SOLAR AND STELLAR MAGNETISM

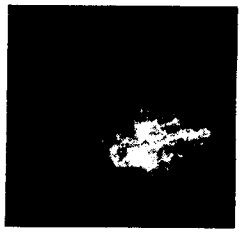
STERREKUNDIG INSTITUUT UTRECHT

April 2, 1999

First edition: April 2, 1999. Based on course CAS1 taught at the Sterrekundig Instituut Utrecht by R.J. Rutten, S.K. Solanki, J.P. Goedbloed and P. Hoyng.

Authors: A. Bik, R.F.S. Collaris, N.L.J. Cox, R. van Deelen, M. Gieles, S. Van Loo, A. van Oosten, D.B. Roy, R.J. Rutten, P.J. Sloover, J. Wiersma, N.P.A. Zagers.

Editors: J.M. Krijger and R.J. Rutten.



A TRACE exposure of a solar active region. The *Transition Region and Coronal Explorer* was launched per Pegasus rocket on April 2, 1998 into a sun-synchronous pole-to-pole orbit. The 30 cm imaging telescope uses combinations of different multi-layer coatings on the four quadrants of the primary mirror and different transmission filters in a filter wheel to project a quarter of the apparent solar disk on a CCD camera in a choice of passbands, varying from white light to EUV wavelengths and from photospheric formation to coronal formation at 2 million Kelvin. The CCD camera has 1024×1024 pixels of 0.5 arcsecond so that the resolution is 1 arcsecond. This EUV image (wavelength unspecified on the website) shows coronal loops above an active region. Taken from `http://vestige.lmsal.com/TRACE/`.

Without magnetic fields the Sun would be as boring a star as the nighttime astronomers think it is.

R.B. Leighton

Magnetic fields are to astrophysics what sex is to psychoanalysis.

H.C. van de Hulst

Contents

Preface

These lecture notes result from a 30-hour undergraduate course titled "Solar and Stellar magnetism" given at the Sterrekundig Instituut Utrecht during the first months of 1999. The motivation for the course was the presence of Prof. S.K. Solanki (ETH Zürich) as Minnaert Guest Professor during this period. The lectures were split in ten sessions of three 45-minute "hours" each:

- R.J. Rutten: introduction;
- R.J. Rutten: stellar magnetism;
- S. K. Solanki: solar magnetic fields;
- S. K. Solanki: magnetic field diagnostics;
- S. K. Solanki: convection, granulation, sunspots;
- S. K. Solanki: flux tubes;
- J. P. Goedbloed: MHD and spectral theory;
- J. P. Goedbloed: heliosphere and solar wind;
- P. Hoyng: dynamo theory;
- P. Hoyng: dynamo issues.

In addition, Prof. Solanki held the "Minnaert Lecture", an evening lecture intended for a wider audience, on solar irradiance modulation and its possible influence on the terrestrial climate.

The students attending the course were asked to produce lecture notes to the course, including the Minnaert lecture. The lectures consisted of rather wide overviews of the field, rather than the cookbook recipes normally taught in undergraduate courses. The students' task was to summarize these reviews in yet more condensed form. This volume is the result.

J.M. Krijger and R.J. Rutten, editors

Bibliography

Books

There are many books treating topics discussed here. This list specifies major textbooks and some pertinent proceedings chapters. The corresponding Utrecht library numbers are added within brackets. Full references are given at the end of this volume.

- A. Bruzek & C.J. Durrant 1977: *Illustrated glossary for solar and solar-terrestrial physics* [4F00 407]
- T. Cowling 1980: *Magnetohydrodynamics* [4E00 S030]
- P. Foukal 1990: *Solar astrophysics* [4F00 S 010]
- L. Golub & J.M. Pasachoff 1997: *The Solar Corona* [4F40 602]
- J. Jackson 1962: *Classical Electrodynamics* [2A60 S018]
- M. Kivelson & C.T. Russell 1995: *Introduction to space physics* [4F00 606]
- M. Kuperus 1998: *Gasdynamica en magnetohydrodynamica*
- E. Landi degl'Innocenti 1992, in Sánchez, Collados and Vázquez: *Solar observations: techniques and interpretation* [4F00 601]
- E. Parker 1979: *Cosmical Magnetic Fields* [4E20 412]
- E. Parker 1996, in Roca Cortes & Sanchez: *The structure of the Sun* [4F95 R 024]
- E. Priest 1982: *Solar Magnetohydrodynamics* [4F00 502]
- E. Priest & A. Hood 1991: *Advances in solar system MHD* [4E20 601]
- B. Roberts 1967: *Introduction to magnetohydrodynamics* [2H70 318]
- E. Schatzman & F. Praderie 1993: *The stars* [4J00 602]
- C. Schrijver & C. Zwaan 1999: *Solar and stellar magnetic activity*
- M. Stix 1989: *The Sun* [4F00 S 008]
- R. Tayler 1997: *The sun as a star* [4F00 608]
- N. Weiss 1974, in Mestel & Weiss: *Saas Fé school 1974* [2J40 432]
- P. Wilson 1994: *Solar and stellar actvity cycles* [4F20 603]
- C. Zwaan et al. 1982: *Magnetische aktiviteit in zon en sterren*

Websites

The number of pertinent websites grows rapidly. A selection:

`http://spaceart.com/solar/eng/sun.htm`
: popular introduction, solar WWW links

`http://antwrp.gsfc.nasa.gov/apod/astropix.html`
: astronomy picture of the day, including many solar ones in its archive

`http://www.lmsal.com/YPOP/Spotlight/Tour/index.html`
: popular introduction, animations

`http://solar-center.stanford.edu/`
: Stanford solar outreach

`http://umbra.nascom.nasa.gov/images/latest.html`
: daily solar images

`http://sohowww.nascom.nasa.gov/synoptic/`
: daily solar images for SOHO planning

`http://sohowww.nascom.nasa.gov/gallery/`
: SOHO image gallery

`http://vestige.lmsal.com/TRACE/`
: TRACE homepage

`http://sohowww.estec.esa.nl/`
: SOHO homepage

`http://umbra.gsfc.nasa.gov/`
: NASA solar data center

`http://www.nso.noao.edu/diglib/`
: NOAO–NSO solar data library

`http://sohowww.estec.esa.nl/spacesolarastro.html`
: ESA collection of solar websites

`http://www.sel.noaa.gov/index.html`
: NOAA space weather

`http://www.sel.noaa.gov/primer/primer.html`
: space weather primer including bursting oil pipes

`http://www.lmsal.com/solarsites.html`
: Karel Schrijver's collection of solar websites

`http://www.amara.com/science/science.html`
: Amara Graps' astrophysics link collection

`http://adsabs.harvard.edu/abstract_service.html`
: ADS abstract server

`http://adsabs.harvard.edu/article_service.html`
: ADS paper server

`http://www.astro.uu.nl/~rutten/`
: solar physics abstracts with BibTeX data

Chapter 1

Introduction

R.J. Rutten

1.1 Why study solar and stellar magnetism?

Magnetism lies at the root of most solar and heliospheric physics, and likewise at the root of much stellar physics. The intricate structure of the solar field, the activity cycle and the influence of the field on the heliosphere represent major quests of (astro-)physics which bear directly on the human environment. The Sun's magnetic field is generated by enigmatic dynamo processes in the solar interior, is organised into the highly complex patterns of solar activity observed in the solar photosphere, dominates the structure of the outer solar atmosphere (chromosphere, transition region, corona), regulates the solar wind, and affects the whole extended heliosphere. Stellar magnetism plays key roles in the magnetic braking of young-star rotation and in collapsed old-star remains.

Solar and stellar magnetism pose questions that require answers in the context of different human endeavours. In terms of pure science, solar magnetism provides the Rosetta stone of cosmic magnetism, not only for other stars but also with respect to accretion disks, cataclysmic variables, galactic dynamics, active galactic nuclei and other cosmic objects in which MHD and plasma processes control the structure and energy partitioning. The Sun is close enough to study these processes in observable detail. The other stars sample different stellar structure, stellar evolution, and stellar companionship, and so widen the parameter ranges over which cosmic magnetism can be diagnosed.

Solar and stellar magnetism also provide input to physics. Terrestrial plasma confinement machines do not reach the scales, densities and temperatures exhibited in the solar atmosphere. Solar physics therefore complements plasma instability studies in fusion research. To physicists, the Sun represents a non-terrestrial plasma physics laboratory that greatly enlarges the accessible parameter space.

Finally, space weather (the American policy term for solar influences on the near-Earth environment and on the Earth's climate) is set by solar magnetism. Solar activity affects satellites, communication links, and power lines, and its modulation affects satellite orbits, jet stream patterns, and is likely to contribute to minor, possibly major, ice ages.

1.2 Solar magnetism gallery

Another reason to study solar and stellar magnetism is one of sheer beauty. Our Sun permits viewing stellar magnetic structures with the angular resolution necessary to appreciate the intricate topology dictated by magnetic fields and their interaction with the plasma. Even though the smallest observable feature sizes (about the province of Utrecht at best!) exceed the intrinsic physics process scales by factors that range from a few times to many orders of magnitude, solar "fine structure" as we observe it is indeed "fine" structure in the best sense of the word, providing rich morphology and spectacular dynamic change to the awed viewer. The rest of this section provides a small sample, ordered from the largest to the smallest observed scales.

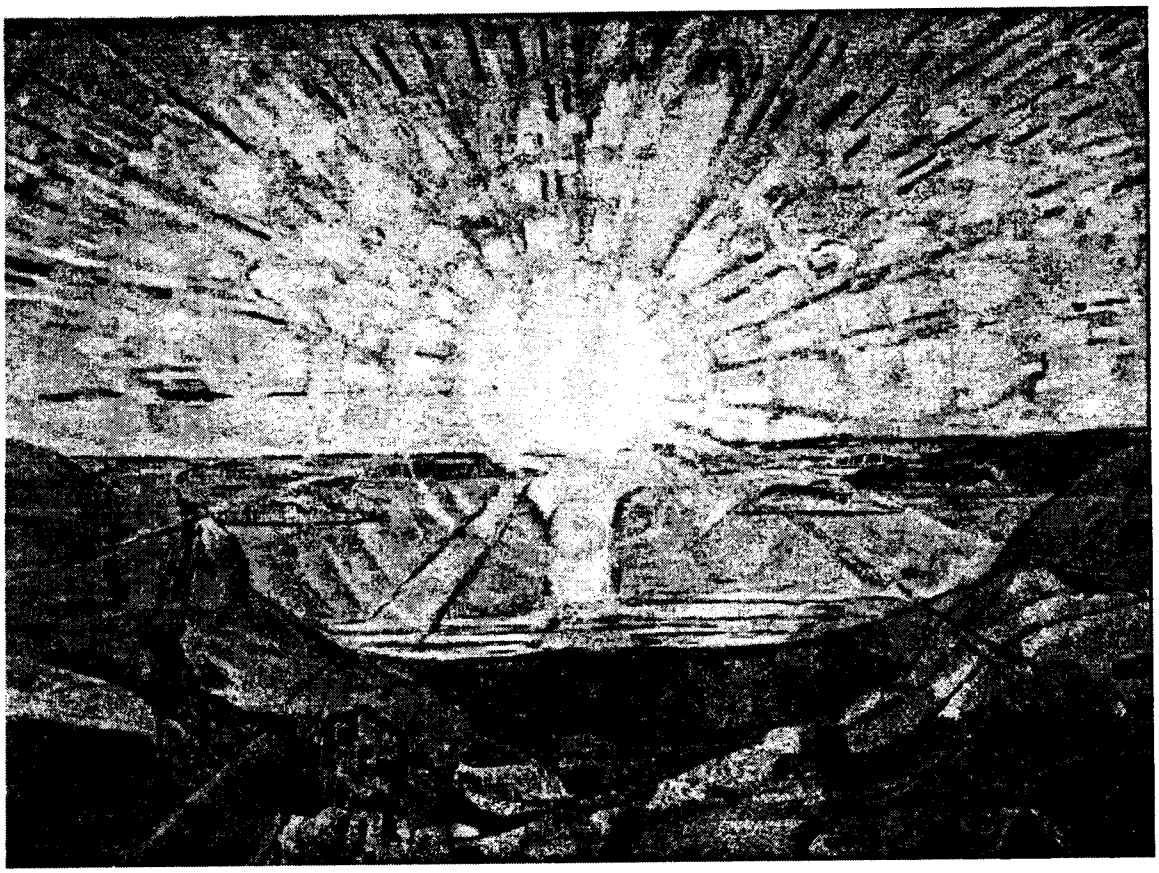


Figure 1.1: Sunset, by Edvard Munch (1863–1944). "The camera cannot compete with the brush and the palette so long as it cannot be used in heaven or hell", said Munch. Whether photographed or painted, sunsets are beautiful without being diagnostically useful within the context of these lecture notes (which isn't the case when the Sun is eclipsed by the Moon rather than the Atlantic Ocean). To optical solar physicists, the painting illustrates undesirable atmospheric refraction and bad optics. Nevertheless, it conveys solar beauty admirably. More on Munch at `http://www.mnc.net/norway/Edvmunch.htm`.

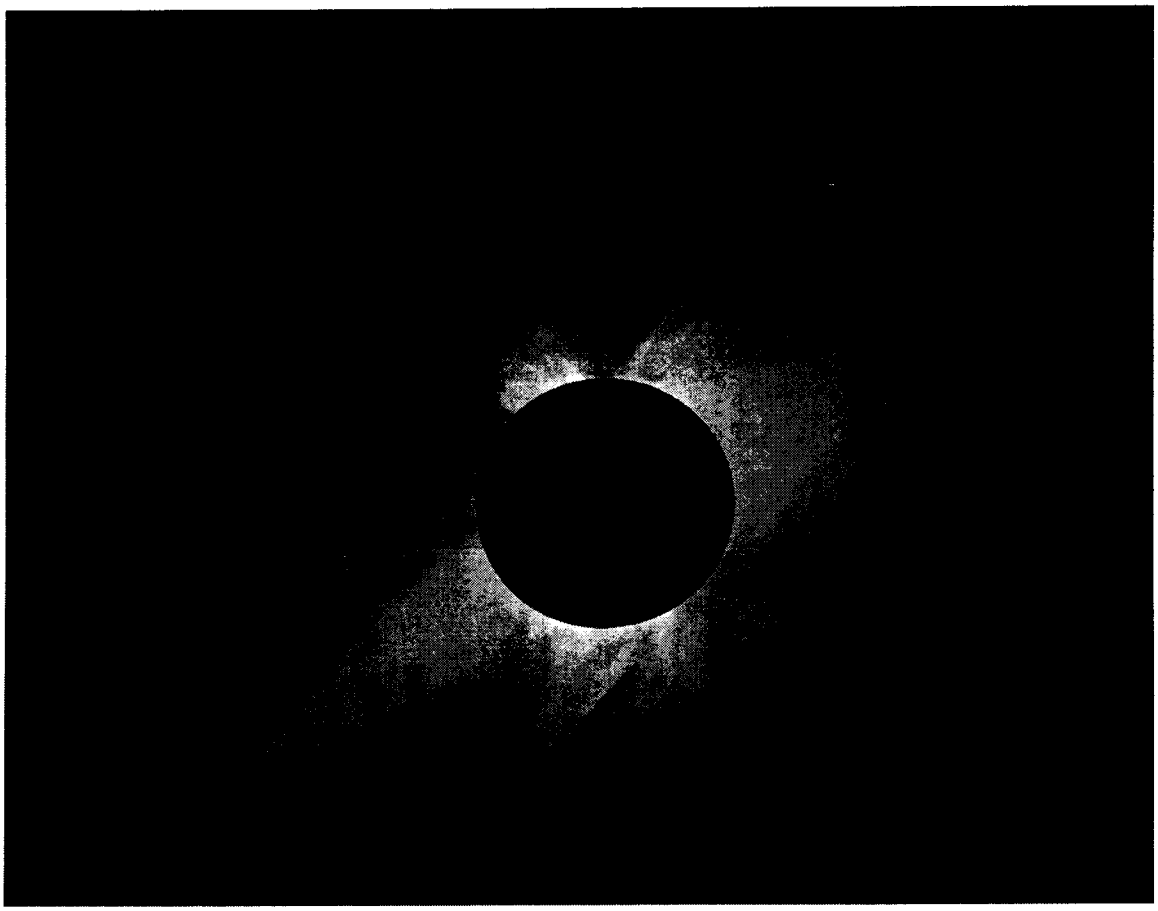


Figure 1.2: The solar corona at the July 11, 1991 total solar eclipse. The Sun's beauty is miraculously enhanced for terrestrial viewers by the apparent size equality of the Sun and the Moon on our sky. The corona seen visually during totality is about a billion times weaker than the uneclipsed solar disk.

How does this faint white glow betray the million-degree coronal temperature? The inner part (the K corona) shines through Thomson scattering off free electrons, using photons that originated in the unseen photosphere. The electron density is very low and the Thomson cross-section is very small, but the corona contains sufficient gas along the line of sight to produce emission that exceeds the background set by scattering off interplanetary dust (the F corona, seen further out) and by scattering in our own atmosphere (the sky background yet farther out). The electron-scattered photons are Doppler-shifted due to the large thermal motions, corresponding to line broadening with Doppler width $\Delta\lambda_D \approx 100$ Å, so that the photospheric Fraunhofer spectrum is washed out. Only shallow dips remain from the very strongest lines (Ca II H & K). The absence of the Fraunhofer lines prompted Grotrian to propose that the coronal gas must be exceedingly hot. The hypothesis was confirmed by Edlén's identification of the few bright emission lines in the visible coronal spectrum as being due not to a new element "Coronium" but to forbidden transitions of highly ionized iron and calcium particles

It is clear by now that the high temperature represents the balance between magnetically controlled heating and X-ray radiation losses — but the precise heating mechanisms are not yet identified in detail. The present course provides a foundation from which you may start your research effort to answer this basic question. This picture was constructed from five different exposures bridging the large contrast between inner and outer corona. Even so, it barely approaches the magnificence of a visually observed total eclipse — as may be confirmed by bringing a pair of binoculars to the right place on August 11, 1999. From `http://spaceart.com/solar/`.

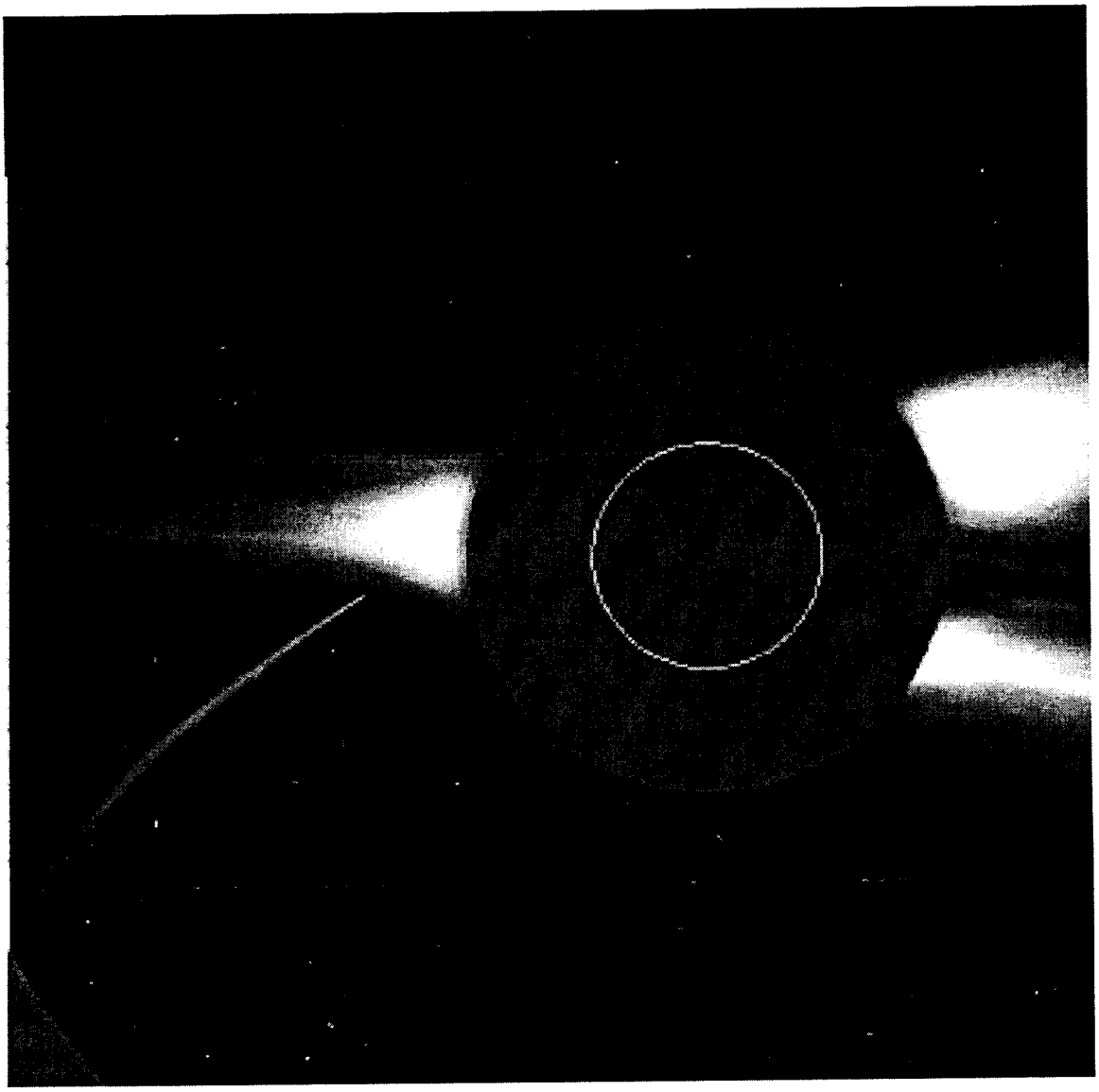


Figure 1.3: The outer corona seen from space, an image from LASCO on-board SOHO taken on December 23, 1996. Outside the Earth's atmosphere an space-walking astronaut may simply use a coin (about a quarter) at arm's length to view a total solar eclipse. The LASCO telescope eclipses the Sun continuously in the same manner, so that the evolution of the coronal structure can be monitored at much higher cadence than from eclipse to subsequent eclipse.

This image shows the corona, including polar plumes all the way out to the edge of the field of view (8.4 million km radius). It also shows Comet SOHO-6 plunging into the Sun (it didn't reappear on the other side). The circle shows the size of the Sun as we see it; the blocking disk is much larger. From `http://spaceart.com/solar/`.

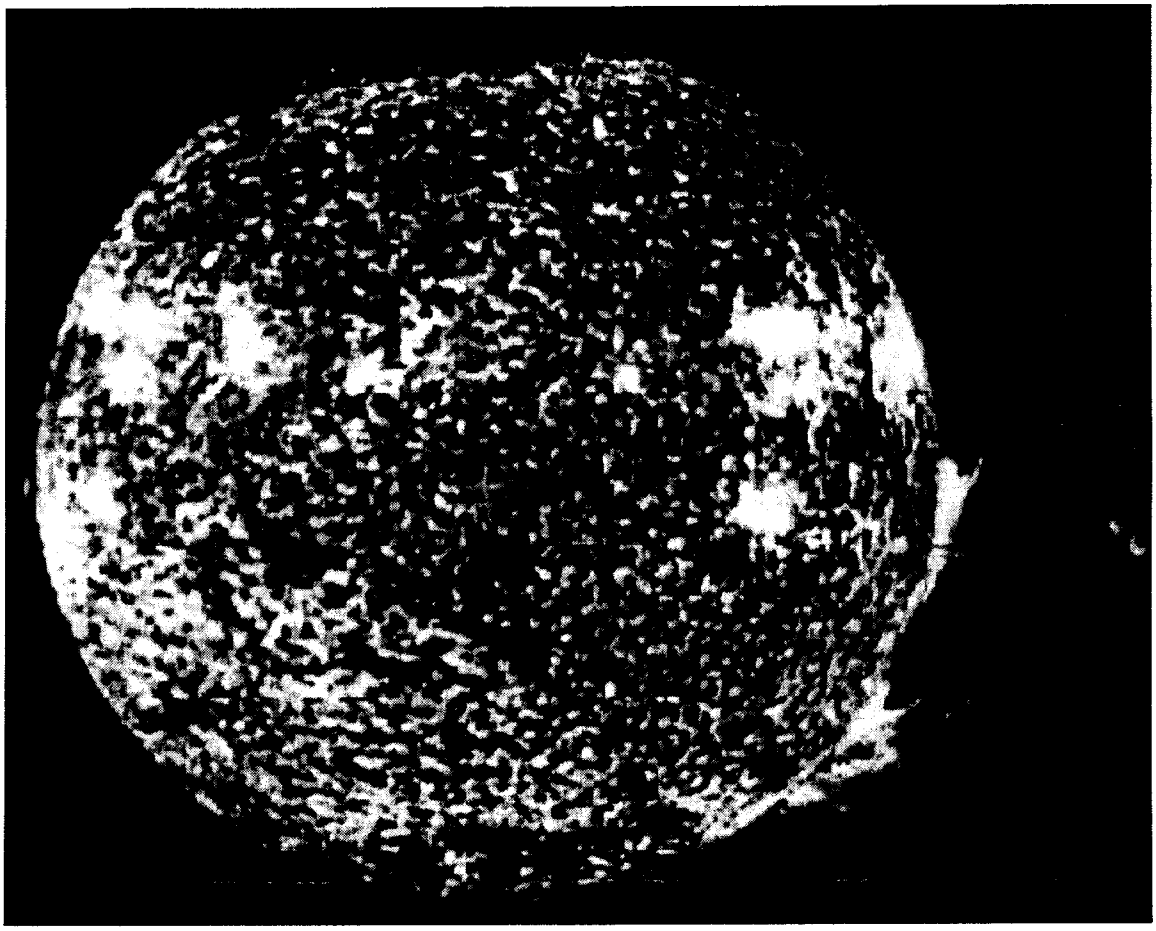


Figure 1.4: The Sun in He II 304 Å. A space platform not only permits eclipses all the time but also permits access to the wide regions of the electromagnetic spectrum that are absorbed by the Earth's atmosphere. This EUV image was taken by an astronaut on-board Skylab, utilising the Apollo Telescope Mount (at that time ATM did not yet mean automated teller machine), a worthwhile leftover from the Apollo Moon program.

The chromospheric network, overlying the borders of the supergranulation cells, appears bright due to magnetic heating, and so do the active regions. A giant prominence juts out far into the corona. Since helium must be once-ionised to show up in the 304 Å resonance line, the gas in the prominence must be much cooler and therefore heavier than the surrounding 1–2 million degree corona.

Obviously, the filament presence and structure maps out magnetic fields. Less obvious is the field geometry, how prominences are suspended, and how they survive as long as they do. The most recent Utrecht solar physics thesis prior to the publishing date of this volume treated solar prominences (Schutgens 1998), but the last word hasn't yet been said about them. From `http://spaceart.com/solar/`.

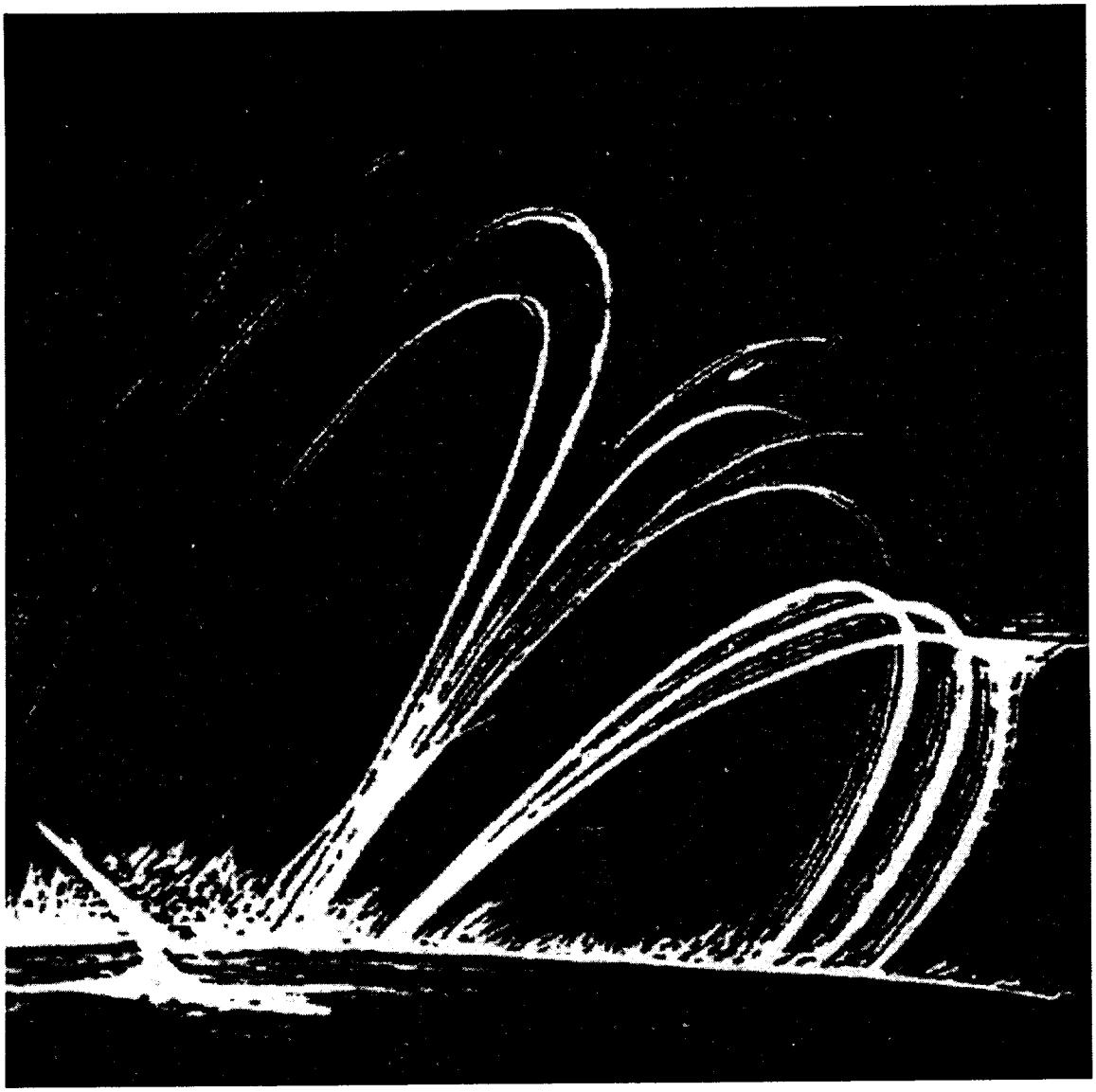


Figure 1.5: Part of a drawing of coronal loops and limb spicules made on October 5, 1871 by A. Secchi who founded a Jesuit observatory in Rome, wrote 700 papers and two books (*The Sun* and *The Stars*), all within three decades, and started spectral classification. He was not only a prolific publicist but also an excellent observer. This drawing was made from spectroheliosope observations selecting radiation in the Balmer Hα line. From Foukal (1990).

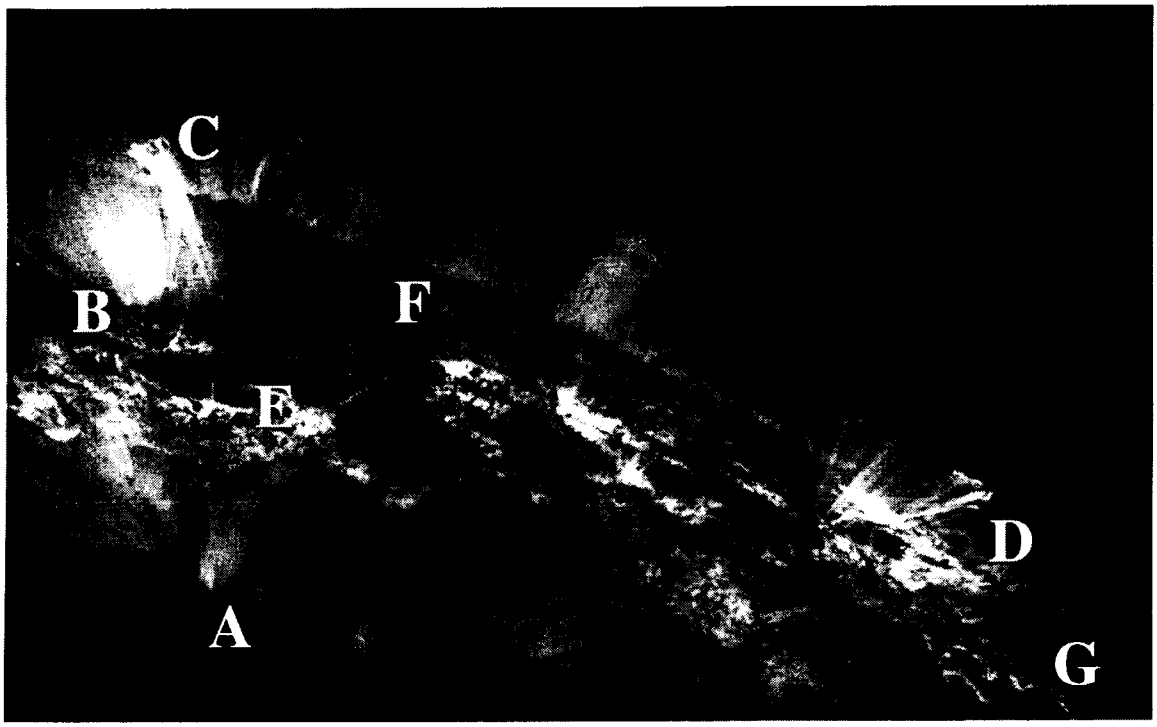


Figure 1.6: The state of the art in coronal loop observation. An image from TRACE (Transition Region and Coronal Explorer) taken in the Fe IX – Fe X line pair at $\lambda = 171$ Å on August 19, 1998 at 06:04 UT, taken from Schrijver and Zwaan 1999.

The Fe IX and Fe X ionization stages exist at temperatures near 1.0 million Kelvin. TRACE images the Sun in 171 Å light at an angular resolution of 1 arcsec (725 km) through a combination of multi-layer coatings on the primary mirror and a transmission filter; there are also other spectral passbands. The field shown here measures 400 arcsec per side.

The TRACE 171 Å images are particularly dramatic because loops at this temperature are relatively scarce. In between the bright loops seen here exists coronal gas of higher temperature that is transparent at 171 Å. Thus, the loops seen here appear as special trees in a forest of transparent other trees — say oak trees scattered though a pine tree forest. In addition, there is much dark material that intercepts radiation from behind. This is much cooler material containing once-ionised helium or perhaps even neutral helium and hydrogen. It scatters 171 Å photons out of the TRACE passband by bound-free absorption and re-emission near the threshold wavelength.

