of the Sun, integrated over all wavelengths, became at least as good as 0.1%. (In previous years, observations from the ground were plagued by uncertainties arising from Earth's atmosphere, which blocks some 2% of the luminosity.) When such precision became available in instruments which also remained stable enough over an entire 11-year cycle, the data indicated that the solar luminosity has a *maximum* value when the number of spots is largest. The excess power output at solar maximum compared to solar minimum is of order 0.1% (see Chapter 1, Figure 1.1).

This is counterintuitive: when there are lots of spots, each dark umbra emits less power than the undisturbed photosphere, and therefore, one would expect the solar output to be a minimum. We shall return to this "problem" in Section 16.1.5.

16.1.4 Spot lifetimes

How long do spots live? Small spots may come and go in a matter of hours. Larger spots require days or weeks to reach maximum size, and days or weeks to decay. Typically, the lifetime is (roughly) proportional to the maximum area of the spot (A millionths): $T(\text{days}) \approx 0.1A$ (Bray and Loughhead, 1979; p. 229). Most spots decay by breaking up into smaller units, and these are then eroded over time by the "pounding" of the convective turbulence around the periphery.

16.1.5 Energy deficits and excesses

Sunspots are localized regions where the process of transporting the solar energy flux upward from the deep interior is subject to a severe deficit. As

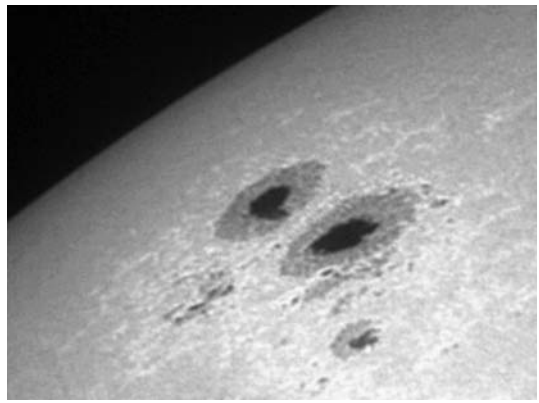


FIGURE 16.4: Faculae in a white-light image of sunspots near the limb of the Sun. Faculae, located in the vicinity of sunspots, appear slightly *brighter* than the undisturbed photosphere. (Photo by Damian Peach. With permission.)

long as the sunspot is present, upward heat transport is blocked in that location. The missing heat is trapped in the convection zone: as a result, the overall energy output from the Sun decreases as long as a large spot is present on the surface. Once the spot decays, the normal outflux of the Sun is restored.

However, sunspots are not the only contributors to perturbations in the solar energy output. Careful photometry of the photosphere *in the vicinity of sunspots* reveals the presence of multiple small features which are slightly *brighter* (by at most a few percent) than the undisturbed photosphere when viewed in white light. These bright point-like features (called "faculae") are much less obvious to the human eye than sunspots. In fact, even with telescopes, faculae are almost impossible to pick out near the center of the solar disk: the easiest place to observe them is in the vicinity of sunspots as the latter approach the limb (see Figure 16.4).

The excess of facular flux above the photospheric value helps to offset *some* of the flux deficit of a large sunspot in the vicinity. In the case of smaller spots, it is possible that facular emission actually *over*compensates for the spot deficit. This might explain why solar power output is slightly larger (by 0.1%) at epochs where spots are most numerous on the solar surface.

16.2 Chromospheric Emission

Another departure from spherical symmetry in the solar atmosphere, which we came across in the preceding chapter, appears when we observe the

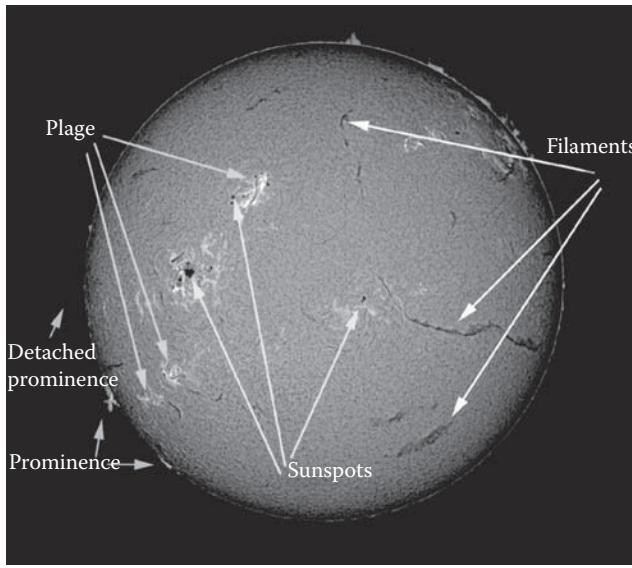


FIGURE 16.5: **(See color insert following page 202.)** The chromosphere: departures from spherical symmetry may include sunspots, plage, filaments, or prominences. (Copyright Peter Ward/Advanced Telescope Supplies. Used with permission.)

chromosphere. When the Sun is viewed in the light of a chromospheric line (such as the CaII K line, or Hα), the Sun is not spherically symmetric. Quantitatively, as we have seen, the differences in brightness between network and cell centers indicate that mechanical energy is being deposited in the network at a rate which exceeds the rate in the cell centers by a significant factor, possibly by as much as an order of magnitude.

Also when we observe the chromosphere, we find that there are features which have a close connection with sunspots in the photosphere (see Figure 16.5). In the chromosphere, the locations of sunspots (as determined from photospheric images) are found to be surrounded by regions of enhanced emission ("plages": see Chapter 15, Figure 15.3). Thus, sunspots are in fact only one component of a more extended physical structure, a "plage," in which chromospheric emission is enhanced. The (white-light) faculae are colocated with plage. The overall feature, including sunspots, plages, and faculae, is called an "active region."

Active regions do not appear randomly at all locations of the solar surface: from centuries of observation, it has been found that there are more favored areas (near the equator, at latitudes of no more than ± 35 degrees), and there are less favored (or even forbidden) areas (near the poles, at latitudes in excess of ± 35 degrees).

Also present in chromospheric lines are features called "prominences." These are structures which were first observed in emission standing above the limb of the Sun: they consist of material which appears to be suspended "in mid-air" (see Figure 16.5).

Prominences can also be viewed on the disk of the Sun when the latter is observed in Hα: in such cases, they appear as dark more-or-less ribbon-shaped features ("filaments": see Figure 16.5) located preferentially at positions where the surface magnetic fields change polarity. The filaments can be "quiescent," i.e., stationary for hours or days, but they can also, at the end of their lifetime, reveal rapid evolution as the prominence material "erupts" rapidly, either because of heating or expulsion of the material into the upper atmosphere.

16.3 Magnetic Fields: The Source of Solar Activity

Why does the Sun depart from spherical symmetry? The answer is well-established: it is due to the presence of magnetic fields. These fields give rise to a variety of observational phenomena. The umbrella term "magnetic activity" is used to cover the magnetically driven phenomena which are such a striking characteristic of the Sun at times. Under the term "magnetic activity," we include the presence of sunspots, faculae, the chromospheric network, and prominences, all of which are more-or-less long-lived phenomena which can be regarded as quasi-stationary in nature. On the other hand, "magnetic activity" also includes phenomena which are by no means stationary, such as flares and coronal mass ejections (CMEs). Both of the latter involve highly time-dependent processes which disturb the solar atmosphere in striking ways, giving rise (at times) to "fireworks displays" which involve the most energetic phenomena in the solar system.

Our aim in this chapter is to describe, in terms of physical processes, how magnetic effects give rise to a rich variety of phenomena in the Sun.

Before discussing the general properties of magnetic fields and their interactions with plasma, and in order to keep the discussion rooted in the Sun, we start with what the observations tell us about the magnetic fields themselves. We need first to understand how astronomers measure the strength of the fields in solar features of various kinds. Once we have a feel for the orders of magnitude of the field strengths in various features, then we will turn to the physics to determine which processes are most relevant in the various phenomena.

To set the scale, we note that the strongest fields on the solar surface occur in the umbrae of large sunspots: the maximum field strength (in Gauss) in a sunspot with area A millionths (Section 16.1.2) is $B \approx 3700\,A/(A+66)$ (Bray and Loughhead, 1979; p. 207). That is, large spots have fields with strengths of a few kilogauss.

16.4 Measurements of Solar Magnetic Fields

There are several approaches to measuring solar fields. One involves remote sensing using optical photons: we observe certain photons coming from a certain feature on the Sun, examine the spectral and polarimetric properties of the photons, and infer the strength and direction of the field in the feature under observation. A second approach uses remote sensing using radio photons: polarization data again contains information on strength and direction. A third approach involves direct measurements of the field *in situ* in the plasma which streams out of the Sun (the "solar wind": see Chapter 18) into interplanetary space, and then extrapolate back to infer the fields at the Sun. Let us consider these approaches.

16.4.1 Remote sensing of solar magnetic fields: optical data

To measure these fields, solar astronomers use a technique which seeks to identify changes in the shape of a spectral line when a field is present.

16.4.1.1 Zeeman splitting

How is a spectral line altered in the presence of a magnetic field? To answer this, let us recall what happens in the *absence* of the field. Each spectral line involves the transition of an electron from one atomic energy level E_1 to another level E_2. In the absence of external magnetic fields, the energy levels are determined by atomic structure. Radiation from an atom is spherically symmetric: there is no preferred direction in the problem. When the atom is observed from any direction, what is observed is a single line with frequency $\nu_o = (E_2 - E_1)/h$, i.e., a single line with a wavelength $\lambda_o = c/\nu_o$.

Now introduce an external magnetic field. Two aspects of the situation change. First, the energies of the atomic levels are altered: this will cause the lines to shift in wavelength. Second, the photons which emerge have properties which are no longer spherically symmetric: observers who are situated in different viewing positions see different spectra.

To understand how magnetic processes affect atoms, we first refer to a basic result of magnetostatics: what happens when one places a magnetic moment in a field? Recall that when iron filings are sprinkled on paper near a bar magnetic, a clear pattern is seen: each iron filing, which is a small magnet in itself, with its own magnetic moment, aligns itself with the local magnetic field lines. Now, every electron has an intrinsic "spin" (with angular momentum $\hbar/2$), and associated with this spin is a magnetic moment $\mu_B = e\hbar/2m_e c$. In a magnetic field B, an electron can settle into one of two states: one, with μ_B *parallel* to the external field, the second, with μ_B antiparallel to the external field. In one of these states, the electron *gains* an energy $+ \mu_B B$,

while in the other state, the electron energy is *reduced* by $-\mu_B B$. Thus, an electron which initially was in an atomic level with a particular energy E_1 now finds that the level "splits" into two levels, with energies $E_1 + \mu_B B$ and $E_1 - \mu_B B$.

What will we observe if we detect the photons which emerge from the above atom? The answer depends on the direction from which we make the observation. Suppose we choose to make the observations parallel or antiparallel to the external magnetic field: that is, we choose to "look straight down the field." Let us also suppose for simplicity that the energy level E_2 does not undergo any splitting in a magnetic field (atomic levels with this property do exist). In that case, what we see is the following: the original single line at a frequency ν_o, i.e., at wavelength $\lambda_o = c/\nu_o$, is now seen to consist of two lines (a "doublet"), at frequencies $\nu_o \pm \Delta\nu$, where $\Delta\nu = \mu_B B/h$.

This conversion of a single line into a doublet as a result of a magnetic field is called Zeeman splitting, after the discoverer (see Figure 16.6).

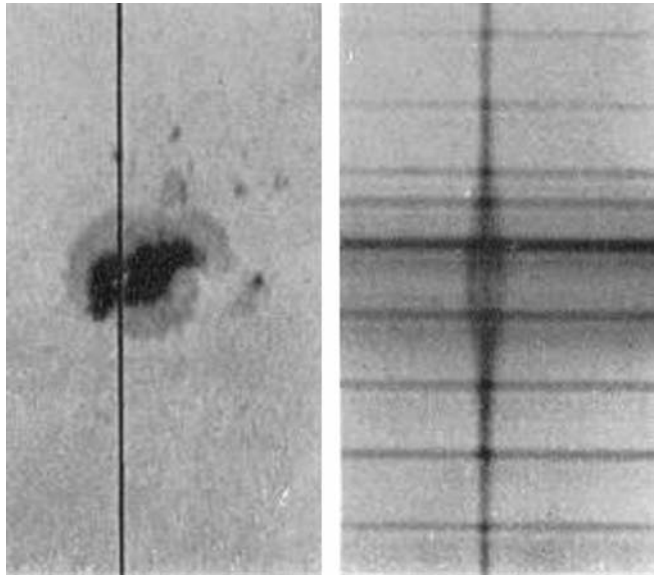


FIGURE 16.6: Zeeman splitting of spectral lines in a magnetic region on the Sun. The vertical slit of the instrument is located as shown in the left-hand image: the slit overlaps with the umbra of a sunspot. On the right-hand side, a (vertical) spectral line which is single in the undisturbed Sun (at top and bottom), become multiple when the slit overlaps the umbra, where magnetic fields are strong. (Image downloaded from the website of High Altitude Observatory, a division of the National Center for Atmospheric Research, funded by the National Science Foundation.)

The wavelengths of the two components of the doublet are $\lambda_o + \Delta\lambda$ and $\lambda_o - \Delta\lambda$, where $\Delta\lambda = \lambda_o^2\ \Delta\nu/c$. Inserting the value for μ_B, we find that in the presence of a magnetic field, the wavelength shift is $\Delta\lambda = \text{const}B\lambda_o^2$ where the constant, equal to $e/4\pi m_e c^2$, has the numerical value of 4.9×10^{-5} cm Gauss^{-1}cm^{-2}. For convenience, if we express wavelengths in units of Å($= 10^{-8}$ cm), we find

$$\Delta\lambda = 4.9 \times 10^{-13} B(G)\lambda_o^2 \tag{16.1}$$

Equation 16.1 is valid for the simplest case, when only the electron's magnetic moment is responding to the external magnetic field. This is referred to as the "normal" Zeeman effect.

The actual Zeeman effect in "real atoms" differs slightly from the above formula because there are other sources of magnetic moments. For example, when an electron orbit has a finite angular momentum, that orbit also has an associated magnetic moment. The process by which an external field interacts with an orbital magnetic moment is quantitatively different from the process by which the external field interacts with the electron spin. As a result, the above expression is only part of the story of the wavelength shift for any given transition. Each "real" transition has a factor g_L associated with it (the so-called Lande g-factor), and the right-hand side of the above expression must be multiplied by g_L. For transitions of various kinds, the numerical value of g_L may be as small as zero, or may be as large as (roughly) three. Moreover, depending on the atomic structure, rather than splitting into two components, a line may split into multiple components, leading to what is sometimes referred to as the "anomalous" Zeeman effect.

To give a numerical example, consider an atom which has a line in the visible part of the spectrum, at (say) $\lambda_o = 5000$ Å. In a field of 3000 G (typical for a sunspot umbra), and assuming $g_L \approx 1$, we find that $\Delta\lambda \approx 0.037$ Å. Therefore, in order to detect a clean splitting of the lines in a sunspot, observers are required to use instruments with a resolving power (defined by the ratio of $\lambda_o/\Delta\lambda$) of more than 100,000. Achieving such a high resolving power requires careful attention to instrumental design.

16.4.1.2 Zeeman polarization: the longitudinal case

The splitting of a single line into two (and only two) components occurs when we make observations *along* the magnetic field. This is referred to as "*longitudinal* Zeeman splitting."

Observations show that the two lines of a doublet are not merely different in their wavelength: they also differ in polarization: the two lines are circularly polarized in opposite senses. To see why this is so, consider a spectral line in which the upper level is not affected by the field (i.e., its Lande g-factor = 0), but the lower level undergoes normal Zeeman splitting. In the lower level, the angular momentum (spin) of the electron has a component along the field of $\pm\hbar/2$, depending on whether the electron spin is in the "up" or "down" position relative to the field.

To understand how a passing photon interacts with an electron in the split level, we note that the angular momentum (spin) of a circularly polarized photon is $\pm\hbar$, depending on whether the photon has right- or left-hand polarization. Consider an electron which is sitting in the "down" position, with spin $-\hbar/2$. If a photon with left-hand polarization (i.e., spin $-\hbar$) passes by that electron, the electron cannot interact with the photon, because if the interaction occurred, then the electron would have to absorb the photon's spin, add it to its own, and enter a state with spin of $-3\hbar/2$. Such a state is not available to the electron: the only available states have spins of $\pm\hbar/2$. As a result, the electron simply ignores the left-hand polarized photon, and the photon passes through unperturbed. But now consider the case where the passing photon is right-hand polarized, i.e., the photon has spin $+\hbar$. Now, the electron *can* interact with the photon, adding the photon's spin $+\hbar$ to its own $(-\hbar/2)$, and ending up with spin $+\hbar/2$. Such a spin *is* allowed: the electron simply transitions to the "up" position. Thus, an electron in the "down" position preferentially absorbs right-hand polarized photons out of the beam. The remaining photons, i.e., the left-hand circularly polarized photons, pass through and reach the observer: the observer therefore, sees left-hand circular polarization as the dominant component of the absorption feature.

How do we know that circularly polarized photons interact with electrons in this way? Because experimental confirmation is available in the laboratory. Specifically, in a thin sheet of iron, the magnetization can be chosen so that the elementary magnets in the iron all tend to be aligned in one particular direction. Then if a circularly polarized photon passes through the sheet, the photon will be scattered preferentially if its polarization has the correct sign to flip an aligned magnet. Of course, such a test does not work with optical photons: the iron sheet prevents them from passing through. But if one uses high energy photons (gamma-rays), then these *can* pass through the iron, and the aligned magnets can be flipped. In fact, this property of photon-electron interactions played a role in establishing the existence of parity violation in weak interactions: a key experiment was performed by Goldhaber et al. (1958).

Returning now to our Zeeman doublet, we recall that a "down" electron has a specific energy shift, depending on the direction of the field. Let us consider a case in which the field points *toward* the observer. Recall that the magnetic moment of the electron is proportional to the electron spin, and the proportionality factor depends on the (negative) charge of the electron. As a result, when the electron spin is sitting in the "down" position, the magnetic moment is sitting "up" relative to the field. Therefore, the electron has an excess energy $+\mu_B B$ relative to the unperturbed state. This means that the lower energy level is no longer the unperturbed value E_l, but a larger value: $E_l + \mu_B B$. The upper energy level is (by assumption) still at the unperturbed value E_2. As a result, the frequency of the transition is no longer equal to $(E_2 - E_1)/h$, but takes on a lower value. A lower value for the

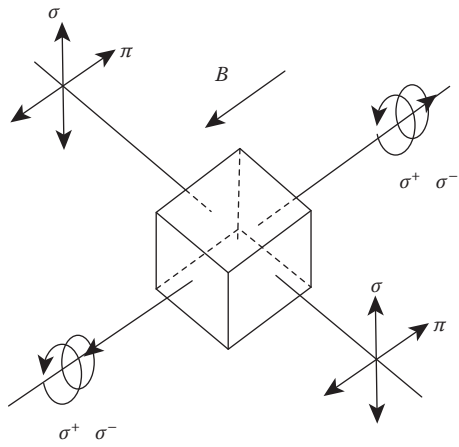


FIGURE 16.7: When observations are made along the direction of the field (vector B), Zeeman splitting leads to components which are *circularly* polarized. When observations are made perpendicular to the field, the split components are *linearly* polarized. (From PHYWE Systeme GmbH and Co, Gottingen, Germany. With permission.)

frequency means a longer wavelength for the photon: therefore, we are discussing the *redward* shifted component of the Zeeman doublet, at wavelength $\lambda_o + \Delta\lambda$.

We conclude that when the field in a certain location of the Sun is pointing *toward the observer*, and we observe a Zeeman line in absorption, the *redward* component will be observed to be circularly polarized in the *left*-handed sense. Analogous arguments can be applied to the "up" electron, showing that the *blueward* component, when seen in absorption, will be *right*-hand polarized.

Conversely, if we observe a region where the solar field is pointing *away from the observer*, the blueward component of the Zeeman doublet will be left-hand polarized (Figure 16.7).

This provides a powerful diagnostic of the *direction* of magnetic fields in the Sun. If I point my telescope at a particular umbra on the Sun, and I observe a Zeeman doublet in absorption, and find that the blueward component, at wavelength $\lambda_o - \Delta\lambda$, is (say) right-hand polarized, then I know that the field lines are pointing toward me. Such fields are directed outward from the Sun. And I also know how strong the field is along the line of sight, by measuring the shift $\Delta\lambda$ and inserting this in Equation 16.1.

16.4.1.3 Zeeman polarization: the transverse case

Magnetic fields in the solar atmosphere are complicated in spatial structure. As a result, although there are certainly possibilities for observing

"straight down the field," this is not always the case. In other locations, the line of sight will turn out to be perpendicular to the field lines.

If I observe the photons which propagate in directions perpendicular to the external magnetic field, the results are different from the longitudinal case. In the perpendicular case, known as the "transverse Zeeman effect," we observe not two, but three components. Two components are still found at wavelengths $\lambda_o + \Delta\lambda$ and $\lambda_o - \Delta\lambda$ (as before), but now there is also a third component at the undisplaced wavelength λ_o. In the case of the transverse Zeeman effect, the polarizations of the components are observed to be *linear* (rather than circular): the undisplaced line is linearly polarized parallel to the field, while the two shifted components are polarized perpendicular to the field.

Because of the polarization properties, by (i) measuring the splitting $\Delta\lambda$, (ii) counting components of the split line, and (iii) measuring polarizations, we can determine the strength of the field, and also the angle of the magnetic field relative to our line of sight. This is a powerful diagnostic for the properties of magnetic fields, especially in the Sun where the fields in an active region can be very complicated, with many changes from one location to another.

16.4.1.4 Babcock magnetograph

The polarization properties of the Zeeman effect provide a practical technique for measuring weak solar magnetic fields. No longer do we have to build an instrument that can cleanly separate the components of the Zeeman doublet. Instead, solar observers (starting with Babcock, 1953) use the trick of observing in one polarization at a certain wavelength which is shifted by $+\Delta\lambda$ on one side of line center, and in the opposite polarization at a wavelength which is shifted by $-\Delta\lambda$ on the other side of line center. By carefully choosing $\Delta\lambda$ so that the observations are made on the steepest part of the absorption line profile, even a slight amount of Zeeman splitting can then be detected. This allows detection of fields as weak as tens of Gauss on the Sun.

An instrument which takes advantage of the circular polarization properties of the Zeeman doublet is referred to as a Babcock magnetograph. It is useful for measuring the *longitudinal* field at all points of the solar disk. Such instruments have been in widespread use for decades for daily monitoring of solar magnetic fields.

Instruments which also measure linear polarization are called "vector magnetographs." These require more extensive data analysis in order to interpret the signals. They are typically used for analysis of the magnetic fields in selected active regions.

16.4.1.5 Orderly properties of sunspot fields

Results from Babcock magnetographs reveal a high degree of order in the fields on the Sun. We have already mentioned that spots typically appear in pairs, a leader and a follower. When a magnetograph is applied to each spot,

a highly ordered behavior emerges. At any given time, essentially all leader spots in (say) the Sun's Northern hemisphere are observed to have the same magnetic polarity. That is, all leader spots in the Northern hemisphere are found to have fields which (say) point *toward* the observer. At the same time, all follower spots in the Northern hemisphere will exhibit fields which point *away from* the observer. And simultaneously, in the Southern hemisphere, the situation will be precisely reversed: the fields in leader spots will point *away from* the observer. When the same observations are repeated 11 years later, the fields are found to be reversed: leader spots in the Northern hemisphere will now be found to have fields pointing *away from* the observer. These rules are referred to as Hale's polarity law, from the observer who discovered the effect. Thus, the true solar cycle is actually 22 years long.

16.4.2 Remote sensing of solar magnetic fields: radio observations

The Sun's corona (Chapter 17) emits radiation over a broad band of radio wavelengths. The source of the radio emission depends on the local conditions, and on the frequency. At frequencies which are in the microwave band, between (roughly) 1 and 20 GHz, two principal emission mechanisms contribute to the radio flux. One is a free-free process, where electrons are accelerated when they pass close to ions in the coronal plasma (see Chapter 3, Section 3.3.1). The second has to do with electrons gyrating in a magnetic field: circular motion involves acceleration, and when a charged particle accelerates, it emits radiation. This "gyro-emission" has a preference to be emitted at certain frequencies, namely, at the "gyrofrequency" ν_B (see Section 16.6.1) and its harmonics.

As we shall see (Section 16.6.1), the value of ν_B depends only on the field strength in the plasma which emits the radiation. Therefore, if ν_B can be derived from observations, we can determine the field strength in the coronal plasma.

Polarization again plays an important role in determining coronal magnetic properties. In some active regions, the free-free emission is observed to be circularly polarized. That is, when the radio flux is measured at a certain frequency ν in right-hand polarization F_R, this flux differs from the flux at the same frequency in left-hand polarization F_L. The degree of polarization $d_p = (F_R - F_L)/(F_R + F_L)$ is observed to have values which may be as large as tens of percent. It can be shown theoretically (e.g., Lee, 2007) that d_p is simply proportional to the ratio of ν_B/ν. Therefore, a measurement of d_p at frequency ν can be converted to $\nu_B \approx d_p\nu$, and thence to the field strength B in the coronal emitting region.

Values of B in the corona span a wide range: in a survey of ten active regions, Schmelz et al. (1994) reported B ranging from as low as 55 G to almost 600 G. Over certain sunspots, the coronal field strength has been reported to be as large as 1800–2000 G (Lee, 2007).

16.4.3 Direct measurements in space: the global field of the sun

Clearly, when we use the term "direct measurement," it is not a question of measuring the fields *in* the atmosphere of the Sun itself. Instead, we rely on a particular property of solar material, i.e., magnetic fields which were at one time situated in the Sun's atmosphere are carried out into interplanetary space by the expanding solar wind. (We shall see below that fields and plasma are effectively "frozen together.") Measurements of the magnetic field *in situ* in interplanetary space are the nearest we can get to "direct measurements" of solar fields. If we can measure the field strength in space, at a certain distance from the Sun (e.g., near 1 AU), we may be able to calculate how strong the fields are back at the surface of the Sun.

Measurements of the magnetic field strength in interplanetary space have been made since the 1960s, when spacecraft first escaped beyond the confines of the Earth's magnetic field and sampled the interplanetary magnetic field (IMF). At first, when there was no clear knowledge as to how strong the IMF might be, some of the early magnetometers were so swamped by the background of magnetic fields caused by the spacecraft itself that they could not reliably identify the IMF. It was soon realized that the detectors had to be sensitive enough to measure fields of order 10^{-5} Gauss, i.e., 1 nanotesla. (For convenience in discussing IMF's, the nanotesla is referred to as 1 gamma ($1\,\gamma$). For comparison, the magnetic field at the Earth's magnetic North pole is $60{,}000\,\gamma$.) Detection of such weak fields required paying special attention to making sure that electric currents in the spacecraft itself did not generate fields which would swamp the IMF's.

It is not always a simple matter to extrapolate the field back to the Sun: one needs to make allowance for certain properties of the solar wind (Chapter 18). Allowing for these, it is found that much of the IMF emerges from the polar regions of the Sun. The fact that the North and South poles of the Sun contain magnetic fields is strongly suggested by certain images of the solar corona, especially those which are taken close to solar minimum. On August 1, 2008, an eclipse of the Sun occurred (Figure 16.8), and on that day, there were no sunspots visible on the surface. (For evidence of the lack of sunspots on that day, see http://sidc.oma.be/news/105/welcome.html.) Thus, the Sun contained none of the strong fields associated with active regions and sunspots on August 1, 2008. This gives the best opportunity to detect the weaker fields associated with the Sun as a whole. In Figure 16.8, the upper and lower parts of the Sun exhibit bright and dark streaks which are reminiscent of how iron filings line up when they are scattered near the North and South poles of a bar magnet.

Extrapolation of IMF data indicate that the radial component of the solar magnetic field near the North and South poles may range from six to 12 G (Hundhausen, 1977). These numbers are subject to revision if the solar wind properties have not been incorporated correctly. The solar polar fields are

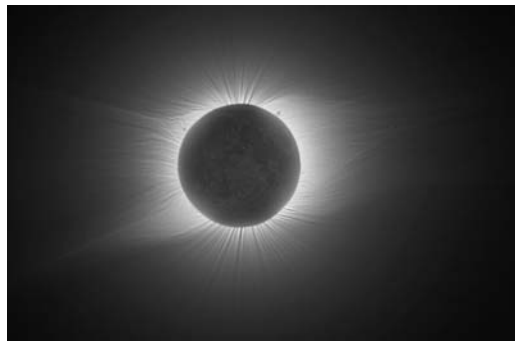


FIGURE 16.8: **(See color insert following page 202.)** Image of the Sun obtained during the eclipse of August 1, 2008, when sunspots were absent. The Sun is hidden by the Moon, and some of the Moon's surface features are faintly illuminated by sunlight scattered off the Earth. Notice the nearly-radial streaks emerging from the North and South poles of the Sun (top and bottom of the image): these are indications of near-radial magnetic fields at North and South poles. (From Dr. H. Luethen. With permission.)

stronger than the fields at the Earth's magnetic poles by up to one order of magnitude.

The magnetometers which have been flown on spacecraft are such that not only can the *magnitude* of the IMF be measured, but also its *direction*. Such measurements indicate that at any instant of time, there is a preferred direction for the field at the Sun's North pole, and simultaneously the field at the South pole has the opposite direction. The preferred directions at the two poles remain constant for years on end. But at intervals of time which range from as short as (about) nine years to as long as (about) 12 years, the field directions at the solar poles reverse sign. The polar reversals do not always occur in coincidence: they may be separated by periods of months or maybe a year. During such periods, both poles of the Sun have the same magnetic polarity. However, for 90% of the cycle, there is a clearly defined polarity for the global field of the Sun. The fact that the global polarity switches in (about) 11 years indicates that the true magnetic cycle of the Sun has a period of (about) 22 years.

The fields of 6–12 G which exist at the poles of the Sun, and which extend far out into interplanetary space, represent the *global* magnetic field of the Sun. The fact that the field reverses sign every 9–12 years indicates that the global magnetic field of the Sun is subject to periodic behavior. This is quite different from the Earth's magnetic field, which, although not strictly constant, nevertheless retains a more-or-less constant value over time-scales of many thousand of years.

Can the polar fields of the Sun be recorded by Zeeman techniques? Fields with strengths of at most 12 G are so weak that they are close to the limit of

observability for Babcock magnetographs. Moreover, fields in the polar regions of the Sun tend to be radially directed. As a result, observations from Earth see these fields as mainly transverse to the line of sight. Therefore, Babcock magnetographs, which are sensitive to the field components along the line of sight, are not ideally suited to detecting the global polar fields of the Sun. Nevertheless, estimates of mean polar fields are reported as part of the output from solar magnetic observatories, e.g., the Wilcox Solar Observatory (see http://wso.stanford.edu/gifs/Polar.gif). The polar fields are found to have strengths of at most 2–3 G. These are certainly weaker than the 6–12 G estimates mentioned above: perhaps the solar wind corrections should be re-examined, or perhaps the Wilcox data are averaging over polar field strengths in different ways. Whatever the source of the discrepancy, the Wilcox data confirm the space-based discovery that the polarities of the Sun's North and South poles reverse every 11 years (or so).

16.5 Empirical Properties of Solar Magnetic Fields

We now have information about two apparently distinct components of magnetic fields in the Sun. One is global and quite weak, while the others (especially in spots) are highly localized and very strong. The localized fields are stronger than the global field by factors of at least 100, and in some cases by almost 1000. In the umbra of a sunspot, the direction of the field is found to be essentially vertical, i.e., perpendicular to the solar surface. As we move from the umbra outward into the penumbra, the field lines are observed to tilt more and more toward a horizontal direction: the horizontal fields give rise to dark and bright "striations" (or "zebra-stripes") which are the hallmark of the penumbra when it is observed at high angular resolution (Figure 16.2).

