Lectures on Solar Physics: From the core to the heliopause

Hannu Koskinen and Rami Vainio

Spring 2011

Preface

These lectures are intended for senior undergraduate and post-graduate level students in physics and astrophysics. As the target group includes physicists, no previous knowledge in astronomy/astrophysics is required. Compared to previous editions of the of these notes, however, we have cut down the material on solar interioir models somewhat, as there are courses in astronomy offered at the Department of Physics that are dealing with these topics.

The course will deal with the physical principles of the structure of the Sun, energy production and transfer, rotation, oscillations, convection, magnetism, magnetohydrodynamics, solar activity, solar wind, and heliosphere.

These lecture notes follow mostly two famous text-books:

M. Stix, *The Sun: An Introduction*, Springer, 1989; 2nd edition, 2002.
(The 2nd edition contains about 30 % more material than the 1st edition.)

E. R. Priest, *Solar Magnetohydrodynamics*, Reidel, 1982.

Important material is taken from:

K. R. Lang, *The Sun from Space*, Springer, 2000.

We also utilize the recent textbook:

C.J. Schriver & G.L. Siscoe (eds.), *Heliophysics – Space Storms and Radiation: Causes and Effects*, Cambridge University Press, 2010.

For students wishing to learn more about the basic concepts of astrophysics there is a very recommendable source in Finnish:

H. Karttunen, K. J. Donner, P. Kröger, H. Oja, and M. Poutanen, *Tähtitieteen perusteet*, Ursa 1995. The latest edition is from 2010.
(The book is also available in English and German.)

Contents

Chapter 1

Overview of the Sun

The Sun is sometimes told to be a typical, rather boring main sequence star. While it is a rather average late type star, it is yet far from uninteresting. On the contrary, the Sun is a very complicated object exhibiting highly variable and complex activity that cannot be simulated in laboratories and even the most powerful computers are still far away from the capacity of a detailed modelling of the largely variable parameter ranges from the hot interior of the Sun to its cool surface and again to the hot corona. The Sun is not unimportant either. For an astronomer it is the only star that can be observed in great detail and, of course, its existence and properties are critical to us on Earth.

The modern picture of the Sun started to develop in the dawn of modern physical sciences when *Galileo Galilei*, one of the first developers and users of the telescope, found sunspots on the solar disc in about 1610–1613. However, after this the development remained rather slow. In 1802 *Hyde* discovered that the solar spectrum contained several absorption lines which were later catalogued by *Fraunhofer*. In 1844 *Schwabe* showed that the sunspot activity varies in an 11-year cycle. In 1859 *Carrington* and *Hodgson* independently observed a solar flare in white light. They noted that 17 hours after the flare a magnetic storm commenced in the near-Earth environment. The secondmost common element in the universe was identified in 1868 in the solar spectrum by *Lockyer* and later given the name helium.

Most of our present understanding of the Sun did not exist before the 20th century. Among the first major advances were *Hale's* measurements of intense magnetic fields in the sunspots in 1908, showing that whatever generated the solar activity, it was closely related to magnetism and was highly variable. One key enigma remained, however. At the end of the 19th century lord *Kelvin* had demonstrated that the largest imaginable energy source for the solar radiation, the gravitational binding energy of the Sun would not be sufficient for more than 20 million years at the present solar

luminosity, which already at that time was considered much too short. To solve this problem understanding of the nuclear forces had to be obtained, and in 1938 *Bethe* and *Critchfield* explained the dominant proton-proton reaction chain that powers the Sun. For this and other discoveries of energy production in stars Bethe was awarded the Nobel Prize in Physics in 1967.

From the 1960's the Sun has been possible to observe from space. The era since the early 1990's can be described as a golden age of solar research. The X-ray images from the Japanese/US Yohkoh (sunbeam in Japanese) satellite launched in 1991 have made the hot active Sun visible for a whole sunspot cycle. (Yohkoh observations ended in 2001 and the spacecraft burned in its re-entry to the atmosphere in 2005.) In 1995 the European Space Agency (ESA) and NASA of the United States launched the joint SOHO spacecraft to the Lagrangian libration point L1 where it still, after 13 years, makes continuous observations of the Sun in particular in UV and optical wavelengths. High-resolution measurements of the Zeeman effect provide detailed observations of the solar magnetic field and detailed Doppler measurements give unprecedented information about solar oscillations facilitating mapping of the interior structure of the Sun using a method called helioseismology. In 1998 the NASA Small Explorer series satellite, called TRACE, started making very high-resolution observations of small-scale phenomena in the solar atmosphere and corona and in 2002 another Small Explorer, RHESSI, was launched for studies of particle acceleration and explosive energy release in solar flares. In 2006, NASA launched a two-spacecraft mission called STEREO. The satellites orbit the Sun on nearly the same orbit as the Earth, one (called STEREO-A, "ahead") has a slightly inner and the other one (STEREO-B, "behind") a slightly outer orbit, so that STEREO-A advances the Earth on its orbit while the STEREO-B lags behind. The rate of separation of the S/C is about 45 degrees/year. The separation allowed for stereoscopic observations of the solar atmosphere and activity during the first year of the mission. In 2006, also another Japanese/US solar mission called Hinode (sunrise in Japanese) was launched. It carries an X-ray telescope, an EUV imaging spectrometer and an optical telescope, which can be used, e.g., for obtaining the solar magnetic field vector at high resolution. Finally, in 2010, NASA launched the new flagship mission for solar observations, *Solar Dynamics Observatory* (SDO). It carries instrumentation for remote sensing of the solar atmosphere at different wavelengths and for helioseismology. Thus, it will carry on the work of SOHO.

Some of the websites of these missions are:

Yohkoh:	http://www.lmsal.com/SXT/
SOHO:	http://sci.esa.int/soho
TRACE:	http://vestige.lmsal.com/TRACE/
RHESSI:	http://hesperia.gsfc.nasa.gov/hessi/
STEREO:	http://stereo.gsfc.nasa.gov/
Hinode:	http://solar-b.nao.ac.jp/index_e.shtml
SDO:	http://sdo.gsfc.nasa.gov/

These space missions have provided a wealth of new data for further research and paved way to future, even more advanced, missions to study the Sun. ESA, for example, is planning to send the Solar Orbiter spacecraft to an inclined orbit around the Sun reaching down to a distance of about 62 solar radii, i.e., some 28 % of the Sun-Earth distance (http://sci.esa.int/solarorbiter). It will carry an advanced set of in-situ and remote sensing instrumentation. Presently the mission is scheduled for launch not earlier than 2017, but its status in ESA's science program is still a candidate mission, and decision on its selection in the science program are expected in the fall 2011. NASA, on the other hand, will launch by 2018 a mission called Solar Probe Plus (http://solarprobe.jhuapl.edu/). This spacecraft will fly through the solar corona reaching a minimum distance of 8.5 solar radii from the surface of the Sun.

Also the ground-based solar observatories contribute to the present progress in understanding the Sun. The large radio telescopes are able to map the plasma emission and the cyclotron radiation from the energetic electrons in solar eruptions and in the visible wavelengths penetrating through the atmosphere, the Earth is still the most cost-efficient place to make the observations. And in 2002 the Sudbury neutrino observatory conclusively showed that the solution to the long-standing solar neutrino problem really lies in the physics of neutrinos and thus there is no need for any fundamental changes in the models of the solar interior.

An extra boost to solar research has come during the last 15 years from an emerging sector of space research, space weather. The term refers to temporally changing conditions in the Sun, solar wind, magnetosphere, ionosphere, and atmosphere, which can be hazardous to technological systems in space and on ground and may threaten human life or health. The Sun is the driver of space weather and in order to reduce the hazardous consequences either by system design or reliable forecasting we need to learn much more details of solar physics. Space weather requires continuous monitoring of the Sun and solar wind. A useful website for real-time information is

http://www.swpc.noaa.gov/

1.1 Basic facts about the Sun

Let us begin by summarizing some basic numbers about the Sun. These will be discussed in greater detail later in the text.

- Age = 4.5×10^9 years
- Mass, $m_\odot = 1.99 \times 10^{30}$ kg ($\approx 330\,000\ m_E$, mass of the Earth)
- Radius, $r_\odot = 696\ 000$ km ($\approx 109\ R_E$, the Earth's radius)
- Average density = 1408 kg/m^3 = 1.408 g/cm^3
- Average distance from the Earth (1 AU) = 150×10^6 km (215 $r_\odot$)
- Gravitational acceleration on surface = 274 m/s^2
- Escape velocity on surface = 618 km/s
- Luminosity = 3.84×10^{26} W
- Rotation period at equator = 26 days
- Mass loss rate $\approx 5 \times 10^9$ kg/s
  – radiation: 4×10^9 kg/s;
  – solar wind: 1×10^9 kg/s
- Effective black body temperature = 5778 K

Spectral classes

Stars are divided to several spectral classes according to decreasing effective temperature: O, B, A, F, G, K, M, R, N,$\cdots$. These are further divided to subclasses. The Sun belongs to class G2.

Exercise: Hertzsprung-Russell diagram

With help of literature, get acquainted with the so-called Hertzsprung-Russell (H-R) diagram.

Structure of the Sun

Figure 1.1 shows the large scale structure of the Sun as we know it today. The energy is produced in the hot central core. Outside the core the energy flows outward radiatively to the distance of about 0.72 $r_\odot$. There the radiation is no more efficient enough and convective motion takes over the energy transfer. The thin surface of the Sun absorbs almost all the the energy and radiates it out as a black body at the temperature of 5778 K.

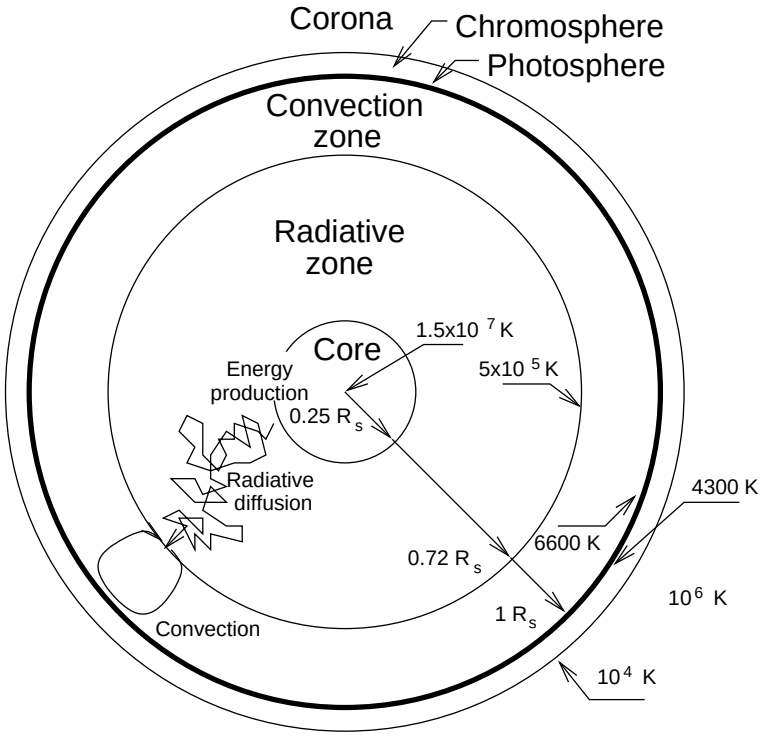

Figure 1.1: The Sun.

1.2 Where is the Sun?

The Sun is located in a spiral arm of our galaxy, the Milky Way. For us it is of importance to determine the distance to the Earth. From Kepler's third law we get

$$\frac{a^3}{T^2} = \frac{Gm_\odot}{4\pi^2}\left(1 + \frac{m_E}{m_\odot}\right), \tag{1.1}$$

where a is the semimajor axis of the Earth's orbit, T the orbital period, and G the gravitational constant ($6.673 \times 10^{-11}\,\mathrm{m^3 s^{-2} kg^{-1}}$). If we knew $m_\odot$ and T we would get a (because $m_E \ll m_\odot$).

Traditionally a was determined by triangulation with two planetary bodies. From Kepler's law we can derive

$$\left(\frac{a_1}{a_2}\right)^3 = \left(\frac{T_1}{T_2}\right)^2 \frac{1 + m_1/m_\odot}{1 + m_2/m_\odot}\,. \tag{1.2}$$

Now we need the masses m_i in units of $m_\odot$. These are obtained from mutual perturbations of the the planetary orbits.

Since 1961 more accurate determination has been obtained using radar echos, but not from the Sun as it is not a very homogeneous reflector. Instead echos from other planets are used and put into Kepler's law. This gives us the **light time** for the unit distance τ_{AU}

$$\tau_{\mathrm{AU}} = 499.004782 \pm 0.000006\ \mathrm{s}\,. \tag{1.3}$$

Using the **exact** value of the velocity of light $c = 299\,792\,458$ m/s we get the length of **the astronomical unit** (AU):

$$1\,\mathrm{AU} = 149\,597\,870 \pm 2\ \mathrm{km}\,. \tag{1.4}$$

Thus, for practical purposes the mean distance to the Sun is 149.6 million kilometers, which we will hereafter use as the value of AU. Note that the Earth's orbit is elliptical:

Perihelion in January:	147.1 million km
Aphelion in July:	152.1 million km

Exercise

What distance on the center of the solar disc does one arcsec (1") correspond to (at perihelion, at aphelion, on average)?

1.3 Mass of the Sun

If a, T, $m_E \ll m_\odot$ are known, Kepler's law gives $Gm_\odot$ with an accuracy of about 8 significant numbers. But G is the most inaccurately known natural constant, whose official error margin was increased so late as in 1998! Consequently we know the present solar mass within an error of about 0.15 per cent:

$$m_\odot = (1.989 \pm 0.003) \times 10^{30} \text{ kg}\,. \tag{1.5}$$

At present mass is lost 4×10^9 kg/s through radiation and 10^9 kg/s is carried away by the solar wind.

Exercise

How much mass has the Sun lost during its lifetime assuming the present loss rate?

1.4 Size of the Sun

The angular semidiameter of the solar disc is 960.0 ± 0.1" (i.e., the angular diameter is 32'$\approx 0.5°$). The surface is defined to be a little deeper in the atmosphere. In calculations of these lectures we use for the radius

$$r_\odot = 696\,000 \text{ km}\,. \tag{1.6}$$

Thus the mean density is 1408 kg/m^3 and for the gravitational acceleration on the surface we find

$$g_\odot = \frac{Gm_\odot}{r_\odot^2} = 274 \text{ ms}^{-2}\,. \tag{1.7}$$

1.5 Luminosity

The solar constant S expresses the total irradiance at the mean distance of the Earth. S can be measured directly and is usually given as

$$S = 1367 \pm 3 \,\text{Wm}^{-2}\,. \tag{1.8}$$

S is related to **the luminosity** of the Sun $L_\odot$ by

$$L_\odot = 4\pi \,\text{AU}^2\, S = (3.844 \pm 0.010) \times 10^{26} \,\text{W}\,. \tag{1.9}$$

Note that **the solar constant is not constant**. The luminosity of the newly-born Sun was about 72% of its present value. Furthermore, the present "solar constant" varies by a factor of

- 10^{-6} over minutes
- 2×10^{-3} (0.2 %) over several days
- 10^{-3} over solar cycle (exact value uncertain)

The physical reasons and apparent periodicities of these variations are not fully understood.

Accurate determination of S requires that it is observed above the dense atmosphere of the Earth, which absorbs most of the radiation in ultraviolet (UV) and infrared (IR) wavelengths. Recent inter-calibrations between various space observations indicate that the on average $S \approx 1366$ Wm^{-2} around solar minima and $S \approx 1367$ Wm^{-2} near solar maxima. However, around solar maxima the irradiance varies by several Wm^{-2} and thus the conservative error estimate in (1.8) is appropriate.

Note that the solar variability seems about a factor of three weaker than typical variations in other Sun-like stars. It is possible that the luminosity variations in the solar polar regions cannot be measured quite correctly from the nearly equatorial direction whereas the average viewing angle of other stars is about 30° off the equatorial plane. However, most likely the present Sun is less variable than typical Sun-like stars.

Luminosity can be given in terms of effective temperature defined by

$$L_\odot = 4\pi r_\odot^2 \sigma T_{\rm eff}^4 \tag{1.10}$$

where $\sigma = 5.6704 \times 10^{-8}\,\mathrm{Wm^{-2}K^{-4}}$ is **the Stefan-Boltzmann constant**. For the Sun $T_{\rm eff} = 5778 \pm 3$ K.

Exercise

The apparent brightness of a star is usually given in terms of magnitude. Explain the concepts of **apparent magnitude** and **absolute magnitude**. What is the absolute magnitude of the Sun?

Exercise

Derive the effective temperature starting from Planck's law

$$B_\lambda = \frac{2hc^2}{\lambda^5 (e^{hc/\lambda k_{\rm B} T} - 1)}\,. \tag{1.11}$$

1.6 Solar spectrum

1.6.1 Spectrum of a star

In 1859 Kirchhoff formulated the general laws governing the production of spectrum:

1. The ratio of emissivity to absorptivity is independent of the composition of the material and depends only on the temperature and wavelength.
2. An opaque body radiates a continuous spectrum.
3. A transparent gas radiates an emission spectrum that is distinct for each chemical element.
4. An opaque body surrounded by a gas of low emissivity shows a continuous spectrum crossed by absorption lines corresponding the spectrum of the gas.
5. If the gas has high emissivity, the continuous spectrum will be crossed by bright lines

Kirchhoff considered solid bodies as continuous emitters but the laws are valid for stars as well. The increasing density toward the stellar surface makes them opaque because the various absorption processes jointly block the radiation at all frequencies. This takes place in the photosphere from which the continuous black-body spectrum originates. We will later discuss the opacity of the Sun more in detail.

The same spectral line may show both as an emission and as an absorption line. An important example in the Sun is the **hydrogen Balmer series line** at 656.3 nm (Hα). In the photosphere the line is an absorption line whereas in the tenuous chromosphere it is an emission line. Thus by using a narrow band-pass filter at this frequency we observe the chromosphere without the photosphere background.

1.6.2 Irradiance, energy flux, and intensity

The solar irradiance $S(\lambda)$ is the energy flux observed at a given distance (in our case 1 AU) per unit area, time, and wavelength interval. It is related to the energy flux $F(\lambda)$ at the solar surface simply by

$$r_\odot^2 F(\lambda) = \mathrm{AU}^2\, S(\lambda)\,. \tag{1.12}$$

The second important quantity is **the intensity** $I(\theta, \phi, \lambda)$ (or $I(\theta, \phi, \nu)$), i.e., the energy emitted per unit area, time, wavelength/frequency interval, and solid angle (in SI-units $\mathrm{Jm^{-2}s^{-1}m^{-1}sr^{-1}}$ or $\mathrm{Jm^{-2}s^{-1}Hz^{-1}sr^{-1}}$). θ is the polar angle from a given direction and ϕ the azimuthal angle around the same direction. Note that the wavelength/frequency dependence is often denoted by $I_\lambda(\theta, \phi)$ or $I_\nu(\theta, \phi)$.

Exercise

Show that

$$\lambda I_\lambda = \nu I_\nu \,. \tag{1.13}$$

In the following, we assume no azimuthal variations and drop the ϕ-dependence. Thus the intensity depends on angular distance θ from the direction perpendicular to the solar surface. The integral of $I(\theta, \lambda)\cos\theta$ over all outward directions ($\cos\theta > 0$) yields **the energy flux** $F(\lambda)$

$$F(\lambda) = 2\pi \int_0^{\pi/2} I(\theta, \lambda)\cos\theta \sin\theta \, d\theta \equiv \pi \overline{I}(\lambda) \,. \tag{1.14}$$

We often denote $\mu = \cos(\theta)$ when the integral above reads

$$\overline{I}(\lambda) = 2 \int_0^1 I(\mu, \lambda)\mu \, d\mu = 2I(1, \lambda) \int_0^1 \frac{I(\mu, \lambda)}{I(1, \lambda)} \mu \, d\mu \,. \tag{1.15}$$

To measure $F(\lambda)$ we must either measure $\overline{I}(\lambda)$ directly from all parts of the solar disc, or the **central intensity** $I(1, \lambda)$ and the **limb darkening** function $I(\mu, \lambda)/I(1, \lambda)$. The latter method has the advantage that only a relative measurement of the diffuse light is needed.

1.6.3 Visible spectrum

Most of the solar energy is irradiated in the visible and near-infrared parts of the spectrum. Figure 1.2 shows the visible spectrum. The red side of the spectrum is almost continuous black-body spectrum with some strong absorption lines, e.g., Hα at 656.3 nm. On the blue side there are more absorption lines.

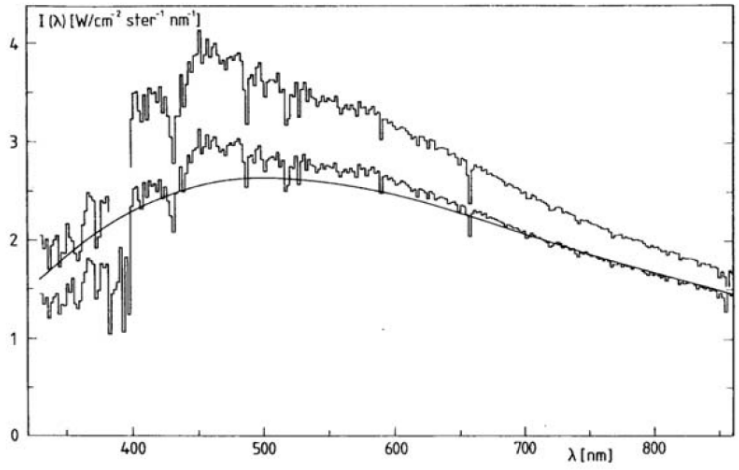

Figure 1.2: Central intensity (upper curve) and mean intensity (lower curve) at the visible wavelengths. The solid line is the black-body spectrum at the temperature 5777 K.

1.6.4 Infrared spectrum

About 44% of the electromagnetic energy is emitted at $\lambda > 0.8\,\mu$m. The spectrum is approximately thermal and can be represented by **the Rayleigh-Jeans law**

$$S(\lambda) \simeq 2ck_BT\lambda^{-4}(r_\odot/\mathrm{AU})^2\,. \tag{1.16}$$

The infrared spectrum is efficiently absorbed by water vapor in the Earth's atmosphere.

1.6.5 Radio spectrum

Radio wavelengths are longer than 1 mm. Instead of wavelength, frequency is often used to characterize the emissions. Recall the simple conversion formula $\lambda(\mathrm{m}) = 300/f(\mathrm{MHz})$. Thus, e.g., 1 mm $\leftrightarrow$ 300 GHz. Figure 1.3 presents the solar radio emissions in terms of flux per frequency interval. The figure illustrates that the Sun is strongly variable at these wavelengths. The reason for this variability is that the radio emissions originate from non-thermal plasma processes in the solar atmosphere, chromosphere, and corona. During strong solar disturbances the radio emissions can exceed the quiet levels by several orders of magnitude. Note also that the slope at longest wavelenghts for the quiet Sun corresponds to higher temperatures (10^6 K) than the main black body radiation. This tells that the chromosphere and corona are much hotter than the Sun itself.

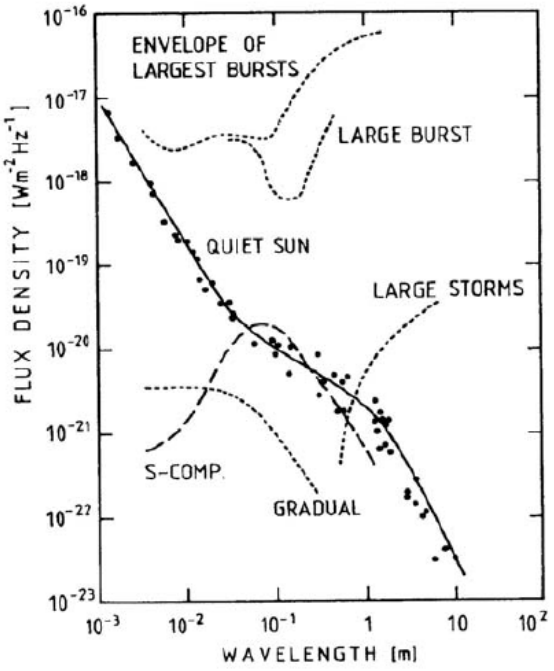

Figure 1.3: Solar radio emissions. Dots and the solid curve represent the quiet Sun, the dashed line (S-comp.) is a slowly varying component correlated to the solar cycle, and the dotted lines illustrate the rapidly varying events in the solar atmosphere.

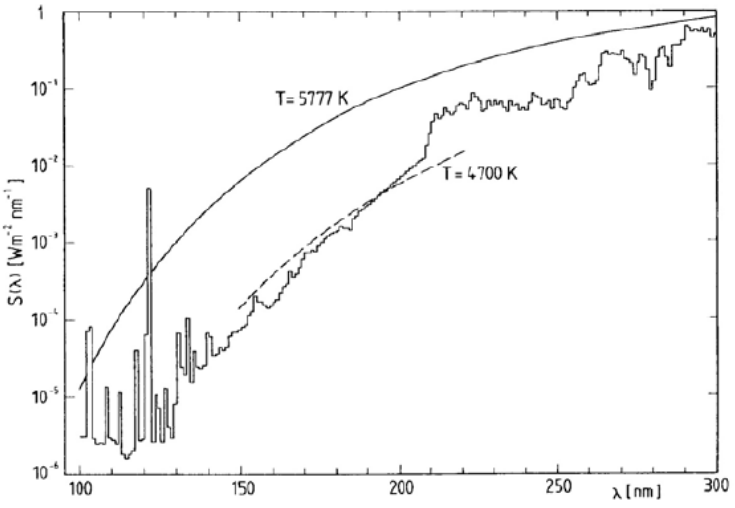

Figure 1.4: Solar UV spectrum down to 100 nm.

1.6.6 Ultraviolet and shorter wavelengths

Figure 1.4 illustrates the UV spectrum. Absorption lines are dominant down to 210 nm. At shorter wavelengths the intensity is reduced to correspond to the temperature of 4700 K. This reduction is due to absorption by the ionization of Al I. (Recall the standard notation: Al I represents the non-ionized aluminum, Al II is the same as Al^+, Al III is Al^{2+}, etc.) Below 150 nm emission lines start to dominate the spectrum. The strongest is the hydrogen Lyman α line centered at 121.57 nm. Its average irradiance 6 mWm^{-2} is as much as all other emissions below 150 nm together.

At short wavelengths the spectrum becomes highly variable illustrating nonuniform distribution of the emission sources in the solar atmosphere. This nonuniformity is both spatial and temporal. The wavelength band below 120 nm is called extreme ultraviolet (EUV). These emissions come both from neutral atoms and from ions up to very high ionization levels, e.g. Fe XVI (i.e., Fe^{15+}) in the solar corona. This makes it possible to study a wide range of temperatures from 8000 K to 4×10^6 K from the chromosphere to the corona. This is utilized by several instruments of the SOHO and TRACE spacecraft.

Solar flares increase the EUV and soft X-ray (0.1 nm – 10 nm) spectrum quite considerably (Figure 1.5). Also hard X-rays and γ-rays are emitted in these processes.

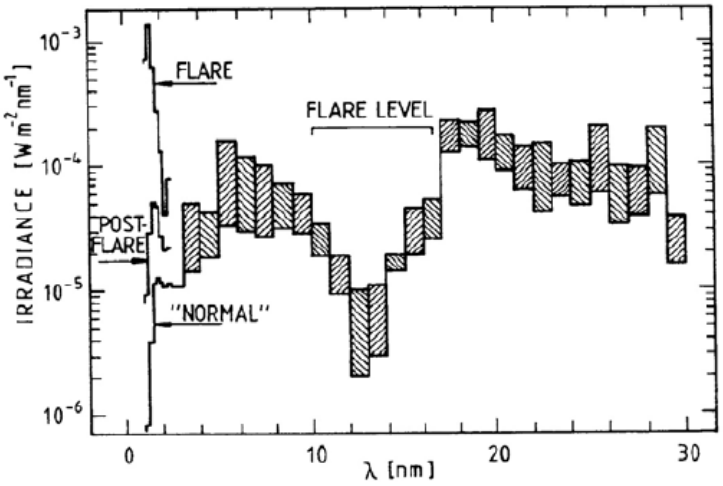

Figure 1.5: EUV and soft X-ray spectrum. The shaded intervals represent the variation without flares. The flare level is indicated. Note also the strong effect of flares in at wavelengths below 2 nm.

Chapter 2

The standard model

We can get information of the internal structure of the Sun through three main means: energy flux, neutrinos, and solar oscillations. Of these the energy flux is a strongly averaged source of information and neutrinos are very difficult to observe. However, during the last 30 years observations of solar oscillations have become an increasingly important tool for detailed studies of the interior of the Sun. We will return to the oscillations and solar seismology in chapter 4. In this chapter, we will give a brief review on the model of the solar interior, called the standard model.

2.1 History of the Sun

A useful solar model must yield correct luminosity, radius, and age of the present Sun. In order to create the model we need to understand something of the evolution history of the Sun. The reason is that we cannot directly measure the helium content in the core, i.e., how much of its hydrogen reservoirs the Sun has so far burned to helium. Note that as the hydrogen is fused to helium the **mean molecular mass** μ increases. In order to keep the thermal pressure $\rho\mathcal{R}T/\mu$ large enough to withstand gravity the temperature and/or density of the core must increase. Due to the properties of the nuclear reactions, this implies an increasing energy production and thus increasing luminosity. Thus our Sun is getting brighter, slowly but unavoidably.

The best way to determine the age of the solar system is to study meteorites. They are likely the oldest bodies of the solar system. An important clue to their age is given by ^{87}Rb that decays to ^{87}Sr with a half-life of 4.8×10^{10} yr. Comparing the relative abundances of ^{87}Rb and ^{87}Sr to the abundance of the stable ^{86}Sr in samples from many different meteorites the age of the solar system has been estimated to be $(4.55 \pm 0.05) \times 10^9$ yr

(according to recent studies the error margin may even be a factor of 10 smaller). But when did the Sun ignite?

A plausible scenario for a formation of a Sun-like star is the following:

1. Take an interstellar gas cloud with a mass of the order of $10^4\, m_\odot$. Some perturbation, e.g., an interstellar shock wave, leads to gravitational collapse if the self-gravity due to the perturbation exceeds the internal pressure of the gas. This condition is called the **Jeans criterion**:

$$\frac{Gm_c}{r} > \frac{\mathcal{R}T}{\mu}\,, \tag{2.1}$$

where $\mathcal{R}$ is the gas constant and μ, T, m_c, r are parameters of the cloud.

In reality the collapse is not spherically symmetric due to rotation and magnetic field. Angular momenta per unit mass are

for typical interstellar clouds	10^{18} m^2/s,
for the present solar system	10^{16} m^2/s.

Typical magnetic flux densities in interstellar clouds are $0.1 - 1$ nT, which, if compressed to the size of a solar system, yield about 10^6 T. Even the highest (natural) concentrations of magnetic field in the present solar system are below 1 T. Thus virtually all initial angular momentum and magnetic flux have disappeared. The magnetic field provides an effective lever arm for the torque to remove angular momentum. This phenomenon is called magnetic braking and it plays a role also in the solar rotation as we will see later.

Exercise

Calculate the free-fall time $t_{\rm ff}$, i.e., the time that a spherically symmetric cloud with initial density ρ and negligible internal pressure needs for complete collapse. Show that for $\rho = 10^{-20}$ kg/m^3 , $t_{\rm ff} \approx 3 \times 10^7$ yr.

2. During collapse the cloud must fragment to form some $10^3 - 10^4$ stars. All this appears to happen during $t_{\rm ff}$.

3. Next the fragments collapse. The center of a collapsing fragment becomes optically thick and heats up until hydrostatic equilibrium is reached. This is called a **protostar**.

4. The protostar evolves quickly ($< 10^6$ yr) to the main sequence of the H-R diagram. A cool star is born. Its parameters are:

$T_{\rm eff}$	$\approx$	3000 K
$T_{\rm core}$	$<$	10^6 K (not yet hydrogen burning)
r	$\approx$	$4\ r_\odot$
L	$>$	$L_\odot$ (several times)

5. The cool star contracts gravitationally. This heats its core until hydrogen burning ignites. The present uncertainties in the dating of the meteorites are less than the length of the whole pre-main-sequence evolution of the Sun but we do not know the exact time of meteorite formation during this period. Thus a reasonable conservative estimate for the age of the Sun is

$$t_\odot = (4.57 \pm 0.05) \times 10^9 \,\mathrm{yr}\,. \tag{2.2}$$

2.2 Basic equations

In this section we formulate a set of equations that govern the solar structure and evolution.

2.2.1 Conservation laws

Continuity of mass

Continuity of mass is usually given in the form $\partial m/\partial r = 4\pi\rho r^2$. However, in case of the Sun it is more convenient to consider the mass m interior to a sphere of radius r as the independent variable

$$\frac{\partial r}{\partial m} = \frac{1}{4\pi\rho r^2}\ . \tag{2.3}$$

We assume here that the Sun is in almost steady state, but in general $r = r(m,t)$ and thus we retain the partial derivatives. The reason to take m as the independent variable is that, except at the very beginning, the mass loss has been negligible and we know $m_\odot$ throughout the whole period to which our model will be applicable. Thus we can place the boundary conditions at $m = m_\odot$. On the other hand, we do not know the radius, except for the present Sun, and the model has to give it as a result.

Conservation of momentum

For most of the present discussion we can consider the Sun in the hydrostatic equilibrium, $\partial P/\partial r = -\rho g$. Using the continuity of mass and $g = -Gm/r^2$ we obtain

$$\frac{\partial P}{\partial m} = -\frac{Gm}{4\pi r^4}\,. \tag{2.4}$$

Note that this equation does not describe the collapse of the protostar. The time-dependent momentum equation is

$$\frac{\partial^2 r}{\partial t^2} = -4\pi r^2 \frac{\partial P}{\partial m} - \frac{Gm}{r^2}\,. \tag{2.5}$$

Energy balance

Let $L(m)$ be the luminosity generated inside the sphere of mass m, ε the rate of energy generation per unit mass, and S the entropy per unit mass (i.e., the specific entropy). Then

$$\frac{\partial L}{\partial m} = \varepsilon - \underbrace{\frac{\delta q}{\delta t}}_{\text{heating rate}} = \varepsilon - T\frac{\partial S}{\partial t}\,. \tag{2.6}$$

The first from of the equation states that the energy produced inside the star is converted to luminosity and heat. The second form of the equation relies on the definition of specific entropy, $dS = \delta q/T$. During the main sequence evolution the interior of the Sun is very close to thermal equilibrium and the heating/cooling effect ($\partial S/\partial t$) is small.

2.2.2 Energy transport

The energy flux F is by definition the luminosity per unit area. We consider energy transport by radiation (F_R) and convection (F_C)

$$F = F_R + F_C = \frac{L}{4\pi r^2}\,. \tag{2.7}$$

Radiative transfer

Let θ be the angle to the local vertical direction, I_ν the intensity, κ_ν the absorption coefficient, and S_ν the source function. Then the radiative transfer is governed by

$$\cos\theta\,\frac{dI_\nu}{dr} = -\kappa_\nu\rho(I_\nu - S_\nu)\,. \tag{2.8}$$

In the interior of the Sun the photon mean free path is very small and thus the Sun is in **local thermodynamic equilibrium** (LTE). This means that the distribution of atomic levels is described by the Boltzmann distribution and the particle distributions are Maxwellian, all for the same local temperature, T, which depends slowly on radius. Thus we can replace the source function by **the Kirchhoff-Planck function**

$$B_\nu(T) = \frac{2h}{c^2}\,\frac{\nu^3}{e^{h\nu/k_{\rm B}T} - 1}\,. \tag{2.9}$$

Note that we cannot write $I_\nu = B_\nu$, because for a perfectly isotropic radiation field there is no net transport. However, we may expand $I_\nu(r,\theta) = \sum_{n=0}^{\infty} I_\nu^{(n)}(r)\cos^n\theta$, substitute the expansion to Eq. (2.8) and collect the

terms of different order in $\cos\theta$. Retaining only the first two terms gives (exercise)

$$I_\nu = B_\nu - \frac{\cos\theta}{\kappa_\nu \rho}\,\frac{dB_\nu}{dr}\,, \tag{2.10}$$

which is an extremely good approximation as the photon mean free path $(1/\kappa_\nu\rho)$ is much smaller than the gradient scale length of B_ν (determined by that of the temperature). The energy flux is then

$$F = \int_0^\infty F_\nu \, d\nu\,, \tag{2.11}$$

where

$$\begin{aligned} F_\nu = \int \cos\theta \, I_\nu \, d\Omega &= -\frac{4\pi}{3\kappa_\nu\rho}\,\frac{dB_\nu}{dr} \\ &= -\frac{4\pi}{3\kappa_\nu\rho}\,\frac{dB_\nu}{dT}\,\frac{dT}{dr}\,. \end{aligned} \tag{2.12}$$

Setting $F = L/4\pi r^2$ and using (2.3) to replace dr by dm we obtain the temperature gradient

$$\frac{\partial T}{\partial m} = -\frac{3\kappa L}{256\pi^2\sigma r^4 T^3}\,, \tag{2.13}$$

where κ is the **opacity** (Rosseland mean absorption coefficient) defined by

$$\frac{1}{\kappa} = \frac{\int_0^\infty \frac{1}{\kappa_\nu}\,\frac{dB_\nu}{dT}\,d\nu}{\int_0^\infty \frac{dB_\nu}{dT}\,d\nu}\,. \tag{2.14}$$

Thus more energy is transported at frequencies where the matter is more transparent and where the radiation field is more temperature-dependent.

Transport by convection

Radiative transport dominates the solar energy transport in the radiative zone out to the distance of about 0.7 $r_\odot$ form the center of the Sun. There the radiation transport becomes too inefficient and the medium becomes unstable for convection. The convection zone reaches up to the solar surface where the convection cells are seen as granules. Convection is a very efficient means of energy transport and the temperature gradient

$$\frac{\partial T}{\partial m} = \left(\frac{\partial T}{\partial m}\right)_C \tag{2.15}$$

is close to the adiabatic gradient, obtained from the adiabatic equation of state, $T \propto P^{1-\frac{1}{\gamma}}$ where γ is the adiabatic index of the gas, by differentiation as

$$\left(\frac{\partial T}{\partial m}\right)_{\rm ad} = \left(1 - \frac{1}{\gamma}\right) \frac{T}{P} \frac{\partial P}{\partial m} \,.$$

Convection will be discussed in more detail in Chapter 5 of these lectures.

The energy transport equation is the fourth equation of the solar model giving gradients of r, P, L, and T. In addition, constitutive relations are needed for the remaining variables ρ, S, ε, and κ.

2.2.3 Equation of state

The pressure arises from momentum transfer by particles and photons. In the case of the Sun the pressure has two components, the gas pressure (P_G) and the radiation pressure (P_R). The radiation pressure is important only in the atmosphere and solar wind. For the present solar model we, thus, apply the perfect gas law

$$P = \frac{\rho \mathcal{R} T}{\mu} \,, \tag{2.16}$$

where $\mathcal{R}$ is the gas constant and μ is the mean molecular weight (in atomic mass unit, u).

The constituents of the gas are usually denoted by

X	fractional abundance of H
Y	fractional abundance of He
Z	fractional abundance of heavier elements

The abundances are fractional abundances in mass, i.e., a mass element δm contains $X\,\delta m$ of H, $Y\,\delta m$ of He and $Z\,\delta m$ of heavies. Consider, first, a hypothetical case, where all heavy atoms are fully ionised and hydrogen and helium remain neutral. For neutral hydrogen and helium, the total number of particles in the sample is $X\,\delta m/\mathrm{u}$ and $\frac{1}{4}Y\,\delta m/\mathrm{u}$. The number of particles (ion+electrons) due to one fully ionised atom in the gas is $z+1$, where z is the nuclear charge, so the total number of particles contributed by all the heavy atoms, with their mass numbers approximated as $A \approx 2(z+1)$, is $\approx \frac{1}{2}Z\,\delta m/\mathrm{u}$. Thus, the total number of particles in the sample of mass δm is

$$\delta N = (X + \tfrac{1}{4}Y + \tfrac{1}{2}Z)\delta m/\mathrm{u}$$

so the mean molecular weight in the case of completely neutral hydrogen and helium and fully ionised heavies is

$$\mu_0 = \frac{\delta m}{\delta N} = \frac{1}{X + \frac{1}{4}Y + \frac{1}{2}Z}\,\mathrm{u}\,. \tag{2.17}$$

(Do not confuse with μ_0 of electrodynamics!) From now on, we will regard the molecular mass as measured in u and omit the unit from notation.

Ionization of H and He adds particles reducing μ as

$$\mu = \frac{\mu_0}{1+E}\,, \tag{2.18}$$

where E is the number of electrons set free by ionizing H and He divided by the number of all other particles. E is given by

$$E = \mu_0[\eta_{\rm H}X + (\eta_{\rm He} + 2\eta_{\rm He^+})Y/4]\,. \tag{2.19}$$

Here the degrees of ionization $\eta_{\rm H}$, $\eta_{\rm He}$, $\eta_{\rm He^+}$ are given by **the Saha equations**

$$\begin{aligned}
\frac{n_{\rm H^+}}{n_{\rm H}} &= \frac{2(2\pi m_e)^{3/2}(k_{\rm B}T)^{5/2}}{u_{\rm H}h^3P_e}\exp(-\chi_{\rm H}/k_{\rm B}T)\\
\frac{n_{\rm He^+}}{n_{\rm He}} &= \frac{2(2\pi m_e)^{3/2}(k_{\rm B}T)^{5/2}}{u_{\rm He}h^3P_e}\exp(-\chi_{\rm He}/k_{\rm B}T)\\
\frac{n_{\rm He^{++}}}{n_{\rm He^+}} &= \frac{2(2\pi m_e)^{3/2}(k_{\rm B}T)^{5/2}}{u_{\rm He^+}h^3P_e}\exp(-\chi_{\rm He^+}/k_{\rm B}T)
\end{aligned} \tag{2.20}$$

where the χ's are the ionization energies of H, He, and $\rm He^+$. The u's (don't confuse with atomic mass unit u), are partition functions of particles with bound electrons

$$u_i = \sum_j g_{ij}\exp(-E_{ij}/k_{\rm B}T)\,. \tag{2.21}$$

g_{ij} is the statistical weight of the jth state, and E_{ij} is the energy of that state, relative to the ground state. For an isolated atom or ion (2.21) has an infinite number of terms and diverges. However, most of the terms have energies close to the ionization energy and cut-off in a dense plasma where particles perturb each other and lower the ionisation potential. In practice the partition functions are commonly approximated by the statistical weights of the ground state:

$$u_{\rm H} = 2\ ;\ u_{\rm He} = 1\ ;\ u_{\rm He^+} = 2\,.$$

Note that the model for μ assumes that all heavy ions are fully ionised, which is, of course, not true. It is, however, a reasonable approximation given that the fractional abundance of heavies is small so the number of the electrons set free by their ionisation is not very large.

Corrections to perfect gas law

In good plasma the electrostatic energy calculated from the mean distance between the particles, $e^2/4\pi\epsilon_0\langle r\rangle$, is much smaller than the average thermal energy $3k_{\rm B}T/2$. In the Sun this relationship is ≤ 0.1. It is not negligible but small enough for the following **Debye-Hückel treatment** to be useful.

In the neighborhood of an ion the density of any other species with charge eQ ($Q = -1$ for electrons) deviates from the mean density $\langle n_z\rangle$ according to **the Boltzmann distribution**

$$n_Q = \langle n_Q\rangle \exp\left(-\frac{eQ\varphi}{k_{\rm B}T}\right) , \tag{2.22}$$

where the potential φ is determined by all charged particles. The potential is found to be

$$\varphi = \frac{eQ}{4\pi\epsilon_0 r}\exp\left(\frac{-r}{\lambda_{\rm D}}\right) , \tag{2.23}$$

where

$$\lambda_{\rm D} = \left(\frac{\epsilon_0 k_{\rm B}T}{e^2\sum Q^2\langle n_Q\rangle}\right)^{1/2} \tag{2.24}$$

is the Debye length. Expanding the potential to the first order we find the electrostatic energy density

$$U_{\rm ES} = \frac{1}{2}\sum eQ\langle n_Q\rangle\varphi_{\rm ES} = -\frac{e^3(\sum Q^2\langle n_Q\rangle)^{3/2}}{8\pi\epsilon_0(\epsilon_0 k_{\rm B}T)^{1/2}} < 0\,. \tag{2.25}$$

Thus the electrostatic correction to the pressure $P_{\rm ES} = U_{\rm ES}/3$ is negative, i.e., for given density and temperature $P < nk_{\rm B}T$. The maximum correction is found at the depth, where $T \approx 5\times10^4$ K and $P \approx 10^8$ Pa, where the correction is -5.9%. This happens somewhere half-way up in the convection zone.

Another important correction is due to electron degeneracy. Electrons are fermions and cannot be packed too closely together in the six-dimensional phase space ($\mathbf{r}$,$\mathbf{p}$). This correction becomes significant at $P \approx 10^{14}$ Pa, i.e., inside $r \approx r_\odot/2$ and reaches its maximum of 1.7% in the center of the Sun.

2.2.4 Entropy

Only the change of the entropy dS is of interest. Entropy can be expressed in the standard way as a function of two state variables. Here, we regard entropy as a function of P and T. From elementary thermodynamics we

have the relationship[1]

$$dS = \left(\frac{\partial S}{\partial T}\right)_P dT + \left(\frac{\partial S}{\partial P}\right)_T dP = \left(\frac{\partial S}{\partial T}\right)_P dT - \left(\frac{\partial S}{\partial T}\right)_P \left(\frac{\partial T}{\partial P}\right)_S dP$$
$$= c_P \left(\frac{dT}{T} - \nabla_a \frac{dP}{P}\right), \tag{2.26}$$

where

$$c_P = T\left(\frac{\partial S}{\partial T}\right)_P \tag{2.27}$$

is the specific heat at constant pressure, and

$$\nabla_a = \left(\frac{\partial \ln T}{\partial \ln P}\right)_S \tag{2.28}$$

is the adiabatic temperature gradient.

Now we must determine c_P and ∇_a in terms of T and P. After some calculation[2] we find

$$\nabla_a = \frac{P\delta}{T\rho c_P}, \tag{2.29}$$

where

$$\delta = -\left(\frac{\partial \ln \rho}{\partial \ln T}\right)_P. \tag{2.30}$$

Here we can use the ideal gas law to get (exercise)

$$\delta(P,T) = 1 - \left(\frac{\partial \ln \mu}{\partial \ln T}\right)_P,$$

where $\mu(P,T)$ is given by the Saha ionization equilibrium.

[1]Let $z(x,y)$ be a function of x and y. Thus, $dz = (\partial z/\partial x)_y dx + (\partial z/\partial y)_x dy$. Now, set $dz = 0$ on the left-hand side to obtain

$$(\partial x/\partial y)_z = -\frac{(\partial z/\partial y)_x}{(\partial z/\partial x)_y}.$$

[2]Consider specific enthalpy, $H = U + PV$, where U is the specific internal energy and $V = 1/\rho$ is the specific volume. Using $dU = T\,dS - P\,dV$ (see below), we get

$$dH = dU + V\,dP + P\,dV = T\,dS + V\,dP$$

so $T = (\partial H/\partial S)_P$ and $V = (\partial H/\partial P)_V$. Integrability of $H(P,S)$ requires that dH is an exact differential, i.e., $(\partial T/\partial P)_S = (\partial V/\partial S)_P$. Thus, $\nabla_a = (P/T)(\partial T/\partial P)_S = (P/T)(\partial V/\partial S)_P = -(P/T)(\partial S/\partial V)_P^{-1}$, and since

$$\left(\frac{\partial S}{\partial V}\right)_P = -\rho^2\left(\frac{\partial S}{\partial \rho}\right)_P = -\rho^2\left(\frac{\partial S}{\partial T}\right)_P\left(\frac{\partial T}{\partial \rho}\right)_P = -\frac{\rho^2 c_P}{T}\left(\frac{\partial T}{\partial \rho}\right)_P = -\rho c_P\left(\frac{\partial \ln T}{\partial \ln \rho}\right)_P$$

we get (2.29).

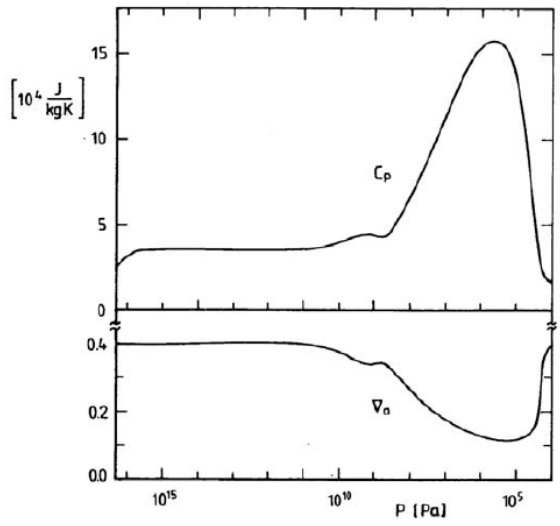

Figure 2.1: Specific heat and adiabatic temperature gradient in the solar interior as functions of pressure.

To find c_P we use the relation $T\,dS = dU + P\,dV$, where U is the specific internal energy and $V = 1/\rho$ the specific volume (recall that "specific" means per unit mass). Then

$$c_P = \left(\frac{\partial U}{\partial T}\right)_P + \frac{P\delta}{\rho T}\,. \tag{2.31}$$

Finally we need U in terms of P and T. Considering only the main contributions, the kinetic energy and hydrogen and helium ionization we have

$$U = \frac{3\mathcal{R}T}{2\mu} + \frac{1}{\rho}[n_{\mathrm{H}^+}\chi_{\mathrm{H}} + n_{\mathrm{He}^+}\chi_{\mathrm{He}} + n_{\mathrm{He}^{++}}(\chi_{\mathrm{He}} + \chi_{\mathrm{He}^+})]\,, \tag{2.32}$$

where the Saha equations are again used to derive $U(P,T)$.

Now we can determine the specific heat and adiabatic temperature gradient, insert them in the expression for dS and use this in the computations of the solar model (Fig. 2.1). Note that the deviations from the perfect gas expressions $c_P = 5\mathcal{R}/2\mu$ and $\nabla_a = 2/5$ occur mainly in the layers where H and He are partially ionized.

Exercise

Do the detailed calculation to derive equation (2.31).

2.2.5 Energy production

From the equation of state we know $\rho(P,T)$. Thus we can make a transformation from $\varepsilon(P,T)$ to $\varepsilon(\rho,T)$. This is useful, as variables (ρ,T) are more

natural in energy production calculations than (P, T).

Solar energy is produced mostly by hydrogen burning where 6.683 MeV energy per nucleon is produced. Actually the whole fusion chain involves 4 protons which finally fuse into one alpha particle, i.e., a total release of 26.732 MeV. Of all nuclei, protons have the smallest charge which is important to get two particles sufficiently close to each other. In the case of protons an electrostatic barrier of about 1 MeV must be overcome. This is significant considering that the interior temperature of the Sun is about 1.5×10^7 K, i.e., 1.3 keV only. Thus very high density is required in order that sufficiently many close encounters can take place.

There are two main reaction chains for stellar fusion. In the Sun 99% of energy comes from **the pp chain** and about 1% from **the CNO cycle**. In the discussion of these we adopt the notation X(a,b)Y where

X	is the target nucleus
a	is/are the incident particle(s)
b	is/are the emitted particle(s)
Y	is the residue nucleus

Furthermore we denote p = ^{1}H, d = ^{2}H, α = ^{4}He, and let * denote an exited state. Note that tritium is not important in the Sun whereas it is the main fuel component together with deuterium in present tokamak experiments on energy production through controlled thermonuclear fusion.