Most loops reach higher than the local pressure scale height so that their emission is concentrated in their lower parts (*e.g.,* near A and B). Some of the fainter tops are seen as well. The loops are optically thin, so that lines of sight along their tops produce more emission, as in the relatively dense post-flare loops near D.

The conduction of energy downwards along the invisible hotter loops leads to the finely speckled "moss" covering plage areas, for example near E. The latter areas also abound in short "grass" stalks of cool dark gas; these are short-lived ejections that appear only momentarily on high-cadence TRACE movies. The limb has similar dark spicule-like structures (above F); these are longer than the classical limb spicules observed in Hα. Near G is a filament, temporarily wrapped in a shimmering sheet of intricate small-scale brightenings.

All TRACE data are available at `http://vestige.lmsal.com/TRACE/Data/datacentprivate.shtml`.

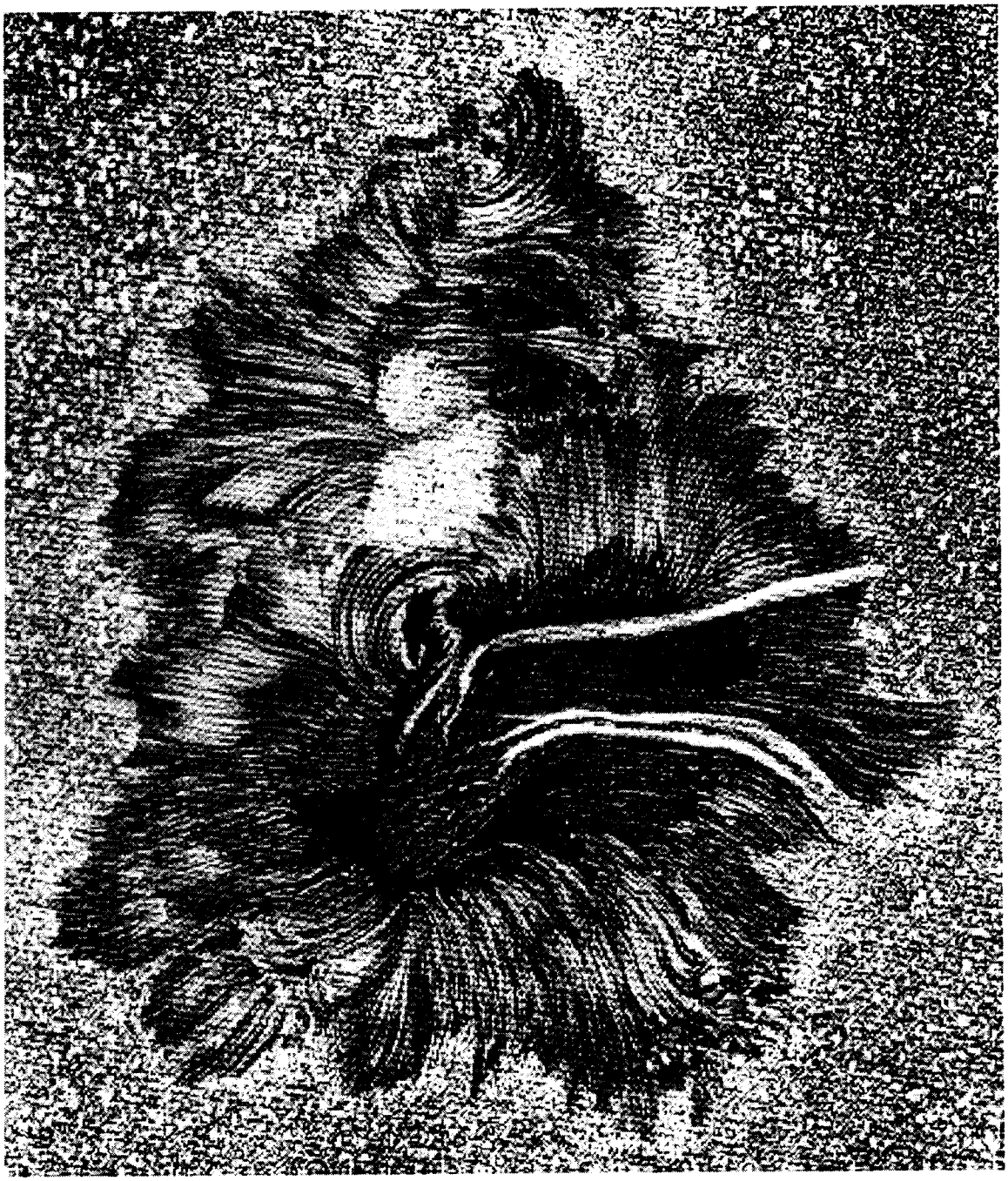


Figure 1.7: Part of a sunspot drawing made by S.P. Langley on September 21, 1870. He observed with a 13-inch (33 cm) telescope at Allegheny Observatory. Note the umbral light bridges and the penumbral filaments. Drawings are of course somewhat less objective than photographic or CCD recordings, but the beauty of a sunspot is conveyed particularly well. From Foukal (1990).

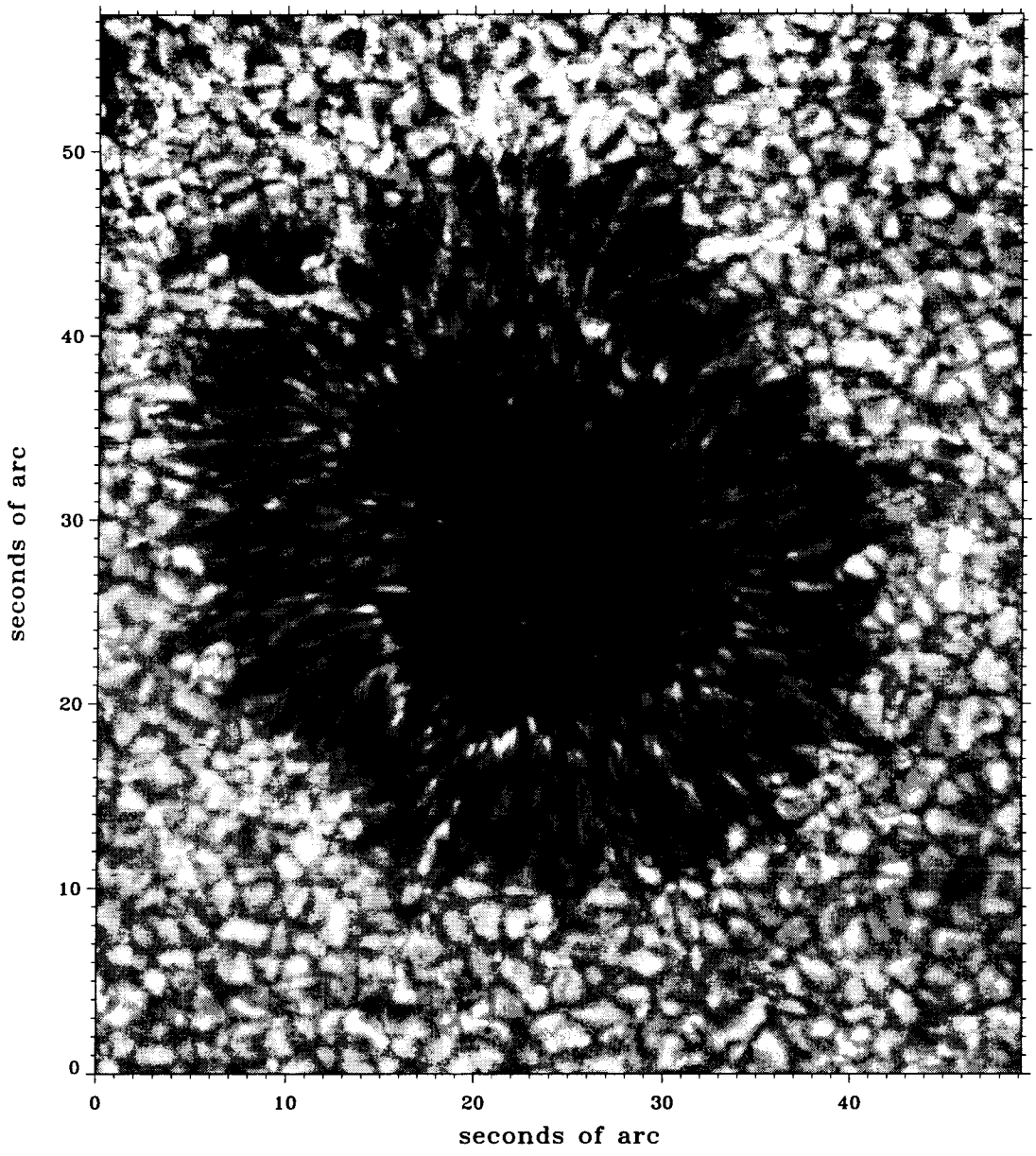


Figure 1.8: Sunspot at high resolution, taken with the Swedish Vacuum Solar Telescope on May 12, 1998. This display is a small part of a 2K×2K pixel G–band image that as a poster adorns walls in many astronomy institutes (including the Sterrekundig Instituut Utrecht and all *Volksterrenwachten* in The Netherlands). It was taken by Göran Scharmer and Tom Berger in the presence of the editors of this volume. A fast tracker system kept the telescope beam into place during the brief (about 15 ms) exposure ("tip-tilt wavefront correction"). The seeing was of a quality that is rare even at La Palma. From `http://www.astro.su.se/groups/solar/`.

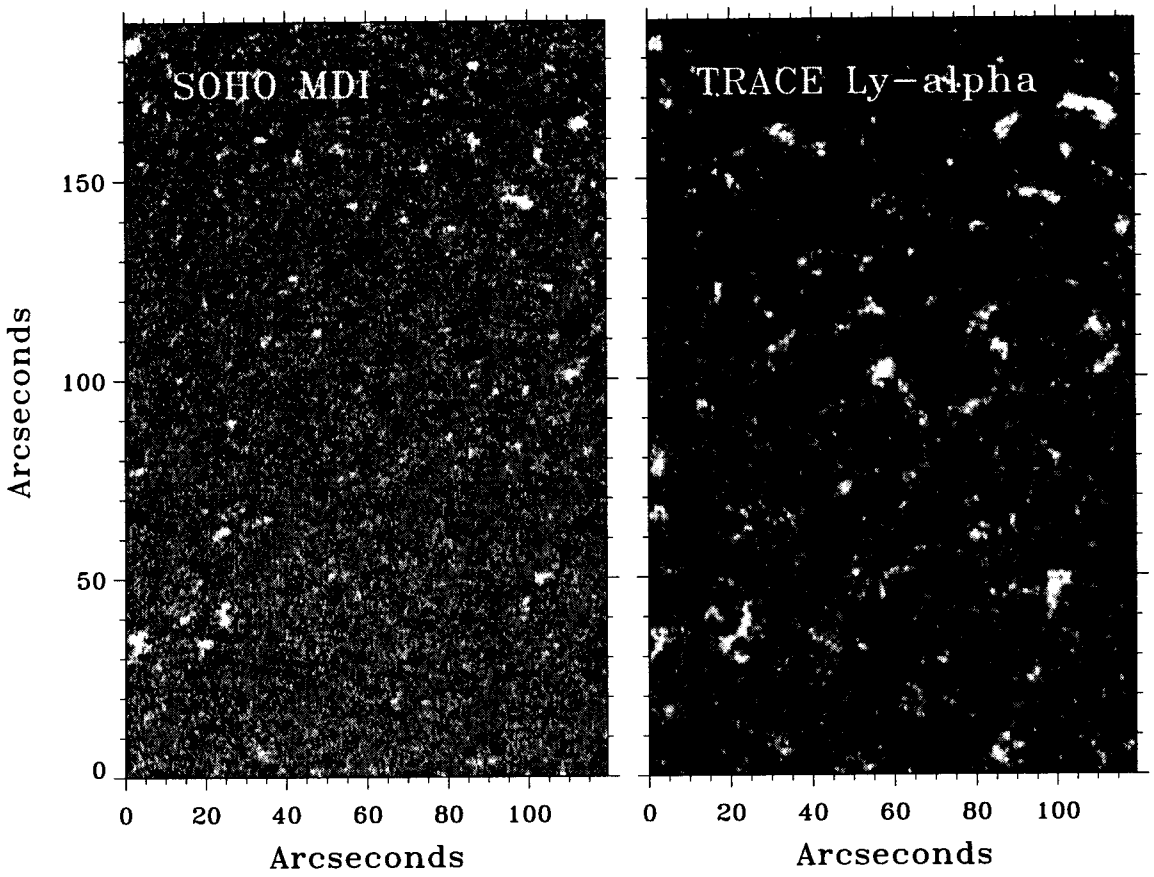


Figure 1.9: The quiet-Sun outside active regions. These two images show the network in close-up. The lefthand one is a magnetogram taken with MDI onboard SOHO; the righthand one is a simultaneous and co-spatial Ly α filtergram taken with TRACE.

The MDI (Michelson Doppler Interferometer) takes magnetograms using the weak photospheric Ni I 6768 Å line by measuring, per image pixel, the difference between the left and right circularly polarized amplitude at two wavelengths (Stokes V signal). The magnetic network appears as sparse patches of positive and negative polarity (black and white). Each patch consists of a cluster of thin vertical fluxtubes with intrinsic magnetic field strength of about $B = 1400$ Gauss at the bottom of the photosphere.

The H I Ly α image at right shows the brightness pattern of radiation emerging from the middle chromosphere. The locations where the chromosphere is heated appear bright. The emission pattern corresponds closely to the underlying photospheric field pattern at left, irrespective of polarity, for all network patches. The remaining "internetwork" areas show small-scale grainy structure in both panels. It is mostly due to non-solar detector noise in the MDI magnetogram, but in the Ly α image at right it shows solar acoustic oscillation patterns. The latter probably cause the so-called "basal emission" in chromospheric lines that is observed for the least active cool stars. Most of the chromospheric emission comes from the magnetic patches, however, and clearly betrays heating that, whatever the precise mechanism may be, is directly related to the presence of strong magnetic field.

Data taken on June 14, 1998, shortly before SOHO was lost for an extended period. Data reduction by C.J. Schrijver and H.J. Hagenaar, taken from Rutten (1999a).

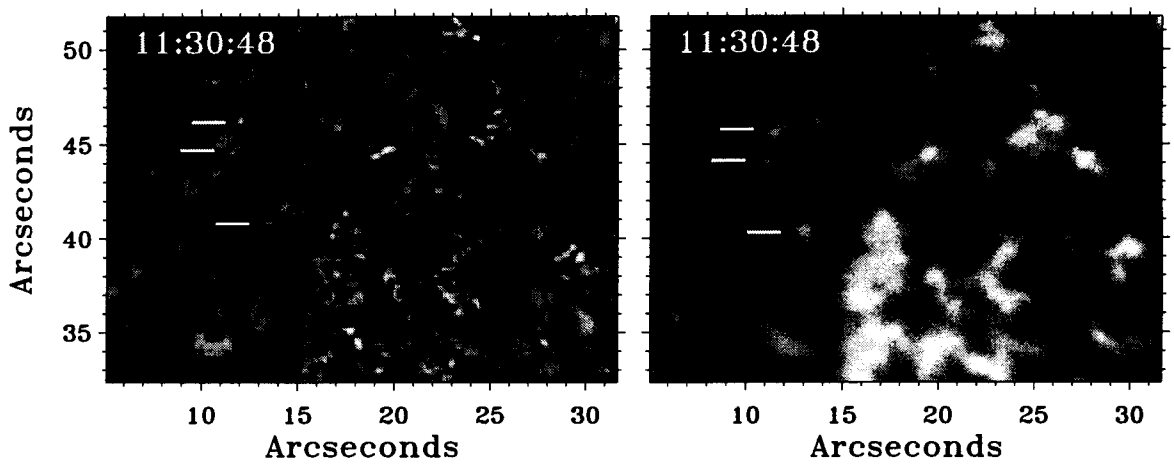


Figure 1.10: Magnetic elements in the solar photosphere (left) and chromosphere (right) at the smallest scale observed so far. These magnificent cospatial high-resolution images were taken with the Swedish Solar Vacuum Telescope on La Palma by R.A. Shine (Lockheed-Martin).

The lefthand image was taken in the Fraunhofer G band around $\lambda = 4305$ Å which contains a few dozen dark CH lines that contribute high contrast. The image was corrected for wavefront distortions introduced by the "seeing" in the Earth's atmosphere using phase-diverse speckle restoration. In this technique sequences of image pairs are taken with one frame in focus and the other one deliberately out of focus. The two frames per pair sample instantaneous atmospheric phase retardation. The sequential pairs sample different atmospheric behavior. The subsequent computer restoration (an iterative process taking weeks even on modern workstations) by Löfdahl *et al.* (1998) delivered resolution at the telescope diffraction limit of 0.2 arcsec.

At such high angular resolution, individual fluxtubes appear as tiny bright "filigree grains" located in intergranular lanes. Most of these grains lie together in network patches; the markers indicate three "loose" ones a bit away from the magnetic network.

The righthand panel shows a nearly simultaneous Ca II K filtergram, using a filter that transmits the Ca II K line core with 3 Å bandwidth. The Ca II K core has its $\tau = 1$ characteristic escape depth about 1000 km above the photospheric level from which the G-band photons emerge. It suffers from strong resonance scattering over about the same distance, so that the details in this filtergram are much fuzzier than in the G-band image (1 arcsec = 725 km on the Sun). Nevertheless, this image pair demonstrates that the core of the Ca II K line brightens considerably at the fluxtube locations, even for the three loose ones away from the clusters.

Thus, one may use Ca II K brightness as a proxy magnetometer, indicating the presence of kilogauss field elements without spatial cancelation when positive and negative polarity patches are added up due to low angular resolution. This property has been used extensively in measuring and monitoring stellar magnetism, also in the Utrecht theses of Middelkoop (1982), Oranje (1985), Schrijver (1986) and Rutten (1987) supervised by C. Zwaan. From Lites, Rutten & Berger (1999).

Chapter 2

Solar magnetism

Lecturer S.K. Solanki, notes by A. Bik

2.1 Introduction

The Sun is an unique object in astrophysics. It is the only star whose surface we can resolve with current telescopes, so that we can observe more physical processes in great detail on the Sun than on any other star. From observations of the Sun, we know that the atmosphere of the Sun is very structured. The interplay between magnetic forces, turbulent motions and heating processes gives many interesting phenomena we can observe in detail.

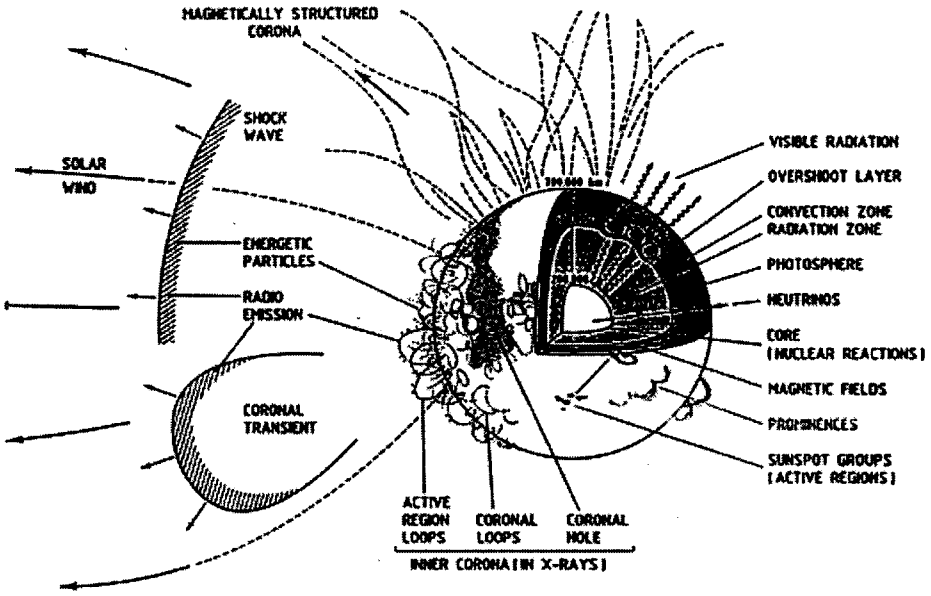


Figure 2.1: A sketch of a lot of features visible at the surface and in the atmosphere of the Sun.

The magnetic field of the Sun is much more complex than the magnetic field of the Earth. The Earth magnetic field can, within a good approximation, be described as a dipole magnetic field. The deviations can be described as perturbations on an ideal dipole field. On the other hand, the magnetic field of the Sun is much more complicated.

Magnetic fields influence the dynamics and thermal structure of the solar atmosphere in many ways. The manifestations of the magnetic field are detected in wavelengths varying from radio to X-rays. Direct observations of the magnetic field, however, are now largely restricted to the visible and infrared wavelengths (Zeeman effect) and some measurements of cyclotron radiation at radio wavelengths.

In the first section we will discuss an active region on the solar surface, starting with the features in the photosphere, visible in optical wavelengths, *e.g.*, sunspots, pores and plage. After that we go higher up in the atmosphere to the chromosphere and the corona and describe features like streamers, solar flares and coronal mass ejections. In the last section we will describe the magnetic features of the quiet sun, *e.g.*, supergranulation and facules.

2.2 Active regions

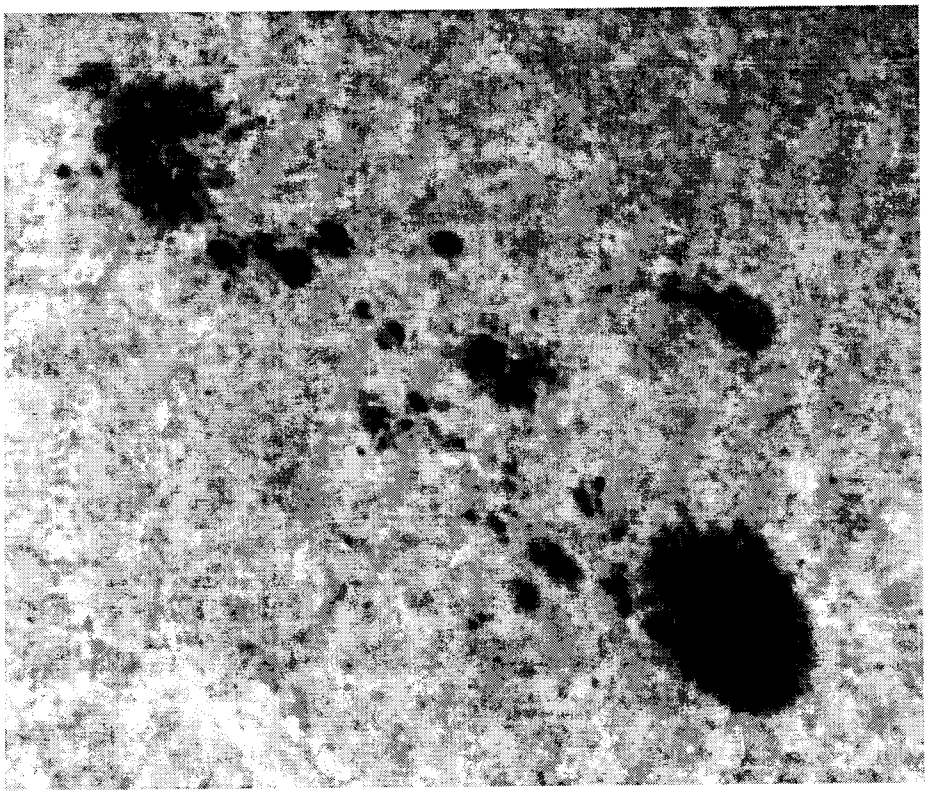


Figure 2.2: An optical image of an active region at the surface of the sun. From Bib Bear solar Observatory (`http://www.bbso.njit.edu`).

When we look at an active region (Figs. 2.2 and 2.6) we notice many different features. A schematic overview of these is given in Figs. 2.3–2.4.

An active region is a mix of magnetic field and plasma, which is constantly changing in structure (Cox *et al.* 1991). An active region consist of a wide range of different features like sunspots, plage and filaments. All these different features can be observed in different wavelengths.

2.2.1 Magnetic features in the photosphere

The magnetic field of the Sun is produced in the interior of the Sun by a dynamo process. The dynamo is located at the bottom of the convection zone (Chapt. 6). Bipolar regions at the surface of the Sun are formed when magnetic flux emerges there in the form of a loop (which is a fluxtube in the shape of an Ω).

The largest flux concentrations, or fragments of the original fluxtube, block the convective energy transport and are visible as *s*unspots. Most of them exist in pairs, one with a negative polarity, and one with a positive polarity. Because of the Maxwell equation $\nabla \cdot B = 0$, which states that there are no magnetic monopoles, the magnetic flux can only emerge in a bipolar form. There are also "open" field lines, but these are either "closed" at the boundary of the heliosphere, or connected with the interstellar magnetic field lines.

In a sunspot (Chapt. 11), we distinguish between two regions. The inner part of the sunspot, the *umbra* and the outer part of the sunspot, the *penumbra*. The umbra is much darker than the penumbra. At wavelengths corresponding to green light, the luminosity of the umbra is 8% of the normal, quiet, photosphere's luminosity. The luminosity of the penumbra is about 80% of the photospheric value (Zwaan *et al.* 1982). The direction of the magnetic field is also different, the magnetic field in the umbra is nearly vertical, while the magnetic field in the penumbra is more horizontal.

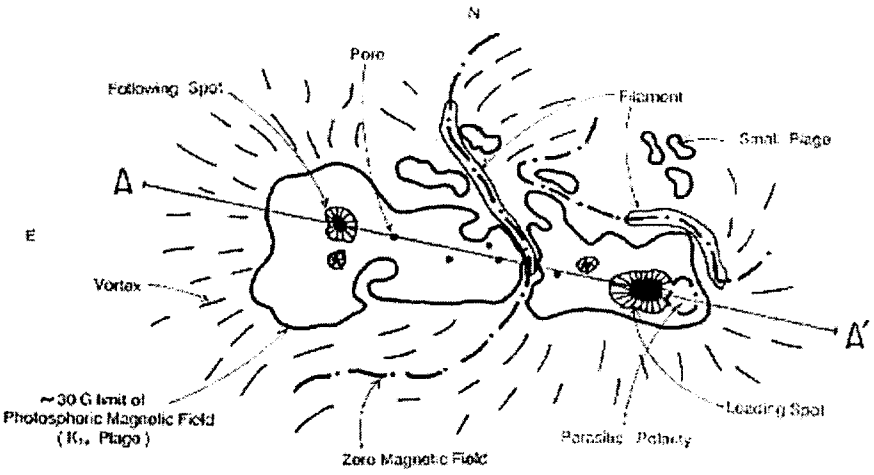


Figure 2.3: A sketch of an active region at the surface of the sun. Taken from Cox *et al.* (1991).

A somewhat smaller dark feature, visible in the optical, is the *pore*. A pore can be described as a sunspot without a penumbra. Typical diameters of pores are in the range of 1500 – 6000 km, much smaller than sunspots (5000 – 60,000 km). At diameters of about 5000 – 6000 km there is an overlap between pores and sunspots. Most of the magnetic features in this size range are sunspots with incomplete penumbrae.

The magnetic field in the pore is relatively vertical, although it becomes somewhat inclined near the outer edge. The maximum strength of the magnetic field varies from 1900 – 2500 Gauss near the center of the pore and is a few hundred Gauss lower at the edge. This is smaller than the strength of the magnetic field in the umbrae of sunspots (2500 – 3300 Gauss).

The lifetime of pores is on the order of hours to days. Pores, and also sunspots, grow rapidly through merging of smaller magnetic features. The pores, in their turn, can evolve to sunspots.

When an active region is formed, first the *plage* appears. Plage are bright regions and consist of 10^3 to 10^6 *fluxtubes*, small magnetic elements. The strength of the magnetic field in one fluxtube is of the order of 1500 Gauss.

The magnetic field is not homogeneous over the plage. The fluxtubes are very small (100 km diameter). The filling factor, the fraction of the area which is covered by magnetic field, can vary over a large range, but is generally much smaller than unity. Typical values for the filling factor in plage are 10 – 20 %.

2.2.2 Magnetic features in the chromosphere and the corona

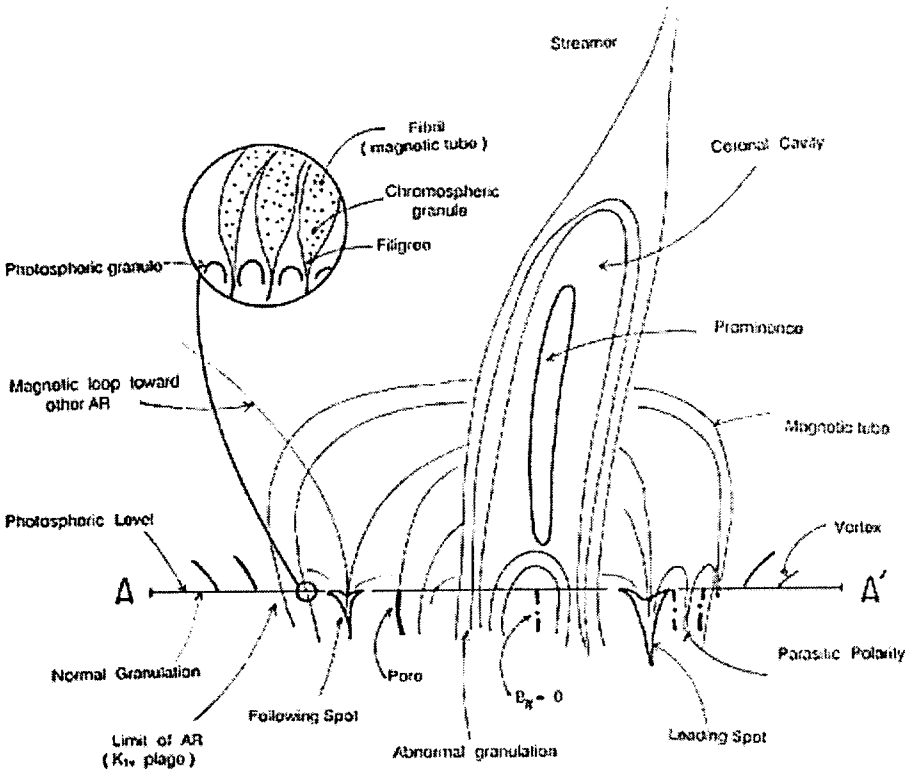


Figure 2.4: The vertical view of an active region. Taken from Cox *et al.* (1991).

If we go higher up in the atmosphere, the magnetic field becomes more dominant. In a fluxtube, pressure equilibrium is valid:

$$\frac{B^2}{8\pi} + P_i = P_s. \tag{2.1}$$

The term $B^2/(8\pi)$ is the magnetic pressure, P_i represents the gas pressure within the fluxtube, P_s is the gas pressure outside the fluxtube.

The external density and pressure decrease rapidly (exponentially) with height, as do the internal density and pressure. To obtain pressure equilibrium (2.1), the field strength must also drop rapidly with height. Since the magnetic flux must be conserved, the fluxtube must therefore expand with height.

In the chromosphere, the temperature outside the fluxtube is lower than the temperature inside. The pressure scale height is proportional to the temperature. So the external

pressure drops more rapidly than the internal pressure and the magnetic fluxtubes expand very rapidly and fill the whole chromosphere.

If you take a cross section in height, a canopy structure (Fig. 2.5) becomes visible. In the photosphere and the lower chromosphere, the fluxtubes are small, while in the higher chromosphere and the corona, space is completely filled with magnetic field. Above this height the field strength decreases only slowly as we go outwards, while the gas pressure keeps dropping exponentially. Hence the magnetic field dominates the force balance and dynamics in the upper chromosphere and corona.

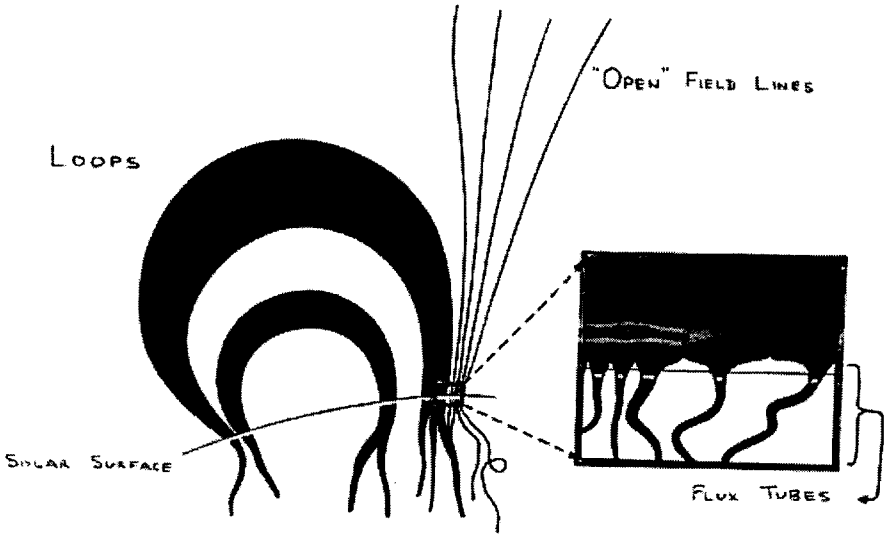


Figure 2.5: A sketch of the canopy structure of the fluxtubes.

One of the structures visible in the Hα image are the *filaments*. In an active region filaments are located above the separation of the opposite polarities in one active region.

Some filaments lie between two active regions, outlining the magnetic neutral line between them. This neutral line is the line where the polarity reverses and the strength of the vertical component of the magnetic field is equal to zero.

We can distinguish between two sorts of filaments, the first filament we described above; the plage filament. This filament is located in an active region. The other filament is the quiescent filament. This filament is not located above an active region, but above the quiet sun.

Because the life time of a plage is often shorter than that of the associated filament, these may exist longer than the plage itself. After the plage disappears, the filament appears as a quiescent filament.

If we see the filament at the edge of the solar disk, we see it as a structure in emission. Then it is called a *prominence*. Prominences (or filaments) lie in the corona of the sun, but the spectrum of a prominence shows a spectrum typical for the chromosphere, *e.g.*, emission lines of hydrogen, helium and various metals. A prominence has a temperature of about 10,000 K, much cooler than the surrounding gas, which has a temperature of 2 million K. So if we take an Hα image, we see the prominence in emission, because the surrounded gas in the corona doesn't emit radiation at that wavelength.

