Chapter 1

Introduction

The Sun is familiar to all of us, but what are perhaps less familiar are the dark spots that sometimes appear on its surface (Figure 1.1). These 'sunspots' were first observed telescopically by Galileo and others in the seventeenth century, but even today, many of their properties remain unexplained.

Sunspots are created when strong magnetic fields, generated deep within the Sun, rise up to the solar surface. The magnetic field partially inhibits the convection that is normally found at the surface, and this weakened convection is less able to transport heat into the sunspot, making it cooler, and hence darker, than its surroundings.

What is less well understood is the detailed nature of this 'magnetoconvection' that is taking place within a sunspot. As we shall see in the following sections, sunspots contain detailed fine structure which is ultimately the result of magnetoconvective processes. The aim of this thesis is to try to better understand these processes by studying magnetoconvection in various situations, beginning with simple models and working up to more complex numerical simulations.

We will start by giving, in this chapter, a broad introduction to the subject of sunspots, covering both observational and theoretical topics. We will then give a brief summary of the remainder of the thesis, indicating how it relates to sunspots and to some of the outstanding theoretical questions about them.

1.1 Internal structure of the Sun

Before describing sunspots, it may be helpful to give a quick outline of the structure of the Sun itself. The Sun is a (more or less) spherical ball of gas, 696 000 km in radius,

Figure 1.1: *Full disc image of the Sun taken by the Swedish 1-m Solar Telescope on 15 July 2002. Courtesy Royal Swedish Academy of Sciences.*

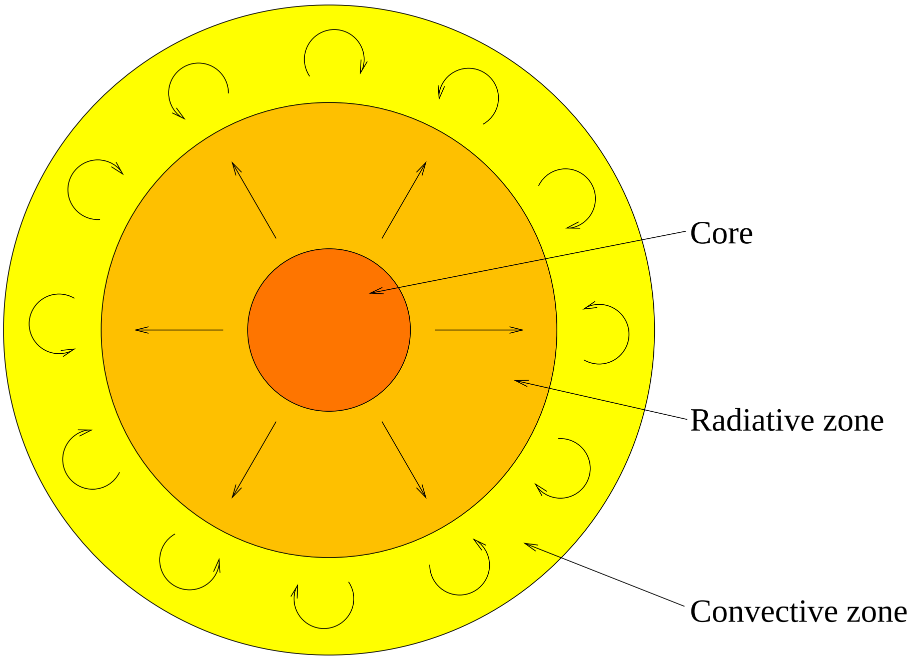

Figure 1.2: *Sketch of the internal structure of the Sun.*

composed mostly of hydrogen and helium. A cross-section is shown in Figure 1.2. At the centre is the core, containing approximately the inner 25% (by radius), where the density reaches values up to 150 times that of water, and the temperature reaches some 15 million degrees Kelvin. Under these extreme conditions, the nuclear fusion reactions that generate the Sun's energy can take place. Above the core, there is the radiative zone (which technically also includes the core), where heat energy is transported towards the surface by radiation, and the convective zone (including the outer 30% by radius), where convection becomes the principal means of energy transport. This convection is indeed observed at the solar surface, where it is known as 'granulation'.

The interior of the Sun cannot be observed directly, of course (although helioseismology can provide indirect measurements of the interior), and so the internal structure is calculated using models. Indeed, models of stellar structure have been quite successful in explaining the observed properties of the Sun and other stars. The basic idea in these models is to assume that the star is in a static equilibrium state. There are then five equations to be solved: one for mass continuity, one for hydrostatic force balance (the pressure gradient must balance gravity), one for energy conservation (the net outward heat flux must balance the energy generation from nuclear reactions), one for heat transport (this gives the temperature gradient required to transport a given heat flux), and finally an equation of state (relating the temperature, pressure and density of the gas). These can be solved (together with appropriate boundary conditions) to determine the internal structure of stars of various masses and compositions.

For the heat transport equation, one must consider the three different methods of heat transfer: conduction, radiation and convection. For stars like the Sun, conduction turns out to be negligible, so only the last two need to be considered. The equations describing radiative heat transfer are well understood; convection, on the other hand, is a highly nonlinear process and there is no simple formula relating the convective heat flux to the temperature gradient in a compressible fluid. Therefore, convection is usually approximated by a 'mixing length' formalism, where a fluid element is assumed to rise adiabatically by a given length (the mixing length) before giving up its excess heat to the surroundings. The mixing length itself is left as a free parameter, usually expressed as a proportion of the local pressure scale height.

The solution of such models yields (for the Sun) the structure shown in Figure 1.2. One can also use a similar approach to model subsurface structure in sunspots; the mixing length theory for convection can be adapted (in a simple way) for magnetoconvection, by taking a reduced mixing length parameter in order to model the reduced convective efficiency. This will be discussed further in section 1.3.

1.2 Sunspot observations: past and present

The earliest recorded observations of sunspots go back to Chinese astrologers, who were apparently observing them at least as early as the 11th century BC. The largest spots would have been visible to them with the naked eye at sunrise or sunset, or reflected on the surface of still waters. There are also various records of sunspot observations in the Western world, apparently going back as far as the ancient Greeks. However, Western religious and philosophical thinking dictated that the Sun was a celestial body, and therefore perfect in every way, and spotless. Thus, the sunspot observations were usually ignored or forgotten, and it was not until the invention of the telescope, in the seventeenth century, that the existence of sunspots became widely acknowledged in the West.

These first telescopic observations were made by Galileo, Christoph Scheiner, and David and Johannes Fabricius, around 1610. (There is still apparently some controversy as to who of these made the first observation.) One of Galileo's sunspot drawings is shown in Figure 1.3. Even from these early observations, it was clear that sunspots are composed of two distinct regions: a dark central area and a lighter outer part. Today, the inner region is known as the *umbra* (from the Latin for 'shade' or 'shadow'), and the outer area is called the *penumbra* (from Latin *paene*, 'nearly' or 'almost', + *umbra*).

Further progress was slow, in part because there followed, between the years of 1645 and 1715, a period of extremely low sunspot activity, which later became known as the Maunder Minimum. The main discovery in the eighteenth century was that of the Wilson depression, which refers to the fact that the visible surface of a sunspot is located at a deeper vertical level in the Sun than the normal photosphere. It was detected by Alexander Wilson in 1769 by careful observations of sunspots near the edge (limb) of the solar disc.