In the following tables we label the various reactions by their reaction rate symbols (λ). Q' denotes the energy delivered to the thermal bath and Q_ν the energy of the released neutrino (not to be included in ε).

pp chain

The proton-proton chain has three branches

	Reaction	Rate symbol	Q' (MeV)	Q_ν (MeV)
ppI	p(p,$e^+\,\nu$)d	λ_{pp}	1.177	0.265
	d(p,γ)^{3}He	λ_{pd}	5.494	
	^{3}He(^{3}He,2p)α	λ_{33}	12.860	
ppII	^{3}He(α,γ)^{7}Be	λ_{34}	1.586	
	^{7}Be(e^-,ν)^{7}Li	λ_{e7}	0.049	0.815
	^{7}Li(p,α)α	λ'_{17}	17.346	
ppIII	^{7}Be(p,γ)^{8}B	λ_{17}	0.137	
	^{8}B(,$e^+\,\nu$)^{8}Be*	λ_8	8.367	6.711
	^{8}Be*(,α)α	λ'_8	2.995	

Of these ppII is an alternative to (λ_{33}) and ppIII an alternative to (λ_{e7}). The branching ratios depend on reaction rates. For the present Sun the ratios are I:(II+III) = 85.2:14.8 and II:III = 14.8:0.019. Note that the last digits in this table depend on small details reaction coefficients used in model calculations. This table is from the 2nd edition of the text-book by Stix and differ slightly from the same table in the 1st edition. The most notable change is in the mean energy of the ^{8}B neutrino (6.711 MeV instead of 7.2 MeV).

While the ppIII branch is of minor importance in the total energy production it is the sequence which produces neutrinos that are easiest to detect on the Earth due to their high energy which facilitates the use of large water and heavy water detectors. The neutrino observations will be discussed further in section 2.5.

Whatever branch the pp chain takes the total energy per produced α particle is the same (26.732 MeV). Note that to get two ^{3}He nuclei in ppI chain reactions p(p,$e^+\,\nu$)d and d(p,γ)^{3}He are needed twice. Thus ppI destroys 4 protons to create one α particle and emits 2 positrons, 2 neutrinos and 2 photons.

CNO cycle

In stars whose cores are hotter, about 1.9×10^7 K, the CNO cycle dominates.

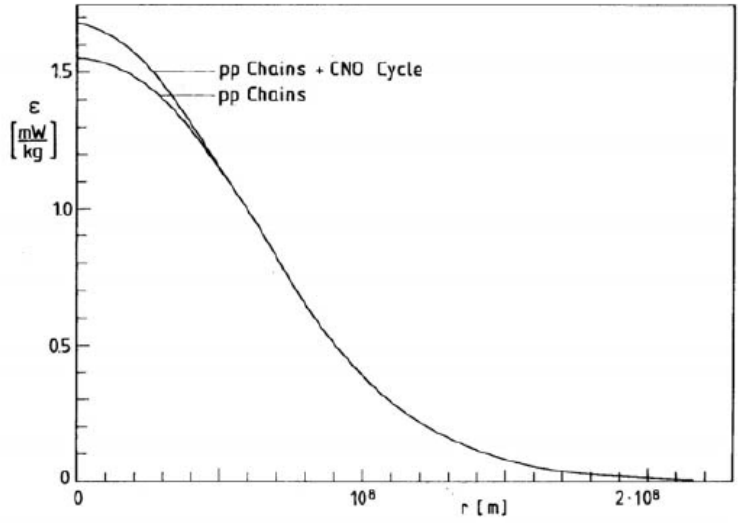

Figure 2.2: Nuclear energy generation in the Sun. Energy production is confined inside $r < 2 \times 10^8$ m, i.e., one fourth of the radius, but only 1.5% of the volume.

Reaction	Rate symbol	Q' (MeV)	Q_ν (MeV)
^{12}C(p,γ)^{13}N	λ_{p12}	1.944	
^{13}N(,e$^+$ ν)^{13}C	λ_{13}	1.513	0.707
^{13}C(p,γ)^{14}N	λ_{p13}	7.551	
^{14}N(p,γ)^{15}O	λ_{p14}	7.297	
^{15}O(,e$^+$ ν)^{15}N	λ_{15}	1.757	0.997
^{15}N(p,α)^{12}C	λ_{p15}	4.966	

Again 4 protons are consumed to produce one α particle. Carbon, nitrogen, and oxygen act as catalysts in this process.

Now the remaining task is to calculate the energy production ε as function of distance from the center. This requires careful quantum mechanical calculation of reaction rates r_{ik} between species i and j. Finally

$$\varepsilon = \sum Q'_{ik} r_{ik} \,, \tag{2.33}$$

where Q'_{ik} denote the energies of the reactions. This calculation is too long to be discussed in these lectures. The result is given in Figure 2.2.

2.2.6 Opacity

Also the opacity is natural to give in variables (ρ, T). Recall the definition

$$\frac{1}{\kappa} = \frac{\int_0^\infty \frac{1}{\kappa_\nu}\frac{dB_\nu}{dT}\,d\nu}{\int_0^\infty \frac{dB_\nu}{dT}\,d\nu}\,. \tag{2.34}$$

Thus the first task is to determine κ_ν by calculating the absorption/scattering of photons by atoms, ions, and electrons. This requires again lengthy time-dependent quantum mechanical calculations from which we get the cross sections σ for the various interactions. Below we list the results for the most important contributions.

Bound-bound absorption

Bound-bound absorption simply means a transition of an atom or an ion to a higher state of energy by absorption of a photon having the same energy as energy difference between the two states. The cross section is given by

$$\sigma_{\rm bb}(\nu) = \frac{e^2}{4\epsilon_0 m_e c} f\phi(\nu)\,, \tag{2.35}$$

where f is called **oscillator strength**, containing the transition probability, and $\phi(\nu)$ is the line profile. The two most important contributions to the line profiles are Doppler broadening

$$\phi_{\rm D}(\Delta\nu) = \frac{1}{\sqrt{\pi}\Delta\nu_{\rm D}} \exp\left[-(\Delta\nu/\Delta\nu_D)^2\right] \tag{2.36}$$

and collision broadening

$$\phi_{\rm C}(\Delta\nu) = \frac{\gamma}{(2\pi\Delta\nu)^2 + \gamma^2/4}\,. \tag{2.37}$$

In these equations $\Delta\nu = \nu - \nu_0$ is the distance from the line center, γ the constant of collisional damping (=twice the collision frequency), and $\Delta\nu_{\rm D} = (2\mathcal{R}T/A)^{1/2}\nu_0/c$ the Doppler width (A is the atomic weight). The Doppler broadening dominates at the core of the line and the collisional damping at the wings of the line.

Bound-free absorption

Bound-free absorption means absorption of a photon by ionization. The determination of the cross section requires complicated analysis. For example,

the cross section per a hydrogen-like atom in the state n is given by

$$\sigma_{\rm bf}(\nu) = \frac{m_e e^{10} Z^4 g_{\rm bf}(n,\nu)}{48\sqrt{3}\pi\epsilon_0^5 c h^6 n^5 \nu^3} \tag{2.38}$$

where $g_{\rm bf}$ is the so-called **Gaunt factor** arising from the quantum mechanical calculation of the ionization probability.

The energy of the photon $h\nu$ must exceed the ionization energy. For example the ionization energy for H is 13.6 eV. This corresponds to the frequency of 3.3×10^{15} Hz, or wavelength of 9.1×10^{-8} m = 91 nm, which is in the EUV range.

Free-free absorption

A free electron can absorb a photon in the presence of a third particle which can take the recoil momentum. This is an inverse process to bremsstrahlung. The cross section depends on the electron distribution in the velocity space

$$d\sigma_{\rm ff}(v,\nu) = \frac{e^6 Z^2 g_{\rm ff}(v,\nu)}{48\sqrt{3}\pi^2\epsilon_0^3 c h v m_e^2 \nu^3}\, dn_e(v) \tag{2.39}$$

where $g_{\rm ff}$ is again a Gaunt factor. If the small contribution from partial electron degeneracy is neglected, the electrons can be assumed to be Maxwellian and the the total cross section is

$$\sigma_{\rm ff}(\nu) = \frac{e^6 Z^2 n_e \overline{g_{\rm ff}}}{24\pi^2\epsilon_0^3 c h m_e^{3/2} (6\pi k_{\rm B} T)^{1/2} \nu^3} \tag{2.40}$$

The free-free transition is possible for all energies. Thus σ is a continuous function of ν.

Exercise

Calculate $\sigma_{\rm ff}(\nu)$ for a Maxwellian electron distribution. How is the Gaunt factor determined in quantum mechanics and what is the average Gaunt factor $\overline{g_{\rm ff}}$ here?

Scattering by electrons

This is the classical **Thomson scattering**. It is independent of photon energy and the cross section is constant

$$\sigma_{\rm s} = \frac{e^4}{6\pi\epsilon_0^2 m_e^2 c^4} \approx 6.65 \times 10^{-29}\,{\rm m}^2 \tag{2.41}$$

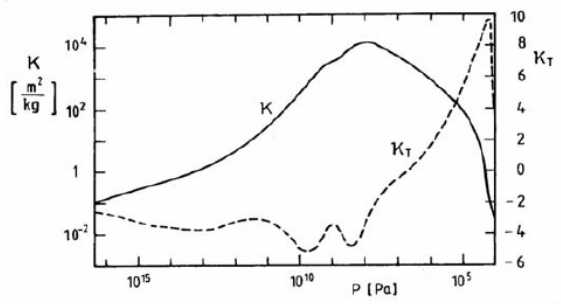

Figure 2.3: Opacity κ and its double-logarithmic temperature derivative $\kappa_T = (\partial \ln \kappa / \partial \ln T)$ as functions of pressure.

which is often given as

$$\sigma_{\rm s} = \frac{8\pi}{3} r_0^2 \tag{2.42}$$

where $r_0 = e^2/4\pi\epsilon_0 m_e c^2 \approx 2.818 \times 10^{-15}$ m is **the classical electron radius**.

If the wavelength of the photon is longer than $\lambda_{\rm D}$, the electron scattering cross section is reduced. This is important only in the core of the Sun where the opacity may be reduced by a few per cent. This may look like a minor factor, but because it reduces the central temperature, which has an effect on the branching ratios of the pp chains, this is a critical parameter in the calculation of energetic neutrino flux by ^{8}B decay.

Using these cross sections it is finally possible to calculate the radiative opacity κ per unit mass. The calculation is by no means simple. The result as a function of pressure is given in Figure 2.3.

2.3 Summary of the model

2.3.1 Differential equations and constitutive relations

Let us summarize the solar model discussed in previous sections. We have four first order differential equations:

$$\frac{\partial r}{\partial m} = \frac{1}{4\pi\rho r^2} \tag{2.43}$$

$$\frac{\partial P}{\partial m} = -\frac{Gm}{4\pi r^4} \tag{2.44}$$

$$\frac{\partial L}{\partial m} = \varepsilon - T\frac{\partial S}{\partial t} \tag{2.45}$$

$$\frac{\partial T}{\partial m} = \begin{cases} -\dfrac{3\kappa L}{256\pi^2\sigma r^4 T^3} & \text{in stable layer} \\ \left(\dfrac{\partial T}{\partial m}\right)_C & \text{in unstable layer} \end{cases} \tag{2.46}$$

In addition we have four constitutive relations

$$\rho = \rho(P,T) \tag{2.47}$$

$$dS = dS(P,T) \tag{2.48}$$

$$\varepsilon = \varepsilon(\rho,T) \tag{2.49}$$

$$\kappa = \kappa(\rho,T) \tag{2.50}$$

2.3.2 Boundary conditions

In order to integrate the differential equations we need four boundary conditions. Two of them are convenient to be imposed at the center of the Sun ($m = 0$): $r(0) = 0$; $L(0) = 0$.

The other two are imposed on the surface. We define the **optical depth** by

$$\tau(r) = \int_r^\infty \kappa\rho\, dr' \,. \tag{2.51}$$

We will show below that the temperature equals the effective temperature at the optical depth $\tau(r_s) = 2/3$ and we take this as the definition of the surface for the interior solar model. We know the surface conditions r_s, P_s, L_s, T_s for the present Sun only, but we can derive two relationships for them, which are then the two remaining boundary conditions. In the whole atmosphere $L = L_s$ (the energy source is deep in the interior), $r = r_s$ (the atmosphere

is geometrically thin), and $m = m_\odot$ (the atmosphere is light). We change the variable of pressure equation $m \to \tau \Rightarrow$

$$\frac{\partial P}{\partial \tau} = \frac{G m_\odot}{r_s^2 \kappa} \,. \tag{2.52}$$

Integrating this through the atmosphere and assuming $P(\tau = 0) = 0$ we get the third boundary condition

$$P_s = \frac{G m_\odot}{r_s^2} \int_0^{2/3} \frac{1}{\kappa} \, d\tau \,, \tag{2.53}$$

once we can evaluate the integral. For this we need $\kappa(\tau)$, and since $\kappa = \kappa(T)$ in the atmosphere, we need $T = T(\tau)$.

To get this, we integrate the equation of radiative transfer over all frequencies making a simplifying assumption of a frequency-independent (grey) κ. For the integrated intensity we use the expansion

$$\int_0^\infty I_\nu \, d\nu \equiv I(\tau, \theta) = I_0(\tau) + \cos\theta \, I_1(\tau) \tag{2.54}$$

and the integrated equation of radiative transfer is

$$\cos\theta \frac{dI}{d\tau} = I - B \,. \tag{2.55}$$

This method is called the **Eddington approximation**. It is too coarse for the study of the atmosphere itself but good enough for finding a proper boundary condition for the model of solar **interior**. The functions I_0 and I_1 can be determined by defining momenta in angular space

$$\begin{aligned} J &= \frac{1}{4\pi} \int I \, d\Omega \\ F &= \int I \cos\theta \, d\Omega \\ K &= \frac{1}{4\pi} \int I \cos^2\theta \, d\Omega \,. \end{aligned} \tag{2.56}$$

Using the expansion of I above we find $J = 3K = I_0$ and $F = 4\pi I_1/3$. Using (2.55) we find

$$\begin{aligned} \frac{dF}{d\tau} &= 4\pi(J - B) \\ \frac{dK}{d\tau} &= \frac{1}{4\pi} F \,. \end{aligned} \tag{2.57}$$

Because F is by definition the constant energy flux, $J = B$. Integration of $dK/d\tau$ leads to

$$I_0 = \frac{3}{4\pi} F\tau + b\,. \tag{2.58}$$

Finally the constant of integration b is obtained from the condition that the far from the Sun (i.e. at $\tau = 0$) the net radiation to the upper half space is the total flux F, i.e. $b = F/2\pi$. On the other hand $I_0 = B = \sigma T^4/\pi$ and we find the fourth boundary condition

$$T^4 = \frac{L_s}{4\pi\sigma r_s^2}\,(3\tau/4 + 1/2)\,. \tag{2.59}$$

This gives the atmospheric temperature as a function of optical depth, and can be substituted in (2.53) to evaluate the pressure on the surface. This expression also shows that at $\tau = 2/3$ the temperature is $T_{\rm eff}$.

Exercise

Go through the details of the above calculations.

Finally, the solution of the standard model equations is usually obtained by iteration. One first takes a good guess at r_s and L_s, calculates T_s and P_s using (2.53) and (2.59), and then integrates inwards to $m = 0$. Usually, the model values do not agree with $r(0) = 0$ and $L(0) = 0$, but one may improve the guesses of r_s and L_s and iterate until they do.

2.4 Results of the model

Now we have a model which is generally called **the standard solar model**. Of course, its details depend on the exact computations of the constitutive relations and the assumption of the age of the Sun. It is clear that actual computations are a very demanding task. Here we present only some of the main features of the calculations.

Figure 2.4 shows the main sequence evolution of the Sun in the H-R diagram. Figure 2.5 presents the evolution of the solar radius and luminosity, and the predicted neutrino counting rate for the ^{37}Cl and ^{71}Ga experiments. (The solar neutrinos are discussed separately in the next section.) Finally, figure 2.6 presents a table of the model results for the present Sun in numerical form. These numbers are consistent with the age of the Sun of 4.57×10^9 yr. This table is very useful, e.g., to convert between pressure and depth. (The column Γ_1 gives the polytropic index that is 5/3 except just below the surface of the Sun.)

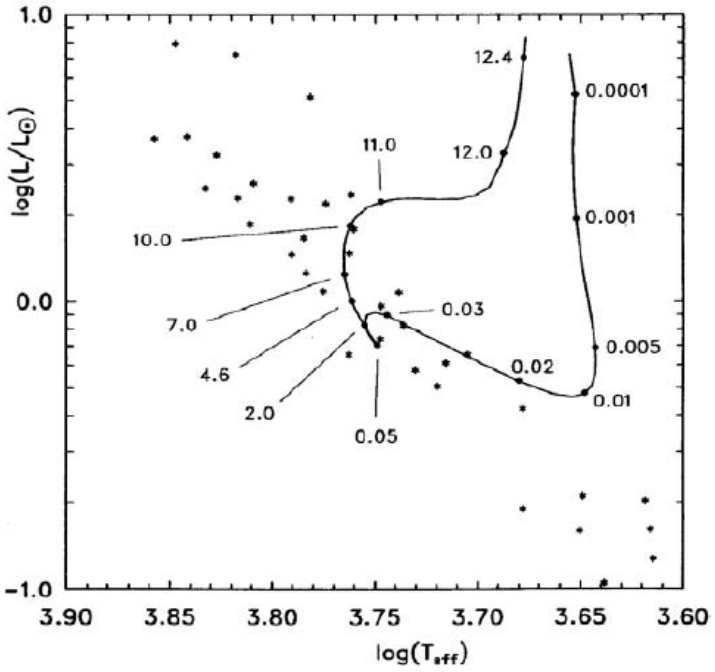

Figure 2.4: The evolution of the Sun in Hertzsprung-Russell diagram.

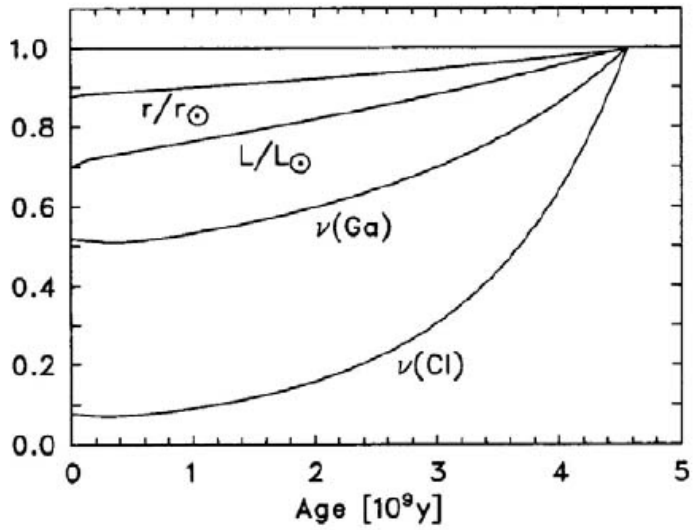

Figure 2.5: Evolution of the solar radius and luminosity and the predicted neutrino flux normalized to the present Sun.

Table 2.4. A model of the present Sun ($Z + n$ means $Z \times 10^n$)

$m/m_\odot$	$r/r_\odot$	P [Pa]	T [K]	ρ [kg/m^3]	$L/L_\odot$	X	μ	Γ_1
1.0000	1.0000	9.550+03	5.778+3	2.512−4	1.000	.735	1.247	1.636
1.0000	.9999	1.726+04	8.241+3	3.159−4	1.000	.735	1.243	1.373
1.0000	.9997	3.121+04	1.032+4	4.380−4	1.000	.735	1.195	1.215
1.0000	.9994	5.641+04	1.163+4	6.704−4	1.000	.735	1.139	1.194
1.0000	.9991	1.020+05	1.274+4	1.056−3	1.000	.735	1.089	1.191
1.0000	.9988	1.843+05	1.382+4	1.689−3	1.000	.735	1.044	1.196
1.0000	.9985	3.331+05	1.493+4	2.721−3	1.000	.735	1.004	1.203
1.0000	.9981	6.022+05	1.612+4	4.402−3	1.000	.735	.967	1.212
1.0000	.9976	1.089+06	1.743+4	7.103−3	1.000	.735	.932	1.223
1.0000	.9971	1.968+06	1.890+4	1.144−2	1.000	.735	.898	1.236
1.0000	.9965	3.557+06	2.059+4	1.836−2	1.000	.735	.865	1.252
1.0000	.9959	6.429+06	2.257+4	2.928−2	1.000	.735	.832	1.275
1.0000	.9952	1.162+07	2.495+4	4.624−2	1.000	.735	.800	1.299
1.0000	.9944	2.101+07	2.783+4	7.272−2	1.000	.735	.769	1.327
1.0000	.9934	3.797+07	3.141+4	1.127−1	1.000	.735	.741	1.363
1.0000	.9923	6.864+07	3.589+4	1.727−1	1.000	.735	.715	1.404
1.0000	.9910	1.241+08	4.166+4	2.610−1	1.000	.735	.692	1.461
1.0000	.9894	2.243+08	4.929+4	3.881−1	1.000	.735	.674	1.528
1.0000	.9874	4.054+08	5.953+4	5.669−1	1.000	.735	.660	1.585
1.0000	.9850	7.328+08	7.288+4	8.210−1	1.000	.735	.652	1.601
1.0000	.9820	1.325+09	8.938+4	1.191+0	1.000	.735	.644	1.583
1.0000	.9783	2.394+09	1.093+5	1.732+0	1.000	.735	.637	1.581
1.0000	.9737	4.328+09	1.342+5	2.513+0	1.000	.735	.630	1.604
.9999	.9681	7.823+09	1.661+5	3.619+0	1.000	.735	.624	1.633
.9999	.9611	1.414+10	2.073+5	5.187+0	1.000	.735	.620	1.654
.9998	.9523	2.556+10	2.601+5	7.413+0	1.000	.735	.618	1.662
.9996	.9415	4.620+10	3.273+5	1.058+1	1.000	.735	.616	1.666
.9993	.9281	8.352+10	4.127+5	1.509+1	1.000	.735	.615	1.667
.9987	.9117	1.510+11	5.209+5	2.152+1	1.000	.735	.614	1.668
.9978	.8917	2.729+11	6.580+5	3.069+1	1.000	.735	.613	1.668
.9963	.8676	4.933+11	8.315+5	4.378+1	1.000	.735	.613	1.667
.9940	.8387	8.916+11	1.051+6	6.245+1	1.000	.735	.612	1.666
.9904	.8057	1.586+12	1.320+6	8.826+1	1.000	.735	.612	1.665
.9746	.7101	5.959+12	2.230+6	1.955+2	1.000	.735	.611	1.665
.9423	.5988	2.200+13	3.153+6	5.180+2	1.000	.714	.614	1.665
.9019	.5193	5.750+13	3.822+6	1.120+3	1.000	.712	.616	1.666
.8569	.4611	1.205+14	4.403+6	2.039+3	1.000	.710	.616	1.667
.7872	.3991	2.754+14	5.158+6	3.984+3	1.000	.708	.617	1.667
.7173	.3539	5.144+14	5.831+6	6.589+3	1.000	.706	.618	1.668
.6342	.3117	9.321+14	6.583+6	1.058+4	.999	.704	.619	1.668
.5445	.2741	1.585+15	7.382+6	1.607+4	.995	.700	.621	1.668
.4452	.2376	2.627+15	8.296+6	2.378+4	.980	.694	.624	1.668
.3388	.2013	4.255+15	9.362+6	3.444+4	.940	.680	.631	1.668
.2430	.1686	6.401+15	1.046+7	4.720+4	.856	.655	.643	1.668
.1629	.1391	8.995+15	1.156+7	6.186+4	.721	.617	.663	1.668
.0897	.1072	1.254+16	1.280+7	8.193+4	.503	.554	.699	1.668
.0310	.0705	1.730+16	1.419+7	1.108+5	.219	.460	.761	1.668
.0034	.0321	2.183+16	1.530+7	1.417+5	.028	.369	.833	1.668
.0002	.0127	2.313+16	1.559+7	1.513+5	.002	.343	.855	1.668
.0000	.0000	2.338+16	1.565+7	1.533+5	.000	.338	.860	1.668

Figure 2.6: A model for the present Sun.

Energy transport through the Sun

The energy produced by the nuclear fusion is transported from the core by gamma rays. Due to the high density the photons are continuously absorbed and re-emitted by the gas in the radiative zone and the outward energy diffusion is so slow that it takes some 170 000 years to reach the base of the convection zone at about 0.71 $r_\odot$. At this distance the temperature has fallen from 15 million K in the core to about 2 million K.

At the bottom of the convection zone radiation becomes less and less efficient to carry energy outward and transport by convection dominates above 0.75 $r_\odot$. Convection is a much faster way of energy transport than radiation in an opaque medium. It takes only about 10 days for the heated gas to climb through the convection zone. During this process the gas also cools rapidly. The solar surface is effectively a black body that absorbs all energy coming from the convection zone and radiates it out at the temperature of 5778 K.

The future of the Sun

According to figure 2.5, the radius and luminosity of the Sun have grown at an almost constant rate. If the luminosity would now drop to 0.72 $L_\odot$, where it was 4.5 billion years ago, the Earth would become ice-covered and increasing the luminosity back to the present level would never thaw the ice due to the high albedo of the ice-covered planet. However, there is no indication that the whole Earth would ever have been ice-covered. The solution to this "faint-young-Sun paradox" most likely lies in the evolution of the atmosphere. The greenhouse effect has probably been more efficient in the ancient atmosphere than it is today. That means that the early atmosphere has been more efficient to absorb the solar energy than the present.

On the other hand the Sun will keep on becoming brighter slowly but unavoidably, and this will have consequences on Earth. Note that this is a matter of millions or billions of years and should not be confused by the cycles of ice-ages the time scales of 40 000–100 000 years or the slow cooling trend during the previous millenium, which ended by the rapid warming during the second half of the 20th century. While there is a correlation between the temperatures and the solar activity, the ice-ages are more likely related to the changes in the Earth's orbital motion and orientation, whereas the recent "global warming" is most likely due to the enhanced greenhouse gas emissions due to fossil fuel burning. In any case, in the long run the solar irradiation increases and at some point the atmospheric temperature will have risen so much that the oceans begin to boil. There is some controversy when this will happen as the solar models give different time scales for the

last phases of the hydrogen burning and the strength of future greenhouse effect is unknown. The estimates vary between 1 and 3 billion years.

So far the Sun has burned about half of its hydrogen content in the core. After 5 billion years more it must begin to burn hydrogen in the outer layers. At this time the Sun will leave its place in the main sequence of the H-R diagram. At the age of about 12 billion years the both the luminosity and the radius of the Sun have increased by a factor of about 10 and the Sun has become a red giant. Hereafter the evolution is much faster than now. After some 100–200 million years the red giant is assumed to flash for a while out to 100 $r_\odot$ (i.e., beyond the present orbit of Mercury) at the luminosity of about 1000 $L_\odot$. 100 million years more and the helium burning in the core begins which causes several flashes within the last 1–2 million years.

When the helium is burned out, radiation pressure blows the outer layers of the sun into the interstellar space forming a planetary nebula. The hot inner core becomes a white dwarf some 12.3×10^9 years after the birth of the Sun. The white dwarf is too light to compress further and there will be no fusion of heavier elements. It will slowly cool down to lower and lower temperatures.

2.5 Solar neutrinos

One of the most famous problems of solar physics has been the so-called solar neutrino problem. In very simple terms it means that the standard solar models predict a larger neutrino flux than has been observed.

As already noted the pp-chain was suggested as the dominating solar energy production scheme in 1938. Before the first neutrino observations in 1967 the solar models had evolved so far that one could predict with confidence a production of a copious amount of 2×10^{38} neutrinos per second. Figure 2.7 shows the calculated neutrino fluxes at 1 AU. The β-decay channels produce continuous spectra, whereas the electron capture of ^{7}Be produces two lines at 862 keV and 348 keV, as the ^{7}Li nucleus can be either in the ground state or in the first excited state. Note that $Q_\nu = 815$ keV given in section 2.2.5 is the weighted average of these two lines.

In addition to the reactions described in section 2.2.5 also the so-called pep line is shown in Figure 2.7. It is due to the reaction p(pe$^-$,ν)d. Although this reaction occurs only at about 0.25% of p(p,e$^+$ ν)d, it is important because it produces a neutrino whose energy is relatively high (1.442 MeV). Only ^{8}B neutrinos (which also come from a very weak reaction branch) have higher energy than the pep neutrinos. The energy is important, because it easier to detect higher-energy neutrinos than the main part of the spectrum.

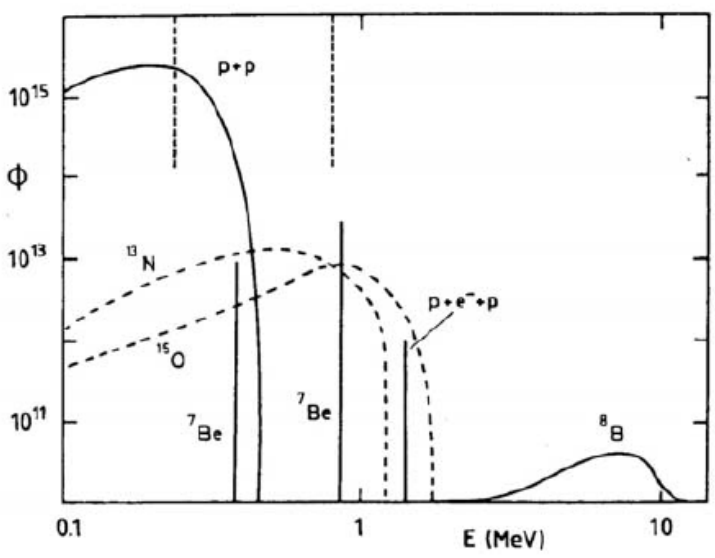

Figure 2.7: Predicted energy spectrum of the solar neutrino flux at 1 AU. the units are neutrinos/(m^2s MeV) for the continua and neutrinos/(m^2s) for the lines. The dotted vertical lines in the top mark the thresholds for the ^{71}Ga (233 keV) and ^{37}Cl (814 keV) experiments.

There are three basic types of solar neutrino detectors. The ^{71}Ga experiment has the lowest energy threshold (233 keV), the threshold of the ^{37}Cl is 814 keV, and large water detectors have the highest threshold of about 5 MeV.

The first solar neutrino experiment was made in the Homestake gold mine in South Dakota beginning in 1967. It used 615 tons of the ordinary cleaning fluid, tetrachloroethene, C_2Cl_4, whose chlorine is converted to argon in the reaction

$$\nu_e + {}^{37}\mathrm{Cl} \rightarrow e^- + {}^{37}\mathrm{Ar}\,. \tag{2.60}$$

The neutrino has sufficient energy to destroy the molecule the radioactive argon atoms remain in the vessel. The half-lifetime of ^{37}Ar is 35 days, which sets the practical experiment lenghts to about 100 days: after this time, the number of ^{37}Ar atoms has almost reached saturation level and the atoms are freely dissolved in the liquid.

The number of reactions is then obtained by counting the argon atoms. It is not an easy task as, on average, one ^{37}Ar atom is produced in every 2.17 days. After about two months of observations 30 atoms are extracted among more than 10^{30} atoms altogether. This is made with the success rate exceeding 90%!

The next step was the Kamiokande water experiment in Kamioka, Japan,

that started operating in 1987. Water detectors detect neutrinos through Čerenkov light from elastic $\nu - e^-$ scattering if the recoil energy of the electron is at least 5 MeV. While the water tank can observe only the ^{8}B neutrinos, it is possible to determine their arrival direction and thus it is known which neutrinos really come from the direction of the Sun.

The lowest energy threshold has the reaction

$$\nu_e + {}^{71}\mathrm{Ga} \rightarrow e^- + {}^{71}\mathrm{Ge}\,, \tag{2.61}$$

which produces radioactive germanium (half-life 11.4 days). In 1990 the Soviet-American Gallium Experiment (SAGE) began to operate in Caucasus and in 1991 the multinational GALLEX experiment in Gran Sasso, Italy.

Due to the extremely weak interaction between neutrinos and ordinary matter, a special unit for the solar neutrino flux has been taken into use: **the solar neutrino unit** (snu). 1 snu corresponds to the capture of one neutrino per second per 10^{36} target atoms. Different model calculations give somewhat different fluxes. Typical predictions for observable fluxes for the gallium, chlorine, and water detectors are

^{71}Ga:	130 ± 7 snu	(GALLEX and SAGE)
^{37}Cl:	7.7 ± 1.1 snu	(Homestake)
Water:	1.0 ± 0.15 snu	(Kamiokande)

The chlorine detectors observe about 2.5 snu, i.e., about one third of the predicted flux. The gallium detectors have the best record, some 50% of the predicted value, and the water detectors fall between these two. This is the famous **solar neutrino problem** that was realized already in the first chlorine experiments.

The neutrino deficit has been attributed to several different causes. In principle, it could be an observation problem. This was a popular explanation as long as the Homestake observations were alone. But after the gallium and water detector results, this appears very unlikely. For a long time it was thought that the origin of the problem would be in erroneous solar models. Quite a number of attempts to correct the models have been made but with no success. Too simple fixes easily lead to problems somewhere else in the models.

One strategy to look for **non-standard solar models** has been to reduce the temperature, which would reduce the ^{8}B neutrino flux and help at least with the original chlorine experiment problem. This is a reasonable idea, as the the ^{8}B neutrino rate is proportional to T_c^{18}, where T_c is the temperature in the center. Lowering the central temperature by 7% only would be sufficient, but this would directly lead to problems with the ^{7}Be neutrino fluxes whose temperature dependence is $\propto T_c^8$. Furthermore the lower-temperature solar models would meet problems also with other observations, in particular the solar oscillations.

Another attempt has been to reduce the relative abundance of heavy elements (low Z). This would reduce κ and thus the temperature gradient. However, this does not seem a plausible way for solar-type stars in the first place, and the solar oscillation observations contradict this suggetion as well.

One more suggestion was a rapidly rotating core which would lower the thermal pressure. Also this is inconsistent with oscillation results that indicate that the core rotates at nearly the same angular speed as the surface.

Furthermore, a strong internal magnetic field could lower the central temperature but it should be very intense to have a significant, say 10%, contribution to the pressure of the core (**exercise**: calculate the required magnetic flux density). The origin of the field would need to be in the primordial cloud but its life-time against Ohmic dissipation would have been much less than the age of the Sun.

The fact that also the gallium experiments fall a factor of 2 short of the predicted neutrino flux indicates that the lowering of the temperature and shifting the neutrino peak to lower energies is not the right solution. Since the turn of the millenium, it has become evident that the solution really is in the **physics of neutrinos**. The neutrinos appear in three different flavors ($\nu_{\rm e}$, ν_μ, ν_τ). Certain non-standard **elementary particle** models predict that the neutrinos have finite masses. In such a case the appearance of the neutrinos can oscillate between the flavors. The nuclear reactions in the Sun produce electron neutrinos ($\nu_{\rm e}$) only and the predicted fluxes are for electron neutrinos. If a large enough fraction of solar neutrinos would transform to other neutrino flavors before reaching the Earth, this would be an elegant solution to the whole problem.

In 1998 the Superkamiokande observations indicated oscillations between muon neutrinos (ν_μ), produced by cosmic rays in the atmosphere of the Earth, and tau neutrinos (ν_τ). This did not solve the problem of solar neutrinos directly but it gave strong evidence that neutrinos are not massless particles.

In the summer of 2001 the first results from the new heavy water (D_2O) detector at the Sudbury Neutrino Observatory (SNO) indicated that the solar neutrino problem really was coming to its final solution. The point is that this 8-kiloton water detector with 1 kiloton heavy water at its heart is able to distinguish between the total neutrino flux and the $\nu_{\rm e}$ flux.

Ordinary water detectors detect neutrinos through Čerenkov light from **elastic** $\nu - {\rm e}^-$ scattering if the recoil energy of the electron is at least 5 MeV. All neutrino flavors can scatter in such a way but the combined cross section of $\nu_\mu - {\rm e}^-$ and $\nu_\tau - {\rm e}^-$ reactions is a factor of 7 smaller than the cross section of $\nu_{\rm e} - {\rm e}^-$ scattering. Thus in the ordinary water detectors the other neutron flavors are practically hidden in the observational error margin. (If

all flavors of neutrinos would have the same flux, the observed flux would be $\frac{1}{3}(1+\frac{2}{7}) \approx 43\%$ of the prediction based on a flux of electron neutrinos alone.) However, in heavy water the deuteron can be broken in two different ways by neutrinos. The first is the **quasi-elastic** process

$$\nu + \mathrm{d} \rightarrow \nu + \mathrm{p} + \mathrm{n}\,, \tag{2.62}$$

where the neutrino can be of **any flavor**. This leaves a neutron whose subsequent capture can be recorded by the release of an identifiable γ. The second process is **inelastic** and involves ν_e only

$$\nu_\mathrm{e} + \mathrm{d} \rightarrow \mathrm{e}^- + \mathrm{p} + \mathrm{p}\,. \tag{2.63}$$

This is an inverse-β-decay which can again be observed through Čerenkov radiation.

The first results reported in 2001 did not yet actually utilize reaction (2.62) as the neutrons could not yet be identified reliably enough. However, the rate of ν_e-only reaction (2.63) has been compared to the very large data base from Superkamiokande including all neutron flavors. This indicates that indeed the difference in the fluxes corresponds to the combined effect of ν_μ and ν_τ.

The efficiency to capture the neutrons of process (2.62) was increased by adding 2 tons of NaCl (table salt) to the D_2O part of the detector. The successful results were published in September 2003 and refined in 2008 (see `http://www.sno.phy.queensu.ca/`).

The very good consistency of the standard solar model with solar oscillation observations and the failure of the non-standard solar models to solve the neutrino problem together with the new neutrino observations indicate that "the solar neutrino problem" has been solved and the solution lies in the properties of the neutrinos. This is an excellent illustration of the strength of the physical method. Not only is the answer in our reach but also the failed attempts to find the solution within the solar models have contributed enormously to our detailed knowledge of the interior of the Sun. Had the problem never been there, we would most likely have been content with much less detailed models of the Sun itself.

Chapter 3

Solar atmosphere

In this chapter we discuss the average structure of the visible parts of the Sun, the photosphere and chromosphere. Solar activity and processes related to solar magnetism will be discussed later.

3.1 Radiative transfer–LTE

Let us consider the radiative transfer using the (frequency-dependent) optical depth as the independent coordinate

$$d\tau_\nu = -\kappa_\nu \rho \, dr \, . \tag{3.1}$$

Note that in the previous chapter we introduced the **specific** absorption coefficient κ_ν. The **volume absorption coefficient** $k_\nu = \kappa_\nu \rho$ is also often called simply the absorption coefficient or the **extinction coefficient**.

The equation of radiative transfer is given as

$$\mu \frac{dI_\nu}{d\tau_\nu} = I_\nu - S_\nu \, , \tag{3.2}$$

where $\mu = \cos\theta$ and S_ν is the source function. The equation can be solved by multiplying it first by $\mu^{-1} \exp(-\tau_\nu/\mu)$ and integrating

$$\begin{aligned} I_\nu(\tau_\nu, \mu) &= I_\nu(\tau_{0\nu}, \mu) \exp(-(\tau_{0\nu} - \tau_\nu)/\mu) \\ &\quad + \frac{1}{\mu} \int_{\tau_\nu}^{\tau_{0\nu}} S(\tau'_\nu) \exp(-(\tau'_\nu - \tau_\nu)/\mu) d\tau'_\nu \, , \end{aligned} \tag{3.3}$$

where $\tau_{0\nu}$ is the optical depth at some reference level. At the observation point $\tau_\nu = 0$ and deep inside the Sun $\tau_{0\nu} = \infty$. Integrating from the

observation point to the solar interior we get the total emergent intensity

$$I_\nu(0,\mu) = \frac{1}{\mu}\int_0^\infty S(\tau_\nu)\exp(-\tau_\nu/\mu)d\tau_\nu\,. \tag{3.4}$$

From a known source function we can thus calculate the intensity. On the other hand, by inverting (3.4) we can derive the source function based on observed intensity at different frequencies. However, before doing this we must discuss the assumed thermodynamic state of the solar atmosphere.

Thermodynamic equilibrium is the simplest thermodynamic state. It is described by one single temperature T everywhere. The particles have a Maxwellian distribution and the states of ionization and excitation are given by the Saha and Boltzmann equations for the same temperature. The radiation field is homogeneous and the isotropic black-body radiation is given by the Kirchhoff-Planck function $B_\nu(T)$. Because there is no temperature gradient, this is not a realistic description anywhere on the Sun.

Local thermodynamic equilibrium (LTE) is a more useful concept. It means that within a finite region a single temperature can be used to describe the particle temperature, population of atomic states and the local ratio between emission and absorption. An important assumption is that the source function in the equation of radiative transfer is $S_\nu = B_\nu(T)$. In the solar atmosphere LTE is a useful approximation for the continuum radiation in the visible and infrared wavelengths and for the weak spectral lines as well as for the wings of most spectral lines. Strong lines and cores of several weak lines, however, depart from LTE. The departures are due to non-thermal distribution of atomic levels while the electrons still maintain their Maxwellian distribution at a single electron temperature T_e. The difference is due to the fact that the atomic levels are thermalized by radiative transitions which are rare in rarefied gases, whereas the electron distribution is thermalized by electron collisions.

To calculate the line absorption coefficient in LTE we limit the discussion Doppler and collisional line broadening (Eqs. 2.36 and 2.37). The line profile is the **convolution** of these

$$\phi(\nu) = \frac{\gamma}{\sqrt{\pi}\Delta\nu_\mathrm{D}}\int_{-\infty}^{\infty}\frac{\exp(-(\nu-\nu')^2/\Delta\nu_\mathrm{D}^2)}{(2\pi(\nu'-\nu_0))^2+\gamma^2/4}\,d\nu'\,. \tag{3.5}$$

With suitable change of variables this can be expressed in terms of the **Voigt function**

$$H(a,v) = \frac{a}{\pi}\int_{-\infty}^{\infty}\frac{\exp(-y^2)}{(v-y)^2+a^2}\,dy \tag{3.6}$$

as

$$\phi(\nu) = \frac{1}{\sqrt{\pi}\Delta\nu_\mathrm{D}}H(a,v)\,. \tag{3.7}$$

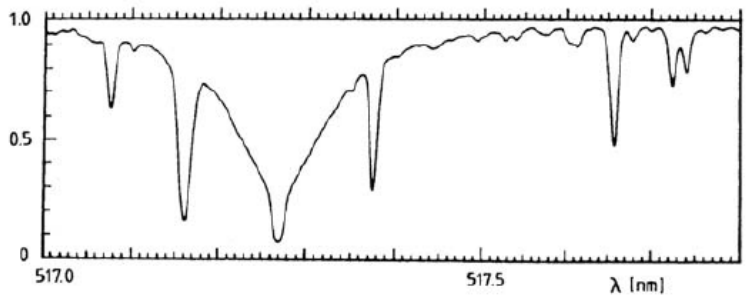

Figure 3.1: The Mg b_2 line. The Doppler profile dominates near the line center and the wings are shaped according to the collisional effect. This line is, in fact, so strong that the LTE approach is no more quite accurate.

The cross section per atom (corresponding to Eq. 2.35) is

$$\sigma = \frac{e^2 f}{4\epsilon_0 m_\mathrm{e} c\sqrt{\pi}\Delta\nu_\mathrm{D}} H(a,v)\,. \tag{3.8}$$

Exercise

Identify the changes of variables above, show that the integral over v of $H(a,v)$ is $\sqrt{\pi}$ and that for small a the Voigt function is normalized with respect to the line center, i.e.,

$$H(a,0) = 1 + \mathcal{O}(a^2)\,. \tag{3.9}$$

Finally to obtain the line absorption coefficient σ must be multiplied by the number of absorbing particles per unit mass and the result is

$$\kappa_l = \frac{\sigma}{\mu m_\mathrm{H}} \frac{n_i}{\sum n_i} \frac{n_{ij}}{n_i} \frac{n_{ijk}}{n_{ij}}\,, \tag{3.10}$$

where i denotes the particle species, j the ionization state, and k the exitation state. The last two fractions are determined from Saha and Boltzmann equations. The actual calculation of the emergent line intensity requires furthermore knowledge of the oscillator strength f and damping constant γ.

Figure 3.1 is an example of an absorption line where both the Doppler effect and the collisional broadening are visible.

3.2 Radiative transfer–Statistical equilibrium

The deviations from the LTE increase with increasing altitude from the solar surface. The electrons are still Maxwellian and the temperature in this

discussion is the electron temperature ($T = T_e$).

Consider now, for simplicity, an atom with one electron for which there are N bound states and a continuum state (state of ionization).

3.2.1 Line radiation

Denote the lower atomic level by L and the upper level by U. The energy between these levels is $h\nu = E_U - E_L$. In this simple case there are three different radiative transitions:

Spontaneous emission

Let there be n_U atoms per unit volume in the upper state. Let A_{UL} be the constant of proportionality for a spontaneous U $\rightarrow$ L transition. Thus $n_U A_{UL}$ photons are spontaneously emitted per unit time and unit volume (thus $[A_{UL}] = \mathrm{s}^{-1}$). The atomic levels have a finite width and the atoms are in random motion. Thus the emitted photons have the frequency distribution $\chi(\nu)$ around the center frequency ν_0. The photons are emitted with equal probability to all directions. Consequently, the number of emissions per unit time, volume, frequency interval, and solid angle is

$$n_U A_{UL} \chi(\nu)/4\pi \,. \tag{3.11}$$

A_{UL} is **the Einstein coefficient for spontaneous emission**. Typically $A_{UL} \approx 10^8\,\mathrm{s}^{-1}$. It can be interpreted as the inverse life-time of the upper state against spontaneous emission.

Induced (or stimulated) emission

If an atom is exposed to radiation of frequency $\nu = (E_U - E_L)/h$, it may emit a photon. The number of emitted photons is proportional to n_U and to the intensity of the incident radiation, I_ν. Denote the line width in this case by $\psi(\nu)$ and denote the constant of proportionality by $B_{UL}/4\pi$. The factor 4π is here for the convenience: Because the induced emission is in the same direction as the inducing radiation it does not introduce this factor as was the case with the spontaneous emission. The number of induced emissions per unit time, volume, frequency interval, and solid angle is

$$n_U B_{UL} I_\nu \psi(\nu)/4\pi \,. \tag{3.12}$$

B_{UL} is **the Einstein coefficient for induced emission**. Note that now $B_{UL} I_\nu$ has the same dimension as A_{UL}.

Absorption (radiative excitation)

This is the inverse of induced emission. The number of absorbed photons per unit time, volume, frequency interval, and solid angle is

$$n_{\rm L} B_{\rm LU} I_\nu \phi(\nu)/4\pi \,. \tag{3.13}$$

$B_{\rm LU}$ is **the Einstein coefficient for radiative excitation**.

The Einstein coefficients are **atomic constants** and they can be determined by quantum mechanical calculations and laboratory experiments.

In the special case of thermodynamic equilibrium the upper and lower states are populated according to the Boltzmann distribution

$$\frac{n_{\rm U}}{n_{\rm L}} = \frac{g_{\rm U}}{g_{\rm L}}\, e^{-h\nu/k_{\rm B}T} \tag{3.14}$$

and $I_\nu = B_\nu(T)$. Furthermore, the so-called **principle of detailed balance** is valid, i.e., the number of upward transitions is equal to downward transitions. Under these circumstances it is straightforward to show

$$g_{\rm U} B_{\rm UL} = g_{\rm L} B_{\rm LU} \tag{3.15}$$

$$\frac{g_{\rm U}}{g_{\rm L}} A_{\rm UL} = \frac{2h\nu^3}{c^2} B_{\rm LU} \,, \tag{3.16}$$

where $g_{\rm U}$ and $g_{\rm L}$ are the statistical weights (degeneracies) of the upper and lower states. As the Einstein coefficients are atomic constants these relations are generally valid and do not depend on the assumption of thermodynamic equilibrium.

Exercise

Prove the above relations between $B_{\rm UL}$, $B_{\rm LU}$, and $A_{\rm UL}$ (although Einstein was the first to do this, it is not too difficult for you!).

3.2.2 Continuum radiation

The continuum processes are (in this simple example) **photoionization** and **radiative recombination**. Photoionization occurs when the energy of an incident photon exceeds the ionization energy of the atom. For detailed calculations one has to determine the photoionization cross section for each atomic level, $\alpha_j(\nu)$. The number of photoionizations from level j per unit time, volume, frequency interval, and solid angle is

$$n_j \alpha_j(\nu) I_\nu / h\nu \,. \tag{3.17}$$

The inverse process of photoionization is radiative recombination, to which there are two contributions. One is an induced process relative to incident radiation, $\beta_j(\nu)I_\nu$, and the second a spontaneous process described by $\gamma_j(\nu)$. The number of recombinations to level j per unit time, volume, frequency interval, and solid angle is

$$n_{\rm C}(\gamma_j(\nu) + \beta_j(\nu)I_\nu)/h\nu\,, \tag{3.18}$$

where $n_{\rm C}$ is the number density of atoms in the continuum state.

3.2.3 Collisions

Collisional transitions have no direct influence on the radiation. However, they change the populations of different atomic levels and must be taken into account in statistical calculations.

In LTE the collisional bound-bound transition coefficients ($C_{\rm UL}$ and $C_{\rm LU}$) are related to each other as

$$C_{\rm UL} = \frac{g_{\rm L}}{g_{\rm U}} \exp\left(\frac{E_{\rm U} - E_{\rm L}}{k_{\rm B}T}\right) C_{\rm LU}\,. \tag{3.19}$$

In the same way the the transition coefficients to and from the continuum state ($C_{{\rm C}j}$ and $C_{j{\rm C}}$) can be shown to be related in LTE as

$$C_{{\rm C}j} = \left(\frac{h^2}{2\pi m_{\rm e} k_{\rm B} T}\right)^{3/2} \frac{n_{\rm e} g_j}{2u_{\rm C}} \exp\left(\frac{E_{\rm U} - E_{\rm L}}{k_{\rm B}T}\right) C_{j{\rm C}}\,, \tag{3.20}$$

where $n_{\rm e}$ is the number density of free electrons, g_j the statistical weight of the state j, and $u_{\rm C}$ the partition function of the ionized state.

These coefficients depend on atomic processes **and** particle distributions. Thus the assumption of LTE is essential here.

3.2.4 The source function

Assuming that we know the various transition coefficients (from theoretical and experimental work in atomic physics) the remaining task is to combine all above contributions to the equation of radiative transfer, which we this time write as a function of the radial co-ordinate

$$\mu \frac{dI_\nu}{dr} = -\kappa\rho I_\nu + \varepsilon\,. \tag{3.21}$$

The absorption coefficient κ and the emission term ε are divided into the line and continuous parts ($\kappa = \kappa_l + \kappa_{\rm C}$; $\varepsilon = \varepsilon_l + \varepsilon_{\rm C}$).

In the line radiation each photon carries an energy $h\nu$ and thus the RHS of the equation of radiative transfer reads

$$\frac{h\nu}{4\pi}[n_{\rm U}(A_{\rm UL}\chi + B_{\rm UL}I_\nu\psi) - n_{\rm L}B_{\rm LU}I_\nu\phi]\,. \tag{3.22}$$

Defining the source function as $S_l = \varepsilon/\rho\kappa_l$ we find

$$\kappa_l = \frac{h\nu}{4\pi\rho}(n_{\rm L}B_{\rm LU}\phi - n_{\rm U}B_{\rm UL}\psi) \tag{3.23}$$

$$S_l = \frac{n_{\rm U}A_{\rm UL}\chi}{n_{\rm L}B_{\rm LU}\phi - n_{\rm U}B_{\rm UL}\psi}\,. \tag{3.24}$$

The corresponding functions for the continuum states $(\kappa_{\rm C}, S_{\rm C})$ are derived from the transition coefficients to and from the level j $(\kappa_{{\rm C}j}, S_{{\rm C}j})$.