The active regions surrounding sunspots have fields of hundreds of G. In the quiet Sun, average magnetic fields (averaged over a field of view of, say, 10 arc sec^2) are weaker, and those fields are not uniformly distributed: the field is highly clumped into compact flux ropes. In the center of each flux-rope, the field may reach values of 1–2 kG. This leads to the graphic term "pin-cushion" which is sometimes used to describe the nature of the magnetic field structure in the quiet Sun (Parker, 1974). Given a field of view of an instrument, the average magnetic field strength that will be measured for that area depends on how many "magnetic pins" happen to lie within the field of view. Active regions are locations where there are enhanced areal densities of the "magnetic pins." Sunspot umbrae are locations where the areal density of "magnetic pins" reaches a maximum: a very efficient process sweeps in, and holds together, the vertical magnetic flux ropes which are the defining characteristic of an umbra.

Both for the global field and for the localized strong fields, there is a 22-year cycle in the direction of the magnetic fields. For several years prior to solar maximum, there is found to be a close correlation between the direction of the global field in (say) the Northern hemisphere and the preferred polarity of leader spots in the Northern hemisphere at the same time. These empirical results indicate that there is a close physical connection between the Sun's weak global fields and the strong fields which appear from time to time on the surface in highly localized structures.

In order to understand why such a connection exists, we need to understand how magnetic fields and the gas in the Sun act and react upon each other. To achieve this understanding, we need to take a long step back from the Sun, and come down to the level of individual charged particles. We need to spend considerable effort (in Section 16.6) on the physical processes which govern the interactions of magnetic fields and charged particles before we apply those processes to specific phenomena on the Sun (Section 16.7). Investment of effort at this stage will pay off well when we return to the rich variety of solar magnetic phenomena.

16.6 Interactions between Magnetic Fields and Ionized Gas

To understand the physical process whereby magnetic fields and gas interact with each other, we need to understand the forces that magnetic fields exert on charged particles.

In a gas which is electrically neutral, such as the air that we all breathe on Earth to stay alive, magnetic fields have no interesting dynamical effects. The motions of the gas on Earth (i.e., the winds) are not at all affected by the Earth's magnetic field. But things are very different in the Sun's atmosphere. It is the fact that the gas in the Sun's atmosphere is electrically charged that opens up the possibility of interesting interactions between field and gas.

16.6.1 Motion of a single particle

A particle with electric charge e, when placed in an electric field $\mathbf{E}$ is acted on by a force $e\mathbf{E}$. When a charged particle moves with (vector) velocity $\mathbf{V}$ in a (vector) magnetic field $\mathbf{B}$, the particle experiences a motional electric field $\mathbf{E}_m = (1/c)\mathbf{V} \times \mathbf{B}$. As a result, the charged particle is subject to the Lorentz (vector) force $(e/c)\mathbf{V} \times \mathbf{B}$. Here, we use boldface to denote that a quantity is a vector, with magnitude and direction. The magnitude of $\boldsymbol{B}$ is written as B. The symbol "$\times$" between two vectors denotes the cross-product of the two vectors: the magnitude of the cross-product $\mathbf{V} \times \mathbf{B}$ is equal to $|\mathbf{V}||\mathbf{B}| \sin\theta$, where θ is the angle between the vectors. The Lorentz force acts in a direction

which is perpendicular to the particle velocity, and also perpendicular to the magnetic field.

As regards the units, if we express the charge e in electrostatic units, and $\mathbf{B}$ in Gauss, then the Lorentz force is in units of dynes.

For a particle of mass m, the equation of motion in a magnetic field is

$$m\frac{d\mathbf{V}}{dt} = \frac{e}{c}\mathbf{V} \times \mathbf{B} \tag{16.2}$$

The solutions of Equation 16.2 have certain distinct properties. First, if the particle is moving in a direction that is parallel, or antiparallel, to the field, then the right-hand side has zero value. This means there the magnetic field exerts no force on the particle. As a result, there are no constraints on the motion of a charged particle along a magnetic field. If a field line stretches from the Sun to the Earth, an electron or a proton can propagate freely along that line between Sun and Earth, and back again.

Second, in all other cases, when there is a component of particle velocity that is *perpendicular* to the field, then the Lorentz force on the particle is finite in magnitude, and acts in a specific direction. Suppose the field is in the $+z$-direction, and the particle moves exactly perpendicular to the magnetic field, in (say) the +y-direction. Then taking the vector product $\mathbf{V} \times \mathbf{B}$, we find that the force acts in the $+x$-direction. If the electric charge is positive, then the particle motion will be deflected toward the x-direction. Once an x-component of velocity occurs, the Lorentz force $\mathbf{V} \times \mathbf{B}$ will contain a component in the $-y$-direction. This will eventually reduce the y-velocity to zero, at which point the x-velocity will have its maximum magnitude. However, the y-velocity will not stop there. Instead, the y-component of the velocity will increase in the $-y$-direction, causing the x-component of the force to become negative. This will cause the x-velocity to decrease, eventually falling to zero, at which point we are back to the initial condition.

The net effect of the Lorentz force is that the particle moves in a circular path in the $x - y$ plane. Because the electric charge enters into the Lorentz force, a positively charged particle moves along the circle in one direction, while a negatively charged particle moves along the circle in the opposite direction. In both cases, the Lorentz force is directed toward the center of the circle.

The essential aspect of the effects of magnetic fields in the solar atmosphere can be understood by considering the following question: how large is the circular path of a particle in the magnetic fields which exist in the solar atmosphere?

The answer to that question comes by balancing the forces that act on a particle which is moving in a circular orbit: in this case, the two forces are the centrifugal force and the Lorentz force.

Balancing these forces in an orbit of radius r_g we find

$$\frac{|e|}{c}|\mathbf{V}||\mathbf{B}| = \frac{m|\mathbf{V}|^2}{r_g} \tag{16.3}$$

where we indicate the magnitude of the electron charge, with the understanding that opposite charges circulate in opposite directions. Solving the above equation leads to an expression for r_g:

$$r_g = \frac{mc|\mathbf{V}|}{|e||\mathbf{B}|} \tag{16.4}$$

The particle is confined to a circular motion, and cannot depart from the field line in the perpendicular direction by a distance of more than r_g. The subscript g denotes that we are dealing with the "gyro-radius," for it identifies the size of the orbit on which the particle is forced to "gyrate" around the field.

Equation 16.4 refers to nonrelativistic motion. In the case of an ultrarelativistic particle, where the time-dilation factor γ is $\gg 1$, $|\mathbf{V}|$ can be set equal to c, the KE equals γmc^2, and the mass m in Equation 16.4 must be replaced by γm.

Recalling that motion parallel and antiparallel to the field is unconstrained, the true "orbit" of a charged particle in a magnetic field is a helix: the particle is free to move up and down the field line, but cannot depart form the field line by more than r_g.

A key aspect of Equation 16.4 is that r_g is proportional to the particle's speed $|\mathbf{V}|$. As a result, the time required for a charged (nonrelativistic) particle to traverse one gyration circle is $t_B = 2\pi r_g/|\mathbf{V}|$, which is *independent* of the speed of the particle. The associated frequency $\nu_B = 1/t_B = eB/2\pi mc$, called the "gyro-frequency" depends only on B. If a value can be estimated for ν_B in any locality on the Sun (such as an active region), that will give us information about the field strength in that active region (see Section 16.4.2).

Now we come to the question which is at the heart of understanding magnetic effects in the Sun. How large is a typical r_g in the solar atmosphere?

Consider a thermal proton in the photosphere, where $T = 6000$ K. The mean thermal speed of the proton is $|\mathbf{V}| \approx 10^6$ cm sec^{-1}. Inserting proton mass and charge, we find that in a field of B Gauss, $r_g(\text{cm}) \approx 100/B$. Thus, in a region of the photosphere where the field has a strength of 10 G, protons are constrained to gyrate no more than 10 cm away from the field line. In a field of 1000 G, they can move only 1 mm away. Electrons are even more tightly constrained: if we consider thermal electrons, the above gyro-radii must be reduced by factors of 43.

In the corona, the temperatures are larger, of order 10^6 K (see Chapter 17). This leads to gyro-radii which are about 10 times larger than in the photosphere. Proton gyroradii in the corona, in regions with $B = 10$ G, are of order 100 cm.

The most striking aspect of these gyroradii is how small they are compared to any of the relevant length-scales in the Sun: the solar radius ($\approx 10^{11}$ cm), the scale height in the atmosphere ($\approx 10^7$ cm in the photosphere), or the size of a granule ($\approx 10^8$ cm). In all cases, the gyroradii of thermal protons and electrons are miniscule compared to any of the relevant length-scales in the Sun.

The conclusion is clear and inevitable: *charged particles in the Sun cannot move freely in directions which lie perpendicular to the magnetic field.* In the solar atmosphere, *the ionized gas is "tied" tightly to the field lines.*

In principle, neutral gas is not affected by the field: at first sight, this might be taken to mean that in the photosphere (where hydrogen is at least 99.9% neutral), the fields might have little effect. But this is not the case. The presence of even a few ions and electrons gives the field "something to hold on to": and then the charged particles communicate the magnetic forces to neutrals in collisions. If there is only one ion for every 1000 neutrals, then that ion has to collide eventually with 1000 neutrals in order to pass on the magnetic forces which the ion is responding to. There may be some local spatial separation between neutrals and ions as the collisions do their work, but the separations (referred to as "ambipolar diffusion") occurs over length scales which are small compared to the length scales of features in the solar atmosphere, such as granules. For all intents and purposes, even in the photosphere, where the gas is more neutral (in an electrical sense) than anywhere else in the Sun, the gas is still "tied" to the field.

16.6.2 Motion of a conducting fluid

So far, we have considered the interaction between a magnetic field and a single charged particle. Now we move up to the macroscopic case, where we consider a fluid (plasma) composed of many individual charged particles. In such a fluid, electrons and ions can move in different directions: the result is that, in the plasma, a finite current can flow. The current density $\mathbf{j}$ is given by $e(N_i\mathbf{V}_i - N_e\mathbf{V}_e)$ where $\mathbf{V}_i$ and $\mathbf{V}_e$ are velocities of ions and electrons, and N_i and N_e are number densities of ions and electrons in the plasma.

Each cubic cm of the solar atmosphere contains N_i ions, on each of which the Lorentz force equals $+(e/c)\mathbf{V}_i \times \mathbf{B}$. Each cubic cm of the solar atmosphere also contains N_e electrons: on each electron, the Lorentz force equals $-(e/c)\mathbf{V}_e \times \mathbf{B}$. The equation of motion for 1cm^3 of solar material, with total mass $\rho = N_i m_i + N_e m_e \approx N_i m_i$ and bulk velocity $\mathbf{V}$, now includes not only the terms with which we are familiar from hydrodynamics (pressure gradient, and gravity), but also a term which describes the Lorentz force acting on that cubic cm:

$$\rho \frac{d\mathbf{V}}{dt} = -\nabla p - \rho \mathbf{g} + \frac{1}{c}\mathbf{j} \times \mathbf{B} \tag{16.5}$$

This equation describes the dynamical effects that a magnetic field exerts on the medium. Let us look in detail at the nature of the magnetic forces: they have interesting properties which will help us to understand why the solar atmosphere contains a variety of magnetic phenomena.

16.6.2.1 Magnetic pressure and tension

The gas pressure p enters into Equation 16.5 because the gradient of p exerts a well-known force on the gas. This is true even in the absence of

magnetic effects, such as in the Earth's atmosphere. (Localized winds blow from high to low pressure.) On small length scales, the force due to ∇p can be considered isotropic without serious error. However, this is not true of the Lorentz force. The term in $\mathbf{j} \times \mathbf{B}$ in Equation 16.5 can be rewritten in a way that brings out the fact that a magnetic field also exerts a force which is not the same in all directions.

To see this, we use Maxwell's equation curl $\mathbf{B} = (4\pi/c)\mathbf{j}$ to replace $\mathbf{j}$ in the Lorentz force in Equation 16.5. Then the Lorentz force becomes $(1/4\pi)$ curl $\mathbf{B} \times \mathbf{B}$. This can be rewritten, using vector identities, as the sum of two components $\mathbf{L}_1 + \mathbf{L}_2$, where $\mathbf{L}_1 = -\nabla(B^2/8\pi)$, and $\mathbf{L}_2 = (\mathbf{B} \cdot \nabla)\mathbf{B}/4\pi$.

Comparing with Equation 16.5, $\mathbf{L}_1$ has the same form as the term $-\nabla p$. This suggests (at first sight) that the magnetic field gives rise to a pressure analogous to gas pressure. The magnitude of the "magnetic pressure" is $p_{\rm mag} = B^2/8\pi$. (Expressing B in units of Gauss, $p_{\rm mag}$ is in dyn cm^{-2}.) If this were the only term we needed to consider, we would be tempted to think that the magnetic pressure at any position might behave just like the gas pressure.

But the Lorentz force also includes $\mathbf{L}_2$. The expression for $\mathbf{L}_2$ can also be written as the sum of two components, $\mathbf{L}_{2a} + \mathbf{L}_{2b}$. Let us define a unit vector $\hat{\mathbf{e}}$ along the direction of the magnetic field. Then $\mathbf{L}_{2a}$ can be written as $\hat{\mathbf{e}}(\hat{\mathbf{e}} \cdot \nabla)B^2/8\pi$: this is a force which acts along the vector $\hat{\mathbf{e}}$, i.e., *along* the field lines. As regards the magnitude of this component, the magnitude is equal and opposite to $-\nabla(p_{\rm mag})$. As a result, $\mathbf{L}_{2a}$ *cancels* the component of $\mathbf{L}_1$ which lies *along the field direction*. The net effect is that although $p_{\rm mag}$ at first sight appears to be analogous to the gas pressure, with equal pressures in all directions, this is not the complete picture of the Lorentz force. In a more complete picture, we find that the gradient of magnetic pressure *along the field direction* is zero. Thus, the magnetic field gradient exerts a force only in the direction *perpendicular* to the field direction. This is a striking indication of anisotropy in the presence of a magnetic field.

The component $\mathbf{L}_{2b}$ can be written $(B^2/4\pi)(\hat{\mathbf{e}} \cdot \nabla)\hat{\mathbf{e}}$. By considering the unit vector and its gradient, it can be shown that $\mathbf{L}_{2b}$ is a vector which lies in a direction *perpendicular* to the field lines. The magnitude of $\mathbf{L}_{2b}$ is equal to $B^2/4\pi R_{\rm curv}$, where $R_{\rm curv}$ is the radius of curvature of the field lines. The vector $\mathbf{L}_{2b}$ points *toward the center of curvature* of the field lines. The term $B^2/4\pi$ represents a *magnetic tension* T_m along the field lines.

The fact that magnetic fields give rise to both tension and pressure (although in *different* directions) should alert us to the fact that magnetic fields will have effects which may have no analogs in the simpler world of gas dynamics (such as in the atmosphere we live in on Earth). The fact that the Lorentz force is highly anisotropic is important for understanding magnetic activity in the Sun.

From a dimensional point of view, the units of pressure are equivalent to the units of energy density. Therefore, the energy density of the magnetic field is equal to $W_{\rm mag} = B^2/8\pi$. If B is in units of Gauss, $W_{\rm mag}$ is in units of ergs cm^{-3}.

16.6.2.2 The equations of magnetohydrodynamics (MHD)

The equations which describe the interaction between a moving fluid and a magnetic field are those of magnetohydrodynamics (MHD). We shall find in MHD an analog to the result (Equation 16.4) that an individual charged particle in the solar atmosphere is confined close to the field lines.

To obtain the MHD equations, we start with Maxwell's equations:

$$\nabla \cdot \mathbf{B} = 0 \tag{16.6}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E} \tag{16.7}$$

In a resistive medium, the current $\mathbf{j}$ which flows in response to an electric field is proportional to that field (according to Ohm's law). The constant of proportionality is determined by how effective the plasma is at conducting the current flow. In terms of $\mathbf{j}$, Ohm's law can be written as follows:

$$\mathbf{j} = \sigma_e(\mathbf{E} + \mathbf{E}_m) = \sigma_e\left(\mathbf{E} + \frac{\mathbf{V} \times \mathbf{B}}{c}\right) \tag{16.8}$$

In Equation 16.8, σ_e is the electrical conductivity of the medium.

From Equation 16.8, the external electric field can be written as $\mathbf{E} = \mathbf{j}/\sigma_e - (1/c)\mathbf{V} \times \mathbf{B}$. Inserting this in Equation 16.7, we find $\partial\mathbf{B}/\partial t = -c\nabla \times (\mathbf{j}/\sigma_e) + \nabla \times (\mathbf{V} \times \mathbf{B})$. Replacing $\mathbf{j}$ with $(c/4\pi)\nabla \times \mathbf{B}$, and using Equation 16.6, we finally have the following equation for the magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V} \times \mathbf{B}) + \eta_e \nabla^2 \mathbf{B} \tag{16.9}$$

The quantity η_e in Equation 16.9, defined by $\eta_e = c^2/4\pi\sigma_e$, is the magnetic diffusivity, which is a measure of the electrical resistivity of the medium.

Equation 16.9 describes the interactions between magnetic fields and a fluid medium. There are two distinct terms on the right-hand side of Equation 16.9. We consider the effects of these two terms separately.

First, suppose that the conductivity of the medium is so large that $\eta_e \to 0$. Then the second term on the right-hand side of Equation 16.9 becomes negligible compared to the first. The surviving term, which includes the velocity of the medium, can be shown to have the following property: if you choose a particular parcel of fluid which contains a magnetic field $\mathbf{B}$ within its area, and follow that parcel around, the amount of magnetic flux ($= \int \mathbf{B} d\mathbf{A}$) enclosed by that parcel of fluid remains constant. That is, magnetic flux neither enters nor leaves the parcel of fluid as it moves. The phrase which is commonly used to describe this behavior is that the field and the fluid are "frozen together." This is the equivalent, in the fluid limit, of the tightly bound nature of single particle motion: the limit of infinite conductivity is formally equivalent to the limit in which the radius of gyration is so small (compared to other lengths in the problem) that the gyroradius can be taken to be zero.

Second, suppose the velocity of the fluid is zero. Then the first term on the right-hand side of Equation 16.9 vanishes. We are left with an equation in which the time derivative of the field is related to the second spatial derivative of the field. This is a diffusion equation: it describes how the magnetic field decays as time goes on. If the spatial properties of the field are such that significant changes in field strength occur over length-scales of L, we can approximate ∇^2 in Equation 16.9 by $1/L^2$. Defining the time-scale $\tau_d = L^2/\eta_e$, we see that Equation 16.9 can be written $\partial \mathbf{B}(t)/\partial t = -\mathbf{B}(t)/\tau_d$. The solution is $\mathbf{B}(t) = \mathbf{B}(0)\exp(-t/\tau_d)$, i.e., the field decays on time-scale τ_d. That is, a field in a stationary medium does not remain constant with time, but decays on a characteristic time-scale

$$\tau_d = \frac{4\pi\sigma_e L^2}{c^2} \tag{16.10}$$

What does this decay time signify? The energy in the electric current is dissipated by resistive effects at a rate $\mathbf{j}\cdot\mathbf{j}/\sigma_e$ such that in the time-scale τ_d, resistive dissipation within 1cm^3 leads to a reduction in the magnetic energy W_{mag} in that cubic cm by an amount of order $B^2/8\pi$. Thus, resistive dissipation causes the field strength to decay on a time-scale of τ_d.

The conductivity which enters into Equation 16.10 is the electrostatic conductivity: when we express the field in G, and the charge on the electron in electrostatic units ($|e| = 4.8 \times 10^{-10}$), the conductivity σ_e has units of sec^{-1}.

16.6.2.3 Magnetic diffusion time-scales in the Sun

The electrical conductivity of the gas in the solar atmosphere is determined by the rate at which electrons (and ions) undergo collisions with the ambient medium when the electrons (and ions) attempt to carry the current. In the limit of complete ionization (e.g., in the corona, or deep below the surface), the collisions are determined by Coulomb effects. In such a case, the conductivity is given by the Spitzer formula $\sigma_e \approx 10^7 T^{3/2}$ sec^{-1} (Spitzer, 1962: note that we have converted Spitzer's formula from electromagnetic units (e.m.u). to electrostatic units (e.s.u.) using the factor c^2). In the upper chromosphere ($T = 10^4$ K) and in the corona ($T = 10^6$ K), typical values of σ_e are 10^{13} sec^{-1} and 10^{16} sec^{-1}, respectively.

In the photosphere and low chromosphere, where the degree of ionization may be much less than unity, σ_e is not as large as the Spitzer value. In a partially ionized gas, σ_e is proportional to the ratio of the number densities of electrons to neutrals. According to Bray and Loughhead (1979, Table 4.7), at optical depth $\tau = 1$ in the photosphere, and at $\tau = 1$ in the umbra of a sunspot (where the degree of ionization is lower than in the photosphere), σ_e has numerical values of order 10^{12} sec^{-1} and 10^{11} sec^{-1}, respectively.

Now we can evaluate the time-scale for a field to decay in the Sun. Suppose we consider one of the smallest identifiable magnetic units on the Sun: a pore. The characteristic length scale is comparable to granule diameters, i.e., $L = 10^8$ cm. Using this in Equation 16.10, along with the photospheric value

of $\sigma_e(10^{12}\ \mathrm{sec}^{-1})$, we find $\tau_d \approx 10^8$ sec, i.e., about 3 years. This is to be compared with the observed life-times of pores: at most a few hours. At the other extreme, consider the largest spot ever observed: with an area $A =$ 6300 millionths of the visible hemisphere, the associated linear scale L is about 8×10^9 cm. According to Equation 16.10, the decay time in the photosphere would be 10^{12} sec, i.e., some 30,000 years. However, the observed lifetime was less than one year.

Thus, the decay of the field by resistive diffusion is several orders of magnitude too long to be consistent with the observed lifetimes of pores and spots. The conclusion is that when we consider a pore (or a larger structure), resistive dissipation is not a significant contributor to the decay of the structure. As a result, the second term in Equation 16.9 is very small compared to the first term: it is as if $z_e \to 0$. Therefore, the magnetic field in the pore, and in other magnetic structures with length scales as large as (or larger than) granules, can be considered as "frozen" into the gas. This means that wherever the gas (or field) goes, the field (or gas) must go also.

One of the interesting features about MHD in the context of solar physics is that sometimes the field dominates the gas, and at other times the gas dominates the field. Which of the two is dominant in any given situation depends on the relative energy densities. In both cases, however, the field and gas are effectively frozen together.

16.7 Understanding Magnetic Structures in the Sun

Now let us see how the effects of MHD operate in a variety of solar features. The solar atmosphere provides a number of interesting situations where we may profitably study the effects of MHD in different limiting conditions.

16.7.1 Sunspot umbrae: inhibition of convection

In an "ordinary" convection cell (i.e., granule), when no magnetic field is present, the circulation of the gas (which is responsible for upward transport of heat) occurs in several stages. (1) Hot matter starts its circulation at depth H, and rises vertically to the photosphere in the bright center of the granule. (2) As the matter approaches the photosphere, it expands (due to reduced ambient density), and spreads out horizontally. In this phase, the material cools off, mainly by radiative losses into space. (3) The cooled material finds a location where it can sink vertically: this occurs in the dark intergranular lanes, and the gas returns eventually to depth H. (4) The material eventually, as a result of fluctuations, absorbs some excess heat, and this begins the circulation of a new cell. Each granule lives long enough to allow roughly one complete circulation to occur.

In the photosphere, with densities of $(2-3)\times10^{-7}$ gm cm^{-3}, and granulation flow velocities $V \approx (1-3)$ km sec^{-1}, the energy densities of the flows ($\approx 0.5\rho V^2$) are 10^{3-4} ergs cm^{-3}. These convective properties are ultimately determined by the requirement that the gas must transport outward the flux of energy which is provided by nuclear reactions deep inside the Sun.

Now we ask: what happens to the circulation in a granule when a magnetic field is present? In an umbra, the magnetic field lines are mainly vertical, and have strengths as large as 3000 G. Since ionized gas can move freely along field lines, stages (1) and (3) of the granule circulation are unaffected by the field. However, stage (2) is severely impeded: matter which contains even the small degree of ionization of photospheric gas is effectively frozen to the field lines. Since the latter are vertical, they restrain the gas from flowing horizontally. The energy density of the field ($W_{\rm mag} = B^2/8\pi$) is $\approx 4\times10^5$ ergs cm^{-3}, i.e., 40–400 times greater than the energy densities of the granular flows. Because of the excessive magnetic energy density, the "frozen fields" are capable of preventing the horizontal flows in stage (2) of granule flow.

The overall effect is that granule circulation is inhibited by the umbral field. Now, it is precisely that circulation which allows convection to transport heat to the solar surface. Shutting down the circulation has the effect that convection cannot function properly in an umbra. To be sure, radiation is available to carry some heat upward, but this is not very effective. As a result, the upward heat flux decreases below the normal value by a significant factor. The umbra becomes darker (by many tens of percent) than the photosphere.

16.7.2 Pores: the smallest sunspots

Sunspots are dark because a vertical magnetic flux tube inhibits convection in one or more granules. But what happens if the flux tube has a diameter which is smaller than a typical granule diameter, i.e., smaller than about 1000 km? In such a case, the flux tube is too small to impose control over the complete circulation of the granule. The gas flows would "move over" into a nonmagnetic area, and convection could proceed uninhibited. Thus, the smallest sunspots (pores) must have diameters which are at least as large as a single granule, i.e., $\geq$1000–2000 km.

16.7.3 Sunspots: the Wilson depression

Magnetic fields exert a pressure *perpendicular* to the field lines. In an umbra, where the magnetic field tube is vertical, the field exerts a pressure of $B^2/8\pi$ in the horizontal direction. The gas inside the flux tube also exerts a pressure $p_{\rm in}$. In order for the flux tube to be a stable structure, the sum of these forces must be balanced by the gas pressure $p_{\rm ext}$ in the external (nonmagnetic) medium:

$$p_{\rm ext} = p_{\rm in} + \frac{B^2}{8\pi} \tag{16.11}$$

Let us consider some typical numerical values for the photosphere. At $\tau \approx 1$ in the undisturbed photosphere, $p_{\rm ext} \approx 10^5$ dyn cm^{-2} (see Chapter 5, Table 5.3). A sunspot in which the vertical field has $B = 1$ kG requires, for stability, that $p_{\rm in} = p_{\rm ext} - B^2/8\pi$ have the value 0.6×10^5 dyn cm^{-2}. If the sunspot has $B = 1.5$ kG, then $p_{\rm in}$ must be $\approx 0.1 \times 10^5$ dyn cm^{-2}, i.e., an order of magnitude *smaller* than the pressure in the undisturbed photosphere. Truly, the flux tube is a region which has been almost "evacuated" of gas.

Recalling that in a sunspot umbra, the field can be as strong as 3 kilogauss, we see that in order to contain such a field, $p_{\rm ext}$ must be at least as large as 3.6×10^5 dyn cm^{-2}. In fact, to allow for the presence of *any* finite gas pressure inside the spot, $p_{\rm ext}$ must be even larger, perhaps $(4-5) \times 10^5$ dyn cm^{-2}. Such high pressures are simply not available in the photosphere: these pressures exceed the photospheric value by factors of e^{1-2}. Therefore, in order to contain the sunspot fields, we must rely upon gas pressures which lie one to two scale heights *deeper than* the photosphere. This corresponds to distances of 150–300 km below the photosphere.

In fact, the "umbral photosphere," i.e., the region where $\tau_* \approx 1$ in the umbra, does *not* lie at the same physical depth as $\tau = 1$ in the undisturbed photosphere. Instead, because of reduced gas density in the umbra, one can "peer in" more deeply into the umbra than in the undisturbed photosphere. As a result, the level where $\tau_* \approx 1$ in the umbra lies deeper *by a few hundred kilometers* than the level where $\tau \approx 1$ in the undisturbed photosphere. This is referred to as the "Wilson depression," named in honor of an eighteenth century observer who first observed the effect in sunspots near the limb.

If we imagine that we could (somehow) stand in the undisturbed photosphere, and look horizontally into a sunspot, what would we see? We would see a medium where the gas pressure is greatly reduced compared to the gas in which we are "standing." In other words, the gas in the spot would be "missing." Where did the missing gas go? The answer is: the cooler conditions have caused the gas to "slump" to greater depths.

To see why this is so, consider the operation of hydrostatic equilibrium in the vertical direction. Below the photosphere, the gas pressure increases exponentially with depth, but the magnetic field does not increase in strength very much. As a result, we quickly arrive at depths where $p_{\rm ext} \gg B^2/8\pi$. At such depths, the field can no longer interfere significantly with convection. This means that sunspots are relatively shallow phenomena, probably retaining their identity to depths of no more than $Z_b \approx$ 1–2000 km (i.e., 1–2 Mm) below the photosphere (Parker, 1979).

Consider the "base of a sunspot" to occur at depth Z_b below the photosphere. At Z_b we may assume that $p_{\rm in}(Z_b) \approx p_{\rm ext}(Z_b)$. At smaller depths z (closer to the surface), the pressure *inside* the sunspot is given by $p_{\rm in}(z) = p_{\rm in}(Z_b) \exp[(z - Z_b)/H_{\rm ss}]$, where $H_{\rm ss}$ is the pressure scale-height *inside* the sunspot. In the undisturbed photosphere, the vertical profile of pressure follows a similar law, except that the scale height H_p is larger, because of the higher temperature: the gas in the undisturbed photosphere at depth z is given by $p_{\rm ext}(z) = p_{\rm ext}(Z_b) \exp[(z - Z_b)/H_p]$.

Comparing the two pressures, we see that the pressure in the sunspot at depth z is *less* than in the undisturbed photosphere by the ratio

$$\frac{p_{\rm in}(z)}{p_{\rm ext}(z)} = \exp\left[(z - Z_b)\left(\frac{1}{H_{\rm ss}} - \frac{1}{H_p}\right)\right] \tag{16.12}$$

Near the photosphere, we know (Chapter 5) that the scale height H_p is about 150 km. In the photosphere of the sunspot, where the effective temperature is reduced by 30% relative to the photosphere (Section 16.1.1), $H_{\rm ss}$ is smaller than H_p by some 30%. Thus, $H_{\rm ss} \approx 100$ km. Therefore, at the photosphere $(1/H_{\rm ss} - 1/H_p) \approx 0.0033$ km^{-1}. At depth Z_b, where the sunspot is not distinguishable from the surrounding gas $(1/H_{\rm ss} - 1/H_p) = 0$. As a rough approximation, we may consider that the mean value of $(1/H_{\rm ss} - 1/H_p)$ between the bottom of the sunspot (Z_b) and the surface is about one-half of the above value, i.e., about 0.0017 km^{-1}. Setting $z = 0$ and $Z_b = 1000$–2000 km, we find that $p_{\rm in}(0)/p_{\rm ext}(0) \approx \exp(-1.7 \text{ to } -3.4)$. Thus, gas pressures in the surface layers of a sunspot are reduced compared to the photosphere by at least five, and possibly by more than ten.

Therefore, when we view an umbra from the photosphere, and find that there is "gas missing" from the umbra, the explanation is that the missing gas has "slumped downward" because of the smaller scale height, which cannot support as much overlying weight.

16.7.4 Prominences

Near the topmost portions of a prominence, the magnetic field lines are mainly horizontal. Material on such field lines can be supported against gravity. Ionized material is not permitted to move vertically because that would involve motions perpendicular to field lines. Motion along the field lines is permitted, but this can be impeded if there is a dip in the field lines.