When larger telescopes became available, in the nineteenth century, it became possible to observe fine details of umbral and penumbral structure. Telescope technology has been improving continually since then and images of stunning detail are now being obtained by instruments such as the Swedish 1-metre Solar Telescope, which came into service in May 2002. Images obtained from this telescope are displayed in Figures 1.4

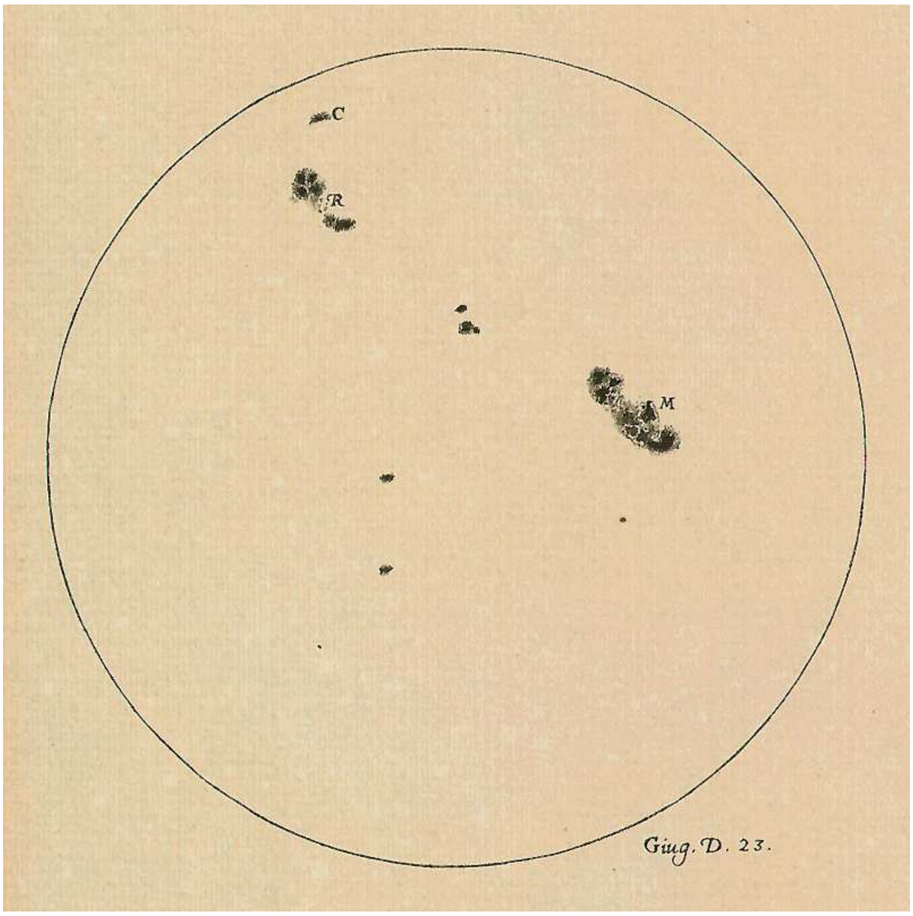

Figure 1.3: *One of Galileo's sunspot drawings, from 23 June 1613.*

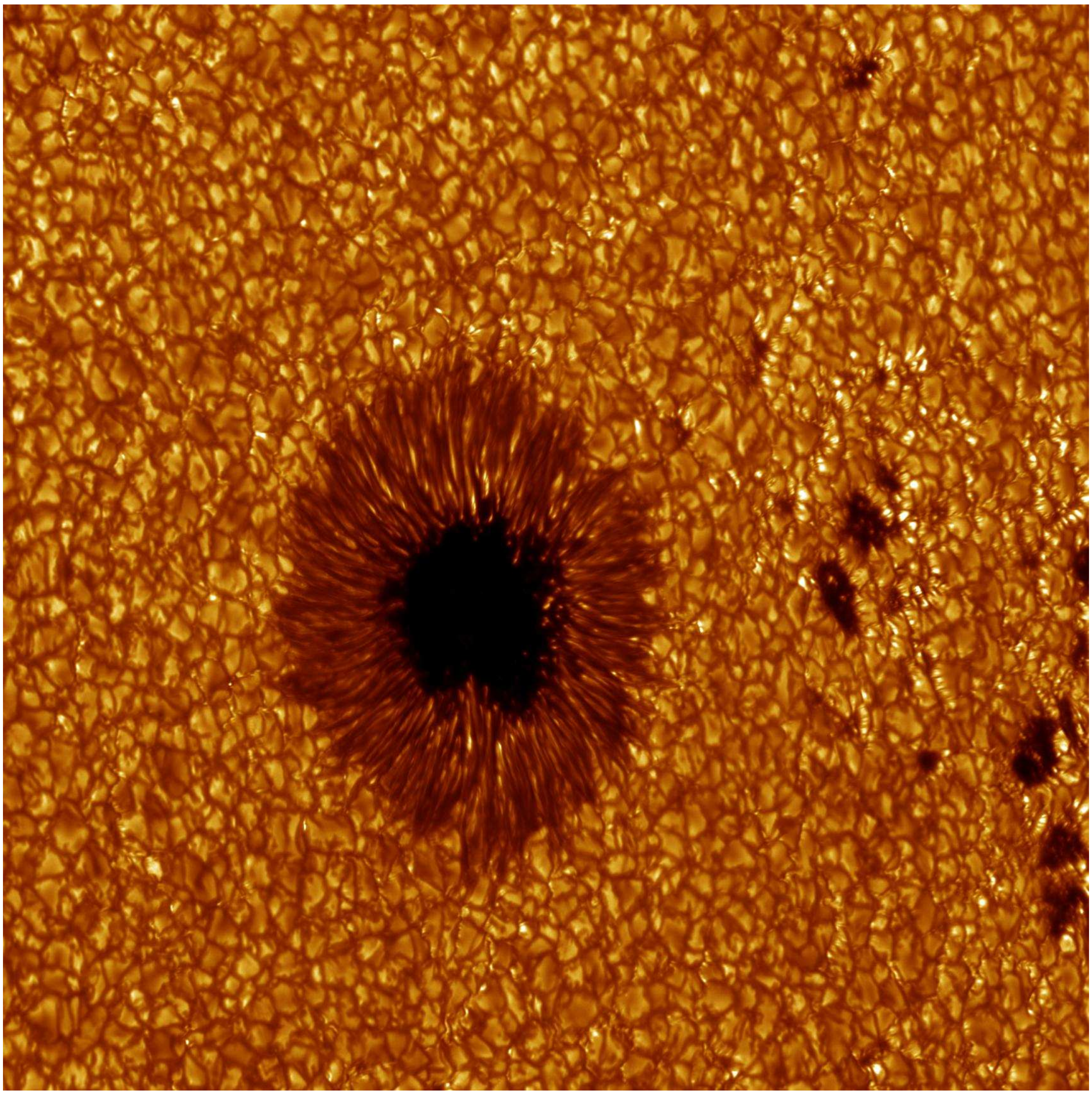

Figure 1.4: *An image taken by the Swedish 1-m Solar Telescope showing a small regular sunspot. Note also the pores visible at the right-hand side. Courtesy Royal Swedish Academy of Sciences.*