Finally the combined source function is the weighted sum

$$S = \frac{\kappa_l S_l + \kappa_{\rm C} S_{\rm C}}{\kappa_l + \kappa_{\rm C}}\,. \tag{3.25}$$

To evaluate the source function we need to know the population numbers $n_{\rm U}$, $n_{\rm L}$, and n_j. In reality there are several possible transition levels and often more than one electron to be considered. Thus the actual calculation of S is a formidable task. Note also that S depends on the intensity through the intensity-dependent transitions. This makes the equation of radiative transfer generally non-linear, which means that the solutions require iterative methods.

3.3 Atmospheric models

The solar atmospheric models describe the altitude-dependence of T and P (often as functions of τ) in the same way as the solar interior models. In this section we neglect all horizontal variations. This is an assumption that is not valid at higher altitudes.

3.3.1 Limb darkening

Figure 3.2 illustrates a remarkable feature of the Sun: The limb of the Sun is much darker than the center. Thus it looks like the center would be hotter than the limb which does not sound physical.

We have already encountered a primitive model for the atmospheric temperature in section 2.3.2, the Eddington approximation. It gave the temperature as a function of τ as

$$T^4 = \frac{L_s}{4\pi\sigma r_s^2}\,(3\tau/4 + 1/2)\,. \tag{3.26}$$

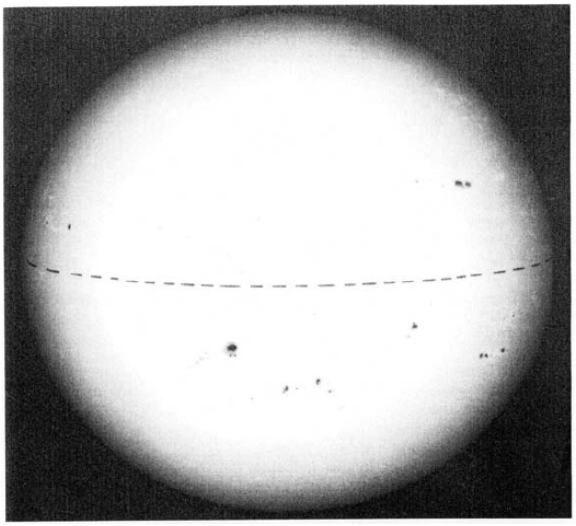

Figure 3.2: The limb of the Sun is much darker than the center.

This approximation predicts limb darkening. The intensity at the observation point ($\tau = 0$) varies as a function of $\mu = \cos\theta$ as

$$\frac{I(0,\mu)}{I(0,1)} = \frac{2+3\mu}{5} \ . \tag{3.27}$$

When we look toward the center of the Sun, we see into deeper layers (closer to the surface) whereas closer to the limb we see into shallower layers. Thus we can expect that the temperature decreases above the surface. Near the disc center ($\mu \approx 1$) the Eddington approximation is quite good but toward the limb the darkening is stronger, in particular the blue edge of the visible spectrum. Thus the temperature structure is more complicated.

Exercise

Show equation (3.27) for the limb darkening in the Eddington approximation.

3.3.2 Examples of model results

In this section we consider some results of solar atmospheric model calculations. In the following we use the wavelength instead of frequency ($T = T(\tau_\lambda)$ and $\kappa_\lambda = \kappa(\lambda)$).

Most of the solar radiation is in the visible and IR continuum. It is convenient to relate the quantities to certain fixed wavelength in the continuum,

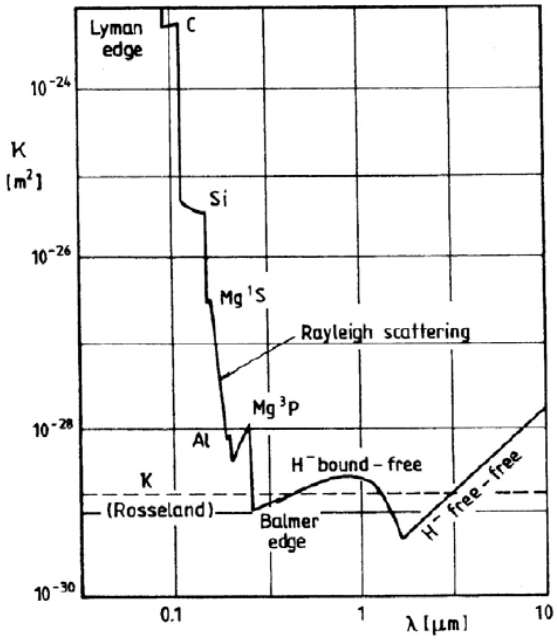

Figure 3.3: Continuum absorption in the solar atmosphere at $\tau_{500} = 0.1$.

e.g., $\lambda = 500$ nm. Figure 3.3 shows the continuum absorption coefficient at the optical depth $\tau_{500} = 0.1$.

The atmospheric continuum absorption at visible and near IR wavelengths is relatively small and mostly due to transitions of the H^- ions. In UV the sharp edges correspond to the ionization energies of various elements. The Lyman edge in figure 3.3 is at 91.2 nm corresponding to the photon energy 13.6 eV, i.e., the energy required to ionize a hydrogen atom in its ground state ($n = 1$).

Figure 3.4 shows the temperature profile translated to a function of altitude. The temperature minimum is at 500 km. The region of decreasing temperature is the photosphere. Above it the **chromosphere** extends up to about 2000 km. Chromosphere has got its name after the colorful flash it exhibits just at the beginning and end of a total solar eclipse. The most prominent color is the red Hα. At the upper end of the chromosphere the temperature begins to rise more rapidly. There is no strict definition of the upper limit of the chromosphere, it is often defined to be at the temperature of 25000 K. Above the chromosphere there is a thin **transition region** to coronal temperatures of the order of 10^6 K. Note that the higher up the model reaches the more it deviates from the LTE approximation. The chromosphere and corona will be discussed in more detail later.

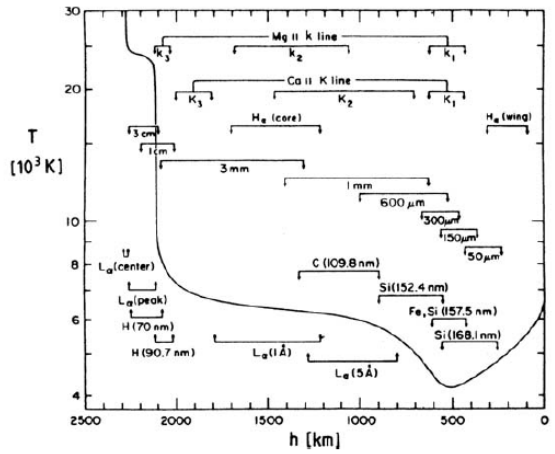

Figure 3.4: Temperature as a function of altitude from the solar surface according to extensive calculations in statistical equilibrium by Vernazza et al. (1981). Shown are also some of the altitude ranges from where some of the most important continua and lines originate.

3.4 The chemical composition of the Sun

The abundances of various elements in the Sun can be determined by careful analysis of the solar spectrum and model calculations.

The main constituent of the universe and the main cause of the continuum absorption in the solar atmosphere is hydrogen. Thus it is reasonable to express the other abundances relative to hydrogen. In the enclosed table the abundances are given in terms of

$$\log A = 12 + \log(n_i/n_{\rm H}) \tag{3.28}$$

i.e., the logarithmic abundance normalized to $n_{\rm H} = 10^{12}$ atoms per unit volume. Note that the abundances are given by numbers of particles not by weight.

Most of the abundances have been determined from the photospheric absorption lines. One of the main uncertainties in this procedure is the oscillation strength f in the calculation of absorption coefficients. For example, the abundance of iron increased by a factor of 10 in 1968 when it was realized that previously used oscillation strengths were a factor of 60 too large! Only a few years ago new measurements of transition probabilities of singly ionized lutetium (LuII) reduced lutetium abundances by a factor of 4, which brought the solar and meteorite values close to each other.

There has been problems also with the abundances of the light atoms lithium, beryllium, and boron in the Sun. Recent determinations have lifted

Table 3.1: Abundances of chemical elements in the Sun compared to abundances in meteorites.

Element	Photosphere	Meteorites	Element	Photosphere	Meteorites
1 H	12.00	–	42 Mo	1.92 ± 0.05	1.97 ± 0.02
2 He	10.93 ± 0.004	–	44 Ru	1.84 ± 0.07	1.83 ± 0.04
3 Li	1.10 ± 0.10	3.31 ± 0.04	45 Rh	1.12 ± 0.12	1.10 ± 0.04
4 Be	1.40 ± 0.09	1.42 ± 0.04	46 Pd	1.69 ± 0.04	1.70 ± 0.04
5 B	2.55 ± 0.30	2.79 ± 0.05	47 Ag	0.94 ± 0.25	1.24 ± 0.04
6 C	8.52 ± 0.06	–	48 Cd	1.77 ± 0.11	1.76 ± 0.04
7 N	7.92 ± 0.06	–	49 In	1.66 ± 0.15	0.82 ± 0.04
8 O	8.83 ± 0.06	–	50 Sn	2.0 ± 0.3	2.14 ± 0.04
9 F	4.56 ± 0.3	4.48 ± 0.06	51 Sb	1.0 ± 0.3	1.03 ± 0.07
10 Ne	8.08 ± 0.06	–	52 Te	–	2.24 ± 0.04
11 Na	6.33 ± 0.03	6.32 ± 0.02	53 I	–	1.51 ± 0.08
12 Mg	7.58 ± 0.05	7.58 ± 0.01	54 Xe	–	2.17 ± 0.08
13 Al	6.47 ± 0.07	6.49 ± 0.01	55 Cs	–	1.13 ± 0.02
14 Si	7.55 ± 0.05	7.56 ± 0.01	56 Ba	2.13 ± 0.05	2.22 ± 0.02
15 P	5.45 ± 0.04	5.56 ± 0.06	57 La	1.17 ± 0.07	1.22 ± 0.02
16 S	7.33 ± 0.11	7.20 ± 0.06	58 Ce	1.58 ± 0.09	1.63 ± 0.02
17 Cl	5.5 ± 0.3	5.28 ± 0.06	59 Pr	0.71 ± 0.08	0.80 ± 0.02
18 Ar	6.40 ± 0.06	–	60 Nd	1.50 ± 0.06	1.49 ± 0.02
19 K	5.12 ± 0.13	5.13 ± 0.02	62 Sm	1.01 ± 0.06	0.98 ± 0.02
20 Ca	6.36 ± 0.02	6.35 ± 0.01	63 Eu	0.51 ± 0.08	0.55 ± 0.02
21 Sc	3.17 ± 0.10	3.10 ± 0.01	64 Gd	1.12 ± 0.04	1.09 ± 0.02
22 Ti	5.02 ± 0.06	4.94 ± 0.02	65 Tb	-0.1 ± 0.3	0.35 ± 0.02
23 V	4.00 ± 0.02	4.02 ± 0.02	66 Dy	1.14 ± 0.08	1.17 ± 0.02
24 Cr	5.67 ± 0.03	5.69 ± 0.01	67 Ho	0.26 ± 0.16	0.51 ± 0.02
25 Mn	5.39 ± 0.03	5.53 ± 0.01	68 Er	0.93 ± 0.06	0.97 ± 0.02
26 Fe	7.50 ± 0.05	7.50 ± 0.01	69 Tm	0.00 ± 0.15	0.15 ± 0.02
27 Co	4.92 ± 0.04	4.91 ± 0.01	70 Yb	1.08 ± 0.15	0.96 ± 0.02
28 Ni	6.25 ± 0.04	6.25 ± 0.01	71 Lu	0.06 ± 0.10	0.13 ± 0.02
29 Cu	4.21 ± 0.04	4.29 ± 0.04	72 Hf	0.88 ± 0.08	0.75 ± 0.02
30 Zn	4.60 ± 0.08	4.67 ± 0.04	73 Ta	–	-0.13 ± 0.02
31 Ga	2.88 ± 0.10	3.13 ± 0.02	74 W	1.11 ± 0.15	0.69 ± 0.03
32 Ge	3.41 ± 0.14	3.63 ± 0.04	75 Re	–	0.28 ± 0.03
33 As	–	2.37 ± 0.02	76 Os	1.45 ± 0.10	1.39 ± 0.02
34 Se	–	3.41 ± 0.03	77 Ir	1.35 ± 0.10	1.37 ± 0.02
35 Br	–	2.63 ± 0.04	78 Pt	1.8 ± 0.3	1.69 ± 0.04
36 Kr	–	3.31 ± 0.08	79 Au	1.01 ± 0.15	0.85 ± 0.04
37 Rb	2.60 ± 0.15	2.41 ± 0.02	80 Hg	–	1.13 ± 0.08
38 Sr	2.97 ± 0.07	2.92 ± 0.02	81 Tl	0.9 ± 0.2	0.83 ± 0.04
39 Y	2.24 ± 0.03	2.23 ± 0.02	82 Pb	1.95 ± 0.08	2.06 ± 0.04
40 Zr	2.60 ± 0.02	2.61 ± 0.02	83 Bi	–	0.71 ± 0.04
41 Nb	1.42 ± 0.06	1.40 ± 0.02	90 Th	–	0.09 ± 0.02
			92 U	< -0.47	-0.50 ± 0.04

the beryllium and boron abundances in the Sun to almost as high as elsewhere in the solar system, but the lithium depletion seems to remain. These elements can be destroyed by nuclear reaction with protons. The reactions are:

$$^6\mathrm{Li}(\mathrm{p},^3\mathrm{He})\alpha;\ ^9\mathrm{Be}(\mathrm{p},\alpha)^6\mathrm{Li};\ ^{10}\mathrm{B}(\mathrm{p},\alpha)^7\mathrm{Be};\ ^{11}\mathrm{B}(\mathrm{p},\gamma)3\alpha$$

The lithium and beryllium burning reaction rates become significant at temperatures of 2.5×10^6 K and 3×10^6 K, respectively. The low lithium abundance and the "normal" beryllium abundance in the photosphere can be used as an independent means to locate the bottom of the convective layer. It should be below the layer where $T = 2.5 \times 10^6$ K to allow the lithium burning but above the layer where $T = 3 \times 10^6$ K not to destroy beryllium. This yields a depth of about 2×10^8 m, which is consistent with results of helioseismology.

The lithium problem is not quite fully understood yet. Model calculations indicate that before the young Sun reached the main sequence, the bottom of the convective layer may have reached a depth where $T = 3.5 \times 10^6$ K and all lithium should have disappeared. On the other hand, several other Sun-like stars seem to have higher or even "normal" lithium abundances.

Note also that if the helium abundance is measured from the spectral lines, it cannot be determined much more accurately than $\log A = 10.8 \pm 0.2$, which is not very good for such a common element. The reason is that the important helium lines are in the UV and IR parts of the spectrum and they are produced in the chromosphere and corona under conditions that are far from LTE. The (interior) solar models calibrated with helioseismic observations give a much more accurate helium abundance $\log A = 10.93 \pm 0.004$.

Chapter 4

Solar Oscillations

As we saw in the context of the solar neutrino problem in Chapter 2, the analysis of solar oscillations has become a decisive tool in validation of internal solar models. The applicability of this method, known as **helioseismology**, extends much further as we shall see.

4.1 Observations of solar oscillations

Global oscillations of the solar atmosphere at the period of about 5 min (3 mHz) were found in 1960 using spectroheliograph recordings. When the red and blue wings of a spectral line are measured simultaneously the relative intensity of the two observations indicate the motion of the observed region toward or away from the observer.

In 1975 the oscillations were found to have a spectrum of discrete frequencies. This opened the way for the new field of helioseismology to develop. Presently it is the most important method in studies of the interior of the Sun. Analysis of very low frequencies require very long continuous time series, which is possible by using global ground-based networks of solar observatories or the SOHO or the SDO spacecraft. SOHO carries three instruments for observations of solar oscillations, each utilizing different techniques and wavelengths: VIRGO is an irradiance instrument, MDI a Michelson Doppler imager, and GOLF an instrument observing global oscillations of the integrated disc. SDO has the HMI instrument that can be used for helioseismology. On ground several networks using complementary observations have been established. GONG employs Michelson interferometers to measure the Doppler shift of a nickel line and TON uses telescopes with a narrow-band filter to measure intensity oscillations of the CaII K-line. BISON and IRIS measure the integrated sunlight with Doppler shift mea-

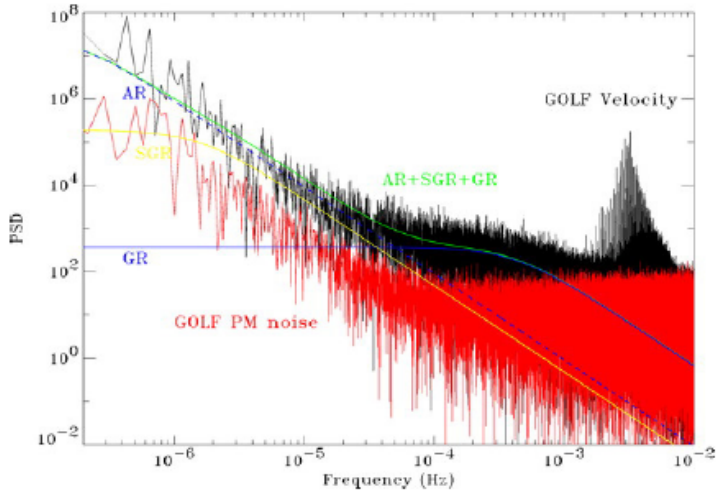

Figure 4.1: The power spectral density of solar irradiance as observed by GOLF. The peak at $\nu \sim 3$ mHz consists of p modes and the signal above background at lower frequencies (10–1000 μHz) could be from the g modes.

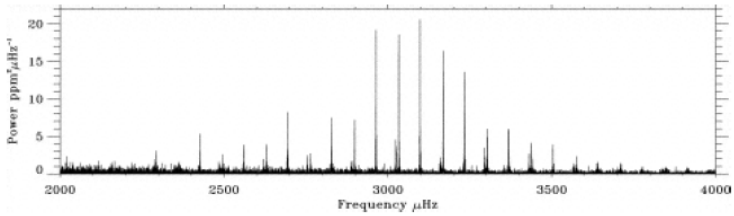

Figure 4.2: A p mode spectrum from VIRGO observations.

surements of a line of K at $\lambda = 769.9$ nm (BISON) and of Na at $\lambda = 589.6$ nm (IRIS).

Figure 4.1 is an example of power spectra of the irradiance observations by the GOLF instrument. The frequency interval around 3 mHz (5 min oscillation period) is illustrated more closely in Figure 4.2 as measured by VIRGO. These modes are called ***p* modes** because the restoring force of the oscillations is the pressure gradient. The lower-frequency ***g* modes** (or gravity waves) in Figure 4.1 are less clear. Their restoring force is the gravity and they are observed mostly in the solar atmosphere (and also in the terrestrial atmosphere). As will be shown in section 4.3 these modes can propagate in the radiation zone and in the solar atmosphere but not in the convection zone. Thus it is unclear if any signal from the internal g modes can be detected or not.

4.2 Spectral analysis

From elementary Fourier analysis we are familiar with the presentation of oscillatory data in terms of frequency ($\nu = \omega/2\pi$) and wavenumber ($k = 2\pi/\lambda$). Note that we use the word frequency (ν) for the oscillation frequency but in theoretical formulations the angular frequency (ω) is more convenient.

The length of the time series T determines the maximum frequency resolution as $\Delta\omega = 2\pi/T$ (or $\Delta\nu = 1/T$). Thus if we want, say, 0.1 μHz resolution, we need a time series of 116 days! This also determines the lowest frequency that is possible to determine. For example, the solar rotation of 27 days corresponds to a frequency of about 0.4 μHz. The solar rotation is not the only reason why such high frequency resolution is required. It has been found that the solar activity shifts the frequencies by $0.45 \pm 0.04\,\mu$Hz, the higher frequencies observed during the solar minimum.

The high-frequency limit is given by the time resolution Δt of the observation. The highest analyzable frequency is the **Nyquist frequency** $\omega_{\rm Ny} = \pi/\Delta t$, i.e., data must be sampled at the rate that is twice the highest frequency (ν) to be investigated. Furthermore, the input data must be carefully filtered near the Nyquist frequency to avoid **aliasing**, i.e., the folding of higher frequencies to the investigated frequency interval.

Same rules apply to spatial sampling (to determine the wave number). In summary

$$\Delta\omega = 2\pi/T \le \omega \le \pi/\Delta t \tag{4.1}$$

$$\Delta k_x = 2\pi/L_x \le k_x \le \pi/\Delta x\,. \tag{4.2}$$

Consider now the vertical velocity v measured at the position (x, y) on the solar disc. Its **Fourier transform** f is defined by

$$v(x, y, t) = \int f(k_x, k_y, \omega) \exp[i(k_x x + k_y y + \omega t)]\, dk_x\, dk_y\, d\omega\,. \tag{4.3}$$

The **power spectrum** is defined by

$$P(k_x, k_y, \omega) = f f^*\,, \tag{4.4}$$

where the asterisk denotes the complex conjugate. Note that in figures 4.1 and 4.2 the power spectra are computed in the frequency domain only.

Figure 4.3 illustrates how the power of solar oscillations is distributed in the $(\omega, k_{\rm h})$ plane as observed from the ground in the 1970's ($k_{\rm h} = (k_x^2 + k_y^2)^{1/2}$ is the horizontal wave number). The contours of equal power form several ridges, each corresponding to a fixed number of nodes in the radial direction. The number of the nodes n for a given rigde is called the **radial mode**

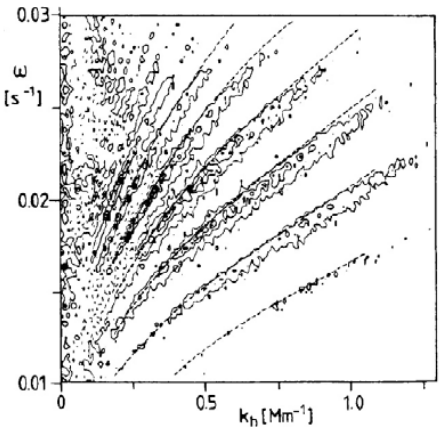

Figure 4.3: The p and f mode ridges analyzed from early ground based observations of solar oscillations. The dashed lines are theoretical predictions.

number. As the restoring force for these oscillations is ∇P, these ridges are called p mode ridges, except for the lowest ridge that is called the f mode (f for fundamental). The f mode is a non-compressional surface wave similar to the surface wave on deep water. It has no radial nodes ($n = 0$).

The Cartesian coordinates are useful only within small areas on the solar disc whereas the global mode structure must be analyzed using spherical coordinates. The velocity is expanded in terms of spherical surface harmonics as

$$v(\theta, \phi, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm}(t) Y_l^m(\theta, \phi)\,, \tag{4.5}$$

where

$$Y_l^m(\theta, \phi) = P_l^{|m|}(\theta) \exp(\mathrm{i}m\phi) \tag{4.6}$$

and P_l^m are the associated Legendre functions. Now l is the total number of node circles on the sphere and m is the number of node circles through the poles.

At frequencies much higher than the solar rotation there is no preferred axis of symmetry nor physical poles. Thus the eigenfrequencies must not depend on m and we can consider the case $m = 0$. At the frequency resolution $\leq 1\,\mu$Hz the rotation introduces a splitting of the modes ($m \neq 0$) which is of importance when oscillations are used in the determination of the internal rotation rate of the Sun.

We will later see that the horizontal wave number k_h is related to l by $k_\mathrm{h} r_\odot = [l(l+1)]^{1/2}$. Instead of the (ω, k) representation the mode structure is usually given in the (l, ν) space. If $\hat{a}(\nu)$ is the Fourier transform of $a_{l0}(t)$,

the power is

$$P(l,\nu) = \hat{a}\hat{a}^* \,. \tag{4.7}$$

An example of such presentation is in Figure 4.4 which is from SOHO MDI observations.

Note that the radial node numbers (different ridges) are easily resolved at large l whereas the resolution of different l from each other is not that perfect.

The amplitudes of single modes are ≤ 30 cm/s and the smallest amplitudes, that are presently measurable, are of the order of 1 mm/s. The integrated amplitude of 5-min oscillations is 0.5–1 km/s. Taking into account randomly distributed phases of the modes and ever-changing spatial pattern of the oscillations the total amplitude is a result of 10^7 or more single modes.

4.3 Physics of solar oscillations

4.3.1 Basic equations

Assume that the Sun is in hydrostatic equilibrium. Consider perturbations which are fast enough to be assumed adiabatic. Then

$$\frac{\delta P}{P_0} = \Gamma_1 \frac{\delta \rho}{\rho_0} \,, \tag{4.8}$$

where P_0 and ρ_0 are functions of r, and δP and $\delta\rho$ are **Lagrangian** perturbations (i.e., perturbations of a fixed parcel of material). The adiabatic exponent

$$\Gamma_1 = \left(\frac{\partial \ln P}{\partial \ln \rho} \right)_S \tag{4.9}$$

is related to the **adiabatic sound speed** c_s by

$$c_\mathrm{s}^2 = \Gamma_1 \frac{P_0}{\rho_0} \,. \tag{4.10}$$

The sound speed also depends on r. We neglect the rotation effects assuming that $\Omega/\omega \simeq 10^{-4}$. This is a good approximation as long as we are not studying the details of the rotation itself. Furthermore, the oscillations are assumed linear. Nonlinear terms are of the order of v/c_s where v is the velocity amplitude, which is less than 1 m/s for each mode we are interested in. Nonlinearities might become a concern if we were interested in the propagation of these oscillations into the solar atmosphere, which is not the case here.

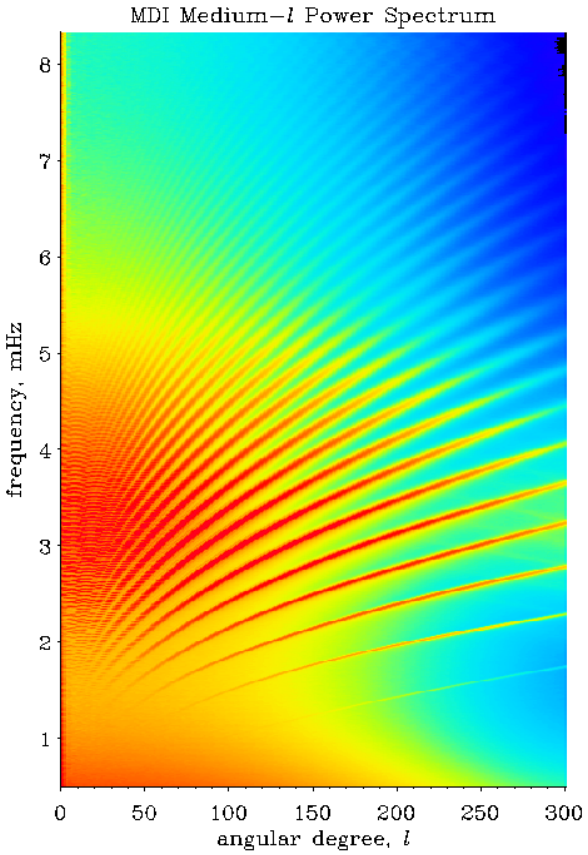

Figure 4.4: p and f mode ridges from 60-day observations by MDI.

Let $\boldsymbol{\xi}$ be the vectorial distance of a gas parcel from its equilibrium position and let the subscript 1 denote the **Eulerian** perturbations, i.e., perturbation at a fixed position in the Sun. The Lagrangian (δf) and Eulerian perturbations are related to each other by $\delta f = f_1 + \boldsymbol{\xi} \cdot \nabla f_0$. The linearized continuity and momentum equations read now as

$$\rho_1 + \nabla \cdot (\rho_0 \boldsymbol{\xi}) = 0 \tag{4.11}$$

$$\rho_0 \frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} + \nabla P_1 - \frac{\rho_1}{\rho_0} \nabla P_0 + \rho_0 \nabla \Phi_1 = 0\,, \tag{4.12}$$

where Φ_1 is the perturbation of the gravitational potential

$$\nabla^2 \Phi_1 = 4\pi G \rho_1\,. \tag{4.13}$$

Thus we have 6 equations (4.8, 4.11, 4.12, 4.13) for 6 unknowns: ρ_1, P_1, Φ_1, and the three components of $\boldsymbol{\xi}$.

Exercise

Derive equations (4.8), (4.11), and (4.12).

4.3.2 Spherical harmonic form

By taking curl of (4.12) we find

$$\mathbf{r} \cdot \nabla \times \frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = 0\,, \tag{4.14}$$

i.e., the curl of $\boldsymbol{\xi}$ has no vertical component. Writing $\nabla\times$ in spherical coordinates and recalling that we are looking for oscillating solutions, i.e., $\partial/\partial t = \mathrm{i}\omega \neq 0$ we find

$$\frac{\partial}{\partial \theta}(\sin\theta\, \xi_\phi) - \frac{\partial \xi_\theta}{\partial \phi} = 0\,. \tag{4.15}$$

Thus the horizontal components of the perturbation ξ_ϕ and ξ_θ can be expressed as derivatives of a single scalar function and we can write the components of $\boldsymbol{\xi}$ as

$$\boldsymbol{\xi} = e^{\mathrm{i}\omega t} \left[\xi_r(r), \xi_h(r) \frac{\partial}{\partial \theta}, \frac{\xi_h(r)}{\sin\theta} \frac{\partial}{\partial \phi} \right] Y_l^m(\theta, \phi)\,. \tag{4.16}$$

Similarly

$$(\rho_1, P_1, \Phi_1) = e^{\mathrm{i}\omega t} [\rho_1(r), P_1(r), \Phi_1(r)] Y_l^m(\theta, \phi)\,. \tag{4.17}$$

The spherical harmonic functions are governed by the differential equation

$$L^2 Y_l^m \equiv -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y_l^m}{\partial\theta} - \frac{1}{\sin^2\theta}\frac{\partial^2 Y_l^m}{\partial\phi^2} = l(l+1)Y_l^m\,. \tag{4.18}$$

Substituting these expressions in the equations of the previous section and eliminating ξ_h and ρ_1 we obtain

$$\frac{1}{r^2}\frac{d}{dr}(r^2\xi_r) - \frac{\xi_r g}{c_{\rm s}^2} + \frac{1}{\rho_0}\left(\frac{1}{c_{\rm s}^2} - \frac{l(l+1)}{r^2\omega^2}\right)P_1 - \frac{l(l+1)}{r^2\omega^2}\Phi_1 = 0 \tag{4.19}$$

$$\frac{1}{\rho_0}\left(\frac{d}{dr} + \frac{g}{c_{\rm s}^2}\right)P_1 - (\omega^2 - N^2)\xi_r + \frac{d\Phi_1}{dr} = 0 \tag{4.20}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi_1}{dr}\right) - \frac{l(l+1)}{r^2}\Phi_1 - \frac{4\pi G\rho_0}{g}N^2\xi_r - \frac{4\pi G}{c_{\rm s}^2}P_1 = 0\,, \tag{4.21}$$

where N is the **Brunt-Väisälä frequency**

$$N^2 = g\left(\frac{1}{\Gamma_1 P_0}\frac{dP_0}{dr} - \frac{1}{\rho_0}\frac{d\rho_0}{dr}\right)\,. \tag{4.22}$$

As the rotation and other possible sources of anisotropy are neglected, m does not appear in the equations. For each l there is $(2l+1)$-fold ($m = -l, \cdots, +l$) degeneracy of eigenfrequencies (cf. atomic physics).

Note that in the equation $L^2 Y_l^m = l(l+1)Y_l^m$ the operator L^2 is $-r^2$ times the angular part of the Laplacian. If we had used Cartesian coordinates ∇^2 would yield $-k_{\rm h}^2$. Thus $l(l+1) \simeq (k_{\rm h} r)^2$. Because the Cartesian coordinates are useful only on a small part of the solar disc, $k_{\rm h} r_\odot \gg 1$ and $l \simeq k_{\rm h} r_\odot$.

4.3.3 Local treatment

Before looking for the solutions of the differential equations we make a few simplifying approximations. First, it has been shown that neglecting the perturbations of the gravitational potential Φ_1 introduces only a small error, which is very small for large degrees l or very small frequencies. This was explained by Cowling in 1941 as follows:

The integral of the Poisson equation is

$$\Phi_1(\mathbf{r}) = -G\int\frac{\rho_1(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}'\,. \tag{4.23}$$

Now the contributions $\rho_1(\mathbf{r}')$ from different places in the Sun tend to cancel each other leaving a negligible net effect. Thus it is safe to drop the perturbations of the gravitational potential from the analysis. Note that the

effect of (unperturbed) gravitation is maintained in the analysis through the gravitational acceleration g.

Of course Cowling did not study the internal solar oscillations, they were found much later. The formalism is taken over from studies of gravitational atmospheres (stellar, solar, terrestrial) and pulsations of much more strongly oscillating stars (e.g., the Cepheids).

This approximation is called the **Cowling approximation** and we are left with two differential equations

$$\frac{1}{r^2}\frac{d}{dr}(r^2\xi_r) - \frac{\xi_r g}{c_\mathrm{s}^2} + \frac{1}{\rho_0}\left(\frac{1}{c_\mathrm{s}^2} - \frac{l(l+1)}{r^2\omega^2}\right) P_1 = 0 \tag{4.24}$$

$$\frac{1}{\rho_0}\left(\frac{d}{dr} + \frac{g}{c_\mathrm{s}^2}\right) P_1 - (\omega^2 - N^2)\xi_r = 0\,. \tag{4.25}$$

The next simplification is to assume that the parameters of the Sun do not vary much over one vertical wavelength of the perturbation, i.e., they are locally constant. This is an exact treatment in an isothermal atmosphere where g, c_s, N are constants and ρ_0 and P_0 have exponential height-dependence (barometric stratification). This is a useful first approximation also in the solar interior. Note that in this approximation the **scale height**

$$H \equiv -\frac{\rho_0}{d\rho_0/dr} = \left(\frac{g}{c_\mathrm{s}^2} + \frac{N^2}{g}\right)^{-1} \tag{4.26}$$

is also constant.

Denote the l-dependent term as

$$S_l^2 = \frac{l(l+1)}{r^2} c_\mathrm{s}^2 \tag{4.27}$$

and look for oscillatory solutions

$$\xi_r \sim \rho_0^{-1/2} \exp(\mathrm{i}k_r r) \tag{4.28}$$

$$P_1 \sim \rho_0^{1/2} \exp(\mathrm{i}k_r r)\,. \tag{4.29}$$

This leads to the dispersion equation

$$k_r^2 = \frac{\omega^2 - \omega_\mathrm{A}^2}{c_\mathrm{s}^2} + S_l^2\,\frac{N^2 - \omega^2}{c_\mathrm{s}^2\omega^2}\,, \tag{4.30}$$

where $\omega_\mathrm{A} = c_\mathrm{s}/2H$.

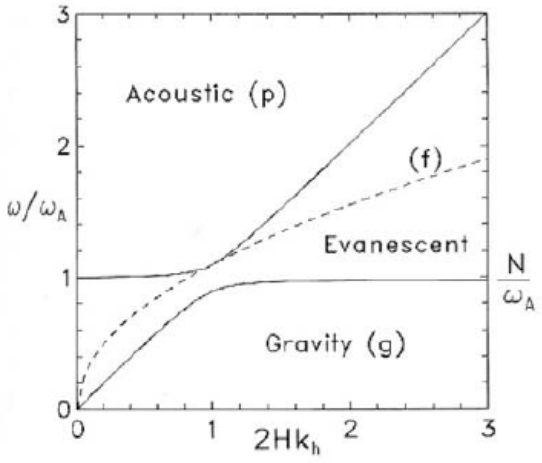

Figure 4.5: Curves of $k_r = 0$ in the (ω, k_h) space.

Exercise

Derive the above dispersion equation.

The solutions of equation (4.30) are oscillatory if $k_r^2 > 0$ (k_r real) and evanescent if $k_r^2 < 0$ (k_r imaginary). As shown in Figure 4.5 there are two regions of real solutions in the (ω, k_h) space. At large ω the dispersion equation simplifies to

$$\omega^2 = c_s^2(k_r^2 + k_\mathrm{h}^2)\,, \tag{4.31}$$

i.e., ordinary acoustic waves. The restoring force is the pressure gradient and the waves are called p modes. ω_A is called the **acoustic cut-off frequency**, below which there cannot be acoustic oscillations.

Exercise

Explain physically why acoustic modes cannot propagate at frequencies below ω_A.

The second oscillatory branch is found at low frequencies:

$$k_r^2 = N^2 \frac{k_\mathrm{h}^2}{k_r^2 + k_\mathrm{h}^2} = N^2 \sin^2\theta\,, \tag{4.32}$$

where θ is the angle between the propagation vector and the vertical direction. The restoring force is gravitation and the modes are called **internal gravity waves**, or ***g* modes**.

In the evanescent region between the p and g modes there is one particular solution with $\nabla \cdot \boldsymbol{\xi} = 0$ at the frequency

$$\omega = \sqrt{g k_{\mathrm{h}}}\,, \tag{4.33}$$

which is the dispersion equation for surface waves on deep water. This is the f mode.

It is a useful exercise to calculate ω_{A} and N for the solar atmosphere. It turns out that the observed p modes lie in the evanescent region. Thus they are not atmospheric waves but really oscillations in the Sun. The solar surface is one of the reflecting boundaries for these oscillations.

Because g, c_{s}, etc., are not really constants inside the Sun, the dividing lines of Figure 4.5 depend on the depth and the waves are refracted and reflected analogously to the electromagnetic wave propagation in a medium of slowly changing refractive index (e.g., radio wave propagation in the ionosphere).

The fact that the modes are trapped between two evanescent regions is the cause of the discrete oscillation frequencies. The high-frequency limit of equation (4.30) is

$$k_r^2 = (\omega^2 - S_l^2)/c_{\mathrm{s}}^2\,. \tag{4.34}$$

Thus the reflection level r_t is defined by $\omega^2 - S_l^2$ and we get an implicit equation

$$r_t = [l(l+1)]^{1/2} c_{\mathrm{s}}(r_t)/\omega\,. \tag{4.35}$$

Figure 4.6 shows the depth of total reflection p-mode waves at different frequencies as a function of l. The right-hand panel of the figure shows how the modes with small degree penetrate all the way into the core of the Sun whereas high-degree modes are trapped in the outer layers of the Sun.

The g modes are evanescent in the convection zone ($N^2 < 0$) and can propagate only in the radiative zone and in the solar atmosphere. g modes are expected to be excited deep in the Sun and it has been speculated that a small signal of their presence could somehow survive through the convection zone. Thus far they have not been positively identified in observational data.

4.4 Helioseismology

Helioseismology (and more generally stellar seismology of pulsating stars) is analogous to seismological studies of Earth's interior. There are two approaches to the use of oscillation data: direct modelling and inversion.

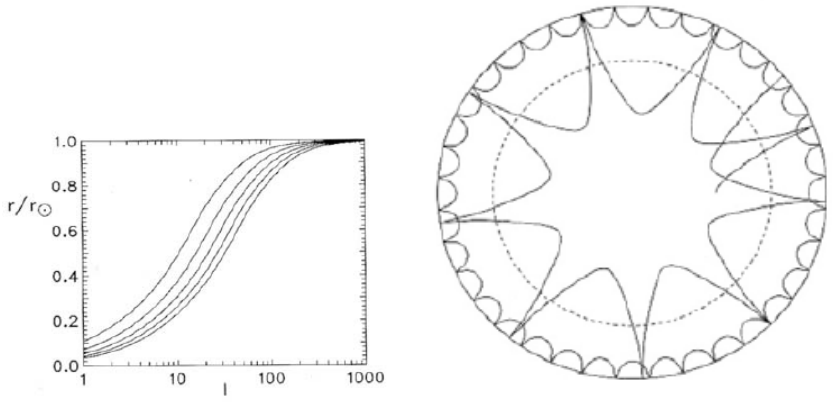

Figure 4.6: Left: Calculation of the depth of internal reflection of oscillations at frequencies 1.5 (top), 2.25, 3, 3.75, and 4.5 (bottom) mHz as a function of l in the standard solar model. Right: Acoustic ray paths for two different p-modes.

4.4.1 Direct modelling and inversion

In direct modelling the depth-dependent coefficients of the equations describing the oscillations are computed from a solar model. Then these equations are solved subject to appropriate boundary conditions. This results in eigenfunctions and eigenfrequencies, i.e., the values of ω for which non-zero solutions exist.

Now the solutions likely deviate somewhat from the observed frequencies. The direct approach is to make small corrections to those parameters of the solar model which have the largest uncertainties, e.g., corrections to the ideal gas law, the fractional abundance of helium and heavier elements (Y, Z) in different layers of the Sun, the depth of base of the convection zone, etc. Note that in these calculations the Cowling approximation is no more sufficiently accurate.

An example of results of direct modelling is the rejection of the low-Z models, which were suggested to solve the neutrino problem. In fact the oscillation data would fit better to higher Z than used in the standard models. Furthermore, internally mixed models, another proposed solution of the neutrino problem, have also lost their support after increased accuracy of the analysis oscillation data. In general, the details of the oscillation spectra lend support to the standard solar models.

The electrostatic correction to the equation of state may have seem to be of secondary importance when discussed in section 2.2.3. However, direct modelling of the oscillations in the 1980's showed that it is an essential part of the solar models and nowadays included in the standard solar models.

The iterative approach to the direct modelling is a primitive version of inversion. There are also **mathematical inversion methods** which are often more powerful but also more complicated tools than the direct modelling. The task is to find a function which appears in the integrand of a given definite integral. This is possible if the integrand depends, in addition to the unknown function, on a parameter and the integral itself is a function of the same parameter. We do not dwell deeper into the complexities of inversion theory here but discuss a few solar examples where recent SOHO data are playing an important role.

4.4.2 Speed of sound

Calculation of the speed of sound is an example where the mathematical inversion can be made analytically. This is not always the case.

In 1982 *Duvall* plotted the ratio $(n+\alpha)\pi/\omega$, where n is the order of the mode (the radial mode number, i.e., the number of the p mode ridge) and α is a constant, as a function of $\omega/k_{\rm h}$. In this representation all ridges collapse to a single curve for a properly chosen α, provided n has the right zero point (f mode). This result is known as **Duvall's law**.

This result follows from the fact that the total phase difference between the internal reflection point r_t (4.35) and the outer reflection point $r_\odot$ is

$$\triangle\psi = \int_{r_t}^{r_\odot} k_r \, dr \,, \tag{4.36}$$

For an eigensolution $\triangle\psi = \pi(n+\alpha)$. From the dispersion equation of the p modes for large ω we find

$$\pi(n+\alpha)/\omega = \int_{r_t}^{r_\odot} \left(\frac{1}{c_s^2} - \frac{l(l+1)}{r^2\omega^2} \right)^{1/2} dr \,. \tag{4.37}$$

Now we want to find c_s by inversion. As the integrand depends on the parameter $l(l+1)/\omega^2$ and r_t depends on the same parameter, the whole integral depends on it and thus the inversion becomes possible. (Note that this does not guarantee that it is analytical although it so happens in this special case.)

Substituting

$$u = l(l+1)/\omega^2 \tag{4.38}$$

$$\xi = (r/c_s)^2 \tag{4.39}$$

in (4.37) we can write the integral as

$$F(u) = \int_u^{\xi_\odot} (\xi - u)^{1/2} \, \frac{1}{r} \, \frac{dr}{d\xi} \, d\xi \,, \tag{4.40}$$

where $\xi_\odot = \xi(r_\odot)$. Now $F(u)$ is a function known from the oscillation spectra (lhs of (4.37)), u is the parameter on which both the integral and integrand depend, and ξ contains the sought function $c_\mathrm{s}(r)$. Differentiating with respect to u we get

$$-2\frac{dF}{du} = \int_u^{\xi_\odot} \frac{dG/d\xi}{(\xi - u)^{1/2}} \, d\xi \,, \tag{4.41}$$

where $G = \ln r$. This is an example of **Abel's integral equation**. Its solution, or inversion, is

$$G(\xi) - G(\xi_\odot) = -\frac{2}{\pi} \int_{\xi_\odot}^{\xi} \frac{dF/du}{(u - \xi)^{1/2}} \, du \,. \tag{4.42}$$

As $G = \ln r$ we finally get

$$r = r_\odot \exp\left(-\frac{2}{\pi} \int_{\xi_\odot}^{\xi} \frac{dF/du}{(u - \xi)^{1/2}} \, du \right) . \tag{4.43}$$

This is an implicit equation for $\xi(r)$ and thus for $c_\mathrm{s}(r)$. Now we can determine the sound of speed from the empirically known function $F(u)$ without knowing anything of the solar model. Of course the actual computation must be done numerically, but that is the case with all practical problems.

Figure 4.7 shows inversion results from SOHO/MDI observations. The quantity plotted is the deviation of the squared sound speed from one of the recent solar models (Christensen-Dalsgaard et al. 1996). Results from the GOLF instrument on SOHO agree with MDI results very well.

The relative errors in Figure 4.7 are remarkably small but still the deviations from the model say something significant of the solar interior. The largest deviations from the model are just below the base of the convection zone (larger c_s than in the model) and in the outer part of the core (smaller c_s than in the model). These are locations of significant variation of the mean molecular mass μ. Because the lower helium abundance increases the sound speed, and vice versa, it has been speculated that there would be a deficit (with respect to the model) of helium just below the convection zone. Similarly, there may be an overabundance of helium in the outer core, which would decrease the sound speed.

Due to the uncertainty of μ in the center of the Sun helioseismology cannot give the central temperature $T_\mathrm{c} \propto \mu c_\mathrm{s}^2$ as accurately as it gives the

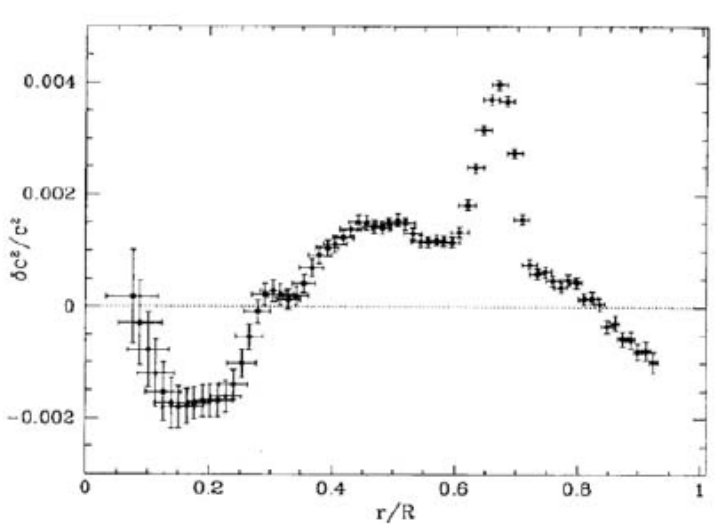

Figure 4.7: The relative difference between the sound speed squared as inferred from 2 months of MDI data, and determined from the standard solar model. Note that while the error is given as a square, the sign gives the direction of the deviation. Horizaontal bars show the resolution in depth.

speed of sound. The present error estimate of temperature determined this way is about 2% around $T_{\rm c} = 1.57 \times 10^7$ K.

Note that when longer time series from SOHO have become available, the curve seems to settle closer to the model prediction in the central parts of the Sun than the curve derived from the earlier ground-based observations. The reason for this is that the low-degree modes, which penetrate deepest into the Sun, are difficult to resolve from ground-based observations.

4.4.3 Solar radius and the depth of the convection layer

The solar radius is possible to measure using the lowest rigde in the (ν, l) diagrams. It is the f mode that does not have radial nodes. Its dispersion equation is extremely simple $\nu = \sqrt{gk_{\rm h}}/2\pi$. The horizontal wave number is given by $k_{\rm h} \simeq \sqrt{l(l+1)}/r$. The f mode frequencies are essentially independent on the internal structure of the Sun. The mass variable $m(r)$ varies very little on the observable layers and thus we can write $g = Gm_\odot/r^2$ and we find that

$$\nu \propto r^{-3/2}\,. \tag{4.44}$$

Now the frequencies can be determined with high precision and $Gm_\odot$ is known at an accuracy of some 8 numbers.

The problem with this method is that the optical depth in the solar atmosphere that corresponds to the surface must be defined precisely. In practice the calculation is done by calculating the f mode eigenfrequencies assuming somewhat different solar radii. Fixing the model radius at optical depth $\tau = 2/3$ (Chapter 2) the radius is $(6.957 \pm 0.001) \times 10^8$ m.

A critical distance to be determined is the depth of the convection zone. The critical parameter determining the depth is the ratio between the so-called mixing-length and the pressure scale height $\alpha = l_m/H_P$ (see next chapter where convection is discussed more in detail). The p mode ridges of high degree l are very sensitive to α. Consequently, helioseismology has contributed signficantly to the determination of the base of the convection zone. First it moved the base from $0.85\,r_\odot$ to $0.74\,r_\odot$ in the 1980's and with improving solar models down to about $0.71\,r_\odot$ in the 1990's. The parameter α has been increased from 1.38 to 1.81 since the end of the 1980's. The present estimate for the location of the base of the convection zone is $r_v = (0.713 \pm 0.001)\,r_\odot$ which corresponds to the depth of (199700 ± 700) km from the surface of the Sun.

4.4.4 Internal rotation

Another inversion problem is to determine the internal rotation rate of the Sun. This is a much more difficult task than the determination of the sound speed because the rotation shows in the splitting of the spectral lines for a given l. Because the rotation rate is less than 0.5 μHz, very good frequency resolution is required. However, the task is doable. Figure 4.8 shows a result based on SOHO MDI data.

An interesting and a very important conclusion from this figure is that the differential rotation disappears near, or below, the base of the convection zone. This is a feature that may be important in the generation of the solar magnetic field as modern dynamo theories locate the generation process near the bottom of the convection zone. Another important result is that there is no sign that the core of the Sun would rotate faster than the outer layers. This is corroborated by the ground-based observations as well. A fast rotating core was one of the suggestions to get around the neutrino problem but also it failed.

4.4.5 Time-distance method

Time-distance seismology is a traditional method of geophysics. In solar physics it has been used only since mid-1990's. The method belongs to the toolbox of **tomography**.

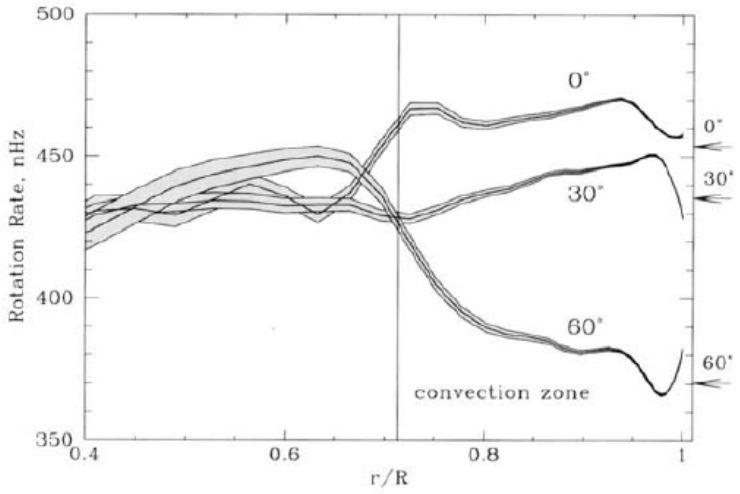

Figure 4.8: Internal rotation rate derived from MDI data. The radial profiles are calculated for four different latitudes. The gray regions indicate the estimated error in the inversion procedure. The errors are larger than in figure 4.7.