On the other hand, filaments are absorption features. We observe them as dark struc-

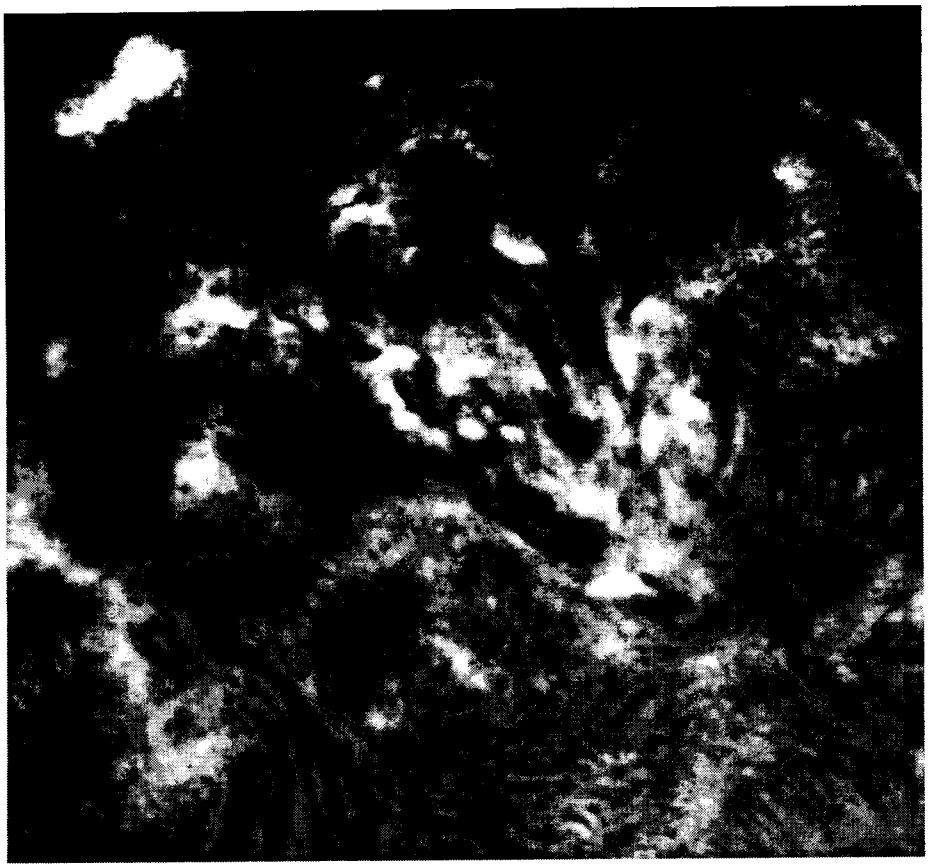


Figure 2.6: An Hα image of an active region of the sun. Taken from `http://www.bbso.njit.edu` (Big Bear Solar Observatory).

tures against the solar disk. The particle density in a filament is very low (10^7 particles cm^{-3}), so the collisional excitation is very small, and negligible. Much of the radiation coming from the filament is produced by scattering of photons from the photosphere. Scattering occurs in all directions, so the photospheric radiation is weakened. A filament thus appears dark.

There are some differences between plage filaments/prominences and quiescent filaments/prominences. When the filaments are seen at the edge of the disk as prominences the plage prominences are more compact and brighter than the quiescent ones. The height where the prominence appears, is for the quiescent prominences somewhat higher than the plage prominence ($1.5 - 2 \cdot 10^4$ km).

The magnetic field of the plage filaments is in the order of 20 Gauss with an azimuthal angle of 20 degrees with respect to the body of the filament. The orientation of the magnetic field is in most cases "potential like". It can be thought of as forming an arcade of loops, which have a depression near their tops. This depression holds the matter in the prominence or filament up agains the gravitational force.

In order to study the bulge of material in the corona, we need to observe in the UV, while we need to observe in X-rays to study the high-energetic, hot plasma. The magnetic field which emerges from the interior of the sun, completely dominates the corona. Under these conditions, the magnetic field is said to be nearly force-free. Meaning that in the force balance, all the non-magnetic forces are small in comparison with the magnetic

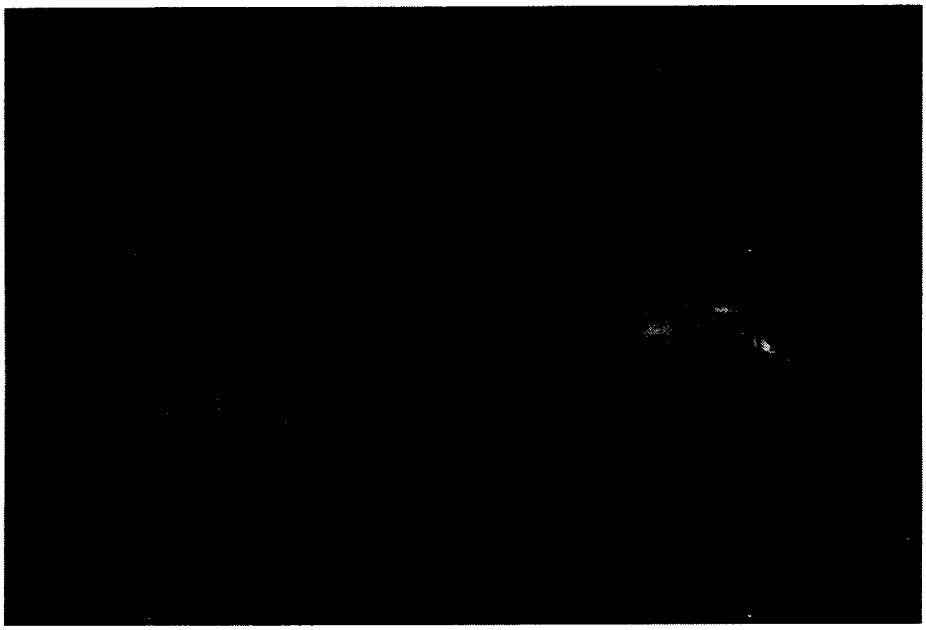


Figure 2.7: An image, taken with the EIT camera from SOHO, in far UV (171 Å) of an active region of the sun, many coronal loops are visible. Taken from `http://sohowww.estec.esa.nl` (SOHO Homepage).

pressure and tension forces, the latter produced by the curvature of the magnetic field.

As mentioned we distinguish two sorts of field lines, the "open" field lines and the "closed" ones. The "closed" field lines lie within the corona, while the "open" field lines extend out into the interplanetary space, and can be closed at the boundary of the heliosphere, or connected with interstellar magnetic field lines.

Due to the fact that field dominates matter, we can observe several loops of matter kept together by the field lines.

The smaller loops are only visible as bright knots in X-rays. The large loops only occur in the zone of active regions, (0 – 60 degrees latitude), because they have their footpoint in active regions. The small magnetic loops can appear at all latitudes and are not limited to the zone where the active regions occurs. Their footpoints are small bipolar magnetic regions.

The largest magnetic loops reach the corona of the Sun and are visible as *coronal loops*. The coronal loops often become unstable and explode into a *coronal mass ejection* (CME). Coronal mass ejections expand away from the Sun at velocities of about 2000 km s^{-1}. They carry up to 10^{13} kg of matter into the interplanetary space.

Another explosive feature occurring in the corona is the *solar flare* (Holman and Benedict 1996b). Flares are detectable from Hα to X-rays. The lifetime of a flare varies from a couple of minutes for the small solar flares to a day for the giant ones. A solar flare occurs when magnetic energy that has built up in the solar atmosphere is suddenly released. The total amount of radiation energy is of the order of 10^{25} J for the largest solar flares. A same amount of energy is used to accelerate mass away from the sun.

Solar flares normally occur in active regions, extending out into the corona of the sun. The corona has a temperature of a few million Kelvin. The temperature of a solar flare is much higher, it varies form 10 or 20 million K for the small ones to 100 million K for the largest flares. CME's were once thought to be initiated by solar flares, but now it is known that most CME's are not associated with solar flares (Holman and Benedict 1996a).

The other type of field lines are the open field lines. The open field lines extend out through the whole interplanetary space. The gas pressure close to the Sun is much higher than the pressure further out in the heliosphere, so there is a pressure gradient. Due to this gradient, there will be a radial force working on the gas in the corona.

Close to the sun, the magnetic field dominates matter, so the solarwind travels along the field lines. However, at a sufficient distance from the Sun ($\approx$ 20 solar radii), the velocity of the solar wind is so large that the kinetic energy becomes larger than the magnetic energy. Beyond this point, the Parker point, the wind dominates the field again.

Coronal holes appear at places where open field lines exist. Around activity minima this happens at the poles of the sun. Coronal holes emit much less radiation in X-ray than other parts of the sun. According to Phillips (1992), coronal holes have a great significance for the connection between the Sun and the interplanetary space. The coronal holes are the sources of high-speed particle streams; the so-called fast solar wind.

Figure 2.8: An image with the LASCO coronograph of SOHO, a beautiful helmet structure is visible. Taken from `http://sohowww.estec.esa.nl` (SOHO Homepage).

In an image with a coronograph (Fig. 2.8), a beautiful *helmet streamer* (called such because of the similarity with a helmet with a spike) is often visible. It is a composition of closed and open field lines. The closed field lines are located in the inner part of the streamer, and the open field lines in the outer part. Gas can stream out along the open field lines.

2.3 The quiet sun

Magnetic field is not confined to active regions, but is present over the whole solar disk (Fig. 2.9). It is visible as a "pepper and salt" structure; small dots on a magnetogram, again with a bipolar character. So, this bipolar structure is not only present in active regions, but also in quiet regions of the sun.

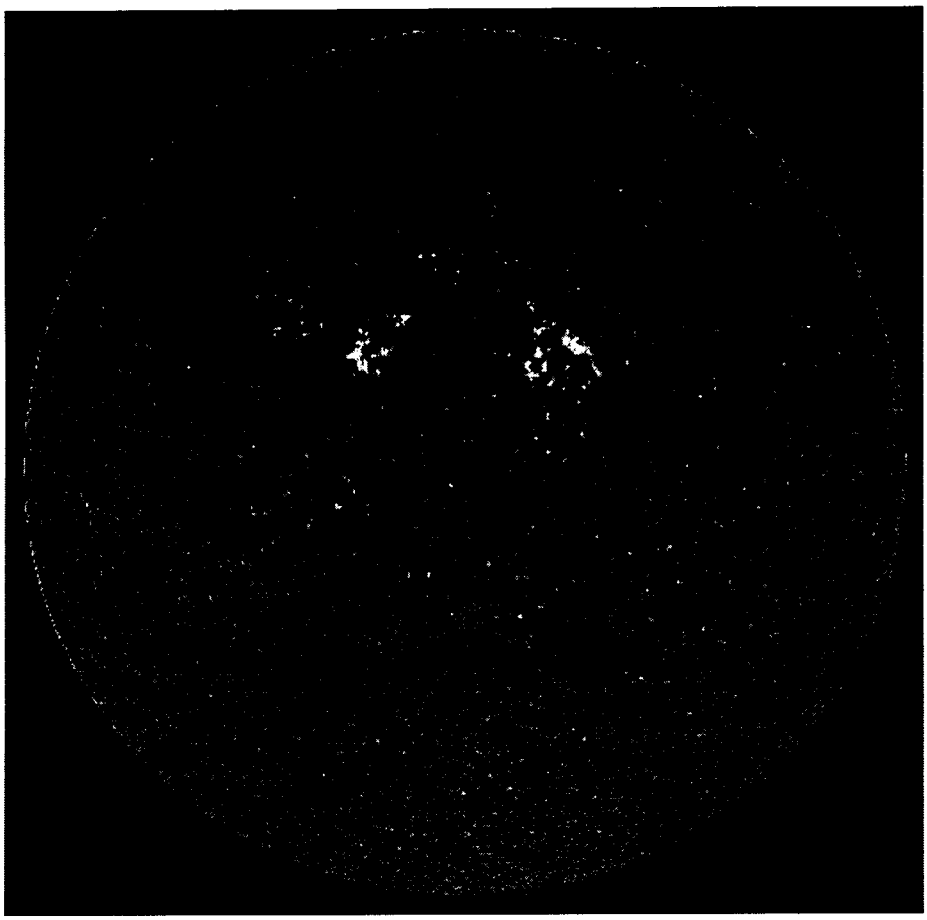


Figure 2.9: A magnetogram with MDI (SOHO). Magnetic field is present over the whole solar disk, not only in the active regions. The magnetogram signal of the quiet Sun is much smaller than the signal of an active region, because the spatially averaged field strenght is much smaller. The magnetic field has a "pepper and salt" structure. Taken from `http://sohowww.estec.esa.nl` (The SOHO Homepage).

2.3.1 Magnetic phenomena in the photosphere

In the quiet regions the field is organized in cell patterns, the so called *supergranulation cells.* The fluxtubes, into which the magnetic field is mainly concentrated, are located at the boundaries of the supergranulation cells.

The supergranulation has its origin in convective motions at intermediate depth in the solar convection zone. The magnetic field is "frozen in" the solar plasma due to its high conductivity. So the magnetic field is carried by the horizontal flows to the boundaries of the cells.

There is an other component of the magnetic field we didn't discuss, it is the so called *weak field.* This field is not organised in cells, and not only distributed in the boundaries of the supergranulation cells, but present over the whole solar disk. The theory is that the weak fields are turbulent fields with a strength of ($B \ll 1$ kGauss), while others are fluxtubes of lower field strength than those in the network (200 – 800 Gauss). The features are so small that we can't resolve them with the present day detectors. We can only measure the average over the area in the detector element, while the filling factor in a detector element is less than unity.

2.3.2 Magnetic phenomena in the chromosphere and the corona

In the Hα image (Fig. 2.10) of an area near the limb of the sun, we observe several short, dark structures, forming what looks like hedgerows around supergranules. These *spicules* mark the location of the magnetic field. At the limb of the sun, these features however appear as bright features. Spicules are jets of plasma, shooting out with velocities of around 10 km s^{-1} from the boundaries of the cells, where the magnetic field is located. In the spicules matter at photospheric temperature is ejected from the chromosphere into the corona.

Figure 2.10: Visualization of the supergranulation cell pattern on the quiet sun. The image is taken in Hα. The spicules and the bright footpoints are visible at the boundaries of the super granulation cells. Taken from Stenflo (1994).

At the lower end of the dark spicules, tiny, bright emission points are visible in Hα.

This are the footpoints of the corresponding magnetic field concentrations. These bright points are located at the boundaries of the supergranulation cells and form the so called *network*. These bright spots are present from the higher photosphere to the chromosphere-corona transition zone. The higher up we go in the atmosphere, the coarser the structures becomes. This indicates that the structures diverge with height. This is what we expect, because the magnetic field lines diverge with height due to the drop of the density. Note, however, that the magnetic field is expected to diverge much more than is seen in brightness images. Obviously the heating of the plasma in the upper chromosphere and transition region is confined mainly to the center of the magnetic concentrations.

Another phenomenon observable in Hα are the quiescent filaments. We already discussed these phenomena in section 2.2.2.

2.4 Conclusion

In this chapter we described a small part of the magnetic features occurring at the solar surface and in the atmosphere of the sun. In the first section we described the morphology of an active region where the field lines break out of the interior of the sun. Sunspots, large dark features, occur at the foot points on the solar surface. The places where the magnetic field emerges are the footpoints of several features in the chromosphere and the corona of the sun, like coronal loops or coronal mass ejections.

If we look at the quiet sun, i.e., areas where no active region is present, the magnetic field is organised in cell patterns, determined by the supergranulation cells. The other magnetic field component of the quiet Sun is the weak field, not organized in cell patterns, but distributed over the whole solar disk. In the chromosphere and the corona quiescent filaments, spicules, the network, active-region plage, X-ray emitting coronal loops, and helmet streamers are observable consequences of the magnetic field.

Chapter 3

Stellar magnetism

Lecturer R.J. Rutten, notes by N.P.A. Zagers

3.1 Introduction

The classification of stars in the H–R diagram is based on their luminosity and effective temperature. It is very well possible to understand main stellar properties such as the existence of a main sequence without taking magnetic fields into account. Apparently, at least on the main sequence, gross stellar properties (*e.g.*, luminosity and effective temperature) are not influenced very much by magnetic fields.

For this reason, there has been no strong "push" to search for stellar magnetic fields. Also, the detection of stellar magnetic fields is far from easy. It took a long time after the discovery of magnetic fields on the Sun before the existence of magnetic fields on stars was proven by direct measurements. Since then, magnetic fields have been detected directly or indirectly on stars throughout the H–R diagram.

A brief overview: T Tauri stars have magnetic fields which often interact with an accretion disk. Late M stars are fully convective and are fast rotators. They show strong magnetic fields. F to M main sequence stars have a convective outer region and possess solar-type fields. All these types of stars have in common that their magnetic fields are generated in a convective layer.

Some stars of type A and B have magnetic fields as well. Their field structure is roughly bipolar. The exact origin of their fields is not clear. Most likely their fields are fossil fields from some earlier stage in their evolution. Other examples of remnant fields are the fields found on white dwarfs and neutron stars. Their fields have been amplified enormously when these stars contracted.

This text presents direct and indirect evidence for stellar magnetic fields and the interpretation of these observations. The emphasis is on solar-like or cool-star magnetism. A separate section is on a relatively new technique, Zeeman Doppler imaging. This promising technique allows mapping the surface distribution of the magnetic field.

This text is mostly based on the books by Böhm-Vitense (1989) and Schatzman and Praderie (1993), and the reviews by Donati (1999) and Landstreet (1992). References to the original literature are found there.

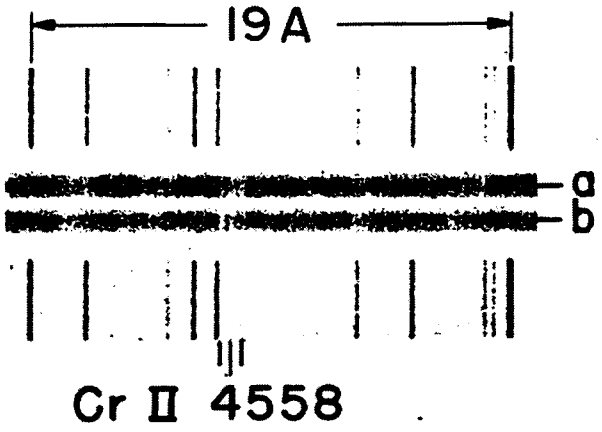


Figure 3.1: An example of a Zeeman triplet in a stellar spectrum, indicated by the three tick marks just above the digit 4 in 4558. The spectrum is a small part of the spectrum of HD 215441. Fig. 14.4 on page 121 of Böhm-Vitense (1989).

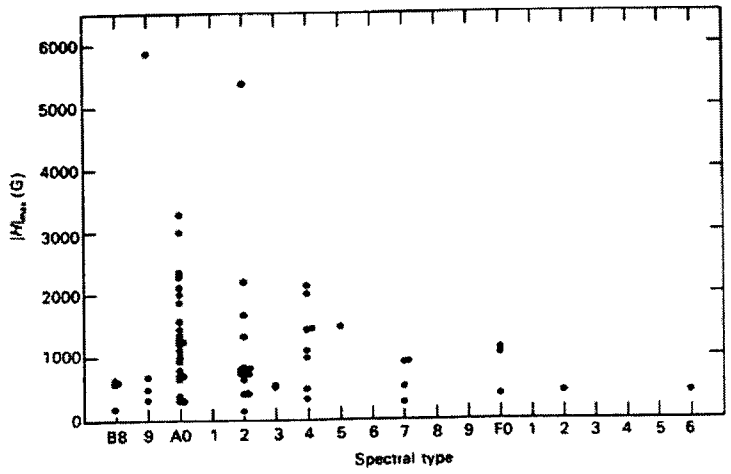


Figure 3.2: The results of measurements of stellar magnetic fields by Babcock. The method of Babcock works best for stars of type A. Fig. (14.9) on page 125 of Böhm-Vitense (1989).

3.2 Direct measurements of stellar magnetic fields

Any direct measurement of a magnetic field is based on the change of shape or polarization of spectral lines by the Zeeman effect (Donati 1999). The first evidence for a stellar magnetic field was found by Babcock in 1947. Fig. 3.1 shows an example of direct evidence for stellar magnetic fields, a Zeeman triplet.

Trying to detect magnetic fields by searching for Zeeman splitting is not very productive. The Zeeman broadening has to compete with the rotational broadening. Only strong fields on stars having a small rotational velocity can be detected. In Fig. 3.1 the rotational velocity is 5 $\mathrm{km\,s^{-1}}$ while the field strength is 3.4 10^4 Gauss. This value for the field strength is large compared to the values found on the Sun. The rotational velocity is small compared to values found for most stars.

In order to detect weaker fields Babcock designed a sensitive magnetograph (*e.g.*, Böhm-Vitense 1989). The light which enters the magnetograph contains both left and right circular polarized light. The magnetograph splits these components and makes two

parallel spectra, one of each component. A comparison of the two spectra shows that lines that are influenced by the Zeeman effect are slightly shifted relative to each other. The broadening of each line due to rotation is no longer spoiling the measurement.

The results of Babcock's survey are presented in Fig. 3.2. Magnetic fields are found from spectral type B8 to F6. Most of the successful measurements are for stars of type Ap, stars of type A with peculiar abundances for the rare earth elements. Babcock found that all the A stars with a magnetic field are Ap stars. All the Ap stars with a small rotational velocity have a magnetic field.

The longitudinal magnetic field found on Ap stars varies with the rotational period in a sinusoidal way. Most Ap stars are also periodic light and spectrum variables as well. The period of these variations is also the rotational period. These effects are all clues leading to the oblique or rigid rotator model. Ap stars are believed to have a roughly bipolar field, not aligned with the axis of rotation.

On the Sun we find a much more complicated field, with no net polarity. For stars with a similar field structure the method of Babcock fails, because the effects of regions with a different polarity cancel out to zero. This is the reason that no fields are found for spectral types later than F6 with this method.

Schatzman and Praderie (1993) explain on their page 327 ff how the magnetic fields of stars of later spectral types are determined. Donati (1999) refers to this method as Stokes I spectroscopy. The method is based on comparing photospheric lines as similar as possible except for the Landé factor g. Lines with large g are more broadened by a magnetic field than lines with small g. With this method fields between 600 and 3000 Gauss have been found on stars between G0V and dM 3.5e.

Besides stars on the main sequence, magnetic fields have been detected on other types of stars. Donati (1999) quotes a detection of a magnetic field on T Tauri stars measured with Stokes I spectroscopy. The review by Landstreet (1992) contains a section on magnetic fields measured on white dwarfs and neutron stars. Table 2 on page 59 of his review gives results for 27 white dwarfs. The field strengths vary between a few MGauss and 350 MGauss. On neutron stars magnetic fields as strong as 10^{13} Gauss are found.

3.3 Indirect evidence for magnetic activity

3.3.1 X-ray emission

X-ray images of the Sun clearly show bright regions. They appear bright because of their high temperature. These regions are definitely not in thermodynamical equilibrium. It is generally accepted that the magnetic field plays an important role in the heating process. Fig. 3.3 shows an H–R diagram containing X-ray emitting stars as detected by the *Einstein* satellite. X-ray emitting stars occur on the entire main sequence. This does not mean that every main sequence star has a corona heated by a magnetic field. Shocks in an otherwise cool wind are also able to produce X-rays. This may occur in the wind of hot O and B stars (*e.g.*, Schatzman and Praderie 1993, Linsky 1985). In cool stars the X-ray emission is caused by a hot corona. This indicates the existence of magnetic fields on such stars.

3.3.2 Ca II H & K and Mg II h & k lines

Fig. 3.4 shows examples of the Ca II H & K lines as observed on the Sun. There is a strong increase of the intensity in the core of the lines in an active region. The same effect is

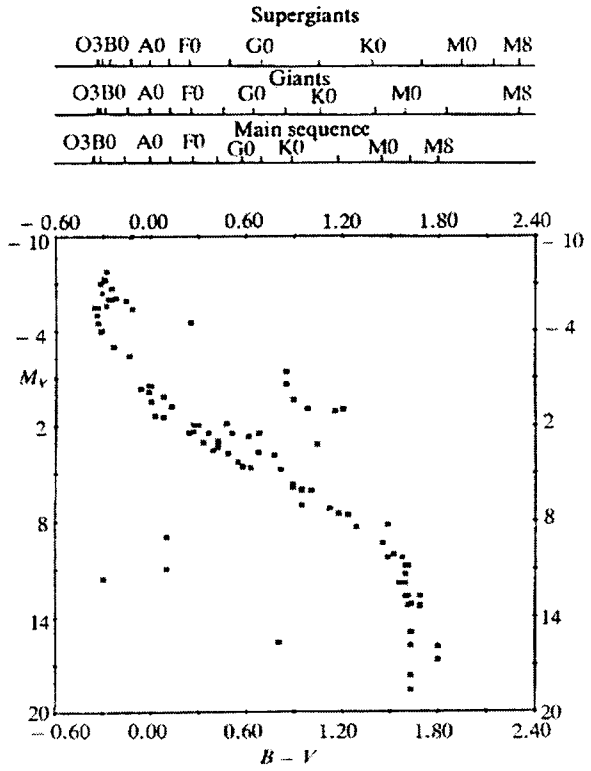


Figure 3.3: H–R diagram for stars detected as soft X-ray sources by *Einstein*. The absolute visual magnitude was determined from the luminosity class or from parallax data. Fig. 1 on page 163 of Vaiana (1981).

observed for the Mg II h & k lines. The fluxes of the H & K emission lines of the disk-integrated Sun vary with amplitudes of the order of 20% through the eleven-year sunspot cycle (*e.g.*, Schatzman and Praderie 1993). This means that the integration of the disks of the stars is not a problem. Effects similar to those on the Sun are expected in the spectra of stars as well.

Fig. 3.5 shows the modulation of the mean H & K flux for a number of stars. The variation on a time scale of days is interpreted as a rotational modulation. The variation on a time scale of years is comparable with the solar eleven-year cycle. On time scales of years some data spread results from faster rotational modulation.

In a project maintained at the Mount Wilson observatory 91 late-type stars have been observed since 1966. After 24 years of observing 85% of these stars showed variation. The typical cycles found are 8 to 12 years. 10-15% are constant in time. Perhaps these stars are in a period comparable to the Maunder minimum.

The bars in Fig. 3.6 represent the variations found in the 91 stars that were part of the Mount Wilson H & K project. For comparison "snapshot" measurements of 396 stars are plotted as well. The overall behavior is an increase of the H & K index with the stellar temperature. In the middle of the data is a gap, known as the Vaughan-Preston gap.

The Ca II H & K and Mg II h & k activity indicators increase as the rotation rate Ω increases. This correlation is clearly visible in Fig. 3.7. The Rossby number N_R is defined as the ratio of the rotation period $1/\Omega$ to the convective turnover time $\tau_c = l/V$, where l is the mixing length and V is a typical convection velocity. The activity is correlated with the Rossby number as well, see Fig. 3.8.

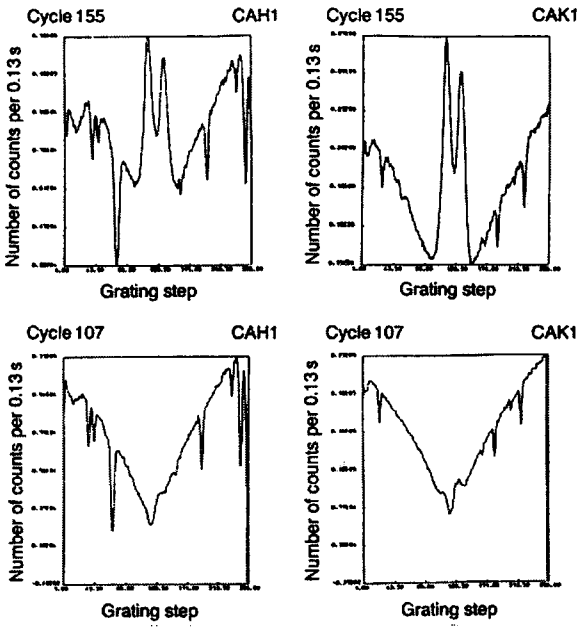


Figure 3.4: The Ca II H & K lines in an active (upper panel) and quiet (lower panel) region of the Sun, of size $2'' \times 10''$. The core of the lines increases in intensity in an active region. Fig. 7.5a on page 314 of Schatzman and Praderie (1993).

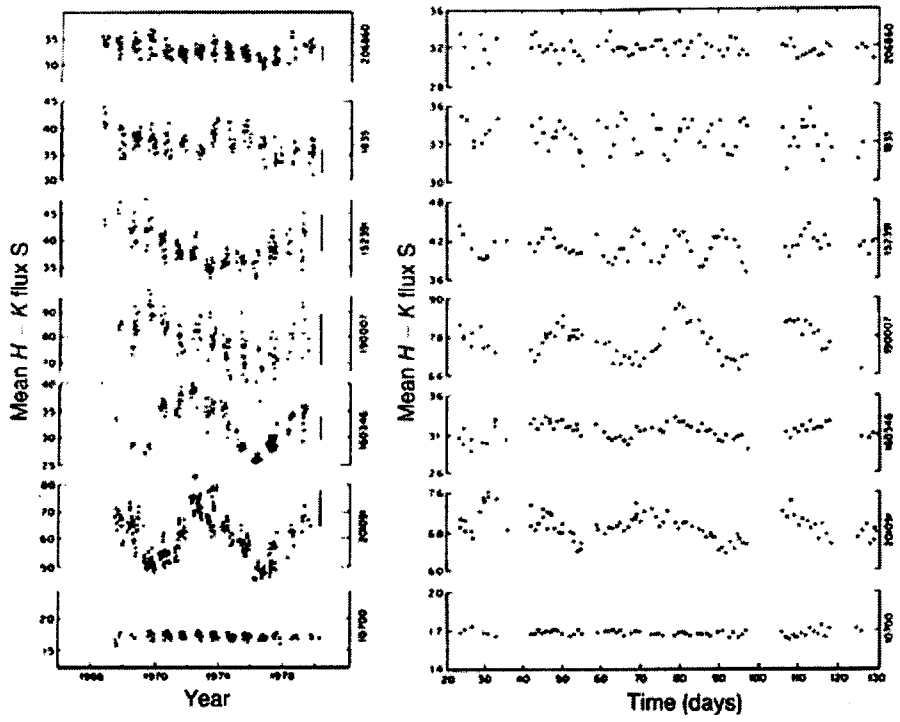


Figure 3.5: Right: rotational modulation of the Ca II H & K lines. Left: Long term variations of the same lines. Fig. 1 on page 279 of Vaughan *et al.* (1981).

The stellar age controls the activity of stars as well. Young stars are in general more active. However, rotation is a more important factor than age. This can be inferred from

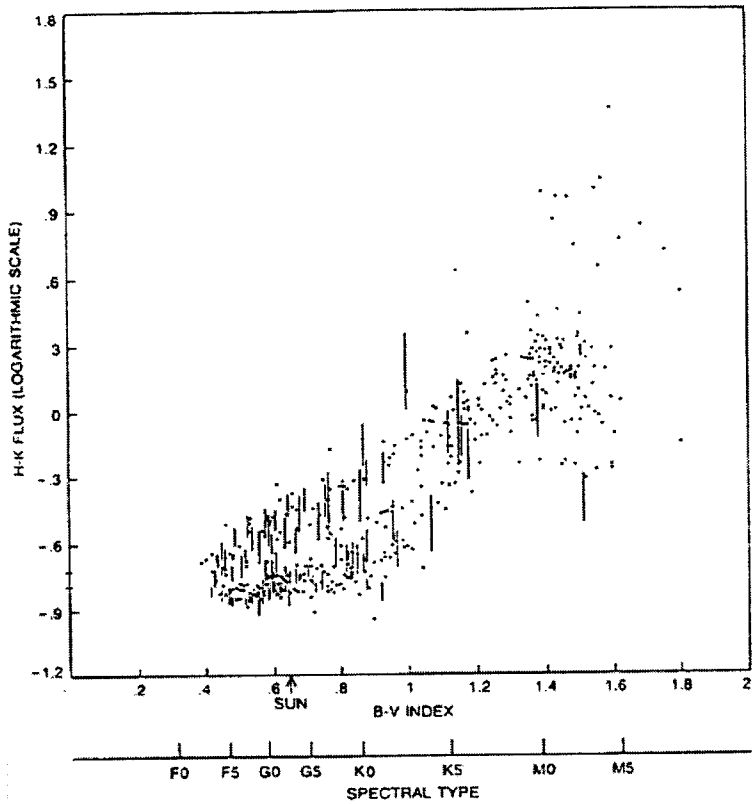


Figure 3.6: This plot shows the variability of the H & K flux vs. the spectral type (the bars). For comparison the instantaneous flux of about 400 stars is plotted as well. Fig. 7.1 on page 105 of Wilson (1994).

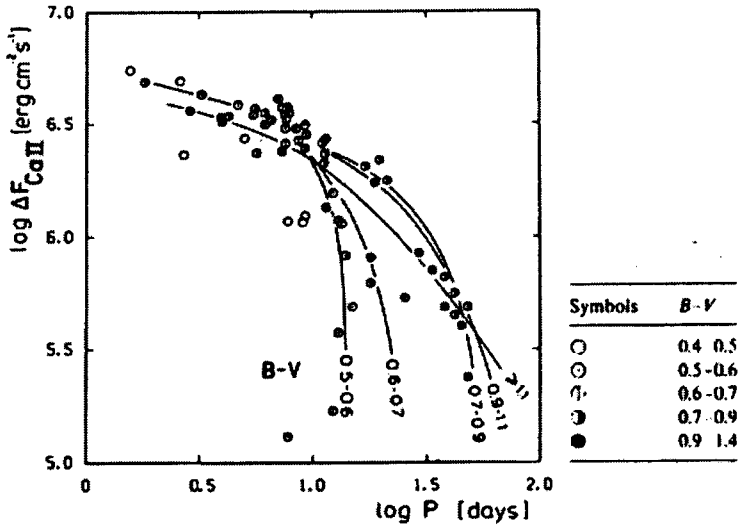


Figure 3.7: The Ca II HK activity versus the stellar rotation period in main sequence stars. Fig. 11.3 of Schrijver and Zwaan (1999).

the extreme activity of RS CVn binaries. These are close binaries containing stars of nearly equal mass, with the hotter component of spectral type F, G, or K and luminosity class V or IV. Due to tidal interaction forces these systems show synchronous rotation.

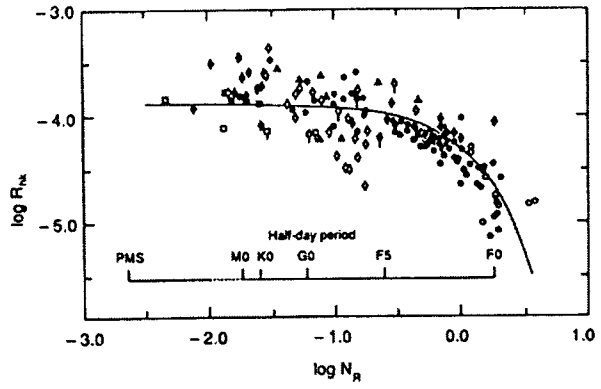


Figure 3.8: The normalized Mg II flux vs. the Rossby number. The squares denote pre main sequence stars, the diamonds represent cluster stars, the circles are late-type stars, and the triangles are active chromosphere stars, *e.g.*, spotted variables and RS CVn stars. Fig. 7.2 on page 109 of Wilson (1994).