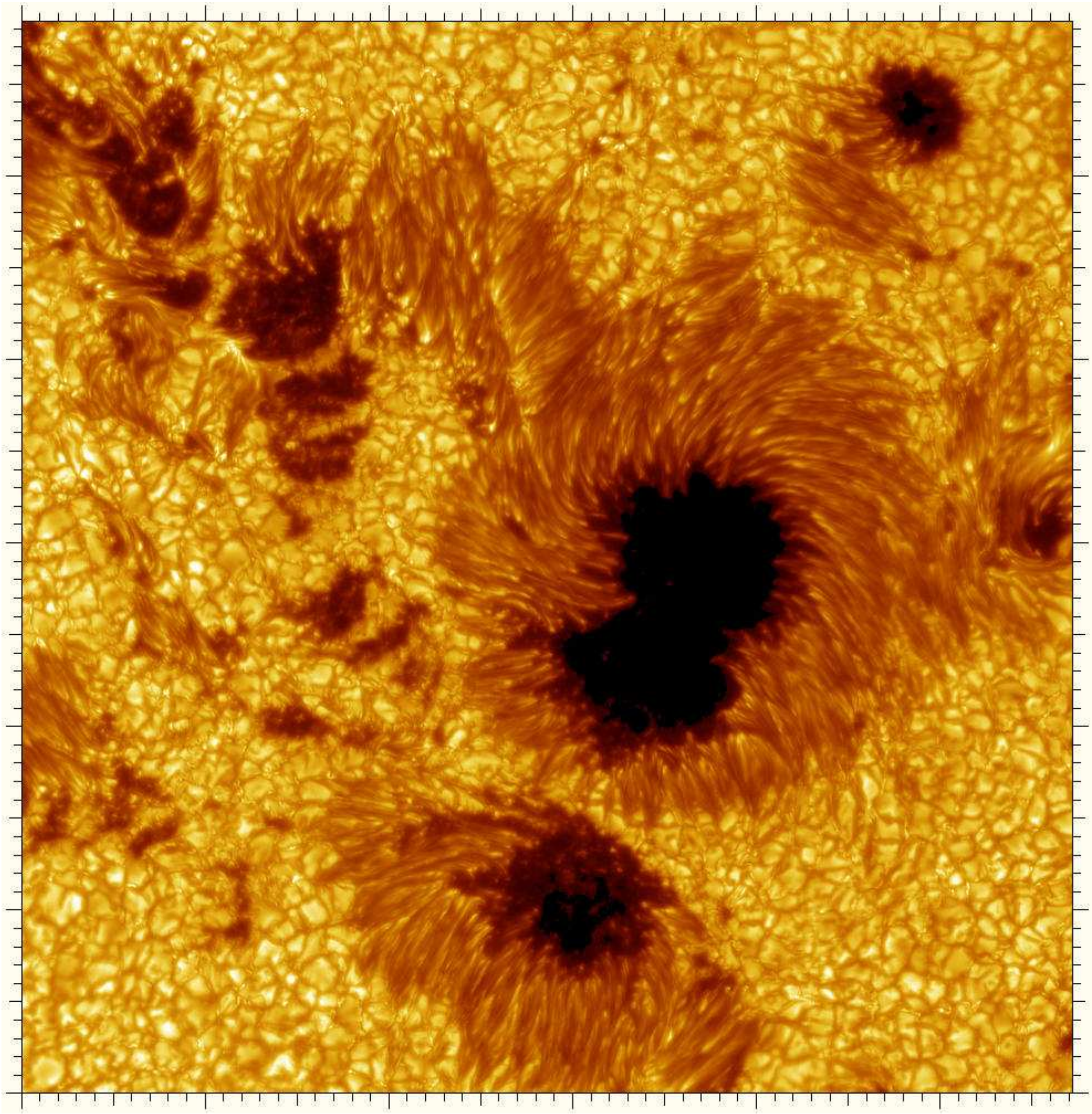

Figure 1.5: *A close-up view of the sunspot group visible in Figure 1.1. These sunspots have a more irregular structure.*

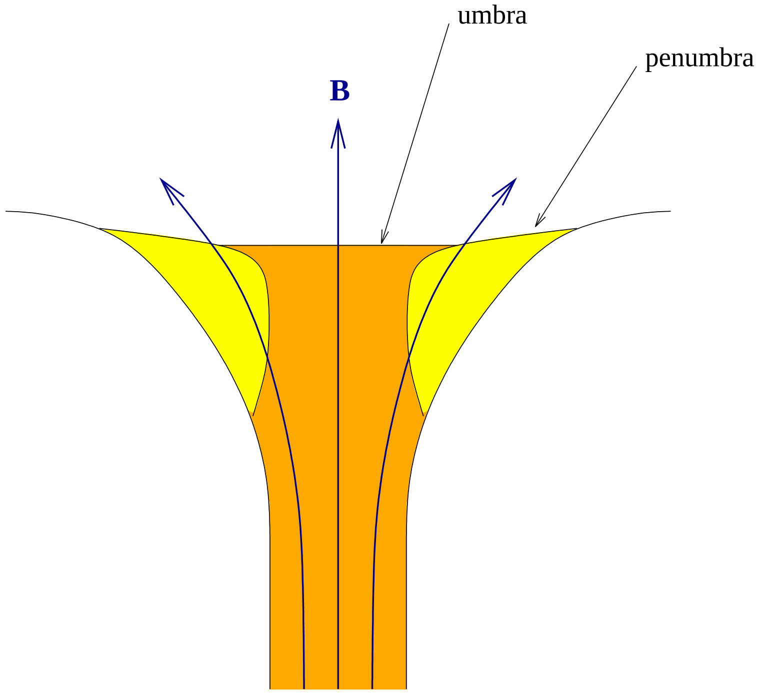

Figure 1.6: *A sketch of a cross-section through a typical sunspot.*

and 1.5. Sunspots, with many small-scale structures and fine details, can be seen; we will return to these fine structures in more detail in section 1.4. Note also the presence of smaller dark patches, known as pores; these are similar phenomena to sunspots, but they are smaller and they do not have penumbrae.

Our modern understanding of sunspots begins with George Ellery Hale's discovery, in 1908, that sunspots are associated with strong magnetic fields. He found this by measuring the so-called Zeeman effect, the splitting of spectral lines by a magnetic field, using an instrument that he himself invented, the spectroheliograph. In due course it was realized that this magnetic field would inhibit convection, explaining why sunspots are cooler and darker than the normal solar surface.

The magnetic field strength at the centre of a large sunspot might reach 0.3 tesla (T).[1] For comparison, the Earth's magnetic field at the surface has a strength of about 5×10^{-5} T; the field strength of a typical bar magnet might be 0.01 T; a medical MRI unit produces fields of about 1.5 T.

1.3 Overall structure of sunspots

Sunspots of many different sizes are observed. The smallest have radii of 1800 km or so, while very large spots can sometimes attain radii of 30000 km or more (Solanki, 2003). However, all well-developed regular sunspots (such as the one shown in Figure 1.4) have a similar structure, and so, at least in the case of the more regular, circular spots, we can talk about the structure of a typical sunspot.

Figure 1.6 shows a rough sketch of a vertical slice through a sunspot. The basic physical picture is that there is a vertical column of magnetic flux (indicated by the orange and yellow regions) surrounded by the ordinary field-free solar plasma. The orange region marks areas where the field remains near-vertical, which gives rise to the umbra at the surface. In the yellow regions, the field has become significantly inclined to the vertical, and it is these inclined fields that (somehow) allow the penumbra to form.

Over the years, a number of models have been put forward to try to account for the large-scale structure of sunspots.[2] One of the first of these models was due to Schlüter and Temesvàry (1958), and a great many have been developed since; an extensive review is given in sections 4.4 to 4.8 of Solanki (2003).

To make the problem mathematically tractable, these models typically assume that the spot is circular and axisymmetric; this is not true for real sunspots, at least on small scales (e.g. the penumbra is considerably non-axisymmetric on fine scales), but it is a reasonable assumption for modelling the large-scale overall structure. The other main assumption made is that the sunspot is in equilibrium with its surroundings (both mechanically and thermally). Again, this can be justified in an average sense, since while real sunspots do have dynamic features, occurring on timescales of perhaps an hour or so, the total lifetime of a spot is much longer than this (large spots can persist for several months).

To describe the equilibrium state we basically need to ensure both a hydrostatic and a thermal equilibrium. The equation for the hydrostatic force balance is

$$-\nabla p + \rho \boldsymbol{g} + \boldsymbol{j} \wedge \boldsymbol{B} = 0. \tag{1.1}$$

The horizontal force balance in a sunspot is therefore between the gas pressure gradient, which exerts an inward force (because the cool interior is at a lower pressure than the

[1]Magnetic field strengths are also commonly quoted in gauss (G); 1 G = 10^{-4} T.

[2]These models can also be used for pores, if the penumbra is omitted.

hot exterior), and the Lorentz force ($\boldsymbol{j} \wedge \boldsymbol{B}$), which is in an outward direction (this can be thought of as a gradient of magnetic pressure $B^2/2\mu_0$, with a higher magnetic pressure inside the spot than outside). In the vertical direction, one must balance the gas pressure gradient against gravity (as well as, possibly, Lorentz forces).