The travel time of an acoustic wave on the Sun can be measured using the covariance function

$$\Psi(\tau, |\mathbf{r}_1 - \mathbf{r}_2|) = \int f(\mathbf{r}_1, t) f^*(\mathbf{r}_2, t + \tau)\, dt\,, \tag{4.45}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are two points on the solar surface, $f(\mathbf{r}_i, t)$ is an oscillation signal at $\mathbf{r}_i$, and τ is the time delay. The signal f can be the velocity or intensity variation. For given $|\mathbf{r}_1 - \mathbf{r}_2|$ the covariance function attains maxima for delay times τ that correspond to the travel times of the signal from $\mathbf{r}_1$ to $\mathbf{r}_2$ via acoustic rays with $1, 2, \ldots$ reflections at the base of their propagation cavities.

The travel time of the acoustic wave varies due to the variation of $c_\mathrm{s}(\mathbf{r})$ and local flows with velocity $\mathbf{v}(\mathbf{r})$ along the path

$$\tau = \int \frac{ds}{c(\mathbf{r}) + \mathbf{v}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})}\,, \tag{4.46}$$

where s is a coordinate along the path and $\mathbf{n}(\mathbf{r})$ is a unit vector tangent to the path. Write $c(\mathbf{r}) = c_0(\mathbf{r}) + \delta c(\mathbf{r})$ and consider small variations from $c_0(\mathbf{r})$, which is the sound speed of a horizontally uniform reference state. Then

$$\tau = \int \frac{ds}{c_0(\mathbf{r})} - \int \frac{\delta c(\mathbf{r}) + \mathbf{v}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})}{c_0^2(\mathbf{r})}\, ds\,. \tag{4.47}$$

This integral is again an inversion problem. This time there is no parameter dependence but a solution is possible because a large number of travel times and wave paths are evaluated at the same time (tomography). Now the variations in sound speed and flow variations can be separated by measuring the travel times τ^+ and τ^- of signals traveling in opposite directions between two points. δc is the same for both measurements but the effect of local flow is opposite. Thus we have two inversion problems

$$\frac{1}{2}(\tau^+ + \tau^-) = -\int \frac{\delta c(\mathbf{r})}{c_0^2(\mathbf{r})}\, ds \tag{4.48}$$

$$\frac{1}{2}(\tau^+ - \tau^-) = -\int \frac{\mathbf{v}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})}{c_0^2(\mathbf{r})}\, ds\,. \tag{4.49}$$

This technique has been applied in studies of supergranulation that is a network of convective cells of 20 000–30 000 km size on the surface of the Sun. The flow structure giving rise to the visible structure on the surface continues to the depth of several thousand kilometers showing that supergranulation really has its origin in the convective motion. More details and an example are included in Chapter 5.

The acoustic signal at a given point can be reconstructed from signals at other points on the solar surface. This results in a possibility of building a holographic **acoustic image** of the Sun. During the last few years this method has been successfully applied in studies that reach deep into the interior or even to the far side of the Sun. With the aid of helioseismology we can really see into the Sun.

4.4.6 Excitation of the oscillations

Very little is known how the oscillations are excited in the first place. Some proposed mechanisms are associated with radiation "hitting" the base of the convection zone, or processes driven by the upward and downward motions of the gas parcels in the convection zone. Once the oscillations are excited they propagate all over the Sun according to their dispersion equation and guided by the environmental parameters.

Although we do not understand the origin of oscillations, their properties have clearly become the most powerful tool in probing the solar interior.

Chapter 5

Convection and rotation

We have already encountered the convection zone several times. In the solar model of Chapter 2 we postponed the determination of the energy transfer term $(\partial T/\partial m)_{\rm C}$ in "the unstable layer" to the present chapter. In discussion of oscillations we learned that the bottom of the convection zone is a special region where even the most modern solar models show some deviations from conditions derived from the oscillation data. Also the differential rotation extends through the convection zone whereas the radiation zone seems to exhibit quite steady rotation.

5.1 Convection

That convection becomes the dominating energy transport process at some distance from the center of the Sun is due to the outward increasing opacity. A major factor here is that when the temperature decreases, more and more ions recombine. Partially ionized and neutral atoms absorb radiation much more efficiently than fully ionized plasma. At the bottom of the convection zone radiation cannot any more remove all energy coming from inside. This starts to heat the gas and it becomes convectively unstable. In recent models this begins at the distance of 0.71–0.72 $r_\odot$ from the center. On the other hand, once convection is initiated, it is a very efficient mechanism for distributing the heat as known from everyday experience (e.g., heating of a room).

The radiation zone is stably stratified but the convection zone is unstable: gas parcels move up, dissolve, and finally the cool gas returns back around the upward moving gas parcels. Note also that the radiation does not stop completely at the base of the convection zone. About 0.05 $r_\odot$ into the convection zone $F_{\rm C}$ exceeds $F_{\rm R}$ and within the last 0.1 $r_\odot$ from the surface

practically all energy transport is convective. Finally the convection reaches the photosphere and causes the granular structure of the solar surface.

The whole convection zone is continuously mixed, which makes it chemically homogeneous. This does not mean that the mean molecular mass μ would be constant as, in particular very close the surface the degree of ionization drops rapidly. However, within most of the convection zone $\mu \approx 0.61$ (section 2.4).

The solar convection is turbulent and problems of turbulent convection are in general difficult pieces of hydrodynamics and thermodynamics. Turbulence is one of the major problems in physics, for which no general description has been found. However, the continuously increasing power of computers has made extensive modelling and numerical "experiments" possible and our present understanding of turbulent phenomena relies heavily on numerical studies. Below we concentrate on some of the very basic physical concepts of the convection zone only.

5.1.1 Schwarzschild criterion

The concept of convective instability can be described considering the vertical temperature gradient dT/dr. Let the gas be vertically stratified and in hydrostatic equilibrium. Let an elementary parcel of material be displaced vertically so slowly that it remains in horizontal pressure balance with its surroundings. The parcel will feel a **buoyancy force** if

$$\delta\rho_{\rm i} < \delta\rho\,, \tag{5.1}$$

where $\delta\rho_{\rm i}$ is the change of density inside the parcel and $\delta\rho$ the change of the ambient density. From the perfect gas law $P = \rho\mathcal{R}T/\mu$ we obtain

$$\frac{\delta P_{\rm i}}{P} = \frac{\delta\rho_{\rm i}}{\rho} + \frac{\delta T_{\rm i}}{T} \tag{5.2}$$

$$\frac{\delta P}{P} = \frac{\delta\rho}{\rho} + \frac{\delta T}{T}\,. \tag{5.3}$$

The horizontal equilibrium implies $\delta P_{\rm i} = \delta P$ and using (5.1) we find

$$-\delta T_{\rm i} < -\delta T\,, \tag{5.4}$$

i.e., the parcel is unstable and will continue rising if

$$-\frac{dT}{dr} > -\frac{dT_{\rm i}}{dr}\,. \tag{5.5}$$

That is, the parcel will rise if the ambient temperature falls faster than the temperature of the parcel.

In the optically thick solar interior the time scale of the energy (heat) exchange is long compared to the sound travel time across the parcel. Thus simultaneously, when assuming the pressure balance (maintained by the sound waves), the cooling of the parcel can be considered adiabatic. Thus the instability condition is usually expressed as

$$-\frac{dT}{dr} > -\left(\frac{dT_{\rm i}}{dr}\right)_{\rm ad} . \tag{5.6}$$

This is called the **Schwarzschild criterion**. It can be written in the form

$$\nabla > \nabla_{\rm a} , \tag{5.7}$$

where the same notation

$$\nabla = \frac{d\ln T}{d\ln P} \; ; \; \nabla_{\rm a} = \left(\frac{\partial \ln T}{\partial \ln P}\right)_S \tag{5.8}$$

as in section 2.2.4 has been used

Exercises

1. Prove (5.7).
2. Show that for perfect gas the Schwarzschild criterion reads
$$-\frac{dT}{dr} > \frac{\gamma-1}{\gamma}\frac{\mu g}{\mathcal{R}} . \tag{5.9}$$
3. Show that the Schwarzschild criterion is equivalent to $N^2 < 0$, where N is the Brunt-Väisälä frequency.

5.1.2 Mixing-length approach

Modelling of turbulent convection is a formidable task even with modern computing tools. There are some mean-field-type approximative methods to find expressions for the average quantities. The mixing-length theory uses the **mixing-length** (l) as a parameter to determine an average scale length of the convective motion. Physically l is the distance over which a moving gas parcel can be identified before it dissolves into its environment. The length scale is chosen based on (better or worse) physical intuition. In mixing-length calculations it is defined by

$$l = \alpha H_P , \tag{5.10}$$

where $H_P = -(d\ln P/dr)^{-1}$ is the pressure scale height and α a constant between 1 and 2.

The results of mixing-length theory depend on the excess temperature ΔT of a gas parcel which has risen a distance of δr. In the first order this is

$$\begin{aligned} \Delta T &= \left[\frac{dT_\mathrm{i}}{dr} - \frac{dT}{dr}\right]\delta r \\ &= \left[\left(\frac{d\ln T}{d\ln P}\right) - \left(\frac{d\ln T_\mathrm{i}}{d\ln P}\right)\right]\frac{T\,\delta r}{H_P}\,. \end{aligned} \tag{5.11}$$

Next it is assumed that the mean value of the displacement for many parcels is $\langle\delta r\rangle = l/2$. Denote the average speed by v. The average convective flux is then

$$\begin{aligned} F_\mathrm{C} &= \langle v\,\rho c_P \Delta T\rangle \\ &= \frac{\alpha}{2}\rho c_P v T\left[\left(\frac{d\ln T}{d\ln P}\right) - \left(\frac{d\ln T_\mathrm{i}}{d\ln P}\right)\right] \\ &= \frac{\alpha}{2}\rho c_P v T(\nabla - \nabla')\,, \end{aligned} \tag{5.12}$$

where $'$ denotes the logarithmic temperature gradient of the gas parcel.

The velocity v can be determined from the buoyancy force on the parcel's density deficit $\Delta\rho$ which yields an acceleration

$$\frac{\partial^2\delta r}{\partial t^2} = -g\,\frac{\Delta\rho}{\rho} = -g\,\delta\frac{\Delta T}{T}\,, \tag{5.13}$$

assuming a pressure equilibrium between the parcel and its environment. δ was given in section 2.2.4 as $\delta = -(\partial\ln\rho/\partial\ln T)_P$. Assuming that ∇, ∇', δ, g and H_P are constants over the mixing length, we can integrate (5.13) and get

$$\left(\frac{\partial\delta r}{\partial t}\right)^2 = \frac{g\delta}{H_P}(\nabla - \nabla')(\delta r)^2\,. \tag{5.14}$$

Now using again $\langle\delta r\rangle = l/2$ and multiplying (5.14) by a factor of $1/2$ to account for some expenditure of buoyancy due to friction, the mean convection velocity is found to be

$$v = l\left[\frac{g\delta}{8H_P}(\nabla - \nabla')\right]^{1/2}\,. \tag{5.15}$$

Thus v also depends on l or α and certain arbitrariness remains in the theory, which is typical for fluid dynamics in general. α is "calibrated" by considering the amount of energy transferred radiatively in the same region. As discussed in section 4.4.3 the model, from which the numbers of these lecture notes have been taken, uses $\alpha = 1.81$.

After some tedious calculation the mean convective temperature gradient (with m as the independent variable) is found to be

$$\left(\frac{\partial T}{\partial m}\right)_\mathrm{C} = -\,\frac{T(\partial\ln T/\partial\ln P)}{4\pi\rho r^2 H_P} \tag{5.16}$$

completing the solar model equations of Chapter 2. Details of the model results in the convection zone are given in the table of Figure 5.1.

The mean convective speed is zero at the base of the convection zone. At first, the speed increases very slowly, reaching 100 m/s at about the same distance where $F_{\rm C}$ reaches $F_{\rm tot}$ (at about $0.94\,r_\odot$). In the upper part of the convection zone the average convective speed accelerates, reaching 1 km/s at $0.999\,r_\odot$ and 2 km/s just below the photosphere. These speeds are very small compared to depth of the convection zone (200 000 km). The recent analyses of SOHO oscillation data are consistent with speeds of this order in a thin layer of some 7000 km, i.e., $0.01\,r_\odot$ under, the solar surface. The motion appears to be rather disordered but clearly turbulent (see Figure 5.3 in the next section).

Note that $F_{\rm C}$ is determined more accurately than v due to the "intuitively" introduced factors of $\frac{1}{2}$ in $\langle \delta r \rangle$ and in formula (5.15). As discussed in section 4.4.3, the model from which the numbers of these lecture notes have been taken uses $\alpha = 1.81$. It is the "calibration" of this number against the radiative flux $F_{\rm R}$ which determines the accuracy of $F_{\rm C}$. In the energy transfer calculations this is quite sufficient but if we are interested in the details of the turbulent convection, as good as possible knowledge of v is essential.

5.2 Granulation

The average convection speed is related to the speed of heat convection, but the solar matter does not flow away from the convection zone. What goes up must come down and the result is a turbulent structure of convection cells. On the surface of the Sun this is manifested by a **granular pattern** covering the entire surface, with the exception of the sunspots. The bright granules are upward moving parcels of hotter gas and the intergranular lanes, which are about 100 K cooler, downward moving, gas.

The granules are small and weak structures. Thus their observation from the ground requires a relatively large telescope with image size of about 20 cm, corresponding to the focal length of about 20 m, and the seeing must be good. Also here the space observations have become very important. Figure 5.2 shows how the granules look like. They form a network of apparently random polygonal shapes.

The granular cell size is defined as the distance between the centers of adjacent granules. The mean cell size is about 1.75" (1250 km) but the distribution is wide (mean width of about 1.5"). Also the concept of granular diameter is in use. It is defined by the area of the region where a granule exceeds a given, e.g., the mean, intensity. The distribution of the diameters

$r/r_\odot$	P [Pa]	T [K]	$\rho\left[\frac{\text{kg}}{\text{m}^3}\right]$	η_{H}	η_{He}	η_{He^3}	$\Delta\nabla$	ΔT [K]	$v\left[\frac{\text{m}}{\text{s}}\right]$	F_{C}/F
1.000	9.55+03	5.78+3	2.51−4	.00	.00	.00	−1.1−1	0.0+0	0	0.00
1.000	1.18+04	6.23+3	2.87−4	.00	.00	.00	9.7−2	7.7+0	152	0.00
1.000	1.34+04	6.74+3	3.01−4	.00	.00	.00	4.1−1	4.6+2	1181	0.05
1.000	1.45+04	7.21+3	3.05−4	.00	.00	.00	5.4−1	1.5+3	2172	0.32
1.000	2.19+04	9.26+3	3.52−4	.02	.00	.00	2.0−1	1.6+3	2427	0.96
1.000	3.33+04	1.05+4	4.57−4	.05	.00	.00	9.7−2	9.2+2	2042	0.99
.999	5.04+04	1.14+4	6.16−4	.09	.00	.00	5.7−2	5.9+2	1783	1.00
.999	7.64+04	1.22+4	8.44−4	.13	.00	.00	3.7−2	4.1+2	1581	1.00
.999	1.16+05	1.30+4	1.17−3	.17	.00	.00	2.5−2	2.9+2	1412	1.00
.999	1.76+05	1.37+4	1.63−3	.21	.00	.00	1.7−2	2.2+2	1266	1.00
.999	2.66+05	1.45+4	2.28−3	.25	.00	.00	1.3−2	1.7+2	1139	1.00
.998	4.04+05	1.53+4	3.18−3	.28	.00	.00	9.2−3	1.3+2	1026	1.00
.998	6.12+05	1.62+4	4.46−3	.32	.00	.00	6.9−3	1.0+2	925	1.00
.998	9.28+05	1.71+4	6.24−3	.36	.00	.00	5.2−3	8.0+1	837	1.00
.997	1.41+06	1.80+4	8.72−3	.40	.00	.00	3.9−3	6.4+1	758	1.00
.997	2.13+06	1.91+4	1.22−2	.44	.00	.00	3.0−3	5.3+1	688	1.00
.997	3.23+06	2.03+4	1.70−2	.48	.01	.00	2.4−3	4.3+1	626	1.00
.996	4.90+06	2.16+4	2.37−2	.52	.01	.00	1.9−3	3.6+1	571	1.00
.996	7.42+06	2.31+4	3.28−2	.56	.02	.00	1.5−3	3.1+1	524	1.00
.995	1.13+07	2.48+4	4.51−2	.61	.04	.00	1.2−3	2.7+1	481	1.00
.995	1.71+07	2.68+4	6.21−2	.65	.07	.00	9.3−4	2.3+1	441	1.00
.994	3.18+07	3.03+4	9.89−2	.72	.15	.00	6.8−4	1.9+1	393	1.00
.993	5.94+07	3.47+4	1.56−1	.78	.30	.00	4.9−4	1.5+1	350	1.00
.992	9.01+07	3.84+4	2.09−1	.81	.43	.00	3.9−4	1.4+1	326	1.00
.991	1.37+08	4.28+4	2.79−1	.84	.57	.00	3.2−4	1.2+1	305	1.00
.990	2.07+08	4.81+4	3.68−1	.87	.70	.00	2.6−4	1.1+1	286	1.00
.988	3.14+08	5.47+4	4.82−1	.89	.80	.00	2.1−4	1.0+1	269	1.00
.987	4.76+08	6.28+4	6.27−1	.91	.86	.01	1.6−4	9.3+0	251	1.00
.985	7.21+08	7.25+4	8.13−1	.93	.88	.03	1.2−4	8.1+0	232	1.00
.983	1.09+09	8.37+4	1.05+0	.94	.85	.09	9.0−5	6.8+0	212	1.00
.981	1.66+09	9.65+4	1.37+0	.94	.77	.19	6.5−5	5.6+0	194	1.00
.976	3.09+09	1.19+5	2.04+0	.95	.57	.40	4.1−5	4.4+0	171	1.00
.971	5.77+09	1.49+5	3.00+0	.96	.38	.61	2.6−5	3.5+0	153	1.00
.967	8.75+09	1.73+5	3.88+0	.96	.28	.72	2.0−5	3.1+0	142	1.00
.962	1.33+10	2.02+5	4.99+0	.96	.21	.79	1.5−5	2.7+0	132	1.00
.956	2.01+10	2.37+5	6.42+0	.96	.16	.84	1.1−5	2.3+0	123	1.00
.949	3.05+10	2.78+5	8.24+0	.97	.13	.87	8.1−6	2.0+0	114	1.00
.942	4.62+10	3.27+5	1.06+1	.97	.11	.89	5.9−6	1.8+0	105	1.00
.932	7.00+10	3.85+5	1.36+1	.97	.09	.91	4.3−6	1.5+0	98	1.00
.922	1.06+11	4.53+5	1.74+1	.97	.08	.92	3.2−6	1.3+0	90	.99
.910	1.61+11	5.34+5	2.24+1	.97	.07	.93	2.3−6	1.1+0	84	.99
.896	2.44+11	6.29+5	2.87+1	.97	.07	.93	1.7−6	9.7−1	78	.98
.880	3.70+11	7.42+5	3.68+1	.97	.06	.94	1.2−6	8.4−1	72	.97
.862	5.61+11	8.75+5	4.73+1	.98	.06	.94	9.1−7	7.2−1	67	.95
.841	8.50+11	1.03+6	6.07+1	.98	.05	.95	6.6−7	6.2−1	62	.91
.818	1.29+12	1.22+6	7.79+1	.98	.05	.95	4.7−7	5.2−1	57	.86
.778	2.41+12	1.56+6	1.14+2	.98	.04	.96	2.7−7	3.8−1	48	.71
.732	4.56+12	2.01+6	1.66+2	.98	.04	.96	1.1−7	1.9−1	34	.34
.717	5.44+12	2.15+6	1.85+2	.98	.04	.96	5.6−8	1.1−1	26	.15
.710	5.96+12	2.23+6	1.95+2	.98	.04	.96	1.9−8	3.8−2	15	.03

Figure 5.1: Model results for the convection zone.

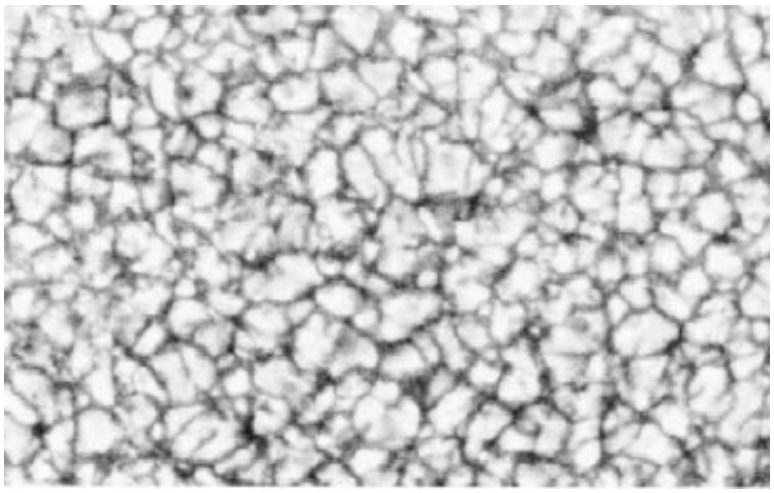

Figure 5.2: The granular surface of the Sun.

is a monotonically decreasing function of the diameter with a half width of about 1" (700 km).

The granulation is a non-stationary phenomenon. They grow, decay, and split in an apparently random manner. The lifetime distribution has roughly an exponentially decreasing shape where the number of granules has decayed by 1/e at about 6 min, which can be taken as the typical lifetime. However, the tail of the observed distribution extends well beyond 20 minutes.

Due to their small size and irregular motion it is rather difficult to measure the motion and evolution of the granules accurately. Both vertical and horizontal flow speeds of the granules are of the order of 1 km/s. Thus during its lifetime a single granule can move vertically about the length corresponding to its diameter (700 km), which is also a good estimate for the local pressure scale height. This suggests that the granules represent the convection in a very shallow layer (less than 2000 km) below the solar surface. Recent computer simulations are able to reproduce convective motion in such a shallow layer resembling the observed granular structure. It appears that the granular structure is governed by the details of the cooling process at the surface rather than by the heat flow from the inner parts of the convective zone.

Consequently, the granulation does not directly represent the structure of the convective flow from greater depths. There are observable patterns also in larger scales called **mesogranulation** and **supergranulation**. Mesogranulation was found as late as in 1981 as a network of "active" granules with a mesh size of 10" (7000 km). In this scale the vertical velocity is ± 150 m/s with an rms speed of 60 m/s. The horizontal velocity component varies around 500 m/s but may reach up to 1 km/s. The mesogranulation

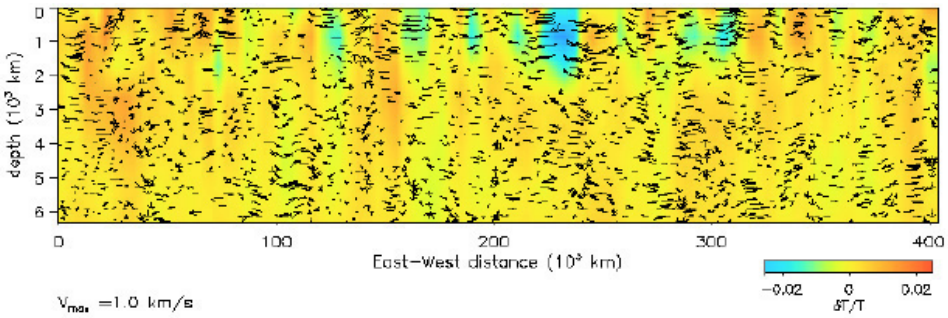

Figure 5.3: Flow pattern below the solar surface derived from SOHO/MDI observations. Note the different horizontal and vertical length scales.

patterns have not been successfully connected to convection cells with a distinct scale. Instead they appear to be a large-scale extension of the ordinary granulation.

Indication of an even larger granulation pattern was obtained already in 1956 and in 1962 *Leighton* and his coworkers identified a cellular pattern with a typical cell diameter of 1.6×10^4 km and a mean spacing of cell centers of 3×10^4 km (40"), which they named supergranulation. Recently it has been found that the distribution of supergranules is similar to the distribution of granules in spite of the large difference in their length scales. This structural similarity supports the view that as the ordinary granules are related to the convection structure in the a very shallow layer below the surface, so are the supergranules related to the convection cells at deeper levels.

The lifetime of supergranules is of the order of 1 day, which is difficult to observe from the Earth as a given granule is difficult to identify after a night. Here SOHO and its Michelson Doppler Imager (MDI) have had a central role. MDI provides excellent pictures of the supergranular pattern on the surface. Furthermore, helioseismic inversion of MDI data allows us to see below the supergranulation. Figure 5.3 is an illustration of the convective motion deduced from tomographic reconstruction MDI observations using the time-distance method introduced in section 4.4.5. The horizontal scale of the structures is that of supergranulation. Horizontal velocities in supergranules are 300–400 m/s whereas the upflows in the central areas have been estimated to 50 m/s and downflows at the boundaries to 100 m/s. The supergranulation is expected to reach to the depth of about 20 000 km where latent heat is released in the recombination process of He^+. The inversion results in this figure do not reach that deep yet, but on the other hand the vertical motion does not seem to stop at the bottom of the investigated layer.

The supergranular structure is associated to the chromospheric emission

network which is clearly visible in the Hα images. It is also a granular pattern covering the whole Sun and sometimes in popular presentations it is mixed with the basic small granules in the photosphere. The connection between the chromospheric network and supergranules is established by the magnetic field which we will discuss in the subsequent chapters.

There are also patterns on the solar surface with length scales of the order of 10^5 km or larger. They are called **giant cells**. No typical patterns of the giant cells have yet been established but they may play a role in the differential rotation of the Sun. Note that the larger the structures become the more important becomes the Coriolis force as discussed in the next section.

5.3 Rotation

The solar rotation was discovered immediately after the advent of telescope in about 1610. In 1630 it was realized that the rotation is not rigid, but the equatorial surface rotates faster than the high-latitude regions. The origin of this **differential rotation** is not yet fully understood, but evidently it is related to the transport of angular momentum.

5.3.1 Axis of rotation

The rotation axis of the Sun is given by two angles: the **inclination** i between the ecliptic plane and the equatorial plane, and the **angle of the ascending node** α of the Sun's equator, i.e., the angle, counted in the ecliptic, between the vernal equinox direction and the direction where the solar equator cuts the ecliptic (from below). Note that the Earth's precession shifts the equinox direction by 0.0196°/year (50"/year). Therefore, α increases by the same rate. Thus the **epoch** must be given with coordinates related to the equinox.

Carrington determined these angles in 1863 as $i = 7.25°$ and $\alpha(1850) = 73.67°$. The latter is still valid but the Greenwich sunspot data from the period 1874–1976 imply $i = 7.12° \pm 0.05°$.

We denote the **heliographic latitude** by ψ. In theoretical discussions the polar angle (co-latitude) $\theta = \pi/2 - \psi$ is often used. There is no physically unique way to define the longitude on the differentially rotating surface. For this purpose Carrington introduced a notation that is still in use. He divided time into intervals of 27.2753 days. These intervals are called **Carrington rotations** and in one year of 365 days there are 13.38 of these. Rotation number 1 was defined to have begun on 9 November 1853. On March 14, 2011, rotation 2108 begun. At the time of commencement of a new rotation

longitude $\phi = 0$ is given to the center of the solar disc. Note that the Carrington rotations are related to the motion of the Earth around the Sun, i.e., the "same place" at the solar equator is toward the Earth after one Carrington rotation. This is called the **synodic period**. The "true" rotation period with respective to the stars is the **sidereal period** of about 25 days. Its exact length is not easy to determine either.

5.3.2 Oblateness

A rotating non-rigid body is not fully spherical. Even the Earth is in this sense elastic and has an **oblateness** $f = (r_{\rm eq} - r_{\rm pol})/r_{\rm eq} \approx 1/300$. The fast rotating giant gas planets Jupiter and Saturn are much more oblate, $f_{\rm J} = 0.065$ and $f_{\rm S} = 0.098$, which can be seen already in rather low resolution pictures. But how oblate is the slowly rotating Sun, whose exact diameter is difficult to measure?

Neglecting the differential rotation and expanding the external gravitational field up to the quadrupole term (the first non-zero correction)

$$\Phi_{\rm ext} = -\frac{Gm_\odot}{r}\left[1 - J_2\left(\frac{r_\odot}{r}\right)^2 P_2(\cos\theta)\right] \tag{5.17}$$

the oblateness expressed as $\Delta r/r_\odot$ is

$$\frac{\Delta r}{r_\odot} = \frac{1}{2}\frac{\Omega^2 r_\odot}{g_\odot} + \frac{3}{2}J_2\,, \tag{5.18}$$

where Ω is the angular velocity of the solar surface, J_2 the quadrupole moment and $P_2(\cos\theta)$ the second Legendre polynomial. Using the Carrington rotation rate, the first term in (5.18) is 10^{-5}.

Exercise

Derive equation (5.18) and calculate Δr in kilometers for the first term on the RHS. Note that the result is below the resolution of present day telescopes.

The observational determination of the oblateness is extremely difficult. In 1967 *Dicke* and *Goldberger* measured the oblateness of $(5.0 \pm 0.7) \times 10^{-5}$. This corresponds to $J_2 \approx 2.67 \times 10^{-5}$. Such a large quadrupole moment would cause a perihelion shift of 3.5" per century in the orbit of Mercury. This result raised some concern about Einstein's general relativity which accounts for the observed 43" per century not explained by classical mechanics. (Recall that the total shift seen from the Earth is about 5600" of which 5026"

is due to the non-inertial reference frame of the observer and 531" the effect other planets.) Later determinations (also by Dicke and co-workers) are, however, much smaller. The present estimates for the oblateness are around 10^{-5}. The internal rotation derived from helioseismology has been estimated to yield for the quadrupole moment $J_2 = (2.18 \pm 0.06) \times 10^{-7}$. Thus the quadrupole moment is negligible in the context of the perihelion shift of Mercury and the general relativity survived one more empirical test.

5.3.3 Rotational history

The Sun has most likely rotated much faster in the past than today. The specific angular momentum of the cloud collapsing to form the Sun was much larger than that of the present solar system. Much of this was probably lost by magnetic braking in a very early phase the evolution. There are stars of similar mass as the Sun rotating faster, which are in the evolutionary phase of entering the main sequence of the H-R diagram. These **T Tauri** stars have surface velocities of about 15 km/s (the present Sun: 2 km/s). Furthermore, in the main sequence the older stars rotate typically more slowly than the younger ones.

According to pre-main-sequence stellar models the Sun was fully convective before the hydrogen burning started. The convection was turbulent and the rapid exchange of momentum between parcels of gas evened out gradients in angular velocity Ω. The total angular momentum J_0 has been estimated to be 8×10^{42} kg m^{2}s^{-1}, whereas the present value is 1.7×10^{41} kg m^{2}s^{-1}.

Matter leaving the Sun carries angular momentum, but the material loss since the time of high J_0 has been negligible. The magnetic field, however, is a very efficient lever arm for a torque. As we will learn later, the magnetic field forces the escaping material to rotate with the Sun out to the so-called **Alfvén radius** $r_\mathrm{A} \approx 12\ r_\odot$. Thus the angular momentum density increases up to r_A, and it is this angular momentum which is conserved in the escaping flow beyond r_A. The rate of angular momentum loss is

$$\frac{dJ}{dt} = \Omega r_\mathrm{A}^2 \frac{dm}{dt}\,. \tag{5.19}$$

How much such **magnetic braking** really has taken place in the history, is difficult to estimate because we do not know the history of the magnetic field and the solar-wind flow, on which r_A depends. Magnetic field is, on the other hand, generated by the solar dynamo and this depends on Ω. Because we hardly can model the present generation rate of the magnetic field, it is understandable that we cannot model its history either. However, as long as the Sun was fully convective the slowing down affected the whole Sun.

When the radiative core developed, the motion of the outer convective zone was disconnected from the interior. The convective part continued to lose angular momentum by magnetic braking but what happened to the core? Because the central core contracted further, it appears as its rotation rate could have accelerated.

However, the recent results of helioseismology do not support the idea of a fast rotating core. The central core may rotate somewhat faster than the radiative zone but something seems to have slowed down the rotation. A strong inward gradient $d\Omega/dr$ would mean strong shear flows which could drive instabilities that, in turn, could transport the excess angular momentum, resulting in smoother $d\Omega/dr$. It has also been speculated that there could be internal magnetic fields. Indeed, already very weak magnetic fields are sufficient to slow down the core. This is one more indication that the magnetic fields are very essential to the solar dynamics.

5.3.4 Differential rotation

Carrington determined the surface rotation rate from sunspot data as a function of the heliographic latitude in (sidereal) degrees per day

$$\Omega(\psi) = 14.25 - 2.75 \sin^{7/4} \psi \tag{5.20}$$

The power 7/4 is a bit awkward. A more modern approach is to expand the rotation rate as

$$\Omega(\psi) = A + B \sin^2 \psi + C \sin^4 \psi + \ldots \tag{5.21}$$

and in most studies only A and B are determined. Here A is the equatorial rotation rate.

In addition to sunspot data, Doppler shifts, edges of coronal holes and surface magnetograms are used in studies of the rotation rate. The different methods yield slightly different results and there is some variability within the individual methods as well. (For example, different sunspot cycles are different.)

It is interesting to note that the larger the structure, that is used to determine the rotation, the more uniform rotation is found. The extreme are observations of large coronal holes which sometimes show very little differential rotation at all. Again helioseismology has revolutionized the studies of differential rotation, as now we can empirically determine the rotation also inside the Sun as illustrated in Figure 4.8.

5.3.5 Theory of rotating convection zone

In a complete description of the convection zone the magnetic field must be included. We postpone this to the next chapter and consider here the role of the angular momentum only.

Assume that the convection zone is a spherical shell of compressible gas. Let the inner and outer radii be r_v and $r_\odot$. Let the turbulent velocity field, denoted by $\mathbf{u} = (u_r, u_\theta, u_\phi)$, be superimposed on a non-uniform rotation $\langle v_\phi \rangle$ and possible meridional circulation $\mathbf{v}_\mathrm{m} = (\langle v_r \rangle, \langle v_\theta \rangle)$. The total velocity is then

$$\mathbf{v} = \langle \mathbf{v} \rangle + \mathbf{u} \,. \tag{5.22}$$

We neglect the molecular viscosity, which is small in the Sun. Thus the momentum equation is the **Euler equation**

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla P - \nabla \Phi \,. \tag{5.23}$$

We further neglect $\partial \rho / \partial t$ and write the continuity equation as

$$\nabla \cdot (\rho \mathbf{v}) = 0 \,. \tag{5.24}$$

Consider now the mean longitudinal motion averaging (5.23) over the angle ϕ. Multiplying the equation by the "lever arm" $s = r \sin\theta$ we get the conservation equation of the angular momentum. Now $\langle v_\phi \rangle = s\Omega$ and finally

$$\frac{\partial}{\partial t}(\rho s^2 \Omega) + \nabla \cdot (\rho s^2 \Omega \mathbf{v}_\mathrm{m} + \rho s \langle u_\phi \mathbf{u} \rangle) = 0 \,. \tag{5.25}$$

$\rho s^2 \Omega$ is the angular momentum density. In the flux term there are two contributions, which transport angular momentum: the **meridional circulation** $\mathbf{v}_\mathrm{m}$ and the **Reynolds stresses**

$$Q_{ij} = \langle u_i u_j \rangle \,. \tag{5.26}$$

The temporal variations of the angular velocity are much smaller than the latitudinal variation. Thus we restrict the discussion to steady models. This implies that the transport processes must balance each other. Furthermore, we assume that there is no flux of matter or angular momentum through surfaces at r_v and $r_\odot$. Thus, at these boundaries

$$\langle v_r \rangle = 0 \;;\; Q_{r\phi} = 0 \,. \tag{5.27}$$

Two main approaches to deal with Q_{ij} are: (1) Direct calculation of the turbulent velocity components u_i, which is very difficult, and (2) mean-field models, which operate with average quantities. As in the case of convection, we discuss the mean-field approach only.

Mean-field models

We start by dividing the Reynolds stresses to diffusive and non-diffusive parts. Note that we have already earlier neglected the molecular viscosity. The diffusion arises from the turbulent motion of fluid elements and we denote the **turbulent diffusivity** by $\nu_{\rm t}$. Diffusion smooths the gradients in the angular momentum according to the diffusion equation, which in this case reads

$$\frac{\partial}{\partial t}(\rho s^2\Omega) = \nabla\cdot(\nu_{\rm t}\rho s^2\nabla\Omega)\,. \tag{5.28}$$

The angular momentum is transported inward and the non-zero Reynolds stresses are

$$(Q_{r\phi}, Q_{\theta\phi}) = -\nu_{\rm t} s\nabla\Omega\,. \tag{5.29}$$

Due to the factor $s = r\sin\theta$ these stresses vanish at the poles. If there were no meridional circulation and no other contribution than (5.29) to the Reynolds stresses, the rotation would be uniform.

If finite meridional circulation is added, the rotation becomes differential. As the Sun rotates, its convective motions are distorted by the **Coriolis force**. In the rotating frame this shows as the term $2\,\mathbf{\Omega}\times\mathbf{v}$ on the LHS of Euler's equation (5.23). The effect of the Coriolis force depends on the scale size; it is stronger for larger structures. The Coriolis effect is estimated by the inverse of the **Rossby number** that is the ratio of inertial and Coriolis forces

$$Ro = \frac{u}{2\Omega l}\,, \tag{5.30}$$

where u and l are the characteristic velocity and length scales. For example, the Rossby number for supergranules ($l\approx 10^7$ m, $u\approx 500$ m/s, and $\Omega\approx 3\times10^{-6}$ s^{-1}) is about 10, and thus the rotation effect is of the order of 1/10. For the giant cells ($l\approx 10^8$ m and $u\approx 100$ m/s) the Rossby number is well below 1 and the effect of rotation to them is significant. As a result the convective heat transport becomes latitude-dependent. The circulation restores the energy balance and transports angular momentum. Figure 5.4 is an example of this kind of modelling of the circulation in the convection zone.

Another way to describe the relationship between the angular velocity and Reynolds stresses is to write formally

$$Q_{ij} \propto \Lambda_{ijk}\Omega_k\,, \tag{5.31}$$

with summation over k. Now the structure of the tensor Λ_{ijk} has to be found by physical reasoning and model computations. The difference from the latitude-dependent heat transport modelling is that the distorted energy balance is not explicitly involved. Of course, also this formalism includes the Coriolis effect. In addition, this approach accounts for possible **anisotropies**

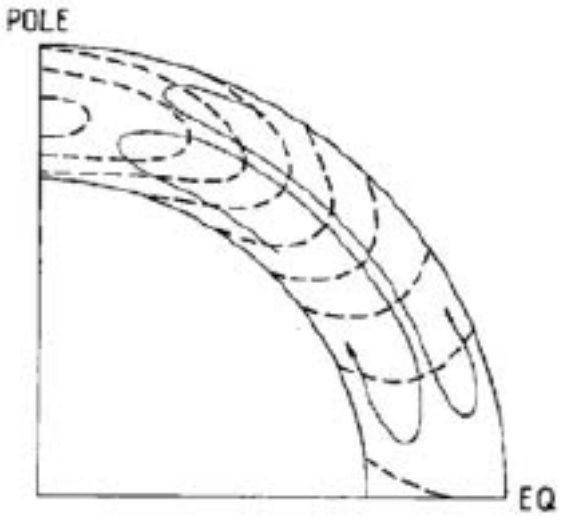

Figure 5.4: Stream lines of meridional circulation (solid curves) and surfaces of constant angular velocity (dashed lines) for a model with latitude-dependent heat transport.

in the turbulence. The turbulent diffusivity is allowed to act differently in the horizontal and vertical directions. A simple way to write the vertical and horizontal Reynolds stresses is

$$Q_{r\phi} = -\nu_{\rm t} r \sin\theta \frac{\partial \Omega}{\partial r} + \Lambda_r \sin\theta\, \Omega \tag{5.32}$$

$$Q_{\theta\phi} = -a\nu_{\rm t} \sin\theta \frac{\partial \Omega}{\partial \theta} + \Lambda_{\rm h} \cos\theta\, \Omega\,, \tag{5.33}$$

where a is a dimensionless parameter to account for different horizontal and vertical smoothing. The vertical part of the lambda tensor Λ_r could be a function of r alone, but the horizontal part $\Lambda_{\rm h}$ must have the angular dependence of at least $\sin^2\theta$. In more refined models both are expanded in terms of trigonometric functions.

It is important to keep in mind that we cannot solve the turbulent convection problems exactly. Thus the coefficients in the expressions for Q_{ij} must be determined by numerical modelling. To do this for the whole convection zone in the scales of the turbulent eddies is far beyond today's computing capabilities. Estimates for the need of grid points in the simulation codes have been of the order of 10^{22}, so the problems are formidable.

Proper understanding of the turbulent convection and differential rotation is also necessary to deal with another major problem in solar physics: the generation and dynamics of the solar magnetic fields. The magnetic field is generated by the motions in the convection zone and it governs most of

the solar activity and all dynamics of the solar atmosphere from the chromosphere to the heliopause beyond the orbits of the planets. Consequently, magnetism and magnetohydrodynamics play an important role in the remaining chapters.

Chapter 6

Solar magnetohydrodynamics

This chapter is a brief review of elements of magnetohydrodynamics (MHD) needed to understand the remaining parts of the lectures. The basic concepts are illustrated here by examples from solar physics but many important phenomena, such as generation of the solar magnetic field, solar flares, properties of the solar wind, etc. are discussed in the subsequent chapters. For students wishing to learn more on plasmas the Lecture notes on Advanced Space Physics, avaiable on-line at `http://theory.physics.helsinki.fi/~plasma_jatko/`, are recommended.

6.1 Basic concepts

6.1.1 Maxwell's equations

The classical electrodynamics is governed by Maxwell's equations

$$\nabla \cdot \mathbf{D} = \rho_q \tag{6.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{6.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{6.3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\,. \tag{6.4}$$

We call the field $\mathbf{B}$ magnetic field. Throughout these lectures $\mathbf{H} = \mathbf{B}/\mu_0$, where μ_0 is the magnetic permeability of free space.

We introduce the concept of the **magnetic field line** to represent a curve whose tangent at every point is in the direction of the field. In MHD

the field lines often get a role of almost physical objects but in reality they are just mathematical constructions.

The magnetic flux tube can be considered as a bunch of magnetic field lines intersecting an area (S) bounded by a simple closed curve. The **magnetic flux** through S is

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{S} \,. \tag{6.5}$$

Consider two cross sections of the same flux tube S_1 and S_2. Because the flux tube is defined by the field lines, no flux crosses the surface of the tube. The condition $\nabla \cdot \mathbf{B} = 0$ implies that $\Phi_1 = \Phi_2$, i.e., the flux is constant along the flux tube. The flux tubes are fundamental building blocks of solar MHD.

In MHD we assume that the plasma speeds are much smaller than the speed of light and neglect the displacement current $\partial \mathbf{D}/\partial t$. This implies that $\nabla \cdot \mathbf{J} = 0$. In MHD the electric field is calculated from $\mathbf{B}$ and from the macroscopic velocity $\mathbf{v}$. Thus we do not need the equation $\nabla \cdot \mathbf{D} = \rho_q$, except if we want to calculate the charge density ρ_q afterwards.

However, we need to relate the electric field and the current through Ohm's law. If the plasma is in motion, Ohm's law is

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \,. \tag{6.6}$$

In the frame moving with the plasma this reduces to $\mathbf{J} = \sigma \mathbf{E}$. Equation (6.6) is sufficient in most discussions below. Sometimes, more general forms are needed in solar physics, e.g.,

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = \frac{\mathbf{J}}{\sigma} + \frac{1}{ne}\mathbf{J} \times \mathbf{B} - \frac{1}{ne}\nabla \cdot \mathcal{P}_\mathrm{e} + \frac{m_\mathrm{e}}{ne^2}\frac{\partial \mathbf{J}}{\partial t} \,, \tag{6.7}$$

where $\mathcal{P}_\mathrm{e}$ is the electron pressure **tensor**. In the weakly ionized photosphere, one must also consider collisions with neutrals resulting in different conductivities in the direction of $\mathbf{E}$ (**Pedersen conductivity**) and perpendicular to $\mathbf{E}$ (**Hall conductivity**).

6.1.2 Fluid equations

Ohm's law relates the electromagnetic variables with the plasma flow. From fluid mechanics we need furthermore the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \,, \tag{6.8}$$

the momentum equation

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla P + \mathbf{J} \times \mathbf{B} + \rho \mathbf{g}\,, \tag{6.9}$$

the equation of state (ideal gas law)

$$P = \frac{\mathcal{R}}{\mu} \rho T\,, \tag{6.10}$$

and an energy equation. Here and in the following discussion the pressure is assumed to be scalar. The action of the magnetic field onto the plasma is established by the Ampère's force density $\mathbf{J} \times \mathbf{B}$ (integrated Lorentz force). In plasma physics the ideal gas law is usually given in the form $P = nk_{\mathrm{B}}T$.

6.1.3 Induction equation

From Ampère's, Faraday's, and Ohm's laws we find the **induction equation**

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \nabla \times (\eta \nabla \times \mathbf{B})\,, \tag{6.11}$$

where $\eta = 1/\mu_0\sigma$ is the **magnetic diffusivity**. Assuming that η is uniform we get the more familiar form

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}\,. \tag{6.12}$$

The first term on the RHS describes convection, the second diffusion. The ratio between these is the **magnetic Reynolds number**

$$R_{\mathrm{m}} = \frac{lu}{\eta}\,, \tag{6.13}$$

where l is the typical scale length of the spatial gradient ("$1/\nabla$") and u is the typical speed. When $R_{\mathrm{m}} \gg 1$, convection dominates, whereas for $R_{\mathrm{m}} \approx 1$, or less, diffusion becomes important.

6.1.4 Conductivity and diffusivity in the Sun

Almost everywhere in the Sun (or in fact in the entire universe) the classical resistivity is very small, i.e., σ is large. Important exceptions in the Sun are the photosphere and lower chromosphere where the ionization is low and collisions with neutrals inhibit the current flow. The photospheric conductivity is about 10 $\Omega^{-1}\mathrm{m}^{-1}$ (= 10 mho/m = 10 S/m, Figure 6.1)

Conductivity of 10 S/m yields $\eta \approx 10^5$ $\mathrm{m}^2\mathrm{s}^{-1}$. For photospheric granules ($l \approx 1000$ km, $u \approx 2$ km/s), we find $R_{\mathrm{m}} \approx 20000 \gg 1$. Thus the diffusion is

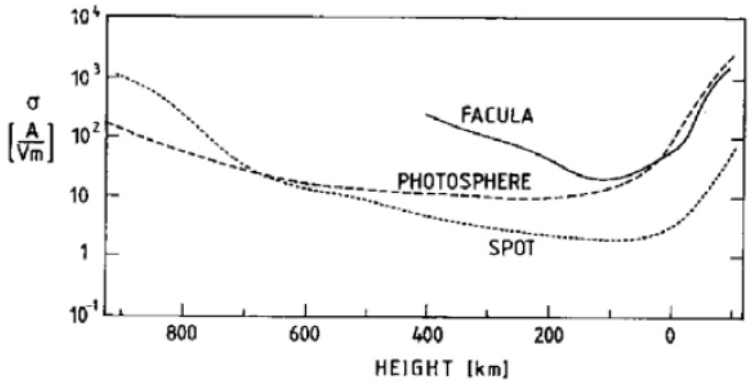

Figure 6.1: Photospheric conductivity. The sunspots are cooler and the conductivity is smaller than in the surroundings.

very weak. This is not quite consistent with the actually observed behavior of the magnetic fields, which seems to imply some 200 times larger diffusivity and correspondingly smaller $R_{\rm m}$. The explanation is that the turbulence in the upper convection zone introduces **turbulent diffusivity** $\eta_{\rm t} \approx 2 \times 10^7$ $\mathrm{m^2 s^{-1}}$, but there is no rigorous way to calculate this number.

The solar gas becomes fully ionized above 2000 km. "Fully" means that the fraction of neutral atoms is negligible but, of course, the heavy atoms have not lost all their electrons. In fully ionized plasma we can use **Spitzer's formula** to calculate the effective electron collision time

$$\tau_{\rm ei}({\rm s}) = 0.266 \times 10^6 \, \frac{T^{3/2}({\rm K})}{n_{\rm e}({\rm m^{-3}}) \ln \Lambda} \,, \tag{6.14}$$

where $\ln \Lambda$ is the Coulomb logarithm. Λ is the so-called **plasma parameter** that gives the number of electrons in a Debye sphere. In the solar interior $\ln \Lambda \sim 5$, in the chromosphere $\ln \Lambda \sim 10$, and in the corona $\ln \Lambda \sim 20$. The conductivity can be calculated as

$$\sigma = \frac{n_{\rm e} e^2 \tau_{\rm ei}}{m_{\rm e}} \,, \tag{6.15}$$

which has the numerical value

$$\sigma({\rm Sm^{-1}}) = 1.53 \times 10^{-2} \frac{T^{3/2}({\rm K})}{\ln \Lambda} \,. \tag{6.16}$$

Using $\ln \Lambda = 20$ the diffusivity is given by

$$\eta({\rm m^2 s^{-1}}) = 10^9 \times T^{-3/2}({\rm K}) \,. \tag{6.17}$$

For a coronal temperature $T = 10^6$ K this yields $\eta = 1$ $\mathrm{m^2/s}$.

Although the diffusivity is often small, it is never exactly zero. In the frame of reference moving with the plasma ($\mathbf{v} = 0$), we have the diffusion-dominated limit. As a simple example, we can consider a one-dimensional current sheet $B(z,t)\mathbf{e}_x$ with the initial condition

$$B(z,0) = \begin{cases} +B_0\,, & z > 0 \\ -B_0\,, & z < 0\,. \end{cases} \tag{6.18}$$

In 1-D the diffusion equation is

$$\frac{\partial B}{\partial t} = \eta \frac{\partial^2 B}{\partial z^2} \tag{6.19}$$

with the solution

$$\begin{aligned} B(z,t) &= \mathrm{erf}\left(\frac{z}{\sqrt{4\eta t}}\right) \\ &= \frac{2B_0}{\sqrt{\pi}} \int_0^{z/\sqrt{4\eta t}} \exp(-u^2) du\,. \end{aligned} \tag{6.20}$$

The total magnetic flux remains constant (=0) but the energy of the field $\int B^2/2\mu_0\, dz$ decreases with time. (Strictly speaking, this configuration is infinite, but we can think that there is an outer boundary somewhere.) It is an easy **exercise** to show that

$$\frac{\partial}{\partial t} \int \frac{B^2}{2\mu_0} dz = -\int \frac{j^2}{\sigma} dz\,. \tag{6.21}$$

Thus the energy is dissipated through **Ohmic heating** in the current sheet.

6.1.5 Ideal MHD

The opposite limit, large R_m, is the convection-dominated case when

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{6.22}$$

$\Rightarrow$

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B}\,. \tag{6.23}$$

This replaces Ohm's law and the theory is called **ideal MHD**. In ideal MHD the magnetic field evolution is determined by $\mathbf{v}$, the electric field is calculated from $\mathbf{v}$ and $\mathbf{B}$, and the current is determined from Ampère's law. No explicit Ohm's law is needed. Note that ideal MHD **does not require** $\sigma =$ "∞". σ may be finite if l and u are large enough.