16.7.5 Faculae

Faculae (singular: facula) are flux tubes with diameters that are *too small* to create pores. Without the ability to impede the circulation in a complete granule, there is no reason why convection should be inhibited. As a result, a facula does *not* appear as a dark feature in the photosphere. However, the presence of fields has the effect (see Equation 16.11) that the pressure inside faculae is lower than in the external gas. The reduced internal pressure allows us to see deeper inside faculae, i.e., faculae also exhibit Wilson depressions, although not as large as in the largest spots. Depressions in faculae are 100–200 km. The deepest gas that we can see in a facula is surrounded by walls of hotter gas extending upward by 100–200 km.

What effect does this have on what we see when we observe faculae? There is almost no observable effect when a facula is observed near the center of the solar disk. However, when a facula is near the limb of the Sun, conditions are different. Now, our line of sight enters the facula at an angle to the vertical.

This allows us to observe granules behind the walls of the tube; those granules lie deeper than the photosphere, and are therefore hotter than the photosphere. As a result, each facula is seen as a localized feature which is brighter than the photosphere. The excess in brightness can be up to a few percent.

Facular excesses play a significant role in explaining the fact that the "solar irradiance" is observed to be *larger* at solar maximum (i.e., when sunspots counts are maximum) than at solar minimum. For each sunspot, there are many faculae in the surrounding active region. Even though the power deficit of a sunspot (many tens of percent) is much larger than in any individual facula, the area occupied by the multiple faculae in the vicinity of the spot is large enough that the facular excesses more than compensate for the spot deficits.

16.7.6 Excess chromospheric heating: network and plages

In Chapter 15, we saw that the chromosphere in cell centers is heated by acoustic waves emerging from turbulent convection. Excess heating of the chromosphere in the network, and in plages, can be understood in terms of an extra source of wave energy. Since the network and plages can be identified with confidence as locations where the magnetic field strengths are larger than in quiet Sun, it is natural to look to the magnetic field as the source of extra wave energy.

The fact that a magnetic field is associated with a tension $T_m = B^2/4\pi$ along the field lines suggests that we consider the analog to transverse waves on a stretched string. Classical mechanics tells us that if a string is under a tension T_m, and the string has a mass density ρ, the speed of transverse waves on the string is $\sqrt{(T_m/\rho)}$. Now, even though a magnetic field line in and of itself has no mass, nevertheless in the solar atmosphere, where field and gas are "frozen together," the gas is tightly tied to the field. Therefore, when the field lines move, the gas also moves along. This confers on the field in effect a mass density equal to that of the ambient gas. Analogous to the stretched string, therefore, a magnetic field line can support a transverse wave mode which propagates at a speed

$$V_A = \sqrt{(T_m/\rho)} = B/\sqrt{(4\pi\rho)} \tag{16.13}$$

This propagation speed is referred to as the Alfven speed, after the Swedish physicist who, in the midst of World War II, discovered this wave mode in a magnetized gas (Alfven, 1942).

Let us look at numerical values for the Alfven speed in the Sun. In the photosphere, where $\rho \approx 3 \times 10^{-7}$ gm cm^{-3}, the numerical value of V_A is given by 515 B cm sec^{-1}, if B is in Gauss. In the upper chromosphere, where $\rho \approx 3 \times 10^{-13}$ gm cm^{-3}, $V_A = 5.15$ B km sec^{-1}. In the low corona, where number densities are 10^{8-9} cm^{-3}, i.e., $\rho \approx 2 \times 10^{-(15-16)}$ gm cm^{-3}, $V_A = 60$–200 B km sec^{-1}. Thus, in the photosphere, even in the umbra of a sunspot, Alfven speeds are no more than 10–20 km sec^{-1}. These speeds are not greatly different from the local sound speed. But in the corona, in active regions where

the fields can be as large as 1000 G or more (Lee, 2007), Alfven speeds may be as large as 60,000 km $\sec^{-1}$. Such speeds greatly exceed the speed of sound, indicating that the fields are definitely in control of the plasma in the corona.

The acoustic waves which heat the quiet chromosphere are longitudinal modes: the spatial displacements in the wave occur in the direction of propagation. In the presence of magnetic fields, the fact that transverse wave modes are also present means that magnetic regions of the Sun have access to an extra (nonlongitudinal) source of wave energy.

Is there a physical mechanism which can generate Alfven waves in a magnetic region? Consider a vertical magnetic field line which passes through a granule in the photosphere. Recalling that Alfven waves are transverse, and that the gas and field are frozen together, we see that the horizontal motions at the top and bottom of the granule will push and pull the field in a direction perpendicular to the field line with a certain velocity δV. By analogy with the formula for the flux of acoustic waves (Chapter 15, Equation 15.1), the flux of Alfven waves can be written as $F_A \approx \rho(\delta V)^2 V_A$. This differs from the flux of sound waves (Chapter 15, Equation 15.1) as regards the propagation speed: it also may differ as regards the amplitude. However, the motions in a convection cell (granule) are such that the horizontal velocity is probably not greatly different from the turbulent velocities ($\delta V \approx 1$ km $\sec^{-1}$) which were cited in connection with the acoustic waves in Equation 15.1. In the presence of a vertical magnetic field, the velocities of the flow may be impeded somewhat by the field. But if the field is weak enough, this impeding may not be serious. As a result, in regions of the photosphere where the field strengths are (say) 10–100 G, the fluxes of Alfven waves could be of order $F_A \approx 10^{7-8}$ ergs cm^{-2} $\sec^{-1}$.

These fluxes are one to two orders of magnitude smaller than the acoustic fluxes $F(\mathrm{ac})$ in the photosphere (see Chapter 15, Section 15.10). At first sight, this might lead to the conclusion that Alfven waves are of no significance in the heating of the solar atmosphere. But this conclusion is not necessarily correct. Compared to the acoustic fluxes *F(chr) which can reach the chromosphere*, (see Chapter 15, Equation 15.2), the Alfven wave fluxes estimated here in the photosphere are not at all negligible. As a result, in photospheric regions where the field strength is at least 10 G, waves in the Alfven mode may provide a significant supplement to the flux of waves which enter the chromosphere. Therefore, since the latter have been shown (Chapter 15) to be energetically significant for the heating of (at least) the lower chromosphere, the extra flux due to Alfven waves in magnetic regions may also significantly contribute to heating the solar chromosphere. Now, empirical evidence shows that plages and network are precisely regions where the field strength is enhanced: therefore Alfven waves may be responsible for the excess chromospheric heating observed in plages and network.

The transverse nature of Alfven waves has the effect that these waves differ in an important respect from sound waves (which are longitudinal). Since Alfven waves are not associated with compressions and rarefactions of gas (as sound waves are), Alfven waves are *not* subject to the steepening which

leads sound waves to form shock waves and thereby dissipate (see Chapter 15, Section 15.11). An important conclusion is that Alfven waves are more difficult to dissipate than sound waves. Therefore, if a region of photosphere creates equal fluxes of sound and Alfven waves, the Alfven waves survive for a longer time in the upper solar atmosphere. In particular, in favorable conditions, Alfven waves might even survive to reach the corona (Chapter 17, Section 17.1.2).

16.7.7 Magnetic fields and gas motions: which is dominant?

In the photosphere, we have now seen two distinct behaviors of the magnetic field. In some cases (sunspots), the field causes the solar surface to be *darker* than normal, while in other cases (network), the field causes enhanced *brightening*.

Can we identify a transition between these behaviors? We have already identified one such transition in the case of pores. There, the critical parameter was the diameter of the flux tube: in order for darkening to occur, the pore must be larger than the diameter of a typical granule.

We can also consider the matter from the perspective of frozen flux. The essence of frozen flux is that gas and field are forced to move together. This raises the question: which one dominates? Does the gas dominate the field or does the field dominate the gas? The answer is: the Sun provides us with the luxury of a "yes and no" answer. Examples of both situations can be found in different features.

The question of dominance can be discussed in terms of the energy densities which are available to gas and field in the photosphere. A moving gas has kinetic energy density $E_d = 0.5\rho V^2$ ergs cm^{-3}. Inserting typical values of density ($2-3\times10^{-7}$ gm cm^{-3}) and velocity (1–3 km sec^{-1}) in the photosphere, we find $E_d \approx 0.1$–1.4×10^4 ergs cm^{-3}. A magnetic field, with energy density $E_{\rm mag} = B^2/8\pi$, is comparable in energy density to E_d if $B \approx 150$–600 G.

Therefore, in any magnetic structures in the photosphere where the local field strengths are in excess of (roughly) 600 G, we expect to find that the gas flows are *not* sufficiently energetic to "push the field around." In such situations, the field will "win out" and impose changes on the gas. Sunspots, where the field suppresses convection and darkness ensues, are an example of this.

In photospheric structures where the field strength is less than (roughly) 150 G, the field is not sufficiently energetic to "push the gas around." In such situations, the gas "wins out," and imposes changes on the field. For example, in the network, moving gas induces wave modes on the field, thereby heating the overlying chromosphere, and causing local brightening.

For more evidence that there is a transition from local brightening ("the gas wins") to local darkening ("the field wins") at fields of a few hundred G, see Ortiz (2005).

16.8 Amplification of Strong Solar Magnetic Fields

Where do the strongest fields in the Sun come from? The answer is that the weak polar fields can be amplified by the differential rotation which is observed on the solar surface. The amplification occurs because the field is frozen into the solar material, and the latter rotates differentially in latitude. As we saw in Chapter 14, Section 14.9.3., the rotation period at the Sun's surface is 25.7 days near the equator, and 31.3 days at 60° latitude, i.e., a difference of about 20%. We have seen (Section 16.4.2) that the solar polar field has a strength of 6–12 G. Assuming that the polar field is a dipole (i.e., its field lines run in the north–south direction), the equatorial field strengths are 3–6 G. Because of field freezing, the equatorial field (which is originally in the north-south direction) will be sheared (i.e., stretched) by differential rotation. Both components of differential rotation (latitudinal and radial) can come into play. But for simplicity, let us consider only LDR. Then a particular field line, after one rotation (i.e., after 25.7 days) will return to the same longitude on the equator, but the high-latitude section of the same field line will lag behind by about 20% of a rotation. After five rotations, i.e., after 4.2 months, the high-latitude section of the field line will have fallen behind by about one full rotation, i.e., the equatorial will have "lapped" the polar section. In one year, the polar portion will have lost 2.8 full rotations on the equatorial portion of the field line.

The excess stretching of the field leads to field lines which become more and more stretched from east to west. That is, although the initial (polar) fields essentially were directed from north to south (i.e., they were poloidal fields), the stretching due to differential rotation leads to increasingly strong fields in the east–west direction (i.e., toroidal fields). It is this tendency for stretched fields to be mainly toroidal which determines that most pairs of sunspots lie almost east–west (see Section 16.1): each pair of spots originates in a strong toroidal (almost) east-west magnetic flux tube.

How strong do the toroidal fields become? It depends on how the area of a flux tube is distorted by the stretching motion. As the field lines are stretched, the area of a flux tube will likely be "squeezed." How much will the squeezing be? This is a complicated problem and it is not easy to give a simple answer. But suppose, for the sake of numerics, that an increase in length by 20% (after one rotation) leads to a reduction in area by (say) $\varphi = 10\%$. To conserve magnetic flux, the reduction in area by 0.1 in 25.7 days would mean that the toroidal field would be larger than the initial value (3–6 G) by $\approx$1.1 after one rotation. After one year, i.e., after 14.2 equatorial rotations, the toroidal field strength near the equator would be increased by $1.1^{14.2}$, i.e., by a factor of 3.9. After two, three, four, and five years, the initial equatorial field of 3–6 G would be amplified by factors of 15, 58, 220, and 870. Thus, the toroidal field strength at the equator would be 2600–5200 G after 5 years. These are comparable to field strengths observed in sunspot umbrae.

Suppose our estimate of φ is too large: suppose a more realistic value is $\varphi \approx 0.05$. Then the toroidal field strength would require about eight years to reach a strength of 1 kG. Thus, depending on the actual value of φ, the continuous operation of LDR *could* result in fields as strong as sunspot fields in time-scales of 5–8 years.

So far, we have considered only LDR as we see it at the surface. But the amount of LDR varies as we examine different depths beneath the surface. In Chapter 14, Figure 14.6, we see that the angular velocity difference between gas at 0 deg latitude and 60 deg latitude is maximum at radial locations of 0.9–0.95 solar radii. As a result, the stretching of poloidal fields will build up faster at depths of 35–70 thousand km below the surface. Another region of strong shear occurs at the interface between convection zone and radiative interior: there, a strong shear occurs over a relatively short interval in the radial coordinate. Magnetic fields which are frozen into such a highly sheared medium may also generate fields of strength $\geq$1 kG in relatively short time-scales.

However, in the radiative interior, where the gas rotates almost as a solid body, there is little or no tendency for the poloidal field to undergo stretching.

In view of these processes, toroidal fields of order 1 kG can be built up in the course of a few years not just in the surface layers, but at all depths throughout the convection zone. Now, the sunspot cycle is observed to last, in fact, some 11 years on average. So the stretching time-scales estimated above are in the right ball-park to allow surface fields to build up to kG strength in the course of (roughly) one-half of the sunspot cycle.

Since stretching of field lines is an inherent process in a medium with frozen fields and differential rotation, the question arises: what eventually stops the process of stretching? Why do the surface fields reach strengths of a few kilogauss and not much more? One reason has to do with buoyancy forces. To see how this operates, consider the application of Equation 16.11 to a stretched flux tube. The internal pressure $p_{\rm in}$ is lower than the ambient pressure $p_{\rm ext}$ by the amount $B^2/8\pi$. How does the temperature inside the flux tube compare with the temperature outside? To answer this, we note that, deep in the interior of the Sun, where radiative transport dominates (see Chapters 8 and 9), photons can carry heat efficiently back and forth between neighboring parcels of gas. These photons are not impeded by the magnetic field. As a result, the *temperatures* inside and outside the flux tube remain essentially equal. Therefore, reduced pressure $p_{\rm in}$ corresponds to *reduced* density $\rho_{\rm in}$ inside the flux tube: $\rho_{\rm in}/\rho_{\rm ext} = p_{\rm in}/p_{\rm ext} = 1 - (B^2/8\pi p_{\rm ext})$. Notice that, in order to avoid negative densities inside the flux tube, the maximum value which the field strength can have is $B_{\rm max} = \sqrt{(8\pi p_{\rm ext})}$.

Because the flux tube contains gas with lower density than in the ambient (nonmagnetic) medium, buoyancy forces act to push the flux tube upward. How strong are the buoyancy forces? In the presence of gravity g and a density difference $\Delta\rho = \rho_{\rm ext} - \rho_{\rm in}$, buoyancy creates an upward acceleration a_b which is given by (see Chapter 7, Equation 7.2):

$$a_b = g\frac{\Delta\rho}{\rho} \approx \frac{gB^2}{8\pi p_{\rm ext}} \tag{16.14}$$

How long does it take for a parcel of gas, subject to the acceleration a_b, to be buoyed up to the surface? To make the time as long as possible, let us consider a parcel of gas starting from as deep as we can reasonably assume, i.e., near the base of the convection zone. Such a parcel starts at a depth $D \approx 2 \times 10^{10}$ cm. What field strength should we use? Well, in the interest of making the rise as fast as possible, let us consider B to have its largest permissible value, $B_{\max}$. In that case, and assuming that the flux tube is free to rise, Equation 16.13 indicates that the full acceleration of gravity ($a_b \approx g \approx 2 \times 10^4$ cm sec^{-2}) would come into play. In such conditions, the parcel would rise to the surface in a time $\tau_r = \sqrt{(2D/a_b)} \approx 1400$ sec. This is less than one-half hour, a very short time indeed in the context of the solar 11-year cycle!

The effects of buoyancy have a well-defined effect in the Sun: they cause flux tubes to move "up and out." And the stronger the field, the faster the buoyancy forces bring it up to the surface. As a result, when we try to impose the condition that the Sun must make a strong (toroidal) field by amplifying its (weak) poloidal field, there is a race against time. On the one hand, differential rotation takes a finite time (5–8 years) to stretch the field and amplify it. On the other hand, as the field becomes stronger, the stronger are the buoyancy forces which want to drive the flux tube "up and out."

As regards numerical values, we note that at the base of the convection zone, where $p_{\text{ext}} \approx 3 \times 10^{13}$ dyn cm^{-2} (see Chapter 7, Table 7.1), $B_{\max} \approx 3 \times 10^7$ G. There is no evidence that fields as strong as this exist anywhere in the Sun: the strongest field ever observed in a sunspot is no more than a few kG. Therefore, in all likelihood, the fields which actually exist inside the Sun are much *weaker* than $B_{\max}$. In view of Equation 16.14, the buoyant accelerations a_b in the "real Sun" are therefore much less than g. Therefore, the time-scales τ_r for buoyant rise are much longer than 1400 sec. For a given value of the starting depth D, the time scale τ_r scales as $1/\sqrt{a_b} \sim 1/B$ (see Equation 16.14). Therefore, other things being equal, a flux tube containing fields of 3 kG ($= 10^{-4} B_{\max}$) will be buoyed up from the base of the convection zone on a time-scale which is 10^4 times longer than the time-scale for $B_{\max}$. This results in a time-scale of ≈ 1 year for the flux-tube to rise to the surface.

16.9 Why Does the Sun Have a Magnetic Cycle with $P \approx 10$ Years?

The Sun's magnetic cycle occurs over a time interval P of 9–12 years, as the sunspots come and go in numbers. The spots are the regions where the field grows to its largest values (3 kG), and the poles are the locations where the weak global fields (6–12 G) are easiest to identify.

In order to understand why the Sun has a cycle, let us start by considering the global field of the Sun at time t_o. Let the global field at t_o be directed in such a way that the Sun's North pole has field lines which point *outward*

from the Sun. Also at t_o, suppose for simplicity that there are no sunspots on the surface, i.e., the Sun is at sunspot minimum, and ready to start a new cycle. Let us see if we can understand the directions of the fields which occur in sunspots in this new cycle.

Differential rotation operates on the poloidal field, and in the course of 5–8 years, the poloidal field lines are stretched out so as to form strong toroidal fields ($\geq$1 kG) beneath the surface. This can be considered as the growth phase of the cycle. At certain locations and at certain (unpredictable) times, when something causes the local toroidal field to become unstable, a section of toroidal field rises up and breaks through the solar surface. The time-scale for buoyancy to bring up a field of strength 3 kG to the surface is <1 year. The breakthrough forms a pair of sunspots (leader plus follower) with a definite polarity. Given the *outward* direction of the Sun's North pole field in the Northern hemisphere at time t_o, the leader during the growth phase will also have field lines pointing *outward* from the Sun. But the follower will have field lines pointing *inward*. Conversely, in the Southern hemisphere, at the same time, pairs of spots will exhibit leaders with *inward* field lines and followers with *outward* field lines. That is, during the growth phase of the cycle, the leader spots in a given hemisphere retain the same sense of the magnetic field as exists at the pole in that hemisphere at time t_o. This helps us to understand Hale's polarity law.

Each spot pair is surrounded by an active region, which retains the overall polarity of the leader and follower spots. After a certain amount of time, which may be as long as weeks or months, each active region decays. The decay occurs due to the incessant "pounding" effects of supergranule motions beating against the magnetic flux tubes, shredding the accumulated flux on length-scales of a supergranule diameter (23–35 thousand km: Chapter 15, Section 15.5) and spreading it out in a diffusive manner over larger and larger areas.

To understand the diffusive decay of active regions, we note that the action of horizontal flows of gas in supergranules act as elements of random walk, pushing field lines around on the surface of the Sun. With diameters of order d = 30,000 km, and horizontal velocities V of order 0.3 km sec^{-1} in the supergranules, the associated diffusivity $D \approx dv$ is of order 10^{14} cm^2 sec^{-1}. In the presence of such a random walk, the time-scale required to cover a distance L is $\tau_d \approx L^2/D$. Therefore, in order for a flux tube to be transported from equator to pole, i.e., across a distance of $L \approx (\pi/2)R_\odot \approx 10^{11}$ cm, the time required is of order 10^8 sec, i.e., $\approx$ 3 years.

In the process of diffusive spreading, the slight tilt which is characteristic of spot pairs (i.e., they do not lie exactly East–West), and the predominant sense of their polarities, becomes important: the follower spots lie at slightly higher latitudes. As a result, in the process of shredding and spreading in (say) the Northern hemisphere, the effect is overall to ensure that fields from the follower spots diffuse toward the North pole. However, because they are followers, their magnetic polarity is *opposite* to the global polarity which existed at time t_o

at the North pole. As time goes on, and as more of the shredded and spread-out follower flux accumulates "up North," the North polar field is weakened, eventually passing through zero strength, and *the North pole takes on the opposite polarity.*

The analogous process is at work in the Southern hemisphere, causing the South pole also eventually to reverse sign. Because the reversal of the field depends on diffusive processes, the reversal does not occur at a precise instant of time: rather, it happens gradually over time. As a result, the polar reversals need not occur at exactly the same time in both hemispheres. But eventually, by time $t_o + P$, *both poles reverse their polarities*, and the Sun enters a new cycle.

What is the value of the cycle time-scale P? The physical properties of the Sun itself set the various time-scales which go into determining P. First, there is a time-scale on which fields can be amplified by differential rotation (5–8 years); second, there is a time-scale for the fields to be buoyed up by gravity to the surface ($\approx$1 year); third, there is a time-scale for the turbulence to shred and disperse the fields up to the polar regions ($\approx$3 years). Combining these three time-scales, we see that a time interval of 9–12 years could encompass all of the elements which contribute to a solar cycle.

This helps us to see why the Sun has a magnetic cycle whose length is measured not in millennia or centuries, and not in minutes or hours, but in time-scales of about 10 years. The time-scale of the solar cycle is determined by the Sun's own differential rotation, its gravity, the diffusivities of supergranule flows, and the linear extent of its surface.

16.10 Releases of Magnetic Energy

We have seen that magnetic fields have energy densities equal to their pressures, i.e., $W_{\rm mag} = B^2/8\pi$ ergs cm^{-3}. In favorable circumstances, the energy in the field can be converted into other forms. The two most prominent classes of events in the Sun which owe their existence to release of magnetic energy are flares and CMEs. We will discuss flares in Chapter 17, in the context of the solar corona, and CMEs in Chapter 18, in the context of the solar wind.

Exercises

16.1 Calculate the Zeeman splitting of a line at $\lambda = 6000$ Å in fields of $1,100$, and 10^4 G.

16.2 Consider an electron, a proton, and a lead nucleus gyrating in the Earth's magnetic field ($B = 1$ G) with a variety of energies. Calculate the radius of gyration for each particle in cases where the kinetic energy is (a) 1 eV, (b) 1 MeV, and (c) 1 GeV.

16.3 In the space between the stars (the interstellar medium: ISM), energetic particles (galactic cosmic rays) gyrate about a field with a strength of about 3×10^{-6} G. Determine the relativistic γ factor for an ultrarelativistic proton which has a radius of gyration of 10 AU in this field.

16.4 Calculate the Alfven speed in the ISM, where the gas number density is 1 proton cm^{-3}. (The mass of a proton is 1.67×10^{-24} gm.).

16.5 Calculate the range of Alfven speeds in the interplanetary medium near Earth: the field strengths range from one to 10×10^{-5} G, and the number densities range from one to 10 protons cm^{-3}.

References

Alfven, H. 1942. "Existence of electromagnetic-hydrodynamic waves," *Nature*, 150, 405.

Babcock, H. W. 1953. "The solar magnetograph," *Astrophys. J.*, 118, 387.

Bray, R. J. and Loughhead, R. E. 1979. "The properties of sunspot groups," *Sunspots.* Dover, New York, pp. 225–246.

Goldhaber, M. Grodzins, L., and Sunyar, A. W. 1958. "Helicity of neutrinos," *Phys. Rev.*, 109, 1015.

Hundhausen, A. J. 1977. "An interplanetary view of coronal holes," In *Coronal Holes and High Speed Wind Streams.* ed. J. B. Zirker. Colorado Associated University Press, Boulder, CO, 301.

Lee, J. 2007. "Radio emissions from solar active regions," *Space Sci. Rev.*, 133, 73.

Ortiz, A. 2005. "Solar cycle evolution of the contrast of small photospheric magnetic elements," *Adv. Space Res.*, 35, 350.

Parker, E. N. 1974. "Magnetic fields in the Sun," *Bull. Amer. Astron. Soc.*, 6, 18.

Parker, E. N. 1979. "Sunspots and the physics of magnetic flux tubes. I. the general nature of the sunspot," *Astrophys. J.*, 230, 905.

Schmelz, J. T., Holman, G. D., Brosius, J. W., and Willson, R. F. 1994. "Coronal magnetic structures observing campaign: 3. Coronal and magnetic field diagnostics derived from multiwaveband active region observations," *Astrophys. J.*, 434, 786.

Spitzer, L. 1962. "Encounters between charged particles," *Physics of Fully Ionized Gases*, 2nd ed. Interscience, New York, pp. 120–154.

Vaquero J. M. 2004. "A forgotten naked-eye sunspot recorded by Galileo," *Solar Phys.*, 223, 283.

Chapter 17

The Corona

Since ancient times, people who are fortunate enough to witness a total eclipse of the Sun have been able to see a remarkable phenomenon with the unaided eye: this feature becomes visible during the short interval of time (no more than 7–8 minutes) when the brilliant light from the photosphere of the Sun is blocked totally by the Moon. When the total eclipse begins, the disappearance of the solar photosphere does not mark the onset of the complete darkness of night. Instead, witnesses see a faint residual brightness in an extended region surrounding the dark side of the Moon. At its brightest, the faint light has an intensity which is no more than 5×10^{-6} times the brightness at the center of the solar disk (van de Hulst, 1950). The faint light is the corona, the Latin word for "crown," because it appears that the Sun is "wearing" a (faint) covering on top of its brilliant (but hidden) photosphere.

In certain eclipses, the corona is seen to extend more or less uniformly all the way round the dark Moon. But in other eclipses, the corona is seen to be concentrated in localized regions. These different behaviors can be seen in Figure 17.1. In the left image, obtained in 1988, the corona is brighter at low latitudes, with bright elongated "streamers" radiating outward close to the equatorial plane of the Sun. At the same time, one notices that there exist obvious decreases in brightness (or even a total absence of brightness) near the North and South poles of the Sun. At those locations, there are dark areas which can be seen extending all the way in to the edge of the Moon's disk. The dark areas in the polar regions are referred to as "coronal holes": the holes are regions where the material is less dense than elsewhere. In the right-hand image, obtained in 1980, bright features can be seen in the corona at all azimuths: not that all areas are equally bright, but at least there is little or no evidence for regions where the corona is completely dark. The difference in the Sun between 1980 and 1988 has to do with the magnetic fields: in 1980, the Sun's magnetic activity was at a maximum (in the 11-year cycle), while in 1988, the magnetic activity was close to minimum.

To the unaided eye, the corona has a pearly white color. For this reason, the corona as seen in Figure 17.1 is referred to as the "white-light corona." Subsequently, we will compare this corona with what is observable in other spectral regions, especially in X-rays. Instrumental measurements show that the spectrum of the white light corona is more or less identical to sunlight, apart from the presence of some emission lines which have no counterpart in the photosphere. The corona appears brightest at locations close to the

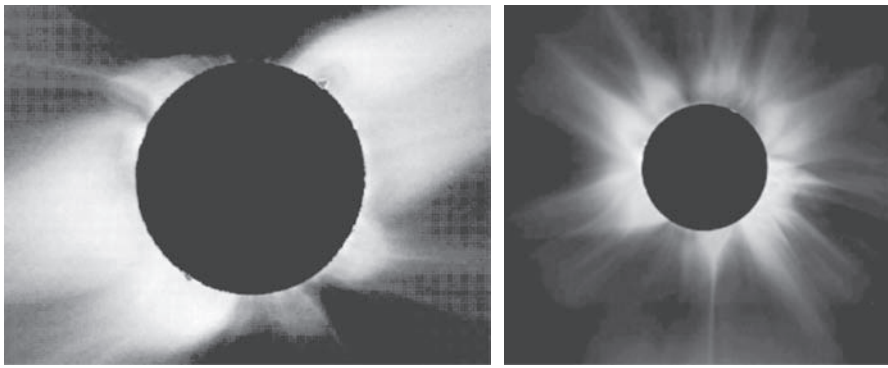


FIGURE 17.1: **(See color insert following page 202.)** The corona as seen during eclipses in 1988 (left) and 1980 (right). Both images are from the website of the High Altitude Observatory, a division of the National Center for Atmospheric Research, funded by the National Science Foundation.

surface of the Sun, and the brightness decreases with increasing distance from the photosphere: this can be seen by inspection of Figure 17.1. At greatest extent, the corona (or at least segments of it) remains visible out to radial distances that are as large as one solar radius or more.

The origin of the corona i.e., the identification of the physical process(es) which heat(s) the corona, is a long-standing problem in solar physics. Magnetic fields play a key role in structuring the corona, and in heating it. In this chapter, we summarize the physical parameters which have been determined for the corona in various locations on the Sun. These provide boundary conditions which have to be explained by solar researchers.

17.1 Electron Densities

Quantitative measurements of the brightness of the corona during eclipses indicate that the brightest regions of the corona, near the edge of the Moon's disk, have intensities of no more than $5{\times}10^{-6}$ times the brightness at the center of the (uneclipsed) solar disk. The intensity falls off rapidly with increasing radial distance: at a radial location of two solar radii, the coronal brightness has fallen off by a factor of at least ten, and in some locations (near the poles) by more than a factor of 100. The steep radial profile of the brightness of the coronal light contains important information about the radial profile of electron densities in the corona. However, the brightness is not the only information-containing parameter in the white light corona.

A remarkable feature of the white light corona (at least out to radial distances of a few solar radii) is that the light is polarized. Coronal light in which the electric vector is in the radial direction (pointing directly away from, or toward, the Sun) differs in intensity from coronal light in which the electric vector is in the tangential direction. The difference is by no means small: at radial locations of one to three solar radii, the degree of polarization can be as large as 50%. This significant polarization indicates that coronal light arises as a result of scattering of radiation off free electrons. The radiation which is being scattered comes from the nearby abundant supply of photons: the photosphere. By fitting the observed values of not only the brightness but also the degree of polarization as a function of radial distance, it is possible to extract the radial profile of electron density.

Results obtained by these methods indicate (Newkirk, 1967) that the electron densities in the inner corona fall off steeply as the observing instrument is pointed progressively farther from the Sun. For example, at radial locations of two, five, and ten solar radii, the densities are smaller than the near-Sun values by factors of about 100, 10^4, and 10^5. The radial profile of density is different in the polar regions from what it is at the equator: the corona is denser (at a given radial location) in the equatorial plane than in the polar regions. The radial profile of density is also different at the minimum of solar activity from what is observed at the maximum of solar activity: the corona (at a given radial location) is denser at solar maximum. The coronal density also varies depending on whether one observes inside a streamer or outside.

However, in all cases the density profile decreases monotonically as one observes farther from the Sun. The coronal regions which lie closest to the Sun, where the densities are highest, are referred to as the "base of the corona." The numerical values of the density in the various cases listed above are found to lie between 10^8 and 10^9 cm^{-3} at the base of the corona. We shall see later that other methods of estimating densities yields results which overlap with this range.