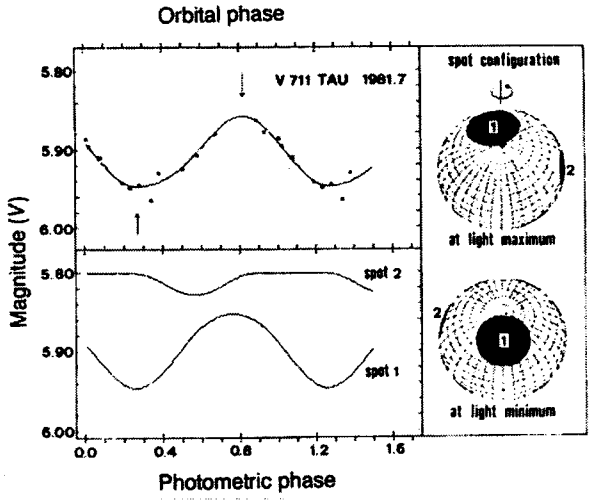


Figure 3.9: Variations in the magnitude of RS CVn star V711 Tau. The variations in the magnitude are modeled by two dark spots. Fig. 10 on page 147 of Rodono *et al.* (1986).

The fast rotation causes the stars to be very active.

3.3.3 Photometric variation

The active Sun shows dark spots and bright active regions as well. The disk coverage of the sunspots is usually small. However, the small effect sunspots have on the luminosity are measurable.

The magnitude variations in some stars have been interpreted as caused by star spots, dark regions in the stars photosphere. An example is given in Fig. 3.9. Stellar spots are usually much larger than sunspots. One reason for this is that in order to be detectable, *i.e.*, causing large enough magnitude variations, the spots have to be large.

If the star spot interpretation is correct, the story is not complete yet. What remains unclear is the exact nature of these large dark regions. It is not evident that they are simply larger versions of sunspots. Perhaps they are of a completely different nature. Their relation to any magnetic field is not clear for the same reason. And finally, there is

a possibility that what we observe as a single large spot is actually a collection of smaller spots.

3.4 Doppler imaging

The principle of Doppler imaging is explained in Fig. 3.10. It shows the effect of a single spot rotating along with the star. The line profile is Doppler broadened by the rotation of the star. If the broadening is too small the technique may not work. The effect of the spot is a small emission "bump" in the spectrum.

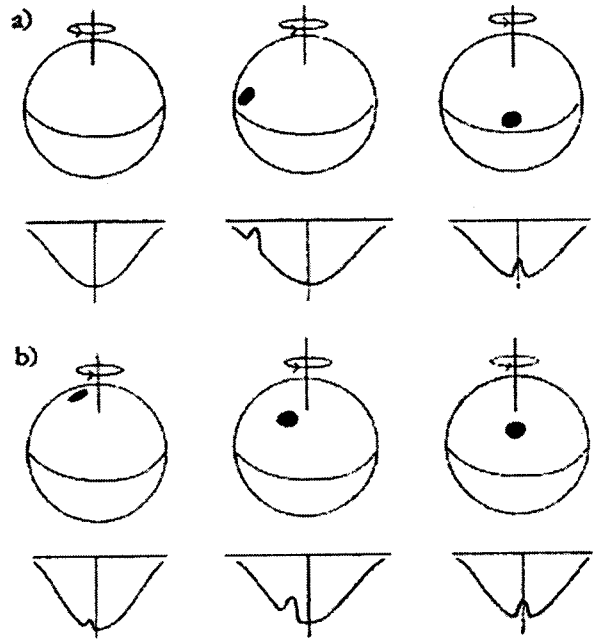


Figure 3.10: The principle of Doppler imaging. Top: an example of a spot near the equator. Bottom: an example of a spot near the pole. Fig. 1 on page 20 of Rice (1996).

First consider a spot near the equator. This spot will be visible only part of the time. When it becomes visible near the edge, the bump appears in the spectrum at the far blue wing of the spectral line. As the spot rotates along, its Doppler shift will change. The bump moves through the line profile to the far red wing. A spot near the pole will be visible for a longer time, perhaps even permanently. The emission bump caused by a polar spot is less blue- or red-shifted, *i.e.*, never reaches the far wing of the line profile. A permanently visible spot moves in both directions.

The longitude of the spot follows from the time of passage through the center of the line profile. The latitude can be inferred from the way the bump travels through the profile. When there are more spots of different sizes the modification of the spectra gets more complicated. The essential part of the reconstruction method is to find the image of the star which reproduces the observed spectra best.

This means that some sort of error function, indicating whether the reconstructed image is "good" or "bad" must be minimized. Using the correct error function and convergence criteria is very important. To gain more confidence in the reconstruction algorithm, the algorithm is tested on a fictitious star before real data is used.

First a fictitious star, in this case the "Vogtstar", is covered with spots as is shown in Fig. 3.11. For the fictitious star a synthetic spectral line flux is calculated. An example of such a spectral line flux is given in Fig. 3.12. The reconstruction algorithm is tested on the synthetic spectrum. Fig. 3.13 is an example of a reconstructed image. Tests like this validate the use of the algorithm on real data. The same test can be performed for binary stars, see Figs. 3.14–3.16. Examples of images reconstructed with the Doppler imaging technique are given in Fig. 3.17 and Fig. 3.18.

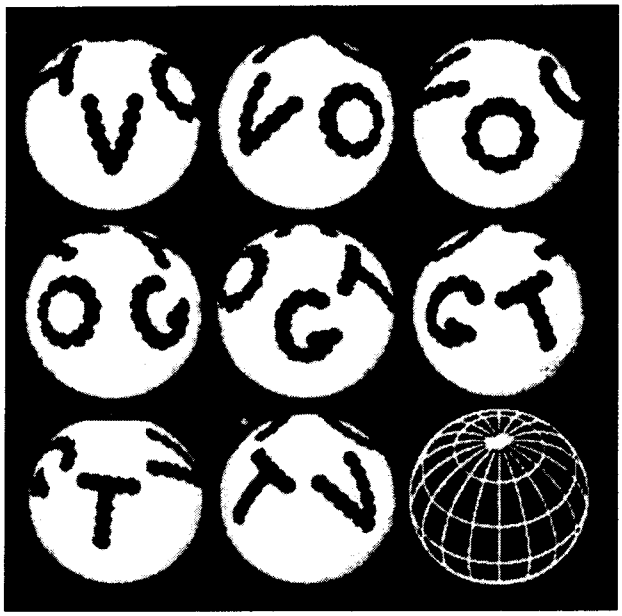


Figure 3.11: The figure shows the "Vogtstar" used as test data for a Doppler imaging reconstruction algorithm. Fig. 4 on page 503 of Vogt *et al.* (1987).

It is also possible to use the deformation of spectral lines by Zeeman broadening as input for similar image reconstruction algorithms. Rather than just using the first moment of the spectral lines, as is done in Stokes I spectroscopy, the full spectral resolution and polarization information of a large number of lines is used. This technique is called Zeeman Doppler imaging. Not only can it detect spots, with this method the surface distribution of the magnetic field can be found as well (Donati 1999, Rice 1996). An example is given in Fig. 3.19.

Both Doppler imaging techniques are capable of showing the detailed surface distribution of star spots. The Zeeman Doppler technique can recover the structure of the magnetic field as well. At the present time, observations have been made for only a small number of stars. Especially the Zeeman Doppler technique has not been fully exploited yet.

However, observations have shown that Zeeman Doppler imaging gives useful results. There are plans for a survey of various types of magnetic stars for a considerable period. This should give information on the evolution of the surface distribution of the magnetic field and differential rotation. Observations of this kind are important for testing stellar dynamo theories.

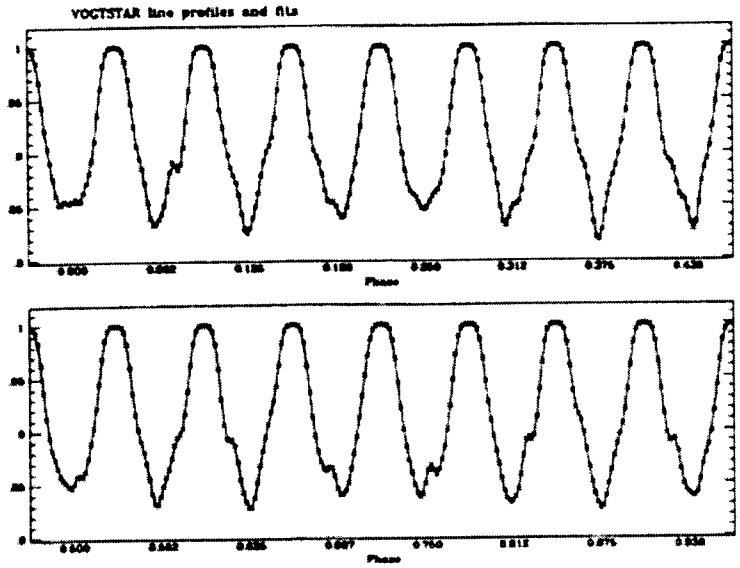


Figure 3.12: The synthetic spectral line flux for the "Vogtstar" test image of Fig. 3.11. The theoretical profiles are shown as the crosses. Ten inversion software's final fit is shown by the solid line. Fig. 5 on page 504 of Vogt *et al.* (1987).

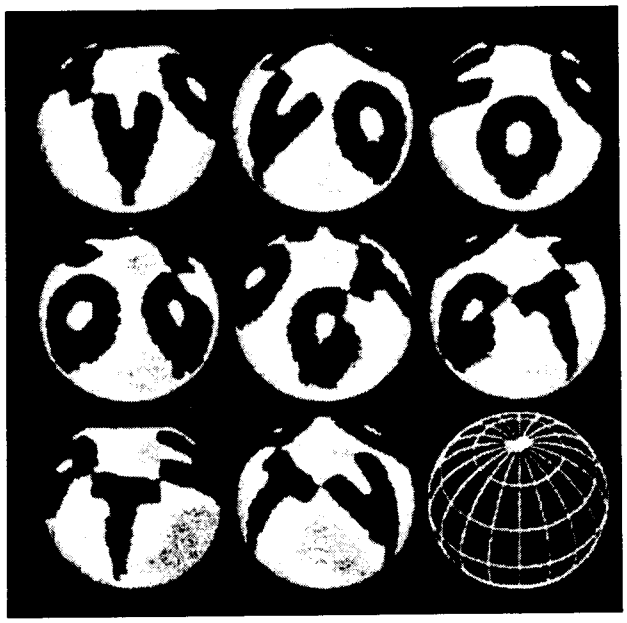


Figure 3.13: The reconstructed image of the "Vogtstar". The image should be compared with the input image in Fig. 3.11. Fig. 6 on page 505 of Vogt *et al.* (1987).

3.5 Cool star activity

Only one star, the Sun, can be observed in great detail. On the other hand we can observe the gross (magnetic activity related) properties of a lot of stars with different depth of the convection zone, effective temperature, gravity, rotation rate, age, etc. The solar paradigm states that the Sun can serve as a guide to understand stellar magnetism in general.

The solar paradigm is of great help to make progress in understanding solar and stellar

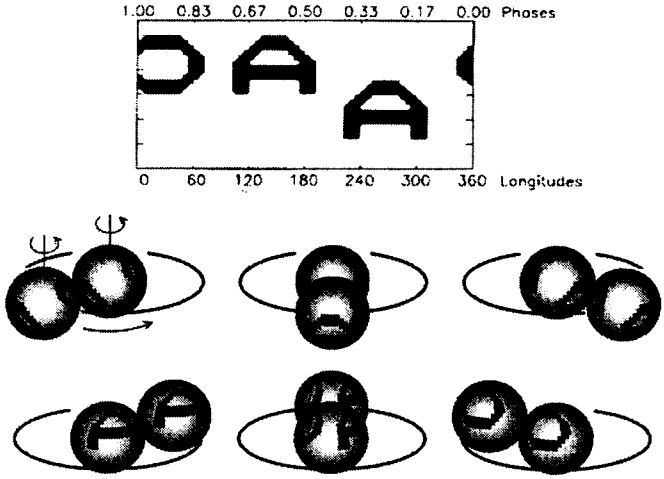


Figure 3.14: An example of a fictitious binary star, used as test data for a Doppler imaging reconstruction algorithm for binary stars. Fig. 1 on page 46 of Piskunov (1996).

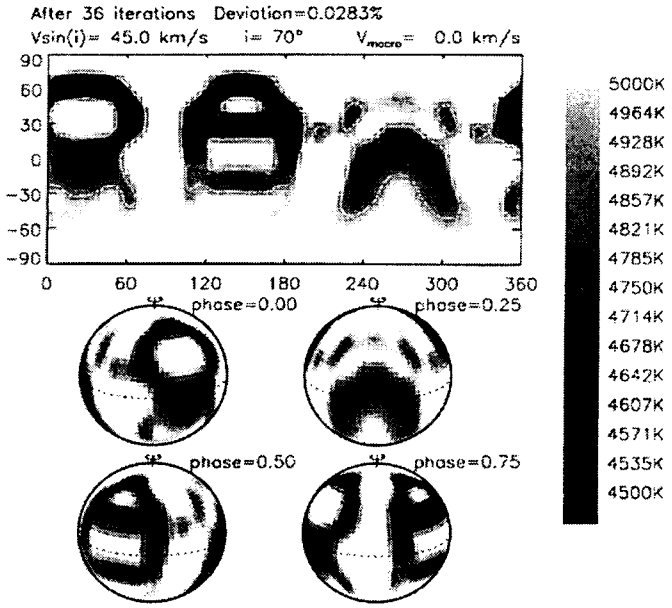


Figure 3.15: The reconstruction of a single star from the fictitious binary star in Fig. 3.14. Fig. 2 on page 47 of Piskunov (1996).

magnetic phenomena. Consider for example the increased Ca II H & K emission in active regions. From solar observations it first became clear that the effect is related to magnetic activity. The next step was to apply it to stellar observations.

The solar–stellar comparison may work the other way as well. The Sun remains just a single star. Its depth of the convection zone, effective temperature, gravity, rotation rate, age, etc. can not be changed. Stellar observations are essential to obtain information on the dependence of magnetic activity on such parameters.

In addition, the magnetic events on the Sun may not provide a good example for similar events on stars in any case. Consider for example the huge spots found from

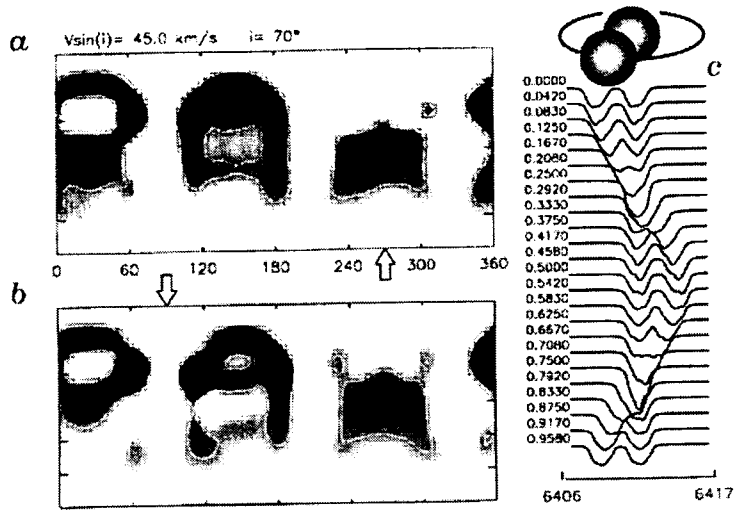


Figure 3.16: The reconstruction of the fictitious binary star in Fig. 3.14. The arrows mark the phase of the eclipse for each component. Fig. 3 on page 48 of Piskunov (1996).

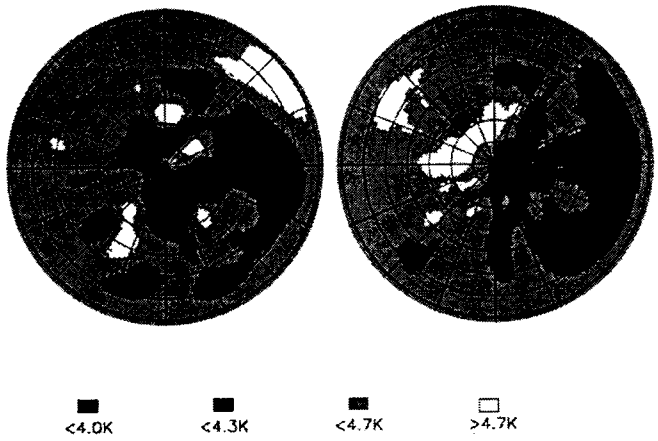


Figure 3.17: Two images of V410 Tau taken at Nov. 1992 (left) and Dec. 1993 (right) obtained with Doppler imaging. The rotational pole is at the center, the lines of latitude and longitude have a spacing of 22.5°. Fig. 2 on page 31 of Rice (1996).

luminosity variations. It seems tricky to assume that these large spots are simply larger versions of sunspots. The Sun being just a single star is perhaps the weakest point of the solar paradigm. Many other stars are members of multiple systems.

Although some critical remarks are possible, the solar paradigm seems to work. Of course it will only work for stars similar to the Sun. How should "similar to the Sun" be defined? Linsky (1985) defines a solar-like star as

> (...) a star which has a turbulent magnetic field sufficiently strong to control the dynamics and energetics of its outer atmospheric regions.

Fig. 3.20 shows schematically which types of stars are definitely or probably solar-like. The most important criterion is that a magnetic field is detectable. Ca II H & K variability on a rotational time scale and X-ray emission are good indicators. Another important criterion is that the magnetic field is generated by a dynamo effect in a convective outer region.

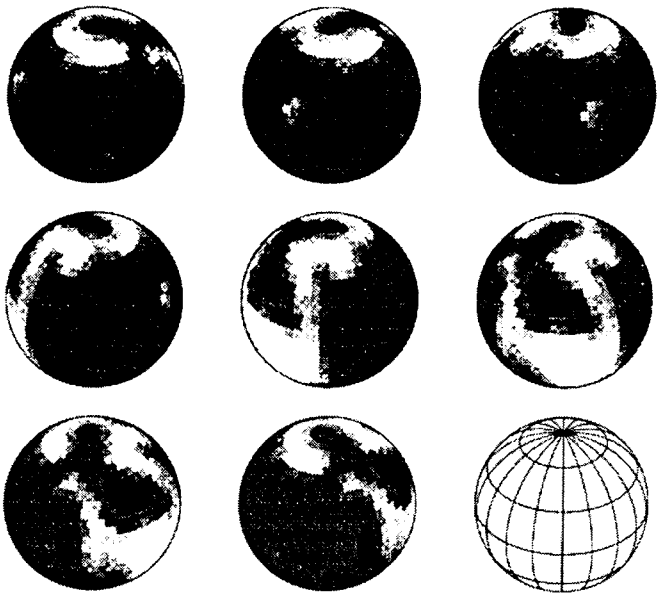


Figure 3.18: The silicon distribution on γ^2 Ari. Maximum $W_\lambda = 550$ mÅ (white), minimum $W_\lambda = 5$ mÅ, average $W_\lambda = 170$ mÅ. Fig. 2 on page 310 of Hatzes (1996).

These conditions are all met for G, K, and M stars, so they are definitely solar-like. Dwarf stars of spectral type A4 – F7 are also good candidates, but their fields have not been detected directly yet.

Other regions of the H–R diagram may possibly contain solar-like stars as well, but the best correspondence to the solar case is found in the cool dwarfs of type G, K, and M. Observations of cool star magnetism will be most useful for a better understanding of solar-like magnetism.

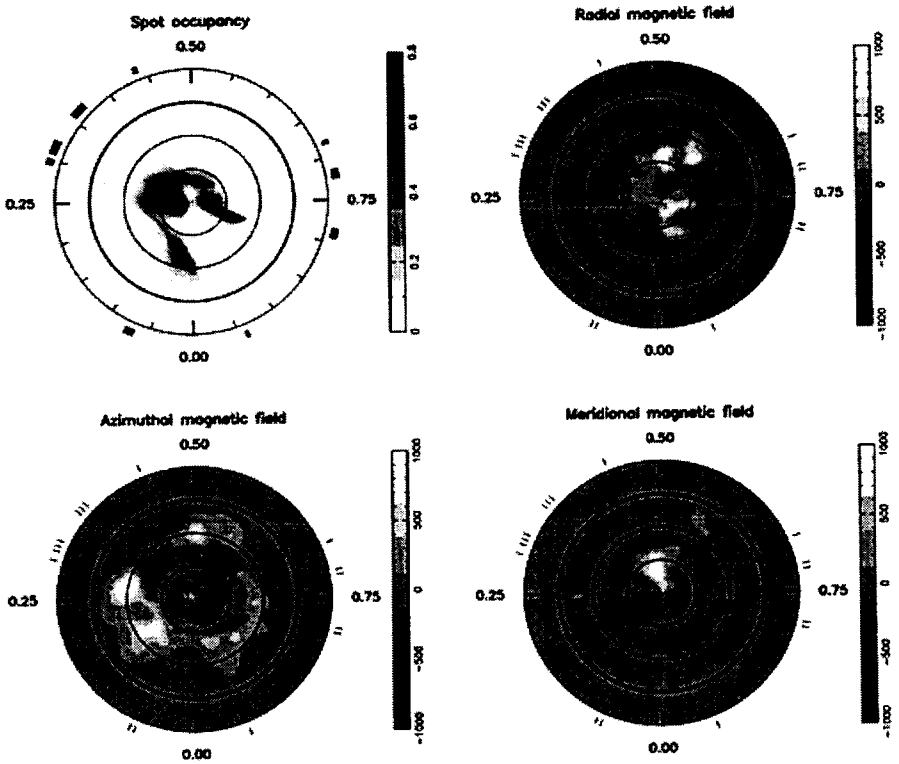


Figure 3.19: An example of an image obtained with Zeeman Doppler imaging. The plot shows a flattened polar view of the brightness (upper left), radial field (upper right), azimuthal field (lower left) and meridional field (lower right) of HR 1099 at epoch 1995.94. The concentric circles indicate parallels with a spacing of 30°, the bold one indicates the equator. The magnetic fields are in Gauss. Fig. 2 on page 32 of Donati (1999).

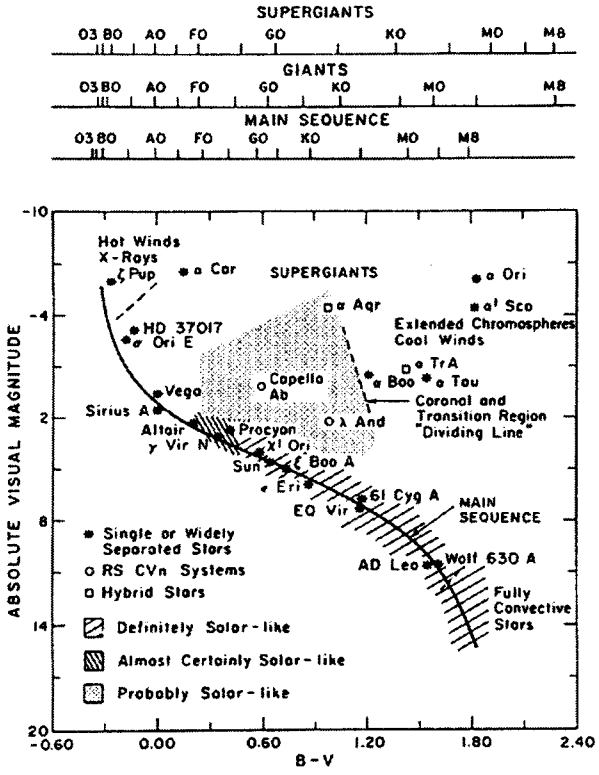


Figure 3.20: An H–R diagram showing schematically which types of stars or definitely or probable solar-like. Fig. 1 on page 341 of Linsky (1985).

Chapter 4

Basic magnetohydrodynamics

All lecturers, notes by R.F.S. Collaris

4.1 Introduction

Many stellar objects are strongly influenced by, if not controlled by, magnetic fields. In order to discuss the magnetic activity in stellar objects (like the Sun) it is necessary to have some knowledge of the properties of magnetic fields in general and of their role in the Universe.

In order to understand the following chapters we begin with a chapter on "Basic magneto hydrodynamics" examining the influence of magnetic fields on stellar objects. It starts by stating "Maxwell's" equations. Next I simplify these equations through some constraints and assumptions resulting in the induction equation. I use this powerful equation to work out the flux conservation and diffusion concepts. Finally I consider magnetic pressure and tension resulting in phenomena like flux tubes and current sheets. Keep in mind that there is an extensive literature on this subject and that this chapter by no means covers all of it. Only the essentials needed for this course are treated here.

4.2 Units

In electromagnetism several systems of units are used. It would clearly be advantageous if a single system could be accepted by all, even though the transition to standard units may be a painful one. The rationalized MKS units are now part of the internationally adopted SI system and in these units Maxwell's equations are simpler. However, because of the simplicity of the defining equations I use the Gaussian cgs units.

In Gaussian cgs e, $\vec{E}$, $\vec{j}$ and $\vec{D}$ are measured in esu, while $\vec{B}$ and $\vec{H}$ are measured in emu, respectively electrostatic units and electromagnetic units. Other units are defined in terms of these by the equations in Fig. 4.1. Note that ϵ and μ are dimensionless and are taken to be unity in plasma physics so that $\vec{H}$ equals $\vec{B}$ and $\vec{D}$ equals $\vec{E}$. A useful conversion table is found on page 618 of Jackson (1962).

Quantity	Defining equation	Unit
length	–	1 cm
mass	–	1 g
time	–	1 s
force	$F = m\ddot{x}$	1 dyne = 1 g cm s^{-2}
charge	$F = \frac{e_1 e_2}{r^2}$	1 statcoulomb = 1 cm $(\text{dyne})^{1/2}$
electric field	$E = \frac{e}{r^2}$	1 statvolt cm^{-1} = 1 $(\text{dyne})^{1/2}$ cm^{-1}
electric displacement	$\mathbf{D} = \varepsilon\mathbf{E}$	1 statvolt cm^{-1}
current	$\frac{F}{l} = \frac{2I_1 I_2}{r}$	1 abamp = 1 $(\text{dyne})^{1/2}$
magnetic induction	$B = \frac{2I}{r}$	1 gauss = 1 abamp cm^{-1}
magnetic field	$\mathbf{H} = \mathbf{B}/\mu$	1 oersted = 1 gauss

Figure 4.1: Definition of Gaussian units. From Priest (1982).

4.3 Equations

4.3.1 Maxwell's equations

Although these equations are called Maxwell's equations, Maxwell "only" added some terms to the already existing equations resulting in:

$$\frac{\partial \vec{B}}{\partial t} = -c\nabla \times \vec{E} \tag{4.1}$$

$$\nabla \times \vec{B} = \frac{4\pi}{c}\vec{j} - \frac{1}{c}\frac{\partial \vec{E}}{\partial t} \tag{4.2}$$

$$\nabla \cdot \vec{E} = 4\pi\rho \tag{4.3}$$

in Gaussian units. The first equation is actually Faraday's equation the second is largely Ampére's equation and the third is Poisson's equation. As a constraint to these equations I use:

$$\nabla \cdot \vec{B} = 0. \tag{4.4}$$

This tells us that magnetic field lines should always connect head to tail except in the case of infinity. This "simple" formula states that there can not be any magnetic mono poles. We are not interested in electromagnetic waves so equation (4.2) becomes:

$$\nabla \times \vec{B} = \frac{4\pi}{c}\vec{j}. \tag{4.5}$$

We could also approximate equation (4.2) by using the typical scale length: $E/L \sim B/(tc)$ (Eq. 4.1) and the fact that $v \ll c$, resulting in:

$$\frac{1}{c}\frac{\partial \vec{E}}{\partial t} = \frac{\vec{E}}{tc} \sim \frac{\vec{B}L}{(tc)^2} = \frac{\vec{B}v^2}{Lc^2} \ll 1. \tag{4.6}$$

Equation (4.5) implies that $\nabla \cdot \vec{j} = 0$ since the divergence of the rotation equals zero, which means that the electrical currents are closed loops and no charge will accumulate. Thus, equation (4.3) becomes:

$$\nabla \cdot \vec{E} = 0. \tag{4.7}$$

Note that the ratio of electrostatic to magnetic density is:

$$\frac{E^2}{B^2} = \frac{E^2}{B^2} = \frac{L^2}{(tc)^2} = \frac{v^2}{c^2} \ll 1. \tag{4.8}$$

In order to obtain the induction equation we should add Ohm's equation:

$$\vec{j} = \sigma(\vec{E})' = \sigma(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}). \tag{4.9}$$

In this equation we used that $v \ll c$ and therefore $\gamma = 1$ where γ is the relativistic factor, *i.e.,* we only consider the non-relativistic (and non-quantum) case. When we also use $\sigma = \infty$ valid for ideal plasma we again find $E \sim (v/c)B$. This is sometimes referred to as "the infinite conductivity limit", though "the large length scale limit" would have been a better name. The current $\vec{j}$ may be found through equation (4.5) or (4.9) but because $v \sim L/t$ is so large $\vec{j}$ is negligible in equation (4.9). Here σ is the conductivity and $\eta \equiv c^2/(4\pi\sigma)$ is the magnetic diffusivity, not to be confused with the electric resistivity $1/\sigma$. Note that $\vec{v}/c \times \vec{B}$ is the induced electric field. Combining the equations (4.1), (4.9) and (4.5) results in:

$$\frac{\partial \vec{B}}{\partial t} = -c\nabla \times \vec{E} = \nabla \times (\vec{v} \times \vec{B}) - \nabla \times (\frac{c^2}{4\pi\sigma} \nabla \times \vec{B}), \tag{4.10}$$

and using:

$$\nabla \times \nabla \times \vec{B} = \nabla(\nabla \cdot \vec{B}) - (\nabla \cdot \nabla)\vec{B} = -\nabla^2 \vec{B} \tag{4.11}$$

gives:

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times \vec{v} \times \vec{B} + \eta \nabla^2 \vec{B} \tag{4.12}$$

which is called the induction equation and expresses the change of the magnetic field in "time". It thus defines the behavior of the magnetic field once $\vec{v}$ is known. The first term on the right is caused by convection, carrying the magnetic lines of force bodily with the fluid. The second by diffusion of the magnetic field. We may observe the effect of $\vec{v}$ on $\vec{B}$ and neglect the effect of $\vec{B}$ on $\vec{v}$ whenever $\beta \gg 1$ (defined in Sect. 4.4.2) and pressure gradients dominate the Lorentz force (see the next sections). This equation lies at the heart of all MHD phenomena. One (simple) example is the disk dynamo which will be treated in later chapters.

4.4 Stellar aspects

Plasma is the most common state of matter in the universe and plasma physics is the tool to treat the dynamics of astrophysical plasmas like solar magnetic activities, solar wind, accretion disks etc.. A great argument to use plasma physics on astronomical elements may be found in Goedbloed and Poedts (1997).

4.4.1 Debye length and charge neutrality

A plasma is a gas made out of charged and neutral particles. The Coulomb interaction between charged particles is cut off by a "cloud" of oppositely charged particles around them. In cold plasmas without any thermal movement this "cloud" would completely neutralise the Coulomb interactions. However we are working with "normal" plasmas which do have thermal energy. Thus the thermal movements of the particles may cause charge fluctuations on the Debije scale. De Debije length is defined by:

$$\lambda_D = \left(\frac{kT_e}{4\pi n e^2}\right)^{1/2}, \tag{4.13}$$

where T_e is the electron temperature in Kelvin and n the electron density per cm^3. A plasma is an ionized gas with a typical scale length L such that $\lambda_D \ll L$. Proper shielding by a cloud of particles occurs only if there are enough particles in the cloud:

$$N_D = \frac{4}{3}\pi n(\lambda_D)^3 = 1380\ T^{3/2}/n^{1/2}. \tag{4.14}$$

These conditions are easily satisfied in astrophysical environments. We may estimate λ_D in the photosphere, by choosing T_e to be 6000 K and n to be 10^{-23} particles m^{-3}, to be 0.16 cm. The typical sizes of solar phenomena range from km to Mm (e.g. a granule is 1 Mm and a sunspot is 30 Mm).

4.4.2 Reynolds number and plasma beta

We are interested in plasmas in which L is the length scale of the system. Thus looking at equation (4.12) $|\nabla \times \vec{v} \times \vec{B}| \sim vB/L$ and $|\eta \nabla^2 \vec{B}| \sim \eta B/L^2$. The Reynolds number defined by dividing the convection part through the diffusion part is:

$$R_m = \frac{vL}{\eta} \tag{4.15}$$

Obviously implying that whenever $R_m \gg 1$ convection dominates and coupling between plasma flow and magnetic field is strong, but, when $R_m \ll 1$ diffusion dominates and coupling is weak. This magnetic Reynolds number is very large in astrophysical circumstances; *e.g.*, for umbrae $L = R = 10^4$ km and $v_\perp = 100$ m s^{-1}, $\eta_{\text{photosphere}} = 10^8$ cm^2 s^{-1} gives $R_m = 10^5 \gg 1$. So all visible changes in the solar atmosphere are caused by moving material.

The plasma beta (β) is defined by the plasma pressure divided by the magnetic pressure:

$$\beta = \frac{8\pi p}{B^2} \tag{4.16}$$

In the photosphere $\beta \gg 1$ but in the corona $\beta \ll 1$ (the so called low-β plasma). For more parameters see Priest (1982).