In ideal MHD the plasma and the magnetic field are **frozen-in** to each other. This means that if two plasma elements are threaded by a common magnetic field line at a given time, they continue to do so. This is a very powerful result because it allows us to consider the plasma motion in terms of the magnetic field evolution. However, one should avoid considering the field-lines as some kind of spaghetti having material existence. Furthermore, many of the most important dynamical phenomena in the solar physics are associated with break-down of the frozen-in conditions.

Exercise

Prove the above **frozen-in theorem**.

6.2 Magnetohydrostatics

6.2.1 Static equilibrium

Consider next MHD plasma in a time-independent ($d/dt = 0$) equilibrium. Assuming scalar pressure the momentum equation reduces to

$$\mathbf{J} \times \mathbf{B} = \nabla P - \rho \mathbf{g} \,. \tag{6.24}$$

The magnetic force can be written using Ampère's law as

$$\begin{aligned} \mathbf{J} \times \mathbf{B} &= -\frac{1}{\mu_0} \mathbf{B} \times (\nabla \times \mathbf{B}) \\ &= -\nabla \left(\frac{B^2}{2\mu_0} \right) + \frac{1}{\mu_0} \nabla \cdot (\mathbf{B}\mathbf{B}) \,. \end{aligned} \tag{6.25}$$

The first term on the RHS is the gradient of the magnetic energy density, i.e., the **magnetic pressure**

$$P_B = \frac{B^2}{2\mu_0} \,. \tag{6.26}$$

The magnetic force can also be given as divergence of the **magnetic stress tensor** $-P_B \mathbf{I} + \mathbf{B}\mathbf{B}/\mu_0$, where $\mathbf{I}$ is the identitity tensor. In a coordinate system where, e.g., z axis is aligned with the magnetic field lines, we can write the stress tensor as $P_B(-\mathbf{e}_x\mathbf{e}_x - \mathbf{e}_y\mathbf{e}_y + \mathbf{e}_z\mathbf{e}_z)$, which shows that the magnetic field produces a tension of P_B in the direction of the field lines and a pressure P_B in the directions perpendicular to the field.

Exercises

1. Show that the magnetic pressure tries to compress the plasma if the magnetic field varies as a function of position, and that the magnetic tension tries to shorten the field lines if the field lines are curved.

2. Sketch the following field configurations and calculate the direction of magnetic forces in each case

   (a) $\mathbf{B} = x\mathbf{e}_y$

   (b) $\mathbf{B} = \mathbf{e}_x + x\mathbf{e}_y$

   (c) $\mathbf{B} = y\mathbf{e}_x + x\mathbf{e}_y$

   (d) $\mathbf{B} = r\mathbf{e}_\theta$

If the gravity is negligible as compared to the magnetic force, the **magnetohydrostatic equilibrium** can be described by

$$\nabla P = -\frac{1}{\mu_0}\mathbf{B} \times (\nabla \times \mathbf{B})\,. \tag{6.27}$$

Assuming plasma isotropic and nearly homogeneous (no magnetic stress) the sum of the magnetic and plasma pressures is constant

$$\nabla\left(P + \frac{B^2}{2\mu_0}\right) = 0\,. \tag{6.28}$$

The **plasma beta**

$$\beta = \frac{2\mu_0 P}{B^2} \tag{6.29}$$

expresses the ratio of the plasma and magnetic pressures.

The gravity effects can be neglected if the gradient scale length is much smaller than H/β where H is the pressure scale height

$$H = \frac{P_0}{\rho_0 g}\,, \tag{6.30}$$

in which P_0 and ρ_0 are typical values of pressure and density.

Assume then that the gravity is directed in the negative z direction and the magnetic field is oriented according to Figure 6.2.

Along the field lines $\mathbf{J} \times \mathbf{B} = 0$ and the equilibrium condition reduces to

$$\frac{dP}{ds} = -\rho g \cos\theta\,. \tag{6.31}$$

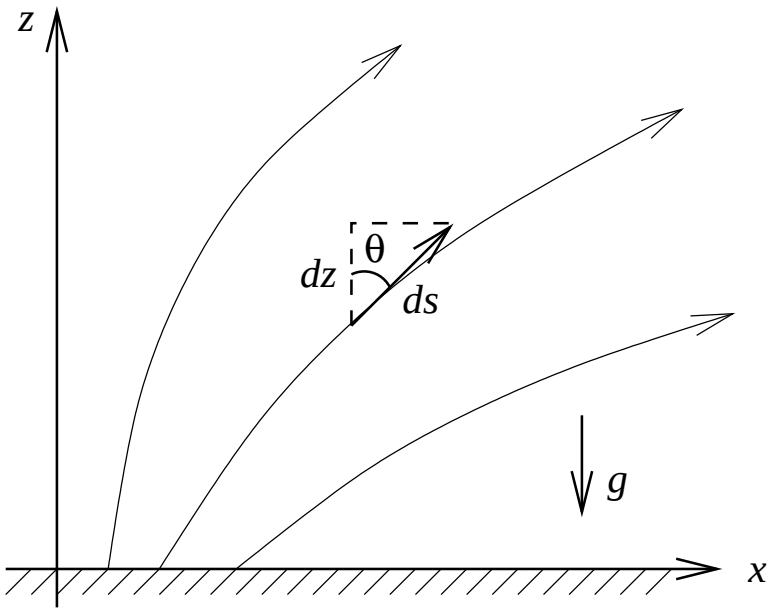

Figure 6.2: Magnetic field lines and direction of gravity.

Because $\cos\theta\, ds = dz$ we get

$$\frac{dP}{dz} = -P\frac{\mu g}{\mathcal{R}T} \tag{6.32}$$

$\Rightarrow$

$$P = P_0 \exp\left(-\int_0^z \frac{\mu g}{\mathcal{R}T(z)} dz\right) , \tag{6.33}$$

where P_0 is the pressure at the base of the field line. If T is uniform this reduces to the simple exponential relationship

$$P = P_0 \exp(-z/H) . \tag{6.34}$$

In the photosphere $T \approx 5000$ K and $H \approx 150$ km. At the maximum temperature of the corona ($T \approx 2 \times 10^6$ K) the scale height is about 10^5 km. (Note that this treatment only gives you the pressure along a single field line.)

6.2.2 Force-free fields

If in magnetohydrostatic equilibrium $\beta \ll 1$, the pressure gradient is negligible and

$$\mathbf{J} \times \mathbf{B} = 0 . \tag{6.35}$$

i.e, the electric current must flow along the magnetic field. Because a current causes a magnetic field around it, the self-consistent field-aligned current consists of spiraling magnetic field lines and is often called a **flux rope**. Another name is **force-free field** because the magnetic force on plasma is zero. Note that the force-free equilibrium is always an approximation to the momentum equation, but often a very good approximation.

The equation $\mathbf{J} \times \mathbf{B} = 0$ is difficult to solve. The problem lies in its nonlinearity: using Ampère's law it can be written as

$$(\nabla \times \mathbf{B}) \times \mathbf{B} = 0\,. \tag{6.36}$$

If $\mathbf{B}_1$ and $\mathbf{B}_2$ are the solutions to this equation, it does not necessarily follow that $\mathbf{B}_1 + \mathbf{B}_2$ would be another solution.

We can express the field-alignment by

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = \alpha(\mathbf{r})\mathbf{B}\,, \tag{6.37}$$

where α is a function of position. Taking divergence we get

$$\mathbf{B} \cdot \nabla\alpha = 0\,, \tag{6.38}$$

i.e., α is constant along the magnetic field. If α is constant everywhere the above sum $\mathbf{B}_1 + \mathbf{B}_2$ of two solutions is also a solution for the force-free field and the equation

$$\nabla \times \mathbf{B} = \alpha\mathbf{B} \tag{6.39}$$

is linear. Taking a curl of this we get the Helmholtz equation

$$\nabla^2\mathbf{B} + \alpha^2\mathbf{B} = 0\,, \tag{6.40}$$

which has known solutions. That the field fulfils the Helmholtz equation is a necessary but not sufficient condition for the field to be force-free as also the boundary conditions must be specified correctly.

6.2.2.1 Potential fields

A particularly important special case of force-free magnetic fields is the current-free configuration $\nabla \times \mathbf{B} = 0$. Then the magnetic field can be expressed as the gradient of a scalar potential $\mathbf{B} = \nabla\Psi$. Because $\nabla \cdot \mathbf{B} = 0$, the magnetic field can be found by solving the Laplace equation

$$\nabla^2\Psi = 0 \tag{6.41}$$

for appropriate boundary conditions. Thus we can use the well-developed methods of potential theory to find the solutions in current-free configurations.

6.2.2.2 Linear force-free model of a coronal arcade

As an example of linear force-free fields we construct a model of a coronal arcade. Let the configuration look like an arc in the xz-plane and extend uniformly in the y-direction. Let the structure be sinusoidal in the x-direction with wave number k. As there is a second spatial derivative in the Helmholtz equation, the same z-dependence is retained after two derivations for sinusoidal and exponential functions. Because the field should vanish at high altitude, we choose the z-dependence as $\exp(-lz)$. This fulfils the Helmholtz equation if $\alpha^2 < k^2$. In order to have the structure above the solar surface we consider $z > 0$. Let us then seek solutions of the form

$$\begin{aligned} B_x &= B_{x,0}\sin(kx)e^{-lz} \\ B_y &= B_{y,0}\sin(kx)e^{-lz} \\ B_z &= B_0\cos(kx)e^{-lz}\,. \end{aligned} \tag{6.42}$$

Now the equation $\nabla \times \mathbf{B} = \alpha\mathbf{B}$ yields

$$\begin{aligned} lB_{y,0} &= \alpha B_{x,0} \\ -lB_{x,0} + kB_0 &= \alpha B_{y,0} \\ kB_{y,0} &= \alpha B_0\,. \end{aligned} \tag{6.43}$$

And the field can be expressed as

$$\begin{aligned} B_x &= (l/k)B_0\sin(kx)e^{-lz} \\ B_y &= (\alpha/k)B_0\sin(kx)e^{-lz} \\ B_z &= B_0\cos(kx)e^{-lz}\,, \end{aligned} \tag{6.44}$$

where k, l, and α must be related by

$$l^2 = k^2 - \alpha^2\,. \tag{6.45}$$

The projection of the magnetic field lines on the xy-plane are straight lines parallel to each other

$$B_y = \frac{\alpha}{(k^2-\alpha^2)^{1/2}}B_x\,, \tag{6.46}$$

whereas the projection to the xz-plane are arcs as we wished (Figure 6.3)

This is an example which becomes much simpler if the current is so weak that we can neglect it and use potential theory. We can look for separable solutions in 2-D Cartesian space by writing $\Psi = X(x)Z(z)$. From the Laplace equation

$$\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial z^2} = 0 \tag{6.47}$$

we find

$$\frac{1}{X}\frac{d^2X}{dx^2} = -\frac{1}{Z}\frac{d^2Z}{dz^2} = -k^2\,, \tag{6.48}$$

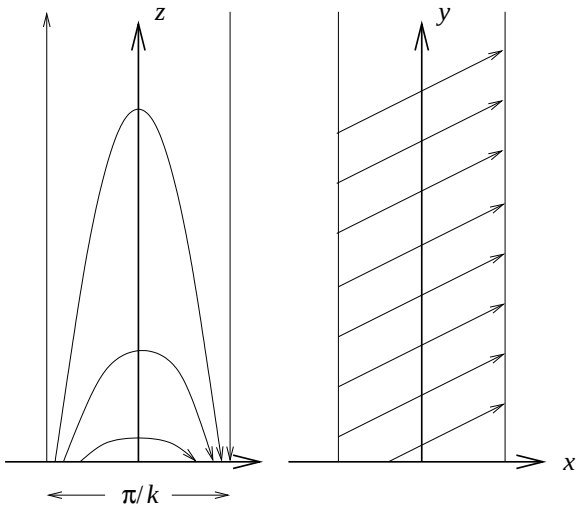

Figure 6.3: Linear force-free configuration above the surface of the Sun.

where k is a constant. This is fulfilled, e.g., by $\Psi = (B_0/k)\sin kx \exp(-kz)$ from which we find the field configuration

$$B_x = \frac{\partial \Psi}{\partial x} = B_0 \cos kx \exp(-kz) \tag{6.49}$$

$$B_z = \frac{\partial \Psi}{\partial z} = B_0 \sin kx \exp(-kz)\,. \tag{6.50}$$

In the xz-plane this looks the same as the above force-free solution but there is no distortion of the arcs in the y-direction.

6.3 Alfvén waves

In MHD the perturbations propagate in form of **magnetohydrodynamic waves**. They are expressed in terms of two characteristic speeds. The **sound speed** is the speed of sound waves

$$c_{\rm s} = \sqrt{\gamma p/\rho} = \sqrt{\gamma k_{\rm B} T/m}\,. \tag{6.51}$$

Here γ is the adiabatic index of the gas, which for an ideal gas can be given as $\gamma = (N+2)/N$, where N is the number of the degrees of freedom of the gas. [$N = 3$ and $\gamma = 5/3$ for a monoatomic ideal gas.] The **Alfvén speed** is the speed of the Alfvén waves in the direction of the magnetic field

$$v_{\rm A} = \sqrt{\frac{B^2}{\mu_0 \rho}} = \sqrt{\frac{2P_B}{\rho}}\,. \tag{6.52}$$

A combination of these speeds is the **magnetosonic speed**, which is the speed of compressional waves, so-called magnetosonic waves, perpendicular to the magnetic field

$$v_{\rm ms} = \sqrt{c_{\rm s}^2 + v_{\rm A}^2}\,. \tag{6.53}$$

6.3.1 Dispersion equation of MHD waves

Consider a compressible, non-viscous, perfectly conductive fluid in a magnetic field. This is described by the following set of equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{6.54}$$

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla P + \mathbf{J}\times\mathbf{B} \tag{6.55}$$

$$\nabla P = c_{\rm s}^2 \nabla \rho \tag{6.56}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \tag{6.57}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{6.58}$$

$$\mathbf{E} + \mathbf{v}\times\mathbf{B} = 0\,. \tag{6.59}$$

From these we can eliminate $\mathbf{J}$, $\mathbf{E}$, and P, ending up with seven equations for seven variables ($\rho,\ \mathbf{v},\ \mathbf{B}$).

Assume that in equilibrium the density ρ_0 is constant and $\mathbf{v} = 0$. Furthermore, let the background magnetic field $\mathbf{B}_0$ be uniform. Considering small perturbations to the variables we can linearize the equations by picking up the first order terms and find an equation for the velocity perturbation $\mathbf{v}_1$

$$\frac{\partial^2 \mathbf{v}_1}{\partial t^2} - c_{\rm s}^2 \nabla(\nabla\cdot\mathbf{v}_1) + \mathbf{v}_{\rm A}\times\{\nabla\times[\nabla\times(\mathbf{v}_1\times\mathbf{v}_{\rm A})]\} = 0\,, \tag{6.60}$$

where we have introduced the Alfvén velocity as a vector

$$\mathbf{v}_{\rm A} = \frac{\mathbf{B}_0}{\sqrt{\mu_0\rho_0}}\,. \tag{6.61}$$

Considering plane wave solutions in the form $\mathbf{v}_1(\mathbf{r},t) = \mathbf{v}_1 \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)]$ we get the dispersion equation for ideal MHD waves

$$\begin{aligned} &-\omega^2\mathbf{v}_1 + (c_{\rm s}^2 + v_{\rm A}^2)(\mathbf{k}\cdot\mathbf{v}_1)\mathbf{k} \\ &+(\mathbf{k}\cdot\mathbf{v}_{\rm A})[((\mathbf{k}\cdot\mathbf{v}_{\rm A})\mathbf{v}_1 - (\mathbf{v}_{\rm A}\cdot\mathbf{v}_1)\mathbf{k} - (\mathbf{k}\cdot\mathbf{v}_1)\mathbf{v}_{\rm A})] = 0\,. \end{aligned} \tag{6.62}$$

Exercise

Derive the above dispersion equation.

6.3.2 MHD wave modes

6.3.2.1 Perpendicular propagation

Let $\mathbf{k} \perp \mathbf{B}_0$. Then $\mathbf{k} \cdot \mathbf{v}_\mathrm{A} = 0$ and the dispersion equation reduces to

$$\mathbf{v}_1 = (c_\mathrm{s}^2 + v_\mathrm{A}^2)(\mathbf{k} \cdot \mathbf{v}_1)\mathbf{k}/\omega^2 \,. \tag{6.63}$$

Clearly $\mathbf{k} \parallel \mathbf{v}_1$, and we have found the wave propagating at the magnetosonic speed

$$\omega/k = \sqrt{c_\mathrm{s}^2 + v_\mathrm{A}^2} \,. \tag{6.64}$$

Assuming harmonic behavior also for the magnetic field the convection equation reduces to

$$\omega \mathbf{B}_1 + \mathbf{k} \times (\mathbf{V}_1 \times \mathbf{B}_0) = 0 \,, \tag{6.65}$$

which yields the magnetic field of the wave

$$\mathbf{B}_1 = \frac{V_1}{\omega/k} \mathbf{B}_0 \,. \tag{6.66}$$

This wave is called **compressional Alfvén wave** or **fast Alfvén (MHD) wave**.

6.3.2.2 Parallel propagation

For $\mathbf{k} \parallel \mathbf{B}_0$, the dispersion equation reduces to

$$(k^2 v_\mathrm{A}^2 - \omega^2)\mathbf{V}_1 + \left(\frac{c_\mathrm{s}^2}{v_\mathrm{A}^2} - 1 \right) k^2 (\mathbf{V}_1 \cdot \mathbf{v}_\mathrm{A}) \mathbf{v}_\mathrm{A} = 0 \,. \tag{6.67}$$

Here we find two different wave modes. If $\mathbf{V}_1 \parallel \mathbf{B}_0 \parallel \mathbf{k}$, we get the dispersion equation for the sound wave

$$\omega/k = c_\mathrm{s} \,. \tag{6.68}$$

The second solution is a transversal wave with $\mathbf{V}_1 \perp \mathbf{B}_0 \parallel \mathbf{k}$. Now $\mathbf{V}_1 \cdot \mathbf{v}_\mathrm{A} = 0$ and we find a wave propagating at the Alfvén speed

$$\omega/k = v_\mathrm{A} \,. \tag{6.69}$$

The magnetic field component of the wave is now

$$\mathbf{B}_1 = -\frac{\mathbf{V}_1}{\omega/k} B_0 \,. \tag{6.70}$$

The wave magnetic field is perpendicular to the background field. This mode does not perturb the density or pressure. The mode causes shear stress on the magnetic field $(\nabla \cdot (\mathbf{BB})/\mu_0)$ and is called **shear Alfvén wave**.

6.3.2.3 Propagation at oblique angles

The Alfvén waves can propagate at all angles with respect to the background magnetic field. To investigate this we have to insert the angles into the dot products of the dispersion equation. Select the z-axis parallel to $\mathbf{B}_0$ and the x-axis so that $\mathbf{k}$ is in the xz-plane. Denote the angle between $\mathbf{k}$ and $\mathbf{B}_0$ by θ. Then the dispersion equation becomes

$$\begin{aligned} V_{1x}(-\omega^2 + k^2 v_{\rm A}^2 + k^2 c_{\rm s}^2 \sin^2\theta) + V_{1z}(k^2 c_{\rm s}^2 \sin\theta\cos\theta) &= 0 \quad (6.71)\\ V_{1y}(-\omega^2 + k^2 v_{\rm A}^2 \cos^2\theta) &= 0 \quad (6.72)\\ V_{1x}(k^2 c_{\rm s}^2 \sin\theta\cos\theta) + V_{1z}(-\omega^2 + k^2 c_{\rm s}^2 \cos^2\theta) &= 0\,. \quad (6.73) \end{aligned}$$

The y-component yields a linearly polarized mode with the phase velocity

$$\omega/k = v_{\rm A}\cos\theta\,. \tag{6.74}$$

This is an extension of the shear Alfvén wave. It cannot propagate strictly perpendicular to the magnetic field. The non-trivial solutions of the remaining pair of equations are

$$\left(\frac{\omega}{k}\right)^2 = \frac{1}{2}(c_{\rm s}^2 + v_{\rm A}^2) \pm \frac{1}{2}\sqrt{(c_{\rm s}^2 + v_{\rm A}^2)^2 - 4c_{\rm s}^2 v_{\rm A}^2\cos^2\theta}\,. \tag{6.75}$$

These are called **fast** (+) and **slow** (–) Alfvén (or MHD) waves.

6.4 Shock waves

Quite often the disturbances on the solar surface and above develop to shock structures. Shocks are particularly important in the corona and solar wind. When dealing with shocks one has to be careful with the frame of reference. For example, a shock in the solar wind may propagate either backward or forward in the solar wind frame. But because the solar wind speed is supersonic, the shock in both cases most likely passes a spacecraft in downwind direction. On the other hand a shock forming in front of a planet, e.g., the terrestrial bow shock, is stationary in the rest frame of the planet but a spacecraft may cross it in either direction. In the solar wind frame the bow shock propagates fast against the solar wind flow.

6.4.1 Hydrodynamic shocks

We start studying the shocks of the ordinary hydrodynamics. Small-amplitude sound waves propagate retaining their shape. If the amplitude for some reason becomes large, nonlinear terms in the wave equation make the crest of the wave to move faster than the trough. The wave steepens and finally the

excess energy of the wave is dissipated as heat. Mathematically the steepening is given by a convective term. If the convection and dissipation balance each other, the shock waves can propagate long distances.

We consider the shock wave in in its own rest frame. We assume it to be very thin in the relevant hydrodynamics scales. The "ahead" or "upstream" region is denoted by subscript 1 and the "behind" or "downstream" by 2. The internal energy per unit mass is $U = P/[(\gamma - 1)\rho]$. We choose the special frame of reference, where the flow of the gas is parallel to the shock normal. Then, conservation of mass, momentum, and energy gives the relationships

$$\begin{aligned} \rho_2 v_2 &= \rho_1 v_1 && (6.76)\\ P_2 + \rho_2 v_2^2 &= P_1 + \rho_1 v_1^2 && (6.77)\\ P_2 v_2 + (\rho_2 U_2 + \frac{1}{2}\rho_2 v_2^2) v_2 &= P_1 v_1 + (\rho_1 U_1 + \frac{1}{2}\rho_1 v_1^2) v_1 \,. && (6.78) \end{aligned}$$

These equations are often written using the notation $[f] = f_1 - f_2$, e.g., $[\rho v] = 0$. They are called **Rankine-Hugoniot relations** and they can be expressed as jumps of various parameters over the shock layer, such as

$$\begin{aligned} \frac{\rho_2}{\rho_1} &= \frac{(\gamma+1)M_1^2}{2 + (\gamma-1)M_1^2} && (6.79)\\ \frac{v_2}{v_1} &= \frac{2 + (\gamma-1)M_1^2}{(\gamma+1)M_1^2} && (6.80)\\ \frac{P_2}{P_1} &= \frac{2\gamma M_1^2 - (\gamma-1)}{\gamma+1}\,, && (6.81) \end{aligned}$$

where $M_1 = v_1/c_{s1}$ is the **sonic Mach** number on the upstream side ($c_{s1} = \sqrt{\gamma P_1/\rho_1}$). Thermodynamics tells us that the entropy, $S = c_V \log P/\rho^\gamma$, must increase, $S_2 \geq S_1$. The equality holds for same conditions on both sides, i.e., when the shock ceases to to be there. From these conditions we can infer the following properties of hydrodynamic shocks

1. $M_1 \geq 1$, i.e., $v_1 \geq c_{s1}$ ahead the shock
2. $v_2 \leq c_{s2}$, flow is subsonic behind the shock
3. $P_2 \geq P_1$ and $\rho_2 \geq \rho_1$, the shock is compressive
4. $v_2 \leq v_1$ and $T_2 \geq T_1$, the flow is slowed down and the gas heated up
5. $1 \leq \rho_2/\rho_1 < (\gamma+1)/(\gamma-1)$, the maximum density ratio is $(\gamma+1)/(\gamma-1)$, but the pressure increases $\propto M_1^2$.

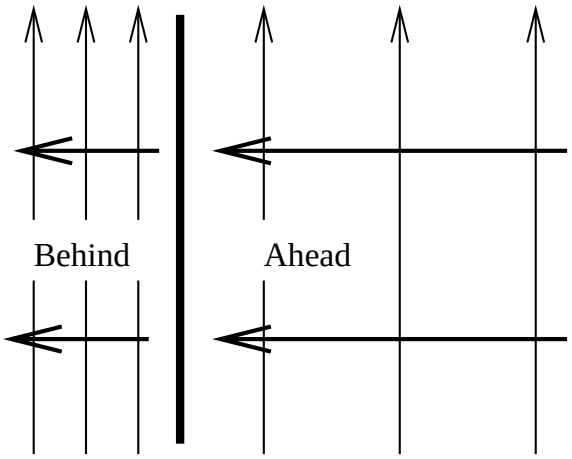

Figure 6.4: Perpendicular shock. The thick arrows indicate the flow direction and the thinner lines the magnetic field direction.

6.4.2 Perpendicular MHD shock

In MHD the magnetic field complicates the shock structure considerably. There are three different wave modes: the slow, intermediate (Alfvén), and fast mode. The Alfvén mode can have a large amplitude without steepening and does not form a shock, whereas the slow and fast modes can develop to shocks.

Also the angle (θ) between the shock normal and the magnetic field is important. The simplest case is the perpendicular shock ($\theta = \pi/2$). Its magnetic field lines are in a plane parallel to the shock and the shock is similar to the hydrodynamic one.

Now the Rankine-Hugoniot relations are derived for mass, momentum, energy, and magnetic flux conservation

$$\rho_2 v_2 = \rho_1 v_1 \tag{6.82}$$

$$P_2 + \rho_2 v_2^2 + \frac{B_2^2}{2\mu_0} = P_0 + \rho_1 v_1^2 + \frac{B_1^2}{2\mu_0} \tag{6.83}$$

$$\left(P_2 + \frac{B_2^2}{2\mu_0}\right) v_2 + \left(\rho_2 U_2 + \frac{1}{2}\rho_2 v_2^2 + \frac{B_2^2}{2\mu_0}\right) v_2 = \left(P_1 + \frac{H_1^2}{2\mu_0}\right) v_1 + \left(\rho_1 U_1 + \frac{1}{2}\rho_1 v_1^2 + \frac{B_1^2}{2\mu_0}\right) v_1 \tag{6.84}$$

$$B_2 v_2 = B_1 v_1 \,. \tag{6.85}$$

From these we find the jumps

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} = \frac{1}{X} \tag{6.86}$$

$$\frac{B_2}{B_1} = X \tag{6.87}$$

$$\frac{P_2}{P_1} = \gamma M_1^2 \left(1 - \frac{1}{X}\right) - \frac{1 - X^2}{\beta_1} \,, \tag{6.88}$$

where $X = \rho_2/\rho_1$ is the positive root of

$$2(2-\gamma)X^2 + [2\beta_1 + (\gamma-1)\beta_1 M_1^2 + 2]\gamma X - \gamma(\gamma+1)\beta_1 M_1^2 = 0 \,. \tag{6.89}$$

Note that, in addition to the upstream Mach number, the upstream plasma beta β_1 is a characteristic parameter of the shock.

The perpendicular shock has the following properties

1. Because $1 < \gamma < 2$, (6.89) has only one positive root.
2. The magnetic field reduces X below the hydrodynamic value.
3. The shock is compressive ($X \geq 1$).
4. Upstream flow is super-fast-magnetosonic: $v_1 \geq v_{\rm ms} \equiv \sqrt{c_{\rm s1}^2 + v_{\rm A1}^2}$
5. The magnetic compression is limited to $1 < B_2/B_1 < (\gamma+1)/(\gamma-1)$.

6.4.3 Oblique shocks

For other angles θ (Fig. 6.5) the Rankine-Hugoniot relations must be written for the two components in the plane determined by $\mathbf{v}$ ind $\mathbf{B}$. It is most convenient to transform to a shock frame of reference, where the upstream field and flow are parallel to each other. Continuity of the transverse electric

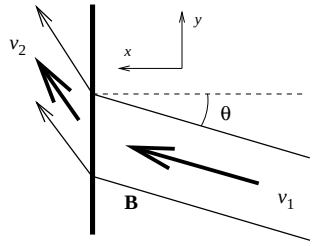

Figure 6.5: The shock geometry for arbitrary orientation.

field implies then that the same holds in the downstream region. This frame is called the de Hoffmann – Teller frame. Using the notations of Figure 6.5, the jump conditions become (exercise):

$$\frac{v_{2x}}{v_{1x}} = \frac{\rho_1}{\rho_2} = \frac{1}{X} \tag{6.90}$$

$$\frac{v_{2y}}{v_{1y}} = \frac{v_1^2 - v_{\rm A1}^2}{v_1^2 - X v_{\rm A1}^2} \tag{6.91}$$

$$\frac{B_{2x}}{B_{1x}} = 1 \tag{6.92}$$

$$\frac{B_{2y}}{B_{1y}} = \frac{(v_1^2 - v_{\rm A1}^2)X}{v_1^2 - X v_{\rm A1}^2} \tag{6.93}$$

$$\frac{P_2}{P_1} = X + \frac{(\gamma-1)X v_1^2}{2c_{\rm s1}^2}\left(1 - \frac{v_2^2}{v_1^2}\right), \tag{6.94}$$

Now the compression ratio $X = \rho_2/\rho_1$ is found as a solution of

$$(v_1^2 - X v_{\rm A1}^2)^2\{X c_{\rm s1}^2 + \frac{1}{2}v_1^2\cos^2\theta[X(\gamma-1) - (\gamma+1)]\} \tag{6.95}$$
$$+\frac{1}{2}v_{\rm A1}^2 v_1^2 \sin^2\theta\, X\{[\gamma + X(2-\gamma)]v_1^2 - X v_{\rm A1}^2[(\gamma+1) - X(\gamma-1)]\} = 0\,.$$

The slow, Alfvén, and fast mode discontinuities are shown in Figure 6.6. It is important to note that the parallel (to shock normal) component of the magnetic field (B_x) does not change over the shock. The intermediate (Alfvén) mode depicted in the figure is not really a shock (it is not a solution of the above equations). It is called the **rotational discontinuity**. In this solution, $X = 1$ and B_y and $v_y$change sign over the discontinuity. An example of that is a reconnecting current sheet discussed a little later. In MHD, a so-called intermediate shock exists as a solution to the Rankine-Hugoniot equations, but it has the property that the solution is non-evolutionary; it means that even a small perturbation can disintegrate the solution to two or more discontinuities. It's unclear, whether the intermediate mode can exist in space plasmas.

The slow and fast shocks have following properties

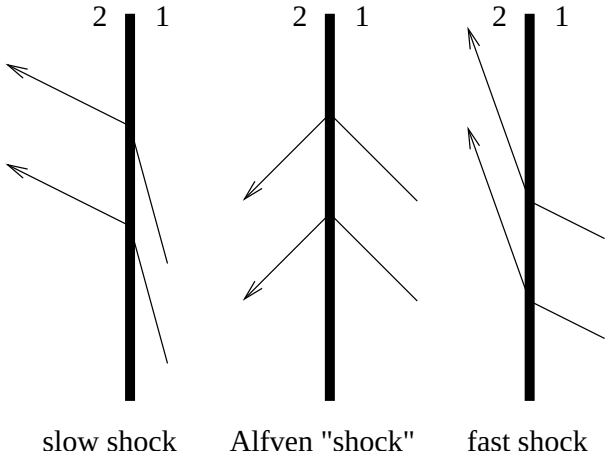

Figure 6.6: The magnetic field lines through MHD shocks. Recall that the Alfvén "shock" is not a shock.

1. They are compressive
2. B_x remains unchanged over the shock
3. They conserve the sign of B_y
4. At the slow shock $B_2 < B_1$
5. At the fast shock $B_2 > B_1$
6. v_{1x} exceeds the slow/fast speed ahead the shock while v_{2x} is smaller than the slow/fast speed behind the shock
7. $v_{2x} < v_{1x}$
8. At the limit $B_x \to 0$, the fast shock becomes perpendicular shock whereas the slow shock becomes a tangential discontinuity ($v_x \to 0$) with arbitrary jumps in v_y and B_y subject to total pressure balance over the shock.

The limiting case $\theta = 0$ is called **the parallel shock** and the directions in between are often denoted as quasi-parallel and quasi-perpendicular. The quasi-parallel shocks are more complicated than the perpendicular shocks because individual particles can be reflected from the shock and move long distances upstream leading to instabilities not described by the MHD theory.

6.5 Instabilities

Stability and the loss of it, the instability, are important elements of all plasma physics, including the solar physics. In this section we summarize

some of the basic concepts and ideas that are useful to understand instabilities in solar physics.

6.5.1 Concept of instability

The basic idea of instability is illustrated by a mechanical analogue of a ball which may be in a valley, on the top of a hill, on a plateau, or in a localized "crater" of a mountain. The physical system consists of both the ball and the landscape. The ball could represent the plasma particles and the landscape the fields confining the plasma, or something more abstract.

If the ball is in such a deep valley that no realistec perturbation can lift it away, the system is **stable**. After the initial perturbation the ball returns back to its equilibrium position. It may oscillate around the bottom of the valley for a long time if the damping of the oscillation is slow.

If the ball is at the summit of the mountain, any perturbation moves it permanently away from the hill and the system is **unstable**. After the perturbation the system finds a new equilibrium which may be completely different from the original. The question, how did the ball get to the unstable position in the first place may look tricky. However, in physical systems the mountain may evolve, e.g., as changing magnetic field configuration.

A ball on a plateau represents a **meta-stable** (or marginally stable) state. The perturbation puts the ball in motion but the system looks the same until the ball reaches either an uphill or a downhill slope. E.g., the solar wind-magnetosphere system is never in stable equilibrium but may look the same for a long time. Sooner or later the system, however, evolves to an unstable state.

Finally, the stability often depends on the strength of the perturbation. A ball in the crater remains there unless the perturbation is strong enough to pull the ball from the crater and let it roll down the hill. This case can, somewhat loosely, be called a **nonlinear instability**.

Our ability for a rigorous analysis of plasma stability is often limited to the linear regime, i.e., we can determine whether plasma is stable or unstable to small perturbations. If we can linearize the relevant plasma equations, we often can write the dispersion equation in the form

$$\omega = \omega(k)\,, \tag{6.96}$$

as was the case with the Alfvén waves in section 6.3. Now the frequency is, in general, a complex number $\omega = \omega_\mathrm{r} + \mathrm{i}\omega_\mathrm{i}$. Assuming the sign convention $A(t) \propto \exp(-\mathrm{i}\omega t)$ for the time dependence of the wave amplitude, $\omega_\mathrm{i} < 0$ corresponds to a damped (stable) solution and $\omega_\mathrm{i} > 0$ a growing (unstable) solution.

For a small damping rate ($|\omega_\mathrm{i}| \ll \omega_\mathrm{r}$) the perturbation is a **normal mode** of the plasma but often the damping takes place very quickly and the mode is **overdamped**. For unstable solutions it is impossible to say in advance to how large amplitude a wave can grow. If nothing quenches the growth, the system develops to a major configurational change. However, often the growth leads to a situation where some plasma particles start to interact more strongly with the growing wave, e.g., in the form of heating. An example, is the so-called **quasi-linear saturation** that is a frequent phenomenon in microscopic instabilities described by the Vlasov theory (see again the lecture notes in Advanced Space Physics).

A way of categorizing plasma instabilities is to divide them between microscopic (kinetic) and macroscopic (configurational) instabilities. A **macroinstability** is something that can be described by macroscopic equations in the configuration space. A **microinstability** takes place in the ($\mathbf{r}$,$\mathbf{v}$)-space and depends on the actual shape of the distribution function $f(\mathbf{r}, \mathbf{v}, t)$, whose evolution is described by the Boltzmann or Vlasov equation. Here we limit the discussion to macroscopic phenomena.

The instabilities do not arise without **free energy**. The free energy may come from the magnetic configuration, from anisotropic plasma pressure, from streaming of plasma particles with respect to each other, etc. Recognition of the free energy source is essential to understanding the instability because different free energy sources may lead to quite different consequences.

6.5.2 Examples of macroscopic instabilities

6.5.2.1 Rayleigh-Taylor instability

The Rayleigh-Taylor instability is perhaps the simplest example of macroscopic instabilities caused by plasma inhomogeneities. However, its analysis requires consideration of electron and ion dynamics separately, and thus the treatment is beyond the MHD theory (see, e.g., Lectures in Advanced Space Physics; the elements of the required tools can be found in any basic plasma physics text, e.g., Koskinen, Johdatus plasmafysiikkaan ja sen avaruussovellutuksiin, Limes ry., 2001).

What we need here is elementary knowledge of particle drift motion. In a homogeneous magnetic field ($\mathbf{B}$) the charges make circular motion about the field lines. Positively charged particles rotate in the left-hand sense and negative charges (electrons) in the right hand sense when looked into the direction of the magnetic field. Assume then that electric ($\mathbf{E}$) and/or gravitational ($\mathbf{g}$) fields are imposed on the background magnetic field. They put the charged particles into drift motion perpendicular to the magnetic

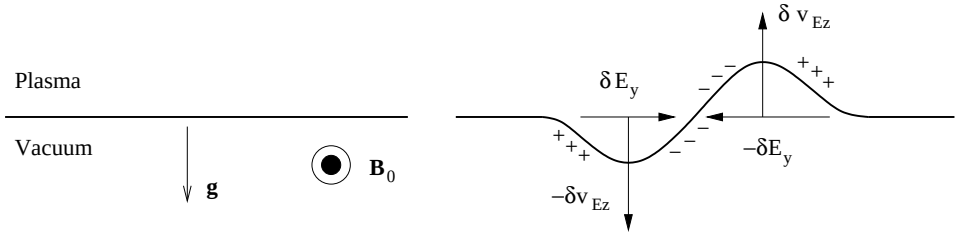

Figure 6.7: Principle of the Rayleigh-Taylor instability. The forming bubbles become unstable against the same instability and form smaller structures and gradually disappear.

field with velocities

$$\mathbf{v}_{\rm e} = \frac{\mathbf{E}\times\mathbf{B}}{B^2} \tag{6.97}$$

$$\mathbf{v}_{{\rm g}\alpha} = \frac{m_\alpha\,\mathbf{g}\times\mathbf{B}}{q_\alpha\,B^2}\,, \tag{6.98}$$

where α labels the different particle species. These are called the **electric** (or E×B) and **gravitational drifts**. They are averages over the gyroradii of the particles. Note that the electric drift is the same for all charged particles and thus does not carry net current. The gravitational drift depends both on the charge (with sign) and mass of the particles. Consequently it carries electric current and due to the larger mass of ions than electrons, the ions are the main current carriers.

Consider heavy plasma supported against the gravitational force by magnetic field. This kind of setting is likely to occur in the solar atmosphere. Let the boundary and the magnetic field be in the (x, y)-plane, $\mathbf{B}_0 = B_0\mathbf{e}_x$, and let the gravitational acceleration $\mathbf{g} = -g\mathbf{e}_z$ act downward and density gradient $\nabla n_0 = [\partial n_0(z)/\partial z]\mathbf{e}_z$ point upward.

Consider next a small sinusoidal perturbation to this boundary. The gravitational field causes ion drift and current in the $-y$-direction. This leads to an electric field perturbation that is in the $+y$-direction in the region where there is plasma below the original boundary and in the opposite direction in the part where the perturbation has lifted the plasma boundary upward. Now, in the downward perturbed region the E×B-drift is downward and in the upward perturbed region upward. Thus the E×B-drift enhances the original perturbation and the system is unstable. The result is that the gravitationally supported plasma falls down and dilute bubbles rise up.

Assuming collisionless plasma the dispersion equation can be written in the form

$$\frac{\omega_{\rm ci}}{\omega}\frac{1}{k_\perp L_n}\left(1-\frac{\omega}{\omega+gk_\perp/\omega_{\rm ci}}\right)-\left(1+\frac{m_{\rm i}}{m_{\rm e}}\right)\frac{\omega_{\rm ci}^2}{\omega^2}\frac{k_\parallel^2}{k_\perp^2}+1=0\,, \tag{6.99}$$

where $\omega_{\rm ci}$ is the ion cyclotron frequency, the perpendicular and parallel directions are with respect to the magnetic field, and L_n the undisturbed density scale length

$$L_n^{-1} = \frac{d \ln n_0(z)}{dz} > 0\,. \tag{6.100}$$

To find exact solutions to this equation is a little tedious. The highest growth rate is found for exactly perpendicular propagation ($k_\parallel = 0$) because then the electric field will lead to largest vertical drift. Assuming further weak gravitational effect ($\omega \gg k_\perp g/\omega_{\rm ci}$) the first order solution is

$$\omega^2 = -\frac{g}{L_n}\,, \tag{6.101}$$

which includes a purely growing solution with growth rate

$$\gamma_{\rm 0rt} = \left(\frac{g}{L_n}\right)^{1/2}\,. \tag{6.102}$$

Expanding the dispersion equation to the second order in $k_\perp g/(\omega_{\rm ci}\omega)$ we could find an oscillating solution but still the growth rate is much larger than the oscillation frequency. If we, on the other hand, would let $k_\parallel \neq 0$, we would find that there are solutions only in a very narrow cone around the perpendicular direction.

In the partially ionized parts of the solar atmosphere the ion-neutral collisions may become important. Electrons can still be taken as collision-free but the ion-neutral collision rate $\nu_{\rm in}$ and the pressure force must be taken into account. The analysis becomes slightly more complicated than in the collisionless case. The collisional growth rate is found to be

$$\gamma_{\rm rt} = \gamma_{\rm 0rt}\left[1 - \exp\left(-\frac{\gamma_{\rm 0rt}}{\nu_{\rm in}}\right)\right]\,, \tag{6.103}$$

which at the limit of vanishing collisions yields $\gamma_{\rm 0rt}$. At the limit of large collision frequency the growth rate is

$$\gamma_{\rm rtn} = \frac{g}{\nu_{\rm in} L_n} = \frac{\gamma_{\rm 0rt}^2}{\nu_{\rm in}}\,. \tag{6.104}$$

Note that in astrophysics the Rayleigh-Taylor instability is often called the **Kruskal-Schwarzschild instability**.

6.5.2.2 Kelvin-Helmholtz instability

The Kelvin-Helmholtz instability has its neutral fluid counterpart in the wind blowing over water. The wind causes ripples on the surface of water.

The same happens in collisional and collisionless plasmas, e.g., the solar wind flow along a planetary magnetopause causes K-H instability and propagation of K-H waves on the bounding surface.

We consider the instability in the ideal MHD scale. At the narrow boundary there may be some viscous effects (e.g., so-called anomalous viscosity due to wave-particle interactions) but they are higher order corrections to this discussion. Let the magnetic field and the flow be tangential to the boundary and let the velocity change, or even reverse, across the boundary. Assume further scalar pressure, linearize around background $\mathbf{B}_0$ and n_0, and consider small displacements $\delta\mathbf{x}$ defined by $\delta\mathbf{v} = d\,\delta\mathbf{x}/dt$. The strategy is to linearize the induction and momentum equations and find an expression for $\delta\mathbf{x}$ (see Lectures in Advanced Space Physics).

Let the boundary be in the (x, z)-plane and assume plane wave solutions for both $\delta\mathbf{x}$ and to the perturbation of the total pressure $\delta P_{\rm tot}$ with wave number $\mathbf{k} = k_x\mathbf{e}_x + k_z\mathbf{e}_z$ and frequency ω. Now the displacement of the boundary is given by

$$\delta\mathbf{x} = \frac{\delta P_{\rm tot}}{m_{\rm i} n_0[\omega^2 - (\mathbf{k}\cdot\mathbf{v}_{\rm A})^2]} \tag{6.105}$$

and $\delta P_{\rm tot}$ by

$$\delta P_{\rm tot} = P_0 \exp(-k|y|)\exp[-i(\omega t - k_x x - k_z z)]\,, \tag{6.106}$$

where $k^2 = k_x^2 + k_z^2$. The y-dependence is selected to make the wave evanescent outside the boundary because there is free energy only at the boundary.

We consider the boundary as a tangential discontinuity in a pressure balance, i.e., a boundary through which the there is no plasma flow and $B_{\rm n} = 0$ but $\mathbf{V}_{\rm t}$, $\mathbf{B}_{\rm t}$, and n, and P may be jump over the boundary. We further require that the normal component of the displacement must be continuous. Denote the two sides of the boundary by 1 and 2 and select a coordinate system where the plasma stream has velocity $\mathbf{v}_0$ in region 1 and the fluid in region 2 is in rest. Because the total pressure $(P + B^2/2\mu_0)$ is continuous, the continuity of the normal component of the displacement yields the dispersion equation for the K-H waves

$$\frac{1}{n_{02}[\omega^2 - (\mathbf{k}\cdot\mathbf{v}_{\rm A2})^2]} + \frac{1}{n_{01}[(\omega - \mathbf{k}\cdot\mathbf{v}_0)^2 - (\mathbf{k}\cdot\mathbf{v}_{\rm A1})^2]} = 0\,. \tag{6.107}$$

The unstable modes are Alfvén waves. The dispersion equation has an unstable solution

$$\omega_{\rm kh} = \frac{n_{01}\mathbf{k}\cdot\mathbf{v}_0}{n_{01} + n_{02}} \tag{6.108}$$

corresponding to the complex root

$$(\mathbf{k}\cdot\mathbf{v}_0)^2 > \frac{n_{01} + n_{02}}{n_{01}n_{02}}[n_{01}(\mathbf{k}\cdot\mathbf{v}_{\rm A1})^2 + n_{02}(\mathbf{k}\cdot\mathbf{v}_{\rm A2})^2]\,. \tag{6.109}$$

The K-H instability occurs thus for sufficiently large $\mathbf{v}_0$. For small $\mathbf{v}_0$ k would have to be too large, i.e., the wavelength too short for MHD.

6.5.2.3 Flux tube instabilities

Flux tube instabilities are particularly important in solar physics (as well as in many laboratory devices). Their stability analysis is usually based on the **energy principle**. It means that the energy content of the system is calculated in presence of small perturbations. If the energy variation ΔW is negative, the system is unstable. The calculations are usually quite cumbersome.

There are three basic modes of instabilities in flux tubes carrying a current along the tube axis. It is a useful **exercise** to draw the different configurations.

The **pinch instability** shows that a linear pinch is unstable for any localized inhomogeneity. Clearly the azimuthal field produced by the current is increased in regions where the tube is pinched and decreased outside . Thus the pinching self-amplifies the instability. This instability can take place in the active regions of the solar corona.

The **kink instability** resembles the pinch effect. If the tube is kinked, there is an inward pressure gradient in the inner curve and outward pressure gradient in the outer curve. Again the perturbation is self-amplifying, i.e., unstable. Also this process may be excited in the solar corona in association to reconnection of various current sheet or flux tube configurations.

The **helical instability** is probably very common in the solar corona where strongly twisted flux tubes occur frequently. This instability requires strong enough field-aligned current to flow through the structure. We will return to this type of phenomena in the context of solar prominences.

6.6 Magnetic reconnection

The concept of reconnection was introduced by Giovanelli in the 1940's to explain rapid energy release in solar coronal loops. In ideal MHD the magnetic field and plasma flow are frozen-in to each other. This means that plasma elements, which are on the same magnetic field line, remain magnetically connected to each other whereas plasma elements not magnetically connected to each other remain on different field lines. Whenever such a connectivity changes, we can speak about reconnection. This is the most general viewpoint to reconnection. As a local electric field along the magnetic field can break the frozen-in flow, this "definition" of reconnection does not require a current sheet between the reconnecting fields.

Being closely associated to frozen-in flow the reconnection is often described in terms of moving magnetic field lines that become cut and reconnected by some, often unknown, physical mechanism. While the picture of spaghetti-like moving magnetic field lines often is a very useful picture, it sometimes seems to take over the real physics. This led Alfvén to denounce his own frozen-in concept as pseudopedagogical. He reminded that there is no need to assume moving field lines because in a single-particle picture all particles drift across the magnetic field lines and if the magnetic field configuration changes, this is the result of changing or moving sources of the magnetic field, the currents. This is an extreme view which does not make the physical analysis of reconnection easier. A more pragmatic, but still physical, point of view is to use the frozen-in picture wherever it is valid and accept that in reconnection the magnetic connectivity changes, but to keep in mind that the physical mechanisms that violate the frozen-in flow are essential.

In MHD the evolution of the magnetic field is described by the induction equation (6.12). As seen above the collisional diffusivity in solar plasmas is typically very small. The diffusion may, however, not need to be determined by classical collisions. Also anomalous resistivity due to wave-particle interactions or plasma turbulence may give rise to finite η. Note also that the induction equation is derived assuming the simple MHD form of Ohm's law. If other terms in generalized Ohm's law (6.7) are taken into account, also the induction equation becomes more complicated. The Hall term $(\mathbf{J} \times \mathbf{B})$ thaws the ions but not the electrons, whereas the electron flow may thaw due to the electron pressure term $(\nabla \cdot \mathcal{P}_{\rm e})$ or inertial effects $(\propto \partial \mathbf{J}/\partial t)$.

The convective term $\nabla \times (\mathbf{v} \times \mathbf{B})$ describes the ideal, never breaking, frozen-in flow. Physically the magnetic flux is rearranged by the diffusion process. The diffusion time is given by $\tau_{\rm d} = L^2/\eta$ where L is the gradient scale length. In space plasmas L is in general large and η small making the diffusion a very slow process, indeed. However, when two ideal plasma systems flow toward each other with different magnetic field orientation a thin current sheet develops over which the gradient becomes sharper. Thus the diffusion rate increases, and if some physical process simultaneously enhances the diffusion coefficient η, diffusion may rearrange the magnetic field very quickly. This is what is usually understood by reconnection. In this sense it is a limiting case of diffusion which can break a thin current sheet that separates plasmas of different magnetic connectivity. This viewpoint to reconnection is much more limited than the general idea of any mechanism breaking the frozen-in flow. However, at the same time it is technically more practical and applicable to many observed configurations in space (solar flares, planetary magnetopauses, magnetospheric tails, etc.).

Reconnection appears to take place in explosive manner in solar eruptions, i.e, eruptive prominences, flares, coronal mass ejections. However,

the transition from slow diffusion to fast reconnection is among the most challenging problems in space physics. In quasi-static two-dimensional configurations (such as current sheets) reconnection can be described, at least qualitatively, rather well but three-dimensional and time-dependent reconnection are much more difficult. But even in two dimensions the microscopic diffusion and thawing of the electrons are not yet completely understood.

6.6.1 Sweet-Parker model

Assuming oppositely ($\pm x$) directed straight magnetic fields on both sides of a current layer it is easy to calculate at which speed the plasma can flow toward the boundary for a given diffusivity η. The separating current flows in the $+y$-direction. In equilibrium $\partial\mathbf{B}/\partial t = 0$ and thus

$$\nabla \times \mathbf{E} = \frac{\partial E_y}{\partial z} = 0\,, \tag{6.110}$$

i.e., E_y constant. Far from the diffusion region

$$E_y = vB_0\,. \tag{6.111}$$

At the current sheet $B = 0$ and Ohm's law gives

$$\mathbf{E} = \mathbf{J}/\sigma \quad \Rightarrow \quad E_y = j_y/\sigma\,. \tag{6.112}$$

Let the thickness of the current sheet be $2l$. Ampère's law yields now

$$j_y = \frac{B_0}{\mu_0 l} \tag{6.113}$$

and the thickness of the current sheet is

$$2l = \frac{2}{\mu_0 \sigma v}\,. \tag{6.114}$$

Thus there is a simple relation between the inflow speed and current sheet thickness

$$l = \eta/v\,. \tag{6.115}$$

With increasing inflow speed the current layer finally becomes so thin that the MHD picture is no more valid. However there is another problem here. Even if the diffusion would be able to consume the magnetic flux, what happens to the plasma piling up at the current sheet?