17.2 Spatial Structure in the White Light Corona

A striking feature of both images of the corona in Figure 17.1 is the nonuniformity in the coronal brightness. If an instrument is used to measure the local brightness around a circle at a fixed radial location (say, $r = 1.2R_\odot$), i.e., if one measures the brightness at different latitudes, the brightness will be found to be larger in certain latitudes, and smaller at other latitudes. The corona at solar minimum (1988) is an extreme case of this: there are regions near latitudes of +90 and −90 degrees (at the top and bottom of the image) where the brightness goes to essentially zero. In 1980, the brightness variations as a

function of latitude are not as extreme. However, even the unaided eye can see that in the right-hand image in Figure 17.1, there exist brighter and less bright regions of the corona at a fixed radial location. In other words, the corona exhibits large-scale spatial "structure." The brightest structures are referred to as "streamers."

In both images, several streamers can be identified which start off as features which are quite broad near the Sun, but become narrower as the height increases above the solar limb. In some cases, the narrowing proceeds to such an extent that, at a certain height, the streamer narrows down to a "cusp," above which a narrow straight-line feature (a "spike") extends farther from the Sun. A particularly clear example can be seen at the bottom of the 1980 image in Figure 17.1, where the spike can be traced with the eye almost to the edge of the field of view. Because of the similarity to certain military helmets which were worn in the World War I era, these structures in the solar corona are called "helmet streamers."

A question that is of interest as regards the physics of the Sun is the following: what are the typical linear sizes of the structures? Visual inspection of both images in Figure 17.1 reveals that within one quadrant of the Sun's limb, there is room to "fit in" perhaps two to three helmet streamers. This suggests that the base of each streamer may extend up to as much as a few tens of degrees of solar latitude. Recalling that the radius of the Sun is 7×10^{10} cm, and therefore, this corresponds to 57 degrees of latitude, we see that the bases of helmet streamers span regions of the Sun which are not small compared to a solar radius. The linear extent of helmet streamers at their base may therefore be as large as several times 10^{10} cm.

The fact that the corona contains significant spatial structure contains valuable information about the origin of the corona as a whole. We shall see below, when we examine images of the Sun obtained in X-rays, that spatial structure plays a key role in determining coronal properties. We shall return to this point once we have described how the temperature and density of the corona are measured.

17.3 Electron Temperatures

The physical parameters of the corona, specifically its density and temperature, cannot be derived by the techniques which were used to study the photosphere. There, the theory of radiative transfer was the tool which provided information as to the variation of physical quantities as a function of optical depth. In the corona, the gas is so rarefied that optically thin conditions prevail at all wavelengths which are visible to the unaided eye. So we have to rely on different techniques if we wish to determine numerical values for the key physical parameters such as density and temperature.

17.3.1 Optical photons

The first reliable estimates of temperatures in the coronal gas were obtained by Edlen (1945), who identified certain emission lines in the *visible spectrum* of the corona.

The optical emission lines in the coronal spectrum had been known for some years before Edlen's work, but they had not been identified. There was even talk of assigning them to a hitherto unknown element "coronium." But in an impressive feat of spectroscopic detective work, Edlen showed that the strongest coronal emission line in the red part of the visible spectrum (at $\lambda = 6375$ Å) originates as a transition between two fine-structure energy levels which exist in the Fe X ion, i.e., iron with nine electrons removed. Edlen also showed that the strongest coronal emission line in the green part of the visible spectrum (at $\lambda = 5303$ Å) originates in a transition between two fine-structure levels in the Fe XIV ion, with 13 electrons removed. The removal of nine or 13 electrons from iron atoms is an energetic process: the electrons in the plasma must have a temperature of at least 1 MK. (In what follows, we use the abbreviation MK for million degrees Kelvin.)

Since Edlen's discovery of these large temperatures, the key question about the solar corona has been: how does the Sun manage to heat up electrons in its atmosphere to temperatures which are at least 200 times hotter than the photosphere? Already when we discussed the chromosphere (Chapter 15), we raised the question as to how *chromospheric* gas could become heated above the photospheric temperature. In the chromosphere, the problem was relatively mild: we "only" had to explain why the temperature should increase above the photospheric value by a factor of about two. In that case, mechanical heating due to acoustic waves was found to be adequate to provide much of the chromospheric heating, at least in the low-to-mid chromosphere. When we come to the corona, we are faced with an analogous problem, except that now we have to account for an increase in temperature by a factor of at least 200. To be sure, we are dealing with gas that has a lower density than the gas in the low-to-mid chromosphere: as a result, even a relatively small flux of mechanical energy may be all that is needed to boost the temperature to the MK mark. Nevertheless, the questions remain: what is the source of the mechanical energy? and how much is needed?

Although Edlen's major achievement was to help to determine the *temperature* in the corona, his study also helped to set an upper limit on the electron *density* in the corona. To see why this is so, we note that the two coronal emission lines which were analyzed by Edlen were found to be forbidden lines: the transitions occurred in both cases between the $P_{3/2}$ ground level and a fine-structure $P_{1/2}$ level which lies about 2 electron-volts (eV) above the ground level. Electric dipole transitions are not allowed between such levels according to the common selection criteria which apply to LS coupling in an atom. But magnetic dipole transitions *can* occur: Edlen found that their radiative probabilities are in the range 10–500 sec^{-1}. This requires that the Fe X and Fe XIV ions must be preserved free from collisions for time-scales of up to 0.1 sec. This

sets an upper limit on the local electron density. The mean free time between collisions τ_c is given generally by the formula $1/(N_e \sigma V)$, where σ is the collision cross-section. For Coulomb collisions, $\sigma \approx 2 \times 10^{-4}/T^2 \approx 2 \times 10^{-16}$ cm^2 in the corona (with $T = 1$ MK). The electron thermal velocity has a mean value of $7 \times 10^5 \sqrt{T}$ cm sec^{-1}, i.e., $V \approx 7 \times 10^8$ cm sec^{-1} in the corona. Therefore, in order to ensure that the mean free time between collisions is longer than 0.1 sec, N_e should not exceed 10^9 cm^{-3}. As we have already seen (Section 17.1), the density at the base of the corona is indeed observed to be not significantly in excess of this limit.

In passing, we note that the fine-structure splitting of about 2 eV between the $P_{3/2}$ and the $P_{1/2}$ levels in Edlen's ions is very large compared to the fine-structure splitting that we normally see in optical spectra. A famous pair of lines in the yellow part of the solar optical spectrum, the lines which were labeled the D lines by Fraunhofer, lie at wavelengths of 5890 and 5896 Å. The separation of the lines occurs because of the fine structure splitting between $P_{3/2}$ and $P_{1/2}$ levels in neutral sodium, analogous to Edlen's lines. However, in the case of the yellow lines, the fine-structure splitting is only 0.002 eV, i.e., three orders of magnitude smaller than in Fe X and Fe XIV. The difference arises because of the highly stripped nature of the ions in Edlen's study.

17.3.2 X-ray photons

At temperatures of order 1 MK, the mean thermal energies of the particles in the coronal plasma are of order 0.1 keV. In such a plasma, much of the radiation emerges in spectral lines with energies of order 0.1 keV, extending in energy up to a few times this value.

Photons with energies of 0.1–1 keV have wavelengths of 100–10 Å. Such photons are referred to by astronomers as "soft" X-rays. ("Hard" X-rays are those with energies of $\geq$10 keV.) Soft X-rays are strongly absorbed in the Earth's atmosphere, and therefore, cannot be observed from the ground. Direct detection of the strongest lines in the coronal X-ray spectrum had to await the launching of rockets and spacecraft which would carry X-ray instruments, with spectral capabilities, into regions of space above the Earth's atmosphere. Such instruments were first launched in the late 1940s, using rocket technology which had been developed during World War II. Solar X-ray astronomy came into its own in the 1960s with the launch of a series of satellites called Orbiting Solar Observatories (OSO), and also with the flight of the Skylab space station (in orbit in 1973–1974). The last in the OSO series, OSO-8, was launched in 1975.

An example of an X-ray spectrum of a region on the Sun is shown in Figure 17.2. The wavelength range (13–18 Å) corresponds to photon energies between (roughly) 0.5 keV and 1 keV.

It is striking to notice in Figure 17.2 how completely different the spectrum of the Sun is at X-ray wavelengths compared to what we see when we view the Sun in visible light. In the latter, there is a strong continuum (which we can

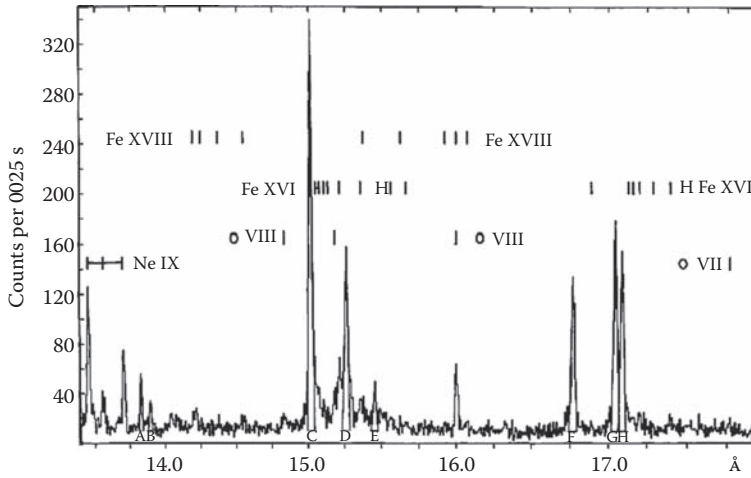


FIGURE 17.2: X-ray spectrum of the Sun between wavelengths of 13 and 18 Å obtained during a University of Leicester rocket flight December 6, 1970 (Parkinson, 1973). (Used with permission of K. Pounds, Professor of Space Physics, University of Leicester, and Director of the Research Group in change of the 1970 rocket experiment.)

see with our own eyes when a rainbow is visible), from which many *absorption* lines remove light (see Chapter 3, Figure 3.4): this is characteristic of the radiation which emerges from the optically thick photosphere. But when we examine the coronal spectrum in Figure 17.2, we see that the continuum (if there is one at all) is very weak, and the spectrum is dominated by a multitude of strong *emission* lines.

By comparing the wavelengths of the lines in the spectrum with tables of lines observed in laboratory plasmas, it is possible to identify the origin of many of the emission lines from the solar corona in Figure 17.2. Interestingly, and in corroboration of Edlen's pioneering work on the interpretation of optical photons, many of the observed X-ray lines can be assigned to highly stripped stages of ionization of some of the more abundant elements in the Sun, including oxygen, neon, and iron. For the lines which are present in Figure 17.2, the emitting element with the highest levels of ionization is Fe, including Fe XVI and Fe XVIII, i.e., iron which has lost two and four more of its electrons than the most highly ionized iron that was discussed by Edlen.

Prominent in the solar X-ray spectrum are the Lyman-series lines of the hydrogenic ions of several elements. The Lyman lines occur when an electron makes a transition into the ground state (the $n = 1$ level) from levels with $n = 2, 3, 4, 5 \ldots$. A hydrogenic ion is one in which only one electron remains in bound orbit around the nucleus. In an element of atomic number Z, the Bohr model of the hydrogen atom indicates that Lyman lines are predicted

to lie at wavelengths which are proportional to $1/Z^2$. Thus, the first four lines in the Lyman series are predicted to lie at wavelengths of 1216 Å/Z^2, 1026 Å/Z^2, 973 Å/Z^2, and 950 Å/Z^2, respectively. In the case of oxygen, the hydrogenic ion O VIII has $Z = 8$. As a result, the wavelengths of the first four Lyman lines of O VIII are predicted to be 19.0 Å, 16.0 Å, 15.2 Å, and 14.8 Å. In Figure 17.2, the last three of these lines are labeled as O VIII lines: the longest wavelength line lies off-scale to the right. Other spectra of the Sun, obtained over a wider wavelength range, and obtained during flares (see Section 17.18) show Lyman-α lines of the hydrogenic ions of O at 19.0 Å, Ne(Z = 10) at 12.16 Å, Mg(Z = 12) at 8.4 Å, Al(Z = 13) at 7.2 Å, Si(Z = 14) at 6.2 Å, S(Z = 16) at 4.75 Å, and Ca(Z = 20) at 3.0 Å.

The presence of highly stripped ions is a clear indication of high temperatures for the electrons in the plasma which emits the lines. How high are the temperatures? To estimate these, we can do the following thought experiment: suppose we were to strip the last remaining electron off O, Ne, and Mg, how much energy would that require? To answer this, we note that, according to the Bohr theory of the atom, the ionization potentials $I(Z)$ required to strip *all* Z electrons off an element with atomic number Z are larger by factors of Z^2 than the ionization potential of hydrogen (13.6 eV).

Now when we applied the Saha equation to a medium with low electron pressure (such as occur in the chromosphere and corona), we found that hydrogen begins to undergo significant (50%) ionization at temperatures of 7100–7200 K (see Chapter 4, Section 4.3). Analogously, in order to generate significant populations (50%) of the hydrogenic ions of O, Ne, and Mg, we need to solve the Saha equation $\theta I - 2.5 \log T = -\log p_e$(see Chapter 4, Equation 4.6) for cases with $I = I(Z)$. In the low corona, the electron pressure does not differ greatly from that in the upper chromosphere (see Section 17.9). Therefore, if we set $\log p_e = 0$ (as in Chapter 4, Section 4.3, for the upper chromosphere), we shall not make a serious error. The logarithmic temperature term is slowly varying, and so the solution for the temperature of 50% ionization in each ion is roughly $T \sim Z^2$. Since the appropriate T for 50% ionization is about 7000 K for hydrogen ($Z = 1$), when we set $Z = 8, 10$, and 12, we find that 50% of O, Ne, and Mg are in the hydrogenic state when the temperature has values of roughly at $T = 0.5$–1 MK.

Thus, the observational detection of Lyman-α lines of hydrogenic O and Ne (and also other elements, including Mg) in the solar X-ray spectrum provided significant corroboration that Edlen (1945) was correct in his identification of the red and green emission lines in the visible spectrum of the corona as arising in highly stripped iron.

17.4 Temperature of Line Formation

Each emission line in the X-ray spectrum of the Sun, arising as it does from a specific element (say, Fe), and from a specific stage of ionization of

that element (say, Fe XIV), is emitted from gas in which the temperature is not strictly uniform, but spans a finite range. However, if we observe a strong line from Fe XIV emitted by the coronal gas, this tells us that the range of electron temperatures in that gas cannot be arbitrarily broad. If the temperature were to be too low, the fraction of Fe which is in ionization stage XIV would be negligible, and so all lines originating in the Fe XIV ion would be weak. If the temperature were to be too high, one or more electrons would be stripped from the ion, forming Fe XV or higher, and X-ray lines from Fe XIV would no longer be emitted in significant quantities. As a result, there exists a finite range of temperatures over which any given X-ray line is emitted with significant efficiency. In fact, detailed atomic structure calculations show that any given ion (say, Fe XIV) is present with maximum relative abundance over a finite range of temperatures. Most importantly, the relative abundance of (say) Fe XIV peaks at a certain temperature: according to the calculations of Jordan (1969), the peak abundance of Fe XIV occurs at about 2 MK. For Fe X, the peak abundance occurs at $T = 1$ MK (see Figure 17.3).

Now it is true that at $T \approx 2$ MK, gas containing iron will contain *some* iron ions which have lost "only" eight or 10 or 12 electrons, while other iron ions will be present which have lost 14 or 16 electrons. However, the dominant ion of iron in that gas is (according to Figure 17.3) Fe XIV. As a result, spectral lines which originate in transitions between energy levels of the Fe XIV ion (such as Edlen's "green line" in the visible spectrum) will be maximally strong in gas with $T \approx 2$ MK.

Because each line originates in a certain stage of ionization, there is a more-or-less well-defined temperature at which each line is emitted with peak efficiency. Given the existence of such a peak, it is reasonable to refer to "the temperature of formation" $T_i(f)$ of each line (labeled by i) in the X-ray spectrum.

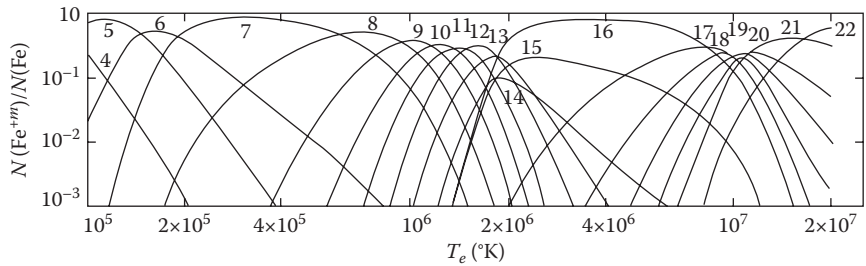


FIGURE 17.3: Fractional abundances of different ionization stages of iron as a function of electron temperature. The ordinate is the ratio of the density of Fe in ionization stage $m+$ to the total density of Fe. Each curve in the figure is labeled with the integer m. Thus, $m = 0$ would correspond to neutral iron, Fe I, and $m = 9$ and 13 correspond to Fe X and Fe XIV respectively. (From Jordan, C. 1969. *Monthly Not. Royal Astron. Soc.*, 142, 501. Used with permission from Blackwell Publishing.)

Note that the temperatures we refer to in discussing ionization processes are *electron* temperatures: they are measures of the mean thermal speeds of electrons in the plasma. Why do the temperatures refer to electrons? Because it is the fast motion of passing (free) electrons which determines whether or not an iron ion in that gas, in the presence of those free electrons, can retain its own bound electrons, or whether it would be energetically favorable to move to a higher ionization stage, or to a lower one. And it is also the passing (free) electrons, with their characteristic thermal speeds of order kT, that are responsible for inducing transitions of bound electrons in (say) Fe XIV from a lower to an upper energy level.

17.5 Pressure Scale Heights in the Corona

Now that we know the coronal temperature, it is useful to consider the concept of scale height (as defined in Chapter 5, Section 5.1). This concept applies to a medium which is in hydrostatic equilibrium: the low corona (within a radial distance of one to two solar radii of the surface) is such a medium. (However, as we shall see in Chapter 18, the more distant corona is definitely *not* in hydrostatic equilibrium.) Applying the values of coronal temperature 1–2 MK, and mean molecular weight of 0.5 (since hydrogen is essentially completely ionized), we find that the scale height in the corona H_c is of order 0.5–1.0×10^{10} cm.

The magnitude of the scale height helps us to understand why the observed brightness of the corona (at most a few millionths of the brightness of the photosphere) corresponds to electron densities of 10^{8-9} cm^{-3} (as reported in Section 17.1 above). In order to see this, let us recall what happens when we make an observation of coronal brightness from the Earth. Because of the steeply falling radial profile of coronal density, when an Earth-based observer makes measurements of coronal brightness, the line of sight can be thought of as a column which passes through the corona. At one point, that column passes more closely to the Sun than at other points of the column. Let this point be at radial location r_p. The local number density of electrons at that location is $N_e(r_p)$. The column density of electrons along the line of sight is therefore given by $N_e(r_p)$ times an effective path-length L through the corona. The reason we see the corona at all is because the $N_e(r_p)L$ electrons per sq. cm. in that column are able to scatter photons (from the photosphere, hidden behind the Moon during a total eclipse) into our line of sight. The density is stratified with a scale-height, H_c, with average values of about 7×10^9 cm in the low corona. In view of the stratification, our line of sight has an effective path-length of a few times H_c, say $L \approx 2 \times 10^{10}$ cm. Electrons along our line of sight scatter light from the photosphere with a cross-section which equals the Thomson value (see Chapter 3, Equation 3.1): $\sigma_T = 6.6 \times 10^{-25}$ cm^2. The

combination $N_e(r_p)L\sigma_T$ indicates the fraction of photospheric light which is scattered into our line of sight by the electrons in the column. With maximum brightness intensities in the corona observed to be $\approx 5 \times 10^{-6}$ relative to the photosphere (van de Hulst, 1950), and equating this relative intensity to $N_e(r_p)L\sigma_T$, we find that $N_e(r_p)$ is $\approx 4 \times 10^8$ cm^{-3}.

17.6 Ion Temperatures

What can we say about the temperature of the ions in the corona? Are the ion temperatures the same as the electron temperatures? In principle, if there are sufficient collisions between ions and electrons, the thermal energy should be equilibrated, and the ion temperatures should be equal to the electrons. This is more likely to happen in denser gas, such as occurs in the dense streamers. In the less dense gas of "coronal holes," equilibration of ion and electron temperatures is more difficult to achieve.

In the event that equilibration occurs, the thermal velocities of (say) iron ions in a coronal region where the electron temperature is 1–2 MK should have mean values $V_{\rm th} = \sqrt{(2kT_e/Am_H)}$ (where $A = 56$ for iron), i.e., $V_{\rm th} = 17$–24 km sec^{-1}. Therefore, if we measure the line width of a coronal iron line, the half-width of the line $\Delta\lambda$ should in principle have a value that is related to λ by the relationship $\Delta\lambda = \lambda V_{\rm th}/c$. However, empirical data reveal that the observed line widths are *larger* than the thermal predictions. What causes the excess line widths? Is it enhanced temperature of ions relative to electrons? Possibly: some extremely high ion temperatures (>100 MK) have been reported in fast solar wind (e.g., Cranmer et al., 2008). But there is another possibility: there might be nonthermal motions ("turbulence") in the corona, and the ion lines might be broadened by those motions. We have already come across the idea of "turbulence" in a very different context: "microturbulence" plays a role in the photospheric spectrum of the Sun (Chapter 3, Section 3.8.2). Quantitatively, however, there is a large difference between photosphere and corona: whereas in the photosphere the amplitude of the turbulence is 1–2 km sec^{-1}, the amplitude in the corona is much larger, up to 60 km sec^{-1} at altitudes of order $0.3R_\odot$ above the surface (Wilhelm et al., 1998). Waves on the magnetic field might explain this coronal "turbulence" (Section 17.17): some models of coronal wind acceleration by magnetic waves predict that the wave amplitudes could be as large as 60 km sec^{-1} (or more) in fast wind (Tu and Marsch, 1997). The presence of such large turbulent motions makes it difficult to determine with confidence how much of the observed broadening of coronal lines can be attributed to thermal motions. Thus it is not always easy to say definitively whether or not the ion temperatures differ from electron temperatures. But if it could be shown that indeed the heavy ions were definitely much hotter than protons or electrons, that might contain important information about the physical process which heats the heavy ions.

17.7 X-ray Line Strengths: The Emission Measure

When an instrument in near-Earth orbit points at the Sun and measures the solar X-ray spectrum quantitatively, the data provide a numerical value for the strength of each line. The line strength can be quantified as the flux of energy F_E (in units of ergs cm^{-2} sec^{-1}) at a distance of $D = 1AU = 215R_\odot$ from the corona. Back at the Sun, the flux F_S emerging from the corona is larger than F_E by the areal factor $(215)^2$. Multiplying F_S by the area of the coronal source, we now have an estimate of the rate E_S (ergs sec^{-1}) at which the coronal source emits energy in any given line. The aim of coronal analysis is to convert E_S into the electron density N_e in the coronal source.

To determine N_e, we start by noting that a particular coronal line from a certain ion is generated by collisions between electrons and that ion: a collision pumps an electron into the upper level, and then the photon emerges when that electron returns to the lower level. Consider 1 cm^3 of plasma in which there exists one ion and one electron. Knowing the quantum properties of the ion, and how fast the electron moves (i.e., temperature T), it is possible to calculate quantum mechanically the rate at which the line would be emitted. Multiplying by the photon energy $h\nu$, this yields an energy emission rate $\Phi(T)$(ergs cm^3 sec^{-1}) from that 1 cm^3 volume at temperature T. At low T, it is hard to excite any atomic levels, and so $\Phi(T) \to 0$. And at the highest T, above (say) 10 MK, all the elements are ionized, and there are no bound levels to radiate lines: only the (weak) continua remain, and $\Phi(T)$ is small. As a result, for each element, $\Phi(T)$ has a peak value at an intermediate temperature. In the solar atmosphere, where a broad mixture of elements are present, each element contributes somewhat differently to the overall $\Phi(T)$ function. The result is a function which peaks at $T = 1$–3×10^5 K (see Figure 17.6). This function is the "radiative loss function" for an optically thin gas.

If the 1 cm^3 volume contains N_e electrons and one ion, the energy emission rate is $N_e\Phi$. If the 1 cm^3 volume contains N_e electrons and N_i ions, and if all the photons can escape without being blocked by other ions, then the energy emission rate is $N_eN_i\Phi(T)$ ergs cm^{-3} sec^{-1}. Finally, if the coronal source we are observing has a volume of V_c, the source emits energy at a rate of $N_eN_iV_c\Phi(T)$ ergs sec^{-1}.

Given a measurement of E_S (ergs sec^{-1}) in a particular line, and knowing $\Phi(T)$ for that line, we can obtain the quantity $N_eN_iV_c$ for the coronal source at temperature T. The ion density N_i of (say) Fe XIV depends on two factors: (i) the abundance of Fe relative to hydrogen in the Sun, and (ii) the fraction of Fe which is in the XIV stage of ionization (Figure 17.3). Item (i) is fixed by a certain choice of chemical abundances for the Sun as a whole. Thus, N_i at a given T is a known fraction φ_i of the local density N_H of the most abundant element, hydrogen. In the high-temperature corona, hydrogen is fully ionized, such that $N_e \approx N_H$. As a result, from the quantity $N_eN_iV_c$, we can obtain the "emission measure" $EM(T) = N_e^2V_c$ of the coronal source at temperature T. The units of $EM(T)$ are cm^{-3}.

EM(T) is a quantitative indication of the amount of material at temperature T in the source region. By choosing the lines we observe strategically, i.e., lines which are formed at different temperatures, we can determine how much material is present in the source as a function of T. Knowing *EM*, the local density N_e can be estimated if the volume of the source region is known. Another way to determine N_e is to measure the ratio of the observed intensities of certain pairs of lines in the X-ray spectrum: the relevant pairs of lines share a common energy level in the parent ion. Since collisions control the rate of photon emission in the various lines, quantum mechanics predicts that the intensity ratio should depend in a well-defined way on the local density.

17.8 Densities and Temperatures: Quiet Sun Versus Active Regions

In the quiet Sun, it has been found (Brosius et al., 1996), that *EM(T)* is maximum at $T = 1$–2 MK, confirming the early work of Edlen (1945). In active regions, *EM(T)* retains significant values up to $T = 4$–5 MK. Thus, active regions contain material which is definitely hotter, by factors of up to two to three, than that which is present in the quiet Sun.

In the quiet Sun, Brosius et al. (1996) report densities of 10^9 cm^{-3}, while in active regions, the densities are large by factors of four to five.

Thus, coronal material in active regions is somewhat *hotter*, and somewhat *denser*, than coronal material in the quiet Sun.

17.9 Gas Pressures in the Corona

Let us notice an important point about the pressure at the base of the corona. Now that we have information about temperatures and densities, we can evaluate empirically the gas pressures at the base of the corona ($p_{cb} = 2N_e kT$). Inserting $T = 1$–2 MK and $N_e = 1$–4×10^9 cm^{-3}, we find coronal base pressures in the range 0.3–2 dyn cm^{-2}.

The physical significance of these pressures becomes apparent when we compare them with the pressure at the top of the chromosphere, $p_{\rm tc}$. In Chapter 15 (last paragraph), we noted that $p_{\rm tc}$ is of order 0.6 dyn cm^{-2}. It is important to notice that the range of $p_{\rm cb}$ *overlaps* with $p_{\rm tc}$ It seems that the pressure at the base of the corona is essentially identical to the pressure at the top of the chromosphere.

This discussion of pressure reminds us that when hydrostatic equilibrium applies, it is useful to think in terms of pressure scale heights, H_p. The scale height is defined to be the vertical distance across which the density

(or pressure) falls off by a factor of e. In terms of this definition, the number n_p of scale heights which separates the top of the chromosphere from the base of the corona is given by $n_p = \ln(p_{\rm tc}/p_{\rm cb})$. Inserting the above ranges of values of $p_{\rm tc}$ and $p_{\rm cb}$, and noting that n_p cannot be negative, we find that n_p ranges from at most 0.7 to a value which approaches zero. In chromospheric gas, H_p is a few hundred km. As a result, the transition from the chromosphere to the corona occurs across a height range which is no more than 100–300 km: it may in fact be close as zero.

On a scale of one solar radius, even a transition over 300 km can be considered relatively abrupt. If we perform the thought experiment of starting in the photosphere, and moving up in altitude through the solar atmosphere, we will find that, after we pass through the chromosphere, the onset of the corona occurs over a height scale of no more than 300 km. Such a transition is essentially discontinuous.

We shall return below (Section 17.15), once we discuss thermal conduction, to discuss a physical reason how such a discontinuity might arise in the solar atmosphere.

17.10 Spatial Structure in the X-ray Corona

Hot material at $T = 4$–5 MK in the corona is highly localized to active regions: it is not found in the quiet Sun. On the other hand, the cooler coronal material, at temperatures of 1–2 MK, can be found in the quiet Sun, which extends over large portions of the Sun's surface. As an illustration, we show in Figure 17.4 an image of the Sun obtained in the 195 Å line of Fe XII, a line which is formed at $T \approx 1.6$ MK (see Figure 17.3). This line is representative of the 1–2 MK material which is found in the quiet Sun.

Visual inspection of Figure 17.4 indicates that the 1–2 MK coronal material is spread more or less everywhere throughout the corona: one can see a "fuzzy glow" which permeates most of the field of view (apart from coronal holes). However, when the Sun is imaged with an instrument which is sensitive to hotter gas ($T > 3$ MK), the result is a more "patchy" picture: see Figure 17.5.

Apparently, as long as we exclude coronal holes, the Sun finds a way to heat material almost everywhere in the corona to temperatures of order 1–2 MK, but heating the gas to temperatures of > 3 MK is a less common phenomenon.

The contrast between the low-to-mid chromosphere and the hotter coronal regions is worth noting. No matter where one looks on the surface of the Sun, one finds chromospheric material at temperatures of 6–7000 K. But the hotter corona is far from spherically symmetric. This suggests that whatever is heating the hottest parts of the corona is distinct from whatever is heating the low-to-mid chromosphere. The spherical symmetry of the latter led us (in Chapter 15) naturally to the conclusion that acoustic waves from

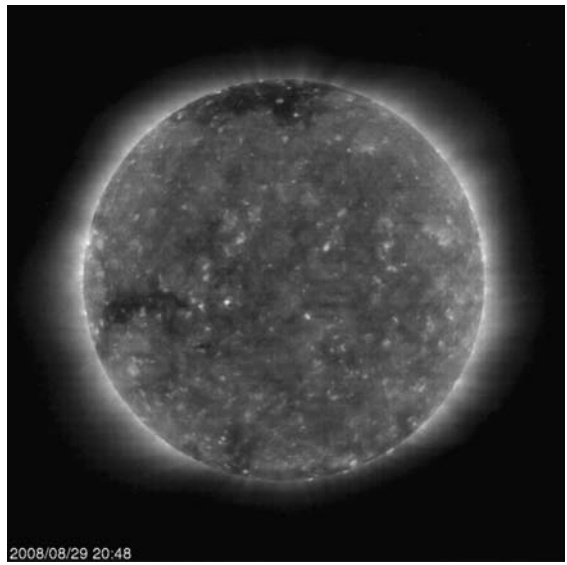


FIGURE 17.4: **(See color insert following page 202.)** Image of the Sun obtained in a line of Fe XII, with formation temperature 1.6 MK. The image was obtained on August 29, 2008, when the Sun was very quiet. (Courtesy of SOHO/EIT consortium. SOHO is a project of international cooperation between ESA and NASA.)

the ubiquitous convection may be heating the low-to-mid chromosphere. In Chapter 15, we provided quantitative evidence in favor of this hypothesis.