Equation 1.1, together with the equation of state, give three equations in the four unknowns p, ρ, T and $\boldsymbol{B}$. (Equation 1.1 counts as two equations because it has both vertical and horizontal components; also $\boldsymbol{B}$ only really counts as one scalar unknown because we have the additional constraint $\nabla \cdot \boldsymbol{B} = 0$.) To close the system, an additional condition must be added. The 'proper' way to do this is to include an energy equation, describing the thermal equilibrium of the sunspot. However, in some models, a simpler approach is taken, in which one simply prescribes the pressure (or some other variable) as a function of depth. This produces a model which is not strictly in thermal equilibrium, but this is not as bad as it sounds, since one can vary the assumed profile until a reasonable match with observations is found.

Even from fairly simple models, it was realized early on (Schlüter and Temesvàry, 1958) that the energy flux through the umbra is too great to be provided by radiation alone, and therefore convection must be occurring below the visible surface of the umbra. In other words, the magnetic field of the umbra must inhibit convection only partially, not completely.

1.3.1 The model of Jahn and Schmidt (1994)

We will now describe a recent model, that of Jahn and Schmidt (1994), as we can use it to illustrate one or two points about sunspot structure. The model is shown in Figure 1.7.

The first thing we need to do is make some assumptions about the structure of the magnetic field. Earlier models approached this by using a self-similarity method, but this does not give a particularly good fit to observations. An alternative approach, and the one used by Jahn and Schmidt (1994), is to assume that $\boldsymbol{j} = 0$ everywhere except on isolated current sheets. It is found that one current sheet enclosing the entire sunspot is insufficient to model the penumbra; Jahn and Schmidt (1994) chose to take two current sheets, an inner one separating the umbra from the penumbra, and an outer one enclosing the entire structure. This approach produces a sharp distinction between umbra and penumbra (as is observed) and allows the different energy transport

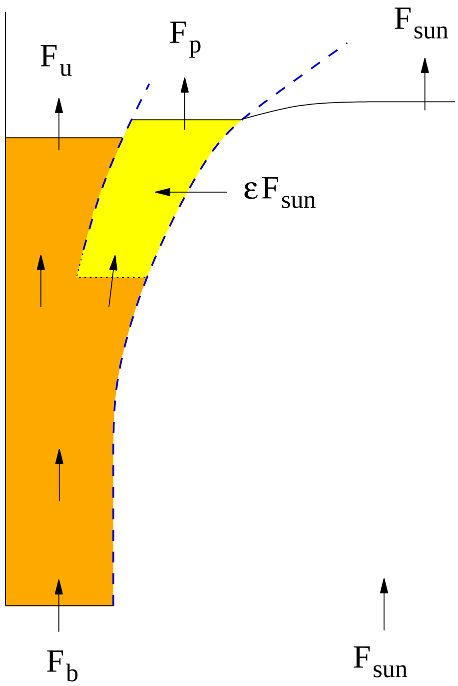

Figure 1.7: *The axisymmetric sunspot model proposed by Jahn and Schmidt (1994). The small black arrows indicate heat fluxes (per unit area). The blue dashed lines indicate the position of current sheets within the model. The yellow region indicates the extent of the penumbra, defined in this model as the region where interchange convection, with associated lateral heat flux of* $\epsilon F_{\rm sun}$*, takes place.*

and thermodynamic properties between the two to be described. It also avoids the complexity of dealing with volume currents in the penumbra which would otherwise be needed (e.g., Jahn 1989). (One criticism, however, is that it produces a sharp jump in the magnetic field strength at the umbra-penumbra boundary, which is not observed.)

An energy equation is included (with convection being described by a mixing-length theory, using a reduced mixing length in the umbra and penumbra in order to simulate the reduced convective efficiency there). When solving this equation, boundary conditions for the energy flux are needed. At the top we can simply match to observed umbral and penumbral energy fluxes; there are also (more indirect) observations that can help with the lateral boundary conditions. In particular, it is observed that the brightness and surface temperature of the umbra both vary during the 11-year solar cycle.[3] This can only be explained if the umbra is thermally well insulated from its surroundings. The penumbra shows a similar intensity variation, although with a much lower amplitude, implying that significant heat flux is being transported from the quiet Sun into the penumbra. This is usually interpreted as a convective process (sometimes called 'interchange convection'), with material (and thermal energy) being exchanged between the penumbra and quiet Sun.

Therefore, Jahn and Schmidt (1994) assume that the flux tube comprising the sunspot is completely thermally insulated from its surroundings, except for the interchange convection that is taking place in the penumbra. This penumbral convection is modelled as a heat flux of ϵ times the normal solar heat flux, which is carried into the penumbra from the external plasma. (It is found that values of ϵ between about 0.6 and 0.7 seem to be required in order to give plausible results.) This interchange convection (and associated heat flux) extends down from the surface to a depth z_{bp}; this depth is computed as part of the model (so that the stratifications of umbra and penumbra match), and it can be thought of as the bottom of the penumbra. Its value is typically found to be about 4000 km in this model.

We see that any model for convection in sunspots would ultimately have to explain how this behaviour for the heat flux arises. In particular, it should explain the differences between convection in the umbra, in which the heat flux is channelled along the field lines, with almost no mixing across the boundary of the flux tube, and in the penumbra, where convective interchanges apparently do occur between the penumbra and the external plasma, with a corresponding heat flux into the spot. (Notice also how

[3] The solar cycle is covered in section 1.6.4.

these convective processes, which occur at small scales, have an impact on the overall large-scale structure of the spot.)

1.3.2 Subsurface structure: 'cluster' and 'monolith' pictures

So far we have considered a sunspot to be a single homogeneous flux tube extending downwards beneath the surface. In fact, this is a simplification, and the flux tube will have more detailed structure than this. There are two competing theories as to what the subsurface structure looks like: either a monolithic column of flux, or a cluster of separate 'mini' flux tubes. (A discussion of the differences between the two models can be found in section 4.3 of Solanki 2003.)

In the monolithic model, the sunspot consists of a single flux tube that remains a coherent structure even down to great depths. The magnetic field within the tube is strong enough to reduce the efficiency of convection, but not so strong as to suppress it completely. In other words, magnetoconvection (as opposed to ordinary field-free convection) occurs inside the flux tube, and this different type of convection is why the heat flux is lower inside the flux tube than outside.

The alternative, cluster, model was first proposed by Parker (1979), who suggested that the flux tube would be subject to the so-called fluting instability, which would cause it to break up into a tight cluster of separate 'mini flux tubes' just below the solar surface. The spaces in between these individual tubes would contain field-free plasma. The field inside the individual tubes would be strong enough to completely suppress convection, but within the field-free regions, convection would be able to take place as normal. Thus, in this model the reduced convective efficiency is caused not by a different form of convection, but simply by a reduced spatial filling factor.

The actual conditions for the fluting instability were set out by an earlier calculation, due to Meyer et al. (1977); it was found that near the surface, the instability would be suppressed by magnetic buoyancy processes, but deeper down, the fluting instability could indeed take place. However, this calculation assumed a static equilibrium configuration, while in reality, convection takes place both inside and outside the sunspot, and this must be taken into account. The central question is therefore whether convection would be able to stabilize the flux tube against fluting. This question remains to be answered, but it has been suggested that the necessary effect could be provided by a so-called 'collar flow', a supergranular-scale convection flow outside the sunspot. (See also section 1.5.4, below).

1.4 Fine structure in sunspots

We now move on from overall properties of sunspots, and turn to their fine structure. Sunspots are now known to have considerable small-scale structure, in both the umbra and the penumbra (cf. Figures 1.4 and 1.5). In addition, the magnetic field has been found to have an extraordinary non-axisymmetric structure in the penumbra.

Further information about sunspots and their fine structure can be found in the recent review articles by Thomas and Weiss (2004) and Solanki (2003).