4.5 High/low conductivity

4.5.1 High conductivity and flux conservation

When assuming an ideal plasma $\sigma = \infty$ ($\eta = 0$) and equation (4.12) becomes:

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times \vec{v} \times \vec{B} \tag{4.17}$$

Now I define the flux as the number of field lines (field strength) through a surface element s of a moving contour L:

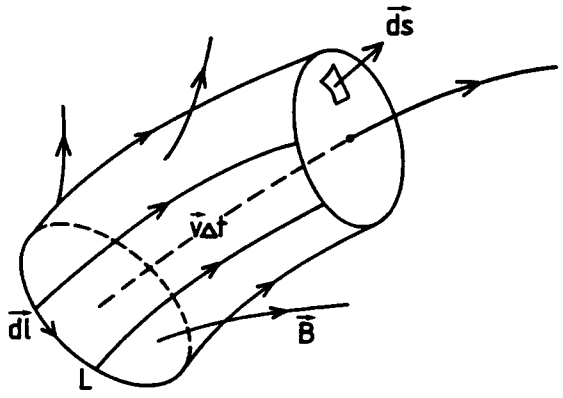


Figure 4.2: Flux conservation through a contour. From Kuperus (1998).

$$\Phi = \int_S \vec{B} \cdot \mathrm{d}\vec{s} \tag{4.18}$$

Here $\mathrm{d}\vec{s}$ is an infinitesimal surface element. The flux changes because of two reasons. In the first place the local magnetic field $\vec{B}$ changes in time:

$$\Delta\Phi_1 = \int_S \frac{\partial \vec{B}}{\partial t} \cdot \mathrm{d}\vec{s} \cdot \Delta t \tag{4.19}$$

and secondly the contour changes shape thus altering the flux:

$$\Delta\Phi_2 = \int_S \vec{B} \cdot \vec{v} \times \mathrm{d}\vec{l} \cdot \Delta t \tag{4.20}$$

Here $\mathrm{d}\vec{l}$ is an infinitesimal line element. I use the vector identities: $\vec{B} \cdot \vec{v} \times d\vec{l} = -(\vec{v} \times \vec{B}) \cdot d\vec{l}$ and $\int_L \vec{v} \times \vec{B} \cdot d\vec{l} = \int_S \nabla \times (\vec{v} \times \vec{B}) \cdot d\vec{s}$ (Stokes theorem) to find:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = \int_S \left(\frac{\partial \vec{B}}{\partial t} - \nabla \times (\vec{v} \times \vec{B}) \right) \cdot \mathrm{d}\vec{s} \tag{4.21}$$

Together with the assumption of an ideal plasma (Eq. (4.17)) we see:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}t} = 0 \tag{4.22}$$

Thus I conclude that the flux through every contour L will stay the same. This is a very important fact for it implies that the field lines are "frozen" into the plasma. Plasma can move freely along field lines, but, in motion perpendicular to them, either the field lines are dragged with the plasma or the field lines push the plasma. The significance of this fact will soon become clear. Observing that:

$$\frac{\mathrm{d}\vec{B}}{\mathrm{d}t} = \frac{\partial \vec{B}}{\partial t} + (\vec{v} \cdot \nabla)\vec{B} \qquad \text{and} \qquad \frac{\partial \vec{B}}{\partial t} = \nabla \times \vec{v} \times \vec{B} \tag{4.23}$$

We find

$$\frac{\mathrm{d}\vec{B}}{\mathrm{d}t} = \nabla \times \vec{v} \times \vec{B} + (\vec{v} \cdot \nabla)\vec{B} = (\vec{B} \cdot \nabla)\vec{v} - \vec{B}(\nabla \cdot \vec{v}) \tag{4.24}$$

Now $\nabla \cdot \vec{v}$ is called the compression or expansion term and so the second term on the right side of equation (4.24) is the change In $\vec{B}$ through compression or expansion. The first term on the same side includes the change of $\vec{B}$ through stretching the field lines by "forcing" them off the surface.

Considering an ideal plasma with frozen-in magnetic fields, we see that by keeping the flux constant, differential movements like compression and shear (forcing the magnetic field lines off the surface) changes the magnetic field.

First I consider shear. We may visualize this by looking at a homogeneous magnetic field (in say the $\vec{z}$-direction) and a velocity field perpendicular to $\vec{B}$ with a gradient along $\vec{B}$ see figure (4.3). We notice that $\vec{B}$ which is indicated by the number of field lines increases. A simple mathematic solution to the given problem: from equation (4.17) with

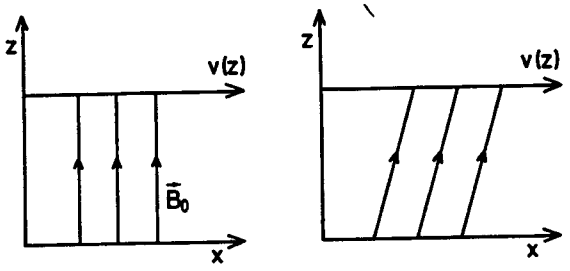


Figure 4.3: A simple example of shear motion. From Kuperus (1998).

$\vec{v}(z) = (v_x(z), 0, 0)$ and $\vec{B} = (B_x, 0, B_z)$ we find

$$\frac{\partial B_x}{\partial t} = B_z \frac{\partial v_x}{\partial z} \qquad \text{and} \qquad \frac{\partial\, B_z}{\partial t} = 0. \tag{4.25}$$

Thus $B_z = B_0$. Take $v(z) = \alpha z$:

$$B_x(t) = \alpha B_0 t \tag{4.26}$$

resulting in:

$$B^2(t) = (B_0)^2 + 0 + (B_x)^2 = (B_0)^2(1 + \alpha^2 t^2). \tag{4.27}$$

This method to increase the field strength is found in systems where turbulence plays an important role and also in systems with differential rotation. In the sun differential rotation causes the poloidal magnetic fields to become (and strengthen the) toroidal magnetic fields, thus creating two toroidal tubes. I now examine compression through a simple example. With $\vec{v}(x) = (v_x(x), 0, 0)$ and $\vec{B} = (0, 0, B_z)$ we find

$$\frac{\mathrm{d}B_z}{\mathrm{d}t} = -B_z \frac{\partial v_x}{\partial x}. \tag{4.28}$$

Now using the continuity equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\rho \frac{\partial v_x}{\partial x} \tag{4.29}$$

We find:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{B}{\rho}\right) = 0, \tag{4.30}$$

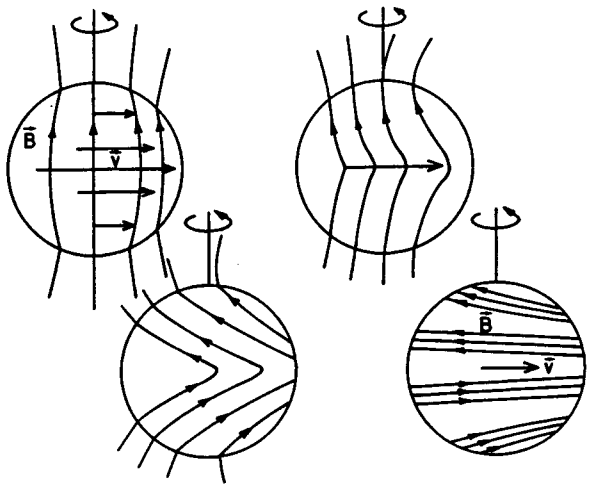


Figure 4.4: Increase of the field by shear motion on a stellar object. From Kuperus (1998).

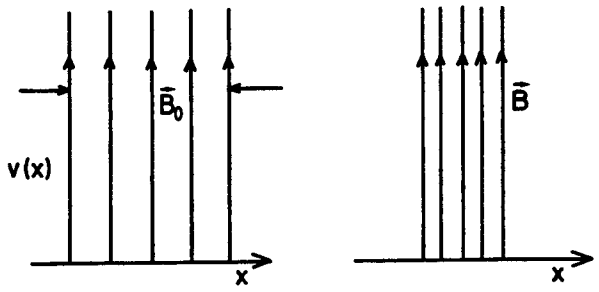


Figure 4.5: A simple example of compression. From Kuperus (1998).

where I used the condition $\vec{v} \cdot \vec{B} = 0$. Thus B/ρ is constant, meaning that whenever matter diverges or converges the field strength does so too. Keep in mind that this only holds if a tube of constant length is considered (see Sect. 4.8.1). A more general example is illustrated in Priest (1982) and Kuperus (1998).

4.5.2 Low conductivity and diffusion

I now look at the static case in which $\vec{v} = 0$ and $\sigma \neq \infty$; thus the plasma is not ideal, or in other words $R_m \ll 1$. Equation (4.12) becomes:

$$\frac{\partial \vec{B}}{\partial t} = \eta \nabla^2 \vec{B}. \tag{4.31}$$

This is called the diffusion equation. The characteristic diffusion time on the characteristic length scale L may be estimated by:

$$\frac{B}{\tau_D} \sim \frac{\eta B}{L^2} \qquad \text{thus} \qquad \tau_D = \frac{4\pi\sigma}{c^2} L^2 \tag{4.32}$$

In astrophysical circumstances τ_D is very large due to the fact that both L and σ are large. One reason we cannot reproduce these conditions in the laboratory is the value for L.

4.5.2.1 The infinitesimally thin current sheet

The smaller the length scale the faster the field diffuses away. As an example I consider the infinitesimally thin current sheet:

$$\vec{B} = B(x,t)\vec{y} \tag{4.33}$$

with

$$B(x,0) = +B_0,\ x > 0\ B(x,0) = -B_0,\ x < 0 \tag{4.34}$$

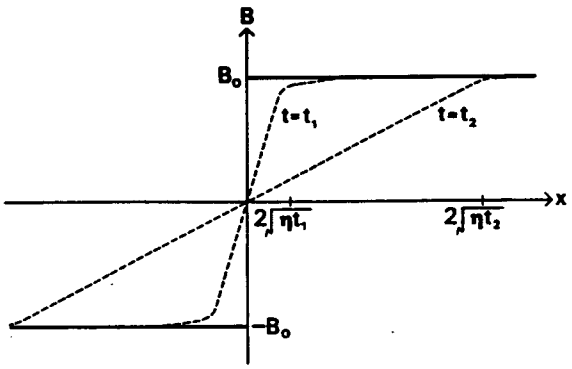


Figure 4.6: A diffusing current sheet: The variation with time of the magnetic field strength. From Priest (1982).

Suppose the plasma in this initially infinitesimally thin current sheet remains at rest and the magnetic field remains unidirectional. The equation (4.31) becomes:

$$\frac{\partial \vec{B}}{\partial t} = \eta \frac{\partial^2 \vec{B}}{\partial x^2} \tag{4.35}$$

which has solution $B(x,t) = B_0\ \mathrm{erf}(\xi)$ where $\xi = x/(4\eta t)^2$ and $\mathrm{erf}(\xi) = 2/\sqrt{\pi}\int_0^\xi e^{-u^2}du$. In time we find the profiles shown in Figure 4.7.

Here we see $|B(x,t)| \sim B_0 x(\pi\eta t)^{-1/2}$ for $|x| \ll \sqrt{4\eta t}$ and $|B(x,t)| \sim B_0$ for $|x| \gg \sqrt{4\eta t}$. Here the profile is undisturbed. The region to which the current is concentrated is known as the current sheet. A rough estimate for its width is $4\sqrt{\eta t}$, which increases with time at a continuously decreasing rate. We can see that the field strength at large distances remains constant, whereas that at small distances decreases monotonically. So I conclude that the field lines do not move outward; rather, the field diffuses away or is annihilated. Magnetic energy is converted into heat by ohmic dissipation, Although the current density $j_z = c\mathrm{d}B/4\pi\mathrm{d}x$ changes in time, the total current in the sheet $J = \int_\infty^\infty j_z\,\mathrm{d}x = 2B_0c/(4\pi)$ remains constant. Of course this is a crude example since we cannot neglect the convection term in the induction equation due to the inward pointing magnetic pressure.

4.6 Reconnection of magnetic field lines

Imagine two oppositely directed fields brought together by fluid motions. The easiest way to look at it is discussed in Tayler (1997): As the field lines approach, the gradient of the

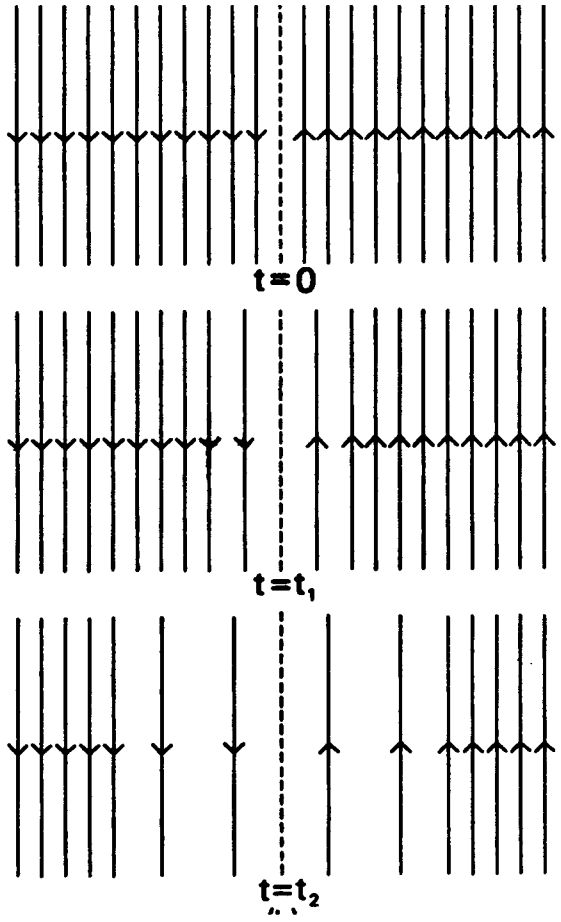


Figure 4.7: A diffusing current sheet: a sketch of the magnetic field lines at three times. From Priest (1982).

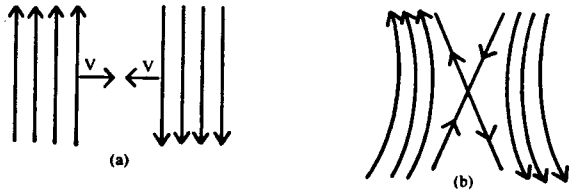


Figure 4.8: A process of steady-state reconnection, showing successive field lines approaching and reconnection. From Tayler (1997).

magnetic field will become extremely strong in the region between them, where there will be what is known as a neutral point; a point where the magnetic field vanishes. As the field gradient increases the resistive dissipation produced by the term $\eta\nabla^2\vec{B}$ will become important even if the conductivity σ is very large. The effect of dissipation is to cause a reconnection of the magnetic field lines which then move apart in a direction perpendicular to their original motion. As a consequence magnetic energy has been released in the neighbourhood of the reconnection point.

Other physical causes of reconnection:

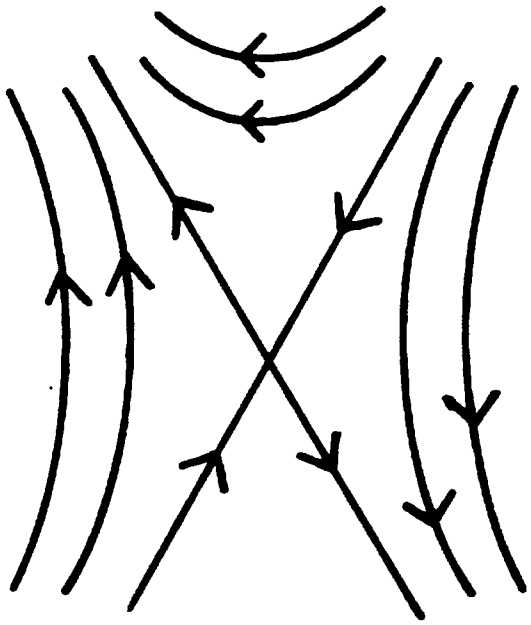


Figure 4.9: The final stage of reconnection resulting in an configuration with an X-type neutral point. From Tayler (1997).

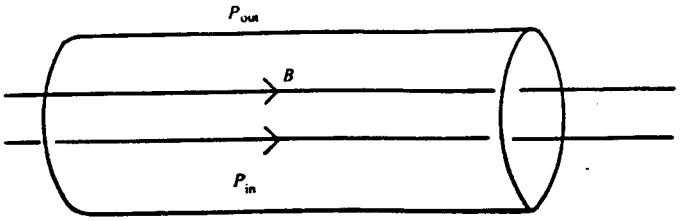


Figure 4.10: Magnetic buoyancy. A tube of magnetic flux with magnetic induction B has an internal gas pressure $p_{\rm in}$ which must be smaller than the external pressure $p_{\rm out}$. Because temperature is uniform, the density in the flux tube is lower than that of its surroundings and it rises. From Tayler (1997).

- a resistive instability such as the tearing mode (see Sect. 4.8.2) in a current sheet or throughout a sheared structure may cause reconnection;
- sudden enhancement of the resistivity at some location may cause reconnection.

4.7 Magnetic pressure and tension

In this section I use our previous findings to determine the effect of magnetic fields on a stellar structures. To this end I use the equation of motion:

$$\rho\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\nabla p + \rho\vec{g} + \frac{1}{c}\vec{j}\times\vec{B} \tag{4.36}$$

The first term on the right is the gas pressure gradient. The second term stands for gravity and the third is the Lorentz force. Of course there may be more terms important in equation (4.36) here but I concentrate on the ones important to this course.

When we combine equation (4.5) with the equation of motion we find:

$$\rho\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\nabla(p + \frac{\vec{B}^2}{8\pi}) + \rho\vec{g} + \frac{1}{4\pi}(\vec{B}\cdot\nabla)\vec{B} \tag{4.37}$$

Here $B^2/(8\pi)$ is the magnetic pressure and $1/(4\pi)(\vec{B}\cdot\nabla)\vec{B}$ the magnetic tension along the magnetic field lines. The existence of magnetic tension along the magnetic field lines may be compared to tension in a string. Therefore we expect that transverse waves may propagate along field lines. These perturbations of magnetic field lines are Alfvén waves and they propagate with the Alfvén speed $v_\mathrm{a} = B^2/(4\pi\rho)$.

We can decompose the magnetic tension assuming $\vec{B} = B\vec{s}$ in terms of the unit vector $\vec{s}$ along the field:

$$\frac{1}{4\pi}(\vec{B}\cdot\nabla)\vec{B} = \frac{B}{4\pi}\frac{d}{ds}(B\vec{s}) = \frac{d}{ds}\left(\frac{B^2}{8\pi}\right)\vec{s} + \frac{B^2}{4\pi}\frac{\vec{n}}{R_\mathrm{c}} \tag{4.38}$$

Where $\vec{n}$ is the principle normal to the magnetic field line and R_c is its radius of curvature. Thus, the smaller the radius of curvature, the larger the tension force becomes. The first term on the right side of this equation is canceled by the component of the magnetic pressure along the field lines which is obvious since we used a vector perpendicular to $\vec{B}$ (*i.e.*, $\vec{j}\times\vec{B}$).

4.7.1 The magnetic buoyancy force

An important conclusion has been reached now: magnetic fields exert pressure. This pressure may be added to the gas pressure. Now I will find through a simple argument that this extra pressure will make the flux tubes (see Sect. 4.8.1) rise upwards through the stellar object. To this end I imagine a simple example in which we have a homogeneous and straight flux tube perpendicular to the gravitational field (see Figure 4.10). Note that inside this flux tube the pressure will be the same as outside (assuming that perturbations will be slow). For a pressure difference will make the flux tube expand or compress until an equilibrium has been reached. In formula:

$$p_\mathrm{in} + \frac{B^2}{8\pi} = p_\mathrm{out} \tag{4.39}$$

Assuming the ideal gas law $p = T\rho\mathcal{R}/\mu$ and that the temperature differences between the inside and the outside of the tube will be small due to the high radiation inside a stellar object, equation (4.39) will result in:

$$\rho_\mathrm{in} - \rho_\mathrm{out} = \frac{\mu}{\mathcal{R}T}\frac{B^2}{8\pi} \tag{4.40}$$

Thus we find the force on the flux tube must be:

$$F_\mathrm{upward} = \int \frac{\mu g}{\mathcal{R}T}\frac{B^2}{8\pi}\,\mathrm{d}V \tag{4.41}$$

This example is somewhat simple since a completely homogeneous tube is very rare. There will almost always be parts were the magnetic field is stronger and these will rise faster causing the tube to bent. These tubes will form a bipolar magnetic surface upon breaking through the stellar surface. While rising up through the interior of the stellar

object, the flux tube will try to keep pressure balance. This means that the thinner the atmosphere the further the tube will expand. This causes the curvature of the magnetic field lines to increase (*i.e.*, the radius of curvature decreases) and the magnetic tension increases too (see equation (4.38)). An equilibrium state will be reached whenever:

$$\frac{B^2}{4\pi R_c} = \frac{\mu g}{\mathcal{R}T}\frac{B^2}{8\pi} \tag{4.42}$$

This will happen when:

$$R_c = 2\frac{\mathcal{R}T}{\mu g} = 2H \tag{4.43}$$

Here H is the scale height. Again this is a simplified example; in order to be able to answer the question whether or not this equilibrium will be reached, one has to consider convection too. Convection causes the tubes to rise adiabatically keeping their temperature higher than the outside temperature; this causes the tube to keep on rising.

4.7.2 Magnetic forces influencing stellar structures

I already discussed some ways in which magnetic fields may influence stellar structures. Note however that there are many ways in which astrophysical magnetic fields can exert their influence. In many cases this will make observations of these stellar objects hard to understand. But sometimes we are able to understand the complex way it works. Again I will not treat all the effects of magnetic fields, but many visible effects will be dealt with in later chapters. Here is a summation of observable phenomena produced by magnetic fields:

- Zeeman broadening;
- line and continuum polarization and depolarization;
- radio emission: gyro synchrotron;
- EUV and X-ray emission: heating, acceleration;
- heating: MHD waves, field annihilation;
- particle acceleration: currents;
- flares: rapid field annihilation;
- structural isolation: flux tubes; network; loops, arcades; prominences;
- thermal isolation: suppression of conduction and convection;
- disruption of convection: abnormal granulation;
- turbulent motions;
- wave motions;
- systematic flows;
- wind acceleration;
- wind structure: fast/slow;
- wind composition: FIP anomalies.

Gyro synchrotron radio emission is produced by particles gyrating about magnetic field lines. Since they are continuously accelerating they have to emit radiation. This kind of physics belongs to the branch of plasma physics called: "the theorie of motion of individual charged particles" treated in Goedbloed and Poedts (1997).

4.7.2.1 Observable phenomena: magnetic braking

Magnetic braking occurs if a rotating magnetized object loses mass which flows along field lines. If the field is strong enough to control the flow (this is true whenever the energy density of the magnetic field is greater than the kinetic energy of the expelled matter), the expelled mass gains angular momentum and the underlying object is slowed down. Mass following these magnetic field lines will eventually arrive at a distance at which their kinetic energy exceeds that of the magnetic fields. There the particles may move on "freely" carrying the magnetic field lines with them.

We can understand that this point (where kinetic energy exceeds magnetic energy) must exists at some radial distance from the object: the mass should have the same angular velocity (initially obtained by leaving the stellar object) in addition to a large enough radial escape speed so as to neglect the object's gravity. Now in order to keep the same angular velocity (forced by the magnetic field) at larger distances the matter has to go faster, thus increasing it's kinetic energy. The magnetic field lines, however, diverge and thus the energy density of the magnetic field decreases with the distance.

The gain in angular speed of the mass, can only come from its angular momentum. Thus the rotation of the object is slowed down. A simple explanation of this phenomenon may be found in Tayler (1997).

4.7.3 Force-free fields

In many applications not all terms in equation (4.36) are equally important. We may for example neglect gravity with respect to the pressure gradient when the height of the structure is less than the scale height. In addition $\vec{v}$ may be neglected when the flow speed of the plasma is much smaller than the sound speed, the Alfvén speed and the free fall speed. We are working within the field of magnetohydrostatics. This results in:

$$\vec{0} = -\nabla p + \frac{1}{c}\vec{j} \times \vec{B} \tag{4.44}$$

Now if β is much smaller than unity, indicating the dominance of magnetic pressure over gas pressure, pressure gradients can be neglected with respect to the Lorentz force. Equation (4.44) reduces to:

$$\vec{j} \times \vec{B} = \vec{0} \tag{4.45}$$

Magnetic fields satisfying this condition are called force-free. In particular when $\vec{j} = 0$ the magnetic field is called current-free or potential. We must note that within these assumptions the Lorentz force is limited since the pressure-gradient is not large enough to compensate and equation (4.44) still holds.

Magneto hydrostatics may be applicable to a variety of solar structures in which $\vec{v}_{\text{plasma}} \simeq 0$. Good examples are: the overall structure of sunspots and prominences, and the large scale structure of the coronal magnetic field, appearing stationary for times long compared with the Alfvén travel time, and regions above an active region.

Examining the equation (4.45) we notice that the electric current flows along magnetic field lines. In spite of the "simple" equation (4.45) general solutions to this equation are still not found.

Some interesting theorems result from the force-free equation and Priest (1982) makes some interesting reading on this subject. However in order to keep this section clear only a few results will be mentioned:

- the induction equation and the force-free equation dictate a slow resistive diffusion of a magnetic field through a series of force-free equilibria. And an initially force-free field diffuses in such a way as to remain force-free;
- if $\vec{j} \times \vec{B}$ vanishes everywhere within a volume V and on its surface S, the magnetic field is identically zero. Thus a non-trivial magnetic field that is force-free within V must be stressed somewhere on S, since the Lorenz force cannot vanish everywhere on S. So if one were to construct a force-free field one must anchor it somewhere on the boundary (e.g. in a high plasma density region).

4.8 Flux tubes and current sheets

Magnetic configurations can be thought of as built up from flux tubes and/or current sheets. We will look at the building blocks simply as isolated entities. Note that this again is a crude assumption since both interact intimately with their surrounding magnetic field.

4.8.1 Flux tubes

The physics of flux tubes is an extensive subject and is discussed in Parker (1979). Again I will only treat a few important aspects (we use the summary found in Priest (1982)). Examples of flux tubes are sunspots and prominences. A magnetic field tube or flux tube is the volume enclosed by the set of field lines which intersect a simple closed curve. The strength of such a flux tube may be defined as the flux crossing a section S.

- The strength (in other words flux) of a flux tube remains constant along its length.
- in order to keep the flux constant the mean field strength of a flux tube increases when it narrows and decreases when it widens;
- whenever a flux tube is compressed or expands $\vec{B}$ and ρ increase or decrease in the same proportion. This too has already been proven in Sect. 4.5.1. For a general proof see Priest (1982);
- an extension of a flux tube without compression increases the field strength. This is easily seen from the previous points: If the plasma is not compressed (thus ρ stays the same for matter is now free to flow in and out the tube) $\vec{B}$ increases with the extra volume created. Increasing of the length of the flux tube may be caused by shearing motions.

Some properties of flux tubes in the solar atmosphere, many of which will be treated in the next chapters, are:

- convection can expel magnetic flux from a convecting eddy and concentrate it to form a vertical flux robe;
- a horizontal flux tube embedded in a gravitationally stratified medium is subject to a magnetic buoyancy force, which tends to make it rise. It can remain in equilibrium as an arch if the feet are anchored at points that are separated by less than a few scale-heights. Also magnetic buoyancy may destabilize an equilibrium magnetic field whose strength declines too rapidly with height;
- a sunspot may consist of either a single large flux tube in equilibrium or a cluster of small tubes held together by magnetic buoyancy and a downdraft;

- a slender sub-photospheric tube in thermal and hydrostatic equilibrium widens with height. If it is cooler than it's surroundings it becomes evacuated over a few scale-heights;
- a magnetic flux tube appearing in a turbulent astrophysical environment cannot be entirely free of twisting, and is often referred to as a flux rope. Only a few revolutions in a long tube are enough to produce qualitative dynamical effects of instability and non-equilibrium. Twisted flux tubes are again extensively examined in Parker (1979). A few characteristics of twisted flux tubes will be present in the next chapters but these will almost always be "simple" and intuitive.

4.8.2 Current sheets

The current density $j \sim cB/(4\pi L)$ mostly is very small since the scale length L is very large. However it is believed that current sheets exist with very small widths and corresponding current densities much larger than the estimate $cB/(4\pi L)$. They do not have long life times and are believed to play an important role in solar flare processes and the like.

A current sheet may be defined as a non-propagating boundary between two plasmas, with the magnetic field tangential to the boundary. The tangential field components are subject to the condition that the total pressure is continuous:

$$p_2 + \frac{{B_2}^2}{2\mu} = p_1 + \frac{B_1^2}{2\mu} \tag{4.46}$$

where subscripts 1 and 2 denote conditions on the two sides of the current sheet. Inside active regions the magnetic field is so strong that for many purposes the plasma pressures outside the current may be neglected.

A current sheet is comparable to a shock wave: it is a discontinuity separating two regions where the equations of ideal MHD hold. Also its width and the details of its interior are determined by diffusive processes. Here the similarity ends for current sheets do not propagate, instead they tend to diffuse away in time and jets of plasma are squirted from their ends at Alfvénic speeds.

Several processes in which current sheets may be formed are also found in Priest (1982):

- the region near an X-type neutral point can collapse;
- the compression of topologically separate parts of a magnetic configuration may produce current sheets at the boundary between them;
- current sheets may develop when a magneto hydrostatic equilibrium becomes unstable.

A full explanation to these causes may be found in Priest (1982). Finally, I discuss some of the basic properties of current sheets:

- in the absence of a flow, a current sheet diffuses away at a speed η/L, where η is the magnetic diffusivity (this can be seen by dividing L by τ_D). As I said before, the magnetic field is annihilated and magnetic energy converted into heat by ohmic dissipation;

- the region outside the current sheet is effectively frozen to the plasma. Plasma and magnetic flux may be brought towards the sheet from the sides at a certain speed. Whenever this speed exceeds the speed of dissipation of the sheet the sheet will become thinner, if it is less the sheet expands and otherwise an equilibrium is maintained;

- the enhanced plasma pressure in the center of the sheet expels material from the ends of the sheet at the Alfvén speed based on the external magnetic field and internal density. Magnetic flux is ejected with the material, and so one effect of the sheet is to reconnect the field lines. Another is to convert magnetic energy into heat and flow energy;

- pairs of slow-mode shock waves (treated in later chapters) propagate from the ends of the current sheet and remain as standing waves in a steady flow;

- current sheets with the conditions given in the example in Sect. 4.5.2.1 are subject to the tearing-mode instability on a time scale of typically the geometric mean of the diffusion time and the Alfvén travel time. A good summary of resistive instabilities is found in Priest (1982).

Chapter 5

Diagnostics with the Zeeman effect

Lecturer S. Solanki, notes by A.M. van Oosten

5.1 Introduction

To get a better understanding about the Sun's magnetic features, we need to know certain parameters: the magnetic field strength B, filling factor, inclination angle and azimuth of **B**, temperature, velocity and pressure. This is not an easy task since all these parameters depend in turn on position on the Sun and time. Also, they often enter in complex and hidden ways into the observed radiation.

A way to obtain information about the magnetic features of the Sun is to observe solar images, or spectra, in polarized light. A useful description of polarized light uses the Stokes parameters, which use net polarizations in addition to the total intensity. The Stokes parameters have the advantage that they can be recorded with the help of only a few measurement (as few as four). They are a powerful tool in resolving magnetic features and determining their properties.

Using the Stokes parameters, empirical models can then be constructed. Such models, sunspot thermal structure, for example, can then be applied to give us a better physical understanding. For example, about the factor by which convective energy transport is inhibited by the sunspot magnetic field and the layers in which convection is important. Or the mechanical heating rate in the upper atmosphere and how it differs from that in other parts of the Sun (quiet Sun, plage, etc.)

In this chapter we discuss how the Stokes profiles can be applied to determine some of the important parameters for modeling solar magnetic features. Diagnostics with Stokes-spectra rely heavily upon the the theory of the Zeeman effect.

5.2 The Zeeman Effect

In this section we will discuss some of the basic properties of Zeeman splitting. This discussion is primarily descriptive in nature. For a rigorous quantum mechanical derivation of the Zeeman effect, I refer to the excellent work of Bransden and Joachain (1989). In the presence of magnetic fields, each atomic energy level, and thus the spectral line associated with it, is split into a number of levels. This splitting is called the Zeeman effect. The magnitude of the splitting is an indication of the strength of the magnetic field. The Zeeman splitting also produces a unique polarization signature in the spectral lines. A

measurement of the polarization thus also provides additional information on the magnetic field.

The splitting itself is associated with the total angular momentum quantum number $J = L+S$, where L is the orbital momentum quantum number and S is the spin quantum number. J can only take on non-negative integer and half-integer values. The number of split levels corresponds to $2J+1$, the magnetic quantum number $m_J = -J, -J+1, ..., J-1, +J$. In the classic picture, Zeeman splitting can be interpreted in terms of the precession of the total angular momentum vector **J** in a magnetic field, similar to the precession of the axis of a spinning top in a gravitational field.

Atomic physicists use the abbreviation "s" for a level with $L = 0$, "p" for $L = 1$, and "d" for $L = 2$, and so on (the reasons for these designations are of historical interest only). It is also common to precede this designation with the integer principle quantum number n. Thus, the designation "2p" means a level that has $n = 2$ and $L = 1$. How strongly the energy levels split depends on the strength of the magnetic field, the landé-factor and the wavelength. Into how many components they split depends on the atomic structure, *i.e.*, on J and the Landé factor. The Landé factor is defined as followed:

$$\mathrm{g} = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} \tag{5.1}$$

So g is entirely dependent upon the, for the Zeeman effect, relevant quantum numbers.

We need to discriminate between the normal Zeeman effect, when a spectral line splits into three equally spaced components, and the anomalous Zeeman effect, when a spectral line splits into an often larger number of components. Here we only consider the case of interest for Zeeman-splitting measurements on the Sun, namely when the splitting is large enough that there is no quantum interference between different m-levels and when it is small enough that we needn't take the Paschen-Back effect into account.

In Fig. 5.1 we see an example of the normal Zeeman effect. Notice that there are (only) three resulting spectral lines, the so called Lorentz triplet. In this case the Zeeman splitting $\Delta\lambda_H$ can be written as:

$$\Delta\lambda_H = kg\mathbf{B}\lambda^2 \tag{5.2}$$

where k is a constant, g is the landé-factor of the line, **B** is the field strength and λ the wavelength.

The "anomalous" Zeeman effect is actually the one most commonly encountered. In the early days of spectroscopy, before the electron spin was discovered, the normal Zeeman effect was predicted, on classical grounds, but observations did not conform to the predictions and were said to be 'anomalous' (Bransden and Joachain (1989)). In Fig. 5.2 we see examples of anomalous Zeeman splitting patterns.

In Fig. 5.3 the formation of a Zeeman triplet can be observed, both in a field, directed along the line of sight and a transverse field. Astrophysical fields are often too weak to separate the σ components fully from the central π component (the lower right figure).