But in the hotter coronal regions, there must be additional localized sources of mechanical energy. Since the hotter coronal regions coincide with active regions, with their locally strong magnetic fields, it is natural to conclude that the heating of coronal material to > 3 MK is associated somehow with magnetic processes. Magnetic fields can provide mechanical energy over and above what is supplied by the acoustic waves from convection. One source of this extra energy is a variety of wave modes which exist in a magnetized plasma and which have no counterparts in a nonmagnetic medium (Alfven waves: Chapter 16, Equation 16.13). But there are also physical processes in a magnetized medium which do not occur in a nonmagnetic medium. The most striking of these ("magnetic reconnection") will be discussed below (Section 17.18.8).

17.11 Magnetic Structures: Loops in Active Regions

The fact that localized magnetic fields are associated with coronal heating receives strong confirmation from images such as Figure 17.5. This image was

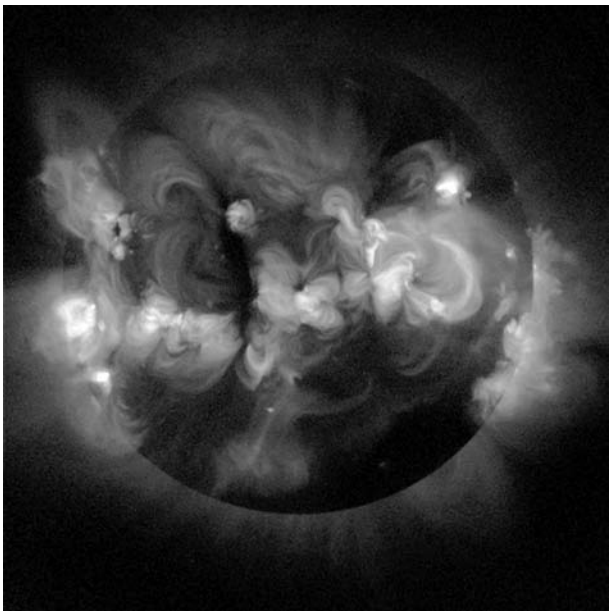


FIGURE 17.5: **(See color insert following page 202.)** Image of the Sun in X-rays emitted by gas at temperatures hotter than (roughly) 3 MK (Yohkoh/SXT). (Image is from the Yohkoh mission of ISAS, Japan. The X-ray telescope was prepared by the Lockheed-Martin Solar and Astrophysics Laboratory, the National Astronomical Observatory of Japan, and the University of Tokyo with the support of NASA and ISAS.)

obtained with the YOHKOH spacecraft, where the imager responds mainly to gas at temperatures of 3 MK and above. Even a casual inspection of the image shows that the strongest emission comes from features which appear to have shapes of a particular type, namely, the emission looks like it is coming from features which are best described by the terms "arcs" or "arches." Each arch has well-defined "foot points" which are rooted in the solar surface, while the central part of the arch loops up to a greater or lesser height above the surface.

One particularly clear example of a loop appears in Figure 17.5 about half way between the center of the Sun and the right-hand limb. That loop has a length which can be estimated visually from Figure 17.5: the foot points are separated by a distance D_p of about one-half a solar radius, i.e., $\approx 3\times10^{10}$ cm. Other loops in X-ray images of the corona can be identified as having smaller foot point separations. In some images, loops can be identified with foot points that are separated by not much more than a couple of times 10^9 cm. How high does a loop with foot point separation D_p extend above the surface of the Sun? At one extreme, magnetic field properties are such that it is unlikely that a

stable loop would extent to heights which exceed D_p by a large factor: such elongated loops would tend to "pinch off" near their base. At the other extreme, it is also unlikely that the loop would reach up only a very low height, much less than D_p: such a squat loop would need to have a shape which was essentially flat on top. In general, a loop with foot point separation D_p is expected to have a maximum height above the surface which is also of order D_p.

The largest loops in Figure 17.5 have lengths which are comparable to the spatial scales which were mentioned above in connection with the white-light corona (Section 17.2). In fact some helmet streamers in the white-light corona are spatially correlated with some of the loops seen in X-rays. The difference is that X-ray observations allow the loops to be observed *on the disk* of the Sun, and also in regions where the loops extend beyond the limb. In contrast, the white-light data can record loops only when these are extended above the limb, where the brilliant photosphere does not overwhelm the (faint) coronal light.

The loop shapes in Figure 17.5 are reminiscent of structure which one observes in the laboratory when iron filings are scattered on a sheet of paper held above a bar magnetic: the filings line up in shapes which include arcs stretching from one end of the magnet to the other. In fact, comparison with solar magnetic data obtained at the same time (with a different instrument) as the X-ray image (such as Figure 17.5) shows that each arc (or loop) has its foot points in an active region. Each loop follows the location of a magnetic flux rope which emerges from the surface of the Sun at a location where the magnetic polarity has a particular sign, and reenters the Sun at another location, where the magnetic polarity has the opposite sign.

Such a loop is said to be "closed," because both foot points are rooted in the denser material of the solar surface. In the presence of a closed loop, solar material is constrained to follow the loop field lines (Chapter 16, Section 16.6.1): the material is forbidden from flowing across the field lines. In this sense, it can be said that coronal gas is "trapped," and density builds up on the loop. As a result, if there are magnetic waves, or other processes, which supply energy to the loop, the energy supply has a "captive audience," i.e., the trapped gas in the loop. This material can be subjected to heating, which will be greater or less, depending on how much mechanical energy the magnetic field can deposit in the gas. This is the ultimate source of the hotter and denser gas which is observed in active regions (see Section 17.8).

17.12 Magnetic Structures: Coronal Holes

Coronal holes differ from the remainder of the Sun not merely in having lower than average density. They also differ in the sense that they exhibit a different magnetic topology. In the holes, the magnetic data indicate that the magnetic field lines, rather than containing closed loops, are in fact *open* to

space. As a result, coronal material is not trapped, but can stream out freely from the Sun. In this situation, there is no "captive audience," and the density does not build up as much as on the closed loops of active regions. Whatever supply of magnetic energy is available goes into accelerating the plasma away from the Sun. So efficient is this acceleration that the fastest solar wind is observed to emerge from coronal holes. This conclusion is true whether we are considering the coronal holes which are permanently present around the North and South poles of the Sun, or the coronal holes which occasionally exist at low latitudes.

17.13 Magnetic Structures: The Quiet Sun

What about the solar image of the cooler coronal gas which appears in Figure 17.4? This image of quiet Sun shows that the material with temperatures of 1–2 MK is found essentially at all locations in the quiet corona.

Is this 1–2 MK material also to be attributed solely to magnetic loops, maybe smaller in size than those which feature prominently in the active Sun (as in Figure 17.5)? In some cases, the answer appears to be "yes." There are certainly many small bright "points" in Figure 17.4: it is easy to imagine that these may be due to small magnetic regions, each with its own localized source of magnetic energy.

But what about the more diffuse, almost spherically symmetric "fuzzy glow" that permeates essentially the entire image? Are magnetic fields also responsible for this material? Maybe so: magnetic data show that magnetic fields can be detected at essentially all locations of the quiet Sun. To be sure, the strength of the fields is not as large as in active regions. But the fields *are* present in the quiet Sun. And when coronal field structures are computed from the surface fields, the calculations suggest that there do exist closed loops in the quiet Sun. (This is in contrast to coronal holes, where the calculations show that the fields are open.) However, from visual inspection of Figure 17.4, it is not obvious that any loop-like structures can be identified with certainty in the "fuzzy glow." We shall revisit this in Section 17.17.2.

17.14 Why Are Quiet Coronal Temperatures of Order 1–2 MK?

At the end of Chapter 15, we noted that the material in the upper chromosphere is less and less efficient at disposing of any mechanical energy which is deposited therein. The cooling mechanism which provided such an effective

thermostatic effect in the low-to-mid chromosphere diminishes greatly in efficiency in the upper chromosphere. The reason for the decreased cooling efficiency is that hydrogen is becoming almost completely ionized, and free protons plus free electrons are much less effective radiators than the bound electrons in hydrogen atoms. (Recall that bound electrons are much better absorbers of photons than are free electrons: see Chapter 3, Section 3.3.1.) As a result, if any mechanical energy is deposited at the top of the chromosphere, the temperature in the gas increases rapidly ("runs away") to much higher values.

What will stop the temperature run-away above the upper chromosphere? The answer is: the run-away will stop when an additional source of cooling (over and above radiation) comes into play in order to dispose of the deposited energy. If and when that happens, a new equilibrium condition becomes possible.

Two additional cooling options are available in the solar corona. One has already been mentioned in the context of coronal holes (Section 17.12): the material begins to flow outward from the Sun. This option is available in coronal holes because the magnetic field lines there are open to interplanetary space. The openness of the field lines is a characteristic signature of coronal holes.

17.14.1 Thermal conduction by electrons

But in the quiet Sun, the outflow option is not always available as a cooling mechanism: the magnetic fields in the quiet Sun's corona are for the most part closed. Here, the new option which becomes available to assist in cooling is the mechanism of *thermal conduction*.

We have already come across this mechanism when we discussed (in Chapter 8) how heat is transported deep inside the Sun. According to Equation 8.1, in the presence of a temperature gradient dT/dr, the flux of heat down the temperature gradient is $F = -k_{\rm th}(dT/dr)$, where $k_{\rm th} = (1/3)\lambda V_t \rho C_v$ is the thermal conductivity. In Chapter 8, we calculated the value of $k_{\rm th}$ in a manner which was appropriate for the deep interior of the Sun: there, the principal agent for the transport of heat was *photons*. Now we need to consider conduction in the corona, where photons are no longer the principal agents of heat transport. In the corona, it is *thermal electrons* which transport the heat most effectively.

Let us see what value of $k_{\rm th}$ results from the electrons in the solar corona. We need to evaluate four parameters. (i) The mean thermal speed of the electrons is given by $V_t = \sqrt{(2kT/m_e)}$. (ii) The mass density of the corona is given by $\rho = N_e \mu m_H$. (iii) The specific heat per gram at constant volume (Chapter 6, Equation 6.3) is given by $C_v = (3/2)k/\mu m_{\rm H}$. (iv) The mean free path λ is given by $1/N_e\sigma$, where the collision cross-section is σ.

To calculate σ for electrons in the corona, we note that collisions occur because of the Coulomb forces. We have already seen (Chapter 11, Section 11.3)

that the Coulomb cross-section is given by $\sigma_c \approx \pi e^4 \Lambda/(kT)^2$. In coronal conditions, the numerical value of Λ is typically 10–20.

Combining the factors in Equation 8.2, we find that the thermal conductivity in the solar corona is given by

$$k_{\rm th} = \frac{1}{3} \frac{k^2 T^2}{N_e \pi e^4 \Lambda} \sqrt{\frac{2kT}{m_e}} \frac{3k}{2\mu m_{\rm H}} N_e \mu m_{\rm H} \tag{17.1}$$

Collecting terms, we see that the density N_e cancels out, and the conductivity depends only on a power law of the temperature, according to $k_{\rm th} = k_o T^{2.5}$. In the corona, inserting appropriate values for the physical constants, we find that the numerical value of the coefficient k_o is about 10^{-6} in c.g.s. units.

The high power of the temperature dependence in $k_{\rm th}$ (index = 2.5) is noteworthy. It has the effect that although thermal conduction by electrons is negligible in the (cold) photosphere (where photons dominate the transport), this is no longer the case in coronal conditions. In the corona, with temperatures which are 200–400 times larger than in the photosphere, the electron thermal conductivity is a million times more effective than in the photosphere. In the corona, thermal conduction is a physical process which cannot be neglected.

Now that we have an expression for the thermal conductivity, we revert to a discussion of energy balance, such as that which was given earlier (given in Chapter 15, Section 15.12) in the context of the chromosphere. Given that energy is deposited at a certain rate every second into 1 cm^3 of the corona, we need to know the rate $E_{\rm cond}$ at which the energy is conducted out of that cubic cm per second by conduction. This rate, i.e., the rate of energy loss per unit volume, is given by the spatial divergence of the heat flux F.

To proceed, we note that in the presence of a closed loop of half-length L, the coronal part of the loop (near the apex of the loop) has a temperature T which is much larger than the foot-points. As a result, the temperature gradient dT/dr along the loop is given roughly by T/L. This leads to a heat flux downward toward the surface $F = k_{\rm th} dT/dr \approx k_o T^{2.5} T/L$. The spatial divergence of this flux, div F can be written roughly as F/L. Therefore, if we express everything in c.g.s. units, the conductive contribution to the rate of cooling is $E_{\rm cond} \approx k_o T^{3.5}/L^2$ ergs cm^{-3} sec^{-1}.

17.14.2 Radiative losses

Although radiative cooling operates with reduced effectiveness in the corona compared to the chromosphere, this is not to say that the radiative losses from the corona should be neglected altogether. They *do* contribute to removing energy which has been deposited in the coronal gas.

In the chromosphere, our discussion of radiative losses was cast in terms of the local opacity (see the discussion in Chapter 15 after Equation 15.9). Now that we are in the corona, it is more convenient (see Section 17.7) to

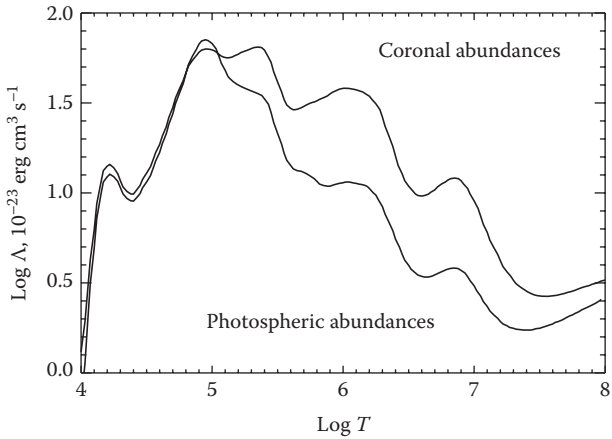


FIGURE 17.6: Radiative loss function (labelled Λ in this figure, rather than $\Phi(T)$) as a function of temperature, for two different sets of elemental abundances. Figure kindly provided by J. Raymond, Harvard-Smithsonian Center for Astrophysics (2008).

express the radiative losses in terms of an optically thin loss function $\Phi(T)$ (sometimes written as $\Lambda(T)$, where Λ is the Greek form of the initial letter of the word "loss"). The function $\Phi(T)$ (or Λ) is defined in such a way that for an optically thin medium at temperature T, the rate of energy loss from 1 cm^3 per second due to radiation is given by $N_e^2\Phi(T)$ (or $N_e^2\Lambda$). The numerical value of $\Phi(T)$ (or Λ) at any given T depends on the particular mixture of chemical elements which are present in the gas: if certain metals which "like" to emit at certain temperatures are abundant, then the value of $\Phi(T)$ (or Λ) will be larger at those temperatures. Results obtained by Raymond (2008) for $\Phi(T)$ for two different mixtures of elements are plotted in Figure 17.6. (Gas with "coronal abundances" differs from gas with "photospheric abundances" mainly in containing more easily ionized elements.)

For both chemical mixtures, the radiative loss function has a maximum value at a temperature of 0.1–0.2 MK. There is a sharp decrease in the loss function at temperatures below 10^4 K, and a more gradual decline (almost monotonic) in the loss function at temperatures between 1 and 10 MK. This behavior is reminiscent of the shapes of opacity curves (Chapter 3, Figure 3.3). This is no coincidence: the opacity due to lines and continua arises from the same atomic levels as are responsible for effective radiative losses in an optically thin plasma.

We note that in the conditions which are appropriate to the quiet solar corona, i.e., at temperatures of a few MK, the radiative loss function *decreases* as the temperature *increases*. We can approximate the slope of the curve in Figure 17.6 at temperatures between 1 and 10 MK by the function $\Phi(T) \sim 1/T^{0.5}$. Visual inspection of Figure 17.6 suggests that, for a gas with

photospheric abundances, the numerical value of $\Phi(T)$ in the temperature range 1–10 MK can be approximated by $\Phi(T) \approx 10^{-19}/T^{0.5}$ ergs cm^3 sec^{-1}.

Then the volumetric loss of energy by radiation $E_{\rm rad} = N_e^2\Phi(T) \approx 10^{-19}$ $N_e^2/T^{0.5}$ ergs cm^{-3} sec^{-1}.

17.14.3 Combined radiative and conductive losses

Now we can return to the energy balance in the corona. If mechanical energy is deposited at a volumetric rate of $E_{\rm mech}$ ergs cm^{-3} sec^{-1}, this can be balanced by the combined effects of radiation and conduction according to

$$E_{\rm mech} = E_{\rm rad} + E_{\rm cond} \approx 10^{-19} N_e^2 T^{-0.5} + 10^{-6} T^{3.5}/L^2 \tag{17.2}$$

In order to arrive at the lowest energy state, the corona adopts a temperature T_c such that the rates of radiative and conductive energy loss are comparable: this makes the cooling time-scale as long as possible. At temperatures which are either higher or lower than T_c, the cooling time-scale would be shorter, and the Sun would have to provide larger fluxes of mechanical energy to compensate for the more rapid cooling.

Equating the two terms on the right-hand side of the above equation at temperature T_c, we find that

$$T_c^4 = 10^{-13} N_e^2 L^2 \tag{17.3}$$

Let us convert from electron density to coronal pressure: $p = 2N_e k T_c$. This leads to

$$T_c^6 = \frac{10^{-13}}{4k^2}(pL)^2 \tag{17.4}$$

Inserting the numerical value of Boltzmann's constant $k = 1.38 \times 10^{-16}$ ergs K^{-1}, we find

$$T_c^6 = 1.3 \times 10^{18} (pL)^2 \tag{17.5}$$

Taking the sixth root of each side, we find

$$T_c = 1050 (pL)^{1/3} \tag{17.6}$$

where the units of T_c are degrees K.

What should we use for the pressure p? In the low corona, p is essentially equal to the pressure at the top of the chromosphere (see Section 17.9). At the top of the chromosphere (see Chapter 15, Section 15.12.4), $p \approx 0.6$ dyn cm^{-2}.

What values are appropriate for loop lengths in the corona? The discussion above about spatial structure indicates that L values range from a few times 10^9 to a few times 10^{10} cm. Substituting these values, we find that at the lower limit, $pL \approx 10^9$, leading to $T_c \approx 1$ MK. At the upper limit, $pL \approx 10^{10}$, leading to $T_c \approx 2$ MK.

It is interesting that simply by considering the lowest energy configuration, with equal loss rates in radiation and conduction, we have arrived at a range of coronal temperatures which spans the observed range in the quiet Sun (1–2 MK: Section 17.8).

To be sure, we have not solved the problem of coronal heating completely: Equation 17.6 depends on the two parameters p and L, which are determined by certain physical processes in the Sun. The value of p is ultimately determined by the flux of mechanical energy which is generated by the convection zone. And the values of L are controlled by the processes which generate magnetic fields in the Sun. So in calculating T_c by means of Equation 17.6, let us not read more than we should into the result: we *can* obtain a numerical result for T_c *only if* we first turn to *empirical results* in order to identify the most appropriate values to insert for p and L.

17.15 Abrupt Transition from Chromosphere to Corona

We have already noted (Section 17.9) that the transition between chromosphere and corona in the Sun occupies a spatial width that is quite narrow. The thickness of the transition is no more than 300 km, and may be much less. Now that we know about thermal conduction in the corona, we return to the question: why is the transition so abrupt?

The answer has to do with the fact that with the hot corona lying above the chromosphere, heat is conducted from the corona down into the chromosphere. The corona adopts a thermal structure such that a certain heat flux $F = -k_{\rm th} dT/dr$ flows downward, supplying energy *from above* to the upper chromosphere. This is distinct from the supply of (acoustic) energy which emerges from the convection zone, and which enters the chromosphere *from below*.

At coronal temperatures, $k_{\rm th} \sim T^{2.5}$ is so large that a given heat flux F can be transported by a small temperature gradient, dT/dr. But as the temperature decreases toward chromospheric values, the value of $k_{\rm th}$ decreases rapidly. In the upper chromosphere, where temperatures are lower than coronal values by factors of 100, $k_{\rm th}$ is a mere 10^{-5} times its coronal value. Therefore, in order to transport the same heat flux F downward, dT/dr must become 10^5 times *larger* in the chromosphere than in the corona. That is, the temperature gradient must become much steeper in the chromosphere than in the corona.

For numerical purposes, we note that in a coronal loop with half-length L, $dT/dr \approx T/L$. Inserting coronal values ($T = 10^6$ K, $L = 10^{9-10}$ cm), we see that in the coronal portion of a loop, dT/dr has a numerical value of typically $10^{-(3-4)}$ K cm^{-1}. In order to transport the same flux F in the much less conductive chromosphere, we need to increase dT/dr by 10^5. This leads to $dT/dr \approx$ 10–100 K cm^{-1}. Of course, the heat flux may not remain strictly

constant all the way down from the corona into the chromosphere: some of the energy may be dissipated along the way by radiative losses. But even if only 1–10% of the coronal heat flux survives into the chromosphere, a temperature gradient of order 1 K cm^{-1} would be required to transport that flux in the upper chromosphere. In the presence of such a gradient, the transition from the upper chromosphere (at $T = 10^4$ K) to a region where the temperature is (say) 10^5 K, would occur across a spatial distance of no more than 10^5 cm, i.e., 1 km.

Our previous estimate of the thickness of the transition ($\leq$300 km), based on a comparison of pressures (see Section 17.9), can easily accommodate such a steep conductive structure. On the scale of the solar radius, the transition region between chromosphere and corona is essentially a discontinuity.

The discussion in the present section is based on a highly idealized treatment of conditions in the corona and chromosphere, as if there is a unique temperature in the chromosphere at all locations, and a unique temperature in the corona. Actually, as we have already seen, there are spatial inhomogeneities in the chromosphere (spicules) and spatial structures in the corona (loops). In places where spicules exist, their lengths (up to several thousand km) allow them to extend well into the corona. As a result, the localized *roughness* of the solar surface is such that it is not accurate to think of a uniform spherically symmetric thin shell of thickness 1 km separating the chromosphere from the corona at all points on the solar surface. Instead, the transition region occurs at different heights above the photosphere in different locations, depending on local conditions. Nevertheless, wherever the transition from chromosphere to corona does occur, it is abrupt, occurring across spatial scales which may be as short as 1 km.

17.16 Rate of Mechanical Energy Deposition in the Corona

Since the corona in the quiet Sun seems to be essentially always at temperatures of 1–2 MK, it seems reasonable to conclude that the corona has reached an equilibrium: as fast as the mechanical energy is deposited in unit volume, the material in that volume disposes of the energy through radiation and conduction losses (at comparable rates).

Now that we know the temperature and density of the solar corona, we can evaluate the radiation loss rate $E_{\rm rad} = N_e^2 \Phi(T) \approx 10^{-19} N_e^2 / T^{0.5}$. In the quiet Sun, where $N_e \approx 4 \times 10^8$ cm^{-3} (Section 17.5) and $T = 1$–2 MK, we find $E_{\rm rad}({\rm QS}) \approx 10^{-5}$ ergs cm^{-3} sec^{-1}. In active regions, where N_e is enhanced over quiet Sun by as much as 4–5 and T is enhanced by 2–5, $E_{\rm rad}$ (AR) is enhanced over quiet Sun values by a factor of at least 5. Thus, $E_{\rm rad}$ (AR)$\geq 5 \times 10^{-5}$ ergs cm^{-3} sec^{-1}. Since conductive losses are essentially equal to the

radiative losses, we see that the rate of input of mechanical energy into the corona $E_{\rm mech}$ must be roughly as follows:

$$E_{\rm mech}({\rm QS}) \approx 2 \times 10^{-5} \text{ ergs cm}^{-3}\text{sec}^{-1} \tag{17.7}$$

$$E_{\rm mech}({\rm AR}) \approx 10^{-4} \text{ ergs cm}^{-3}\text{sec}^{-1} \tag{17.8}$$

It is worthwhile to compare these rates of energy deposition in the *corona* with the rates of mechanical energy deposition in the *chromosphere* (see Chapter 15, Equation 15.3): $E_{\rm mech}({\rm chr}) = 0.3$–$3$ ergs cm^{-3} sec^{-1}. In each cubic centimeter of the corona, mechanical energy is deposited at a rate which is three to five orders of magnitude *smaller* than in the chromosphere. The demands of the chromosphere for mechanical energy are much greater than the demands of the corona. Despite the large reduction in mechanical energy deposition rate in the corona, the greatly reduced density of coronal material (compared to chromospheric densities) allows even this low rate of energy input to heat the coronal gas to temperatures of 1–2 MK.

Now that we know the rate of emission $E_{\rm rad}$ (QS) from unit volume of the quiet corona, we can ask: how much power does the entire quiet corona emit in the form of radiation? The quiet corona is distributed over essentially the entire solar surface, i.e., it has an area $A({\rm QS}) \approx 4\pi R_\odot^2 = 6 \times 10^{22}$ cm^2. With an exponential scale height $H_c \approx 7 \times 10^9$ cm in the low corona (Section 17.5), the volume of the quiet corona is $V_c({\rm QS}) \approx A({\rm QS})H_c \approx 4 \times 10^{32}$ cm^{-3}. Multiplying this by $E_{\rm rad}$ (QS), we find that the radiative power of the quiet corona is $L_{\rm cor} \approx 4 \times 10^{27}$ ergs sec^{-1}. Comparing to the total power output from the Sun (Chapter 1, Equation 1.11), we see that the corona radiates at a rate which is only one millionth of the photospheric radiation rate.

17.17 What Heats the Corona?

The problem of coronal heating is an active topic of research interest. The heating hypotheses can conveniently be divided into two major groups: waves and nonwaves.

17.17.1 Wave heating

One possibility for heating the corona is that a flux of waves (of some sort) is entering the corona and dissipating. According to this viewpoint, by analogy with our discussion of the chromosphere (Chapter 15), the local volumetric rate of energy deposition would be given by the divergence of the flux of energy $F(w)$ in the waves. Also by analogy with our discussion of the chromospheric heating, the divergence can be approximated by the ratio $F(w)/\lambda_d$, where λ_d is a dissipation length (see Section 15.11). In a stratified medium, if dissipation is driven by nonlinear processes (such as shocks), λ_d might be a few times

the local scale height. If this line of reasoning applies to the low corona, this suggests that λ_d may be $(2\text{–}3)\times 10^{10}$ cm. In the context of this "wave model of coronal heating," the necessary wave flux $F(w) = E_{\rm mech}\lambda_d$ would be of order

$$F(\text{QS}) \approx 5\times 10^5 \text{ ergs cm}^{-2} \text{ sec}^{-1} \tag{17.9}$$

and

$$F(\text{AR}) \approx 3\times 10^6 \text{ ergs cm}^{-2} \text{ sec}^{-1} \tag{17.10}$$

Recalling that in the chromosphere, the wave flux $F(\text{chr})$ is estimated to be in the range 10^{7-8} ergs cm^{-2} sec^{-1} (see Chapter 15, Equation 15.2), we see that the coronal flux of waves (if waves are indeed the source of coronal heating) is smaller than the chromospheric wave flux by a factor of at least three, and maybe by as much as 200.

In the chromosphere, the nature of the waves which perform the heating (at least in the low-to-mid chromosphere) can be identified with a fair degree of confidence: acoustic waves emitted by the turbulent convection beneath the photosphere. As a quantitative confirmation of this hypothesis, the theory of sound emission from convective turbulence predicts enough wave flux to perform the required heating (Chapter 15, Section 15.12.2). Moreover, the fact that convection is always present, and occurs at all locations of the solar surface, means that the low-to-mid chromosphere is essentially spherically symmetric on the solar surface.

But what might be the source of waves which could be responsible for *coronal* heating? Could some acoustic waves be responsible? Since coronal heating is stronger in active regions, it is natural to suspect that waves of a magnetic nature might serve the purpose. Let us consider two candidates.

17.17.1.1 Acoustic waves?

What about acoustic waves? It is true that a large fraction of the acoustic flux coming up from the convection zone is dissipated in the chromosphere. But a "large fraction" does not necessarily mean "all." Might there be some acoustic wave flux "left over" at the top of the chromosphere? After all, even if as much as 99.5% of a flux of 10^8 ergs cm^{-2} sec^{-1} were dissipated in the chromosphere, the surviving 0.5%, i.e., 5×10^5 ergs cm^{-2} sec^{-1} would suffice to supply the necessary flux of energy to heat the corona, at least in the quiet Sun. The spherical symmetry of the acoustic flux could help to explain why the "fuzzy glow" of 1 MK gas in the quiet Sun can be found almost everywhere on the surface.

However, there are empirical reasons which make it difficult to accept the acoustic wave heating possibility: the last of the OSO missions (OSO-8, launched in 1975) was used to search for acoustic waves coming up into the corona. The flux was found to amount to no more than 7×10^4 ergs cm^{-2} sec^{-1} (Bruner, 1978). This is almost an order of magnitude smaller than what is required to heat even the quiet Sun corona (Equation 17.9). It seems that there are simply not enough acoustic waves to heat the corona.

17.17.1.2 Alfven waves?

We have seen (Chapter 16, Section 16.7.6) that the flux of Alfven waves in the photosphere $F_{\rm A}(\sim B)$ could be of order 10^{7-8} ergs cm^{-2} sec^{-1} in regions where B = 10–100 G. Now, we have also seen (Equation 17.9) that a wave flux of 5×10^5 ergs cm^{-2} sec^{-1} would suffice to heat the quiet Sun's corona. Comparison with F_A indicates that a field strength of even 0.5 G in the photosphere (quite weak by solar standards) *could* be sufficient to provide enough Alfven wave flux to heat the quiet Sun corona.

However, we must remember some important provisos. First, waves which are generated in the photosphere must survive into the corona, and second, the waves must be dissipated in the corona, i.e., on length scales of a few times 10^{10} cm. The first of these may be difficult to satisfy because Alfven waves tend to be reflected from a jump in density. Now, between the photosphere (where, with mass densities of 2–3 $\times 10^{-7}$ (Chapter 5, Table 5.3), the number densities are of order 10^{17} cm^{-3}) and the coronal base (where number densities are $\approx 10^9$ cm^{-3} Section 17.8), there is a reduction in density by $\varphi_d \approx 10^{-8}$. This jump in density occurs mainly across a length scale of 1–2 thousand km (the chromosphere), with a nearly discontinuous (smaller) jump between chromosphere and corona (see Section 17.15). Waves with periods of a few minutes (such as those emitted by gas circulating in granules), in regions where the Alfven speed V_A is at least a few km sec^{-1}, will have wavelengths $\lambda_w = PV_A$ which exceed 1000 km. In such a case, the Alfven waves will "sense" the change in density from photosphere to corona as essentially a discontinuity. Across a discontinuity where the density jumps by a factor of φ_d, the flux of the transmitted wave is only $4\sqrt{\varphi_d}$ of the incident flux (Alfven and Falthammer, 1963). Using the above estimate of $\varphi_d \approx 10^{-8}$, the fraction of the incident waves which survives into the corona is only 4×10^{-4}. Thus, even if we were to allow the field in the photosphere to be as large as 100 G, in which case the photospheric flux is as large as $F_A \approx 10^8$ ergs cm^{-2} sec^{-1}, the flux transmitted into the corona might be no more than 4×10^4 ergs cm^{-2} sec^{-1}. This is too small to supply the wave flux even for the quiet corona (Equation 17.9).

It is possible that, rather than restricting attention to Alfven waves which are generated in the photosphere, we should consider that the Sun may generate Alfven waves elsewhere. Magnetic reconnection (see Section 17.18.8) events in the chromosphere or in the corona might provide localized sources of Alfven waves. But it is hard to obtain quantitative estimates of wave fluxes from such events.

17.17.2 Non-wave heating: the magnetic carpet

A hypothesis which relies on magnetic fields, but which may also explain the near-spherical symmetry of the quiet Sun corona, is called the "magnetic carpet" (Figure 17.7).