1.4.1 Umbral fine structure

The main feature of note in the umbra is the presence of so-called umbral dots, the first recorded observation of which was by Thiessen (1950). Umbral dots are small bright features visible against the dark background, with temperatures typically about 1000 K hotter than the coolest part of the umbra. The latest observations (Sobotka and Hanslmeier, 2005) appear to have spatially resolved most of the dots; their typical diameter is found to be about 100 km. Umbral dots are also often observed to travel radially towards the centre of the spot; this is particularly true of dots near the edge of the umbra. The dots are generally interpreted to be convective features, since they are warm compared to the umbral background and generally show upward motions.

Convection is not observed directly in the umbra, since there is a 'radiative blanket', a stably stratified region near the surface in which heat is transported purely by radiation; however, we know that convection must be taking place below this layer. The umbral dots presumably represent particularly vigorous rising convective plumes, which are able to penetrate through the radiative blanket and reach the surface. (The theory of umbral magnetoconvection is discussed further in section 1.5.1.)

The pattern of umbral dots also seems to be better explained by the monolithic, as opposed to the cluster, model (section 1.3.2). This is because a cluster of isolated flux tubes, with convection in between, would be expected to produce a network of bright lines rather than isolated bright points; on the other hand, magnetoconvection (in a monolithic flux tube) seems to be able to explain the umbral dots as a pattern of spatially modulated oscillations (see section 1.5.1).

Another feature of umbrae worth mentioning is that they sometimes contain lanes of bright material, crossing from one side to the other, which are known as light bridges. These mark out 'fissures' in the umbra and generally contain weaker, more horizontal magnetic fields than the rest of the umbra. In addition, there are also sometimes 'dark nuclei', dark regions containing few if any umbral dots; these probably correspond to regions of stronger magnetic field (cf. section 1.5.1).

1.4.2 Penumbral fine structure

The penumbra shows a very rich structure. The most prominent feature is the presence of alternating light and dark filaments which extend (approximately) radially outwards from the centre. Modern high resolution images show that the bright filaments are themselves composed of separate 'grains', which are typically about 350 km wide (or less), with lengths ranging from about 350 to 2500 km. The grains often show internal structure with a few dark bands crossing them (Rouppe van der Voort et al., 2004). The grains are also observed to move radially, with the motion being in an inward direction within the inner 60% or so of the penumbra (by radius), and an outward direction in the outer 40%, with typical speeds of about 500 m s^{-1}. The inward-moving grains sometimes penetrate into the umbra, where they become umbral dots.

The magnetic field in the penumbra has a curious 'interlocking-comb' structure, in which the inclination of the field to the vertical varies significantly in the azimuthal direction (see Figure 1.8). This variation appears also to be correlated with the intensity variations between the bright and the dark penumbral filaments. The picture that we have is as follows: in the bright filaments, the inclination to the vertical varies from about 30° in the inner penumbra to about 60° at the outer edge; in the dark filaments, the field is inclined at about 65° to the vertical in the inner penumbra, becoming nearly horizontal at the outer edge of the spot. Indeed some of these field lines even reverse direction and plunge back down beneath the surface.

The two families of field lines also differ in their larger-scale connectivity. It appears that the more nearly horizontal fields (from the dark filaments) remain close to the surface, whereas the more vertical component of the field (from the bright filaments) extends high into the atmosphere, and can even extend many thousands of kilometres across the Sun (Sams et al., 1992; Winebarger et al., 2001).

As far as the vertical structure goes, we know that the more horizontally-oriented

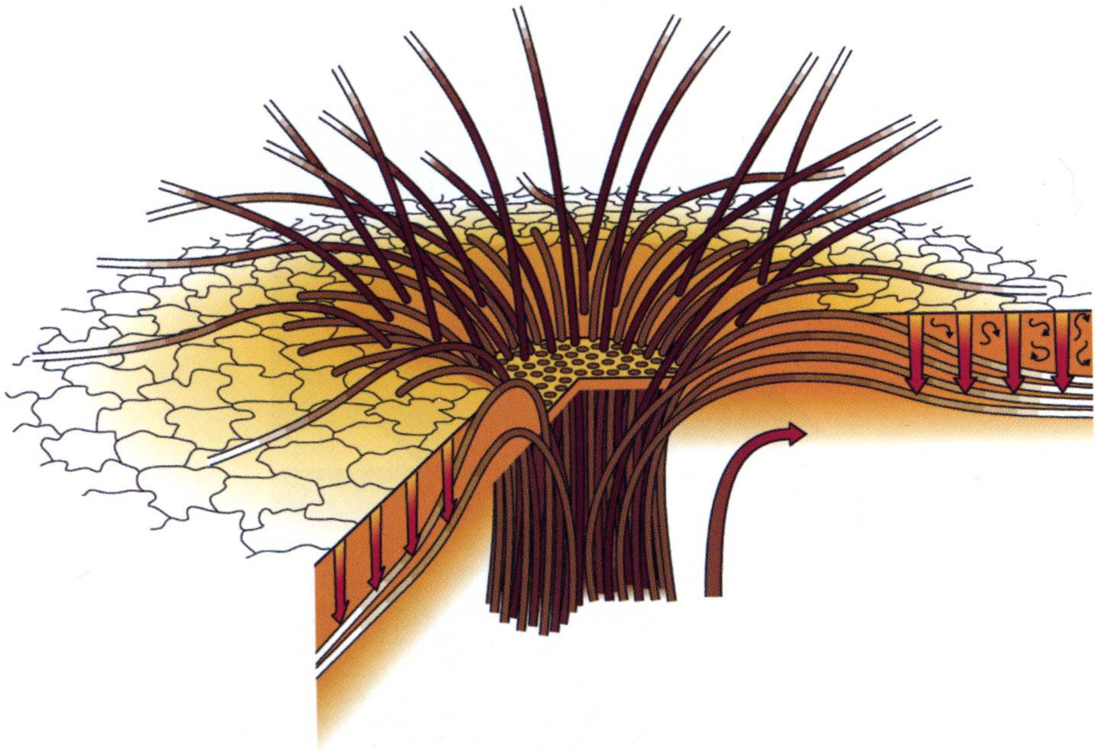

Figure 1.8: *Sketch showing the interlocking-comb structure of the magnetic field of a sunspot penumbra (from Thomas and Weiss 2004). The brown tubes represent the magnetic field lines, of which there are two families, which coexist side by side: the field associated with the bright penumbral filaments rises up into the solar atmosphere, and the field associated with the dark penumbral filaments remains close to the solar surface or sometimes dives down beneath it. The large vertical arrows represent the effect of flux pumping (section 1.5.3) while the large curved arrow represents the sunspot's moat flow (section 1.6.3). Reprinted, with permission, from the Annual Review of Astronomy and Astrophysics, Volume 42 ©2004 by Annual Reviews, www.annualreviews.org.*

field lines do not extend far above the surface. It is less clear how deep down below the surface these fields extend. They could be confined to isolated tubes, enclosed above and below by the more vertical field component, or they could be more like thin vertical slabs that are bounded above but extend downwards for a significant distance.

The observations of penumbral fine structure raise a number of important questions. First of all, how is the extraordinary interlocking-comb magnetic structure formed, and how is it maintained after its formation? Secondly, what is the origin of the observed fine structures – both the bright and dark filaments themselves, and also the smaller-scale details within them? Finally, why is the penumbra so different in appearance to the umbra?

It is likely that the observed penumbral features are ultimately of a convective origin, and therefore we should be able to answer these questions through studies of magnetoconvection. However, at present, the physical processes involved are far from well understood. (We will discuss some of the current theories below – see in particular sections 1.5.2 and 1.6.2.) In fact, the main aim of this thesis is to try to better understand some of these convective processes (and, in the process, to begin to answer some of the above questions), by starting from simple models and working our way up to more complicated (and more sunspot-like) situations. (See section 1.7 below for more details.)