The lines undergoing Zeeman splitting also exhibit polarization effects. Polarization has to do with the direction in which the electro-magnetic fields are vibrating. Using certain filters, called polarizers, certain polarizations can be selected allowing us to better research, for example, sunspots. It is almost only in sunspots that we can see and directly measure the Zeeman splitting in unpolarized radiation. The most complete observational information is of course obtained if we can record the full state of polarization across the spectral lines with high spectral resolution(Stenflo (1994))

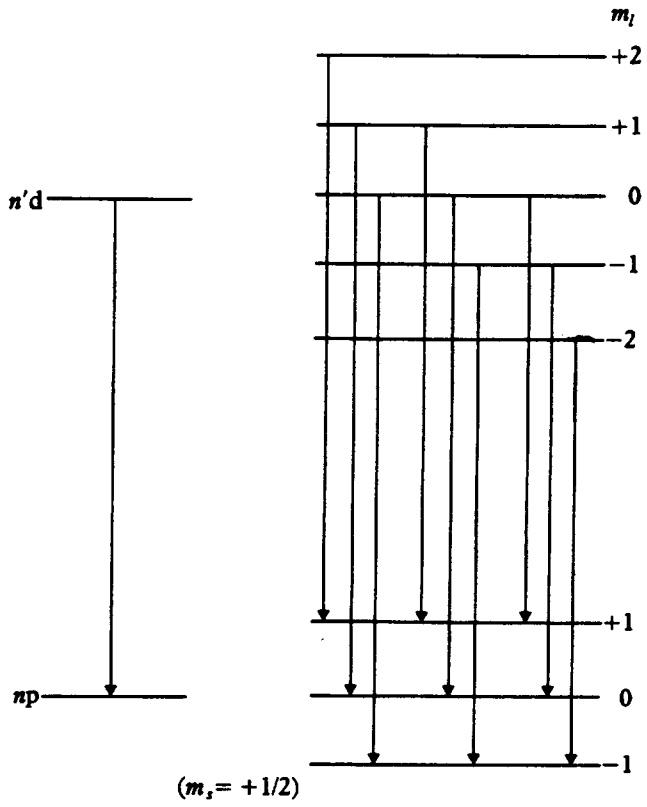


Figure 5.1: The normal Zeeman effect. In nine transitions are possible between the split levels consistent with $\Delta m_l = 0$ or ± 1 and $\Delta m_s = 0$. Of these, there are only three different frequencies and the lines form a Lorentz triplet. The frequencies of transitions associated with $m_s = -\frac{1}{2}$ are the same as those for $m_s = +\frac{1}{2}$.

5.3 The Stokes parameters

The Stokes-parameters are generally denoted I, Q, U, and V. Together they form a 4-vector:

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}. \tag{5.3}$$

The four Stokes parameters each represent the light intensity transmitted by a certain idealized filter (see also Fig. 5.4):

I = total intensity;

Q = intensity of linear polarization at 0^o minus the intensity of the linear polarization at 90^o;

U = intensity of linear polarization at 45^o minus the intensity of the linear polarization at 135^o;

V = intensity of right-handed (clockwise) circular polarization, minus the intensity of left-handed (counter-clockwise) circular polarization.

These six intensity readings uniquely determine the full state of polarization of the incident beam. Examples of the four Stokes profiles of a solar spectral lines for a range of

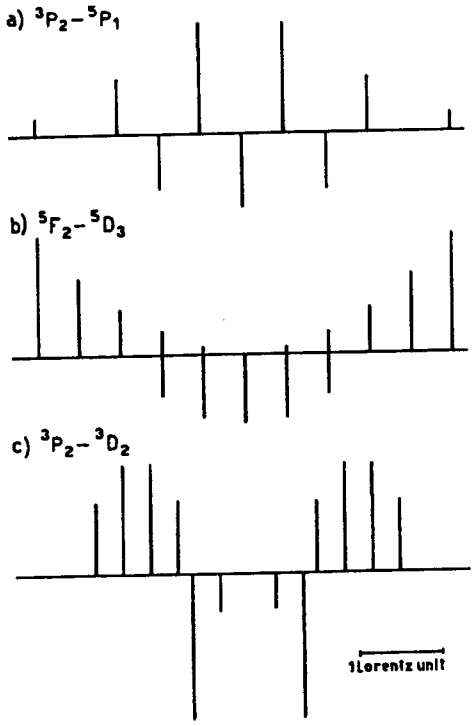


Figure 5.2: Illustration of different anomalous Zeeman splitting patterns. The lengths of the vertical bars are proportional to the strengths of the transitions, their positions indicate their wavelength shifts. The π components point downwards, the σ components upwards. From Landi Degl'Innocenti and Landi Degl"Innocenti (1985)

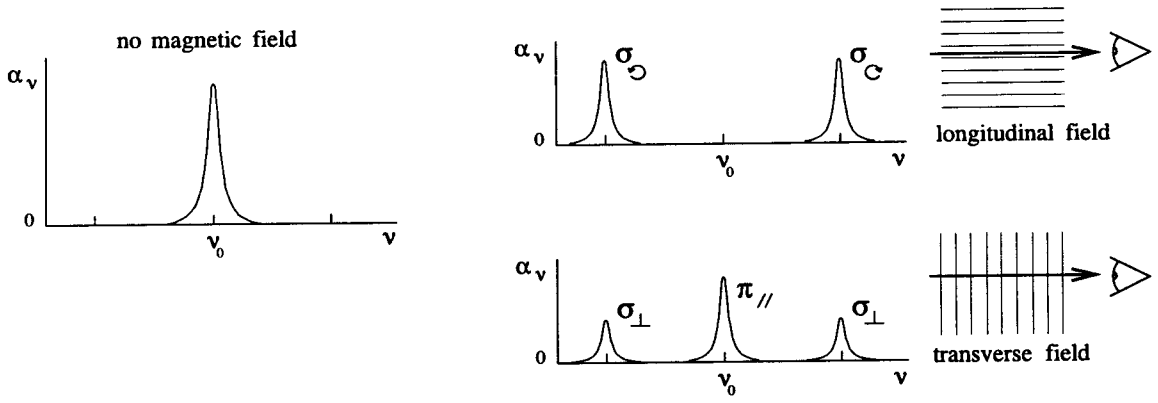


Figure 5.3: Zeeman triplet. Left: extinction profile for a medium without a magnetic field. The graphs on the right are for a medium that is pervaded by a strong, homogeneous magnetic field, respectively showing the longitudinal Zeeman pattern (upper right) and the transverse Zeeman pattern (lower right). α_ν is here the extinction coefficient for a frequency ν. When the line of sight is along the field lines, the extinction profile consists of two symmetrically displaced σ components, applying to lefthand and righthand circularly polarized light, respectively. When the line of sight crosses the field at right angles, the "transverse" Zeeman effect produces three extinction peaks, one at line center which applies to linearly polarized radiation with the Stokes vector parallel to the field vector, and two displaced σ components that extinguish linearly polarized radiation with the Stokes vector perpendicular to the field direction. From Rutten (1999b)

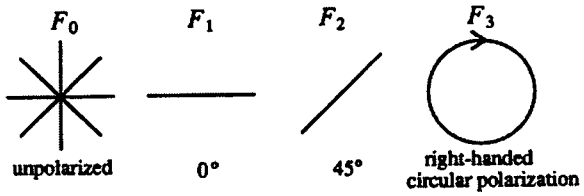


Figure 5.4: A graphical representation of the filters used to obtain the Stokes-profiles

field strengths are shown in Fig 5.6. The Stokes profiles are sensitive to many different atmospheric parameters:

- abundance: affects strength of all profiles of an element;
- temperature: affects the strength of neutral vs. ionized lines;
- magnetic field strength: affects line profile shapes, for example through different Zeeman splitting;
- field inclination: affects the ratios V/Q and Q/U;
- field filling factor: affects the total amplitude $\sqrt{V^2+Q^2+U^2}$;
- velocity:
  - stationary flow: produces line shift and asymmetry;
  - non-stationary flow: produces time-dependent line shift, asymmetry and enhanced line width;
- sizes of magnetic elements: affect the profile shapes and magnetic field appearance in images.

The similar influence of different atmospheric parameters on the Stokes-profiles makes diagnostics a difficult task. Also the influences are often subtle and thus hard to detect. Due to correlated gradients of the magnetic and velocity fields, the observed Stokes V profiles are asymmetric, *i.e.*, the blue wing differs from the red line wing both in area and in amplitude (see Fig. 5.5). An area-asymmetry can only arise if the correlated magnetic and velocity field gradients are along the line of sight.

5.4 Modeling Magnetic Features

If the observed magnetic feature is spatially resolved, one-component modeling is sufficient. However, usually two or more components are needed, since due to seeing, the finest relevant scales on the solar surface usually cannot be resolved.

An unresolved magnetic element has a magnetic filling factor α. The magnetic filling factor is the fraction of area covered by the field within a spatial resolution element. α is proportional to V, U and Q, yet since these parameters depend on so many different quantities (see section 5.3), which usually aren't all reliably known, α is hard to determine. Errors of a factor 2 are not uncommon. $\alpha = 1$ means a magnetic element is spatially resolved. An unresolved element has the following observed Stokes profiles (averages over resolution element):

$$< I > = \alpha I_{\rm m} + (1-\alpha) I_s \tag{5.4}$$

$$< P > = \alpha P_{\rm m}, \tag{5.5}$$

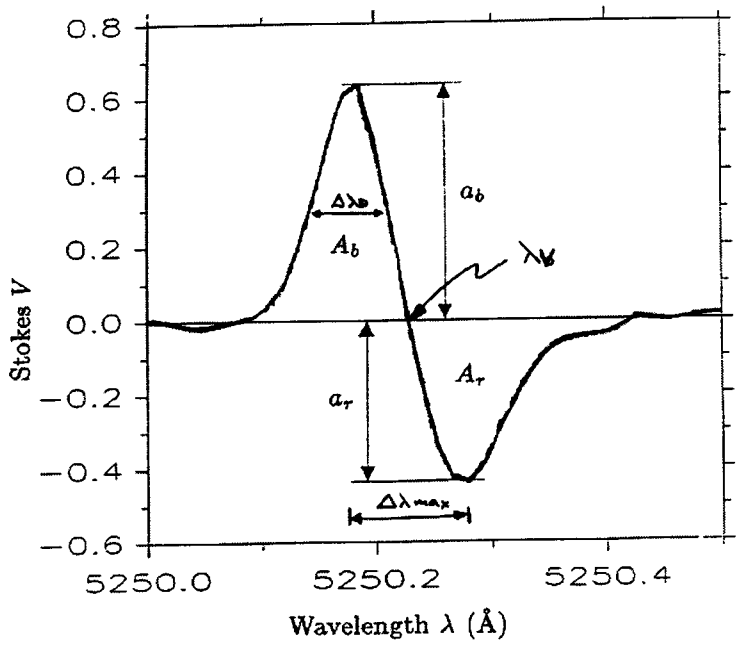


Figure 5.5: Example of an asymmetric stokes profile, where A_b and A_r are the area of the blue and red line wing respectively andd a_b and a_r the amplitude of the respective line wings. λ_v is the line shift due to the velocity of the matter and λ_D the Doppler line broadening.

where $I_{\rm m}$ is the intensity in the magnetic element, I_s is the light coming from outside the magnetic element, $P = Q$, U, V respectively, and α is the magnetic filling factor. The observed profiles of Q, U and V give information on the magnetic element alone, since there is no polarized radiation coming from outside the element There are three main approaches to modeling:

- purely empirical: the model parameters are derived independently from the observations. Reasonable values are assumed for the rest;
- semi-empirical: key quantities obtained from observations. All model parameters are linked self-consistently via a (simple) physical model. The model is then solved through an iterative process;
- theoretical: all magnetic and hydrodynamic quantities are derived from basic physics. The model is then compared directly to the observations, or to semi-empirical models.

5.5 Magnetic Field-strength Diagnostics

We discriminate between three regimes of magnetic field strength:

- the weak case: $\Delta\lambda_H \ll \Delta\lambda_D$;
- the intermediate case: $\Delta\lambda_H \approx \Delta\lambda_D$;
- the strong case: $\Delta\lambda_H >> \Delta\lambda_D$.

where $\Delta\lambda_H = kg\mathbf{B}\lambda^2$ measures the Zeeman splitting of a spectral line. The profile depends on the Doppler width $\Delta\lambda_D \sim \lambda$, the turbulent velocity $\xi_T, \sqrt{T}$ and on the amount of

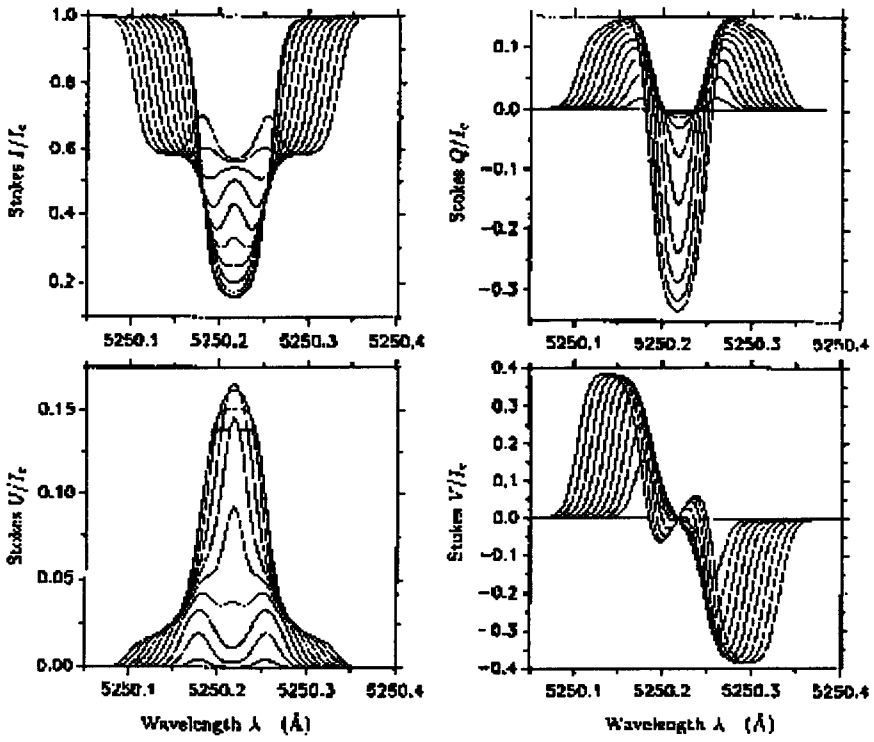


Figure 5.6: Example of a numerical solution of the LTE (Local Thermodynamic Equilibrium) transfer equations for the Fe I 5250.2 A absorption line at the center of the solar disk, using a quiet Sun model atmosphere with a homogeneous magnetic field of inclination $\gamma = 45^o$ and azimuth $\chi = 0$. The ten different curves correspond to field strengths varying from 0.2 to 2.0 kG in steps of 0.2 kG. The Stokes parameters are expressed in units of the intensity I_c of the local continuum. Note that the signal in Stokes U is exclusively due to magneto-optical effects. From Solanki (1993)

saturation in the line width in the absence of a magnetic field. Thus, the regime of a given measurement depends not only on the field strength but also on the spectral line that is measured.

Also, looking at the two relations above, one notices that while $\Delta\lambda_D$ increases with λ, $\Delta\lambda_H$ is proportional to λ^2. Hence the Zeeman-sensitivity increases with increasing wavelength. Higher wavelengths (infrared) could thus provide us with more detailed information on the magnetic structures.

However, as always, a few problems arise when measuring in the infrared spectrum. Firstly, detectors themselves emit infrared radiation at relatively low temperatures. It is thus necessary to use cryogenic detectors, *i.e.*, detectors which are kept at a sufficiently low temperature, so there is as little noise as possible. And secondly, we still lack some knowledge of the infrared spectrum; We miss some spectral identifications, which make it hard to couple the spectral lines to the models, although the situation is rapidly improving.

5.5.1 Weak-Field Case

In the weak field case, the Zeeman splitting is much smaller than the Doppler line width. B can only be measured if the magnetic structure is spatially resolved. If this is not the case, then only $< B >$ can be measured. In the weak-field case the following relations are valid:

$$V = \alpha \Delta\lambda_H \cos(\gamma) \frac{\delta I}{\delta \lambda} \tag{5.6}$$

$$\frac{\cos\gamma}{\sin^2\gamma} \sim \frac{V}{\sqrt{Q^2+U^2}} \tag{5.7}$$

$$\tan 2\chi \sim \frac{U}{Q} \tag{5.8}$$

where γ is the angle of $\mathbf{B}$ to line of sight and χ is the azimuth angle.

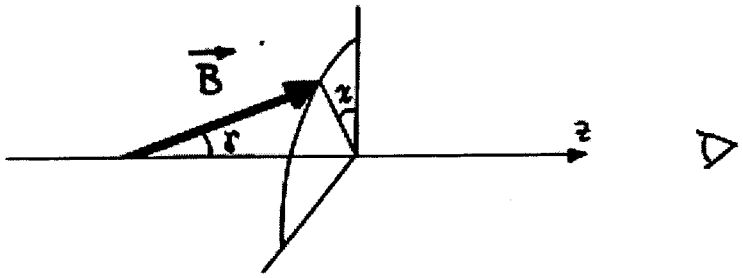


Figure 5.7: The magnetic field vector **B**, showing the definition of the angles γ and χ, with the z-axis being the line of sight.

When the field is resolved, the magnetic filling factor α=1. So, if Q and U are also observed, B, γ and χ can be determined. One has to be cautious though, when using the above relations. They are useful for first rough analysis of observations. The neglection of magneto-optical effects, however, may lead to serious errors in the calculations. In Fig. 5.10 we can see an example, illustrating the weak-field approximation. Also, there is the so called 180^o ambiguity: It is impossible from Q and U alone, to determine the direction of the transverse field, *i.e.*, it is unclear if it is pointing in one direction, or a direction opposite to that.(See also Fig. 5.8)

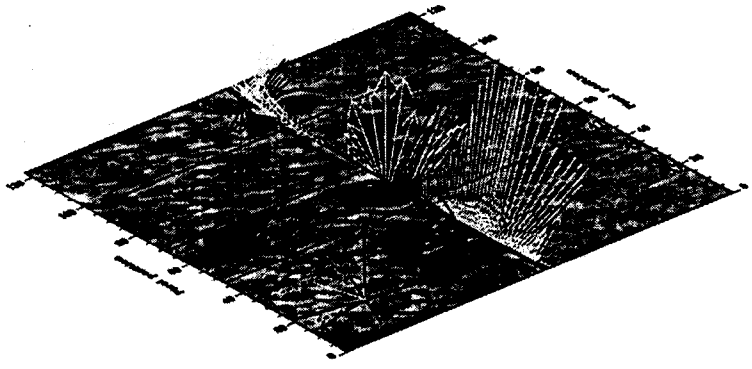


Figure 7: Same as in Fig. 6, but in a 3D representation.

Figure 5.8: Observations of an active region, with directions of the magnetic field drawn in

For weak fields, the following proportionalities are valid: $V \sim B$, $U \sim B^2$, $Q \sim B^2$ For strong fields, however, the Stokes V profile hardly changes with increasing field strength.

This phenomenon is called Zeeman-saturation, and it is a function of the intrinsic field strength alone (see Fig 5.9). For even stronger fields, Q and U also show this property (*i.e.*, Zeeman saturation).

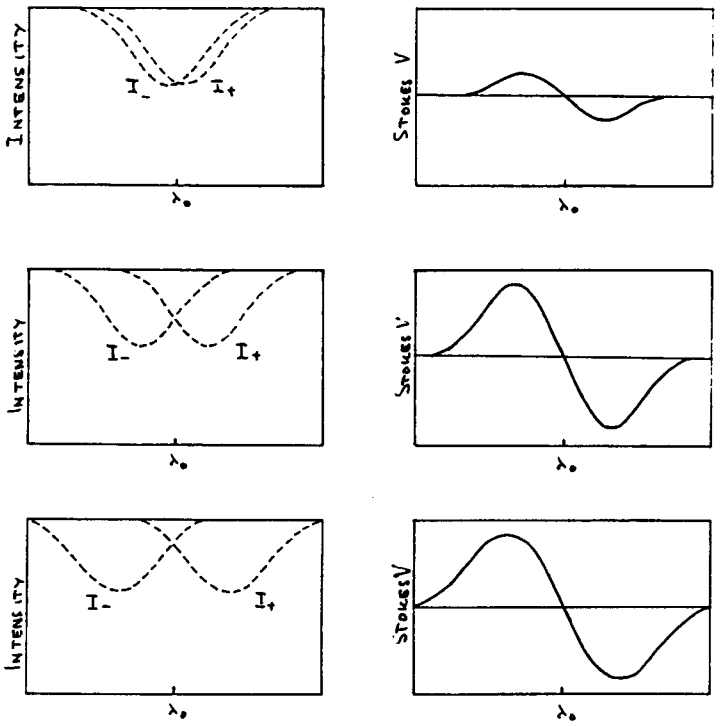


Figure 5.9: Zeeman-Saturation: The Zeeman-splitting $\Delta\lambda_H$ increases downward, due to increasing magnetic field strength. When the field strength B is high, Stokes V hardly rises at all anymore, while for weak fields V is proportional with B.

5.6 Thermal Diagnostics

When the field is spatially resolved, one can determine the temperature relatively easily using standard methods:

- measuring the continuum intensity $I_c(\lambda)$;
- observing Stokes I line depths (a single spectral is sufficient);
- observing profile shapes of strong lines (like the Ca II K line).

When the field is spatially unresolved however, one has to use different methods. The ratio between two Stokes V-profiles with different azimuth χ_e or line weakening w_T can be used. Or the shapes of Stokes Q, U and V can be used: If we look at the Stokes Q profile, we can determine the amount of Zeeman saturation, by determining the height of the peaks. The smaller the peak the more saturation. From the shape of the Stokes V profile of well-chosen spectral lines, we can determine the chromospheric temperature rise.

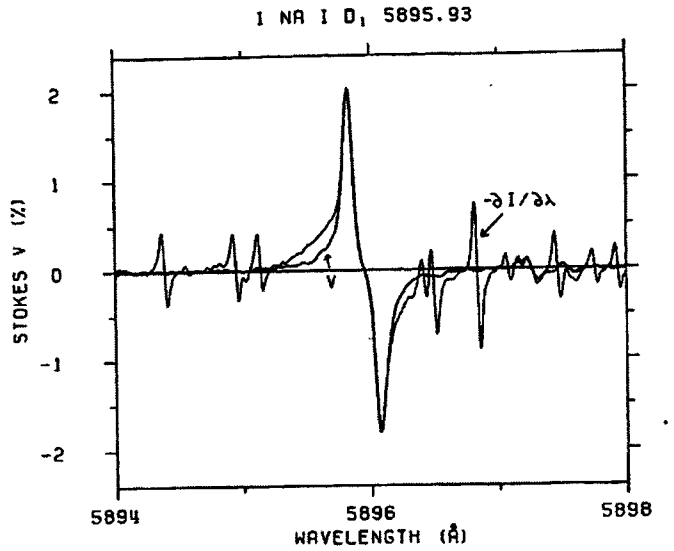


Figure 5.10: Stokes V (thick curve) plotted in units of the continuum intensity together with $\frac{\delta I}{\delta \lambda}$ (thin curve), which has been scaled to agree with the Stokes V amplitude of the $NaID_1$ line. From Stenflo (1984)).

5.7 Velocity Diagnostics

While observing a spectral-line, it can be affected in several ways, often due to the velocity of the particles involved. Firstly, there are wavelength shifts. These can have different causes:

- Steady flows: as matter flows with a constant velocity, it is Doppler-shifted over a constant $\Delta\lambda_v$;
- Non-Stationary velocities: if there is a correlation between the temperature and the velocity of the matter;
- Instrumental Smearing: if the line profile is asymmetric (see section 10.9);

Secondly, there is line broadening, caused by:

- Line of sight velocity gradient: continuously changing velocities along the line of sight, means different Doppler shifts at different depths, broadening the line;
- Turbulent velocity: matter moving in a random direction, causes random Doppler shifts (thermal broadening);
- Instrumental smearing: due to the insufficient spectral resolution of the instrumentation(see section 10.9);
- Zeeman-broadening: a distribution of field strengths within the resolution element will induce Zeeman line broadening (especially in the infrared);

And finally line asymmetry:

- Line of sight velocity gradient: Stokes V area asymmetry (see section 10.10.2);

- Non-stationary or transverse velocity gradients: the asymmetry is due to correlated gradients of the magnetic and velocity fields;

Note that all these different items are not the cause of a spectral line, only the cause of a difference in the position, broadness, or the course of the line (e.g. asymmetry).

5.8 Conclusions

There are techniques available to determine many of the parameters which describe the physical structure of magnetic features. For some quantities no reliable technique appears to exist (*e.g.*, density). Such parameters must be determined via model-assumptions from other quantities. No diagnostic is, at some level, completely model independent, although this is often forgotten. In general, it is better to combine observations made in different spectral lines The more physical quantities are diagnosed simultaneously the better.

Chapter 6

Dynamo Theory

Lecturer P. Hoyng, notes by P.J. Sloover

Previous chapters have treated the magnetic structures observed on the solar surface, in the outer solar atmosphere, and in the solar wind. In this chapter we will take a closer look at the origin of these structures. What generates them, and how can we study the parts of the Sun that are hidden under the thick layers of the Sun's atmosphere? The main reference for this contribution is Hoyng (1992).

6.1 Introduction

Observations show that the Sun's atmosphere is in constant motion. Large bubbles of gas flow up and down and protuberances throw huge amounts of matter into space. With just a set of binoculars and a piece of paper you can make a projection of the Sun and see a well known, but still not fully understood phenomenon, sunspots. These dark regions on the surface are cooler than their surroundings and therefor seem black against the brighter photosphere. These spots are of magnetic origin, and have been observed with reasonable accuracy since the 17^{th} century. The Sun's magnetic field changes periodically with a period of 22 years, the well-known solar cycle. This behaviour can't be explained by means of a static primordial field for two reasons. A primordial field can not change its orientation in time and, if there is no dynamo active, resistive decay would make the field decay within one billion years. Our planet also has a non-primordial field. The resistive decay time of the Earth's core is approximately fifty thousand years. From paleomagnetic evidence we know that the Earth has had a magnetic field for several billions of years. Therefore there must be some process regenerating the field, and this process is what dynamo theory is about. There are different types of approaches to work on the subject: numerical calculations, spectral theory, laminar theory and mean field theory. In this chapter we shall concentrate on the mean field theory.

6.2 Observations

From observations we know that the number of sunspots changes periodically with an average period of 11 years. There are some deviations from this period, the longest period being 17 years and the shortest approximately 7.3 years. The sunspot data show a modulation of about 90 years, the "Wolf-Gleissberg cycle". Tree ring ^{14}C data shows us the solar activity over long periods of time, going back much further than sunspot data.

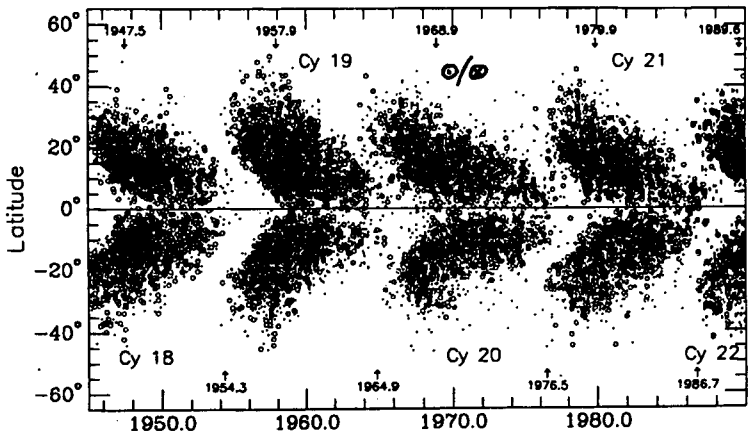


Figure 6.1: The butterfly diagram: a plot of the latitudes of occurrence of active regions versus time. From Harvey-Angle (1993)

From this data we observe that there are also long term cycles, with periods of several hundred years. They also include the so-called Grand Minima, periods of time in which the Sun seems to have no magnetic activity. The ^{10}Be nucleus has also been used to track the Sun's activity back into time. It is formed in the atmosphere and rains down over the Earth's surface. The amount of ^{10}Be formed depends on the intensity of the solar wind, which is directly proportional to the magnetic activity of the Sun. ^{10}Be is found in the polar caps and by taking samples from different depths we can recover the history of solar activity.

During a solar maximum the amount of sunspots is at a peak, and a large fraction of all flux is concentrated in active regions. These groups of two, or sometimes more, sunspots of opposite polarity are located in belts, parallel to the equator. Their bipolar axes are nearly East-West oriented and, with only few exceptions, obey Hale's polarity law. The formation zone of sunspots slowly decreases in latitude as the activity cycle evolves. This leads to the famous butterfly diagram shown in Fig. 6.1. The distribution of spots is not uniform in longitude either. New active regions tend to from in regions of previous activity. The smaller magnetic elements are located at the edges of supergranulation cells, but are otherwise distributed increasingly more homogeneous over the surface. A full magnetic period takes 22 years to complete.

Active regions emerge from deep in the convection zone and decay by turbulent diffusion in one or a few months. The decay of these active regions result in the background small-scale field. The total unsigned flux in this so-called "intranetwork field" shows a slight increase at a solar maximum, but this might be of artificial nature. The detector has a limited resolution and might not be able to resolve all individual granules. Another explanation for the minor increase is found in the decay of regions of high activity, which leads to an increase in the background noise. The large patches, flux tubes and bundles, have only a slightly stronger magnetic field than their surroundings. A typical sunspot has a field strength of about 3000 Gauss, and a small bright network element has a field strength of approximately 1500 Gauss. The difference in flux, 10^{22} Mx for a spot compared to 10^{18} Mx for a network element, is mainly caused by the large area of the spots compared to the small elements.

Two distinct patterns of convective activity are clearly visible at the surface. Granulation (typical length scale $\lambda = 1 \cdot 10^3$ km and time scale $\tau_c = 400$ s) and supergranulation

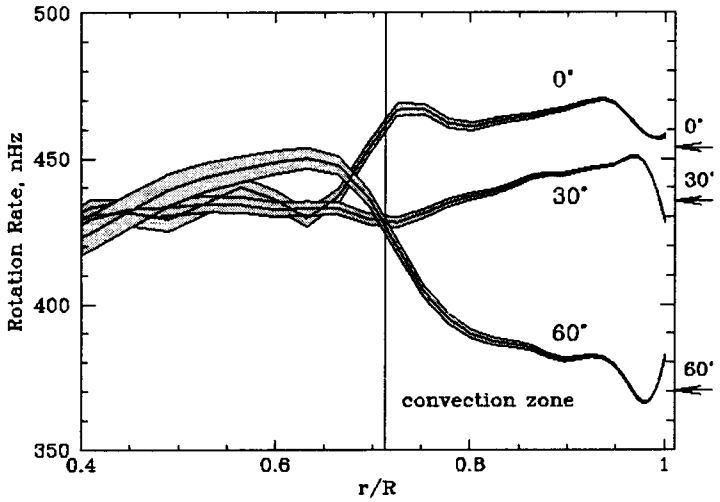


Figure 6.2: Inferred rotation rate as a function of depth and latitude. Evidently the convection zone rotates uniformly along a radius with all depths showing the differential rotation seen at the surface. Below the convection zone is a layer of shear below which the radiative interior seems to rotate rigidly. This shear zone which coincides with the sound speed excess could be the region where the solar cycle dynamo operates. From SOHO (1998)

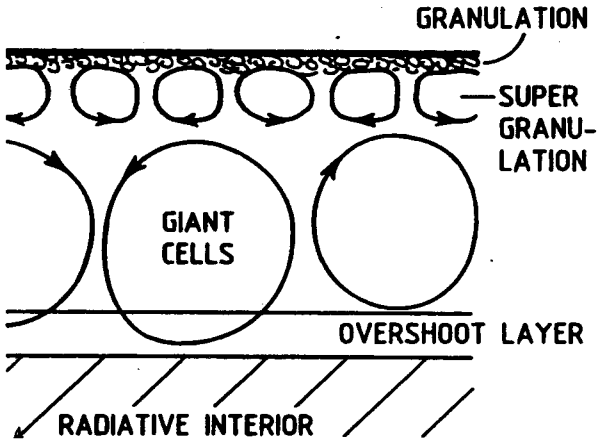


Figure 6.3: A rough sketch of the convection zone. From Hoyng (1992)

($\lambda = 3 \cdot 10^4$ km and $\tau_c = 10^5$ s). Deep in the convection zone, giant convection cells may exist ($\lambda_c = 10^5$ km and $\tau_c = 3 \cdot 10^6$ s). At the bottom of the convection zone exists a transition zone, the so-called overshoot layer. From helioseismology we know that Ω_r is roughly independent of r in the convection zone, and that Ω_r has a large radial gradient close to the overshoot layer. Below the transition layer the Sun rotates like a solid object.

To explain the form of the fields observed on the Sun we come with a first hypothetical picture of the structures involved. The large scale field behaves as if it consists of two tubes of toroidal field deep in the convection zone, one on each side of the equator with opposite polarities. Due to the convective activity, the frozen fields are at times taken up through the surface, where they result in pairs of sunspots. Their bipolarity resulted to Hale's polarity law. Sunspots travel in pairs of opposite polarity and the leading spots on the northern hemisphere have a polarity opposite to the ones on the southern hemisphere. The

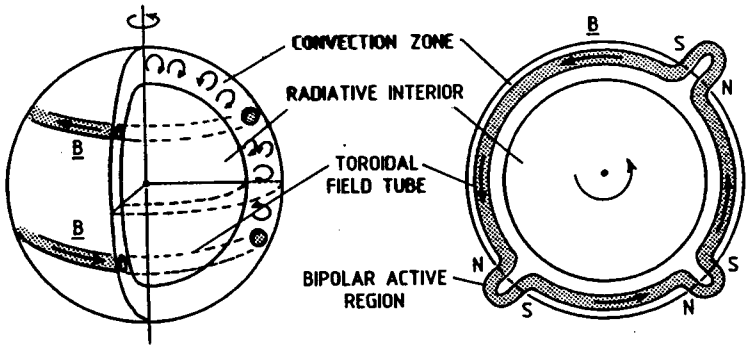


Figure 6.4: Apparent structure of the large scale field in the convection zone. From Hoyng (1992)

tubes apparently move to the equator in about 11 years and disappear as they arrive there. Then, somehow, new tubes are formed at high latitudes, but with opposite polarities. This process can be explained partly by means of differential rotation. Due to the difference in rotation speeds, an arbitrary fieldline is wound up to form two tubes of toroidal field of opposite polarity. However, to explain the periodic changes in polarity we need to take a closer look at the turbulent processes in the convection zone. In Fig. 6.2 is a sketch of the observed differential rotation in the Sun's outer layers.

6.3 Different types of dynamo

6.3.1 A trivial dynamo: the Jupiter-Io system

A distinction can be made between trivial and non-trivial dynamos. A trivial dynamo is a system in which a conductor moves through a given magnetic field. In a real dynamo the magnetic field must be generated by the current it excites. An example of a trivial dynamo is the Jupiter-Io system, see Fig. 6.5. The low orbit of Io causes it to be moulded by the high tidal forces, resulting in vulcanic activity. The eruptions bring huge amounts of ionized material into Io's atmosphere, thus making Io act like a conductor. As it circles through Jupiter's magnetic field (0.02 Gauss), at a speed of 56 km/s it generates a potential difference of 100 kV. The generated current excites radio emission but does not act as a source of the magnetic field, and thus makes it a trivial dynamo.