In Figure 17.7, the background (with local hot spots in white) represents a segment of an image of the Sun taken in the same Fe XII line as

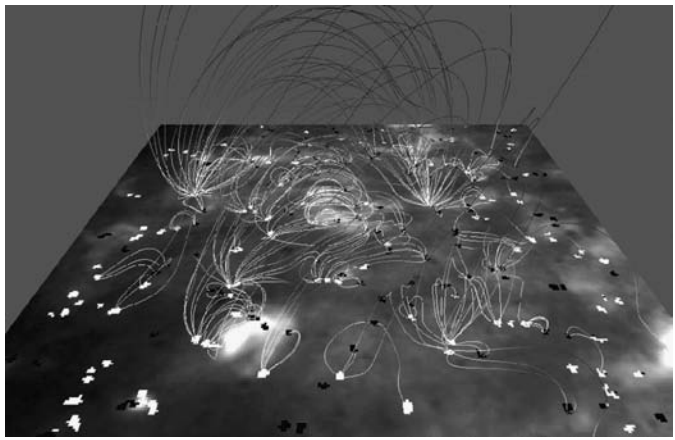


FIGURE 17.7: **(See color insert following page 202.)** Illustration of the "magnetic carpet" in the solar atmosphere. (Image available for public use at the NASA/Goddard website http://stargazers.gsfc.nasa.gov/images/sun_images/sunspots_cmes_occur/aerialcarpt.gif)

in Figure 17.4 above. The background reminds us that the "fuzzy glow" of material at 1–2 MK is present at most locations in the solar corona. Superposed on the image are magnetic field lines. Each field line is rooted in a pixelated feature on the Sun's surface which is either white or dark. White and dark pixels represent opposite magnetic polarity: therefore, each field line begins and ends in opposite polarity spots, looping upward between one white foot point and one dark foot point. As regards linear scale, each pixelated feature is some 10^4 km across. (On this scale, 700 pixels $= 1R_\odot$.)

Any given pixelated feature may have multiple field lines emerging from it, each connecting to a separate pixel of opposite polarity at another location. Thus, although field lines fill up most of the available space at high altitudes, when one approaches the surface, the field lines become clumped, or concentrated into specific "threads" which emerge from specific locations. The analogy to a domestic carpet, with its multiple clumps of thread emerging from a substrate, in which all clumps are "rooted," leads to the nomenclature "magnetic carpet" for the structure which dominates in the low solar corona.

However the analogy with a domestic carpet goes only so far. The Sun's magnetic carpet is *not* a static structure: far from it. Temporal variability is *of the essence* because new magnetic fields are continually emerging from beneath the solar surface. The image in Figure 17.7 is only a single "still frame" taken from a movie which shows dramatic variability. Any particular one of the many white and dark pixelated features which appear in the "still frame" in Figure 17.7 actually emerges at a certain point in time, splits apart or coalesces with neighbors, drifts around the surface because of granule motions, and eventually disappears. The time-scale during which any given feature survives

from emergence to disappearance has been found to be 1–2 days. As the features on the surface evolve, driven by continual emergence of new magnetic fields, the magnetic field loops in the solar atmosphere also evolve, expanding, contracting, distorting in a multitude of ways (Schrijver et al., 1997).

One possible outcome of this complex process is that field lines in the corona can find themselves in situations where they undergo a process known as "magnetic reconnection" (see Section 17.18.8 below). Magnetic energy is released in each reconnection event, and converted into heat. It is possible that the energy which is released as a result of the multitude of reconnection events which occur in the magnetic carpet every second may be the primary source of energy which heats the coronal plasma. And because there is some field present in essentially all parts of the Sun's surface at any given time, coronal heating due to the magnetic carpet could account for the "fuzzy glow" which permeates most of the solar surface (see Figure 17.4).

17.18 Solar Flares

The most spectacular phenomena on the Sun are associated with explosive events called flares. These are transient short-lived brightenings which can be observed in all regions of the electromagnetic spectrum. Here, we first describe the general properties of flares, and then examine the physical processes which are at work.

17.18.1 General

Flares are difficult to see in the optical continuum ("white light flares") because the surface of the Sun is so bright: in a 2.5-year period around solar maximum in 1980, only 12 white light flares were detected (Neidig and Cliver, 1983). Flares are much easier to detect in chromospheric lines: in a two-year period 1978–1979, 15,500 flares were recorded in Hα (Kurochka 1987). Flares are also easy to detect in X-rays, where the brightness of the entire Sun can increase by an order of magnitude or more in a matter of minutes or even seconds, and then decay to the previous level of brightness. A plot of intensity of radiation as a function of time is referred to as a "light curve."

Examples of flare light curves, as recorded by one particular series of X-ray satellites (GOES 10,12) are shown in Figure 17.8. Along the abscissa is plotted the time, covering a range of three days in 2005, when the Sun was moderately active. The ordinate shows the soft X-ray flux measured by GOES in two different wavelength (energy) ranges: 1–8 Å (1.5–12 keV) (upper curves), and 0.5–4 Å (3–25) keV (lower curves). Flares emit more energy at lower energies (upper curves) than at higher energies (lower curves). The labels A, B, C, M, and X along the right-hand side indicate the class of flare, depending on the peak flux observed in the GOES $1 - 8$ Å channel.

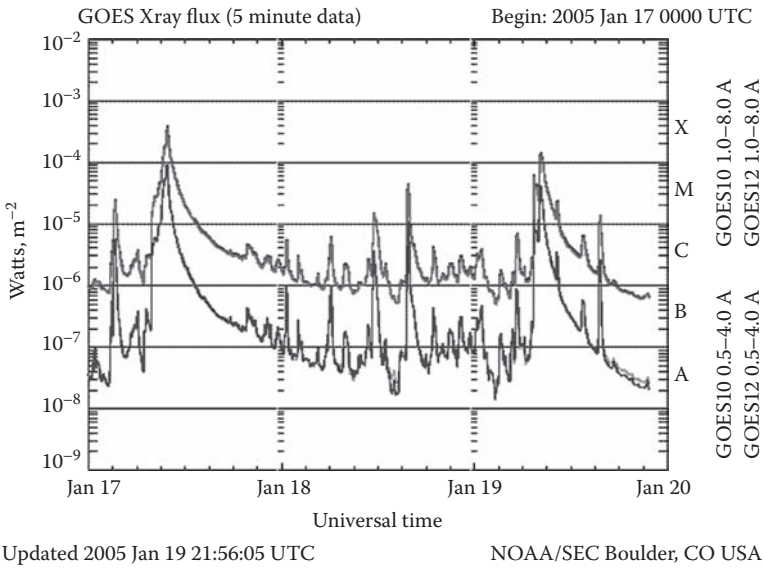


FIGURE 17.8: **(See color insert following page 202.)** Time profiles of soft X-ray flux emitted by the entire Sun in two ranges of wavelengths (ranges of energies) during a three-day interval. Data are in the public domain. (From Space Weather Prediction Center, Boulder, CO; National Oceanic and Atmospheric Administration; U.S. Department of Commerce. With permission.)

A flare is characterized by a rapid increase in radiative flux, followed by a slower decline. When data are obtained simultaneously in other wavelength ranges (radio, optical and UV spectral lines), flares are also apparent as rapid increases in those spectral regions, followed by slower declines. The various time-scales may be different in different wavelength ranges. In the hardest X-rays, the rise time-scale of a flare may be very abrupt, no more than one second. As can be seen in Figure 17.8, the largest flare events observed in soft X-rays may take the better part of one day to return to the "quiet Sun level," whereas the smaller events are complete in a matter of a few minutes.

17.18.2 Plasma temperature and density

In Figure 17.8, results are given for X-rays with two different energies. By comparing the relative fluxes of flare X-rays of different energies, an estimate of flare temperature can be obtained. These temperatures T_f are found to range from 5 to 25 MK: the largest flares typically have higher temperatures (Feldman et al., 1996). In comparison with the quiet (i.e., nonflaring) corona, where $T = 1$–2 MK, we see that flares can involve a temperature increase by as much as ten or more.

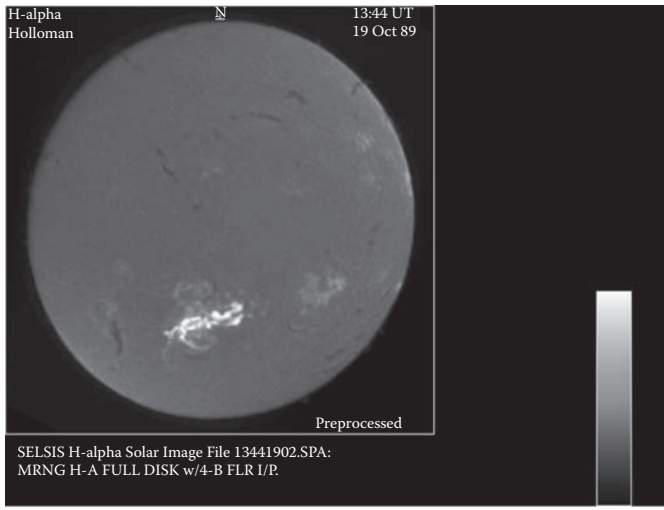


FIGURE 17.9: Image of the full Sun taken in the Hα line during a large solar flare. The flare is the brightest area in the image, occupying a significant fraction of the active region in which the flare occurred. (Image available to the public on the NASA wevsite http://nssdc.gsfc.nasa.gov/image/solar/killerha.gif)

Electron densities N_f in coronal flare plasma are also larger than the densities in the quiet Sun. For example, in a sample of flares analyzed by Moore and Datlowe (1975), the mean N_f was found to be $0.2–0.7 \times 10^{11}$ cm^{-3}, depending on certain assumptions about the flare volume. Compared with the quiet Sun densities $0.4–1 \times 10^9$ cm^{-3} (see Sections 17.5, 17.8), the flare densities are larger by a factor of up to 100.

17.18.3 Spatial location and extent

The X-ray detector on the GOES spacecraft (Figure 17.8) cannot make an image of the Sun. The detector measures flux from the entire Sun. As a result, there is no way to tell on which part of the Sun any given flare occurred. But other instruments, which can make images in optical photons, demonstrate unambiguously that flares occur in active regions, especially in active regions where sunspots with opposite polarities are in close spatial contact. This indicates that flares draw their energies (by a process to be discussed below) from magnetic fields.

When the largest flares are imaged in Hα, they are found to spread out spatially to cover an area which is a significant fraction of the area of the active region (see Figure 17.9). The linear extent of the largest flares on the surface of the Sun can be several times 10^9 cm, with areal coverage of order 10^{19} cm^2. In the vertical direction, flare plasma may extend up to altitudes

which are not greatly different from the extent on the surface. As a result, the total volume V_f of a large solar flare can be 10^{28-29} cm^3.

Although most of the attention in solar flare research is paid to the rapid (and spectacular!) energy release process, there is actually another process which also deserves attention: the build-up and storage of energy in the magnetic field prior to a flare. The latter process is much less dramatic than the flare itself. Build-up and storage occurs slowly over time intervals of hours and days. It occurs in active regions because magnetic field lines are rooted in the photosphere, where convective motions are ubiquitous. At the footpoint of a magnetic loop, convection pushes the field lines in complicated ways, with twistings and stretchings and braidings. These complex motions causes stresses to build up in the field lines, and the stressed fields serve as a reservoir in which free magnetic energy is stored. The build-up of free energy can continue as long as no process occurs to stop it. At certain times and places, a trigger comes into operation, and this releases the stored free energy quickly as a flare. The trigger causes a transition from a long-drawn out storage process (on time-scales of 10^5 sec) to a rapid release process (on time-scales of seconds): the two processes occur on time-scales which differ by as much as five orders of magnitude. We will return to this in Section 17.18.9.

17.18.4 Amount of energy released

In order to appreciate the physics of flares, we first need to determine the amount of energy which is released in one such event.

Solar flares come in a very broad range of energies. Some are so small that the only evidence for an event is a slight short-lived increase in brightness in the chromospheric line Hα. Essentially the flare energy emerges purely in the form of photons. Measurements provide the excess luminosity in Hα over and above the luminosity in the quiet Sun. Integrating the excess luminosity over the lifetime of the flare yields an energy E_α for the flare in Hα photons. However, if we wish to evaluate the total energy of a flare, we have to count up the energy which emerges in other photons, and in other forms.

For example, in large flares, hard X-rays emerge with photon energies of tens or 100s of keV. Such X-rays indicate the presence of fast electrons, also with energies of tens or 100s of keV, i.e., much higher than the typical thermal energies in the corona. (In 1 MK plasma, thermal energies are of order 0.1 keV.) Large flares are quite efficient at accelerating these "nonthermal electrons." From the shape of the X-ray spectrum, information can be extracted about the energy distribution of the nonthermal electrons: typically, the spectrum is found to be a power law in energy:

$$\frac{dN_e}{dE} \sim E^{-\delta} \tag{17.11}$$

where the index δ is found to range between two and five, with most values between three and 4.5 (Brown et al., 1981). To evaluate the total energy

$E_{\rm nt}$ contained in the nonthermal electrons in a flare, we need to multiply dN_e/dE by E and integrate over the energy spectrum from a minimum energy $E_{\rm min}$ to a maximum energy $E_{\rm max}$. The result is

$$E_{\rm nt} \sim \frac{1}{E_{\rm min}^{\delta-2}} - \frac{1}{E_{\rm max}^{\delta-2}} \tag{17.12}$$

Since δ typically exceeds two, the value of $E_{\rm nt}$ is determined mainly by the term which depends on $1/E_{\rm min}^{\delta-2}$. The smaller $E_{\rm min}$ is, the larger $E_{\rm nt}$. It is not easy to determine the value of $E_{\rm min}$ reliably, but it is probably no more than 20 keV. If we set $E_{\rm min}$ = 10–20 keV, then $E_{\rm nt}$ can reach values approaching 10^{32} ergs. Apparently, a flare site is very effective at accelerating many electrons to energies of tens of keV.

As well as the energy which is contained in fast electrons, flares also contain other forms of energy (see, e.g., Sturrock 1980). Thermal energy in flare plasma ($\approx N_f k T_f V_f$) over and above the energy density in an equal volume of "quiet" coronal plasma can be 10^{30} ergs and more. Moreover, some flares inject energetic protons and other nuclei into the solar wind: in large solar flares, the energy contained in the fast protons may be comparable to the energy in fast electrons. Finally, some of the largest flares eject bulk material from the corona into the solar wind: in a flare volume of up to 10^{29} cm^3 (Section 17.18.3), the preflare corona in the active region, with densities of 4×10^9 cm^{-3} contained a mass of order 10^{15} gm. A flare may eject all of this mass. In the largest mass ejections from the Sun, ejecta mass $M_{\rm fe}$ may be as large as 10^{15-16} gm (Jackson and Howard 1993), and the ejection velocity $V_{\rm fe}$ may be as fast as 1000 km sec^{-1}. Thus, the kinetic energy of ejected material, $E({\rm kin}) = 0.5 M_{\rm fe} V_{\rm fe}^2$, can reach values as large as 10^{32} ergs.

Adding up the various channels among which flare energy is distributed, estimates indicate that the largest solar flares are found to have energies of a few times 10^{32} ergs. One of the largest events reported by a space-borne total irradiance monitor (Kopp et al., 2005) had a total energy of $6 \pm 3 \times 10^{32}$ ergs.

At the opposite extreme of energies, it may not really be meaningful to speak of "the smallest solar flare." Depending on the sensitivity of the detector, and on how quiet the Sun is in the wavelength of measurement, one may be able to pick out smaller and smaller events which might qualify as "flares." In hard X-rays (where the "quiet corona" emits at a very low level), events can be identified as flares with E (tot) as small as $\approx 10^{26}$ ergs (Lin et al., 1984). These small events are sometimes referred to as "microflares": the prefix "micro" indicates they have energies which are 10^6 times smaller than the largest flares. If it could ever be demonstrated with confidence that there are events in the Sun with E(tot) as small as (roughly) 10^{23} ergs, they might justifiably be referred to as "nanoflares."

Is there anything on Earth to which we can compare solar flares in terms of energy release? Well, in a nuclear explosion which is rated as equivalent to one megaton (MT) of TNT, the energy released is 4×10^{22} ergs. Each "nanoflare" in the Sun (if such events occur) therefore corresponds to a 2.5 MT explosion.

17.18.5 Numbers of small and large flares

Now that we know that solar flares come in a broad range of energies, an important physics question is the following: do large flares occur as often as small flares? The answer is a clear No: small flares occur much more frequently than large ones. When a large sample of flares is analyzed, it is found that the differential flare distribution, i.e., the number of flares dN_f which have total energies between E_t and dE_t, follows a power law distribution.

$$\frac{dN_f}{dE_t} \sim E_t^{-\beta} \tag{17.13}$$

For example, in a sample of 15500 flares detected in Hα in the years 1978–1979, and converting energy observed in Hα to total energy E_t (including all lines and continua from hydrogen), Kurochka (1987) obtained a good fit to the data with Equation 17.13 using $\beta \approx 1.77$. In the two years of Kurochka's observations, only one (large) flare was observed with $E_t \geq 10^{32}$ ergs, whereas some 10000 (small) flares were observed with $E_t \geq 10^{27}$ ergs. The observing noise was such that Kurochka was unable to identify any flare events with E_t less than (roughly) 10^{27} ergs. For all we know, the Sun may be generating millions of "microflares" and/or "nanoflares" every year, but such events are, up to the present time, generally lost in the observational noise.

The existence of a power law distribution of flare energies contains important information about flare physics: there is no "characteristic energy" associated with solar flares. Instead, there is simply a broad range of energies available to the solar atmosphere when it "decides" to create a solar flare.

This leads us to raise the question: might there be a true physical *upper limit* to the energy of a solar flare? Or is the "empirical upper limit" of $6 \pm 3 \times 10^{32}$ ergs (mentioned above) merely a limitation set by the finite amount of time that the Sun has been observed? If we were to observe the Sun for another 100 or 1000 years, would the Sun occasionally perhaps have a flare with even larger energies? The answers to these questions are currently unknown. We will return to this issue of maximum energy of flares when we discuss coronal mass ejections in Chapter 18.

17.18.6 Do flares pose a significant perturbation on solar structure?

The largest flares, with energy releases of a few times 10^{32} ergs, certainly involve important disruptions of the active region in which they occur: field lines connecting different spots in the active region are rearranged, and some matter may be ejected into space. However, in the larger context of the Sun as a whole, flares represent only a small perturbation.

To see this, we note that even the largest flares have durations which are no more than a fraction of a day (see Figure 17.8). During such an interval (e.g., 50,000 sec), the nuclear reactions in the core of the Sun continue to pour out energy at the standard rate, i.e., 4×10^{33} ergs sec^{-1}. Therefore, over the

duration of a large flare, the Sun puts out 2×10^{38} ergs of nuclear energy. Compared to this, the energy released in even the largest flares is only a small perturbation of the solar energy budget.

According to Equation 17.13, smaller flares occur more frequently. However, the total energy $E_t(\text{fl})$ emitted in the form of flares is obtained by performing the integral of $E_t dN_f/dE_t$ over all flares from the smallest energy E_t (min) to the largest energy E_t(max). In the integral, it is important that the index β in Equation 17.17 has a value which is *smaller than* two. As a result, $E_t(\text{fl})$, which is given by the expression

$$E_t(\text{fl}) \sim E_t(\text{max})^{2-\beta} - E_t(\text{min})^{2-\beta} \tag{17.14}$$

is dominated by E_t(max), i.e., by the *largest* flares.

Equation 17.14 shows that, if it can ever be demonstrated that there exist a class of flares where the value of β is *in excess of* two, then the *smallest* flares in that class would occur so frequently that *they* would dominate the total flare energy budget.

17.18.7 Energy densities

Now that we know, for large flares, that $E_t(\text{fl}) \approx 10^{32}$ ergs, while the flare volume $V_f \approx 10^{28-29}$ cm^3, we can estimate the mean energy density in large flares. The result is $E_t(\text{fl})/V_f \approx 10^{3-4}$ ergs cm^{-3}.

What is the origin of this energy density? Since flares are observed to occur in active regions, it is reasonable to expect that the magnetic field is somehow responsible. In quantitative terms, we have already seen (Chapter 16, Section 16.6.2.1) that magnetic fields have an energy density $W_{\text{mag}} = B^2/8\pi$ ergs cm^{-3} if B is expressed in Gauss.

Magnetic fields have energy densities of 10^{3-4} ergs cm^{-3} in regions where the field strengths are 160–500 G. Now, we have already seen that active regions in the Sun have fields of hundreds of Gauss (Chapter 16, Section 16.5). Thus, the energetics suggest that the magnetic fields in active regions are sufficiently strong that they *could* (at least in principle) supply the energy which is released in a flare. The trick is to identify a mechanism which has the ability to convert magnetic energy into fast electrons and heat. Let us now turn to one such mechanism.

17.18.8 The physics of flares: magnetic reconnection

Magnetic reconnection is a process which converts magnetic energy into plasma kinetic energy, and eventually thermal energy. According to the Sweet–Parker mechanism (named after the two researchers who first proposed a model, see Parker, 1957), reconnection occurs when two regions of magnetic field, with oppositely directed field lines, approach each other with (slow) speed u (see Figure 17.10). In the limit of MHD, where fields and plasma are "frozen" together, plasma plus field comes in from top and bottom, and

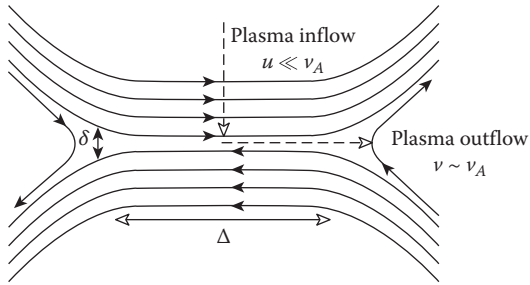


FIGURE 17.10: Schematic of magnetic reconnection. Plasma flows in slowly from top and bottom toward the central region, where field lines of opposite polarity come into close contact. The result is that plasma emerges at high speed (the Alfven speed) toward the left-hand and right-hand sides. (From Hantao Ji (PPPL). With permission.)

plasma plus field "squirts out the side" at (faster) speed v. As a result of the inflow, a plasma sheet builds up between the regions of opposite polarity. In this sheet, the magnetic field changes sign, passing through values close to zero in the process. The region of zero field strength (the neutral line) lies close to the horizontal dashed line in Figure 17.10. In equilibrium, the thickness of the sheet (δ in Figure 17.10), and the length of the sheet (Δ in Figure 17.10) take on values which are determined by mass conservation: $\delta/\Delta \approx u/v$.

The field lines in the outflow are farther apart than in the inflow, indicating that the emergent fields are *weaker* than the ingoing fields. Since magnetic energy is proportional to B^2, the magnetic energy in the outflow is *reduced* compared to the inflow. To conserve energy, the original magnetic energy is converted into another form: kinetic.

The central process is that an ion which is initially gyrating around the incoming field eventually finds itself swept into a region where the field strength is very weak. In that region, the gyroradius, which is proportional to $1/B$ (Equation 16.4), becomes so large that it exceeds the distance between the particle and the neutral line. In such conditions, the ion essentially "breaks loose" from the field, and the conclusions of MHD are no longer valid. The ion is now free to flow in a different direction, e.g., in the outflow direction.

The time-scale t_i required for an inflowing particle to traverse the plasma sheet is δ/u. Because of finite resistivity in the plasma, the magnetic field diffuses out of the plasma sheet on a time scale $t_d \approx 4\pi\sigma_e\delta^2/c^2$ (where σ_e is in electrostatic units: see Chapter 16, Equation 16.10). Parker (1957) suggested that the inflow would adjust itself to a value which would ensure that the two time-scales t_i and t_d are equal.

The influx of energy occurs mainly in magnetic form: the rate of influx per unit length is $uB^2\Delta/8\pi$. The outflux of energy occurs mainly in kinetic form: the rate per unit length is $0.5\delta\rho v^3$. In steady state, these rates are equal: this leads to $v^2 \approx B^2/4\pi\rho$. That is, the outflow proceeds at the Alfven speed V_A

(see Chapter 16, Equation 16.13). Recall that in the corona, there are locations where V_A can be as large as 60,000 km sec^{-1} (Chapter 16, Section 16.7.6).

Now we have enough information to determine, in terms of the local parameters, the inflow speed and the thickness that a steady-state Sweet–Parker reconnection layer must have in any coronal site. For the inflow speed u, we find:

$$\frac{u}{V_A} = \frac{c}{2}\sqrt{\frac{1}{\sigma_e \pi V_A \Delta}} \tag{17.15}$$

In coronal reconnection, Δ will be comparable in scale to the transverse dimensions of the magnetic loops which are reconnecting, i.e., Δ may be of order 10^9 cm. Using the Spitzer conductivity in the corona ($T = 1$ MK), i.e., $\sigma_e \approx 10^{16}$ e.s.u. (see Chapter 16, Section 16.6.2.3), we find that the inflow speed is of order 1 meter sec^{-1}. Compared to the speed at which diffusion would merge the field lines over length scales of order Δ (i.e., $V_d = c^2/4\pi\sigma\Delta$), the value of u at a reconnection site is several orders of magnitude faster. Reconnection really does speed up the process.

For the thickness of the Sweet–Parker reconnection layer, we find

$$\delta = \frac{c}{2}\sqrt{\frac{\Delta}{\sigma \pi V_A}} \tag{17.16}$$

Inserting numerical values, we find that δ is a few meters in the corona.

17.18.9 Triggering a flare

At first sight, we might be tempted to conclude that a structure with a thickness δ of only a few meters could hardly have any significant role to play in the Sun. After all, the magnetic field structures which we see in active regions have linear scales of 10^{8-10} cm, i.e., orders of magnitude larger than δ. But such a conclusion would miss an essential point of reconnection: a reconnection site is a *dynamic* structure where material and fields participate *in a symbiotic way* of being "swept up" together and brought into the reconnection sheet. As time goes on, more and more field can be swept in to be "processed" through the reconnection site. By this means, magnetic energy which was originally distributed over a large volume (as much as 10^{28-29} cm^3) can be converted to kinetic energy during the course of a flare.

However, there is a problem with the time-scale: with $u \approx 100$ cm sec^{-1}, material which was originally located in a flux tube with linear width of 10^{8-10} cm would require some 10^{6-8} sec ($\approx$10–1000 days!) to reach the reconnection sheet. This is much too slow to be relevant for a flare, where rise-times are seen to be as short as seconds. In order for a flare to occur as we see it on the Sun, something must happen to increase the local value of u dramatically. What could this be?

To address this, we note that the Sweet-Parker model is a model which is based on a theory of fluid flow known as magnetohydrodynamics (MHD).

The key word in the previous sentence is "fluid," i.e., a continuous medium in which any small volume element in the fluid contains a very great number of molecules. However, in a situation where the thickness δ of the "fluid" layer becomes smaller than a certain critical length-scale, the assumptions of MHD break down. Two possible critical length-scales in this regard are the ion gyroradius r_g (see Chapter 16, Equation 16.4) and the ion inertial length, $d_i = c/\omega_{\rm pi}$, where $\omega_{\rm pi}$ is the ion plasma frequency ($=\sqrt{(4\pi N_i e^2/m_i)}$). In a plasma where magnetic pressure and thermal pressure are comparable, these two length scales are also comparable, i.e., $r_g \approx d_i$.

Computer simulations of reconnection show that if δ decreases to a value which is so small that $\delta \leq d_i$, the dynamics of the reconnection sheet are no longer dominated by the ordinary electrical conductivity σ_e (as the Sweet–Parker model assumes). Instead, the dynamics becomes dominated by a process known as the Hall effect: this is a collisionless process whereby charged particles drift perpendicular to the E and B fields. When the Hall effect sets in, the simulations show that there is a dramatic increase in the reconnection rate: u increases by as much as 10^6. When that happens, the inflow time-scales for solar magnetic loops become as short as 1–10 sec. This is fast enough to be relevant to triggering solar flares.

Is there a way to check that the computer simulation results are indeed relevant to the onset of flares? Yes, it seems that there is. By evaluating typical values of δ (see Equation 17.20) and d_i for flares on the Sun and on other stars (Cassak et al., 2008), it appears that δ in flare conditions indeed takes on numerical values which are comparable to d_i. Not all active regions can actually satisfy this criterion: in such active regions, the slow Sweet-Parker process would be all that would ever happen. In such active regions, flares would not occur. This is consistent with observations which show that not *all* active regions produce flares: some active regions live out their life in a quiet nonflaring manner.

17.18.10 Consequences of magnetic reconnection

What consequences does reconnection have in the solar corona? Using typical coronal values of field and density, we find that the coronal V_A has values which are typically several hundreds of km sec^{-1} and more. Thus, the gas which emerges from a reconnection site forms a high speed jet which flows out into the ambient corona.

Collisional processes ensure that the jet energy will eventually be deposited in a finite volume around the reconnection site, leading to local heating. How hot will the heated plasma be? If all of the kinetic energy can be converted into heat, then the temperature T_f of the heated plasma will be such that the mean thermal speed $\sqrt{2R_gT/\mu}$ is comparable to V_A. For $V_A = 500$–1000 km sec^{-1}, and $\mu = 0.5$, this corresponds to $T_f \approx 10$–30 MK, consistent with temperatures deduced from X-ray data (Section 17.18.2). Thus, reconnection is a mechanism whereby magnetic energy is converted (ultimately) into the thermal energy of the hot flare plasma.

Another process which occurs at a reconnection site is particle acceleration. The motional electric field $E_f = (1/c)\mathbf{u} \times \mathbf{B}$ at the reconnection site can accelerate electrons: in conditions where E_f exceeds a certain limit E_D (known as the Dreicer field), the electrons experience "runaway" to high energies, 100 keV or more. In a large flare, as we have seen, the fast electrons may actually contain energies of 10^{32} ergs, i.e., a large fraction of the overall flare energy.

Finally, reconnection results in changes in field line connectivity: magnetic field lines which were previously closed, i.e., which previously looped back to the Sun's surface, can become open to interplanetary space. This can allow a volume of plasma which was previously trapped (on closed field lines) to escape from the Sun. This process may occur during events known as coronal mass ejections (see Chapter 18, Section 18.8).

In summary, a flare involves conversion of magnetic energy into three different channels: accelerated particles, thermal energy, bulk motion away from the Sun. Different flares divide up the magnetic energy in different proportions among these channels.

References

Alfven, H. and Falthammer, C.-G. 1963. "Magnetohydrodynamics," *Cosmical Electrodynamics: Fundamental Principles*, 2nd ed. Clarendon Press, Oxford, pp. 73–133.

Brosius, J. W., Davila, J. M., Thomas, R. J., and Monsignori-Fossi, B. C. 1996. "Measuring active and quiet Sun coronal plasma properties with extreme-ultraviolet spectra from SERTS," *Astrophys. J., Suppl.* 106, 143.

Brown, J. C., Smith, D. F., and Spicer, D. S. 1981. "Solar flares observations and their interpretations," in *The Sun as a Star*, ed. S. Jordan, NASA Sci. Tech. Inf. Branch, Washington DC, NASA SP-450, 200.

Bruner, E. C. 1978. "Dynamics of the solar transition region," *Astrophys. J.*, 226, 1140.

Cassak, P. A., Mullan, D. J., and Shay, M. A. 2008. "From solar and stellar flares to coronal heating: theory and observations of how magnetic reconnection regulates coronal conditions," *Astrophys. J. Lett.*, 676, L69.

Cranmer, S. R., Panasyuk, A. V., and Kohl, J. L. 2008. "Improved Constraints on the Preferential Heating and Acceleration of Oxygen Ions in the Extended Solar Corona," *Astrophys. J.*, 678, 1480.

Edlen, B. 1945. "The identification of the coronal lines," *Monthly Not. Royal Astron. Soc.*, 105, 323.