The first attempts to solve this question were made (independently) by *Sweet* and *Parker* in the late 1950's. They considered a geometry given in figure 6.8. The length of the reconnection region $2L$ is assumed to be much longer than its thickness $2l$. Assume, for simplicity, that the inflow

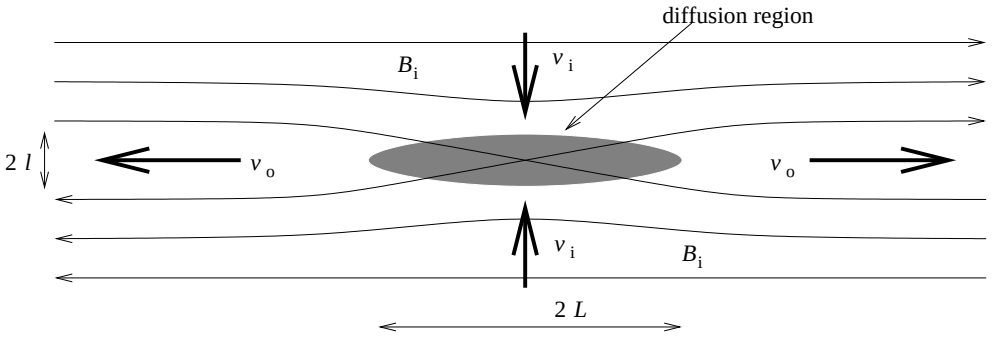

Figure 6.8: Sweet-Parker model of reconnection.

(subscript i) and outflow (subscript o) regions are symmetric. This applies well to the tail current sheet in a magnetosphere but, e.g., in solar flares the asymmetries may be considerably increase the complexity of the problem. The electric field in the inflow is

$$E = v_\mathrm{i} B_\mathrm{i} = v_\mathrm{o} B_\mathrm{o} \,. \tag{6.116}$$

Assume further incompressible flow $\rho_\mathrm{i} = \rho_\mathrm{o} = \rho$. Then conservation of mass implies

$$v_\mathrm{i} L = v_\mathrm{o} l \,. \tag{6.117}$$

Assume that all inflowing electromagnetic energy is converted to the kinetic energy of the outflow. The inflowing Poynting flux is

$$|\mathbf{S}| = |\mathbf{E} \times \mathbf{H}| = \frac{E B_\mathrm{i}}{\mu_0} = \frac{v_\mathrm{i} B_\mathrm{i}^2}{\mu_0} \,. \tag{6.118}$$

The mass flowing into the diffusion region in unit time ρv_i is accelerated to the outflow velocity v_o. Thus the energy change per unit surface is

$$\Delta W = \frac{1}{2} \rho v_\mathrm{i} (v_\mathrm{o}^2 - v_\mathrm{i}^2) \,. \tag{6.119}$$

Equating the energy increase and the Poynting flux and noting that $v_\mathrm{o} \gg v_\mathrm{i}$ we get

$$\frac{v_\mathrm{i} B_\mathrm{i}^2}{\mu_0} = \frac{1}{2} \rho v_\mathrm{i} v_\mathrm{o}^2 \tag{6.120}$$

$\Rightarrow$

$$v_\mathrm{o}^2 = \frac{2 B_\mathrm{i}^2}{\mu_0 \rho} = 2 v_\mathrm{Ai}^2 \,. \tag{6.121}$$

Thus the outflow speed is $\sqrt{2}$ times the Alfvén speed of the inflow. The inflow speed is found from the width of the diffusion region $2l = 2/\mu_0 \sigma v$

$$v_\mathrm{i} = v_\mathrm{Ai} (\sqrt{2}/R_\mathrm{mA})^{1/2} \,, \tag{6.122}$$

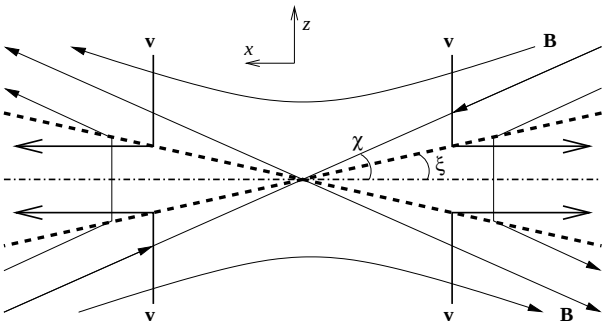

Figure 6.9: Petschek model of reconnection. Most of the plasma is accelerated at the slow mode shocks which make an angle ξ with respect to x-axis.

where $R_{\mathrm{mA}} = \mu_0 \sigma v_{\mathrm{Ai}} L$ is the magnetic Reynolds number calculated for the inflow Alfvén speed (often called the **Lundquist number**). R_{mA} is very large and thus the inflow speed in the Sweet-Parker model is very slow. For example, in solar flares the energy release would take several days, not a few minutes as is observed.

The ratio $v_{\mathrm{o}}/v_{\mathrm{i}}$ is called the **reconnection rate**. It describes the efficiency of reconnection. The reconnection rate in the Sweet-Parker model is of the order of the inflow Alfvén Mach number.

6.6.2 Petschek model

In 1964 *Petschek* realized that not all of inflowing plasma need to pass through the diffusion region. According to figure the flow is deviated also outside the diffusion region at slow mode shocks connected to the diffusion region. Thus the reconnection can process much more magnetic flux and the outflowing plasma can be accelerated to much higher speed.

Diffusion region is essential also in Petscheck's model because reconnection itself with the formation of the X-type magnetic neutral line takes place only in that region. Particles are accelerated on the slow mode shocks and the acceleration rate can be calculated from MHD jump conditions at the shocks. As shown in figure 6.9 the magnetic field decreases across the shock but the flow field becomes more dense. Assuming incompressibility this implies the increase of flow speed in the outflow region. The acceleration depends on the angle ξ between the shock front and x-axis. Another interpretation for the acceleration is that since the magnetic field turns at the shock, there is a current and the acceleration is due to the $\mathbf{J} \times \mathbf{B}$ force.

In the coordinate system of figure 6.9 the shock is stationary but in the plasma frame the shock propagates at the inflow Alfvén speed. This allows us to calculate the angle χ between the x-axis and the inflowing magnetic field

just prior to the shock. In order to have a standing shock in the coordinates of the figure, the component of the inflow velocity in the direction of the shock normal must be the same as the shock velocity in the direction of its own normal in the plasma frame. This implies

$$v_{\rm i} \cos\xi = v_{\rm Ai} \sin(\chi - \xi)\,. \tag{6.123}$$

Assuming a steady state (i.e, E_y = constant) a brief calculation gives the outflow speed

$$v_{\rm o} = v_{\rm Ai} \cos\chi\,, \tag{6.124}$$

which is again of the order of inflow Alfvén speed, this time slightly less. A detailed analysis shows that the ratio between the inflow and outflow speeds is now

$$\frac{v_{\rm i}}{v_{\rm o}} \approx \frac{\pi}{8 \ln R_{\rm mA}}\,. \tag{6.125}$$

Thus the logarithmic dependence on the Reynolds number is much weaker than in the Sweet-Parker model. The estimates for the maximum inflow speed varyf $0.01 - 0.1\, v_{\rm Ai}$, which is much higher than in the Sweet-Parker model ($1/\sqrt{R_{\rm mA}}$). Consequently, the reconnection can handle much more magnetic flux in the Petschek than in the Sweet-Parker model.

Petschek's model has been developed further. In 1970 *Sonnerup* added two fast shocks outside the slow shocks which start the deviation of the plasma flow in the same way as the bow shock in front of a magnetosphere. This way reconnection can handle even more plasma and thus be faster.

There has been some controversy whether the Petschek model describes the correct physics. It gives a description of reconnection that is only very weakly dependent on the properties of the reconnection region and thus on η. At the limit of very low η the Petschek (as well as Petschek-Sonnerup) model predicts that if the inflows are pushed stronger toward each other (larger E_y), the outflow cone will open, not flatten as the intuition and observations indicate. However, it is quite possible that correct microscopic physics enhances the "effective" η to the level that the macroscopic picture of the Petschek model gives a qualitatively correct picture.

In 1986 *Priest and Forbes* constructed a more general mathematical description of MHD reconnection structures of which the Sweet-Parker and Petschek-Sonnerup models are special cases. In this and several other numerical analyses and simulations the boundary conditions have been found to play a decisive role. It may well be the case that the external circumstances dictate if and how the reconnection will take place whereas the local physics of the diffusion region only has to adjust to tear the current sheet and dissipate as much magnetic energy as is required.

To produce the Petschek-type reconnection in numerical simulations of, e.g., the magnetospheric tail current sheet is not trivial. In principle a

numerical MHD code has always some diffusivity (and thus, resistivity) due to the finite sizes of the computing grid and the time-steps. In order to make a realistic magnetotail simulation the resistivity has to be made as small as possible. However, this makes the current sheet region more ideal and harder to reconnect, unless the resistivity (or η) is artificially enhanced in the diffusion region. Thus we find something like a contradiction. In order to have Petschek-type reconnection we must add physics that is different from the assumptions of the model. Even if we found experimentally the shock pattern predicted by the Petschek-Sonnerup description, it would not tell us much of the **physical mechanism** how reconnection takes place because the Petschek-Sonnerup approach is practically independent of the physics of the diffusion region. On the contrary, the importance to determine what causes the enhanced η in the diffusion region, or alternatively the role of other terms in generalized Ohm's law, becomes more urgent.

Exercises

1. Show that in magnetic field configuration

$$B_x = y,\ B_y = x$$

the magnetic pressure and tension forces balance each other. Thus the X-line is stable.

   (a) Stretch the above configuration in y-direction:

$$B_x = y,\ B_y = \alpha^2 x$$

where $\alpha^2 > 1$. Show that now there are net forces toward the X-line along the x-axis and away from the X-line along the y-axis.

2. Show that in the Sweet-Parker model half of the inflowing magnetic energy is converted to heat and the second half to particle acceleration.

Chapter 7

Solar magnetism

The magnetic field of the Sun is very complicated both in time and space. It was first found in sunspots by *Hale* in 1908. Although we can today measure much weaker magnetic fields on the Sun, the sunspots have retained a central role in studies of solar magnetism. Thus we discuss first the phenomenology of sunspots. After a century of intensive study we still lack an accurate physical description of the generation and evolution of the solar magnetic field. The theory of magnetic field generation is a difficult piece of physics and we can describe only some of the basic features of the solar dynamo theory below. For an interested student a recommendable source is M. R. E. Proctor and A. D. Gilbert (eds.), Lectures on Solar and Planetary Dynamos, Cambridge University Press, 1994. As the solar dynamo problem is under active study, during the last ten years a lot of details have been clarified further, but this source is still a valid starting point in the topic.

7.1 Sunspots

There are written records of sunspots at least from the year 300 BC. Their breakthrough into science took place when *Galileo* began observing the Sun with his telescope in 1610, after which the number of sunspots has been counted more or less regularly. As soon as Galileo and others understood that it really was a question of spots on the surface of the Sun, it became clear that the Sun rotates and only some 20 years later the differential character of the rotation was established. Kirchhoff's emission laws explain that the spots are darker because they have a lower temperature than their surroundings. And finally Hale found the magnetic field. After the development of plasma physics and MHD it became possible to understand why the magnetic field is essential in keeping the spots cooler.

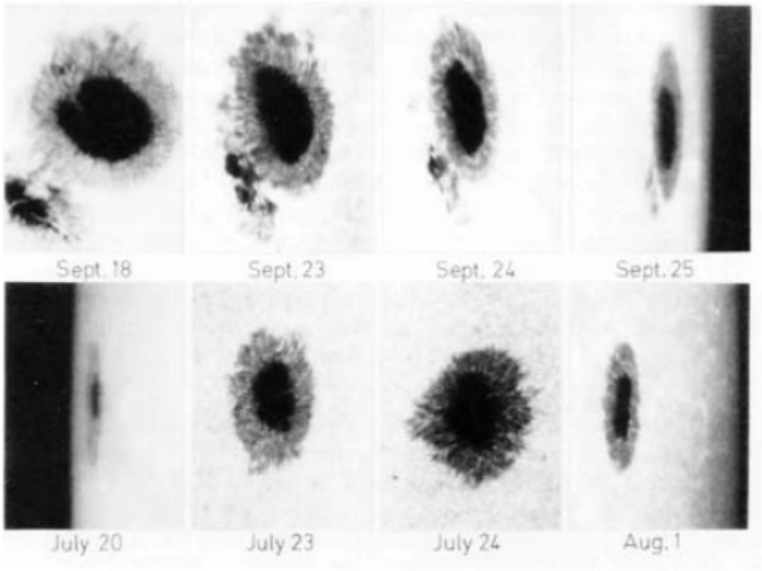

Figure 7.1: Two sounspots moving with the rotating Sun.

7.1.1 Characteristics of sunspots

The magnetic field and the sunspots are intimately related to each other. A spot corresponds to an intense magnetic flux tube emerging from the convection zone to the photosphere. The biggest spots can have a diameter of about 20 000 km. The center of the spot is called the **umbra** whose temperature is about 4100 K and the largest magnetic fields are about 0.3 T (corresponding to 3 kG in the Gaussian units, which are still frequently used in solar and stellar physics). The strong magnetic field is the cause of the low temperature, as the total pressure, including magnetic energy density, must be in balance. We can say that the magnetic field inhibits the hot plasma to reach the surface.

Around the spot there may be a **penumbra** that consists of dark and bright filaments. Young spots do not have penumbrae and in about 50% of the cases the spot development stops before the a penumbra has developed.

The magnetic field is determined by measuring the **Zeeman splitting** of atomic spectral lines. A line that radiates at a wavelength λ_0 without magnetic field becomes split into two or three components when a magnetic field affects the motion of the electrons in the atom. The number of components depends on the orientation of the magnetic field. When the line-of-sight in is the direction of the magnetic field, the observer sees two shifted σ components which have a circular polarization of opposite sense. When the magnetic field is perpendicular to the line-of-sight, the observer sees an unshifted π component linearly polarized perpendicular to the magnetic

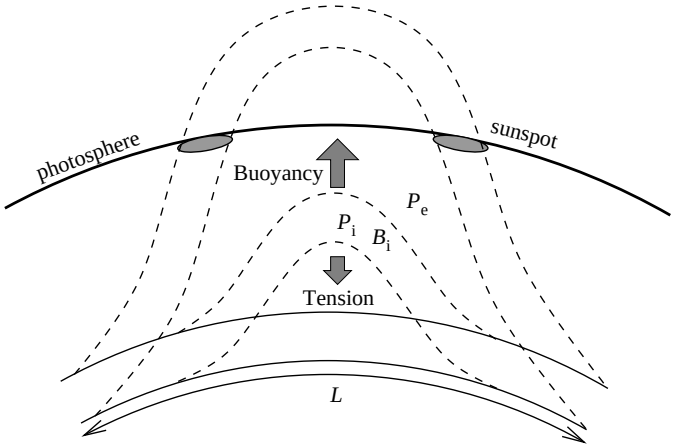

Figure 7.2: A flux tube emerging from the convective zone due to buoyancy force.

field and the two σ components now linearly polarized along the magnetic field. Thus the observation of the degree of polarization is important in the determination of the magnetic field and its direction on the Sun.

The Zeeman effect is very weak. For the three-component split, the shift, $\Delta\lambda$, in wavelength of the two outer components from the central wavelength is (in meters)

$$\Delta\lambda = \lambda^2 \Delta\nu / c \approx 47\, \lambda^2 B\,, \tag{7.1}$$

where the shift in the frequency ($\Delta\nu$) is related to the electron gyrofrequency, $f_{\rm ce} = eB/(2\pi m_{\rm e})$, as $\Delta\nu = f_{\rm ce}/2 = 1.4 \times 10^{10} B$. All quantities are given here in SI units (in most textbooks these formulas are in cgs units, i.e., length in cm and magnetic field in gauss). Even in a strong magnetic field of 0.3 T the shifts $\Delta\lambda/\lambda_0$ at optical wavelengths are of the order of 10^{-6}.

The spots usually appear in pairs or in larger groups. The magnetic field emerging from one spot must return to another. Thus the sunspot pairs have opposite polarity. A simple explanation of sunspot pairs was provided by *Parker* in 1955 and is illustrated in figure 7.2.

Write the pressure balance between the flux tube and the weakly magnetized surrounding as

$$\frac{k_{\rm B} T \rho_{\rm e}}{m} = \frac{k_{\rm B} T \rho_{\rm i}}{m} + \frac{B_{\rm i}^2}{2\mu_0}\,, \tag{7.2}$$

where 'e' and 'i' denote the exterior and interior of the flux tube. Here the kinetic pressure is given in terms of unit mass m in order to avoid confusion between molecular mass μ and the magnetic permeability! As in the example of convective instability (Chapter 5), there is buoyancy if $\rho_{\rm e} > \rho_{\rm i}$. The tube

will rise if the buoyancy exceeds the magnetic tension that tries to pull the tube back,

$$(\rho_{\rm e} - \rho_{\rm i})g > \frac{B_{\rm i}^2}{\mu_0 L}\,, \tag{7.3}$$

where L is the distance between the footpoints of the flux tube. From these equations we get the condition for the minimum separation of the footpoints

$$L > \frac{2k_{\rm B}T}{m} = 2H_P\,, \tag{7.4}$$

where H_P is the pressure scale length. If the two footpoints are too close to each other, the tension wins over the buoyancy. Note that this explanation just illustrates the buoyancy but the exact generation mechanism of the sunspots is not yet properly understood.

The sunspots are obviously related to the convective motion below the solar surface. Small spots are of the size of the mesogranular network whereas the distribution within sunspots groups is of the size of supergranules. However, it is not clear to how deep into the convective zone the spots are connected magnetically.

A sunspot group is surrounded by a region of moderate field strength (about 0.01 T). This **active region** is hotter, and thus brighter, than its surroundings. In active regions there is a lot of weaker magnetic structures and activity. However, this activity does not seem to compensate the energy flux blocked by the spot. That is, when there are large spots in the Sun, the total energy flux is smaller than without the spots and a part of the energy appears to be somehow stored in the convective zone. This has an important consequence for studies of other magnetically active stars, as *transient* drops in their luminosity can be interpreted in terms of stellar spots.

The situation is quite different in the longer time scales. As discussed in section 1.5 the total irradiance during the high solar activity years (i.e., when the number of sunspots is high) is larger than during solar minimum, which seems contradictory to the transient behavior. The exact reason to this is not known. A possible explanation is related to the so-called **faculae**. Faculae are bright spots of about 0.25" diameter, i.e., smaller than the granular structures. However, their total area is 15–20 times larger than the area covered by sunspots. Thus already a small enhancement of the facular brightness can overcompensate the the dimming due to the large sunspots. This explanation is further supported by the fact that the number of faculae is known to increase with increasing average magnetic activity in the regions where they are found.

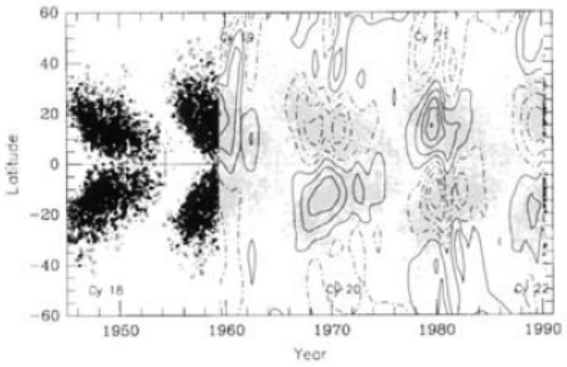

Figure 7.3: The butterfly diagram of sunspot appearance. The contours are $\pm 20\,\mu$T, $\pm 60\,\mu$T, $\pm 100\,\mu$T,..., solid lines indicate positive polarity, dashed lines negative.

7.1.2 The sunspot cycle

The sunspots appear in quasi-periodic patterns, **solar cycles**, of about 11 years. This cyclic behavior was found by *Schwabe* in 1844. When a new cycle begins, spots start to appear at mid-latitudes (around 30–40°) on both hemispheres. The life-time of individual spots is relatively short, from days to weeks, but with the time more and more new spots appear. The new spots appear closer and closer to the equator, resulting in the famous butterfly diagram (figure 7.3). After the maximum occurrence the sunspot number starts to decrease to the solar minimum of practically no sunspots at all.

In 1923 Hale was able to confirm the **polarity rules** of sunspots which he had formulated with his colleagues in 1919:

- The magnetic orientation of leader and follower spots in bipolar groups remains the same in each hemisphere over each 11-year cycle.
- The bipolar groups in the two hemispheres have opposite magnetic orientation.
- The magnetic orientation of bipolar groups reverses from one cycle to the next.

These rules are illustrated in Figure 7.4. As it takes two sunspot cycles to return with the same orientation, the length of the magnetic cycle is actually 22 year. It is sometimes called the **Hale cycle**.

The systematic behavior of bipolar sunspot groups can be understood in terms of a subsurface **toroidal magnetic field**. "Toroidal" means in this context that the magnetic field lines form loops around the solar rotation

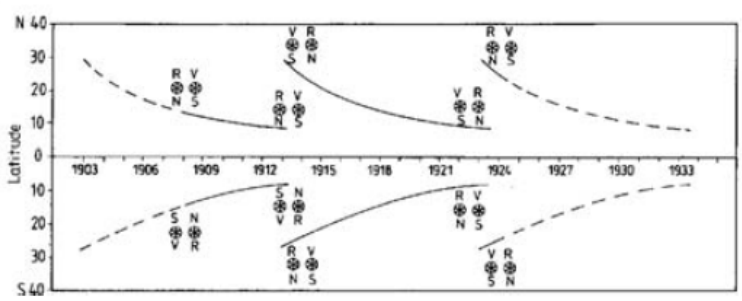

Figure 7.4: The polarity law of sunspot pairs. N and S are the north and south polarities. R and V indicate the red (longer wavelength) and violet (shorter wavelength) σ components of the Zeeman triplet.

axis. Locally this field may be be driven to the surface by convective or magnetic buoyancy and form the bipolar sunspot groups. The total flux of the toroidal field is of the order of 10^{15} Wb. If we assume that it is distributed within the latitudinal range of the sunspots and throughout the convection zone, the mean toroidal field is $B_{\rm t} \approx 0.02$ T. If most of the flux is concentrated in the thin overshooting layer at the bottom of the convective layer, the mean field in that region can be of the order of 1 T.

There is also a **mean poloidal magnetic field** inside the Sun. Here "poloidal" means a field whose field lines are in the meridional planes. The total internal (mean) field is the sum of these two components. In the next section we will see how the differential rotation drags the poloidal field lines to enhance the toroidal component. This takes place during the rising solar activity. In order to establish the cyclic behavior there must be another process to return toward a more poloidal configuration during the decaying activity. This mechanism must also produce a reversed poloidal field compared to the previous minimum configuration.

While the solar cycle is remarkably repetitive, it also shows great variability that cannot yet be predicted properly. The intensity of sunspot activity is usually given by the **relative sunspot number** R introduced by *Wolf* in 1848

$$R = k\,(10\,g + f)\,, \tag{7.5}$$

where g is the number of spot groups and f is the total number of spots (an isolated spot is calculated also as a group). The calibration constant k is individual for each observatory. It accounts for the instrument properties and seeing conditions. R is nearly proportional to the total area of the Sun covered by the spots and a rough measure of the total absolute magnitude of the magnetic flux penetrating the visible hemisphere within the sunspots.

Sunspots have been calculated in the same way since the 1840's. The

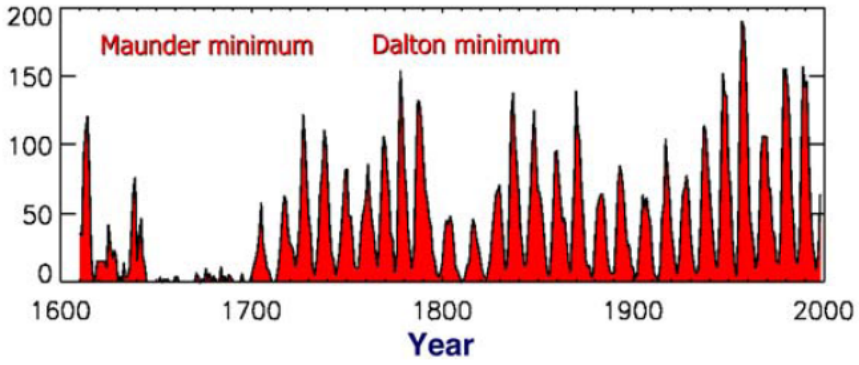

Figure 7.5: Long term sunspot number variation.

record from the earlier years starting in 1610 has been reconstructed from more scattered data. Figure 7.5 shows the entire time series. The cycles have been numbered so that the cycle which peaked in 2000 has number 23. The next cycle, cycle 24, is probably starting to rise in 2009. Note that the present solar minimum has been a relatively long one, and many forecasts of the start of the cycle 24 have been erroneous.

Both the intensity and the shape of the peaks are different from one cycle to another. The strongest recorded maximum took place in 1957 (cycle 19). During the last century there was an increasing trend of the peak sunspot numbers with the exception of cycle 20. Also the most recent maximum (cycle 23 in 2000) was weaker than the previous two. It may be a sign of the so-called **Gleissberg cycle** of about 80 years superposed on the 22-year Hale cycle. In that case the coming maxima would be smaller than the recent ones.

Also the length of the cycles varies up to a few years. However, perhaps the most remarkable feature is that the solar activity seems to have been almost nil during a long period of the 17th century. It is believable that this is not an artifact of poor observations. There simply were almost no spots on the Sun. This coincided with the so-called little ice age when the climate was exceptionally cool. Today it is known that the variable solar activity does affect the climate but it is not really known how.

7.2 Solar dynamo

What is the origin of the magnetic field of the Sun? In principle the field could be a remnant of the magnetic field in the interstellar cloud which once collapsed to form the Sun. Recall that if the cloud's weak field, less than 1 nT, were compressed with the matter without any losses, the resulting

flux density would be huge, some 10^6 T. Much of this was lost in the early evolution but considering the fact that the Ohmic diffusion time for the Sun is of the order of 10^{10} years, the mere existence of the field does not require its generation. The case is different for the planets, e.g., for the Earth $\tau_\eta \approx 10^4$ years, thus the Earth must possess a dynamo of some type, or the only magnetism would be remanence in magnetic materials in the ground.

Not even the 22-year magnetic cycle of the Sun is a fully convincing signature of an active solar dynamo. It might be a sign of some kind of oscillatory behavior of a slowly decaying fossil field. However, the detailed features of the differential rotation and its association to the migration of the sunspots can be considered as the strongest evidence of the dynamo. The present Sun as well as the Earth and other magnetized planets are able to manifold the pre-existing flux through a dynamo process. In the case of the Sun this takes place in the convection zone, most likely near its bottom. The excess magnetic energy is expelled away with the solar wind. The energy sources for the magnetic field generation are the rotation and the heat produced in the core.

7.2.1 The idea of dynamo

Let us again start from the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}\,. \tag{7.6}$$

The convective term involves the plasma motion from which we want to generate new flux whereas the second term describes how the field is diffused away. It is important to understand that both terms are needed in the description of a plasma dynamo. If there were just diffusion, a discussion of dynamo action would, of course, not be reasonable.

If, on the other hand, there would be no diffusivity at all, the mere convection would describe the ideal MHD without creation of any new flux either. The problem of **dynamo theory** is to find solutions for the whole induction equation where the convection and diffusion together result in creation of new magnetic flux, or more exactly, manifolding of the existing flux. This is somewhat analogous to an old-fashioned bicycle dynamo. If you just have the dynamo rotating, not connected anywhere, the only effect would be weak friction that would make the cycling a little harder. But if you connect the dynamo through a load, e.g., a lamp, a current flows in the cable and that current is associated with a magnetic field created by the dynamo action. Note that you already have a magnet in your dynamo but the energy to create the new flux is not drained from the magnetic energy of

the magnet but from the mechanical work you are doing to get the magnet rotating. This way we have natural roles for $\mathbf{v}$, the rotation, and for η, the dissipation in the cable and the lamp. Both are needed!

This analogy shall not be taken too literally. Technological dynamos are multiply-connected systems where the load is external to the dynamo itself. In MHD plasma there are no cables or circuits. The new flux is directly superposed on the pre-existing field in the same simply-connected volume of fluid whose motion creates the flux and the flux is also dissipated in the same volume.

An important concept is the **self-excitation** of the cosmic dynamos. In a bicycle the seed magnetic field is provided by a permanent magnet. We can imagine to set up a self-exciting dynamo by winding the wire connected to the load around the system so that it creates a magnetic field that is in the same direction as the seed field. Thereafter we remove the original magnet and the seed field to the dynamo is now provided by the field generated by the dynamo itself. This is not a perpetuum mobile as the energy source for the magnetic field generation is the motion that has to be strong enough to balance the dissipation.

7.2.2 Kinematic and dynamic dynamos

The plasma dynamo is easiest to describe at the **kinematic** level where the velocity field $\mathbf{v}$ is assumed to be given and not affected by the evolution of the magnetic field. This is a reasonable starting point in a hydrostatic object like the Sun where the pressure and gravitation balance each other and the Lorentz force $\mathbf{J} \times \mathbf{B}$ is negligible, i.e., the field is force-free. Then the induction equation is **linear**, which helps a lot in the analysis.

However, in reality the magnetic force may not be negligible and $\mathbf{v}$ becomes a function of $\mathbf{B}$ and then the induction equation is **non-linear**. In that case the analysis requires a simultaneous solution of the momentum equation, i.e., we have to solve a **dynamic** (magnetohydrodynamic) problem, which in the case of the Sun means a combined solution of the convective motion and magnetic field generation. This requires extensive computer simulations and the problems yet to be solved are difficult.

In fact, the present understanding of the solar dynamo is that the kinematic approach cannot provide a satisfactory description of the magnetic field. Also here the helioseismology has played a role. Figure 7.6 illustrates our understanding of the differential rotation before and after the analysis of solar oscillations. In the old picture the rotation was assumed to resemble concentric cylinders, the outer of which would rotate faster. In that case there would be considerable velocity shear throughout the convection zone. The analysis of rotational modes described in Chapter 4 indicates that the

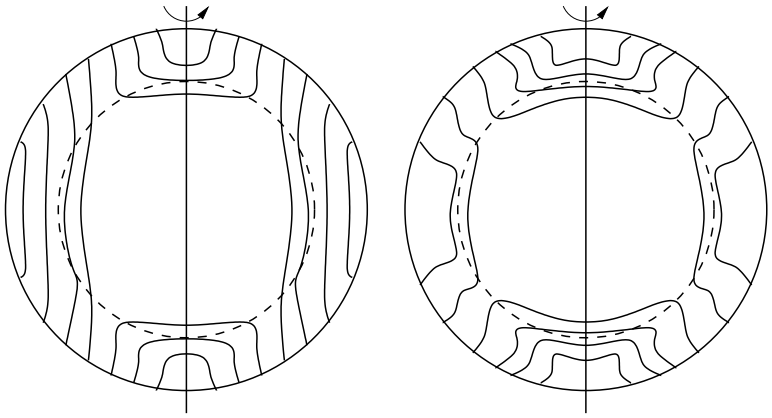

Figure 7.6: Differential rotation as assumed before helioseismological results (left) and the present view (right).

main shear region is close to the bottom of the convection zone. It has turned out that the kinematic approach does not seem suitable in this latter case when the field generation is likely to take place within a relatively thin region near the bottom of the convection zone.

7.2.3 Anti-dynamo theorems

The existence of MHD dynamos is not a trivial matter. A theoretically attractive feature of MHD is that one can prove various theorems which follow from the well-defined basic equations. The frozen-in theorem is one, there are theorems on how force-free fields must be stressed by non-force-free configurations, etc. A special class are the anti-dynamo theorems which constrain the configurations where the dynamo action is possible at all. A short list of examples of anti-dynamo theorems is

- Cowling's theorem: *An axisymmetric magnetic field cannot be maintained by dynamo action.* This means that a dynamo must produce a more complicated configuration than the simple dipole.
- *A two-dimensional magnetic field cannot be maintained by dynamo action.* This means that in any coordinate system (x, y, z) the field $\mathbf{B}$ cannot be independent of one coordinate.
- *An incompressible motion in a spherical volume having a zero radial component everywhere cannot maintain a field.*
- Zeldovich's theorem: *An incompressible motion in which $v_z \equiv 0$ in some Cartesian coordinate system cannot maintain a magnetic field.*

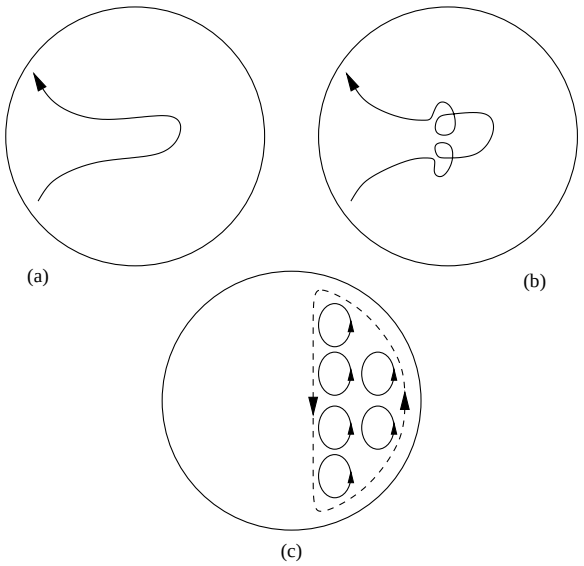

Figure 7.7: Parker's dynamo concept.

The last one is essentially the same as the theorem by Bullard and Gellman, according to which the field cannot be maintained by a pure rotation, no matter how non-uniform. We do not prove any of these here but leave the Cowling's theorem as an **exercise**.

The point of all these theorems is that it is not useful to look for too simple dynamo solutions. In the following we will, however, see that in the *mean-field sense* an axisymmetric field can be generated by a dynamo that employs the non-uniform rotation to produce the toroidal field element and the turbulence of the convective flow to produce the poloidal field.

7.2.4 Parker's turbulent dynamo

In 1955 *Parker* presented a qualitative idea for a dynamo in a differentially rotating and convective star. Let $B_\phi \mathbf{e}_\phi$ denote the toroidal field and $\mathbf{B}_\mathrm{p} = B_r \mathbf{e}_r + B_\theta \mathbf{e}_\theta$ the poloidal field. The toroidal field lines are circles around the rotating axis and the poloidal field lines are curves in a given meridional plane. As the magnetic field in the convection zone is frozen-in the differentially rotating plasma, the rotation produces toroidal field from the poloidal by stretching the field lines (Figure 7.7 a). Thus the production of toroidal field is straightforward to understand.

The production of poloidal field is more complicated. Here Parker invoked the concept of turbulent convection. The upward (downward) moving plasma blobs stretch the toroidal field in the upward (downward) direction. The upward moving plasma blobs tend to spread out as they rise. If the

convection takes place in a rotating frame of reference, the spreading gives rise to vorticity in the blob. Consequently, the blobs start to rotate and the motion becomes **helical**. This is similar to the helical motion of cyclones in the Earth's atmosphere. The helical motion twists the the toroidal field lines (Figure 7.7 b). Projected into the meridional plane the twisted field lines form magnetic loops (Figure 7.7 c). The helical motion in the two hemispheres has opposite sense (again, think the cyclones) but as also B_ϕ has opposite direction the resulting loops have the same sense on both hemispheres (Figure 7.7 b).

The concept of **turbulent convection** is essential here as the turbulent diffusivity is needed to partially break ("reconnect") the flux freezing in order to form the magnetic field loops. Note that the turbulent diffusion also smoothens out the magnetic fields in the of the loops and a mean poloidal field $\mathbf{B}_\mathrm{p}$ is formed from the a toroidal field.

7.2.5 Mean-field approach

Parker's intuitive concept was expressed in a systematic way by *Steenbeck*, *Krause* and *Rädler* in 1966 in terms of the so-called **mean-field electrodynamics** (or mean-field magnetohydrodynamics). This approach is still kinematic. While it may not provide a complete description of the solar dynamo, it illustrates some of the basic principles at analytical level and also introduces the alpha effect which belongs to the basic jargon of MHD dynamo physics.

We assume the velocity field given and write the mean magnetic and velocity field as sums of the average fields ($\langle\mathbf{B}\rangle, \langle\mathbf{v}\rangle$) and the fluctuating parts ($\mathbf{b}, \mathbf{u}$)

$$\mathbf{B} = \langle\mathbf{B}\rangle + \mathbf{b} \tag{7.7}$$

$$\mathbf{v} = \langle\mathbf{v}\rangle + \mathbf{u}\,. \tag{7.8}$$

Now the field $\mathbf{u}$ represents the turbulent motion. The meaning of the average fields is not quite trivial. In studies of turbulent motion "average" means in general ensemble average. However, here we can understand the mean values been taken over longitudes (ϕ) but not over latitudes (θ) because the differential rotation is an essential part of the problem. Note that when taking averages the average of a fluctuating quantity is zero.

If the total velocity field were known, we could solve the kinematic problem directly. However, as usual we cannot express the turbulent motion in analytical form. Thus we have to be satisfied of knowing $\langle\mathbf{v}\rangle$ and **assuming** reasonable statistical properties of $\mathbf{u}$. Substituting the above expressions to the induction equation and separating the mean and fluctuating parts we

obtain

$$\frac{\partial\langle\mathbf{B}\rangle}{\partial t} = \nabla\times(\langle\mathbf{v}\rangle\times\langle\mathbf{B}\rangle+\mathcal{E}-\eta\nabla\times\langle\mathbf{B}\rangle) \tag{7.9}$$

$$\frac{\partial\mathbf{b}}{\partial t} = \nabla\times(\langle\mathbf{v}\rangle\times\mathbf{b}+\mathbf{u}\times\langle\mathbf{B}\rangle+\mathbf{G}-\eta\nabla\times\mathbf{b})\,, \tag{7.10}$$

where

$$\mathcal{E} = \langle\mathbf{u}\times\mathbf{b}\rangle \tag{7.11}$$

$$\mathbf{G} = \mathbf{u}\times\mathbf{b}-\langle\mathbf{u}\times\mathbf{b}\rangle\,. \tag{7.12}$$

$\mathcal{E}$ is **the mean electric field** (in a sense the mean electromotive force) induced by the fluctuating motion. It appears as an extra electric field in the mean induction equation (7.9) and also in Ohm's law

$$\langle\mathbf{J}\rangle = \sigma(\langle\mathbf{E}\rangle+\langle\mathbf{v}\rangle\times\langle\mathbf{B}\rangle+\mathcal{E})\,. \tag{7.13}$$

Thus $\mathcal{E}$ can act as an additional source of the mean current and, through the current, of the mean magnetic field.

The inclusion of $\mathcal{E}$ is the secret of the turbulent dynamo and causes the mean fields to evolve in a different way from the laminar fields satisfying the equations of MHD. If we knew how to compute $\mathcal{E}$, we would have a solution for the mean magnetic field $\langle\mathbf{B}\rangle$. Unfortunately the calculation is, in general, too difficult.

However, there is a linear relationship between $\mathbf{b}$ and $\langle\mathbf{B}\rangle$ and, thus, also between $\mathcal{E}$ and $\langle\mathbf{B}\rangle$. Assuming incompressibility of the turbulent flow ($\nabla\cdot\mathbf{u}=0$), we can expand this relationship (**exercise**) as

$$\mathcal{E}_i = \alpha_{ij}\langle B_j\rangle+\beta_{ijk}\partial_k\langle B_j\rangle\,. \tag{7.14}$$

In the mathematical language the coefficients α_{ij} and β_{ijk} are pseudotensors which relate an axial vector $\langle\mathbf{B}\rangle$ to a polar vector $\mathcal{E}$. In the kinematic approach the coefficients represent the statistical properties of the field $\mathbf{u}$ and are independent of $\mathbf{B}$.

The mean electric field is possible to calculate explicitly if the vector $\mathbf{G}$ can be neglected. This requires that either the magnetic Reynolds number must be small or $u\tau \ll l$, where $(u,\,\tau,\,l)$ are the characteristic scales of $\mathbf{u}$ and $\mathbf{b}$. In the Sun, R_m is large and $u\tau\sim l$. In this respect the **geodynamo** is somewhat easier because in the liquid core of the Earth R_m is small.

Nevertheless, in order to proceed, we make the so-called **first-order-smoothing approximation** and neglect $\mathbf{G}$. After this we may not anymore be able to end up with the correct solution for the Sun but it is instructive to find some solution anyway. Assume further that the turbulence is isotropic,

which reduces the coefficients to $\alpha_{ij} = \alpha\delta_{ij}$ and $\beta_{ijk} = \beta\epsilon_{ijk}$ (where ϵ_{ijk} is the antisymmetric permutation symbol). Under these approximations the mean electric field is

$$\mathcal{E} = \alpha\langle\mathbf{B}\rangle - \beta\nabla\times\langle\mathbf{B}\rangle\,, \tag{7.15}$$

where α and β are determined by the statistical properties (correlations) of the field $\mathbf{u}$ as

$$\alpha = -\frac{1}{3}\int_0^\infty \langle\mathbf{u}(t)\cdot\nabla\times\mathbf{u}(t-t')\rangle dt' \tag{7.16}$$

$$\beta = \frac{1}{3}\int_0^\infty \langle\mathbf{u}(t)\cdot\mathbf{u}(t-t')\rangle dt' \tag{7.17}$$

and substituting to the equation for $\langle\mathbf{B}\rangle$ we find

$$\frac{\partial\langle\mathbf{B}\rangle}{\partial t} = \nabla\times\left(\langle\mathbf{v}\rangle\times\langle\mathbf{B}\rangle + \alpha\langle\mathbf{B}\rangle - \eta_{\rm t}\nabla\times\langle\mathbf{B}\rangle\right), \tag{7.18}$$

where the total diffusivity is $\eta_{\rm t} = \eta + \beta$. The turbulent contribution

$$\beta \approx \frac{1}{3}u^2\tau \approx \frac{1}{3}ul \gg \eta \tag{7.19}$$

dominates over the classical diffusion. In the solar convective zone $\eta_{\rm t} \approx \beta = 10^8 - 10^9$ m^2/s. This reduces the global time scale of diffusive decay to the order of 10–100 years. This is a quite reasonable number considering that the entire solar cycle is 22 years!

The novel feature is that the rate of change of the mean magnetic field is related to the field itself through the coefficient α. This the **alpha effect**. Although the quantitative estimation of α is difficult, its definition expresses Parker's intuitive notion that the turbulent motion is correlated to its own **vorticity**, i.e., the motion is helical. This introduces the concept of **kinetic helicity** to the dynamo theory.

Note that kinetic helicity should not be confused with **magnetic helicity**

$$H_{\rm m} = \int \mathbf{A}\cdot\mathbf{B}\,dV\,, \tag{7.20}$$

where $\mathbf{A}$ is the vector potential, or **current helicity**

$$h_j = \mathbf{j}\cdot\mathbf{B}\,, \tag{7.21}$$

which both are useful concepts in MHD. Magnetic helicity is a measure for the structural complexity of the magnetic field configuration (see e.g., Lectures on Advanced Space Physics). In ideal MHD $H_{\rm m}$ is conserved and

in the case of small η it is nearly so. This may have consequences to the evolution of the magnetic field in the mean-field dynamos. However, further discussion of these problems is beyond these lectures.

This solution looks suspiciously axially symmetric. Have we thus violated Cowling's theorem? The answer is no. Cowling's theorem applies for laminar flows in MHD and there is no alpha effect nor $\mathcal{E}$ in the exact MHD equations. They arise from the mean-field approach to the turbulent flow. The extra term $\sigma\mathcal{E}$ in Ohm's law invalidates the proof of Cowling's theorem (**exercise**).

7.2.6 The kinematic $\alpha\omega$ dynamo

Regardless of whether the above discussion is valid for the solar dynamo, or not, it is important that it has been established that the alpha effect can sustain a dynamo. In fact, if the system does not have much differential rotation, the alpha effect can be responsible also for the production of the toroidal field from the poloidal field. In such a case the process is called an α^2 dynamo. In Parker's concept discussed earlier the alpha effect works together with the differential rotation. Then the dynamo is called an $\alpha\omega$ dynamo.

Use spherical polar coordinates and assume known $\alpha(r,\theta)$ and $\Omega(r,\theta)$. Let α be antisymmetric with respect to the equatorial plane (**Exercise:** Why?)

$$\alpha(r,\pi-\theta) = -\alpha(r,\theta) \tag{7.22}$$

and the angular velocity symmetric

$$\Omega(r,\pi-\theta) = \Omega(r,\theta)\,. \tag{7.23}$$

Furthermore, assume that besides rotation there is no other mean motion

$$\langle \mathbf{v} \rangle = (0,0,\Omega r\sin\theta)\,. \tag{7.24}$$

Separate the mean file to poloidal and toroidal components $\langle \mathbf{B} \rangle = \mathbf{B}_{\rm p} + \mathbf{B}_{\rm t}$ where

$$\mathbf{B}_{\rm p} = \nabla \times (0,0,A(r,\theta,t)) \tag{7.25}$$

$$\mathbf{B}_{\rm t} = (0,0,B(r,\theta,t))\,. \tag{7.26}$$

The mean-field induction equation can also be separated to its poloidal and toroidal parts. Assuming, for simplicity, constant $\eta_{\rm t}$ we get

$$\frac{\partial A}{\partial t} = \alpha B + \eta_t \nabla_1^2 A \tag{7.27}$$

$$\frac{\partial B}{\partial t} = \frac{\partial\Omega}{\partial r}\frac{\partial}{\partial\theta}(A\sin\theta) - \frac{1}{r}\frac{\partial\Omega}{\partial\theta}\frac{\partial}{\partial r}(rA\sin\theta) - \frac{1}{r}\frac{\partial}{\partial r}\left[\alpha\frac{\partial}{\partial r}(rA)\right] - \frac{1}{r^2}\frac{\partial}{\partial\theta}\left[\frac{\alpha}{\sin\theta}\frac{\partial}{\partial\theta}(A\sin\theta)\right] + \eta_{\rm t}\nabla_1^2 B\,, \tag{7.28}$$

where $\nabla_1^2 = \nabla^2 - (r\sin\theta)^{-2}$.

Now the role of the alpha effect becomes clear. With $\alpha = 0$ the poloidal field would decay exponentially and together with the disappearance of A, the same would happen to B. The alpha effect generates poloidal field from the toroidal field. Note that in equation (7.28) both the alpha effect and the differential rotation (actually $\nabla\Omega$, only the derivatives of Ω are involved) produce toroidal field from the poloidal field. In the solar models the alpha effect's role in this phase is usually neglected as it is likely (but not really known) that

$$|\alpha| \ll r_\odot^2 |\nabla\Omega| \,. \tag{7.29}$$

If $|\nabla\Omega|$ would be very small, the alpha effect would take over also the generation of the toroidal field.

Consequently in the $\alpha\omega$ dynamo the oscillation between the mean toroidal and poloidal components takes place as:

$$\ldots \; \mathbf{B}_\mathrm{t} \stackrel{\alpha}{\longrightarrow} \mathbf{B}_\mathrm{p} \stackrel{\nabla\Omega}{\longrightarrow} \mathbf{B}_\mathrm{t} \; \ldots$$

The efficiency of the $\alpha\omega$ dynamo is measured by the dimensionless **dynamo number**

$$D = \alpha_0 \Omega_0 r_\odot^3 / \eta_\mathrm{t}^2 \,, \tag{7.30}$$

where α_0 is the typical value of α and Ω_0 the typical value of internal differences of Ω. If any of these is zero, the field decays exponentially. In order to prevent the decay D must exceed a certain value, which depends on the specific construction of the dynamo in question.

7.2.7 Parker's migratory dynamo

In his 1955 paper Parker also addressed the fact that the sunspots migrate from higher latitudes toward the equator. Thus also the newly created toroidal field is expected to migrate equatorward. In this case it is easier to use Cartesian coordinates in a given location in the northern hemisphere. Let x point southward, y eastward and z upward. Using the mean-field language of the previous section, let α = constant, $\langle \mathbf{v} \rangle = (0, \Omega_0 z, 0)$ with Ω_0 = constant and seek solutions for $A(x,t)$ and $B(x,t)$. Now

$$\mathbf{B}_\mathrm{p} = \left(0, 0, \frac{\partial A}{\partial x}\right) \tag{7.31}$$

$$\mathbf{B}_\mathrm{t} = (0, B, 0) \tag{7.32}$$

and the induction equations are

$$\dot{A} = \alpha B + \eta_\mathrm{t} \, \partial^2 A / \partial x^2 \tag{7.33}$$

$$\dot{B} = \Omega_0 \, \partial A / \partial x + \eta_\mathrm{t} \, \partial^2 B / \partial x^2 \,. \tag{7.34}$$

Seeking a solution of the form

$$(A, B) = (A_0, B_0) \exp[i(\omega t + kx)] \tag{7.35}$$

we obtain the dispersion equation

$$(i\omega + \eta_t k^2)^2 = ik\Omega_0\alpha\,. \tag{7.36}$$

This describes a **dynamo wave**, whose direction of propagation depends on the parameters α and Ω_0. The detailed analysis of the propagation characteristics is left as an **exercise**. The point of this analysis is that for appropriately chosen parameters a simple $\alpha\omega$ dynamo can explain the motion of the sunspots, which is a key observational fact in the whole business of solar magnetism.

7.2.8 Dynamo and the solar cycle

The $\alpha\omega$ cycle qualitatively corresponds to the solar cycle between poloidal and toroidal field configurations and by suitable tuning of the parameters the typical length of the cycle can be reproduced. At the solar minimum there are no sunspots and the large scale magnetic field is as poloidal as possible. The differential rotation destroys this pattern and starts to enhance the toroidal component. This shows in the polarity pattern of the sunspot pairs. On one hemisphere the leading spots have the positive polarity, on the other the following spot is positive following the winding of the toroidal field lines. After the maximum epoch the alpha effect takes over and starts to reorganize the poloidal field but now in the opposite direction.

Although the kinematic $\alpha\omega$ dynamo is unlikely to give a correct representation of the solar dynamo, it is possible to adjust the parameters so that an oscillatory behavior corresponding to the solar cycle can be reproduced. Figure 7.8 is an old (1976) example by Stix. This example assumes rotation and internal circulation circulation patterns different from what we today know them to be. It also gives only a very coarse picture of the magnetic field.

If we assume that the dynamo action takes place near the bottom of the convective zone, we expect to see temporal variations in that region somehow related to the 11- or 22-year periods. Recent helioseismological observations with the GONG and SOHO MDI experiments indicate that there are quasi-periodic oscillations just above and below the bottom of the convective zone at equatorial latitudes, but they have a surprising period of 1.3 years! The oscillation is strongest at $r = 0.72\,r_\odot$, whereas in the uppermost radiative zone at $r = 0.63\,r_\odot$ weaker variations are found at the same frequency but in antiphase. The signal becomes much weaker away

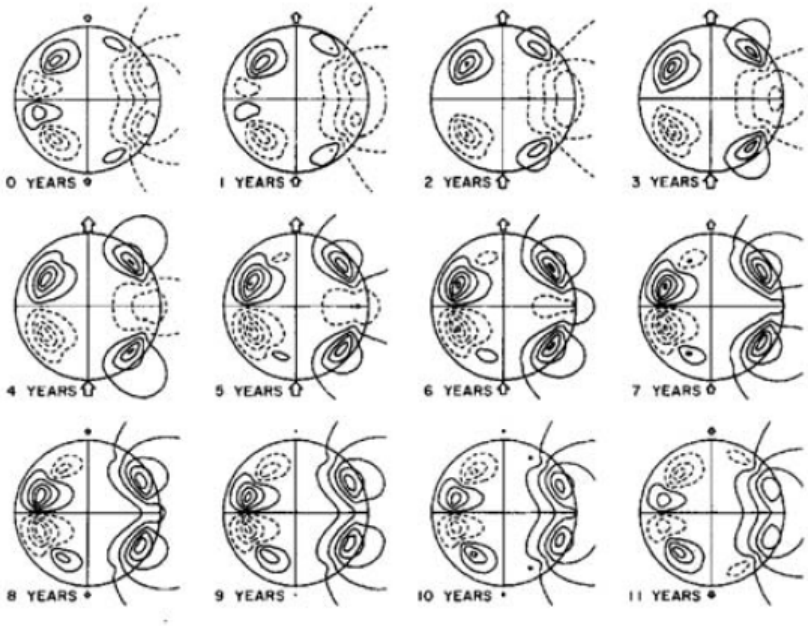

Figure 7.8: An example of oscillatory kinematic $\alpha\omega$ dynamo. The meridional cross sections show contours of constant toroidal field strength on the left, and poloidal field lines on the right. Arrows above the poles indicate the strength of the polar field. The time scale is artificially adjusted to 11 years.

from the equator. These findings are not yet understood, but they may represent the first direct observation of variability at the seat of the solar dynamo.