A final noteworthy feature of penumbral structure is the so-called Evershed flow, discovered by John Evershed in 1909. This is a radial, near-horizontal outflow across the penumbra, and appears to be confined to the dark penumbral filaments, containing the near-horizontal fields. (This is consistent with our expectation from magnetohydrodynamics that the flow and magnetic field would tend to be aligned.) The flow is generally interpreted to be a 'siphon flow' along the 'returning' magnetic flux tubes, which dive back down below the solar surface near the outer boundary of the sunspot. What is happening here is that there is a different magnetic field strength at each end of the tube. At the inner footpoint, the field strength may be around 0.1 T (a typical penumbral value). The outer footpoints of these tubes typically correspond to magnetic features in the photosphere, with field strengths typically around 0.15 T. Therefore, there is a higher magnetic pressure, and hence lower gas pressure, at the outer edge of the tube, and this gas pressure difference drives the outflow along the tube.

1.5 Magnetoconvection

The models of overall sunspot structure show that energy is transported through the spot predominantly by convection, and the form taken by this magnetoconvection will influence the appearance and structure of the sunspot. Indeed, the differences in appearance between the umbra and the penumbra must ultimately be caused by changes in the pattern of magnetoconvection as the magnetic field becomes progressively more inclined to the vertical. In order to better understand these changes it is necessary to develop a theoretical understanding of magnetoconvection.

The study of magnetoconvection is an old subject; the linear theory for incompressible magnetoconvection was described by Chandrasekhar (1961), and a review of linear and weakly nonlinear work is given by Proctor and Weiss (1982). With the availability of modern computers, more recent work has turned to three-dimensional simulations of compressible convection, which can be used to investigate the nonlinear behaviour in a variety of regimes.

These simulations are of two kinds. The first kind attempts to include details of all relevant physics (e.g. radiative transfer and partial ionization effects) and to produce results which are directly comparable with observations. These simulations require large computing resources, and so the number of different cases that can be investigated is limited, but such simulations have been quite successful in modelling certain solar convective processes. For example, the MURAM code (Vögler et al., 2005) has been used recently to produce simulations of umbral dots which compare favourably with observations (Schüssler and Vögler, 2006).

A second approach is to simplify the physics and study more idealized models. This allows individual physical processes to be separated and studied in isolation. This kind of simulation also tends to require less computing power, so that many runs can be performed, and the different types of behaviour (occurring for different parameter values) can be catalogued and studied. A disadvantage is that quantitative comparisons with observations are not possible; one can only gain a qualitative understanding.

In general, one wants to run the more idealized type of simulation first, in order to gain a broad understanding of the physical processes involved, and then to follow up with more realistic simulations in order to make comparisons with observations. In the case of sunspot umbrae, realistic simulations are starting to be carried out (as mentioned above), but for the penumbra, we are still at the stage of trying to gain a qualitative understanding of the physics. Therefore, in this thesis we will focus on the more idealized kind of simulations (leaving out radiative transfer and so forth).

We now summarize some of the calculations of magnetoconvection that have been carried out (of the more idealized kind), and explain their relevance to sunspots.

1.5.1 Simulations with vertical magnetic fields

The simplest form of magnetoconvection calculation considers a computational box containing compressible fluid with an imposed uniform vertical magnetic field. The upper and lower surfaces of the box are taken to be impermeable stress-free boundaries, with an imposed temperature difference between the two (in order to drive convection). In the horizontal directions, periodic boundary conditions are usually used (for reasons of computational convenience).

Examples of this type of calculation include Weiss et al. (1990); Matthews et al. (1995); Weiss et al. (1996); Tao et al. (1998); Rucklidge et al. (2000); Weiss et al. (2002). The results vary depending on the imposed magnetic field strength. For very strong fields, the Lorentz forces are strong enough to completely inhibit convection, and the fluid remains motionless. As the field strength is reduced, we come first into the *strong field* regime, in which convection takes the form of small-scale, steady, hexagonal convection cells. As the field strength is reduced further, the convection becomes weakly time-dependent, taking the form of spatially modulated oscillations, where adjacent plumes alternately wax and wane in amplitude. For weaker field strengths still, we enter a *flux separation* regime, where the magnetic field is separated from the convection. There are regions with large-scale, field-free convection, from which the magnetic flux has been expelled, and there are regions with strong fields and weak, small-scale convection. Finally there is a *weak field* regime where the magnetic flux is confined to intermittent regions of intense fields, with the rest of the domain being essentially field-free.

Calculations with a vertical magnetic field are applicable to the umbra of a sunspot (particularly in the central region). The umbra appears to be in the regime where the time-dependent spatially modulated oscillations are found; the umbral dots are then interpreted as regions where a particularly vigorous oscillation has protruded through the radiative blanket. In addition, the dark nuclei observed in some umbrae may be examples of flux separation (Weiss et al., 2002).

1.5.2 Magnetoconvection in inclined magnetic fields

Less attention has been given to the case of an inclined magnetic field (which would be appropriate for the penumbra, or the outer regions of the umbra). Tilting the field has two main effects. Firstly, the problem is no longer rotationally symmetric about a vertical axis, and this can lead to travelling wave phenomena. Secondly, the actual pattern of convection changes when the field is tilted. In a vertical field, hexagons are found near the onset of convection (as mentioned above); as the field becomes progressively more inclined, the hexagons become elongated along the direction of tilt, and eventually give way to field-aligned rolls. This latter solution is the preferred one in a horizontal magnetic field (Danielson, 1961).

Matthews et al. (1992) have looked at the linear theory for compressible convection in tilted fields, as well as some simple nonlinear models, and they were the first to point out that a tilted field would lead to travelling wave solutions. Hurlburt et al. (1996) have carried out two-dimensional simulations with inclined fields, showing the nonlinear development of these travelling rolls. Hurlburt et al. (2000) report some three-dimensional simulation results, which appear to show a transition between cellular and more roll-like patterns as the tilt of the field is increased; however, they describe only a limited number of results, in small computational boxes, so it is perhaps difficult to interpret what is happening.

In addition, Julien et al. (2000, 2003) have performed an asymptotic calculation of this problem in the limit of very strong magnetic fields. They find a transition to a new form of convection for nearly-horizontal fields, which may explain the differences between the bright and dark filaments in sunspots. However, their work makes certain assumptions and approximations and it is difficult to know if the result will still hold in a more general calculation.

We also point out a calculation by Busse and Clever (1990), which studies the linear instabilities of finite-amplitude convection rolls within tilted magnetic fields. Although the paper is more focused on the laboratory rather than the astrophysical case (for example, low magnetic Reynolds numbers are assumed throughout), some of the results may still be applicable to astrophysical situations.

1.5.3 Flux pumping

A recent calculation (Weiss et al., 2004; Thomas et al., 2002b) has demonstrated another phenomenon that can occur in compressible magnetoconvection: flux pumping. In this calculation, a layer of compressible fluid, with no magnetic field, is simulated until a steady convecting state is reached. A thin layer of magnetic field is then inserted into the centre of the box. This layer does not stay where it is, but is moved about by the turbulent convection; the calculation shows that the magnetic field lines are transported preferentially downwards. This process, known as flux pumping, is essentially caused by the strong asymmetry in compressible convection: there are weak broad upflows but very strong downflows.

The pumping is able to hold down the field lines even against their natural tendency to rise via magnetic buoyancy. The calculations show that flux pumping is a robust process, and it provides a possible explanation for how the returning flux tubes, observed in the penumbra (see above), can remain submerged below the solar surface.

1.5.4 Simulations in cylindrical geometry

An alternative approach is to run simulations in cylindrical geometry. For modelling the small-scale features, the geometry is probably not all that important, but to look at the larger-scale structure of pores and sunspots, the cylindrical geometry may need to be included.

The problem is simplified if one assumes axisymmetry, as was done by Hurlburt and Rucklidge (2000) and Botha et al. (2006) (for example), so that only a two-dimensional computation is required. They used a cylinder containing an initially vertical magnetic field. Their results show that this field is quickly swept into a concentrated flux tube at the centre of the cylinder – a form of flux separation. Convection is suppressed inside the flux tube, while convective cells form towards the edge of the domain. The authors suggest that their solution is a good representation of a pore, but the solution cannot really be applied to sunspots, since their penumbrae are highly non-axisymmetric.