6.3.2 The homopolar disc dynamo

An instructive exercise is the analysis of the homopolar disc dynamo. A conducting disc rotates with angular velocity Ω around a conducting coil, connected with sliding contacts. The coil generates a magnetic field $\vec{B}$, and the required current I is maintained by the potential drop which exists between points P and Q because the disc moves through the magnetic field, see picture. The system obeys the circuit equation: $V_{\rm disc} + V_{\rm coil} = IR$, where R is the total resistance. $V_{\rm disc}$ is proportional to Ω and I: $V_{\rm disc} = \Omega I a$. Where a is a constant, following from this equation:

$$V_{disc} = \int_0^R (\vec{E} + \frac{1}{c}\vec{u}\times\vec{B})\cdot \mathrm{d}\vec{l} = \frac{1}{c}\int_0^R \vec{u}\times\vec{B}\cdot\mathrm{d}\vec{l} = \frac{1}{c}\int_0^R \Omega r\cdot B\cdot \mathrm{d}r = \frac{\Omega}{2c}BR^2 = \frac{\Omega}{2\pi c}\phi_{disc} = a\Omega I \tag{6.1}$$

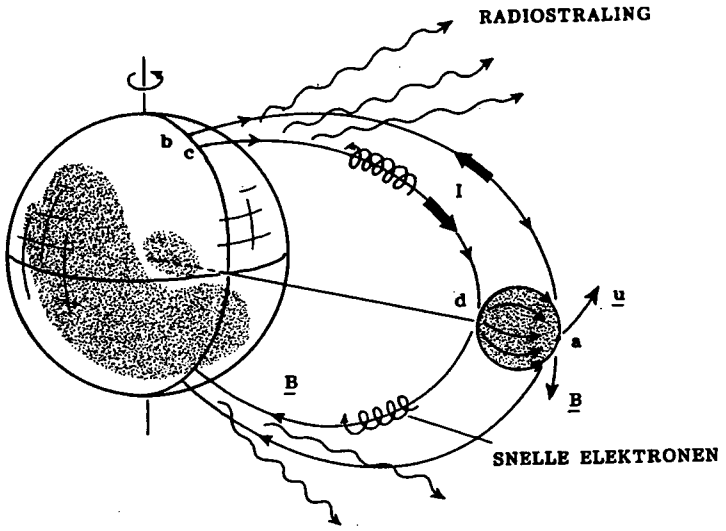


Figure 6.5: Schematic representation of the Jupiter-Io system. From Hoyng (1994)

$$a \equiv \frac{\phi_{disc}}{2\pi c I} = constant \tag{6.2}$$

On the other hand: $V_{\rm coil} = -L \cdot \dot{I}/c^2$, where L is the self-inductance of the coil. The result is: $L\dot{I} = c^2(a\Omega - R)I$. If $\Omega > R/a$ (a critical value), a spontaneous fluctuation δI will grow exponentially. The direction of the fluctuation determines the direction of I. The lorentz force acts as a breaking torque. Angular momentum balance along the vertical axis $\vec{e_z}$, requires $T\dot{\Omega} = M + M_{\rm L}$ (T = moment of inertia; M = driving torque).

$$M_L = \vec{e_z} \cdot \int \vec{r} \times \vec{f}_l \, d\tau = \tag{6.3}$$

$$\vec{e_z} \cdot \int \vec{r} \times (\frac{\vec{J} \times \vec{B}}{c}) d\tau = \tag{6.4}$$

$$\frac{\vec{e_z}}{c} \cdot \int [(\vec{r} \cdot \vec{B})\vec{J} - (\vec{r} \cdot \vec{J})\vec{B}] d\tau = \tag{6.5}$$

$$-\frac{1}{c}\int rJB d\tau = -\frac{1}{c}\int Br dr \int J d\sigma = \tag{6.6}$$

$$-\frac{BI}{c}\int r dr = \frac{I}{2\pi c}\phi_{disc} = -aI^2 \tag{6.7}$$

In the derivation above, $\vec{r} \cdot \vec{B} = 0$ by design, hence $T\dot{\Omega} = M - aI^2$. In the stationary state we have $\dot{I} = \dot{\Omega} = 0$, whence $\Omega = R/a$ and $I = \sqrt{M/a}$. Increasing the driving torque enhances I but does not increase Ω! The power $M\Omega$ equals the electrical power I^2R dissipated in R (friction is neglected).

The disc dynamo shows us two important properties which also apply to astrophysical dynamos. The first is the differential motion, here concentrated in the sliding contacts, and second, reflectionally asymmetric motion. If we reverse the direction of Ω the current will drop to zero. The correct sense of rotation is determined by the sense of the winding of the coil. It is this asymmetry that is not a insignificant detail but an important aspect of the system. There are also some important difference with astrophysical dynamos. The disc dynamo allows only two possible directions for current to flow. In astrophysical

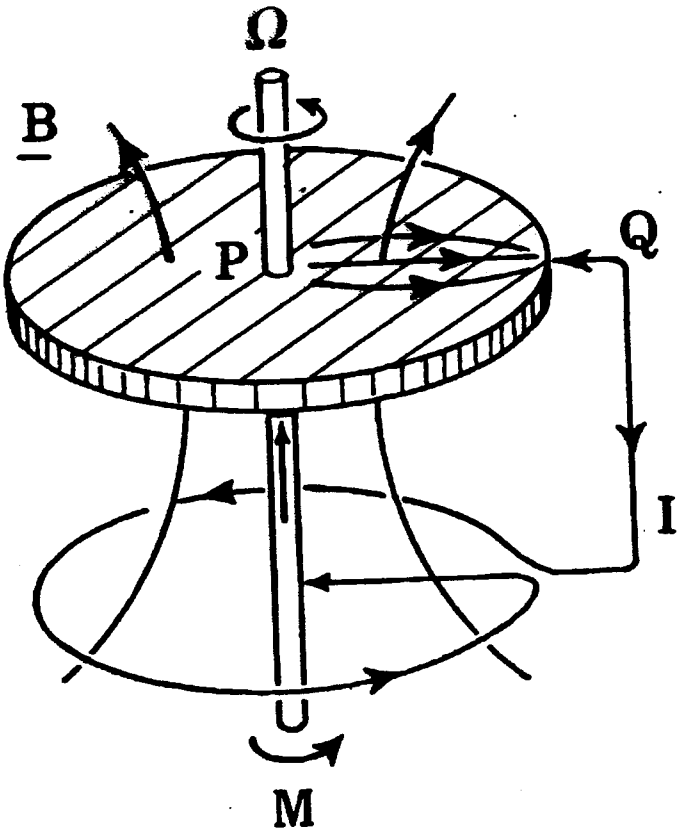


Figure 6.6: The homopolar disc dynamo. From Bullard (1955)

objects the currents can flow over spheres or spherical shells, and have no preferred path. This suggests that a dynamo in a simply-connected body, with no preffered current path, can only exist if the velocity field is sufficiently asymmetric. Also the Reynolds number $R_{\rm m}$ is small in the disc dynamo, which is not the case for most astrophysical objects. Field lines are frozen in objects with large $R_{\rm m}$ where have assumed the field to slip unaffected through the rotating disc.

The importance of the absence of symmetry for a dynamo is reinforced by Cowling's famous theorem. According to this theorem it is impossible to maintain a stationary axisymmetric magnetic field by dynamo action. To prove this, we split $\vec{B}$ in a meridional component $B_{\rm m}$ and an azimuthal component B_ϕ: $B = B_{\rm m} + B_\phi$. In Fig. 6.7 is presented a sketch of $B_{\rm m}$. There must as least be one line, L, for which $B_{\rm m} = 0$. The current density $\vec{J}$ has a nonzero component J_ϕ along L since $\vec{\nabla} \times B_{\rm m} \neq 0$ there. However, $J_\phi \neq 0$ is impossible for a stationary dynamo since:

$$\oint_{\mathcal{L}} J_\phi d\ell = \oint_{\mathcal{L}} \vec{J}_\phi \cdot d\vec{\ell} = \sigma \oint_{\mathcal{L}} \vec{E} \cdot d\vec{\ell} = \tag{6.8}$$

$$\sigma \int_S \nabla \times \vec{E} \cdot d\vec{\sigma} = -\frac{\sigma}{c} \int_S \frac{\partial \vec{B}}{\partial t} \cdot d\vec{\sigma} = 0 \tag{6.9}$$

We can hereby conclude that dynamos are non-axisymmetric or non-stationary (or both). Observations show that the dynamos of the Sun and the Earth are indeed neither exactly axisymmetric nor stationary. This is now seen to be an essential feature rather

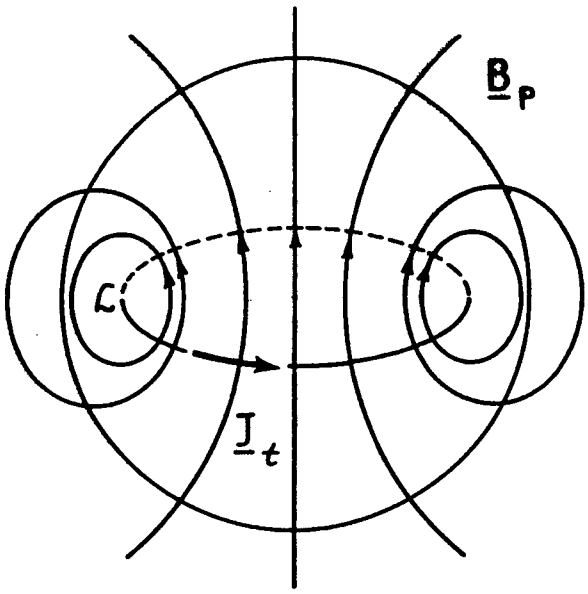


Figure 6.7: A sketch of the meridional field component B_m and a line $\mathcal{L}$ for which $B_m = 0$. From Hoyng (1992)

than an irrelevant detail. These so-called anti-dynamo theorems have shown us that it is difficult to construct a dynamo by taking simple flows $\vec{u}$ interacting with mathematically simple fields $\vec{B}$. A way to deal with these problems, without the use of large computers, is through of the mean field theory. This will be the subject of the next paragraph.

6.4 Mean field theory in a nutshell

The Sun has an extremely complex velocity field, caused by the turbulent processes in the convection zone. Magnetic field lines are twisted and turned into knots and teared into pieces. This makes it difficult to solve the set of mathematical equations that describe the problem. To get a grasp on the problem we introduce the idea of splitting the physical quantities involved into an average (large scale) component and a fluctuating component. When taking an average the fluctuations will cancel each other out and only the large scale component will remain. Averages will be indicated by $< . >$. The averages obey the following set of Reynolds rules:

$$< f + g > = < f > + < g > \tag{6.10}$$

$$< f < g >> = < f >< g > \tag{6.11}$$

$$< c > = c \tag{6.12}$$

$$< . > \quad \text{commutes with } \nabla, \partial/\partial t \text{ and } \int dt$$

In the above f and g are arbitrary functions of $\vec{r}$ and t; c is a constant. When we apply these rules to the physical quantities involved we get:

$$\vec{B} = \vec{B_0} + \vec{B_1} \qquad \vec{E} = \vec{E_0} + \vec{E_1} \tag{6.13}$$

$$\vec{J} = \vec{J_0} + \vec{J_1} \qquad \vec{u} = \vec{u_0} + \vec{u_1}, \tag{6.14}$$

where $< Q_1 >= 0$ and $< Q_0 >= Q_0$, where Q can be $\vec{B}$ or any of the others above. When we take these averages into account in the MHD equations we get this new set of equations:

$$\nabla \times \vec{E}_0 = -\frac{1}{c}\frac{\partial \vec{B}_0}{\partial t} \tag{6.15}$$

$$\nabla \times \vec{B}_0 = \frac{4\pi}{c}\vec{J}_0 \tag{6.16}$$

$$\vec{J}_0 = \sigma\{\vec{E}_0 + \frac{1}{c}\vec{u_0} \times \vec{B}_0 + \frac{1}{c} < \vec{u_1} \times \vec{B}_1 >\} \tag{6.17}$$

$$\nabla \cdot \vec{B}_0 = 0 \tag{6.18}$$

All seems natural, but something interesting has occured in the equations. The quadratic term $\vec{u} \times \vec{B}$ in Ohm's law gives rise to a new term $< \vec{u_1} \times \vec{B}_1 > \neq 0$. This is because $\vec{u_1}$ and $\vec{B}_1$ are statistically correlated. The terms $< \vec{u_0} \times \vec{B}_1 >$ and $< \vec{u_1} \times \vec{B}_0 >$ vanish.

In the process of averaging the equations there is a loss of information. This can be seen by setting $< f >= \mathcal{A}(f)$. Now from $\mathcal{A}^2(f) = \mathcal{A}(1 \cdot \mathcal{A}(f)) = [\mathcal{A}(1))(\mathcal{A}(f)] = \mathcal{A}(f)$ it follows that $\mathcal{A}^2 = \mathcal{A}$ and this makes $\mathcal{A}$ a projection operator. It is not possible to reconstruct f from $\mathcal{A}(f) = f_0$ since $\mathcal{A}$ has at least one zero eigenvalue. The average may be regarded as an average over a large ensemble of systems, each having the same $\vec{B}_0$, $\vec{J}_0$, $\vec{E}_0$ and $\vec{u_0}$, but with different realizations of $\vec{u_1}$, $\vec{B}_1$, $\vec{J}_1$ and $\vec{E}_1$. There is no relationship between $\vec{B}_0$, one ensemble member picked at random, and the field $\vec{B}$ of the dynamo. However, an average over longitude in one system also satisfies the Reynolds rules. $\vec{B}_0$ is then the large-scale, axisymmetric component of $\vec{B}$. In the next chapter, written by Mark Gieles, will be shown that for isotropic turbulence $\vec{u_1}$:

$$< \vec{u_1} \times \vec{B}_1 > \quad = \quad \alpha\vec{B}_0 - \beta\nabla \times \vec{B}_0 \quad = \quad \alpha\vec{B}_0 - \frac{4\pi\beta\vec{J}_0}{c} \tag{6.19}$$

The parameters α and β are determined by the statistical properties of $\vec{u_1}$:

$$\alpha \simeq -\frac{1}{3} < \vec{u_1} \cdot \nabla \times \vec{u_1} > \tau_c \quad ; \quad \beta \simeq \frac{1}{3} < u_1^2 > \tau_c, \tag{6.20}$$

where τ_c is the correlation time of the turbulence. In this chapter we will solely illustrate the consequences of Eq. (6.19) and give an explanation for the physics hidden in the coefficients α and β . Substitution of (6.19) into (6.17) gives rise to the following equation:

$$\vec{J}_0 = \sigma_e\{\vec{E}_0 + \frac{1}{c}\vec{u_0} \times \vec{B}_0 + \frac{\alpha}{c}\vec{B}_0\} \tag{6.21}$$

with

$$\frac{1}{\sigma_e} = \frac{1}{\sigma} + \frac{4\pi\beta}{c^2} \quad \stackrel{\sigma=\infty}{\longrightarrow} \quad \sigma_e = \frac{c^2}{4\pi\beta} \propto < \vec{u_1^2} >^{-1} \tag{6.22}$$

Equations (6.15)–(6.17) and (6.21) form a new set for the mean field. We can now eliminate $\vec{E}_0$ and $\vec{J}_0$ from the equations and come to the dynamo equation, an important result in dynamo theory.

$$\frac{\partial \vec{B}_0}{\partial t} = \nabla \times \{\vec{u_0} \times \vec{B}_0 + \alpha\vec{B}_0 - (\eta + \beta)\nabla \times \vec{B}_0\} \tag{6.23}$$

The physical meaning of this equation will now be presented.

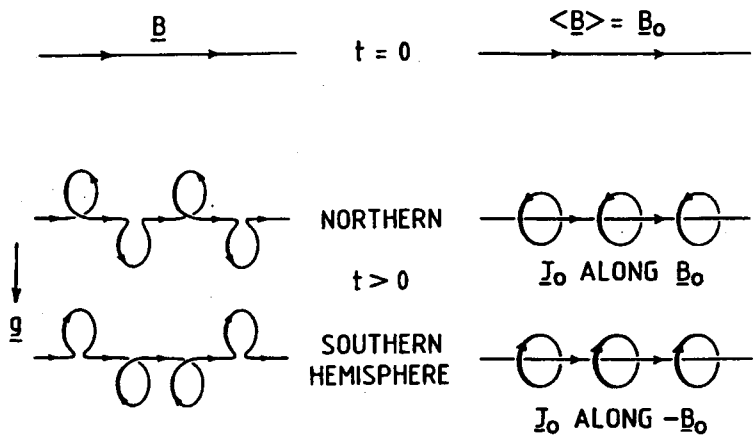


Figure 6.8: The α-effect. From Hoyng (1992)

6.4.1 About the dynamo equation

We first take a look at the first term on the right hand side of the equation. $\partial_t \vec{B}_0 = \nabla \times (\vec{u_0} \times \vec{B}_0)$. This tells us that the mean field is advected by the mean flow, just as the actual field $\vec{B}$ is advected by the actual flow $\vec{u}$. The second term gives rise to a new term, the so-called α-effect. $\partial_t \vec{B}_0 = \alpha \nabla \times \vec{B}_0$ (with α taken a constant). Consider the picture below.

Rising gas bubbles expand laterally as they adept themselves to the changing density. The Coriolis force makes the gas and the field it carries rotate clockwise on the Northern hemisphere. The rotation is anti-clockwise on the Southern hemisphere and for sinking bubbles on the Northern hemisphere. This is depicted on the left side of the picture. On the right side is shown the same process, but after taking averages. It seems new magnetic field is generated in circles on the mantle of the flux tubes rising and sinking in the convection zone. Alternatively, we may say that there is a mean current $\alpha\sigma_e \vec{B}_0/c$ along the flux tube which generates new magnetic field $\vec{B}_0$ around the tube. This is called the "α-current". The last term of the dynamo equation takes into account the effects of turbulent diffusion. Since $\beta >> \eta$, which will be proven in section 7.3.2, the mean field $\vec{B}_0$ will diffuse much faster than we would have expected from the standard MHD equations. The diffusion time of the mean field of the Sun has been reduced from $R_\odot^2/\eta \approx 4 \cdot 10^9$ years to $R_\odot^2/\beta \approx 10$ years, which is of the same order as the solar cycle. That this is not shear coincidence will be shown in the last chapter, by Mark Gieles. A visual explanation on the effects of turbulence on the diffusion of the magnetic field is shown in Fig. 6.9.

The turbulence causes the field lines to get entangled. After averaging the mean field has spread over a considerable volume, and thus even if σ is infinite, the turbulence will cause $\vec{B}_0$ to behave as if it were subject to a finite effective conductivity. The dynamo equation apparently gives a good description of the global magnetic field of the Sun and planets. An (old) numerical solution for the sun is presented in Fig. 6.10.

The solutions of this axisymmetric solar model clearly show the movement of tubes or waves of toroidal mean field to the equatorial plane and the changes of polarity. The functions α and β have been adapted to make a good fit with the well known 22 year period of the solar cycle. The toroidal field component of $\vec{B}_0$ is much larger than the poloidal component. This provides a basis for explaining the butterfly diagram. The idea is that this toroidal field occasionally breaks through the surface by buoyancy. An

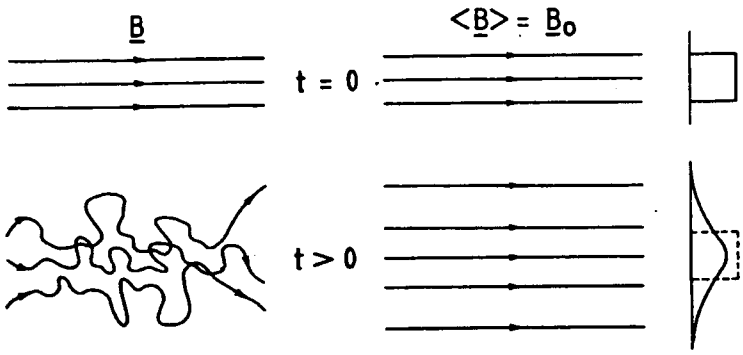


Figure 6.9: A sketch of the effects of turbulent diffusion on the magnetic field.

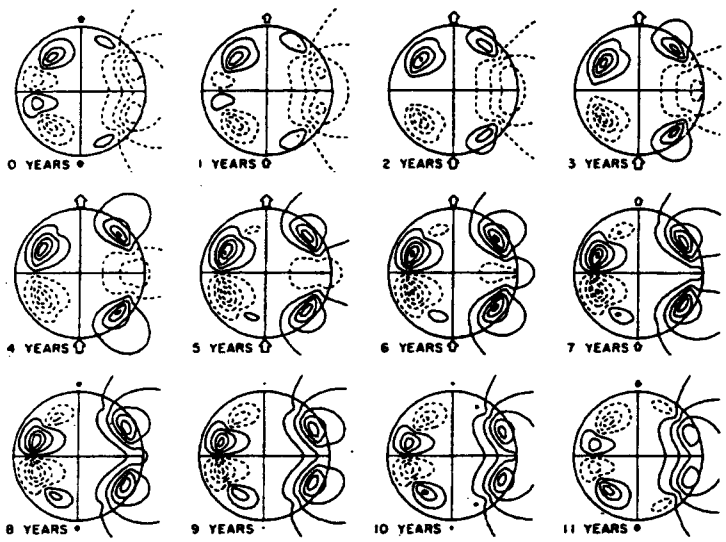


Figure 6.10: Numerical solution of an axisymmetric solar $\alpha\Omega$-dynamo. The shear is radial with $\partial\Omega_r/\partial r < 0$, and $\alpha \propto \cos\theta$. Each frame is a meridional cross section through the Sun. To the left are contours of constant toroidal mean field; on the right are the field lines of the poloidal mean field. Solid curves indicate toroidal field pointing out of the figure and clockwise poloidal field lines. The field at the poles is indicated by vertical arrows. From Stix (1976).

intuitive explanation of this α Ω-dynamo system will be presented in the last paragraph of this chapter.

6.5 A physical explanation for the solar $\alpha\Omega$-dynamo

In the cartoon in Fig. 6.11 one may see, step by step, how the various effects will cause the toroidal flux tubes to move from high regions to the equatorial plane and eventually diminish and change sign.

a. Two toroidal flux tubes of opposite polarity in each hemisphere.

b. The α-effect generates new loops of magnetic field around these tubes. Since α as well as $\vec{B}_0$ change sign between hemispheres, this new poloidal field has the same orientation in both hemispheres.

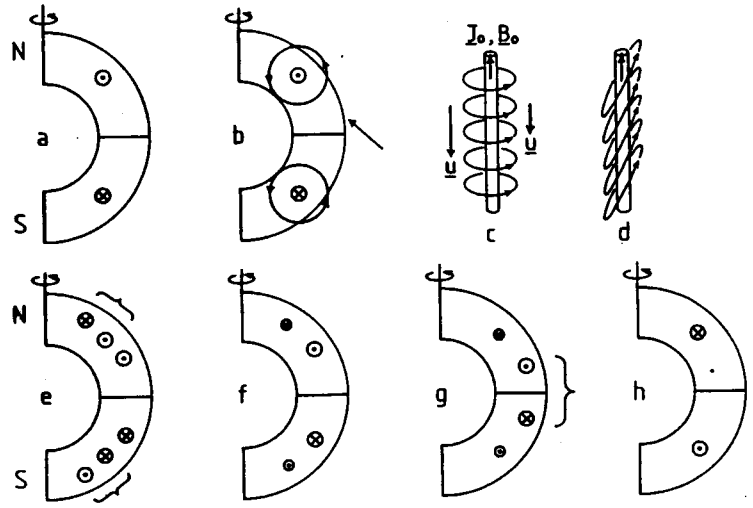


Figure 6.11: Physical explanation of the oscillatory αΩ-dynamo. From Stix (1976)

c. We now observe the system along the direction of the arrow. The differential rotation makes the loops tilt resulting in D.

e. Turbulent diffusion causes the tubes indicated by an } to cancel each other, but some mean field of opposite polarity is left over.

f. The original flux tubes have moved to the equator, leaving a small amount of opposite field in their wake. This process continues and leads to G.

g. The flux tube reproduces itself at the equator side while destroying itself on the poleward side. The destruction is incomplete, leading to a growing wake field of opposite polarity.

h. We are back at our start position, but now with opposite polarities, and a new cycle will commence.

6.5.1 Earth's magnetic field

It is currently assumed that the magnetic field of our planet is generated in an analogous way. The explanation of Fig. 6.12 assumes that there is no differential rotation in the Earth, and that the α-effect generates the poloidal field from the toroidal field, but also toroidal from poloidal.

In the first frame we see the two tubes of toroidal mean field, oppositely oriented on both hemispheres. The α-effect drives a mean α-current J_0 along these tubes, which has the same direction in both hemispheres (opposite on the sun), since $\alpha \propto \cos\theta$. The toroidal α-current generates its own poloidal mean field. This field extends outward and it is this field we observe at the Earth's surface. The α-effect comes into effect for a second time now, generating a mean α-current along those poloidal field lines which are closed inside the sphere. This poloidal mean current is the source of the toroidal field, the field we started of with. The field thus generated is self sustaining, and would even increase if turbulent diffusion wouldn't damp it. The damping allows it to become constant. The dynamo equation apparently possesses stationary, axisymmetric solutions $\vec{B}_0$. This is not

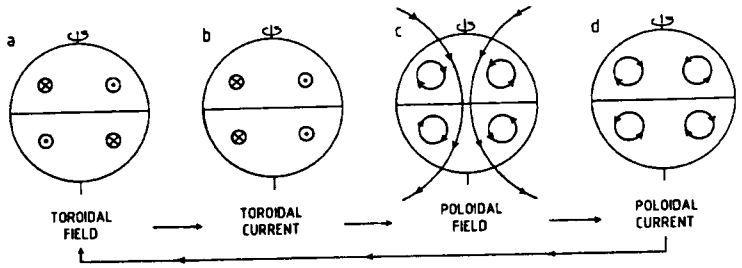


Figure 6.12: Physical explanation for the observed magnetic field on earth. From: Hoyng 1992

in contradiction with Cowling's theorem, since that applies to $\vec{B}$, but not to the mean field $\vec{B_0}$. The actual field B does not need to be axisymmetric at all, even if $\vec{B_0}$ is, so that there is no contradiction. The mean field theory supplies us with a powerful way to simplify the problem. The price we pay for it, however, is a (huge) loss of information.

Chapter 7

Dynamo issues

Lecturer P. Hoyng, notes by M. Gieles

This chapter contains a lot of mathematical derivations, therefore I created Sect. 7.7 which is an appendix with formules to which I will refer now and then to keep the derivation surveyable. Also I include Sect. 7.6 with mathematical steps which I will omit in the text with the same intention, to keep it all surveyable. In this section I will use the same symbols as in the text, so it easy to look some steps up and follow the mathematical derivations. The main reference to this chapter is Hoyng (1992).

7.1 Introduction

In this chapter we are going to have a closer look at the dynamo equation:

$$\partial_t \vec{B}_0 = \vec{\nabla} \times \{(\vec{u}_0 \times \vec{B}_0 + \alpha \vec{B}_0) - (\eta + \beta)\vec{\nabla} \times \vec{B}_0\} \tag{7.1}$$

It was derived in the previous chapter by Jacco Sloover, so the equation is the starting point for further analyses. In Sect. 7.2 we will derive the plane wave solution of the dynamo equation. In Sect. 7.3 we look at a specific limit of the solution, which is the most probable solution of plane dynamo waves in the sun. Sect. 7.4 consists of the mathematical proof of $< \vec{u}_1 \times \vec{B}_1 >= \alpha \vec{B}_0 - \beta \vec{\nabla} \times \vec{B}_0$. This identity has been derived intuitively without mathematical proof in the previous chapter and it will be shown that there is a mathematical proof for it

7.2 Plane wave solution for the dynamo equation

To find the plane wave solution of the dynamo Eq. (7.1) we consider a homogeneous infinite space filled with turbulent gas, such that α and β are constant. The mean flow $\vec{u_0}$ is in the y-direction and taken as a lineair function of x and z so that we can write $\vec{u_0} = u_0(x,z)\vec{e_y}$ and define the constant vector $\vec{a} \equiv \nabla u_0$ (homogenity), see Fig. 7.1. Axisymmetry as in Fig. 6.10 implies now invariance for translation along the y-axis, which is the same as saying that $\partial/\partial y = 0$. This invariance also makes further calculations easier if we choose for $\vec{B_0}$ the gauge

$$\vec{B}_0 = \vec{\nabla} \times (P\vec{e}_y) + T\vec{e}_y = -\vec{e}_y \times \nabla P + T\vec{e}_y, \tag{7.2}$$

where we made use of Eq. (7.55) and take P and T as functions of x, z and t. With this gauge $\nabla \cdot \vec{B}_0 = 0$ and $\vec{B}_0$ is split into its toroidal ($\parallel \vec{e}_y$) and poloidal ($\perp \vec{e}_y$) components.

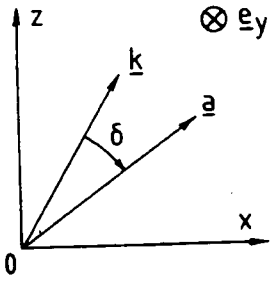


Figure 7.1: Co-ordinate system for analysing the plane wave solutions of the dynamo Eq. (7.1). $\vec{e}_y$ = unit vector along y-axis, $\vec{u}_0 = u_0(x,z)\vec{e}_y$, $\vec{a} = \vec{\nabla}u_0$, $\vec{n} = \vec{k}/k$, δ = angle between $\vec{k}$ and $\vec{a}$.

Now we have a useful expression for $\vec{B}_0$ and insert this expression Eq. (7.2) into the dynamo Eq. (7.1). With the use of Eq. (7.42) – (7.45) we get two differential equations for P and T:

$$\frac{\partial P}{\partial t} = \alpha T + \beta\vec{\nabla}^2 P \tag{7.3}$$

$$\frac{\partial T}{\partial t} = (\vec{e}_y \times \vec{a}) \cdot \vec{\nabla}P - \alpha\vec{\nabla}^2 P + \beta\vec{\nabla}^2 T \tag{7.4}$$

Here we see that new poloidal field can be created from T by α-effect and that it is subjected to turbulent diffusion by the $\beta\vec{\nabla}^2 P$-term. New toroidal field can be created from P by shear flow by the first term on the right hand side of Eq. (7.4), the 'Ω-term', and by α-effect, and it is also subjected to turbulent diffusion by the $\beta\vec{\nabla}^2 T$-term. From Eq. (7.3) we see that that $P \downarrow 0$ if α equals zero by turbulent diffusion, and, subsequently, also $T \downarrow 0$ according to (7.4). However this does not mean that there is no generation of magnatic field: the *mean field* is zero, but, in general, $\vec{B}_1$ is not. Because the terms $(\vec{e}_y \times \vec{a}) \cdot \vec{\nabla}P$ and $\alpha\vec{\nabla}^2 P$ may have widely different relative magnitude, we define three limits for these magnitude, leading to three different dynamos with their own properties:

1. $\alpha\Omega$ – *dynamo* : the α-term is much smaller than the Ω term in Eq. (7.4) and can be ignored. One α-term and the Ω-term remain in (7.3) and (7.4). These dynamos tend to have periodic behaviour and the solar dynamo is believed to be of the $\alpha\Omega$-type.
2. α^2 – *dynamo* : the Ω-term is much smaller than the α-term in Eq. (7.4) and can be ignored, so that the two α-terms remain. This is the same as ignoring the term $\vec{u}_0 \times \vec{B}_0$ in Eq. (7.1). These dynamos often have non-oscillating, stationary solutions. The dynamo of the Earth could be an α^2-dynamo.
3. $\alpha^2\Omega$–*dynamo* : the shear term and the α-term in Eq. (7.4) are of comparible magnitude.

7.3 $\alpha\Omega$-Dynamo waves in the sun

7.3.1 Plane wave solution in the $\alpha\Omega$-limit

Because the solar dynamo is believed to be an $\alpha\Omega$-dynamo, it is interesting to have a closer look at the plane wave solutions of this approximation to get more insight in the physics

of the solar dynamo. We will write P and T as a standard plane wave

$$(P,T) = (P_0,T_0)\exp\{i(\vec{k}\cdot\vec{r}-\omega t)\} \tag{7.5}$$

and insert it in Eq. (7.3) and Eq. (7.4). If we ignore the term $\alpha\vec{\nabla}^2 P$ and write $(\vec{e}_y\times\vec{a})\cdot\vec{k} = -\vec{e}_y\cdot(\vec{n}\times\vec{a})k = -(a\sin\delta)k \equiv -sk$, we can write the result in matrix notation:

$$\begin{pmatrix} i\omega-\beta k^2 & \alpha \\ -iks & i\omega-\beta k^2 \end{pmatrix}\begin{pmatrix} P_0 \\ T_0 \end{pmatrix} = 0 \tag{7.6}$$

To find a nontrivial solution to this equation, the determinant of the 2×2 matrix has to vanish, leading to the dispersion relation :

$$\omega^2 + 2i\beta k^2\omega - (\beta k^2)^2 - i\alpha ks = 0 \tag{7.7}$$

from which we find for ω :

$$\omega = -i\beta k^2 \pm \sqrt{i\alpha ks} = -i\beta k^2 \pm (1+i)\sqrt{\alpha ks/2} \tag{7.8}$$

From this we can conclude that :

$$\text{frequency } \Omega = \mathrm{Re}\,\omega = \pm\sqrt{\alpha ks/2} \tag{7.9}$$

$$\text{growth rate } \Gamma = \mathrm{Im}\,\omega = -\beta k^2 \pm \sqrt{\alpha ks/2} \tag{7.10}$$

The lower sign solution gives a wave that is always damped. This one is not interesting, unless one is interested in an initial value problem. From Eq. (7.9) and Eq. (7.10) we see that a permenant wave is possible with the right choice of α and β. If we do this analysis for the α^2 limit, we find :

$$\Omega = \mathrm{Re}\,\omega = 0; \qquad \Gamma = \mathrm{Im}\,\omega = -\beta k^2 \pm \alpha k \tag{7.11}$$

This is a purely growing or decaying solution and nonperiodic since Re $\omega = 0$.