Chapter 8

Chromosphere and corona

With the modern space-borne X-ray and ultraviolet observations of the Sun by the Yohkoh and SOHO spacecraft studies of the outer layers of the solar atmosphere have gained increasing attention. These regions have turned out to be even more intriguing than anticipated earlier. Although driven by processes inside the Sun, much of the most fascinating solar activity takes place in the upper atmosphere. In this chapter we review some basic properties of this region whereas the most important eruptive phenomena are discussed in the next chapter. Note that this section does not contain many pictures, as the best originals are in color and reproduction of them in black and white does not make very much sense. The interested student is encouraged to look for picture material in the web, e.g., from the spacecraft home pages listed in the introduction.

8.1 Structure of the upper atmosphere

As indicated in figure 3.4 the temperature of the solar atmosphere starts to rise from a minimum at an altitude of about 500 km and there is a very steep increase to 10^4 K above 2000 km. The region between these altitudes is called the **chromosphere**. The second jump above the chromosphere to temperatures of 10^6 K is called the **transition region** to the **corona**. The heating of the hot corona belongs to the most important open questions in solar physics.

The physics of the upper atmosphere is quite different from the physics of the solar interior and the photosphere. Above the photosphere the density drops rapidly and the interparticle collisions become less frequent. This means that the deviations from LTE become increasingly important and the ionization and excitation levels cannot any more be predicted from the Saha

and Boltzmann equations. They have to be determined from the observations. Another consequence is that the plasma pressure becomes smaller than the magnetic pressure and the magnetic field begins to dominate the dynamics of the gas. The large scale properties of the magnetic field can be studied in the MHD approximation but many important features require kinetic treatment.

In studies of the solar interior and the photosphere it was somehow reasonable to begin with average, or steady-state, models. However, such models of the upper atmosphere may be rather useless. They can even be seriously misleading, as e.g., the higher half of figure 3.4. The temperature jump in the transition region is very rapid indeed but its spatial structure is complicated and varies with time.

8.1.1 Chromosphere

The upper solar atmosphere is optically thin in almost the whole visible spectrum. Thus the only opportunities to observe the outer parts of the Sun were, for long time, provided by the fortunate coincidence that the angular diameters of the Sun and the Moon are very nearly the same and there are total eclipses every now and then. The chromosphere got its name from the colorful flash spectrum just before or just after the occultation. This spectrum has several lines that are not visible against the solar disc.

A particularly important chromospheric emission line is the Hα line at 656.3 nm which is an absorption line in the cool Sun but an emission line in the hotter chromosphere. Thus we can produce a picture of the chromosphere using a narrow-band filter around the Hα line. Another important source of information is the Fraunhofer K line of the singly-ionized calcium (Ca II 393.4 nm). Other principal means of observing the chromosphere are radio emissions produced by electrons in the strong magnetic fields and the UV emissions. The latter require space technology which, when available, provides fabulous UV pictures from the hot chromosphere and corona.

Pictures of the chromosphere show a granular structure with scale sizes comparable to the **photospheric supergranules**. These illustrate a phenomenon called **magnetic network**. The small scale magnetic loops in supergranular scales extend from the solar surface to the chromosphere.

Embedded into this structure are small dark dots (**mottles**). These are most likely associated with **spicules** seen on the limb. The spicules look like torches rising from the surface to average altitude of 5000 km. In polar regions **macrospicules** reach up to 20000 km. Thus the spicules are both chromospheric and coronal structures. The upward plasma flow in the spicules is 20–30 km/s and stays remarkably steady all the way. The dark dots follow the supergranular structure. The magnetic flux is convected with

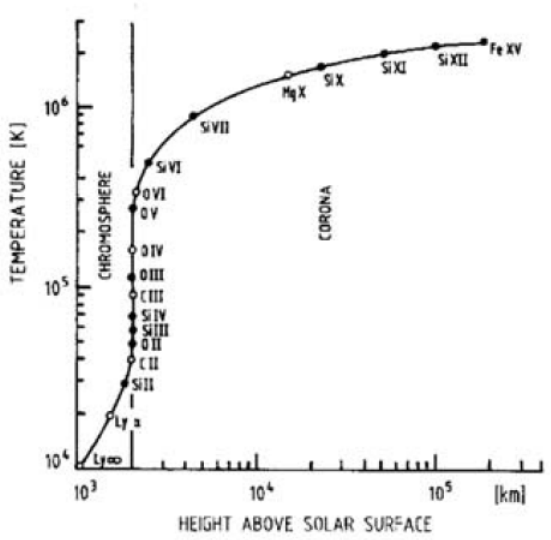

Figure 8.1: The rapid temperature increase in the transition region. The dots and circles indicate the temperature of formation of some ionized atomic species.

the plasma toward the boundaries of supergranules and the acceleration in the spicules is expected to be related to this process. However, details of this are unknown.

An interesting detail is that the chromospheric plasma is completely replaced roughly in one year. This can be calculated by estimating the mass of the chromosphere and the amount of solar wind escaping from the Sun.

8.1.2 Transition region

Figure 8.1 describes the rapid temperature increase from the chromospheric to coronal temperatures. The somewhat higher temperatures in the chromosphere than in Figure 3.4 are due to different methods of determining the temperature in a non-LTE environment. Here the temperature at a given altitude is determined according to observed ion species and the temperature required for this ionization. For example the altitude step from the first ionization of oxygen (O II) to five times ionized oxygen (O VI) covers only a very thin region.

There are two important points to keep in mind concerning the transition region. First, the steep increase of temperature requires equally steep decrease in density because the system must be in pressure balance over such a thin layer. Of course, the magnetic pressure is substantial but it cannot compensate the entire temperature gradient. Second, the transition

region is spatially much more inhomogeneous than figure 8.1 suggests. It is evident that the transition region is not a layer at all but a very structured and temporally evolving thin region. However, no generally accepted model has been found so far.

8.1.3 Corona

The corona could traditionally be observed during solar eclipses only. It is important to understand that the white light seen in those cases does not come from coronal emissions but is photospheric light scattered from the coronal gas.

The early observations indicated two distinct components in the white-light corona: the ***K* corona** and the ***F* corona**. K comes from the German word *Kontinuum* and F from the dark **Fraunhofer lines**. Both components resemble photospheric spectrum but in the K corona the Fraunhofer lines are absent (thus the word continuum). The K component is also highly polarized which indicates that it arises from **Thomson scattering** on free electrons (c.f., section 2.2.6). Actually, there are weak dips corresponding to the Fraunhofer H and K lines also in the K corona. The explanation for the filling of the lines is Doppler broadening due to the high temperature of the scattering electrons. This was the first hint that the corona might be hot, as first suggested by *Grotrian* already in 1931.

The F corona shows the photospheric continuum with the Fraunhofer lines. The light is unpolarized and it is explained as photospheric light scattered on dust particles. The K corona decays faster than the F corona that dominates beyond 2–3 $r_\odot$. It can be observed as **zodiacal light** far in the interplanetary space. Note that the white-light observations of coronal mass ejections far beyond 3 $r_\odot$ discussed in the next chapter are also based on Thomson scattering on electrons in the dense plasma cloud.

It is possible to create an artificial eclipse using an occulting disc. Such a device is called the **coronagraph**. However, there is so much stray light in the terrestrial atmosphere that in practice a coronagraph must be carried outside the atmosphere. Another advantage with space-borne observations is that there are numerous useful coronal emission lines of highly ionized atoms in the UV wavelengths and also X-ray emissions can be detected even against the solar disc. The space technology, including white-light coronagraphs, X-ray detectors, and in particular the SOHO mission, has opened a completely new view to the corona. More recently the progress has bee carried further with the TRACE and RHESSI satellites.

The corona is closely linked to the solar magnetism and illustrates the large-scale structure of the field. At the solar minimum the magnetic field is as regular as it ever can be with north and south poles, i.e., the poloidal field

dominates the large scale structure of the magnetic field. Within the polar regions there are **polar plumes**. These emerge from large **coronal holes** and represent the plasma flowing out with the solar wind. At the solar maximum the polar regions are not easy to recognize because the field is dominated by the irregular contributions from the toroidal field. There can be several coronal holes and the magnetically closed regions often resemble Prussian helmets, and are called **helmet streamers**. Note that the word streamer points to the visible closed structures whereas the escaping plasma streams in the solar wind originates mostly, if not completely, from the coronal holes.

At the time of first spectroscopic observations the spectrum of the corona was not easy to interpret at all. Because helium was found first in the Sun, the existence of a new element **coronium** was suggested to explain some abundant previously unknown spectral features. The reason is that nobody would expect the corona to be so hot. During the years 1939–1941 Grotrian and *Edlén* correctly identified several of the coronal lines to originate from highly ionized atoms. Three of the most conspicuous visible lines represent strong transitions of Fe XIV (530.3 nm), Ca XV (569.5 nm), and Fe X (637.5 nm). Of these Fe X is formed at 10^6 K and Fe XIV at 2×10^6 K.

The coronal spectrum is very rich in UV and X-ray lines. While the white-light observations require coronagraphs many of these short wavelengths can be observed also against the solar disc as they emerge from a much hotter gas. The X-ray detector onboard the Japanese Yohkoh satellite has observed ionized iron up to Fe XXVI during solar flares. This emission is the Lyman α emission of an iron ion with only one electron. The wavelength is 178 pm and the required temperature is about 2×10^7 K. Such and even higher temperatures are not uncommon in solar flares. The X-ray and UV observations give possibility of much more detailed temperature determination than the visible coronagraph observations. Recall that the corona is not in LTE and plasmas of very different temperatures are produced by the active phenomena. Thus, be careful with statements like: "The coronal temperature is ...".

The X-ray pictures of the Sun have revealed the very violent structure of the corona. The coronal holes are clearly seen as dark regions whereas the hot plasma radiating the X-rays is confined in the magnetic bottles of the closed field lines. In addition there are numerous small X-ray bright points. The coronal holes remain colder because they are on open field lines from which the plasma escapes as the solar wind before it is heated to the same temperatures as plasmas in the closed field line regions.

Also radio waves are useful in studies of the corona. They are emitted by electrons gyrating in the strong magnetic field, especially during solar activity producing radio flares. We will return to these in the next chapter.

8.2 Magnetic field in the solar atmosphere

The Zeeman effect makes it possible to produce fairly good magnetograms of the photospheric field. Higher up the increasing temperature broadens the emission lines and observing the splitting becomes difficult with the exception of the relatively cool prominences that have magnetic fields of 1–10 mT that is much weaker than the typical sunspot fields. During solar activity the intensity and polarization of radio emissions can be used to determine the field but unfortunately with rather poor spatial resolution. Typical values at the base of the corona are 10 mT and at 2 $r_\odot$ 0.1 mT (= 1 G). Further out there are in situ measurements down to the perihelia of Helios 1 and 2 satellites at 0.3 $AU \approx 65\ r_\odot$ and, of course, a lot from 1 AU.

A frequently used method to reconstruct the atmospheric field from the photospheric observations is to assume the field to be current-free, i.e., $\mathbf{B} = -\nabla\Psi$. In reality, this may not be quite true, as there probably are currents in the solar atmosphere. Because we do not know enough of these currents, the potential field approximation is the best that can be done. Thus the equation to solve is the Laplace equation

$$\nabla^2\Psi = 0 \tag{8.1}$$

for $r > r_\odot$. The solution in terms of spherical harmonics is

$$\Psi(r,\theta,\phi) = r_\odot \sum_{l=1}^{N}\sum_{m=0}^{l} f_l(r) P_l^m(\theta)(g_l^m \cos m\phi + h_l^m \sin m\phi)\,, \tag{8.2}$$

where

$$f_l(r) = \frac{(r_w/r)^{l+1} - (r/r_w)^l}{(r_w/r_\odot)^{l+1} - (r_\odot/r_w)^l}\,. \tag{8.3}$$

The radial dependence is assumed to be such that at some reference level $r = r_w$ the field is assumed to point in the radial direction, and thus $f_l(r_w) = 0$. This is motivated so that at some distance the radial solar wind outflow (that is the motivation for the subscript w) forces the frozen-in field into its own direction. From coronal images this happens not far from the Sun and the reference level, called the **source surface** is usually placed at $r_w = 2.6\ r_\odot$ (measured from the center of the Sun). The functions f_l are normalized as $f_l(r_\odot) = 1$ and thus the amplitudes of the various multipoles are given by the expansion coefficients g_l^m and h_l^m. The coefficients are found by least-square fitting to the magnetic observations in the photosphere. When large enough number of terms (e.g., $N = 25$) is taken into account the magnetic field pattern fits nicely to simultaneous coronagraph images.

Exercise

Show that (8.2) together with (8.3) is a solution of the Laplace equation.

8.3 Heating of the upper atmosphere

The recent Yohkoh, SOHO, and TRACE observations of the chromosphere and corona have shed much new observational information about the dynamics of the upper atmosphere. However, the rapid heating still remains one of the main insufficiently understood topics in solar physics.

8.3.1 Heating of the chromosphere

That there is some temperature increase in the chromosphere is not so difficult to understand because the rarefied gas starts to deviate from LTE and it does not need to find equilibrium with the lower atmospheric levels if some processes keep on heating it. There is two rich energy sources for the heating, namely the acoustic fluctuations and the magnetic network. The acoustic fluctuations probably play an important role in the chromosphere, as the photosphere of the Sun is not a quiet place at all! The power density of the sound waves in the chromosphere has been estimated to be about 10 Wm^{-2} which may be sufficient to heat the chromosphere up to 10 000 K. However, this is not nearly enough for the coronal temperatures. Going upward magnetic heating, either through MHD waves or current sheet dissipation becomes increasingly important. Note also that the heating must be practically continuous. If the heating would be turned off, the chromosphere would cool down in about 20 minutes.

8.3.2 Heating of the corona

The high temperature of the corona was once a great surprise and its heating still belongs to the toughest problems in solar physics. The acoustic fluctuations do not reach the coronal altitudes and thus the mechanism must be related to the magnetic field. In principle there is no lack of energy, as the amount of required energy is not so much, in particular, if compared to the radiation. The energy flux needed to power the active regions is of the order of 10^4 Wm^{-2} and on the average only a fraction of 10^{-4} of the power in electromagnetic radiation would be sufficient to heat the corona. But the corona is optically very thin, which implies that there is no mechanism to absorb the electromagnetic radiation.

There is enough energy also in the solar magnetic field. The problem is how to convert it into heat, in particular in the narrow transition region but also higher up where the mean temperature still increases from 10^6 K to 2×10^6 K.

We can think of several mechanisms to dissipate the magnetic energy as heat: waves, instabilities, current sheet dissipation, reconnection. In MHD

the energy balance can be expressed writing the Poynting theorem as

$$\int \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S} = \frac{\partial}{\partial t} \int \frac{B^2}{2\mu_0}\, dV + \int \frac{J^2}{\sigma}\, dV + \int \mathbf{v} \cdot \mathbf{J} \times \mathbf{B}\, dV\,. \qquad (8.4)$$

The LHS describes the magnetic energy entering as the Poynting flux through the surface $d\mathbf{S}$ of the volume V where the energy may show up as increasing magnetic energy and be dissipated through Ohmic heating and work (acceleration) by the magnetic force ($\mathbf{J} \times \mathbf{B}$). Again the Ohmic term is not likely determined by classical resistivity but may rise from turbulence and/or wave-particle interactions.

MHD waves are excited by the motion of magnetic and acoustic disturbances in or near the photosphere. Spectral features in the transition region are wider than could be expected for the hot gas. The excess Doppler widening has been estimated to amount to 10^4 m/s, which may be a signature of upgoing MHD waves. When these waves propagate outward they are damped and part of their energy is transformed to heat. The linear damping of the MHD waves is, however, a very slow process. Nevertheless, within the diverging coronal holes the wave heating may be the only alternative because there are no unstable flux tubes nor current sheets. Again, finding the physical mechanisms is difficult. A proposal has been **phase-mixing** of waves of different wavelengths and speeds propagating in the same spatial volume. This can lead to large spatial gradients where the effective resistivity increases and shows as Ohmic dissipation of the wave energy in the Poynting theorem. Phase-mixing is a type of turbulent phenomena where the modern tools of chaotic dynamics are used. Phase-mixing is, of course, not limited to open field line regions.

Another possibility for damping of Alfvén is that the waves have high enough frequencies to be damped by the cyclotron resonance with the plasma ions. Alfvén waves become ion–cyclotron waves at frequencies close to the ion cyclotron frequency, and these waves are very efficiently damped by the ions. As the magnetic field and, therefore, the cyclotron frequencies decrease as a function of radial distance, waves that are created at or near the solar surface because of micro-flaring (see below) and/or turbulent motions can propagate without damping until they reach the distance at which the cyclotron frequency becomes comparable to the wave frequency. Observations of ion temperatures in coronal holes indicate that minor ions like the oxygen species are very hot (up to 100 MK) and that their temperatures are anisotropic, being larger in the perpendicular direction relative to the magnetic field. This is what we expect from the cyclotron heating model, because the ions with the lowest cyclotron frequencies should be heated most efficiently, and because the heating is due to wave electric fields directed perpendicular to the magnetic field. However, some theoretical calculations

predict even too efficient wave damping by the ions with lower cyclotron frequencies than hydrogen, leaving almost no wave energy to heat this species. Thus, the cyclotron heating mechanism in the solar corona is far from being completely undertood.

Even if the waves generated near the solar surface have small frequencies, a phenomenon called turbulent cascading may allow short wave-length fluctuations to be generated from the long-wavelegth ones. The high-wavenumber fluctuations may again be efficiently damped at scales close to the ion Larmor radii. This turbulent heating mechanism in a way combines the ideas of cyclotron heating and phase mixing.

We know from observations that flux tubes in different scales, such as coronal loops, are continuously created and disrupted through various instabilities. The disrupting flux tubes convert magnetic energy into heat and acceleration whenever that happens but the disruptions may be too sparse and localized to explain the heating of the whole corona. These processes may be important during strong solar activity but the corona is hot also during quiet periods.

The Skylab mission revealed in 1973 that there are **X-ray bright points** everywhere on the Sun and later it has been demonstrated that their distribution is uniform over the whole Sun and they exist also during quiet phases of the solar activity. They resemble small flares (to be discussed in the next chapter) and the underlying particle acceleration is most likely due to continuous reconnection processes of the ever-changing magnetic field structures in the low corona.

While large flares can release some 10^{25} J of energy in some 10 minutes, they are too infrequent and can account at most 1% of heat to sustain the 10^6 K temperature of the corona. Thus if small flares should explain the heating, they would need to be very frequent, indeed. It seems likely that a direct scaling down from the large flares may not be sufficient, but the small flares relatively more dissipative.

The EUV observations at the temperatures of 10^5 K (i.e., in the thin transition region) have shown that there are localized hot spots that explode and shoot material upward at the speeds of hundreds km/s. These hot upward plasma jets occur above the lanes of the magnetic network. It has been claimed that the jets would carry enough energy to heat the corona but the observations are still inconclusive.

During the last few years the UV and EUV observations of the SOHO and TRACE satellites have finally shown that there are even larger amounts of (relatively) small explosive events than was previously thought all over the Sun, perhaps some 20 000 every minute. The inner solar atmosphere is very active also during the quiet phases of the solar cycle. The small

activations have been dubbed to **micro-flares** or **nano-flares**. Although this terminology is not quite rigorous, micro can be associated to events of the order of 10^{19} J which you need one million to correspond to a flare and nano to events of 10^{16} J which you need one billion to one flare.

The brightest micro/nanoflares lie above enhanced elements of the magnetic network and the denser events correspond to greater fluctuations. This suggests that the lower corona is not only heated but continuously replenished by chromospheric material that has been heated to coronal temperatures. Thus a substantial part of the energy may come with the heated plasma from below. One scenario would be that the new magnetic field emerges from the Sun in the centers of supergranular cells and is carried to their edges by the convective motion and finally reconnected with the magnetic field from the neighboring cells. The energy released by the reconnection would then power the microflares observed in the overlying low corona.

Recently there has been some argumentation whether the small-scale flares are abundant enough, or not, to account for the coronal heating. Some observations support this interpretation, others do not. However, it is clear that SOHO and TRACE have conclusively shown that there is a correlation between the solar magnetic field and coronal heating. The variability of the small-scale magnetic elements observed in the photosphere (so-called magnetic carpet) has been found to correlate with temperature fluctuations in the corona. Furthermore, observations of the temperature distribution forming **polar plumes** within the coronal holes seem to correlate with photospheric fine-structure associated also to the supergranular structure and magnetic network.

The coronal heating problem is not yet completely solved but there are already so many observational pieces that the final picture is emerging. It may no more need any stroke of genius but collective work by large group of scientists to put all pieces together.

8.3.3 Consequences of the high temperature

The most obvious consequence of the high temperature is the ionization of the atoms. This has a major impact on the MHD of the upper atmosphere because the electric conductivity becomes high and the diffusivity drops dramatically. For an observer it is good news that a suite of emission lines unique to the corona become available. In fact, tuning the (space-borne) instruments to these lines has made most of the discussion of the previous section to something more than pure speculation.

Another consequence is that the free electrons provide large **thermal conductivity**. Inside the Sun the energy transfer is totally dominated by

radiation and convection. In the atmosphere the radiation takes over but in the coronal plasma also the thermal conductivity must be taken into account. Far from the Sun, in the solar wind, the thermal conductivity is a non-trivial issue because in the collisionless magnetized plasma the thermal conductivities of different particle species are different and difficult to determine from theory. Finally the heating is a necessary to drive the supersonic expansion of the solar wind, which will be the topic of the next Chapter.

Chapter 9

Solar wind and the heliosphere

Where is the outer boundary of the Sun? From the previous chapters it is clear that defining the Sun to be the sphere bounded by the photosphere leaves much interesting solar physics out of consideration. In this chapter we learn that within a radius of about $12\,r_\odot$ the Sun with its atmosphere rotates almost as a rigid body. However, the material content of the Sun reaches much farther. As noted at the beginning of this course, the solar wind carries matter away from the Sun about 1 million ton every second. This plasma flow dominates the material content of the interplanetary space beyond the orbit of Pluto until it finally reaches a pressure balance with the interstellar plasma flow. This happens at a boundary called **heliopause** which can be taken as one possible definition of the outer edge of the Sun.

9.1 The expansion of the solar wind

In the 19th century it had become apparent that there was a connection between the solar activity and magnetic disturbances on the Earth, which was not through the electromagnetic radiation. It was, however, very difficult to understand what the connection would be and even Lord Kelvin made a strong statement that any such connection must be fictitious. In 1929 Chapman proposed that the solar flares would emit plasma clouds and if such a cloud would hit the Earth's magnetic field, it would cause magnetic disturbances. But how could such clouds escape from the strong gravitational field of the Sun? After all, the escape velocity on the solar surface is 618 km/s. The kinetic energy of a proton at that speed is 2 keV, which corresponds to a temperature of 2×10^7 K which was too much to be believed in at that time. As we now know, such temperatures do occur in coronal

loops and flares and the escape is no more such a big mystery, although we do not yet know the details how the plasma is heated and accelerated.

In 1951 Biermann noticed that the shapes of cometary tails were consistent with a **continuous** corpuscular outflow from the Sun, unrelated to large flares. Later Alfvén pointed out that the flow must be magnetized plasma and around the year 1962 the existence of continuous magnetized solar wind was confirmed by direct observations outside the Earth's magnetosphere. First indications came from the Russian Lunik III and Venus I spacecraft in 1959, and the definitive proof of its continuous nature was provided by the U.S. Mariner II during 1962–1967.

Today we know that there are two main types of solar wind, a more tenuous wind flowing faster (about 750 km/s) and a denser but slower (about 350 km/s) wind. The details of the source regions and mechanisms are still under investigation, but the general view is that the fast wind originates from large coronal holes at high solar latitudes whereas the slow wind emerges from smaller and less permanent structures at lower latitudes. In addition to these, the CME-related outflow (chapter 10) can be considered as a third independent solar wind type. Solar wind has never disappeared during more than three decades it has been monitored. On May 11, 1999, the slow (300 km/s) wind had for a short time an extremely low density of 0.2 cm^{-3}.

9.1.1 Chapman's static coronal model

An early attempt to describe a continuous flow of energy through the corona by means other than radiation was Chapman's static model in 1957. He considered a sphere of radius r around the Sun and assumed that the thermal flux through the surface (carried by electron heat conduction), $4\pi r^2 \kappa\, dT/dr$, was constant ($\kappa = \kappa_0 T^{5/2}$ is the thermal conductivity). Assuming that $T \to 0$ when $r \to \infty$ this has the solution

$$T = T_0 (r_\odot / r)^{2/7} \,. \tag{9.1}$$

Assuming $T_0 = 10^6$ K this predicts a temperature of 10^5 K at 1 AU, which is in quite good agreement with observations, although it was not known in 1957.

However, the model predicts that the pressure approaches a value that is much too large at large distances from the Sun. Assuming hydrostatic equilibrium

$$\frac{dp}{dr} = -\frac{G m_\odot m n}{r^2} \tag{9.2}$$

and the ideal gas law we obtain

$$p = p_0 \exp\left\{ \frac{7 G m_\odot m n_0}{5 p_0 r_\odot} \left[\left(\frac{r_\odot}{r} \right)^{5/7} - 1 \right] \right\} . \tag{9.3}$$

This means that far from the Sun the pressure approaches a constant that is much larger than the pressure of the interstellar gas. As the temperature decreases toward zero with increasing distance, the density has to increase without bound.

Exercise

Calculate the temperature and and pressure of Chapman's model at 1 AU and at an assumed heliopause distance of 100 AU.

9.1.2 Parker's solution

In 1958 Parker presented another solution to the problem. Parker's solution was based on very strong simplifications and cannot be directly generalized to correspond to the real solar wind. However, its basic idea is important. Parker noted that the corona cannot be in static equilibrium; it must either expand or collapse. He succeeded to predict a supersonic solar wind a few years before the first satellite observations showed that he was essentially right. Parker's argumentation was the following.

Assume time-independent spherically symmetric outward-directed flow. Neglect the magnetic effects and write the continuity equation, momentum equation and equation of state as

$$4\pi r^2 nv = \text{const.} \tag{9.4}$$

$$nmv\frac{dv}{dr} = -\frac{dp}{dr} - \frac{Gm_\odot mn}{r^2} \tag{9.5}$$

$$p = nk_\mathrm{B}T\,. \tag{9.6}$$

Let the expansion be isothermal. This is clearly not true, but it is interesting to see where this leads. The solutions are of the form

$$\left(v - \frac{v_\mathrm{c}^2}{v}\right)\frac{dv}{dr} = \frac{2v_\mathrm{c}^2}{r} - \frac{Gm_\odot}{r^2}\,, \tag{9.7}$$

where $v_\mathrm{c} = \sqrt{k_\mathrm{B}T/m}$ is the isothermal sound speed, i.e., the polytropic index is set to $\gamma = 1$. This equation has a **critical point**: $v = v_\mathrm{c}$, $r = r_\mathrm{c} = Gm_\odot/(2v_\mathrm{c}^2)$. After integration we find the following family of curves

$$\left(\frac{v}{v_\mathrm{c}}\right)^2 - \ln\left(\frac{v}{v_\mathrm{c}}\right)^2 = 4\ln\frac{r}{r_\mathrm{c}} + \frac{2Gm_\odot}{rv_\mathrm{c}^2} + C\,. \tag{9.8}$$

Figure 9.1 illustrates these solutions. Solutions in regions I and II are double valued and, thus, unphysical and those in III have too high (supersonic) source speed. The solution IV crossing the critical point is Parker's

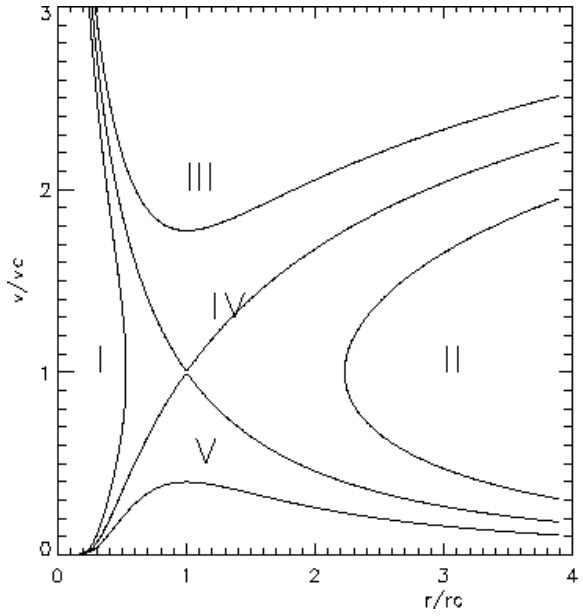

Figure 9.1: Solutions of equation (9.8)

solution for the supersonic solar wind. The critical point fixes the integration constant to $C = -3$. Also V is a physically valid solution, called **stellar breeze**. There are stars which produce subsonic stellar breezes.

Exercise

Calculate the details of Parker's solution.

While elegant, Parker's solution is much too simple for the real solar wind. In fact, as we shall see in the next section, the isothermal polytropic index $\gamma = 1$ leads to a diverging enthalpy. For the most physical assumption $\gamma = 5/3$ there is no critical point and thus the supersonic flow is not described correctly. The wind cools as it expands and thus thermal conduction and heating must be included. Because the plasma is collisionless, ions and electrons cool with different cooling rates and the interaction of plasma with magnetic fluctuations plays different role in electron and ion expansion. Also the detailed models for the fast and slow solar wind are different, because the physical processes in their source regions are different. At 1 AU the observed solar wind properties are summarized in Table 9.1.

Table 9.1: Typical solar wind parameters at 1 AU.

	slow wind	fast wind
v	350 km/s	750 km/s
$n_{\rm e}$	12×10^6 m^{-3}	4×10^6 m^{-3}
$T_{\rm e}$	1.3×10^5 K	1×10^5 K
$T_{\rm p}$	3×10^4 K	2×10^5 K
B	3 nT	6 nT
$v_{\rm A}$	20 km/s	70 km/s

9.1.3 Energy balance

The solar wind transfers energy from the Sun and the energy balance is important. In the corona we must consider kinetic energy, internal energy, gravitational energy, thermal conduction, radiation, and heating. Most of these must also be taken into account in the description of the solar wind acceleration beyond the sound and Alfvén velocities. Here we describe some of the factors to be included.

In a steady state the divergence of the total energy flux must be zero

$$\nabla \cdot \left[\mathbf{v} \left(\frac{1}{2}\rho v^2 + H - \frac{G m_\odot \rho}{r} \right) - \kappa \nabla T + \mathbf{F}_{\rm R} + \mathbf{F}_{\rm H} \right] = 0 \qquad (9.9)$$

Here H is the internal energy (**enthalpy**) and κ the thermal conductivity (in fact, a tensor, but often substituted by a scalar). $\nabla \cdot \mathbf{F}_{\rm R}$ describes the radiation and $\nabla \cdot \mathbf{F}_{\rm H}$ the heating of the upper solar atmosphere. There is a temperature maximum somewhere in the corona. Inside this maximum thermal conduction is inwards, toward the transition region and chromosphere, where it balances the radiative loss through the strong Lyman α line. Outside the maximum thermal conduction is outwards. The Chapman model had only this outward contribution, whereas the original version of Parker's solution neglected it altogether.

The real solar wind departs from a one-fluid behavior already in the corona. Modern model calculations show that ions are heated more efficiently and reach a higher maximum temperature. Further out the ions cool faster than electrons, and at 1 AU their temperature is already smaller. Different contributions to the energy equation have more or less complicated dependencies on the actual magnetic field configuration. We will return to the magnetic field separately.

Further out in the solar wind the radiation and heating become unimportant, but the internal energy of the plasma deserves further consideration. Assume that the coronal gas consists of protons and electrons only, let $n = n(r),\ T = T(r),\ n_{\rm e} \approx n_{\rm i} \approx n$, and neglect, for simplicity, the differences

in the temperatures. Then the pressure is

$$p = n_{\rm e} k_{\rm B} T + n_{\rm i} k_{\rm B} T = 2 n k_{\rm B} T \tag{9.10}$$

and the thermal energy of the gas in a volume V is

$$U = \frac{3}{2}(n_{\rm e} + n_{\rm i}) k_{\rm B} T V = 3 n k_{\rm B} T V \,. \tag{9.11}$$

The gravitational potential is given by

$$\Phi = -\frac{G m_\odot m n V}{r} \,. \tag{9.12}$$

The thermal energy lifts the gas up when the volume V expands. At the same time the internal pressure pushes new gas into this volume performing the work pV. The free energy is the enthalpy

$$H = U + pV = 5 n k_{\rm B} T V \,. \tag{9.13}$$

Assuming a temperature of $T = 2 \times 10^6$ K we find

$$\frac{E}{|\Phi|} \approx 0.5 \,. \tag{9.14}$$

This means that the heating of the corona to this "classic" temperature level does not provide enough free energy to exceed the gravitational potential and the corona should collapse, not expand. Thus there must be some mechanism(s) doing extra work Q on the gas. According to solar wind observations the required energy is about $H + Q = 1.25|\Phi|$. There is no generally accepted theory yet to explain what powers the escape. Most likely it is of magnetic origin and associated to the heating of the ions in the corona.

Assuming that there is enough energy available for the solar wind expansion and neglecting details of $\nabla \cdot (\mathbf{F}_{\rm R} + \mathbf{F}_{\rm H})$, the energy transport equation can be written as

$$n m v r^2 \left(\frac{1}{2} v^2 + \frac{\gamma}{\gamma - 1} \frac{p}{n m} - \frac{G m_\odot}{r} \right) = r^2 \kappa \frac{dT}{dr} + F_\infty \,. \tag{9.15}$$

Here $\kappa = \kappa_0 T^{5/2}$ ($\kappa_0 \approx 10^{-11}$ Wm^{-1}K^{-1}) and F_∞ is the energy flux far from the Sun. This equation takes also the cooling of the expanding (single fluid) wind into account. The internal energy is written in the form showing the polytropic index γ and we immediately see one reason why the Parker solution index $\gamma = 1$ is unphysical.

There are three basically different classes of solutions, depending on the asymptotic behavior of the temperature:

1) $T \sim r^{-2/7}$ heat conduction dominates in the far region
2) $T \sim r^{-2/5}$ kinetic flux dominates in the far region
3) $T \sim r^{-2/3}$ adiabatic expansion

In practice stellar winds belong either to class 1 (cold tenuous winds) or class 3 (hot, dense winds), whereas class 2 is a limiting case between these two. However, the different particle species may fall into different categories. According to observations the proton temperature in the slow solar wind at 1 AU behaves roughly as $T_p \sim r^{-2/3}$ being in the adiabatic class, whereas $T_e \sim r^{-1/3}$, which on the side of thermal conduction.

There are various possible reasons for the different cooling rates. The electrons are bound more tightly to the magnetic field of the solar wind (next section) and electrons and ions react differently to turbulence and wave-particle interactions. Note that while any of these effects may be slow, the spatial and temporal scales are vast compared, e.g., to gyro radii or periods.

9.2 The interplanetary magnetic field

Solar wind carries magnetic field from the Sun to the interplanetary space. This field is called the **interplanetary magnetic field** (IMF). It is maintained by electric currents in the expanding solar wind. While it obviously affects the local properties of the solar wind, it has also consequences on the solar rotation and on the behavior of plasma environments of solar system bodies.

9.2.1 The Parker spiral

The observed structure of the IMF varies considerably from the ecliptic to the poles. To begin, let us consider a cylindrically symmetric case in the equatorial plane. Assume that the flow is radial and let Ω be the angular speed of the solar rotation. Let the angle between the radial direction and the magnetic field be ψ and assume that the IMF is frozen into the expanding solar wind. Close to the Sun the plasma rotates with the body but in the radial expansion the field is wound to a spiral. Let $\mathbf{v}$ be the flow velocity. Its component perpendicular to the field is $v_\perp = v \sin\psi$. This can be imagined as the speed of the field line in this direction (although we should again be careful of not addressing a physical reality to the field lines). The high conductivity ties the field line to the surface of the Sun (or to the source surface, discussed in section 8.2). Thus the speed of the field line perpendicular to the radial direction is $\Omega(r - r_\odot)$, and

$$v \sin\psi = \Omega(r - r_\odot)\cos\psi \tag{9.16}$$

$\Rightarrow$

$$\tan\psi = \frac{\Omega(r - r_\odot)}{v}\,. \tag{9.17}$$

When r increases, this approaches the Archimedean spiral. This is called the **Parker spiral**.

Let $\mathbf{B}$ be radial and constant on the surface of the Sun and write $\mathbf{B}$ and $\mathbf{v}$ in spherical coordinates determined by the rotation axis of the Sun, (r, θ, ϕ):

$$\mathbf{B} = (B_r, 0, B_\phi), \quad \mathbf{v} = (v_r, 0, v_\phi)\,. \tag{9.18}$$

Note that the components of the vectors are functions of r. From $\nabla \cdot \mathbf{B} = 0$ we get

$$B_r = B_0(r_\odot/r)^2\,. \tag{9.19}$$

Thus the radial component of the field decreases as $1/r^2$. To find the azimuthal behavior we can write the force balance as

$$\rho(\mathbf{v}\cdot\nabla\mathbf{v})_\phi = (\mathbf{J}\times\mathbf{B})_\phi\,, \tag{9.20}$$

assuming that $(\nabla p)_\phi = 0$. Using Amperé's law and multiplying by r^3 we obtain

$$r^2\rho v_r\frac{d}{dr}(rv_\phi) = \frac{1}{\mu_0}r^2B_r\frac{d}{dr}(rB_\phi)\,. \tag{9.21}$$

The mass flux $r^2\rho v_r$ and the magnetic flux r^2B_r are constants and we can integrate this equation to get

$$L = rv_\phi - \frac{rB_rB_\phi}{\mu_0\rho v_r}\,, \tag{9.22}$$

where L is a constant of integration. The constancy of L reflects the effect of the magnetic torque on the specific angular momentum.

To express B_ϕ in terms of B_r we consider the frame which rotates with the angular speed Ω. In this frame the velocity vector is $(v_r, 0, v_\phi - r\Omega)$. This vector is parallel to $\mathbf{B}$ and thus

$$B_\phi = \frac{V_\phi - r\Omega}{V_r}B_r\,. \tag{9.23}$$

At large distances $B_\phi \propto r^{-1}$, i.e., it decreases more slowly than the radial component, which explains the spiral formation.

Define now the radial **Alfvén Mach number** M_A

$$M_\mathrm{A} = \frac{v_r}{v_A} = \frac{v_r\sqrt{\mu_0\rho}}{B_r} \tag{9.24}$$

$\Rightarrow$

$$v_\phi = \Omega r\,\frac{M_\mathrm{A}^2\left(\dfrac{L}{r^2\Omega}\right) - 1}{M_\mathrm{A}^2 - 1} \tag{9.25}$$

According to observations $M_{\rm A}$ increases from ~ 0.1 in the corona to ~ 10 at 1 AU. Thus there is a critical point in the expression of v_ϕ where $M_{\rm A} = 1$ at a certain distance $r = r_{\rm A}$ which is called the **Alfvén radius**. At this distance the flow becomes super-Alfvénic. According to observations this takes place at the distance of about $12\, r_\odot$. As the azimuthal speed cannot be singular at that point we find that the constant of integration is

$$L = \Omega r_{\rm A}^2 \,. \tag{9.26}$$

This is equal to the specfic angular momentum of a corotating (with the Sun) particle at radius $r_{\rm A}$.

We can now write the azimuthal velocity as

$$v_\phi = \frac{v_r/v_{\rm A} - 1}{(v_r r^2)/(v_{\rm A} r_{\rm A}^2) - 1}\, \Omega r \,. \tag{9.27}$$

Close to the Sun this reduces to

$$v_\phi \simeq r\Omega \,, \tag{9.28}$$

i.e., rigid rotation with the Sun. On the other hand at large distances

$$v_\phi \simeq r_{\rm A}^2 \Omega / r \,, \tag{9.29}$$

which expresses the conservation of angular momentum from the Alfvén radius outward. Thus, $r_{\rm A}$ is an effective lever arm, with which the solar wind brakes the solar rotation. This effect is called **magnetic braking**.

Out of the equatorial plane ($\theta \neq \pi/2$) the calculation is more complicated. The azimuthal component of the field turns out to be

$$B_\phi \approx -\frac{B_0 r_\odot^2 \Omega \sin\theta}{r v_r} \,. \tag{9.30}$$

Thus, far from the Sun the total magnetic field behaves as

- $B \to r^{-1}$ in the equatorial plane (the spiral becomes tightly wound)
- $B \to r^{-2}$ in the direction of the poles

Between the equatorial plane and the polar direction the field gets a helical structure. At 1 AU the equatorial spiral angle is about 45°.

Exercises

1. Derive equation (9.30)

2. Show that the mass and angular momentum losses are related by

$$\frac{dJ}{dt} = \frac{2}{3}\Omega r_{\mathrm{A}}^2 \frac{dm}{dt} \tag{9.31}$$

   Calculate this number for the present Sun and estimate how long time the present magnetic braking would need to stop the rotation. How much more efficient the magnetic braking is for $r_{\mathrm{A}} = 12\ r_{\odot}$ than it would be for $r_{\mathrm{A}} = r_{\odot}$?

The inclusion of the magnetic field introduced a new critical point analogous to the sonic point in Parker's solution. Actually, in MHD flow there are three critical points corresponding to the distances where the flow speed exceeds the slow, intermediate and fast MHD wave speeds.

9.3 The observed structure of the solar wind

Finally, the real solar wind structure is much more complicated than the simple calculations above would suggest. The escaping flow originates from the coronal holes whose shape and location change all the time. SOHO observations looking through the holes to the photosphere show further that the escape is highly structured within individual holes. In addition to this variability the solar eruptions eject faster or slower plasma clouds to the background solar wind flow. These structures can drive various shock phenomena in the wind to large distances beyond the Earth orbit.

When the solar activity is at its minimum the solar magnetic field is as poloidal as possible and the coronal structure is dominated by two large polar holes with opposite magnetic polarities. The almost radial solar wind flow escapes mostly (if not completely) from these holes and drags the frozen-in magnetic field in such a way that a **heliospheric current sheet** forms near the equatorial plane. However, as the holes have a very asymmetric shape, the current sheet is asymmetric as well (Figure 9.2). When the Sun rotates, the current sheet moves up and down, which led Alfvén to call this structure the ballerina's skirt. The Earth is either above or below the skirt. Depending on whether the field is pointing mostly toward or away from the Sun, the Earth is said to be either in the **toward sector** or the **away sector**. Superposed to this large-scale structure there are large variations in all components of the IMF.

Around solar maximum the solar magnetic field structure is much more complicated and the polar coronal holes are reduced in side. On the other

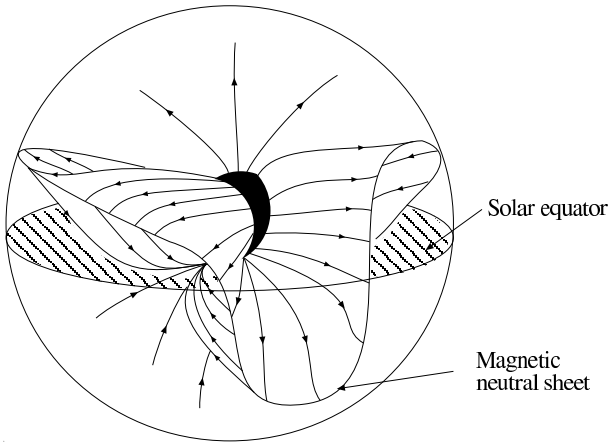

Figure 9.2: The ballerina skirt formation of the solar wind follows the shape and location of the dominating polar coronal holes.

hand, there are more smaller-scale opening and closing structures at lower latitudes. This makes also the solar wind structure more variable, which in turn drives magnetic activity in the terrestrial environment.

While the structure and magnetic field behavior in the polar directions can be inferred theoretically and even depicted in the polar plumes and coronagraph images, it was not until the 1990's when first direct observations of the off-ecliptic solar wind behavior became available through the joint ESA and NASA spacecraft Ulysses, which was the first one to leave the ecliptic.

Due to the large velocity orbital velocity of the Earth (30 km/s), Ulysses had to be sent first to Jupiter to be deflected from the ecliptic plane using the planet's large gravitation field. Consequently, the final orbit of the spacecraft has its aphelion at the distance of Jupiter, 5.3 AU, and perihelion at 1.3 AU, with the highest heliographic latitude of 80°. Ulysses reached this point for the first time above the southern solar hemisphere in September 1994 and above the northern hemisphere in March 1995, and the next polar passages were in 2000 and 2001. Early in 2004 Ulysses was again at its aphelion and began its third orbit. This was also the last one, as the mission was terminated in June 30, 2009.

Note that the large variability of the solar wind speed at 1 AU (table 9.1) is partially due to the variable vertical distance from the heliospheric current sheet. The slowest speeds of the solar wind arise near the edges of the polar coronal holes and from coronal holes elsewhere. This is nicely illustrated by Ulysses observations (Figure 9.3). When the spacecraft was within $\pm 20°$ from the ecliptic, it observed both slow and fast solar wind, but at higher latitudes it encountered only fast tenuous solar wind from the polar coronal

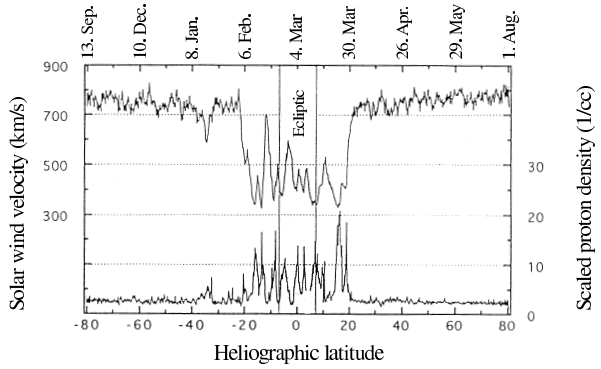

Figure 9.3: Ulysses observations of solar wind speed (upper curve) and proton density (lower curve) as a function of the heliographic latitude.

holes. Note that these observations of solar wind structure correspond to the period of minimum solar activity. During solar maximum, the large polar coronal holes are practically absent.

Ulysses has made a large number of precision measurements of the solar wind and its structure. One of the surprising results was that the polar magnetic field is not as strongly convergent toward the pole as expected. It is less dipolar meaning that the magnetic flux is distributed more evenly over a large area. This has consequences to the entry of cosmic rays to the solar system.

9.4 Shocks in the solar wind

The solar wind is an excellent laboratory for ideal MHD in a great variety of scale lengths. Due to the very large magnetic Reynolds number the solar wind exhibits a very pure form of well-developed turbulence. However, we still lack a complete description of plasma turbulence and this may be a reason to our inability to provide satisfactory explanation for different cooling rates of electrons and protons. Related to turbulence and MHD waves are the shocks introduced in section 6.4. There are several types of shocks in the solar wind. High-speed CMEs drive shocks, the interaction regions between sectors of fast and slow solar wind develop to shock structures with the winding up of the IMF spiral, planets are obstacles to the solar wind flow causing shocks and finally when the solar wind meets the interstellar plasma, it somewhere becomes subsonic and forms a termination shock structure somewhere inward from the heliopause.

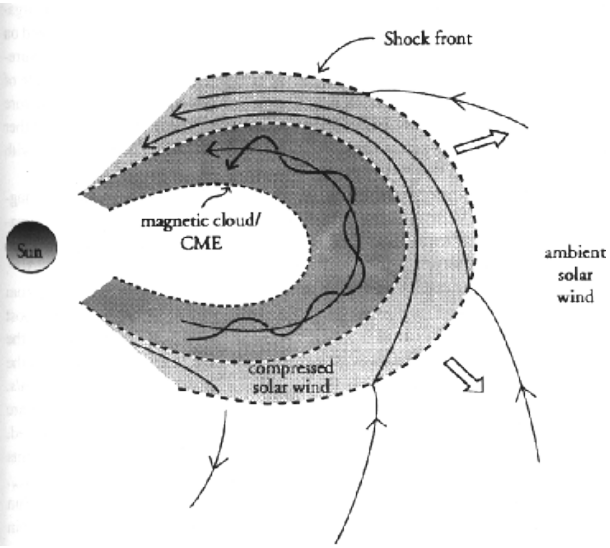

Figure 9.4: A sketch of a shock driven by a CME. Note that the magnetic field of the CME can have different orientations and often has a flux-rope structure with a strong field in the direction perpendicular to the plane of the figure.

9.4.1 Shock structures emerging from the Sun

Perhaps the most important class of solar wind shocks are those driven by the fast CMEs. As noted in the previous chapter, the CMEs originate with largely variable speeds from a few tens of km/s to about 2000 km/s. The slowest ones are soon accelerated close to the speed of the ambient solar wind flow whereas the fast ones are decelerated. The slow CMEs do not drive shocks, except maybe close to the Sun, whereas the fast ones do as long as their speeds relative to the ambient plasma are supersonic (or actually super-Alfvénic). This is clearly the case with a large number of CMEs observed at 1 AU.

Close to the Sun the CME-related shocks (Figure 9.4) are important in the acceleration of solar energetic particles. When a CME and the shock ahead of it hit the magnetosphere of the Earth they shake the system and depending on the magnetic structure of the system they drive the most severe magnetic storms in the terrestrial environment.

Another important class of solar wind shocks develop during the outflow. Figure 9.5 illustrates the formation of these **corotating interaction regions** (CIR). Consider a given direction in the non-rotating frame. Assume that at first slow wind is blowing into this direction. As the Sun rotates, this direction becomes one of fast wind. Thus this faster and less dense flow keeps catching the slow and denser flow. However, both components are

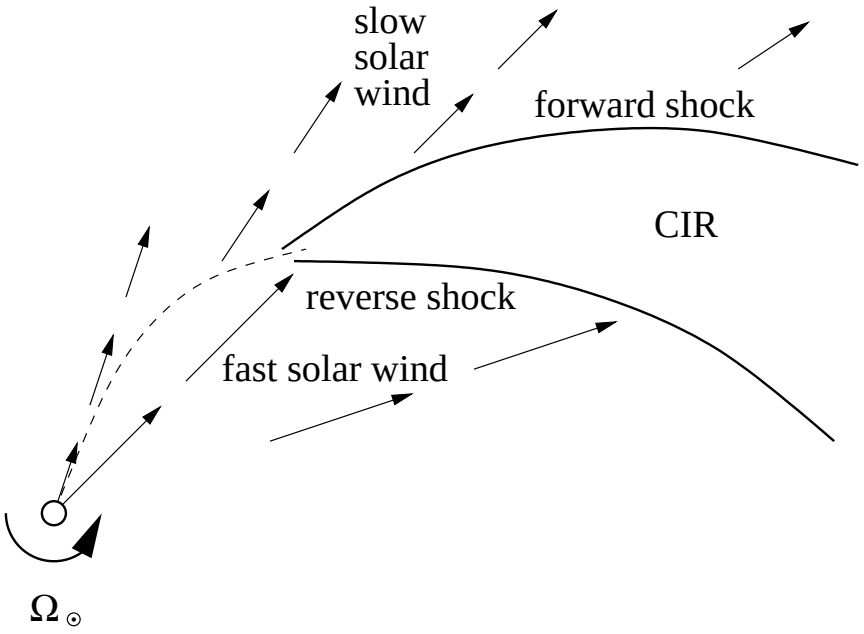

Figure 9.5: Formation of a corotating interaction region.

very ideal MHD flow and cannot mix easily. If the relative flow speeds differ more than the local Alfvén speed a steepening boundary structure forms. Close to the Sun the field lines are still nearly radial and the boundary is more like a tangential discontinuity. Further out the Parker spiral is wound more tightly in the slow wind ahead the structure and less tightly in the fast wind behind. A fully developed CIR exhibits a **forward shock** ahead the structure and a **reverse shock** behind it. Note that one has again to be careful with the frame of reference! The forward shock may sound natural but so is also the reverse shock as in the frame of the reference of the fast flow the shock is propagating toward the observer.