An interesting feature of the simulations is that they tend to produce inflows along the surface just outside the flux tube. These flows could represent the 'collar' flow that would be needed to stabilize the sunspot against the magnetic fluting instability. Because the flow observed at the surface outside of sunspots is always *outwards* (the so-called moat flow), the authors suggest that in sunspots, these collar flows would be hidden from view underneath the inclined outer edge of the flux tube.

The next step would be to extend these calculations to non-axisymmetric cylindrical geometry. Preliminary non-axisymmetric calculations have been presented by Hurlburt et al. (2000), which show that the axisymmetric 'pore' results can be unstable to non-axisymmetric perturbations. This may be related to the formation of sunspots from pores (which is discussed below).

1.6 Formation and decay of sunspots

So far we have considered sunspots as essentially static structures. We now move on to describe the processes leading to their formation and eventual disappearance.

1.6.1 Emergence of magnetic fields at the solar surface

Sunspots are magnetic features, and so before we can explain the processes that lead to the formation of a sunspot, we must first of all explain how magnetic fields reach the solar surface at all.

The study of how astrophysical bodies generate magnetic fields is known as *dynamo theory*. In the case of the Sun, magnetic fields are believed to be generated near the base of the convection zone, where a combination of differential rotation and turbulent convection allows large toroidal flux tubes to be generated.

These flux tubes may subsequently be subjected to instabilities driven by magnetic buoyancy, which will cause part of the tube to start rising up through the convective zone. The tube then becomes shredded and frayed by the turbulent convection, before arriving at the surface. This explains the observation of *bipolar active regions*, which are areas where small patches of magnetic field, of both polarities, start to emerge at the solar surface, as illustrated in Figure 1.9. A large active region will typically be composed of several of these rising flux tubes.

The emergence of the magnetic field is accompanied by the appearance of small pores, located at the points where the flux tubes (or flux tube fragments) intersect the solar surface. These pores then start to move towards each other and coalesce, forming larger and larger pores. This coalescence is driven by magnetic buoyancy: as the individual flux tube fragments continue to rise, they draw closer together, like the strings of a bunch of balloons held in one hand.

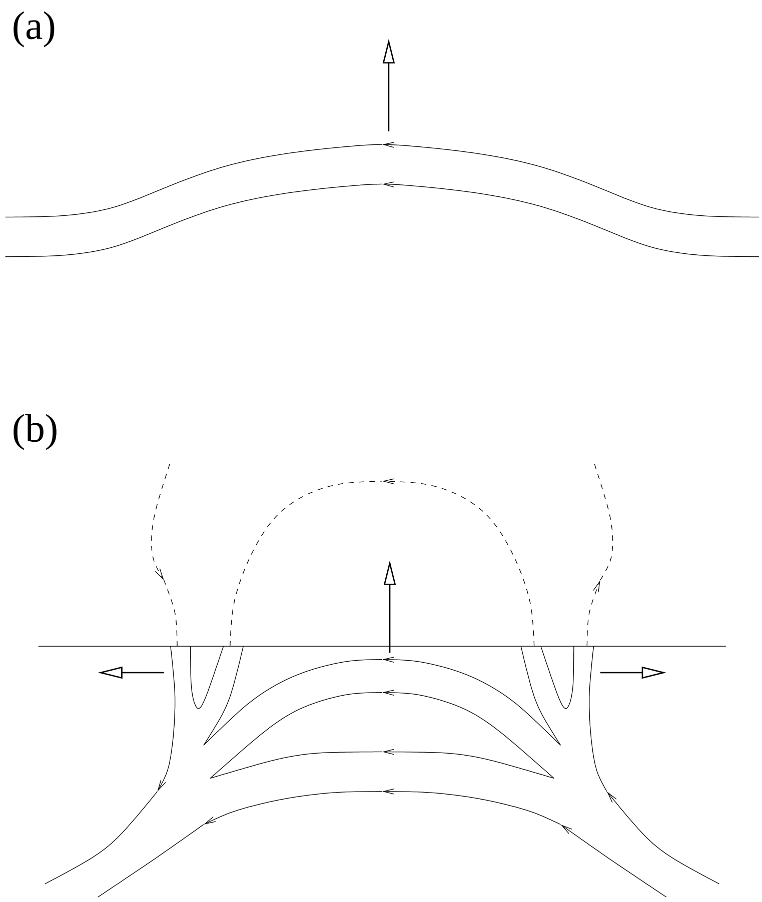

Figure 1.9: *Sketches of a rising flux tube (after Figure 4 of Zwaan 1992). (a) A magnetic flux tube near the base of the convection zone. Part of the tube has started to rise towards the surface. The small arrows indicate the magnetic field direction, while the larger open arrow indicates the motion of the flux tube. (b) The flux tube has reached the surface. On the way up, it has fragmented into several parts. Pores or sunspots will be observed at the points where the fragments intersect the solar surface (the horizontal line). The dashed lines show the magnetic fields that would be seen above the surface. Note how this model naturally produces bipolar active regions containing adjacent groups of sunspots with opposite magnetic polarities.*

1.6.2 Formation of the penumbra

Once the radius of a pore exceeds a certain critical value (approximately 2000 km, according to Rucklidge et al. 1995), a penumbra appears, and the pore develops into a sunspot. The formation of the penumbra is a very rapid event. The penumbra forms sector by sector, with each sector being completed in less than twenty minutes (Leka and Skumanich, 1998), and the whole process is over in under half a day (Zwaan, 1992).

Although the theoretical processes underlying penumbral formation remain largely unknown at present, the observed formation of penumbral sectors, within a short dynamical timescale of 20 minutes or so, does suggest that the transition from a pore to a sunspot is triggered by an instability of some sort. Models of pore structure indicate that the corresponding flux tubes must fan out with height, and as the magnetic flux contained in the pore increases, the radius of the flux tube correspondingly increases, as does the angle of tilt of the magnetic field at the edge of the pore. It is conjectured (Rucklidge et al., 1995) that once this angle exceeds a critical value, an instability sets in which leads to the formation of a penumbra.

One possible candidate is the fluting instability, which we have already discussed earlier in this chapter. The calculation of Meyer et al. (1977) showed that pores and sunspots would be stabilized (at least near the surface) by magnetic buoyancy effects. However, their calculation did not take account of convection, which in reality occurs both inside and outside the pore; therefore, the possibility remains that pores may be susceptible to a modified form of the fluting instability, driven by convection. The conjecture is that such an instability does indeed exist – but only if the field at the outer edge of the pore is sufficiently inclined to the vertical. This condition would be met only once the pore exceeded a certain size, and the resulting instability would drive the transition from a pore to a sunspot.

There are one or two illustrative calculations that demonstrate the possibility of such an instability. In Cartesian geometry, Tildesley (2003b) has found a convectively driven filamentary instability of a highly idealized configuration representing a pore. The nonlinear development of this instability was investigated by Tildesley and Weiss (2004) (see also Tildesley 2003a); it is found that a pattern of bright and dark filaments develops, although these are much wider than the filaments observed in real sunspots. In addition, Hurlburt et al. (2000) have performed a calculation in cylindrical geometry showing that an axisymmetric pore-like configuration is unstable to non-axisymmetric perturbations, again leading to filamentary structure (although again the filaments are rather wide compared with observations).

These calculations suggest that pores could develop convectively driven instabilities at their outer edges. The nonlinear development of this instability has not yet been calculated within a realistic pore model, but it is conjectured (Weiss et al., 2004) that the instability would lead to fluting at the outer boundary, with some families of field lines being displaced upwards, and some downwards. (Hence the term 'convectively driven fluting instability'.) Clearly, more realistic simulations, starting from more realistic pore configurations, are needed to test this hypothesis.