Feldman, U., Doschek, G. A., Behring, W. E., and Phillips, K. J. H. 1996. "Electron temperature, emission measure, and X-ray flux in A2 to X2 X-ray class solar flares," *Astrophys. J.*, 460, 1034.

Jackson, B. V. and Howard, R. A. 1993. "A CME mass distribution derived from Solwind coronagraph observations," *Solar Phys.* 148, 359.

Jordan, C. 1969. "The ionization equilibrium of elements between carbon and nickel," *Monthly Not. Royal Astron. Soc.*, 142, 501.

Kopp, G., Lawrence, G., and Rottman, G. 2005. "The Total Irradiance Monitor (TIM): science results," *Solar Phys.*, 230, 129.

Kurochka, L. N. 1987. "Energy distribution of 15000 solar flares," *Soviet Astron.*, 31, 231.

Lin, R. P., Schwartz, R. A., Kane, S. R., Pelling, R. M., and Hurley, K. C. 1984. "Solar hard X-ray microflares," *Astrophys. J. Lett.*, 28, 421.

Moore, R. L. and Datlowe, D. W. 1975. "Heating and cooling of the thermal X-ray in solar flares," *Solar Phys.*, 43, 189.

Neidig, D. F. and Cliver, E. W. 1983. "The occurrence frequency of white-light flares," *Solar Phys.*, 88, 275.

Newkirk, G. 1967. "Structure of the solar corona," *Annual Rev. Astron. Astrophys.*, 5, 213.

Parker, E. N. 1957. "Sweet's mechanism for merging magnetic fields in conducting fluids," *J. Geophys. Res.*, 62, 509.

Parkinson, J. H. 1973. "New observations of Fe XVII in the solar X-ray spectrum," *Astron. Astrophys.*, 24, 215.

Raymond, J. C. 2008. Harvard-Smithsonian Center for Astrophysics. Personal communication.

Schrijver, C. J., Title, A. M., Van Ballegooijen, A. A., Hagenaar, H., and Shine, R. A. 1997. "Sustaining the quiet photospheric network: the balance of flux emergence, fragmentation, merging, and cancellation," *Astrophys. J.*, 487, 424.

Sturrock, P. A. 1980. "Solar flares: a monograph from Skylab solar workshop II," Colorado Associated University Press, Boulder.

Tu, C. Y. and Marsch, E. 1997. "Two-fluid model for heating of the solar corona and acceleration of the solar wind by high-frequency Alfven waves," *Solar Phys.*, 171, 363.

van de Hulst, H. C. 1950. "The electron density of the solar corona," *Bulletin Astron. Inst. Netherlands*, 11, 135.

Wilhelm et al. 1998. "The solar corona above polar coronal holes as seen by SUMER on SOHO," *Astrophys. J.*, 500, 1023.

Chapter 18

Solar Wind

We have already seen that the solar corona, with its temperature in excess of 1 MK, is the site of some processes which are related to several distinct physical processes: high levels of ionization, effective thermal conduction, and emission of spectral lines over a broad range of the electromagnetic spectrum. Now we turn to a different physical property of the corona which also depends on the fact that the temperature is of order 1 MK. We shall find that, given the global properties of the Sun (specifically, its mass and radius), the corona cannot "stand still," but must undergo expansion.

In order to describe this aspect of the corona, we start by considering the equation of HSE. In previous chapters, we have found that HSE applies to certain locations in the Sun but not to others. For example, HSE applies in the radiative interior and in the photosphere/chromosphere, but not in the convective zone. Now we ask: does HSE apply to the corona?

18.1 Global Breakdown of Hydrostatic Equilibrium in the Corona

In order to apply the equation of HSE, $dp/dr = -\rho g$, to the corona and solar wind, we need to allow for the fact that $g = GM_\odot/r^2$, is no longer a constant. (In the photosphere, it is safe to assume that g is a constant as the height varies: see Chapter 5, Equation 5.2.). Now we must make allowance for the fact that g decreases as we move away from the Sun. This has a dramatic effect on the solution of the HSE.

To see this, let us assume at first that the coronal material is a perfect gas at a constant temperature: i.e., $p = R_g \rho T/\mu$. This allows us to write the HSE as

$$\frac{1}{p}\frac{dp}{dr} = -\frac{A}{r^2} \tag{18.1}$$

where $A = GM_\odot/a^2$, and $a = \sqrt{(R_g T/\mu)}$ is the isothermal sound speed. Notice that the quantity A has the dimensions of a *length.*

The solution of Equation 18.1 is $\ln(p) = (A/r) + \text{const}$. To evaluate the constant of integration, we consider the conditions at the base of the corona,

where $r = r_o$. From our discussion in Chapter 17, we know that r_o is essentially the radial location of the top of the chromosphere, which is in turn given by $r_o = R_\odot$ within a fraction of 1%. At this location, the pressure p_o is known to be 0.3–2 dyn cm^{-2} (see Chapter 17, Section 17.9).

Therefore, the coronal pressure as a function of radial distance is given by

$$p(r) = p_o \exp\left[A\left(\frac{1}{r} - \frac{1}{r_o}\right)\right] \tag{18.2}$$

As $r \to \infty$, Equation 18.2 shows that the pressure does *not* tend to zero. Instead, the pressure $p(\infty)$ approaches the asymptotic value $p_o \exp(-A/r_o)$.

The striking aspect of this solution is that it is a very different result from the one we would get if we were to extend the photospheric solution to infinity (see Chapter 5, Section 5.1). Using the photospheric solution $p(h) = p_o \exp(-h/H)$ where H is the scale height (with g=const), we see that as $h \to \infty$, $p(h)$ goes exponentially rapidly to zero. It is the fact that in the corona, g is *not* a constant, but decreases as r increases, which makes the solution in the corona qualitatively different. In the corona, there is *no* exponentially rapid decrease of pressure toward zero.

In order to evaluate the asymptotic pressure $p(\infty)$, we need to know the value of A. To calculate A, we need to choose a value for μ: what value should we use? In the corona, hydrogen and helium are completely ionized, just as they are in the deep interior; this suggests that we should use the value $\mu = 0.58$ which we used in the hot interior of the Sun (Chapter 7, Section 7.8). Inserting $T = 1$ MK and $\mu = 0.58$, we find $a^2 = 1.43 \times 10^{14}$ cm^{2}sec^{-2}. Combining this with the value of $GM_\odot$ (Chapter 1, Equation 1.9), we find $A \approx 9 \times 10^{11}$ cm, i.e., $A \approx 13 R_\odot$. For a coronal temperature $T = 2$ MK, we find $A \approx 6.5 R_\odot$.

As a result, if the entire corona were to be in HSE, the pressure of the coronal gas at infinity would be smaller than p_o by a factor $e^{-13} \approx 2 \times 10^{-6}$. Inserting $p_o = 0.3$–2 dyn cm^{-2} this would lead to $p(\infty) = (1 - 4) \times 10^{-6}$ dyn cm^{-2}.

What are we to compare this to? The answer is: the Sun does not exist in a vacuum, but is surrounded by the "interstellar medium" (ISM) which exists in the space between the stars. The ISM contains gas, dust, magnetic fields, and energetic particles ("cosmic rays"). The ISM near the Sun contains hydrogen with number densities $n_a \approx 0.14$ cm^{-3}, electrons with $n_e \approx 0.07$ cm^{-3}, and temperatures T of order 10^4 K (e.g., Gayley et al., 1997). The gas pressure in the ISM, $p(\text{ISM}) \approx (n_a + n_e)kT$, is therefore, $\approx 3 \times 10^{-13}$ dyn cm^{-2}. The mass of dust contributes only about 1% of the mass of gas, and contributes negligibly to the pressure. The magnetic fields, with strengths of 1.6–3 $\times 10^{-6}$ G (Gayley et al., 1997), contribute pressures $B^2/8\pi$ of order 0.1–0.4 $\times 10^{-12}$ dyn cm^{-2}. And the cosmic rays contribute pressures of $\approx 10^{-12}$ dyn cm^{-2} (Ip and Axford, 1985). The combined effects of all ISM constituents provide $p(\text{ISM}) \approx 1.4$–$1.7 \times 10^{-12}$ dyn cm^{-2}.

Compared with $p(\infty)$, we see a large difference in the pressure: $p(\infty)$ for a corona in HSE exceeds p(ISM) by five to six orders of magnitude. As a result, it is physically impossible for the ISM to contain the pressure of the solar corona if the latter is in HSE.

The conclusion is that *the corona cannot be in HSE.* If the corona does not have the property of being hydro*static*, what other option is available? The answer is: the corona must be hydro*dynamic.* This means that the coronal material must undergo expansion, moving from high pressure (at the base of the corona) to low pressure (in the ISM). This expansion has nothing to do with evaporation: instead it involves the bulk flow of the coronal material outward from the Sun. This outflow of bulk coronal material from high pressure to low pressure is described by the term s*olar wind*, by analogy with winds on Earth, which (in the absence of Coriolis forces) blow from high to low pressure.

18.2 Localized Applicability of HSE

Are there any conditions in which an entire corona *could* be in HSE? In principle, yes: it could happen if $p(\infty)$ were to have a value no larger than $p(\mathrm{ISM})$. If we could achieve that goal, then the solar coronal pressure *could* be contained by the ISM. One way to achieve that goal, would be to reduce $p(\infty)$ by reducing the coronal temperature, thereby increasing the numerical value of A. According to Equation 18.2, the value of $p(\infty)$ could be made as small as $p(\mathrm{ISM}) \approx 10^{-12}$ dyn cm^{-2} if A were to have a value as large as ≈ 26–$28 R_\odot$. This would require $T \approx 0.46$–0.5 MK.

However, as we have seen, the solar corona is not as cool as this: there are physical reasons (see Chapter 17, Section 17.14.3) for the quiet solar corona to have $T = 1$–2 MK. (The active corona is even hotter.) Given the actual values of the physical constants which enter into the electron thermal conductivity and the radiative losses, we simply are not free to make the coronal temperature as low as 0.46–0.5 MK. Therefore HSE is *not* applicable to the entire corona: given the observed properties of the Sun ($M_\odot$, $R_\odot$), and the empirical coronal temperatures, we are led to the conclusion that hydrodynamic expansion is an *intrinsic global property* of the solar corona.

Although HSE is certainly not applicable *in a global sense* to the corona, this should not be construed to mean that HSE is *absolutely excluded in each and every locality* of the corona. On the contrary, in certain favorable localized regions, it may turn out that circumstances *do* allow HSE to be applied locally.

Two examples can be considered. First, in a closed magnetic loop (Chapter 17, Section 17.11), magnetic forces prevent ionized gas from escaping into the wind: hydrodynamic outflow is not allowed to occur. Within the confines of such a loop, HSE *may* be a good approximation to the profile of density as a function of height.

Second, it can be shown from fluid dynamics that, in the limit where outflow speeds are much less than a (the speed of sound), the hydrodynamic solution approaches the hydrostatic solution. We shall see (Equation 18.8) that the flow speed of the solar wind does (eventually) indeed become as large as a (and larger), but this happens only at radial distances which are at least several $R_\odot$ from the Sun. It is critical to note that, there *is* a finite range of

radial distances, close to the Sun, where the wind speed is actually *much less* than a. Within this region, HSE *can be used* as a reasonable approximation for the density. We shall take advantage of this result in Section 18.7.

18.3 Solar Wind Expansion: Steady State Flow

The breakdown of HSE in the corona means that dp/dr cannot be equal to $-\rho g$. The imbalance of forces between the pressure gradient and gravity causes the coronal gas to accelerate. The conservation of momentum, when applied to unit volume of the corona (in which the mass equals ρ, the local density), leads (see Chapter 7, Equation 7.1, replacing z with r) to the equation

$$\rho \frac{dV}{dt} = -\frac{dp}{dr} - \rho g \tag{18.3}$$

The total time-derivative d/dt can be written as the sum of two terms: $\partial/\partial t + V\partial/\partial r$. In a steady state situation, where the flow does not depend explicitly on time, only the radial gradient term is present. In a situation where only radial gradients are important, we can write $\partial/\partial r$ as the ordinary derivative d/dr. Then inserting $g = GM_\odot/r^2$, Equation 18.3 becomes

$$V \frac{dV}{dr} = -\frac{GM_\odot}{r^2} - \frac{1}{\rho}\frac{dp}{dr} \tag{18.4}$$

Let us consider the simplest case, in which the corona is assumed to be isothermal, i.e., $T = \text{const}$ at all radial locations. With this assumption, we are greatly simplifying the equation for the conservation of energy. In an isothermal corona, the pressure and density are related at all locations by the formula $p(r) = a^2\rho(r)$. This allows us to rewrite Equation 18.4 as follows:

$$V \frac{dV}{dr} = -\frac{GM_\odot}{r^2} - \frac{a^2}{\rho}\frac{d\rho}{dr} \tag{18.5}$$

Turning now to conservation of mass, we note that, at a radial distance r, the rate of mass outflow from the Sun in a spherically symmetric wind is given by $dM/dt = 4\pi r^2 \rho(r) V(r)$. Once the solar wind leaves the corona, and flows out into interplanetary space, no further significant mass can be added to the outflow. Therefore, dM/dt is independent of r, i.e., $r^2\rho(r)V(r) = \text{const}$. Thus the radial derivative of $r^2\rho(r)V(r)$ is zero. Taking logarithms, this means that

$$\frac{1}{\rho}\frac{d\rho}{dr} + \frac{1}{V}\frac{dV}{dr} + \frac{2}{r} = 0 \tag{18.6}$$

Using Equation 18.6, we can replace the final term in Equation 18.5. Then collecting terms in the radial gradient dV/dr, we obtain an equation for dV/dr:

$$\left(V - \frac{a^2}{V}\right)\frac{dV}{dr} = \frac{2a^2}{r} - \frac{GM_\odot}{r^2} \tag{18.7}$$

The structure of this equation indicates that dV/dr can be written as the ratio of two terms, $\mathcal{N}(r)/\mathcal{D}(r)$. The numerator $\mathcal{N}(r)$ is the radial function on the right-hand side of Equation 18.7, while the denominator $\mathcal{D}(r)$ is the radial function $V(r) - a^2/V(r)$.

At a certain radial location, the function $\mathcal{D}(r)$ passes through the value zero. This occurs when the wind speed $V(r)$ becomes equal to the sound speed, $a \approx 120\sqrt{T_6}$ km sec^{-1} (where T_6 is the temperature in MK). In coronae with $T_6 = 1$ and 2, the sound speeds are $a \approx 120$ and $a \approx 170$ km sec^{-1}, respectively. The radial position where $V(r) = a$ is referred to as the "sonic point." In order to prevent dV/dr from becoming infinitely large at the sonic point, $\mathcal{N}(r)$ must also pass through the value zero at the sonic point.

Setting $\mathcal{N}(r) = 0$ at radial location $r = r_s$, we find that the sonic point lies at

$$r_s = \frac{GM_\odot}{2a^2} \tag{18.8}$$

Inserting the numerical value $GM_\odot = 1.327124 \times 10^{26}$ cm^3 sec^{-2} (Chapter 1, Equation 1.9), we find that in a corona with $T = 1$ MK (i.e., $a^2 = 1.43 \times 10^{14}$ cm^2 sec^{-2}), the value of r_s is 4.6×10^{11} cm. In a corona with $T = 2$ MK, $r_s \approx 2.3 \times 10^{11}$ cm. Compared with the solar radius, we see that the sonic point lies at radial locations of $r_s \approx 6.6R_\odot$ and $3.3R_\odot$ for $T = 1$ and 2 MK, respectively.

Thus, in response to the breakdown of HSE, the material in an isothermal corona is accelerated outward, increasing from essentially zero velocity at the base of the corona to a velocity as large as the sound speed at radial locations of a few solar radii. With $T = 1$ MK, the wind reaches a velocity of ≈ 120 km sec^{-1} at $r \approx 6.6R_\odot$. With $T = 2$ MK, the wind reaches a velocity of ≈ 170 km sec^{-1} at $r \approx 3.3R_\odot$.

The larger the coronal temperature, the faster is the acceleration of the wind. To quantify this, we note that if an increase in velocity by $\Delta V = 120$ (or 170) km sec^{-1} were to occur with constant acceleration over a spatial interval of $\Delta x = 4.6$ (or 2.3) $\times 10^{11}$ cm, the corresponding acceleration ($\approx (\Delta V)^2/(2\Delta x)$) would be roughly 160 (or 630) cm sec^{-2}. Thus, the inner solar wind experiences an outward acceleration which, as regards the magnitude, coincidentally is not far from the (downward) acceleration (981 cm sec^{-2}) which is experienced by objects near the Earth's surface.

18.4 Observational Evidence for Solar Wind Acceleration

The outflow of the solar wind can be measured *in situ* by spacecraft near Earth's orbit, and farther out in the solar system. Those measurements indicate that the solar wind accelerates only slowly at radial distances of $200R_\odot$

and more. But no spacecraft has up to now flown through the region where the solar wind undergoes its greatest acceleration, i.e., close to the sonic point at distances of several $R_\odot$.

Is there any way to probe the regions of maximum wind acceleration? Yes, it can be done by remote sensing of distance radio sources as the radio waves from those source propagate through the inner solar wind. As the Sun moves through the sky during the year, it passes close to certain natural radio sources at definite times. For example, the Crab Nebula (a bright supernova remnant) passes behind the Sun in mid-June each year. Also, certain spacecraft are in orbit around the Sun, and from time to time, they pass behind the Sun (as viewed from Earth).

When a distant source is observed with a radio telescope, the line of sight to that source approaches closest to the Sun at a certain point. The radial distance from the Sun's center to that point of closest approach is called the *impact parameter p.* Because the solar wind density falls off with increasing radial distance, a radio observation of that source is most heavily influenced by the properties of the solar wind at the particular radial location $r = p$.

What do the radio data reveal? The most prominent feature is that many radio sources (especially the smallest ones) exhibit rapid fluctuations in intensity as the source comes closer and closer to the Sun. This is the phenomenon of *scintillation (twinkling)*, analogous to that which causes stars to twinkle in the night sky: it is caused by clumps of more or less dense material moving across the line of sight to the distant (small) source. Each clump acts as a miniature lens which can change the intensity of the light. The solar wind is not a smooth flow: instead, it is a turbulent medium, which does, on average, expand outward. But there are, in the wind, clumps of matter akin to the eddies which form in fast-flowing water, or which form in the jet streams in the atmosphere and ionosphere of the Earth.

As the shadow of an individual clump moves across a radio telescope, the intensity which is recorded changes in a way which is determined by the shape of the clump. If two telescopes, separated on the Earth's surface by a distance d, record the same clump (identified by its shape), but separated in time by an interval t, then the speed of the shadow across the Earth's surface can be determined: d/t. This is equal to the speed of the clumps across the line of sight, and therefore gives an estimate of the speed of the solar wind at the impact parameter, i.e., at the radial location $r = p$.

An example of the variation of clump speed as a function of the impact parameter is shown in Figure 18.1. There is obviously a lot of scatter. Why? Because the data in Figure 18.1 were obtained over a time interval of several days, and from day to day, as the Sun rotates, the line of sight may shift from a helmet streamer into a coronal hole into quiet Sun. Therefore, we cannot expect that the line of sight to a given source will continue to pass through exactly the same coronal material throughout an observing run which lasts several days. It would be surprising if there was *not* considerable scatter in the data. But overall, one can see that the outflow speed is small close to the Sun,

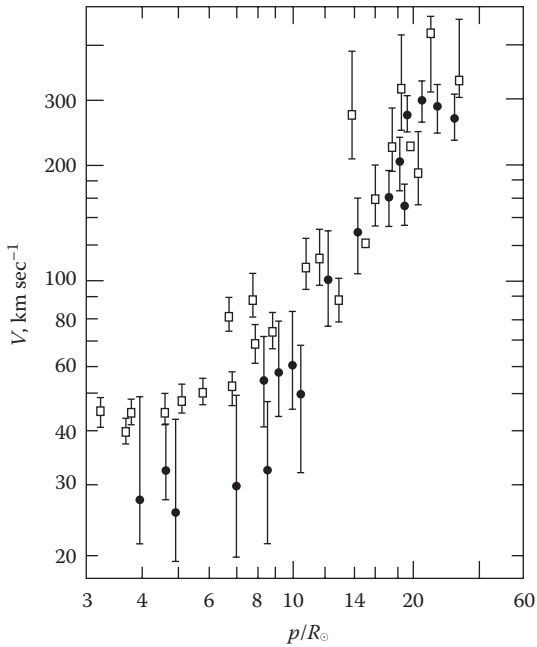


FIGURE 18.1: Estimates of the outflow speed of solar wind clumps as a function of radial distance from the Sun. The abscissa contains the impact parameter p, which is the radial distance from the Sun to the point where the line of sight to a distant source passes closest to the Sun. (From Yakovlev, O. I. and Mullan, D. J., *Irish Astron. J.*, 23, 7, 1996. With permission.)

and the outflow speed increases with increasing radial distance. The outflow speed first reaches roughly 120 km sec^{-1} (i.e., the sound speed) at radial distances of about $10R_\odot$, slightly larger than the $6.6R_\odot$ which we estimated for a 1 MK corona. And by the time p has increased to values of $20R_\odot$ or so, the speed is up to 300–400 km sec^{-1}. Thus, at a radial distance of about 0.1 AU from the Sun, the wind is well on its way to reaching a typical speed of 470 km sec^{-1} at Earth's orbit (i.e., at $215R_\odot$) (see Section 18.6 below).

Thus, radio studies of the solar wind using scintillation provide useful information about the otherwise inaccessible region where the solar wind undergoes its greatest acceleration.

18.5 Energy Equation

So far, we have obtained information about the solar wind by explicitly referring to the conservation of momentum and conservation of mass.

Our assumption $T = \text{const}$ involves a particular solution of the energy equation. But we have not specified, in physical terms, how a constant temperature of 1–2 MK might be maintained out to distances of $(3.3\text{–}6.6)R_\odot$ and beyond.

In order to determine how, in physical terms, the temperature actually varies as a function of radial distance, we need to solve the equation of energy conservation. In order to do that, we would have to include processes which deposit energy in the gas, or remove energy from the gas, or distribute energy through the gas.

As an example of the latter, the effects of thermal conduction are important. In spherical symmetry, the equation of heat conduction in steady state is described by

$$\frac{d}{dr}\left(r^2 k_{th}\frac{dT}{dr}\right) = 0 \tag{18.9}$$

We have already seen (Chapter 17, Section 17.14.1) that in coronal plasma, the thermal conductivity $k_{th} = k_o T^{2.5}$. Inserting this in Equation 18.9, we find, after a first integration, that

$$T^{2.5}\frac{dT}{dr} = \frac{\text{const}}{r^2} \tag{18.10}$$

The solution of this equation is $T \sim r^{-2/7}$. This is a rather slow function of radial distance. For example, if $T = 2$ MK at $r = R_\odot$, then at the sonic point distance $(3.3R_\odot)$, a thermally conducting wind would have $T \approx 1.4$ MK. Thus, contrary to our assumption above of constant T, the temperature would not in fact have remained constant all the way out to the sonic point. Nevertheless, it is also true that T would not have cooled off "drastically." From this perspective, a model which assumes constant temperature (at least out to a few solar radii) is not totally unrealistic.

There are other possibilities for keeping the corona hot. These include the deposition of energy from wave modes of various kinds (Alfven waves, shocks). Another approach is to *assume* that the solar wind material has the property that the pressure at any radial location is related to the density at that location by a simple relation, $p(r) \sim \rho(r)^{(n+1)/n}$. This is nothing less than the "polytrope law" (see Chapter 10, Equation 10.1) which we found to be helpful in describing certain properties of the solar interior.

It turns out that a rich variety of solar wind solutions can be derived by considering various values of the polytropic index n. The isothermal case corresponds to $n = \infty$. The adiabatic case corresponds to $n = 3/2$. In the latter case, $p \sim \rho^{5/3}$, which for a perfect gas ($p \sim \rho T$) corresponds to $T \sim \rho^{2/3}$. Once the solar wind approaches a nearly constant value (see next section), Equation 18.6 implies that $\rho \sim r^{-2}$. In an adiabatic wind, this leads to rapidly declining T as r increases: $T \sim r^{-4/3}$.

However, discussions of the application of the full energy equation to the solar wind lie outside the bounds of this "first course."

18.6 Asymptotic Speed of the Solar Wind

We have seen that according to Equation 18.7, the wind is moving at 120 (or 170) km sec^{-1} at radial locations of 6.6 (or 3.3) $R_\odot$ for a coronal temperature of 1 (or 2) MK. Let us now consider how the speed behaves as we examine the flow at very large distances from the Sun.

In $\mathcal{N}(r)$ (see Equation 18.7), the term in $1/r$ dominates over the term in $1/r^2$ as $r \to \infty$. That is, $\mathcal{N}(r) \to 2a^2/r$. Moreover, at large r, the wind is supersonic, i.e., V exceeds a. As a result, the dominant term in $\mathcal{D}(r)$ as $r \to \infty$ is V. Therefore, at large r, $dV/dr = \mathcal{N}(r)/\mathcal{D}(r) \to 2a^2/rV$. In an isothermal wind, this can be integrated to give the solution

$$\frac{1}{2}V^2 = 2a^2 \log_e(r) + \text{const} \tag{18.11}$$

Recalling that at $r = r_s, V = a$, we can evaluate the constant. This leads to

$$\left(\frac{V}{a}\right)^2 = 1 + 4\log_e\left(\frac{r}{r_s}\right) \tag{18.12}$$

This solution shows that, at large distances, the solar wind speed asymptotically approaches the functional form $V(r) \to 2a\sqrt{(\log(r/r_s))}$. This is a very slow function of radial distance. For example, if $T = 1$ MK, by the time the solar wind reaches Earth orbit, i.e., $r = 215.04\ R_\odot$(see Chapter 1, Section 1.5), the ratio r/r_s has the value ≈33. According to Equation 18.12, this gives $V(1\text{ AU}) \approx 3.9a \approx 470$ km sec^{-1}. At a distance which is 100 times greater than Earth's orbit (i.e., far beyond the orbit of Pluto), V has increased to ≈690 km sec^{-1}.

Turning briefly to energy considerations, we note that if thermal conduction (mainly by electrons) dominates the energy equation in the solar wind, then compared to the coronal temperature at $r = R_\odot$, T at 1 AU would be reduced by factors of $215^{2/7} = 4.6$. Thus, with $T = 1$–2 MK at $r = R_\odot$, T at 1 AU should be 2–4 $\times 10^5$ K. On the other hand, if the solar wind were described by an adiabatic polytrope ($n = 3/2$), with $T(r) \sim r^{-4/3}$ (see Section 18.5), the ion temperatures at $r = 1$ AU would be very low, of order 10^3 K.

Thus, the model of an isothermal solar wind predicts that wind speeds of several hundred km sec^{-1} and temperatures of up to a few times 10^5 K should occur near Earth's orbit. These predictions have been tested by many spacecraft since the 1960s, and have been found to be not too bad *on average*. (For an informative overview of the early experiments to measure solar wind properties, and the scientists involved, see Hufbauer, 1991.) However, the data also indicate that the solar wind has highly variable properties. Examples of solar wind velocity, density, and proton temperature close to the Earth's orbit are shown for a 27-day interval in Figure 18.2. During the interval of the measurements, the Sun was at a low level of activity. The speed varies from

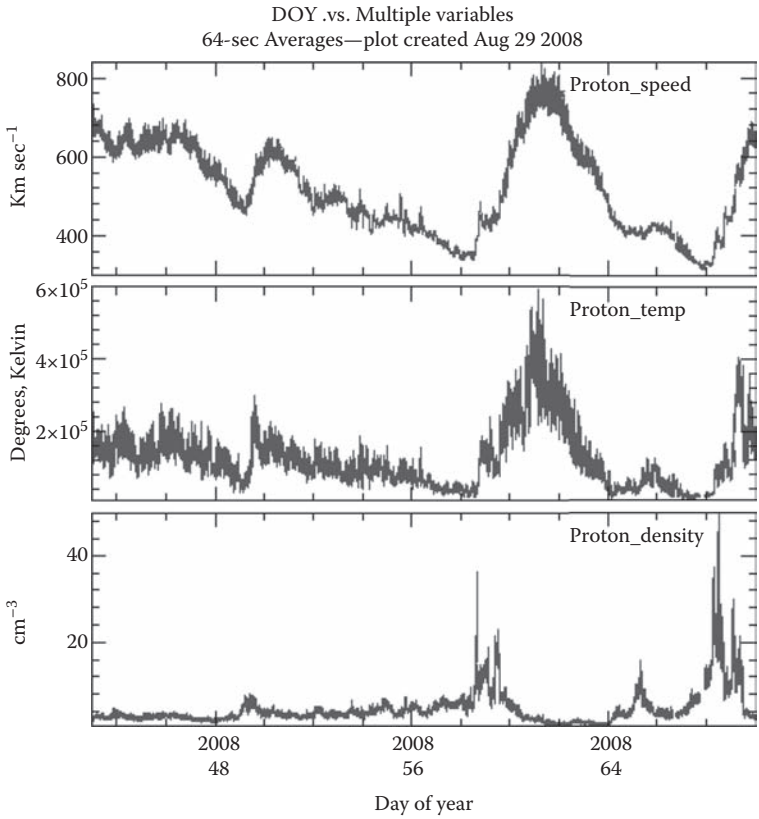


FIGURE 18.2: Solar wind properties (density, speed, temperature) for a 27-day interval early in the year 2008. DOY=day of year. These (and other) data are publicly available for any time interval between 1998 and 2008 from the ACE spacecraft data archive at http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_SWEPAM.html.

roughly 300 to roughly 800 km sec^{-1}. And the proton temperature ranges from a few times 10^4 to a few times 10^5 K.

Despite the large amount of variability, two major categories of solar wind have been identified (Zirker 1981): "slow wind," with mean speeds (near Earth) of 330 km sec^{-1}, and "high speed wind" with mean speeds of 700 km sec^{-1}. The high speed wind has been found to emerge preferentially from coronal holes. Since those holes are always present to some extent at North and South solar poles, the wind over the solar poles is always "high speed." This was confirmed by the Ulysses spacecraft which traversed the solar wind emerging from both solar poles. But there are also times when high speed stream flows in the plane of the Earth's orbit: when a high speed stream flows past the Earth, the geomagnetic field may become disturbed.

At 1 AU, in the plane of the Earth's orbit, as well as identifying the two major components at low and high speeds, it is possible to define an "average" set of plasma characteristics for solar wind flows (e.g., Zirker, 1981). The mean values of velocity, density, and ion temperature in the "average" solar wind at 1 AU are found to be 470 ± 120 km sec^{-1}, 8.7 ± 6.6 cm^{-3}, and $1.2 \pm 0.9 \times 10^5$ K.

Thus, the model of an isothermal corona does a good job at predicting at least the *average speed* of the wind near the Earth. And the proton temperatures overlap well with the predicted values due to thermal conduction.

18.7 Rate of Mass Outflow from the Sun

In order to determine how much mass the Sun loses per unit time, we need to know not only the *speed* but also the *density* of the solar wind. We already know that the speed (in an isothermal corona) is equal to the sound speed at the radial location $r = r_s$. Can we also estimate the density at $r = r_s$? This is more difficult, but we can do it, roughly. To do this, we note that between the surface of the Sun and $r = r_s$, the flow speed has not yet reached values as large as the sound speed. In fact, close to the surface, the speed is much smaller than the sound speed. As a result, the *hydrodynamic* terms in the equation of motion are small compared to the *hydrostatic* terms. In other words, close to the surface, the corona, although in principle expanding, is flowing so slowly that *the material is not far from HSE.* To be sure, the closer we get to $r = r_s$, the farther the conditions depart from HSE. And by the time we arrive at $r = r_s$ and beyond, HSE has broken down altogether.