At 1 AU the typical spiral angle is about 45° and the CIRs become fully developed only beyond the Earth's orbit.

9.4.2 Planetary obstacles in the plasma flow

In the solar wind frame planets and other solar system bodies are objects moving with supersonic speeds. However, if there is nothing in the environment of the body that would deflect the collisionless solar wind before the particles hit the surface of the body, only a minor downstream cavity is formed without supporting a shock structure. This is what happens with the Moon, most other moons of the other planets, and asteroids, and we call this **Moon-like interaction**.

On the other hand, if a planet has a strong enough internal magnetic field, the magnetic force deflects the solar wind particles around a structure that we call the **magnetosphere**. All magnetic planets (Mercury, Earth, Jupiter, Saturn, Uranus, Neptune) are known to form a magnetosphere which is essentially a magnetic cavity in the solar wind. This type of interaction between the body and the solar wind is often referred to as **Earth-like interaction**. Also planets with a dense enough atmosphere (Venus and Mars) form structures that are called **induced magnetospheres**. Now the deflection is not due to the magnetic field but due to the inability of the solar wind plasma to penetrate through the ionized atmosphere. This type of interaction is called **Venus-like**. Finally, also comets, when they are active close to the Sun, have their own type of interaction with the solar wind (**comet-like interaction**). In this case charged particles, which are ionized from the large neutral cloud expanding out from a small cometary nucleus, add new mass to the solar wind. These four interaction types are manifestations of how the Sun affects various Solar System bodies by its gravitation, electromagnetic radiation and with its magnetized plasma, the solar wind.

In the following, we illustrate the solar wind interaction with the magnetosphere in the case of the Earth. Note, however, that in the case of induced magnetospheres the shock formation is not as clean as here, because there is a considerable amount of neutral gas and dust outside barrier against the solar wind. Part of this gas becomes ionized and the ions feel the $-\mathbf{v} \times \mathbf{B}$ electric field and are accelerated to follow the bulk solar wind flow. They are said to be **picked-up** by the solar wind.

When the supersonic solar wind hits the magnetic field of the Earth, it pushes the magnetic field on the dayside and stretches it to a very long tail on the nightside. In the first approximation the ideal solar wind and magnetospheric MHD plasmas cannot mix and a well-defined **magnetopause** forms. The distance from the center of the Earth to the dayside magnetopause can be estimated calculating the pressure balance between the magnetic pressure inside the magnetopause and the solar wind dynamic pressure

$$K \rho_{\mathrm{SW}} v_{\mathrm{SW}}^2 \cos^2 \psi = \frac{B_{\mathrm{MS}}^2}{2\mu_0} \,, \tag{9.32}$$

where ψ is the angle between the magnetopause and the solar wind direction. K is a constant that would be 2 for an elastic collision (pure reflection) and 1 for a purely inelastic collision (absorption). In the fluid deflected around the obstacle K depends on the upstream Mach number and is typically about 0.9. Nominal value for the subsolar distance is about 10 R_{E} (1 $R_{\mathrm{E}} \approx 6370$ km).

As the magnetopause shields the planetary magnetic field from the solar wind, currents must flow on the magnetopause. Similarly the stretching

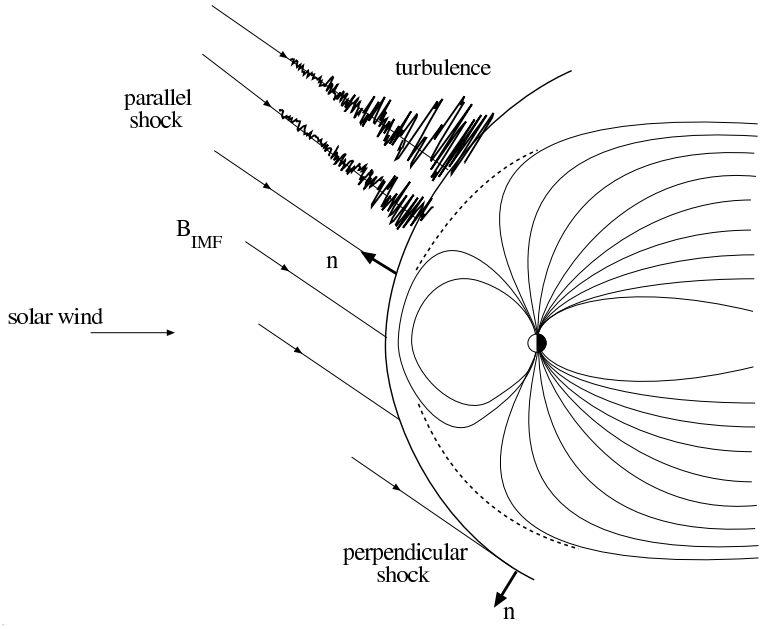

Figure 9.6: A magnetosphere and its bow shock.

of the long magnetotail causes a strong current inside the magnetosphere. Thus the solar wind-magnetosphere interaction must drive currents in the system. Much of the magnetospheric physics deals with various aspects of these currents.

The shock upstream the magnetosphere is called the **bow shock**. For typical solar wind parameters the nose of the shock is about 3 R_{E} upstream from the nose of the magnetopause. The shock is clearly a fast MHD shock and it converts a considerable amount of solar wind kinetic energy to heat and Poynting flux.

Due to the nearly paraboloid shape of the shock a part of it is quasi-parallel and another quasi-perpendicular, depending on the actual direction of the IMF. Thus it is possible to experimentally study the properties of different types of MHD shocks during different solar wind conditions. The quasi-perpendicular side is a well-behaving thin shock whereas the quasi-parallel side is messy on both sides as individual particles can stream back and forth through the shock region along the magnetic field lines without being affected too much by the shock structure.

9.5 The heliosphere and the heliopause

The solar wind dominates the physics of space far beyond the planetary system (sometimes called solar system, but we should perhaps include the

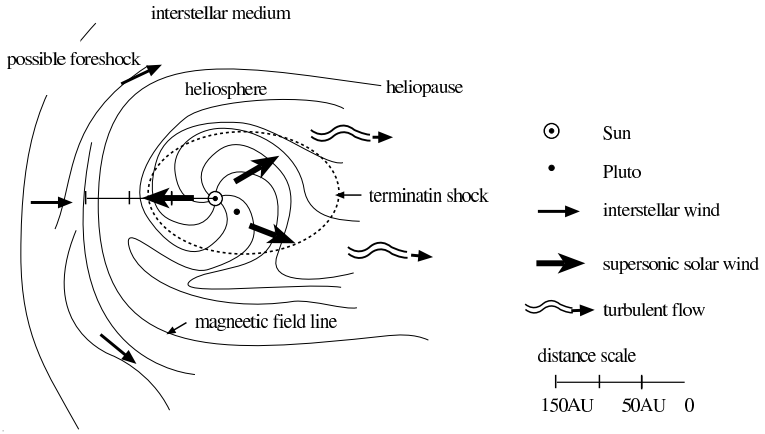

Figure 9.7: The heliosphere.

Oort cloud of comets to the solar system as well). There is no commonly used definition for the heliosphere but in the spirit of these lectures we can take it to be the region from the Sun to the heliopause.

So far we do not have any fully convincing observations of the location of the heliopause. Some flashes of radio waves from the anti-sunward direction have been speculated to be caused by interaction of solar eruptive material with the heliopause, but as the distances are vast and thus the transfer times very long, these are highly speculative. The location of the termination shock, however, is better known: in December 2004, Voyager 1 became the first spacecraft to enter the heliosheath. At that time, its distance to the Sun was 94 AU. On September 1, 2007, Voyager 2 followed and left the supersonic solar wind at 84 AU from the Sun.

We can try to estimate where the heliopause should be. From Lyman-α emissions we can estimate that the neutral hydrogen density is about 0.1 cm^{-3} and it is streaming toward the Sun from a certain direction at the speed of about 25 km/s. However, neutral hydrogen does not interact with the solar wind plasma in the extremely collisionless environment but penetrates almost freely toward the Sun. The degree of ionization of the local interstellar matter is not known very well and the proton density estimates vary in the range 0.001–0.05 cm^{-3}. Also the local interstellar magnetic field is difficult to estimate but it may be as large as 0.3–0.5 nT.

Denote in the following the interstellar parameters by G and the solar wind parameters by S. It is likely that the distance between the termination shock and the heliopause is large and the sheath structure in between may cause some problems. Neglecting this we can derive a first approximation to the heliopause distance in the interstellar upstream direction from the

pressure balance equation

$$\rho_{\rm S} v_{\rm S}^2 + \frac{B_{\rm S}^2}{2\mu_0} = \rho_{\rm G} v_{\rm G}^2 + p_{\rm G} + \frac{B_{\rm G}^2}{2\mu_0}\,, \tag{9.33}$$

where the terms ρv^2 indicate the dynamic pressures on each side and in the interstellar gas also the static pressure is taken into account.

On the solar wind side the speed remains almost constant throughout the heliosphere, the density decreases as r^{-2} and the magnetic field (in the near ecliptic plane) as r^{-1}. Thus the Alfvén Mach number remains almost constant, somewhat below 10. Due to the large speed the dynamic pressure dominates everywhere in the heliosphere.

On the interstellar side everything is much more uncertain. $B_{\rm G} = 0.5$ nT implies a pressure of 10^{-13} N/m^2 (i.e., 0.1 pPa). Assuming a high proton density of 0.1 cm^{-3} together with the known (neutral gas) flow speed of 25 km/s implies a dynamic pressure also of the order of 0.1 pPa, whereas the static pressure, assuming a temperature of 10^4 K, yields a pressure that is of the order of a magnitude smaller. As the plasma density most likely is smaller than 0.1 cm^{-3}, it is possible that the dominating pressure on the interstellar side is magnetic and the total pressure at most 0.1 pPa.

We can estimate the solar wind dynamic pressure using typical parameter values at 1 AU. The speed is roughly constant whereas the density scales as

$$\rho_{\rm S} = \rho_{1\,\rm AU}(1\,\rm AU/r_{\rm HP})^2\,, \tag{9.34}$$

which, when inserted into the pressure balance equation, gives us the distance to the heliopause ($r_{\rm HP}$). From this we find the distance to the heliopause to be about 115 AU. This value is somewhat small, because most MHD models predict the termination shock to be at a distance of 2/3 of the heliopause, and the distance to the termination shock is about 90 AU towards the upwind direction. (Downstream the distance is, of course, larger.) Thus, the total pressure on the interstellar side might be even slightly smaller than our rough estimate above.

It is not known if there is a shock also in the interstellar flow upstream of the heliopause. This depends essentially on whether the solar speed of about 25 km/s in the interstellar plasma is supersonic/Alfvénic or not. If it is, there should be an upstream shock as well.

As noted above, due to low collision frequencies the neutral interstellar hydrogen ($n \approx 0.1$ cm^{-3}) and helium ($n \approx 0.01$ cm^{-3}) can penetrate deep into the heliosphere. On this long journey (25 km/s is very slow speed indeed) they, however, interact with fast solar wind protons and experience **charge-exchange** collision with them. About 80–90% of the ionization is due to this process. Most of the remaining ionization is due to solar UV

radiation. As the newly-born ions in the solar wind are much slower and, in fact, move against the ambient flow they experience a $-\mathbf{v} \times \mathbf{B}$ electric field that accelerates them to the ambient flow. This is a similar pick-up process as in the solar wind interaction with induced magnetospheres. It reduces the solar wind flow slightly and increases its mass flux. Finally, close to the Sun there is an ionization cavity where the density of interstellar gas is very small.

Chapter 10

Solar eruptions

In this chapter we discuss the violent behavior of the Sun, which is observable in a wide range of wavelengths (or energies) but is perhaps most dramatic in radio waves (long wavelengths, low energies), and in UV emissions and X- and γ-rays (short wavelengths, high energies). These emissions are signatures of particle acceleration to very high energies during the eruptions. Two of the most important eruptive phenomena in the atmosphere of the Sun are **flares** and **coronal mass ejections** (CMEs). These are related to each other although it is not yet quite clear how. In order to understand this relation we begin with a more stable phenomenon called **prominence**.

10.1 Prominences

Prominences are perhaps the most beautiful solar phenomena at the limb of the Sun. There are written records of prominences from the Middle Ages and they were rediscovered during the eclipse of 1842 and photographed for the first time in 1860.

Prominences are giant gas clouds above the solar suface supported by the magnetic field. On the limb they can be seen against the dark background because the plasma density is high ($n \approx 0.5 - 1.0 \times 10^{17}$ m^{-3}), a few hundred times larger than the ambient coronal density. Correspondingly the prominences are a few hundred times cooler than the corona (5000 – 10 000 K). As this is also below the chromospheric temperature, the prominences look like dark **filaments** in Hα images against the chromosphere.

The clouds typically have the form of vertical sheets. These sheets are remarkably stable, surviving up to 300 days and they can be used in studies of the solar rotation. There are two basic types of prominences, quiescent and active-region prominences. The quiescent prominences are typically the

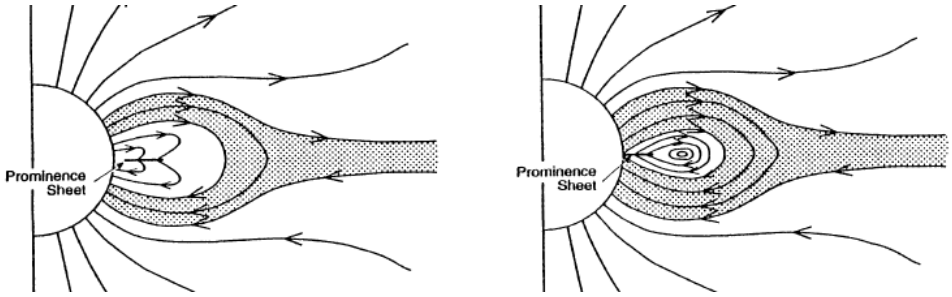

Figure 10.1: Magnetic field configuration in association with prominences. On the left the polarity of the prominence is normal, on the right it is inverse.

larger of these and extend higher up (above 30 000 km). Their magnetic field is 0.5–1 mT. In the active regions the prominences are smaller and mostly below 30 000 km, but their density is somewhat higher and the magnetic field is stronger (2–20 mT). The prominences often decay away but in active regions they can erupt in association with solar flares and/or CMEs. In fact, a disappearance of a filament is often used to determine the location where a CME originated.

The line-of-sight (vertical) component of the magnetic field reverses over a filament but the direction in which the field passes through the prominence may be the same as we would expect for a simple arcade (normal polarity) or opposite to it (inverse polarity). Figure 10.1 illustrates these two configurations. The high-latitude quiescent prominences have mostly the inverse polarity whereas the active-region prominences can have have either of the polarities.

The filament itself is a current sheet with the sheet current (units of A/m) directed either away from the page (normal polarity) or into the page (inverse polarity) in figure 10.1.

Since the temperature in the prominence is only of the order of 10^4 K, the gas can not be in pure hydrodynamic equilibrium: this temperature would require a scale height of about 300 km, which contradicts the observed height of prominences, typically of the order of 50 000 km. Thus, the magnetic force $\mathbf{j}\times\mathbf{B}$ in the current sheet plays an important role in supporting the cool plasma against gravity.

It is not quite clear how the prominences form. First there must be a flux tube rising through the solar surface as an arc. It has been suggested that when the distance between the foot points of the flux tube becomes too long a radiative instability sets in and the cool plasma starts to accumulate in the central part of the arc. Also twisting of the flux tube may be important in the process (Figure 10.2). The twist forms a magnetic dip in the lower part of the horizontal portion of the flux-tube and the cool plasma starts to

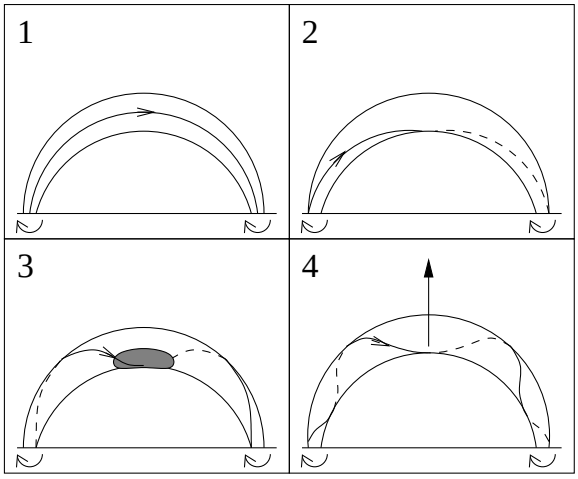

Figure 10.2: The formation of a prominence in a twisted-flux-tube model by Priest et al.

accumulate there. A twist becoming too strong is a possible explanation for the eruption of the flux tube.

10.2 The Sun in radio wavelengths

Before we go to the details of solar flares and coronal mass ejections, we open a new window to the Sun: the radio waves. The importance of radiowaves in solar physics is based on two facts already shown in Chapter 1: The Earth's atmosphere is transparent for electromagnetic waves that are longer than a few mm (i.e., frequencies below about 100 GHz). On the other hand the solar spectrum at these wavelengths is extremely variable reflecting both the high coronal temperatures and violent activity.

During World War II amateur radio operators noticed noise (or hiss) that occurred only in daytime, and the first radar systems operated at meter wavelengths were occasionally jammed by radio inference coming from the Sun. Soon after the war also intense radio outbursts were detected. There is actually a whole zoo of different types of radio emissions (Figure 10.3). They are named mostly according to their observed properties, which may be a nuisance for a student but appropriate for scientific study as the nomenclature is free from (possibly erroneous) physical interpretation. The most important emissions are grouped as Types I–IV.

Type I bursts are also called Type I noise storms. They are long-lived (hours to days) sources of radio emission with **brightness temperatures** from $10^7 - 10^9$ K. Recall that the brightness temperature is the temperature

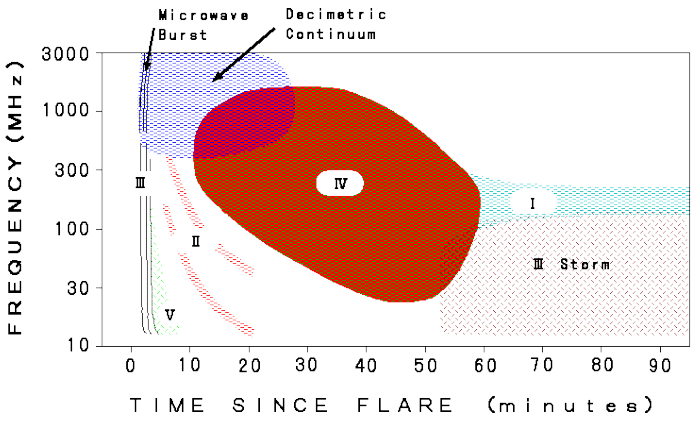

Figure 10.3: Schematic signatures of solar radio events in a dynamic spectrum. The timing is relative to the evolution of a typical solar flare.

that the source would have if it was emitting thermal radiation. In the solar corona the particles emitting electromagnetic waves do not need to be in thermal equilibrium with the surrounding plasma, nor is the radiation mechanism always thermal. Type IV bursts are the most common type of activity on the Sun at meter wavelengths but they are not directly associated with solar flares. They are attributed to electrons accelerated to modest energies of a few keV within large-scale magnetic loops that connect active regions to more distant areas of the Sun.

Type II bursts have been observed at frequencies 0.1–100 MHz. The slow drift to lower frequencies at a rate of about 1 MHz per second suggests an outward motion at about 1000 km/s and has been attributed to outward propagating shock waves.

Type III bursts are the most common flare-associated radio bursts at meter wavelengths, observed in the range 0.1 MHz – 1 GHz. They are characterized by a fast drift from high to low frequency at a rate of up to 100 MHz per second. They are attributed to beams of electrons thrown out from the Sun with kinetic energies of 10–100 keV, or velocities up to $c/2$. This type of emission is called **non-thermal** as the surrounding plasma cannot thermalize to temperatures corresponding to the kinetic energy of the emitting particles.

Type IV bursts are broad-band continuum radiation lasting for up to one hour after impulsive flare onset. The radiation from a Type IV burst is partly circularly polarized, and has been attributed to synchrotron emission from energetic electrons trapped within expanding magnetic loops that travel out into space with velocities from several hundreds of km/s to about one thousand km/s.

As precursors of solar flares impulsive microwave bursts, which last only a few milliseconds, are observed. Their radiation temperatures can reach up to 10^{15} K which requires a coherent radiation mechanism, e.g., a synchrotron maser.

Radio emissions occur in perturbed plasma layers of the solar atmosphere and are mainly due to free electrons moving in the magnetized plasma. An important lesson from basic plasma physics is that a free-space electromagnetic wave can propagate only if its frequency is higher than the local (electron) **plasma frequency**

$$f_{pe} = \frac{1}{2\pi}\sqrt{\frac{n_e e^2}{\epsilon_0 m_e}}\,, \tag{10.1}$$

whose approximate numerical value is given by

$$f_{pe}(\mathrm{Hz}) \approx 9 \times \sqrt{n_e(\mathrm{m}^{-3})}\,. \tag{10.2}$$

In the chromosphere the electron density drops from 2.5×10^{17} m^{-3} to 10^{16} m^{-3}, which correspond to plasma frequencies from 4.5 GHz to 0.9 GHz, or wavelengths $6.7 - 33.3$ cm. In the corona the electron density decreases as

$$n_e(r) = 1.55 \times 10^{14}\, r^{-6} + 2.99 \times 10^{14}\, r^{-16}\,, \tag{10.3}$$

where the distance is in solar radii (from the center of the Sun) and the density in m^{-3}. [Note that this model is valid only as an average density model; variations in coronal density structure in the lateral directions and as a function of time are important as well.] At 2 $r_\odot$ the plasma frequency is thus about 14 MHz corresponding to 21 m. If we observe radio waves longer than 21 m, we know that they must have been emitted at least at $2\,r_\odot$ from the center of the Sun.

This exercise is, in fact, more useful than just the determination of the minimum distance. Shock waves rising from the Sun perturb the electron density and drive plasma oscillations at the local plasma frequency and its first harmonic. [Note that solar radio astronomers call the harmonic emission at $f = 2\,f_{pe}$ the second harmonic.] This produces the Type II emissions that propagate outward and can be identified by their falling tone. From this we can calculate at what distance from the Sun the shock wave was at a given time.

The energy density in radio waves is extremely small. For this purpose the radioastronomers have introduced a particular unit, **Jansky**, 1 Jy = $10^{-26}\,\mathrm{W\,Hz^{-1}\,m^{-2}}$. In solar astronomy a four orders of magnitude larger unit, the **Solar flux unit** sfu is used

$$1\,\mathrm{sfu} = 10^{-22}\,\mathrm{W\,Hz^{-1}\,m^{-2}}\,. \tag{10.4}$$

It is also important to determine the circular polarization state and degree of the radiation as these depend on the emission mechanism and the emission geometry. For example, the electron plasma emission (Type II) is unpolarized whereas cyclotron and synchrotron emissions (Type III and IV) in strong magnetic fields are polarized.

Exercise

Estimate the total power of solar radio waves hitting the surface of the Earth in the wavelength range 10 cm – 1 m.

10.3 Solar flares

Solar flares are the most efficient mechanism to release magnetic energy associated with active region filaments (i.e., prominences), sunspots or sunspot groups. For example, the total magnetic energy of an arcade with a radius of 20 000 km, length 100 000 km, and shear angle 45° is about 6×10^{25} J which is enough for a large flare.

10.3.1 Characteristics of the flares

A flare was first observed in white light by Carrington in 1859. Unfortunately, only large flares can be observed in white light. More characteristic for flares are certain line emissions (again Hα is an important line) and consequences of the rapid energy release. In a flare electrons are accelerated typically to energies of 10–100 keV, sometimes up to 10 MeV, and the highest energy nuclei may reach to hundreds of MeV or even some GeVs per nucleon. These particles emit electromagnetic radiation throughout the spectrum from radio waves to X- and γ-rays. Figure 10.4 shows the development of a typical flare as seen at various wavelengths.

A characterization of the flares and the flare sequence is rather complicated because the different signatures develop in different ways. The chromospheric eruption is easiest to observe using the Hα line. It shows a "flash" of the flare which takes a few minutes. At the early part of this flash the γ-rays, hard X-rays, EUV, and microwaves indicate an **impulsive phase**. This is followed by the **main phase** or **decay phase** which last from 30 min to 1 hour. Just before the flash there is a brief precursor characterized by thermal radiation corresponding to a temperature up to 10^7 K.

There are no reliable methods of predicting flares in advance. Because the flare accelerated particles are of concern for space-flight, the development of the active regions are monitored continuously and information of possible

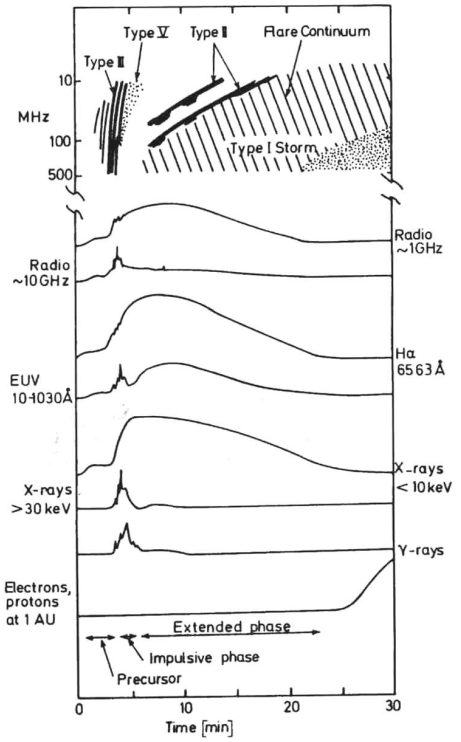

Figure 10.4: Various observational indicators of a flare.

flare activity is spread throughout the world. However, the quality of these predictions compared to the statistical appearance of flares (so-called **skill**) in different phases of the solar cycle is still quite poor. In addition, the most harmful energetic particles (ions at energies > 10 MeV/nucleon) are also accelerated in fast CMEs, some of which are not related to large flares.

Flares and radio waves

The radio emissions arise from the energetic electrons in the magnetized plasma. Figure 10.4 repeats the main radio wave characteristics discussed in the previous section. The most important radio flare emissions are the Type II and Type III bursts. Type III emission drifts rapidly down in frequency which is explained to indicate a fast motion ($10^7 - 10^8$ m/s) of electrons in the 10–100 keV range. This is consistent with the required energy to produce the hard X-rays observed simultaneously. Type II emission appears somewhat later and indicates a lower velocity of 10^6 m/s. Type II is interpreted to be emitted by a shock wave generated by the flare, or an associated CME, which propagates outward through the corona. Note, however, that in many solar eruptions, type III burst emissions continue over an extended time period

(up to hours) after the flare flash phase indicating that particle acceleration in the solar corona occurs over an extended time period as well.

X-ray flares

The solar X-rays are totally absorbed in the atmosphere and can thus be observed only in space. The first observation of the flare X-rays was made by *Peterson* and *Winckler* using a high-altitude balloon in 1958. They observed radiation in the energy range 200–500 keV, lasting less than a minute and coincident in time with a solar radio burst and an Hα flare.

Nowadays solar X-rays are monitored regularly and their intensity as measured by the GOES satellites is readily available at the NOAA site: `http://www.swpc.noaa.gov/` The intensity is indexed in classes A ($10^{-8} - 10^{-7}$ Wm^{-2}), B ($10^{-7}-10^{-6}$ Wm^{-2}), C ($10^{-6}-10^{-5}$ Wm^{-2}), M($10^{-5}-10^{-4}$ Wm^{-2}), and X($> 10^{-4}$ Wm^{-2}). Within the classes the intensity is given as, e.g., M7.5, which corresponds to the flux of 7.5×10^{-5} Wm^{-2}. The largest observed X-ray flare took place on November 4, 2003. It was classified as X27. Note that this is close to the saturation limit of GOES observations and thus perhaps not quite accurate.

There are two basic components in the solar X-ray spectrum: **soft X-rays** between 1 and 10 keV, which are mainly thermal radiation of hot electrons, and **hard X-rays** between 10–100 keV which are non-thermal radiation from electrons accelerated to velocities of a sizable fraction of the speed of light. The distinction between soft and hard X-rays is not only a question of energy range. Also the spectra are different. The soft X-ray spectrum has an exponential shape, whereas the non-thermal hard X-ray spectrum shows a power law at increasing energy, which is less steep. In both cases the dominant radiation mechanism is **bremsstrahlung** of electrons moving among ions (mostly protons) of the ambient gas. As a 1-keV photon has a wavelength of 1.24 nm the X-ray spectra are in the wavelength range 0.01–1 nm.

The soft X-ray flux builds up gradually and peaks a few minutes after the impulsive emission. The Skylab observations in 1973 revealed first clear pictures of soft X-ray loops of a million degree coronal gas. These loops are associated with magnetic structures and their dynamics, i.e., flaring, indicates strong magnetic activity.

The temperatures of flaring soft X-ray loops are about then times hotter than the quiescent non-flaring coronal loops. Thus the plasma is heated up to several times 10^7 K, which is enough to strip almost all electrons from the iron atoms and the radiation comes from the inner shells of species such as Fe XXIII–XXVI (neutral iron has 26 electrons). This implies that there are also line emissions in the X-ray band, e.g., at 0.1778 nm (Fe XXVI),

0.185 nm (Fe XXV), 0.3177 nm (Ca XIX), 0.5039 nm (S XV), 0.917 nm (Mg XI), 1.346 nm (Ne IX). Coronal line emission is collisionally excited, so its intensity is controlled by the differential emission measure,

$$\frac{\mathrm{d}EM(T)}{\mathrm{d}T}dT = \int_0^\infty n_e^2(T,z)\,\mathrm{d}z,$$

where the integral is over the column depth along the line of sight. Thus, line emission is sensitive to the density of the emitting hot plasma. From the lines the densities of the hot loops have been estimated to be $10^{17} - 10^{18}$ electrons per cubic meter.

The hard X-rays often have a double source which is nearly co-located with the magnetic footpoints of the soft X-ray loops and Hα emission in the chromosphere. The two spots flash within 10 s from each other. This indicates that the hard X-rays are produced in the low corona and dense chromosphere by non-thermal electrons injected down along the legs of the coronal loop. This is further supported by the similar time profile of radio waves at centimeter wavelengths also produced by the non-thermal electrons.

Flares and γ-rays

The γ-rays have energies above 100 keV. Nuclear interactions of flare-accelerated protons and helium nuclei (energies 1–100 MeV/nucleon) with nuclei in the dense atmosphere below the acceleration site produce γ-rays at energies between 0.4 and 7.1 MeV. Furthermore, protons with energies above 300 MeV interact with hydrogen in the solar atmosphere and produce mesons. The decay of neutral mesons produces a broad γ-ray peak around 70 MeV whereas the decay of charged mesons leads to bremsstrahlung with a continuum of γ rays extending to several MeV. Also neutrons with energies above 1 GeV are produced.

There are two particularly strong γ-ray lines in the solar spectrum: 511 keV and 2.223 MeV. The former is due to electron-positron annihilation

$$\mathrm{e}^+ + \mathrm{e}^- \rightarrow \gamma + \gamma \ (511\,\mathrm{keV} \ \ \mathrm{each})\,, \tag{10.5}$$

where the positrons have been emitted by radioactive nuclei. The latter is a stronger line and results from the capture of a neutron by a proton:

$$\mathrm{n} + \mathrm{p} \rightarrow \mathrm{d} + \gamma \ (2.223\,\mathrm{MeV})\,. \tag{10.6}$$

Two important **spallation** reactions are

$$\begin{aligned} \mathrm{p} + {}^{16}\mathrm{O} \quad &\rightarrow \quad {}^{12}\mathrm{C}^* + \alpha + \mathrm{p} \\ {}^{12}\mathrm{C}^* \quad &\rightarrow \quad {}^{12}\mathrm{C} + \gamma \ (4.438\,\mathrm{MeV}) \end{aligned} \tag{10.7}$$

and

$$\begin{aligned} \mathrm{p} + {}^{20}\mathrm{Ne} &\rightarrow {}^{16}\mathrm{O}^* + \alpha + \mathrm{p} \\ {}^{16}\mathrm{O}^* &\rightarrow {}^{16}\mathrm{O} + \gamma\ (6.129\,\mathrm{MeV})\,. \end{aligned} \tag{10.8}$$

In the latter reaction also γ-lines at 6.917 and 7.117 MeV are prominent.

As the heated flaring loops are as hot as the center of the Sun (although much less dense), also some fusion reactions lead to γ emissions, e.g.,

$$\begin{aligned} {}^{4}\mathrm{He} + \alpha &\rightarrow {}^{7}\mathrm{Be}^* + \mathrm{n} \\ {}^{7}\mathrm{Be}^* &\rightarrow {}^{7}\mathrm{Be} + \gamma\ (0.431\,\mathrm{MeV}) \end{aligned} \tag{10.9}$$

and

$$\begin{aligned} {}^{4}\mathrm{He} + \alpha &\rightarrow {}^{7}\mathrm{Li}^* + \mathrm{n} \\ {}^{7}\mathrm{Li}^* &\rightarrow {}^{7}\mathrm{Li} + \gamma\ (0.478\,\mathrm{MeV})\,. \end{aligned} \tag{10.10}$$

Furthermore, there are several emission lines of nuclei exited by protons, e.g., ^{14}N (5.105 MeV), ^{20}Ne (1.634 MeV), ^{24}Mg (1.369 MeV), ^{28}Si (1.779 MeV), and ^{56}Fe (0.847 and 1.238 MeV).

10.3.2 Physics of flares

Some years ago it was said that there are more flare models, or theories, than there are flare theorists. With Yohkoh X-ray observations the number of plausible scenarios has, however, dropped. Any proposed flare mechanism must explain the release of magnetic energy arising from an active region. The energy release cannot be simple diffusion because the magnetic diffusion times ($\tau = l^2/\eta$) are very long, of the order of hundreds of years for scale sizes of 10 000 km. By compressing the gradient scale length to 1 km or below we can reach diffusion times of minutes. Thus the role of thin current sheets and reconnection becomes important.

While the concept or reconnection was first introduced to explain solar eruptions by Giovannelli already in 1946–1948, the questions "where", "when", and "how" are still unclear. People familiar with the problems of reconnection in the terrestrial magnetosphere sometimes think that the case for reconnection would be easier in the solar atmosphere than in the magnetosphere, but this is not quite true. The flux tube structure of the solar atmosphere is much more complicated, and there seldom are such simple structures as the magnetotail current sheet. Another important feature is the **line-tying** which means that the footpoints of the flux tubes are tied to the photosphere at both ends. On one hand this is a strongly stabilizing effect, but on the other, it allows large scale motion of the flux tubes.

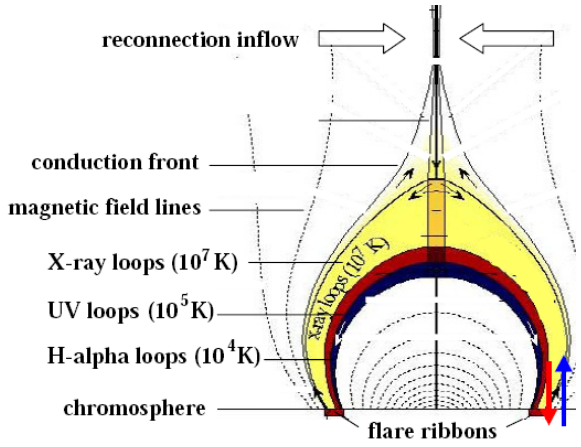

Figure 10.5: Origin of various flare emissions.

The observational possibilities of the solar reconnection are different from the magnetospheric observations. Recently the high-resolution images of X- and γ-ray emissions but the TRACE and RHESSI spacecraft have made the reconnection almost visible. These together with radio emissions tell us of the particles energized in the process and also something about the temperature development. However, there will never be the luxury of making in situ observations of the particles and fields as in the near-Earth space, including the ionospheric observations.

There are several facts that make the comparison of flares and magnetospheric substorms meaningful. In both cases the magnetic Reynolds number is large, so the magnetic field evolution is convection-dominated. While the plasma is much denser in the solar atmosphere, so is the magnetic field and, interestingly enough, the Alfvén speeds are of the same order of magnitude. As discussed in Chapter 6, $v_{\rm A}$ is a key parameter of the elementary reconnection models because it determines the outflow velocity of the plasma. Also the scale sizes are not too far from each other. The length of a prominence flux tube of 100 000 km is equal to 16 $R_{\rm E}$. Thus the substorm in the magnetotail is spatially somewhat larger than a typical flare. The time-scale of the process is proportional to the size divided by $v_{\rm A}$ and thus a typical flare is somewhat shorter than a typical substorm. Of course, the released energy in a flare is several orders of magnitude larger because there is so much more magnetic flux to be released.

Figure 10.5 sketches our present understanding how the various observational features discussed above are expected to arise. Energy is released by explosive reconnection above the top of the coronal loop. The heated loop radiates soft X-rays whereas the hard X-rays and γ-rays and most of the intense radio waves are due to the more energetic non-thermal particles.

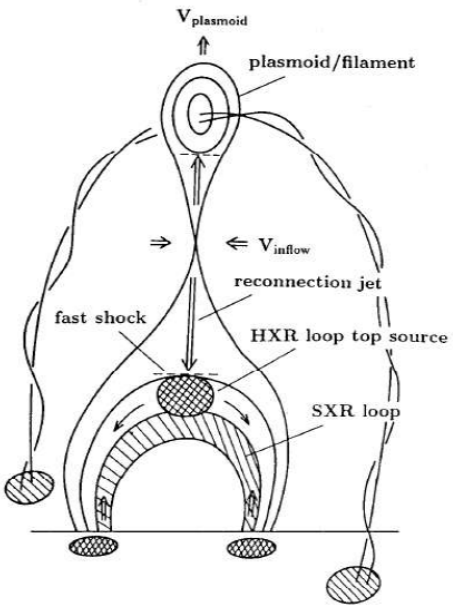

Figure 10.6: Shibata's plasmoid driven reconnection model for flares.

There are several more refined variants of the model. Figure 10.6 illustrates an example of a recent attempt toward a "unified model" of flares by *Shibata*. It also includes the magnetic structure that is released upward in the reconnection process.

If we rotate figure 10.6 by 90 degrees, it is somewhat analogous to a plasmoid or flux rope release from the terrestrial magnetotail in a substorm process. An important difference is that whereas the flux rope through the core of the plasmoid in the magnetotail is connected to the flanks of the magnetosphere, here it is tied into the photosphere but somewhere else than the underlying soft X-ray loop. In Shibata's model the twisted flux rope acts like a piston that stretches the field below. This enforces the plasma flow toward the current sheet leading to explosive reconnection. In this sense the formation and ejection of the plasmoid enhances the reconnection rate. The reconnection then both heats the plasma and accelerates electrons that emit hard X-rays when they are decelerated closer to the surface.

The driver plasmoid must be powered externally. One can imagine that this is a result of magnetic buoyancy in the flux rope in the core of the plasmoid. While this is a suggestive picture, one has to keep in mind that it is still only one of many proposed scenarios.

Finally, the flares are not completely described by the MHD flow theory alone. In the simple quasi-static current sheet models the reconnection appears to be very much driven by the external boundary conditions, i.e., by

the plasma flow toward the current sheet. But what finally determines when the reconnection can set in, is the microscopic physics of the reconnecting region. In case of more complicated geometries this issue may be even more important than in simple two-dimensional examples discussed in Chapter 6. Consequently the interplay between microscopic and macroscopic physics needs to be understood much better than is the case today.

The flares can be of very different sizes. Recall that in addition to the energy scale of $10^{21} - 10^{25}$ J, we discussed in the previous chapter briefly the processes called micro-flares (order of 10^{19} J) of nano-flares (order of 10^{16} J) as possible mechanisms to heat the corona. How similar or different they are to these large-scale flares is an open question.

10.4 Coronal mass ejections

The solar flares were, for a long time, thought to be the main drivers of geomagnetic storms. During the last 20 years or so this picture has been revised in favor of **coronal mass ejections** (CME). Here the terminology is a bit misleading. Most of the **matter** or **mass** in the mass ejections originates from the lower atmosphere and thus is not "coronal mass". The word "coronal" refers to the fact that CMEs are **observed** in the corona. There have been suggestions to rename the phenomenon to solar mass ejections or something else but once a suggestive name has been introduced it often is too late to change it.

CMEs are large plasma and magnetic clouds leaving the Sun. Their masses are in the range $5 \times 10^{12} - 5 \times 10^{13}$ kg and the angular sizes $40 - 50°$. Although they are huge, CMEs are difficult to observe. They remained undiscovered until the early 1970's when they were found with white-light coronagraphs onboard OSO 7 (the first on December 14, 1971) and Skylab. A coronagraph produces an artificial occultation which allows the faint light from the CME be observed around the occulting disc. Coronagraphs can be used also in terrestrial observatiories but there is always so much straylight in the atmosphere that it is very difficult to see a CME from ground.

A CME does not radiate itself. The observed light is produced by Thomson scattering of the solar photons from the electrons in the cloud. CME coronagraphs use white light because most of the solar photons are in the visible wavelengths. The white-light brightness varies in proportion to electron density but not to temperature. Thus the brightness can be used in determination of the density structure of the emitted cloud.

Observations have established that CMEs are a very common phenomenon. The whole Sun occurrence-rate is 0.8 events/day at solar minimum and 3.5

events/day at solar maximum. The former number is from the LASCO observations onboard SOHO during the solar minimum in the beginning of the sunspot cycle 23 and is a factor of 4 higher than previous estimates. This is likely due to the much better sensitivity of LASCO to weak CMEs than of earlier observations.

Already the early observations established that CMEs are more often associated with eruptive prominences than impulsive flares. The exact relationship between the flares and CMEs is unclear but only some 40% of CMEs have an associated flare close to the site of the ejection. However, these flares may take place before, simultaneously, or after the lift-off of the CME. On the other hand, for some 70% of ejections a disappearance of a dark filament or eruptive prominence has been identified. The prominence material can often be identified in the coronagraph images but only very rarely *in situ* by spacecraft in the solar wind close to the Earth. Thus the interaction between CMEs and the ambient solar wind belongs to the central problems in solar and solar wind physics.

When a CME leaves the Sun its speed at 5 $r_\odot$ varies from less than 50 km/s to more than 2000 km/s. At 1 AU the speed only seldom is larger than 750 km/s and never smaller than the minimum solar wind speed of about 280 km/s. Thus the originally slow CMEs are accelerated toward the solar wind speed whereas the very high-speed CMEs are decelerated. The energy carried by a CME is of the order of 10^{24} J which is quite comparable to large flares. However, the energy is mostly in the kinetic energy of the plasma cloud and less in high-energy particles.

CMEs originate from the closed field line regions and this determines their magnetic topology. However, as the magnetic structure has to be torn off from the Sun, the field must open locally. Yohkoh soft X-ray images have shown examples of the escaping cloud after which a soft X-ray arcade remains for several hours like a wound on the Sun at the place from which the CME was ejected. This phenomenon is called a **gradual flare** and it is associated to the restructuring of the magnetic field field after the major ejection.

As determined from the observations of counter streaming electrons along the magnetic field lines when, the CME may reach 1 AU with both ends of the magnetic fields still tied to the Sun. A more typical topology is such that only one end is tied to the Sun, and structures completely detached from the Sun are also observed. Thus the role of reconnection to cut the field lines behind a CME is an important question to study further.

At least 1/3 of all CMEs are nearly force-free flux ropes. Assuming such a field configuration the effect of the CME, when it hits the Earth's magnetopause, depends critically whether the magnetic field in front of the plasma cloud points toward the north or the south. The strongest effect is

expected to take place when the field points toward the south, as then the dayside reconnection opens the magnetopause for the excessive amount of energy and plasma to penetrate into the magnetosphere.

As noted above the CME rate increases with the increasing activity during the solar cycle. The latitudinal evolution is, however, opposite to the evolution of sunspots. During the minimum the CMEs come from very close to the equator whereas the source region widens toward the maximum. Furthermore, the structure of the CMEs becomes more complicated reflecting the more complicated magnetic structure of the active Sun. The CMEs may actually play an important role in the total magnetic flux budget of the Sun. It has been suggested that the reason for their existence is to carry away the excessive flux produced during the solar cycle. When the differential rotation creates toroidal field, the field accumulates at low latitudes in the regions of closed magnetic loops. The persistent solar wind removes flux only from the regions of open field lines and the ejection of large magnetic clouds may take care of getting rid of the closed flux.

Another indication that CMEs are related to the global solar dynamics is that sometimes the Sun appears to expel two CMEs almost simultaneously to opposite directions. There is no accepted explanation to this behavior.

The CMEs are now under intensive study world-wide. The following are some central questions about CMEs. Many of these will find an answer during the coming few years but others may remain with us for a long time.

- What is the physics of the initiation of CMEs and what are the observable signatures of this physics?
- What is the exact physical relationship between CMEs and flares?
- What determines where and when a CME takes place and how could we predict CMEs?
- How are CMEs accelerated?
- What are the details of CME-associated particle acceleration (see the next section)?
- What is the role of CMEs in the long-term (solar cycle) evolution of the corona?
- What is the role of reconnection? Where and when does it take place? Is reconnection a primary driver of the release or its consequence? What are the effects of reconnection on the topology of the CMEs?
- What roles do the magnetic helicity and shear play here? What instabilities do occur, and where and when?

10.5 CMEs, flares and particle acceleration

Both flares and CMEs accelerate charged particles to high energies. In the flares typical proton energies are of the order of 10 MeV but can reach to 1 GeV. Electrons are less energetic, typically 100 keV but may in rare events reach to 100 MeV. Particles are accelerated to all directions, some of them give rise to the X- and γ-rays when they meet other solar particles, some produce radiowaves in the strong magnetic field structures, and some escape from the Sun and become observable, e.g., close to the Earth. In this section we discuss the particle acceleration from the outside point of view.

Note that the fluxes of energetic particles are much less than the flux of the ambient solar wind. A typical flare causes at 1 AU a flux of 10^7 particles $\mathrm{m}^{-2}\mathrm{s}^{-1}$, whereas typical solar wind flux is 5×10^{12} particles $\mathrm{m}^{-2}\mathrm{s}^{-1}$. The energetic solar particles are often called **solar cosmic rays** (SCR) to distinguish them from the **galactic cosmic rays** (GCR) originating outside the solar system. The GCR fluxes at 1 AU are even smaller, about 6×10^2 particles $\mathrm{m}^{-2}\mathrm{s}^{-1}$.

In Chapter 9 we discussed the spiral structure of the interplanetary magnetic field. The bulk of solar wind plasma expands almost radially out from the Sun and causes the spiraling magnetic field as the magnetic field lines are tied to the surface of the rotating Sun. The individual energetic particles are forced to follow the spiraling magnetic field lines. Thus solar energetic particles reaching the Earth mostly originate from the western half of the solar surface. Near the Earth, only particles with highest energies are capable of penetrating through the Earth's magnetic field (e.g., particles of higher energy than roughly $\sim$750 MeV above Helsinki). Finally the cosmic ray particles are stopped by the atmosphere where they cause nuclear reactions and produce neutrons and muons that can be detected on ground.

The solar energetic particle events are divided in two main categories: **impulsive** and **gradual**. The impulsive events are much more common than the gradual events. Table 10.1 lists the main observational characteristics of these two classes.

Particles in impulsive events are thought to be accelerated close to the Sun both by the rapid energy release in the impulsive phase of a flare and by strong wave activity associated to the process. The very high abundance of ^{3}He in impulsive events is a curious fact, as the fraction of ^{3}He nuclei of all helium in the solar atmosphere is about 5×10^{-4}. This indicates that, whatever the acceleration mechanisms are, at least one of them must be very efficient in accelerating this particular species. The same acceleration mechanism may accelerate also ^{4}He as the abundance of this ion species is enchanced (over the gradual event values) as well. This points to gyroresonant

Table 10.1: Properties of impulsive and gradual solar energetic particle events

	impulsive	gradual
Particles:	electron-rich	proton-rich
^{3}He/^{4}He	≈ 1	≈ 0.0005
Fe/O	≈ 1	≈ 0.1
H/He	≈ 10	≈ 100
Duration of X-ray flare	impulsive (minutes) (hard & soft X-rays)	gradual (hours) (soft X-rays only)
Duration of particle event	hours	days
Radio bursts	Types III and IV	Types II and IV
Coronagraph observations	typically nothing	CME in 96% of cases
Solar wind observations	energetic particles	very energ. particles
Longitudinal extent	$< 30^\circ$	$\approx 180^\circ$
Events/year (solar max)	≈ 1000	≈ 100

wave-particle interaction with waves having frequencies below the proton gyrofrequency ($f_{\rm cp}$). The gyrofrequencies of ^{3}He^{++} and ^{4}He^{++} are 2/3 $f_{\rm cp}$ and 1/2 $f_{\rm cp}$, respectively. In the high-frequency branch of the Alfvén waves the electromagnetic ion cyclotron waves propagate at frequencies somewhat below $f_{\rm cp}$ and are a possible candidate for the acceleration. Note that modern instruments have found ^{3}He/^{4}He-ratios of the order of $0.01 - 0.05$ also in several gradual events, which is often interpreted so that gradual events are accelerated from a pool of suprathermal particles containing remnants from previous impulsive flares.

The gradual events have a very strong association with CMEs. Their particles are most likely accelerated by the shock waves that the CMEs drive in the coronal and solar wind plasma. This also explains the much longer duration of the events. One might also expect that the long duration could explain the acceleration to higher energies. However, it is unlikely that the shock acceleration could accelerate the ambient solar wind ions to such high energies, because the minimum rate of the shock acceleration process scales as the cyclotron frequency and this decreases as r^{-2} in the solar wind. Thus acceleration in the corona is likely to be required for particles reaching energies of several tens or hundreds of MeV.

Note that we should avoid confusion between **gradual flares** and **gradual particle events**. As discussed in the previous section, the gradual flares are post-CME effects in the corona and they are probably not responsible for particle acceleration in the gradual particle events.

Especially the details of solar flare particle acceleration belong to the many unsolved questions in solar physics. From observations we know that very strong acceleration takes place and it is quite clear that the energy

must have magnetic origin. In fact, only the very strongly stressed and sheared magnetic field structures have enough energy to explain the rapid acceleration. But from understanding these conditions to actual acceleration mechanisms in the great variety of explosive phenomena is a long way. One of the promising and well studied models of flare acceleration is **stochastic acceleration** by strongly fluctuating fields, including shocks, ion cyclotron waves, turbulence, etc. Several of the different mechanisms may be involved before a particle is lifted from the quasi-thermal background to the observed energy level.