It is further conjectured that once this mildly fluted structure develops, the depressed field lines are then 'grabbed' by the turbulent convection outside the spot, and pulled down even further, via the flux pumping process described above. This would help the initial filamentary structure to develop into a full-blown penumbra, with its interlocking-comb magnetic field, and would explain how the 'returning' flux in the penumbra can form. It would also provide an explanation of why there is hysteresis between pores and sunspots, since once the field lines in the dark filaments are submerged, it would become more difficult to return them to the surface again.

1.6.3 Decay of sunspots

Sunspots have finite lifetimes, varying from hours to months depending on the size of the spot (the larger spots tend to live longer). The decay of a sunspot begins as soon as, or even before, the sunspot is fully formed. There are two main processes by which sunspots decay: gradual decay, and fragmentation.

The gradual decay is marked by a continual reduction in the size of the sunspot and the associated magnetic flux. Observationally, we see that a well-developed sunspot is always surrounded by a so-called *moat flow*, which is essentially a supergranular convection cell centred on the sunspot's position. This leads to a radial outflow around the periphery of the sunspot. Within this outflow are embedded so-called *moving magnetic features*, small flux elements that move radially outwards from the spot. These appear to represent portions of the sunspot's magnetic flux that are slowly chipped away by the convection currents in the moat, leading to the gradual decay of the spot.

Sometimes sunspots can also decay by fragmentation, whereby the spot suddenly splits into several smaller components, the components then disappearing shortly after-

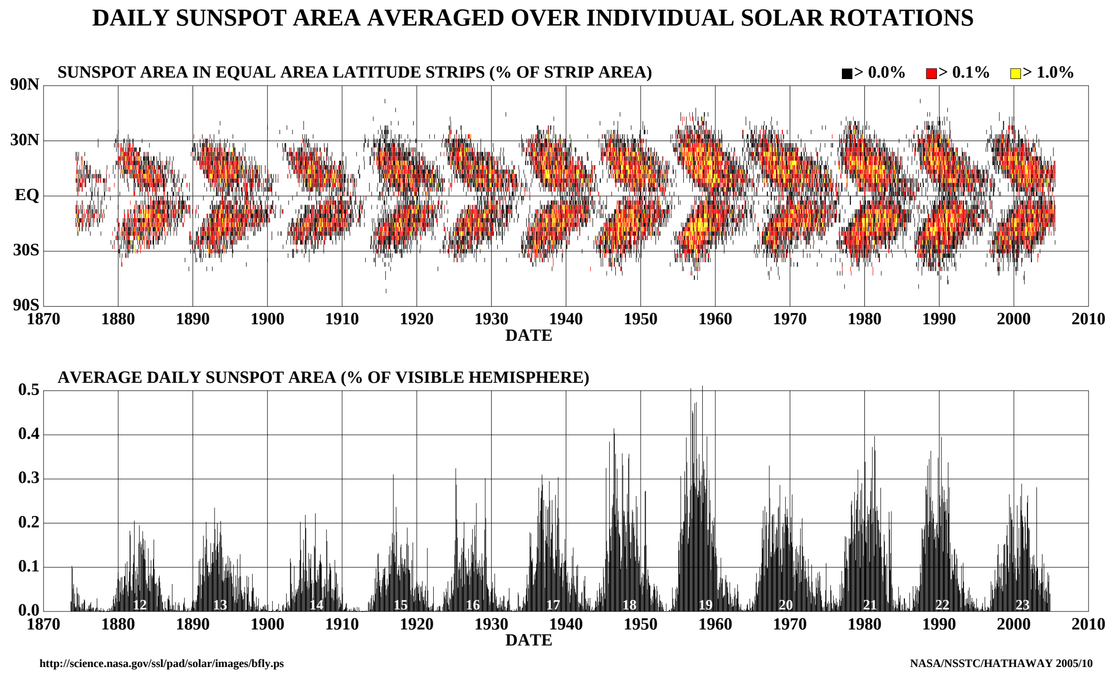

Figure 1.10: *Butterfly diagrams showing (a) sunspot latitudes and (b) total area of the Sun covered by sunspots as a function of time. Courtesy NASA / David Hathaway.*

wards. Often, the lines along which a sunspot splits are the same as the outlines of the pores that came together to form the sunspot in the first place. These lines are also often associated with light bridges.

1.6.4 The 11-year solar cycle

No description of sunspots would be complete without a mention of the 11-year solar cycle. This refers to a cyclic variation both in the number of sunspots observed and in the latitudes at which they appear. This is illustrated in Figure 1.10 which shows the area of the solar surface covered by sunspots at various latitudes for the last century or so. This is usually known as the 'butterfly diagram' (for obvious reasons). Note that the cycle is not completely periodic; the amplitude is modulated from cycle to cycle. There are also periods known as 'grand minima' where sunspot activity becomes weak or non-existent; the most recent of these was the Maunder Minimum of c.1645–1715.

Because sunspots are a surface manifestation of magnetic fields that are actually produced much deeper in the Sun, the sunspot cycle is an indication that the dynamo processes that create these fields operate in a cyclic manner. The solar dynamo is not yet fully understood, but it is widely believed that the Sun produces large amounts of toroidal magnetic flux near the base of the convective zone, where helioseismology indicates that a strong shear layer exists. (A portion of this flux then rises to the surface to create sunspots.) This toroidal magnetic flux is thought to migrate towards the equator as a 'dynamo wave', resulting in the characteristic shape of the butterfly diagram.

1.7 Outline of the thesis

So far, we have given a broad outline of the subject of sunspots, as well as trying to point out some of the unanswered questions in the field. The area that remains most puzzling is the penumbra, with its complex filamentary structure. We currently do not understand the detailed nature of convection within the penumbra, nor do we know how this convection gives rise to the pattern of bright and dark filaments observed at the surface, together with the finer-scale structures found within these filaments. Related to this is the question of why the convection within the penumbra takes on such a radically different form from that within the umbra (as indicated by the difference in appearance of the two regions and the sharp transition between them). In addition, we cannot yet explain how the intricate penumbral structure comes about to begin with, nor how it is maintained in the presence of magnetic buoyancy and other effects.

Clearly, in order to answer these questions, we will need to gain a better understanding of magnetoconvection as it applies to sunspot-like magnetic field configurations. We begin by looking at a simplified problem, that of a uniform magnetic field, inclined at a fixed angle to the vertical. This should shed some light on the form taken by convection in the different parts of a sunspot, where the angle of the field to the vertical varies from nearly zero in the umbra, to 30–60° in the bright penumbral filaments, to 60–90° in the dark penumbral filaments. In Chapter 2 we will investigate the linear stability theory for a simplified model, which will clarify some of the symmetry effects and other aspects of the problem. In Chapter 3 we develop these ideas into weakly nonlinear models, and in the first part of Chapter 5, fully nonlinear numerical simulation results (using the full compressible MHD equations) are presented.[4]

The second part of our work will be to look beyond uniform fields and to produce models in which different parts of the sunspot (e.g. umbra and penumbra) are present within the same model. We do this by setting up a simplified 'sunspot-like' magnetic

[4]A condensed version of Chapters 2 and 3 is given in Thompson (2005).

field at the beginning of the calculation, and then investigating the forms of convection that arise within that field structure.[5] The ultimate aim would be to produce a model showing umbral and penumbral convection, intricate filamentary structure, and all of the other details that are observed on the Sun.

Sadly, we have not been able to achieve this aim. However, our models do at least show a clear difference between the umbral and penumbral forms of convection, as well as a noticeable sharp transition between these two patterns. Thus, they do begin to answer some of the questions posed above; they should also provide a useful starting point for any future research into these problems. (The main feature missing from our results is the interlocking-comb magnetic structure in the penumbra; it seems that more detailed modelling techniques will be needed before this can be reproduced.) The models themselves will be described in Chapter 4 (where a simplified approach based on the Swift-Hohenberg equation is followed) and the second half of Chapter 5 (where the full compressible MHD equations are again used).

[5]This pre-supposes the existence of such a field structure. In other words, we are not attempting to address the question of how a sunspot's field structure is created; we are instead merely trying to understand the convective patterns that arise once such a structure has been formed.