But as a rough approximation, we can use Equation 18.2 (rewritten in terms of particle number density n, assuming constant T) to evaluate the density at the sonic point:

$$n(r_s) = n_o \exp\left[A\left(\frac{1}{r_s} - \frac{1}{r_o}\right)\right] \tag{18.13}$$

Inserting the empirical values $n_o = 10^{8-9}$ cm^{-3} (see Chapter 17, Section 17.1), we find that in a corona with $T = 1$ MK, where $A = 13$ and $r_s \approx 6.6$ (both in units of $r_0 = R_\odot$), the density at the sonic point $n(r_s) \approx 2 \times 10^{3-4}$ cm^{-3}. Repeating the calculation for a corona with $T = 2$ MK, we find a much larger density at the (closer) sonic point $n(r_s) \approx 1 \times 10^{6-7}$ cm^{-3}. Outside the sonic point, in the limit where the velocity is varying only slowly with distance, we expect to have $n(r) \sim r^{-2}$. Thus, between $r = r_s$ and the Earth's orbit ($r = 215.04\ R_\odot$), the density should decrease by a factor of $(215/6.6)^2 \approx 1100$ (for $T = 1$ MK) and by a factor of $(215/3.3)^2 \approx 4200$ (for $T = 2$ MK). Thus, near the Earth's orbit, we expect to find $n(1\ AU) \approx$ 2–20 cm^{-3} (if $T = 1$ MK), and $n(1AU) \approx$ 240–2400 cm^{-3} (if $T = 2$ MK). The empirical densities in

Figure 18.2 are consistent with the estimates for an isothermal corona with $T = 1$ MK, but not for $T = 2$ MK.

In the context of the approximate discussion given here (i.e., HSE remains roughly valid out to a radial distance of r_s, and constant speed outside r_s), we conclude that the Sun may well be able to maintain a coronal temperature of 1 MK out to $r \approx 6.6 R_\odot$, but the Sun is probably *not* able to maintain a temperature as large as 2 MK out to $r \approx 3.3 R_\odot$. Ultimately, the inability of the Sun to maintain coronal gas at 2 MK out to several $R_\odot$ is an indication that the Sun supplies only a finite flux of mechanical energy to the corona. We have already seen in Chapter 17, Section 17.14.3 that the coronal temperature is controlled by the chromospheric pressure, which is in turn controlled by the amount of mechanical energy flux emerging from the Sun.

How large would the mechanical flux have to be in order to maintain a corona at $T = 1$ MK and at $T = 2$ MK? We can estimate a *lower limit* on the necessary energy flux by considering one component only, namely, the kinetic energy. The flux of K.E. in the wind equals the K.E. density $0.5\rho V^2$ times the flow speed V. At $r = r_s$, this K.E. flux $F_K(r_s)$ equals $0.5\rho(r_s)a^3$. Transforming back to the base of the corona, this would correspond to an energy flux crossing the surface $r = 1$ solar radius of $F_K(1) = r_s^2 F_K(r_s)$. Inserting numerical values for the case $T = 1$, we find $F_K(1) = 130$–1300 ergs cm^{-2} sec^{-1}. For the case $T = 2$ MK, the result is much larger, $F_K(1) = 0.5 \times 10^{5-6}$ ergs cm^{-2} sec^{-1}. Now, the solar wind does not escape easily from active regions, where magnetic loops are closed (Chapter 17, Section 17.11). It seems preferable to look to the quiet Sun as the source of mechanical energy to power the "typical" solar wind. According to Equation 17.9, the flux of mechanical energy entering the base of the quiet corona is limited, $F(\mathrm{QS}) \approx 5 \times 10^5$ ergs cm^{-2} sec^{-1}. Comparing with $F_K(1)$ for the case $T = 1$ MK, we see that the quiet Sun would have no problem in supplying the demands of the coronal K.E. flux, with plenty of energy to spare for the radiated flux, plus the internal energy flux ($\sim nTa$), plus the conductive flux ($\sim T^{3.5}$). On the other hand, for the case $T = 2$ MK, even $F_K(1)$ alone may already "soak up" the entire available supply $F(\mathrm{QS})$: there would be nothing available for the radiated flux, or for the increased demands on internal energy flux (increased by $> 10^3$ compared to the $T = 1$ MK case), or for the increased demands on conductive flux (increased by ten). For the case $T = 2$ MK, the numerical values indicate that the Sun simply does not generate enough mechanical energy to "go around."

Let us use the empirical data at radial distances $D = 1$ AU to estimate the solar mass loss rate in the "average" solar wind (Zirker, 1981), where $V(1\ AU) = 470$ km sec^{-1} and $n(1\ AU) \approx 9$ protons cm^{-3}. Allowing for the presence of a few percent helium nuclei as well, the mean gas density at 1 AU is $\rho(1\ AU) \approx 2 \times 10^{-23}$ gm cm^{-3}.

Expressing 1 AU in cm (Chapter 1, Section 1.2), we find that the rate of mass outflow of a spherically symmetric wind, $4\pi D^2 \rho(1\ AU) V(1AU)$, is some 3×10^{12} gm sec^{-1}, i.e., a few metric tons per second. To be sure, the wind is not altogether spherically symmetric: the polar wind is certainly faster on average than the equatorial wind, while the mean density in the polar wind

is smaller. So we do not expect the assumption of spherical symmetry to be reliable to better than a factor of (maybe) two.

As well as mass loss in the solar wind, the Sun is also losing mass as a result of nuclear reactions in the core. The solar power output L of 3.8416×10^{33} ergs sec^{-1} (Equation 1.11) corresponds to a nuclear mass loss rate $(dM/dt)_{nuc} = L/c^2 \approx 4 \times 10^{12}$ gm sec^{-1}. This is close to the mass loss rate in the solar wind. Is this a coincidence? It is hard to say: the mass loss rate in the wind is determined by coronal heating processes which we cannot yet fully identify. Perhaps when all mechanisms are better understood, there may be a deep-rooted reason why the Sun has the property of comparable mass loss rates from core and corona.

Given the lifetime of the Sun ($\approx 4.5 \times 10^9$ years, from measurements of meteorite ages), and assuming that the mass loss rates have remained constant over time, we find that in the course of its lifetime, the Sun's mass has decreased by only a few parts in 10,000.

18.8 Coronal Mass Ejections

No matter when spacecraft observations of the solar wind are made, they essentially always report the presence of an outflow of matter at speeds of a few hundred km sec^{-1}, and with densities within an order of magnitude of 10 cm^{-3}. In the sense that "there is always something there," the solar wind can be considered to be more or less a steady-state phenomenon.

However, from time to time, a major disruption is observed to propagate out through the wind. These events are called "coronal mass ejections," or "CME's" for short. They originate in the solar corona. An example is illustrated in Figure 18.3, where a "bulb" of material expands and breaks open, dumping its contents into the solar wind. The contents propagate outward through the solar wind, maintaining identity for a finite distance. In some cases, CME's survive as far as Earth's orbit, where they can (in certain cases) disturb the Earth's geomagnetic shield. The disturbances in the Earth's field may give rise to large induced voltages in terrestrial power lines. The speed of the CME may be larger than 2000 km sec^{-1}, or may be smaller than 100 km sec^{-1}, depending on the physical properties of the CME.

CME's occur more frequently when the Sun is at maximum activity than at minimum activity. Near solar maximum, several CME's may emerge per day. But near solar minimum, the rate is on average less than one per day. These data indicate that CME's owe their existence to the magnetic fields which are present in the corona.

Is there a typical mass associated with CME's? The data show that small CME's occur more frequently than large CME's. According to data which were obtained for almost 1000 CME's during the years 1979–1981, close to solar maximum (Jackson and Howard, 1993), the number of CME's with mass M

in the sample is given by the expression:

$$N(M) = 370 \exp(-9.43 \times 10^{-17} M) \tag{18.14}$$

The fact that the masses of CME's follow an exponential law indicates that there is in effect an upper cut-off in the mass distribution: for CME's with masses in excess of 10^{16} gm, the number of CME's falls off exponentially rapidly compared to those with masses less than 10^{16} gm. It appears as if the Sun (at least in the years 1979–1981), was capable of producing CME's with masses up to (essentially) 10^{16} gm, but not much more than that. After many years of observations by means of many spacecraft, the most massive CME observed so far has $M_{\max} \approx 7 \times 10^{16}$ gm.

What energies are associated with CME's? The kinetic energy is easy to evaluate. With speeds which are typically in the range 300–1000 km sec^{-1}, the KE of a CME with mass M is typically $0.5M \times 10^{15-16}$ ergs. The existence of an upper cut-off in M sets an upper limit on the KE of CME's: $0.5 \times 10^{31-32}$ ergs. Inserting $M_{\max}$, we find an upper limit on CME energy of a few times 10^{32} ergs.

We have already seen (Chapter 17, Section 17.18.5) that solar flares also have energies which extend up to no more than a few times 10^{32} ergs. To be sure, the flares differ from CME's in the sense that flare energies follow a power-law distribution, whereas CME's have an exponential distribution. Thus, it is more difficult to be sure about an upper cut-off of flare energies. Nevertheless, it is striking that the two most prominent classes of transient energy release from the Sun can apparently produce events which, in their largest manifestations, have comparable maximum energies.

The physical feature which provides a common denominator to flares and CME's is the magnetic field. In the presence of solar gravity, and given the amount of plasma in coronal gas, it appears that the magnetic fields which the Sun produces are limited in the maximum amount of energy they can store. If, in a given active region, that limit is exceeded, the field apparently "must" respond by releasing the stored energy. The form that the released energy takes can be either a flare or a CME or a combination of the two, depending on local conditions.

Do CME's contribute significantly to solar mass loss? Apparently not: it has been estimated (Jackson and Howard, 1993) that even at solar maximum, the Sun loses mass in the form of CME's at a rate which is only about 16% of the total mass loss rate. At solar minimum, the percentage would be smaller.

18.9 How Far does the Sun's Influence Extend in Space?

The solar wind originates in the Sun's corona, and it flows out past the Earth's orbit. Disturbances in the wind (e.g., CME's) can have an effect on our lives on Earth (e.g., radio blackouts, voltage surges). In that sense, the

solar wind allows the Sun to "reach out" far beyond the confines of one solar radius and cause certain events on Earth which have nothing to do with the Sun as a source of heat and light.

The solar wind also influences the surroundings of other planets, especially those with magnetic fields. All four giant planets (Jupiter, Saturn, Uranus, and Neptune) were discovered to have strong magnetic fields when the Voyager 2 spacecraft visited all four in the decade from 1979 to 1989. The planetary magnetic fields trap charged particles into confined orbits, giving rise to an environment of interesting interactions between fields and plasma. Neptune is on average 30 times farther from the Sun than the Earth is, but even so, the solar wind has an influence on the shape of the magnetosphere out there.

How far out does the Sun's influence extend? To answer that, we recall (Section 18.1) that in the ISM, the total pressure p(ISM) (including magnetic fields and cosmic rays) is $\approx$1.4–1.7 $\times 10^{-12}$ dyn cm^{-2}. Now the solar wind outflow, with its energy density $E_w(r) = 0.5\rho(r)V(r)^2$, exerts an outward ram pressure equal to $E_w(r)$. This is the pressure which allows the Sun to "push back" the ISM. But this "pushing back" can work only a long as $E_w(r)$ exceeds p(ISM). Therefore we can determine the maximum radial extent of the Sun's influence by seeking the radial distance r_m at which $E_w(r_m) \approx p$(ISM).

To solve this, we recall that in an isothermal wind, the speed V in the outer wind varies only very slowly as the distance r from the Sun increases (see discussion following Equation 18.12). Since V is essentially constant, the density ρ must fall off as $1/r^2$. As a result, $E_w(r)$ varies essentially as $\sim 1/r^2$.

At $r = 1$ AU, using a mean density of 9 protons cm^{-3} and a mean wind speed of 470 km sec^{-1}, we find $E_w(r = 1AU) \approx 2\times10^{-8}$ dyn cm^{-2}. At a distance of r AU, $E_w(r) \approx 2 \times 10^{-8} r^{-2}$ dyn cm^{-2}. This pressure becomes equal to p(ISM) at $r_m \approx$110–120 AU. Therefore, at distances of order 100 AU, the pressure of the ISM should bring the solar wind to a halt. Depending on the local conditions, the halt may be so abrupt that a shock wave is set up: this is referred to as the "termination shock" of the solar wind.

Thus, the influence of the Sun extends, via the solar wind, to a radial distance which lies well beyond Pluto's orbit, to a distance of order 100 AU. Beyond that distance, the pressure of the ISM prevents the solar wind from escaping. The Sun's "sphere of influence," also known as the "heliosphere," comes to an end at a distance of about 100 AU.

So far, two spacecraft have traveled far enough to make the transition through the termination shock. The Voyager 1 transition occurred in December 2004 at a radial distance $r = 94$ AU. The Voyager 2 transition occurred (in a different direction) in August 2007, at $r = 84$ AU. The different radial distances of the two transitions indicate that the termination shock is not spherical. However, both transitions occurred at distances which are not far from the estimate given above ($\approx$100 AU).

The heliosphere presents a barrier to the galactic cosmic rays (GCR) which are present in the ISM. A fraction of those GCR can reach the Earth's orbit, but only at the expense of "swimming upstream" against the solar wind. In

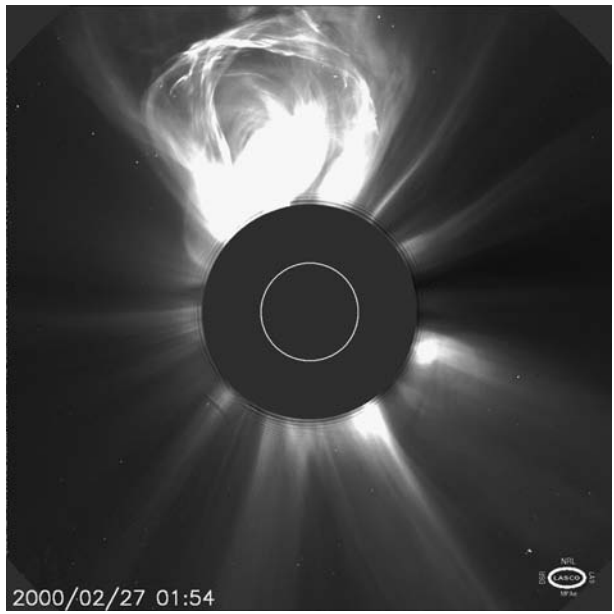


FIGURE 18.3: **(See color insert following page 202.)** A coronal mass ejection ("CME") lifts off from the Sun. Image obtained by the LASCO instrument on board the SOHO spacecraft. LASCO obscures the brilliant photosphere of the Sun behind the dark mask in the center of the image: the white circle inside the mask represents the location of the Sun's photosphere. (Courtesy of the SOHO/LASCO consortium. SOHO is an international collaboration between ESA and NASA.)

the course of penetrating in as far as the Earth's orbit, the GCR scatter off magnetic fluctuations in the wind, thereby losing some energy. As a result, the Sun protects us on Earth from some of the harmful effects of cosmic rays.

During the 11-year activity cycle, as the sunspot numbers go up and down, the magnetic fluctuations in the wind also change their properties. This alters the flux of GCR arriving at Earth by as much as tens of percent (see Figure 18.4). When sunspots are more abundant, and the magnetic fields generated by the Sun are stronger, GCR have a harder time making their way in to the Earth's orbit. When sunspots are less abundant, the fluxes of GCR at Earth are larger.

Given that the "typical" solar wind travels at a speed of 470 km $\sec^{-1}$ at Earth orbit, and not much more than that at greater distances, we can estimate roughly the time t_p required for solar wind to travel from the Sun to the edge of the heliosphere. With a heliospheric radius $r_m \approx 100 AU \approx 1.5 \times 10^{15}$ cm, and an assumed constant speed $V = 4.7 \times 10^7$ cm $\sec^{-1}$, we find t_p is very close to one year. As a result, when magnetic conditions change at the Sun, it takes a full year before the heliosphere as a whole "knows" that

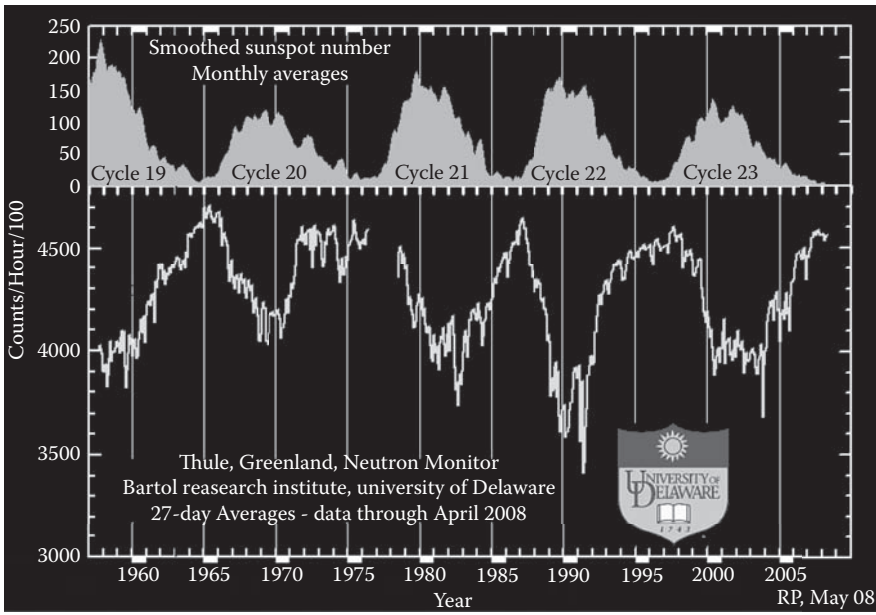


FIGURE 18.4: **(See color insert following page 202.)** Count rates of GCRs over a 50-year interval. Neutron monitors are instruments which respond to GCR with energies of a few GeV. (Plot provided by Roger Pyle via website http://neutronm.bartol.udel.edu)

the magnetic properties of the wind have changed. This leads to a phase shift in the GCR count rates relative to the sunspot counts: the minimum GCR counts occur on average a year or two later than the peak in the sunspot counts.

The size of the heliosphere r_m depends on two independent quantities: the solar wind flux (determined by the Sun), and the ISM pressure (determined by conditions in the ISM). In the course of the Sun's lifetime, the local ISM may have a variety of properties. As a result, the edge of the heliosphere may approach the Earth more closely than the current distance of about 100 AU. In extreme cases, it is even possible that Earth's orbit might lie at or beyond the edge of the heliosphere. See Exercises 1–3.

Exercises

18.1 The Sun is in orbit around the center of the Galaxy, taking some 250 My to orbit once. In the course of an orbit, the Sun encounters various

types of ISM. In some of these, the local density may rise to values of $n = 10^3$ cm^{-3}, 10^5 cm^{-3}, or even 10^7 cm^{-3}. Assuming that the gas temperature remains at the value $T = 100$ K in all clouds, and that all other components of ISM pressure are unchanged, calculate the extent of the heliosphere in each of the above clouds.

18.2 Also in the various components of ISM, the field strength increases roughly in proportion to $n^{2/3}$. For the values of n given in Exercise 1, evaluate the local field strengths. With these new B values, recalculate the extent of the heliosphere.

18.3 Assuming that only n or B varies, what values of n and B are necessary in order that the extent of the heliosphere shrinks to become smaller than the size of the Earth's orbit?

References

Gayley, K. G., Zank, G. P., Pauls, H. L., Frisch, P. C., and Welty, D. E. 1997. "One- versus two-shock heliosphere: constraining models with Goddard High Resolution Spectrograph Lyα spectra towards α Centauri," *Astrophys. J.*, 487, 259.

Hufbauer, K. 1991. "The solar wind, 1957–1970," *Exploring the Sun: Solar Science Since Galileo.* Johns Hopkins University Press, Baltimore, MD, pp. 213–258.

Ip, W. H. and Axford, W. I. 1985. "Estimates of galactic cosmic ray spectra at low energies," *Astron. Astrophys.*, 149, 7.

Jackson, B. V. and Howard, R. A. 1993. "A CME mass distribution derived from SOLWIND coronagraph observations," *Solar Phys.*, 148, 359.

Yakovlev, O. I. and Mullan, D. J. 1996. "Remote sensing of the solar wind using radio waves: Part 2," *Irish Astron. J.*, 23, 7.

Zirker, J. B. 1981. "The solar corona and the solar wind," *The Sun as a Star.* ed. S. Jordan. NASA SP-450, 135.

Appendix

Symbols used in the text, their units, and where they are introduced

a	isothermal speed of sound (cm $\sec^{-1}$) (Chapter 18.1)
a_{hc}	relative acceleration of hot and cold gas (Chapter 7.1)
a_R	radiation constant (ergs cm^{-3} K^{-4}) (Chapter 2.1)
a_λ	limb darkening coefficient (Chapter 2.2)
A	area of solar surface (Chapter 17.16)
Å	Angstrom (unit of length = 10^{-8} cm) (Chapter 2.1)
b_λ	limb darkening coefficient (Chapter 2.2)
$\boldsymbol{B}$	magnetic field vector (Chapter 16.6.1)
B	magnitude of $\boldsymbol{B}$ (Chapter 16.6.1)
c	speed of light ($= 3 \times 10^{10}$ cm $\sec^{-1}$) (Chapter 2.1)
c_s	adiabatic speed of sound (cm $\sec^{-1}$) (Chapter 7.1)
C_p	specific heat at constant pressure (ergs gm^{-1} K^{-1}) (Chapter 6.7.1)
C_v	specific heat at constant volume (ergs gm^{-1} K^{-1}) (Chapter 6.7.1)
d_c	mass column density (gm cm^{-2}) (Chapter 3.3, 5.1)
D	distance of Earth from Sun ($= 1.5 \times 10^{13}$ cm) (Chapter 1.2)
D	horizontal diameter of convection cells (cm) (Chapter 6.5)
e	unit of electric charge ($= 4.8 \times 10^{-10}$ e.s.u.) (Chapter 16.6.1)
$\boldsymbol{E}$	electric field vector (Chapter 16.6.1)
F	flux of radiation at arbitrary depth (ergs cm^{-2} $\sec^{-1}$) (Equation 2.22)
F_s	flux of sound waves emitted by convection (Equation 14.27)
$F_\odot$	flux of radiation at the solar surface ($= 6.3155 \times 10^{10}$ ergs cm^{-2} $\sec^{-1}$) (Chapter 1.8)
g	acceleration due to gravity (cm $\sec^{-2}$) (Chapter 5.1)
g_{ad}	adiabatic temperature gradient (K cm^{-1}) (Chapter 6.8)
g_i	statistical weight of atomic level i (Chapter 3.3.2)
g_s	acceleration due to gravity at the Sun's surface ($= 27422$ cm $\sec^{-2}$) (Chapter 1.6)
g_T	temperature gradient (K cm^{-1}) (Chapter 6.8)
G	Newton's gravitational constant (cm^3 gm^{-1} $\sec^{-2}$) (Chapter 1.1)
h	height in the atmosphere, increasing *outward* from the Sun (cm) (Chapter 5.1)
H	vertical depth of a granule (Chapter 6.5)
H_p	pressure scale height (cm) (Chapter 5.1)
i	$\sqrt{-1}$ (Chapter 13.5.2)

I ionization potential (electron volts) (Chapter 4.1)
I_λ intensity of radiation per unit wavelength (Chapter 2.1)
I_ν intensity of radiation per unit frequency (Chapter 2.1)
J mean intensity at arbitrary depth (Equation 2.23)
k Boltzmann's constant ($= 1.38 \times 10^{-16}$ ergs K^{-1}) (Chapter 1.7)
k_h horizontal wavenumber of oscillation mode in the Sun (Chapter 14.7)
k_r radial wavenumber of oscillation mode in the Sun (Chapter 14.7)
k_λ absorption coefficient at wavelength λ (cm^{-1}) (Chapter 2.3)
k_{th} thermal conductivity (ergs cm^{-1} sec^{-1} K^{-1}) (Chapter 8.1)
K degrees Kelvin: unit of temperature (Chapter 2.1)
$K(\tau)$ related to radiation pressure at optical depth τ (Equation 2.26)
l angular degree of oscillation mode in the Sun (Chapter 13.2)
L half-length of coronal loop (cm) (Chapter 17.14.1)
$L_\odot$ luminosity (=power output) of the Sun ($= 3.84 \times 10^{33}$ ergs sec^{-1}) (Chapter 1.4)
m_H mass of the hydrogen atom ($= 1.66 \times 10^{-24}$ gm) (Chapter 1.7)
M Mach number of granule flows (Chapter 14.8)
$M_\odot$ mass of the Sun (1.99×10^{33} gm) (Chapter 1.3)
n polytropic index (Equation 10.1)
n_a number density of photon absorbers (cm^{-3}) (Chapter 3.3)
n_e number density of electrons (cm^{-3}) (Chapter 4.1)
N column number density (cm^{-2}) (Chapter 3.3)
N number related to adiabatic exponent (Chapter 14.1)
N_e number density of electrons (cm^{-3}) (Chapter 17.7)
N_i number density of ions (cm^{-3}) (Chapter 17.7)
p gas pressure (dyn cm^{-2}) (Chapter 5.1)
p_e electron pressure (dyn cm^{-2}) (Chapter 4.1)
p_{mag} magnetic pressure (dyn cm^{-2}) (Chapter 16.6.2.1)
P_{ac} acoustic cut-off period (sec) (Equation 13.14)
P_g critical period of gravity modes in the Sun (Chapter 1.10)
P_p power density in pressure pulse due to granules (Equation 14.26)
q slowly varying function of optical depth (Chapter 2.8)
r radial coordinate (cm) (Chapter 2.1)
r_g gyroradius (cm) (Equation 16.4)
r_o Emden unit of length (cm) (Equation 10.9)
R_g the gas constant ($= 8.3 \times 10^7$ ergs mole^{-1} K^{-1}) (Chapter 1.7)
$R_\odot$ radius of the Sun ($= 6.96 \times 10^{10}$ cm) (Chapter 1.5)
S_λ the source function at wavelength λ (Chapter 2.4)
t time (sec) (Chapter 7.1)
T temperature (degrees K) (Chapter 2.1)
T_b temperature at the base of the convection zone (Chapter 8.3)
T_{eff} effective temperature (K) (Chapter 1.8)
u_ν energy density of radiation per unit frequency (ergs cm^{-3} Hz^{-1}) (Chapter 2.9)
U internal energy (ergs gm^{-1}) (Chapter 6.7)

V	velocity of vertical gas motion (cm $\sec^{-1}$) (Chapter 7.1)
V_{esc}	escape speed from the surface (Chapter 1.7)
V_{th}	mean thermal speed of particles (Chapter 17.6)
w	auxiliary pressure variable in oscillation calculations (Equation 14.16)
W_{mag}	magnetic energy density (ergs cm^{-3}) (Chapter 16.6.2.1)
x	dimensionless unit of length in the Lane–Emden equation (Chapter 10.4)
X	dimensionless unit of length in oscillation calculations (Chapter 14.2)
y	Lane–Emden function, which describes polytrope structure (Equation 10.6)
z	depth in atmosphere, increasing *inwards* towards the solar center (Chapter 7.1)
z	auxiliary radial displacement variable in oscillation calculations (Equation 14.16)
z_b	depth of the base of the convection zone (Chapter 7.9)
z_r	depth below the surface where a p-mode is refracted upwards (Equation 14.24)
α	mixing length parameter = ratio of H to H_p (Chapter 6.5)
α	dimensionless frequency in oscillation calculation (Equation 14.10)
β	power law index in flare distribution (Equation 17.13)
γ	ratio of specific heats (Chapter 6.7.1)
Γ	generalized exponent for adiabatic conditions in ionization zone (Chapter 6.7.3)
δ	power law index for nonthermal electron energy distribution (Equation 17.11)
ε	oblateness of the Sun (Chapter 1.9)
ε_λ	radiant emissivity at wavelength λ (Chapter 2.3)
η_e	magnetic diffusivity (cm^2 $\sec^{-1}$) (Equation 16.9)
θ	variable in Saha equation = $5040/T$ (Chapter 4.2)
θ	dimensionless unit of pressure in oscillation calculation (Chapter 14.2)
ξ	vertical displacement of a parcel of gas (Chapter 13.5.1)
κ	opacity (cm^2 gm^{-1}) (Chapter 3.1)
λ	wavelength (cm, Å, μm) (Chapter 2.1)
Λ	Coulomb logarithm (Chapter 11.3)
Λ	radiative loss function (Chapter 17.14.2)
μ	$\cos(\psi)$, where ψ = angle between line of sight and Sun's normal (Chapter 2.2)
μ	mean molecular weight (Chapter 1.7)
ν	temporal frequency ($\sec^{-1}$) (Chapter 2.1)
ν_g	critical frequency in the Sun (Equation 1.20)
π	3.14159
ρ	mass density (gm cm^{-3}) (Chapter 3.1)
ρ_b	density at the base of the convection zone (Chapter 8.3)

ρ_c density at the center of a polytrope (Chapter 10.2)
$\bar{\rho}$ mean density of the Sun ($= 1.4$ gm cm^{-3}) (Equation 1.21)
σ collision cross-section for a particle (cm^2) (Chapter 11.3)
σ_B Stefan–Boltzmann constant (5.67×10^{-5} ergs cm^{-2} sec^{-1} K^{-4}) (Chapter 2.1)
σ_c Coulomb collision cross-section (Chapter 11.3)
σ_e electrical conductivity (sec^{-1} [in e.s.u.]) (Equation 16.8)
σ_λ absorption cross-section (cm^2) for a photon with wavelength λ (Chapter 3.3.1)
σ_T Thomson cross-section for photon scattering by a free electron (Equation 3.1)
τ optical depth (Chapter 2.3)
φ fractional abundance of negative hydrogen atoms (Chapter 3.4)
Φ radiative loss function (ergs cm^3 sec^{-1}) (Chapter 17.7)
ψ angle between line of sight and Sun's normal (Chapter 2.2)
ω solid angle (Equation 2.22)
ω angular frequency of solar oscillations (radians sec^{-1}) (Chapter 14.1)
ω_{ac} cut-off frequency for vertically propagating acoustic waves (Equation 13.13)
Ω angular frequency of solar rotation (radians sec^{-1}) (Chapter 1.9)

Index

Physics/Astronomy

Physics of the Sun
A First Course

Dermott J. Mullan

With an emphasis on numerical modeling, **Physics of the Sun: A First Course** presents a quantitative examination of the physical structure of the Sun and the conditions of its extended atmosphere. It gives step-by-step instructions for calculating the numerical values of various physical quantities.

The text explores how the physical conditions in the visible surface of the Sun are determined by the opacity of the material in the atmosphere. It also presents the empirical properties of convection in the Sun and discusses how the physical parameters increase with depth through the convection zone. The author shows how certain types of "real stars" are actually polytropes and offers a simplified version of oscillation equations to highlight the properties of p- and g-modes in the Sun. He also focuses on the initial temperature rise into the chromosphere, why the temperature in the quiet corona has the value it does, and how the physics of magnetic fields help us to understand various striking phenomena that are observed on the Sun.

This text enables a practical appreciation of the physical models of solar processes. Through the included numerical modeling problems, it encourages a firm grasp of the numerical values of actual physical parameters as a function of radial location in the Sun.

Features

- Presents step-by-step explanations for calculating the numerical models of the photosphere, convection zone, and radiative interior
- Describes the structure of polytropic spheres
- Covers the empirical properties and oscillation equations of helioseismology
- Explores the acoustic power in the Sun and the process of thermal conduction in different physical conditions

C3074

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

ISBN: 978-1-4200-8307-1
90000
9 781420 083071

www.crcpress.com