7.3.2 The $\alpha\Omega$-dynamo in the Sun

With the results of Sect. 7.3.1 we are able to apply this to the solar dynamo and try to explain the properties of mean field models such as the one shown in Fig. 6.10.We can determine the period of the waves with Eq. (7.9) and taking $s = a\sin\delta \approx \Delta u_0/R_\odot \approx \Delta\Omega_r$ and $k \approx 1/R_\odot$:

$$P_d = \frac{2\pi}{\Omega} = 2\pi\left\{\frac{\alpha ks}{2}\right\}^{-1/2} \approx 2\pi\left\{\frac{R_\odot}{\alpha\Delta\Omega_r}\right\}^{1/2} \tag{7.12}$$

Here $\Delta\Omega_r$ is the magnitude of the differential rotation, say, the difference between equatorial and polar rotation. Taking $P_d = 22$ years and $\Delta\Omega_r \simeq 6\times 10^{-7}\mathrm{s}^{-1}$, it follows that $\alpha \approx 10$ cm s^{-1}, which is of the same order of magnitude as found in numerical models. Because in the sun the wave is marginally stable, we have to require that the growth rate Γ has to be zero. With Eq. (7.10) we find $\beta k^2 = (\alpha ks/2)^{1/2}$, or $R_\odot/\beta \approx P_d/2\pi$. This means that the turbulent diffusion timescale, defined in the previous chapter as $R_\odot/\beta$, is of the order of the period of the dynamo. From the values we have so far we find $\beta \approx 4\times 10^{13}$ cm^2 s^{-1}. Numerical models of the solar dynamo based on Eq. (7.1) usually find $\beta \approx 10^{13}$

$cm^2\ s^{-1}$. A slightly smaller value is inferred from the observed surface diffusion of solar magnetic fields. Finally, $\eta \approx 2 \times 10^4\ cm^2\ s^{-1}$ in the convection zone near $T = 10^6 K$, confirming that $\beta \gg \eta$. Note that the frequency of the wave is given by $\Omega = (\alpha k s/s)^{1/2}$ and that from marginal stability follows that $(\alpha k s/2)^{1/2} = \beta k^2$. One could say $\Omega = \beta k^2$, but this merely expresses marginal stability. The difference is very important, for example, when we determine the group velocity of the dynamo wave. We will now look how $B_{\rm pol}$ and $B_{\rm tor}$ are related to each other and to $\vec{B}_0$. Therefore we start with inserting Eq. (7.5) in Eq. (7.2) :

$$\vec{B}_0 = (-iP_0\vec{e}_y \times \vec{k} + T_0\vec{e}_y)\exp(i\psi), \tag{7.13}$$

where $\psi = \vec{k}\cdot\vec{r} - \omega t$ is the phase. We see it is a transverse wave, because $\vec{k}\cdot\vec{B}_0 = 0$. From the first line of Eq. (7.6) we obtain

$$\frac{P_0}{T_0} = \frac{\alpha}{\beta k^2 - i\omega} = \frac{\alpha}{2\Omega}(1+i) \tag{7.14}$$

At the last step we assumed marginally stability of the wave. Inserting this in Eq. (7.13) and taking the real part gives us

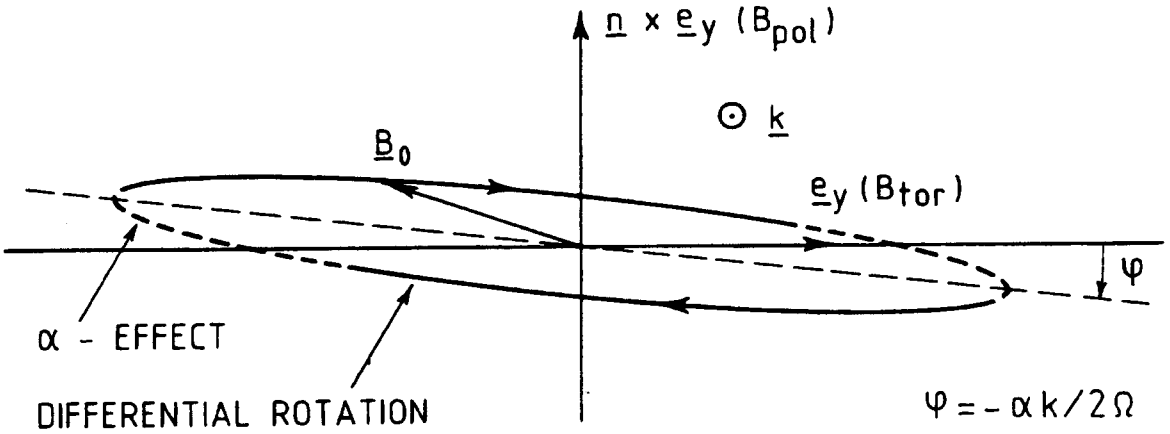


Figure 7.2: An $\alpha\Omega$-dynamo wave. The vector $\vec{B}_0$ runs over a narrow ellipse $\perp \vec{k}$, clockwise for $\alpha > 0$ and anticlockwise for $\alpha > 0$. Differential rotation sweeps $\vec{B}_0$ rapidly from one edge to the other (solid line) and does work , so that $\vec{B}_0$ decreases and then increases again. The α-effect moves $\vec{B}_0$ very slowly around the edge (broken line), after which differential rotation takes over again. During most of the time $\vec{B}_0$ is (anti)parallel to $\vec{e}_y$.

$$\vec{B}_0 \propto \frac{\alpha k}{\Omega\sqrt{2}}\cos(\psi + 3\pi/4)\vec{n} \times \vec{e}_y + \cos\psi\vec{e}_y \tag{7.15}$$

By eliminating ψ we find that $\vec{B}_0$ rotates over a tilted ellipse, see Fig. 7.2. This tilt angle, $\varphi = -\alpha k/2\Omega$ is very small as $\alpha k/2\Omega \approx \alpha P_{\rm d}/4\pi R\odot \approx 10^{-2}$. The relative magnitude of the poloidal field $B_{\rm pol}$ and the toroidal field $B_{\rm tor}$ follows from (7.15):

$$\frac{B_{\rm pol}}{B_{\rm tor}} = \frac{\alpha k}{\Omega\sqrt{2}} \approx 10^{-2} \tag{7.16}$$

The ellipse in Fig. 7.2 is therefore very flat; the toroidal field is about a factor 100 stronger than the poloidal field, a property which also follows from numerical models. Finally we compute the group velocity of the wave:

$$\vec{v}_{\rm g} = \frac{\partial \Omega}{\partial \vec{k}} = \frac{\partial}{\partial \vec{k}}(\alpha \vec{e}_{\rm y} \cdot \vec{k} \times \vec{a}/2)^{1/2} = \frac{\alpha}{4\Omega}\vec{a} \times \vec{e}_{\rm y} \tag{7.17}$$

A wave packet propagates perpendicular to $\vec{e}_{\rm y}$ and $\vec{a}$, along the isoplanes of the 'rotation' $\vec{u}_0$, as was first proven by Yoshimura (1975). See Fig. 7.1 for the geometry. In Fig. 6.10, $\vec{a}$ points inwards and since $\alpha \propto \cos\theta$, $\vec{v}_{\rm g}$ is directed to the equator in both hemispheres, thus explaining the migration of the mean field towards the equator in that model. We can make an approximation for the magnitude of $\vec{v}_{\rm g}$: $v_{\rm g} = \alpha a/4\Omega \approx \alpha \Delta\Omega_{\rm r} P_{\rm d}/8\pi \approx 160$ cm s^{-1}, or 5 $\times 10^5$ km in 11 year, about 50° in latitude, which is of the right order of magnitude.

7.4 Rigorous mean field theory

In this section we are going to prove that $< \vec{u}_1 \times \vec{B}_1 > = \alpha \vec{B}_0 - \beta \vec{\nabla} \times \vec{B}_0$ by means of a rather mathematical derivation. Especially here the reader will need Sect. 7.6 to follow the derivations. The notation here is the one explaned earlier in Sect. 6.4. We will have a closer look at equations of the form:

$$\partial_{\rm t} f = (A_0 + A_1(t))f \tag{7.18}$$

where A_0 is time-independent, but A_1 fluctuates, with zero average. Here f can be e.g. $\vec{B}$. Now we write f as :

$$f = e^{A_0 t} u \tag{7.19}$$

Take the derivative of f and isolate $\dot{u}$:

$$\dot{u} = e^{-A_0 t} A_1 e^{A_0 t} u \equiv \tilde{A}_1 u \tag{7.20}$$

Then integrate this expression to get u, see Eq. (7.46) and Eq. (7.47) for the detailed derivation. From now we will write $f^\tau \equiv f(\tau)$ and $a = u(0)$:

$$u = a + \int_0^t {\rm d}\tau \tilde{A}_1^\tau \left(a + \int_0^\tau {\rm d}\sigma \tilde{A}_1^\sigma u^\sigma \right) \tag{7.21}$$

Take the time average, indicated with $<>$, differentiate with respect to t and use $< \tilde{A}_1 >= 0$:

$$\partial_{\rm t} < u >= \int_0^t {\rm d}\sigma < \tilde{A}_1^t \tilde{A}_1^\sigma u^\sigma > \tag{7.22}$$

While u^σ evolves on a timescale $|\tilde{A}_1|^{-1}$, which is assumed to be much longer than the timescale $\tau_{\rm c}$ over which $\tilde{A}_1$ changes itself : $|\tilde{A}_1|\tau_{\rm c} \simeq |A_1|\tau_{\rm c}| \ll 1$, where $|A_1|$ is the magnitude of A_1. If we consider u^σ and $\tilde{A}_1$ as uncorrelated and split the $<>$ -term in two parts, we make a relative error order $|A_1|\tau_{\rm c} \ll 1$. If we now change the integration variable : $\sigma = t - \tau$, we get :

$$\partial_{\rm t} < u >= \int_0^\infty {\rm d}\tau < \tilde{A}_1^t \tilde{A}_1^{t-\tau} >< u^{t-\tau} > \tag{7.23}$$

We can let the integral run from zero to infinite, because $< >$ is zero beyond $|t-\tau| > \tau_c$. Now we approximate :

$$< u^{t-\tau} >=< u^t > -\tau\partial_t < u^t > \simeq < u^t > \tag{7.24}$$

This approximation is permitted because with Eq. (7.23) we can estimate $\tau\partial_t < u^t > \approx \tau_c \cdot \tau_c|A_1|^2 < u^t >= (|A_1|\tau_c)^2 < u^t >$, so we make a relative error of order $(|A_1|\tau_c)^2 \ll 1$. Now we have succeeded to reduce Eq. (7.23) to a differential equation:

$$\partial_t < u >= \left(\int_0^\infty \mathrm{d}\tau < \tilde{A}_1^t \tilde{A}_1^{t-\tau} >\right) < u > \tag{7.25}$$

From here it is just a lot of algebra from which only the important steps will be mentioned here, again with references to the derivations in Sect. 7.6. First substitute Eq. (7.19) in Eq. (7.25):

$$\partial_t < f >= \left(A_0 + \int_0^\infty \mathrm{d}\tau < {A_1}^t e^{A_0\tau} {A_1}^{t-\tau} > e^{-A_0\tau}\right) < f > \tag{7.26}$$

Ignore $e^{\pm A_0\tau}$, since $|A_0|\tau_c \ll 1$ is assumed, and use Eq. (7.48) – (7.51) and the substitution $f = \vec{B}_0$ and $\vec{A}_1 = \vec{\nabla} \times (\vec{u}_1 \times \cdot)$ to get an expression for $< \vec{u}_1 \times \vec{B}_1 >$:

$$< \vec{u}_1 \times \vec{B}_1 >= \int_0^\infty \mathrm{d}\tau < \vec{u}_1^t \times \vec{\nabla} \times \vec{u}_1^{t-\tau} > \times \vec{B}_0 \tag{7.27}$$

Assume incompressibility $\vec{\nabla} \cdot \vec{u}_1 = 0$ and use Eq. (7.57) and Eq. (7.52) to get for the i^{th} component of $< \vec{u}_1 \times \vec{B}_1 >$ (surpress index 1 on u temporarily):

$$\begin{aligned} < \vec{u}_1 \times \vec{B}_1 >_i &= \varepsilon_{ikl} \int_0^\infty \mathrm{d}\tau \left(< {u_k}^t B_{0,s} \nabla_s \, {u_l}^{t-\tau} > - < {u_k}^t {u_s}^{t-\tau} \nabla_s \, B_{0,l} > \right) && (7.28) \\ &\equiv \alpha_{is} B_{0,s} + \beta_{isl} \nabla_s \, B_{0,l} && (7.29) \\ & && (7.30) \end{aligned}$$

with

$$\begin{aligned} \alpha_{is} &= \varepsilon_{ikl} \int_0^\infty \mathrm{d}\sigma < {u_k}^t (\nabla_s {u_l}^{t-\tau}) > && (7.31) \\ \beta_{isl} &= -\varepsilon_{ikl} \int_0^\infty \mathrm{d}\tau < {u_k}^t {u_s}^{t-\tau} > && (7.32) \end{aligned}$$

Assume isotropy, so the tensor of the third step in Eq. (7.31) is invariant for rotation :

$$< {u_k}^t (\nabla_s \, {u_l}^{t-\tau}) >= \text{const} \cdot \varepsilon_{ksl} \tag{7.33}$$

so with Eq. (7.58):

$$\varepsilon_{ksl} < {u_k}^t (\nabla_s \, {u_l}^{t-\tau}) >= \text{const} \cdot \varepsilon_{ksl}\varepsilon_{ksl} = 6 \cdot \text{const} \tag{7.34}$$

Hence (restoring index 1) :

$$\text{const} = \frac{1}{6} < \vec{u}_1^t \cdot \vec{\nabla} \times \vec{u}_1^{t-\tau} >, \tag{7.35}$$

and

$$< u_k{}^t(\nabla_s\ u_l{}^{t-\tau}) >= \frac{1}{6} < \vec{u}_1^t \cdot (\vec{\nabla} \times \vec{u}_1^{t-\tau}) > \varepsilon_{ksl} \tag{7.36}$$

Similarly to this we find, using Eq. (7.59) :

$$< u_k{}^t u_s{}^{t-\tau} >= \frac{1}{3} < \vec{u}_1^t \cdot \vec{u}_1^{t-\tau} > \delta_{ks} \tag{7.37}$$

Now we have to substitute these results in Eq. (7.31) and Eq. (7.32) and use Eq. (7.53) and Eq. (7.54) to find:

$$\alpha_{is} \;=\; \alpha \cdot \delta_{is} \quad \text{with} \quad \alpha \equiv -\frac{1}{3} < \vec{u}_1 \cdot \vec{\nabla} \times \vec{u}_1 > \tau_c \tag{7.38}$$

and

$$\beta_{isl} \;=\; -\beta \cdot \varepsilon_{isl} \quad \text{with} \quad \beta \equiv \frac{1}{3} < \vec{u}_1^2 > \tau_c \tag{7.39}$$

If we insert this in Eq. (7.29) we find what we were looking for :

$$\begin{aligned} < \vec{u}_1 \times \vec{B}_1 >_i \;&=\; \alpha \cdot \delta_{is} B_{0,s} - \beta \varepsilon_{isl} \nabla_s B_{0,s} \\ &=\; \left(\alpha \vec{B}_0 - \beta \vec{\nabla} \times \vec{B}_0\right)_i \end{aligned} \tag{7.40}$$

so

$$< \vec{u}_1 \times \vec{B}_1 >= \alpha \vec{B}_0 - \beta \vec{\nabla} \times \vec{B}_0 \tag{7.41}$$

This result has been derived under the following conditions:
1). $\tau_c \ll \lambda_c / u_1$, since $|A_0| \sim u_1/\lambda_c$
2). $|A_0|\tau_c \ll 1 \longrightarrow \Omega_{\rm rot}\tau_c \ll 1$
3). turbulence is incompressible ($\vec{\nabla} \cdot \vec{u}_1 = 0$) and isotropic.
None of these conditions is fulfilled in the solar convection zone! We could remove conditions 2) and 3), with as result that α and β become tensors. However condition 1) is essential and cannot be removed. This is a long standing open problem. The fact that simple scalar expressions for α and β work rather well, suggests that the general form of the expression $< \vec{u}_1 \times \vec{B}_1 >= \alpha \vec{B}_0 - \beta \vec{\nabla} \times \vec{B}_0$ remains more ore less as it is.

7.5 Current status

There is still much research being done to the solar dynamo. We can roughly split the research in three subjects.

7.5.1 Interface-dynamo

The idea of these kind of dynamos is that they are believed to operate in subadiabatic overshoot-tachocline regions. The field is believed to be at the base of the convection zone and not in the convection zone itself. The original idea came from Parker (1993) and typical examples can be found in MacGregor and Charbonneau (1997).

7.5.2 Storage, instability and eruption of magnetic flux

Starting point for explaining the solar magnetic fields here is not the solar dynamo, but the presents of strong fluxtubes in the overshootlayer. Models are being made to explain what happens if these tubes get instable and rise through the convection zone and create active regions. The original work started in Utrecht (van Ballegooijen 1982).

7.5.3 Surface distribution of flux

In this research they just look at the distribution of the vertical flux tubes emerging from the sun. A typical example of this research can be found in Wang and Sheeley (1994).

7.6 Mathematical steps belonging to previous derivations

$$\begin{aligned}\vec{\nabla}\times(\vec{u_0}\times\vec{B_0}) &= (\vec{B_0}\cdot\vec{\nabla})\vec{u_0}-(\vec{u_0}\cdot\vec{\nabla})\vec{B_0}=\vec{e_y}(\vec{B_0}\cdot\vec{\nabla})\vec{u_0}\\ &= \vec{e_y}(\vec{B_0}\cdot\vec{a})=-\vec{e_y}(\vec{a}\cdot\vec{e_y}\times\vec{\nabla}P)\\ &= \vec{e_y}(\vec{e_y}\times\vec{a})\cdot\vec{\nabla}P \qquad (7.42)\end{aligned}$$

$$\begin{aligned}\alpha\vec{\nabla}\times\vec{B_0} &= \alpha\vec{\nabla}\times(-\vec{e}_y\times\vec{\nabla}P)+\alpha\vec{\nabla}\times(T\vec{e}_y)\\ &= \alpha\vec{\nabla}\times(\vec{\nabla}P\times\vec{e}_y)+\alpha\vec{\nabla}\times(T\vec{e}_y)\\ &= -\vec{e}_y(\alpha\vec{\nabla}^2P)+\alpha\vec{\nabla}\times(T\vec{e}_y) \qquad (7.43)\end{aligned}$$

Here we used vector operation Eq. (7.56) from which the first three terms on the righthand side equal zero in this case.

$$\begin{aligned}\beta\vec{\nabla}\times(\vec{\nabla}\times\vec{B_0}) &= \beta\vec{\nabla}\times\vec{\nabla}\times(\vec{\nabla}\times(P\vec{e}_y))+\beta\vec{\nabla}\times(\vec{\nabla}\times(T\vec{e}_y)) \qquad (7.44)\\ &= -\beta\vec{\nabla}\times(\vec{\nabla}^2P\vec{e}_y)-\beta\vec{e}_y\vec{\nabla}^2T \qquad (7.45)\end{aligned}$$

To go from (7.44) to (7.45) we used vector operation (7.57).
Integrating Eq. (7.20) gives us an expression for u:

$$u = a+\int_0^t \mathrm{d}\tau\tilde{A}_1^\tau u^\tau \qquad (7.46)$$

Insert this expression for u in the integral on the right hand side to get the next iteration for u:

$$u = a+\int_0^t \mathrm{d}\tau\tilde{A}_1^\tau\left(a+\int_0^\tau \mathrm{d}\sigma\tilde{A}_1^\sigma u^\sigma\right) \qquad (7.47)$$

Assume $e^{(\pm A_0\tau_c)}\simeq 1$ and revert to $< f >= f_0$:

$$\partial_t < f >= \partial_t f_0 = \left(A_0+\int_0^\infty \mathrm{d}\tau < {A_1}^t {A_1}^{t-\tau} >\right) f_0 \qquad (7.48)$$

but also:

$$\partial_t f_0 = \partial_t < f >=< (A_0+A_1)(f_0+f_1) >= A_0f_0+ < A_1f_1 > \qquad (7.49)$$

Now we have :

$$< A_1f_1 >= \int_0^\infty \mathrm{d}\tau < {A_1}^t {A_1}^{t-\tau} > f_0 \qquad (7.50)$$

Now we can make the substitution $f = \vec{B}_0$ and $\vec{A}_1 = \vec{\nabla} \times (\vec{u}_1 \times \cdot)$ so we get the desired expression for $< \vec{u}_1 \times \vec{B}_1 >$ by omitting the first $\vec{\nabla}\times$:

$$< \vec{u}_1 \times \vec{B}_1 >= \int_0^\infty \mathrm{d}\tau < \vec{u}_1^t \times \vec{\nabla} \times \vec{u}_1^{t-\tau} > \times \vec{B}_0 \tag{7.51}$$

$$\vec{\nabla} \times (\vec{u}_1^{t-\tau} \times \vec{B}_0) = (\vec{B}_0 \cdot \vec{\nabla})\vec{u}_1^{t-\tau} - (\vec{u}_1^{t-\tau} \cdot \vec{\nabla})\vec{B}_0 \tag{7.52}$$

$$\int_0^\infty \mathrm{d}\tau < \vec{u}_1^t \cdot \vec{\nabla} \times \vec{u}_1^{t-\tau} > \simeq < \vec{u}_1 \cdot \vec{\nabla} \times \vec{u}_1 > \tau_c \tag{7.53}$$

$$\int_0^\infty \mathrm{d}\tau < \vec{u}_1^t \vec{u}_1^{t-\tau} > \simeq < \vec{u}_1^2 > \tau_c \tag{7.54}$$

7.7 Mathematical tools

$$\vec{\nabla} \times (\phi \vec{F}) = \phi \vec{\nabla} \times \vec{F} + (\vec{\nabla}\phi) \times \vec{F} \tag{7.55}$$

$$\vec{\nabla} \times (\vec{F} \times \vec{G}) = (\vec{G} \cdot \vec{\nabla})\vec{F} - (\vec{F} \cdot \vec{\nabla})\vec{G} + \vec{F}(\vec{\nabla} \cdot \vec{G}) - \vec{G}(\vec{\nabla} \cdot \vec{F}) \tag{7.56}$$

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{F}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{F}) - \vec{\nabla}^2 \vec{F} \tag{7.57}$$

$$\varepsilon_{ksl}\varepsilon_{ksl} = 3\,(\delta_{ss}\delta_{ll} - \delta_{ls}\delta_{sl}) = 6 \tag{7.58}$$

$$\varepsilon_{ikl}\varepsilon_{ksl} = -\varepsilon_{kli}\varepsilon_{kls} = -2\delta_{is} \tag{7.59}$$

Chapter 8

Solar convection and granulation

Lecturer S.K. Solanki, notes by R. van Deelen

The observed surface of the Sun is called the *photosphere*. This is the region from which the observed photons of the visible wavelength originate. Even when the Sun is quiet, i.e during a sunspot minimum, we see it changing continuously.

Magnetic fields are present all over the Sun. In the "quiet" regions, i.e outside the *active regions* (see Chapter 11), the magnetic field gets organized in a cellular pattern; the so-called *network*. The network is organized by the *supergranulation* (see Fig. 2.10) which has its origin in convective motions at intermediate depths in the solar *convection* zone. This zone occupies the outer part of the Sun to a depth of 30% of the solar radius.

The *granules*, smaller convection cells seen at the surface, also organize and move the magnetic field. Particular emphasis is placed on the description of *granulation*, the convective structures which we understand the best.

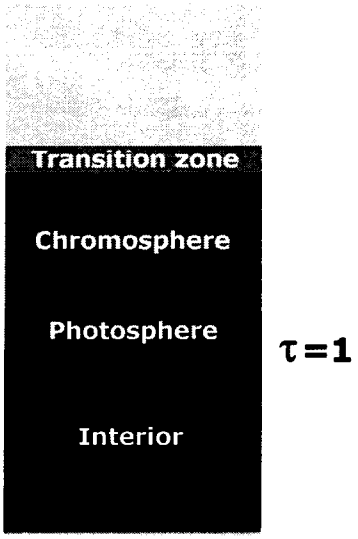


Figure 8.1: The base of the photosphere ($h = 0$ km) is located at $\tau = 1$, the optical depth in the visible continuum. It ends approximately 500 km higher up where the chromosphere starts. The transition zone is at about 2300 – 2600 km.

8.1 Idealized convection

It should be pointed out that an adequate theory for magnetoconvection to describe all the properties of the Sun does not exist, because our understanding of solar and stellar turbulence and convection is limited. Still it is the detailed properties of the turbulence (among other things) that determine what the solar dynamo, that drives the magnetic field, looks like.

We begin by looking at highly idealized convection. When does it occur? Why does the Sun have an outer convection zone and how does the convection manifest itself at the solar surface?

8.1.1 The Schwarzschild criterion

Convection will occur in layers of a star where the temperature gradient is so steep that an arbitrary perturbation (e.g., a little push upwards) of a gas parcel results in a unchecked movement of this gaseous blob.

Let us consider a parcel of gas lifted adiabatically from its equilibrium position somewhere in the solar interior. If the density of the gas in this perturbed parcel is less than the density of the surrounding gas in its perturbed location, the parcel will continue to rise due to buoyancy and expand (pressure equilibrium).

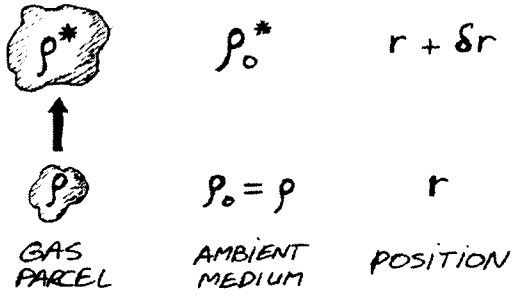


Figure 8.2: Gas parcel gets lifted up due to a buoyancy force acting on it.

So convective instability (see Fig. 8.2) occurs only when:

$$\rho^* - \rho_0^* < 0\,, \tag{8.1}$$

where ρ^* is the gas density in the parcel at the perturbed location $(r + \delta r)$.

When this condition is satisfied a buoyancy force acts on the parcel, which accelerates upwards. If the situation is stable, the force on the parcel decelerates it, finally bringing it to a halt and accelerating it back towards its original location. Thus the parcel will in general show a damped oscillatory motion.

We can make a Taylor expansion of ρ^* and ρ_0^* around ρ and $\rho_0 = \rho$, which we can restrict to first order terms for sufficiently small δr:

$$\rho^* = \rho + \left(\frac{d\rho}{dr}\right)_{\rm ad} \delta r + \mathrm{O}((\delta \mathrm{r})^2)\,, \tag{8.2}$$

$$\rho_0^* = \rho + \left(\frac{d\rho}{dr}\right) \delta r + \mathrm{O}((\delta \mathrm{r})^2)\,. \tag{8.3}$$

Expansions (8.2) of ρ^* around ρ and (8.3) of ρ_0^* around ρ_0 are now introduced into (8.1) and lead to:

$$\left[\left(\frac{d\rho}{dr}\right)_{\text{ad}} - \left(\frac{d\rho}{dr}\right)\right] \delta r < 0\,, \tag{8.4}$$

where $(d\rho/dr)_{\text{ad}}$ is the density gradient under adiabatic conditions and $d\rho/dr$ is the actual density gradient.

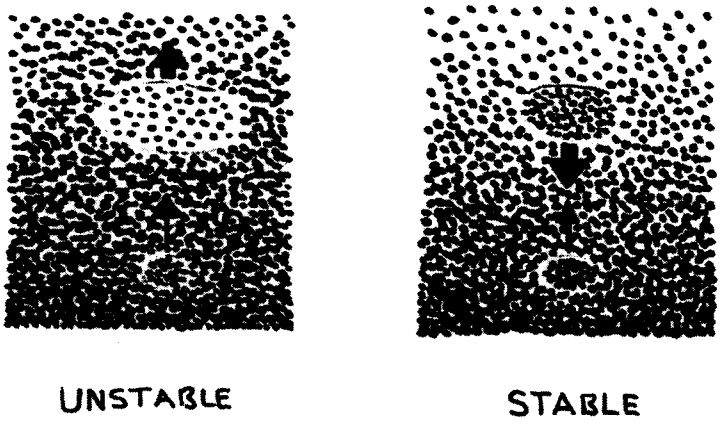


Figure 8.3: The destiny of gas parcels after a little kick; in a convectively *unstable* situation it leads to an accelerating life. For a *stably* stratified medium it follows damped harmonic oscillary motion.

To obtain the Schwarzschild criterion of convective instability it is necessary to transform the density gradients into temperature gradients using the equation of state of an ideal gas:

$$\begin{aligned} P = \rho \Re T/\mu\,, \quad &\text{with} \quad \Re = \text{gas constant} \\ &\text{and} \quad \mu = \text{mean molecular weight.} \end{aligned} \tag{8.5}$$

Imposing pressure equilibrium (Eq. (8.4)) results in:

$$\frac{dT}{dr} < \left(\frac{dT}{dr}\right)_{\text{ad}} + \frac{T}{\mu}\frac{d\mu}{dr} - \frac{T}{\mu}\left(\frac{d\mu}{dr}\right)_{\text{ad}}\,. \tag{8.6}$$

μ changes due to ionization, formation of molecules or thermonuclear fusion (evolution). Now, we're only interested in the convection zone of the Sun, which does not reach down into the solar core where fusion of hydrogen into helium is taking place. Here is $d\mu/dr = (d\mu/dr)_{\text{ad}}$, because the ionization equilibrium is instantaneously adjusted. The timescales for chemical reactions of the parcel with its surroundings are much shorter than the timescale of upward movement ($\tau_{\text{reaction}} \ll \tau_{\text{up}}$). This allows us to simplify (Eq. 8.6) to

$$\frac{dT}{dr} < \left(\frac{dT}{dr}\right)_{\text{ad}}\,, \tag{8.7}$$

which can be rewritten as

$$\nabla > \nabla_{\text{ad}}\,, \tag{8.8}$$

where

$$\nabla = \frac{d \ln T}{d \ln P}\,, \tag{8.9}$$

$$\nabla_{\rm ad} = \left(\frac{d \ln T}{d \ln P}\right)_{\rm ad}\,. \tag{8.10}$$

These conditions are sufficient (Schwarzschild 1906) and necessary (Lebovitz 1966) for convective instability if $d\mu/dr = (d\mu/dr)_{\rm ad}$ or $d\mu/dr = 0$.

Near the solar core, the μ-gradient in (Eq. 8.6) becomes important because of the nuclear fusion taking place there. The timescales for changes in μ-gradients due to nuclear reactions are much longer than the travelling timescale ($\tau_{\rm nuclear} \gg \tau_{\rm up}$).

8.1.2 Why does the Sun have an outer convection zone?

In the absence of convection the energy produced in the solar core is transported outward by radiation. Convection occurs when $\nabla_{\rm rad} > \nabla_{\rm ad}$. In the diffusion approximation of radiative transfer (excellent in an optically thick medium such as the solar interior) the resulting radiative temperature gradient reads:

$$\nabla = \nabla_{\rm rad} = \frac{3}{64\pi}\frac{\kappa\rho H_p L_\odot}{r^2\sigma T^4} \sim \frac{\kappa\rho}{T^4} \tag{8.11}$$

Here $L_\odot$ = Solar luminosity
σ = Stefan-Boltzman constant
κ = Absorption coefficient per gram
H_p = Pressure scale height

Convection occurs in layers at which, e.g. $\nabla_{\rm rad}$ is large. This is the case where κ is large. κ increases rapidly with depth in layers in which elements are partially ionized, due to the increasing number of free electrons with depth, which interact with photons extremely well. Because $\nabla = \nabla_{\rm rad} \sim \kappa$ these zones become convective ($\nabla_{\rm rad} > \nabla_{\rm ad}$). The main zones of partial ionizations in the Sun are:

Just below the solar surface	$H \rightarrow H^+$
At deeper levels (7000 km)	$He \rightarrow He^+$
Deeper still (30 000 km)	$He^+ \rightarrow He^{++}$

At greater depths H and He are completely ionized, so that κ is approximately constant. $\nabla_{\rm rad}$ decreases again, mainly due to $\nabla_{\rm rad} \sim 1/T^4$, so that at sufficiently large depth the heat can be transported by radiation (here $\nabla_{\rm rad} < \nabla_{\rm ad}$). Hence the convection stops below a certain level.

The Schwarzschild criterion divides the Hertzsprung-Russel diagrams into domains of stars with convective envelopes and stars without appreciable convection in or just below the photosphere (see Fig. 3.20). Detailed calculations on stellar interior theory reveal that the Sun and all stars with an effective temperature below about 7000 K must have a convective envelope immediately below the observable photosphere.

8.1.3 Overshooting

The solar gas becomes convectively stable just at the visible surface, so why do we see the granulation pattern at all? At the edge of the convection zone moving parcels of gas

overshoot into stable stratified gas due to their momentum (see Fig. 8.4). Although the buoyancy force now acts as a brake and brings eventually the parcel to a standstill. This only happens 100 – 400 km above the solar surface. This effect is called *overshooting.* Without this overshooting we wouldn't have seen much of the beautiful granulation!

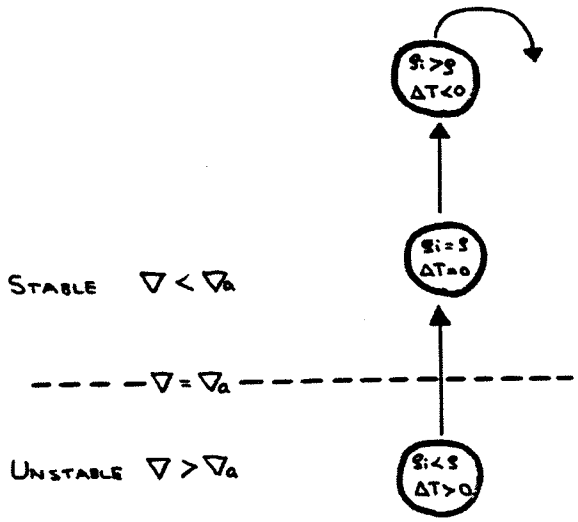


Figure 8.4: An illustration of convective overshoot, i.e. the penetration of convection into a stably stratified medium. The parcels can reach 400 km into the stably stratified photosphere but the mean is about 100 km.

8.2 Granulation and larger convective cells

Since the gas density drops exponentially with height the heated gas parcels cannot continue to rise indefinitely. In particular, when they reach the solar surface they cool due to radiative losses, making them less buoyant. Also, they enter stably stratified gas. Eventually they stop and are pulled downwards again. However, due to the continuing rise of hot gas from below they cannot flow down where they came up. They must first move to the side. An excess pressure therefore soon builds up over the upflows which drives the horizontal motions there, forming the granular velocity cells. Note that as the magnetic field lines are "frozen-in" due to the high electrical conductivity of the solar plasma, they are carried by the horizontal flows to the cell boundaries, where the magnetic flux gets concentrated (see Section 10.6.1 for this so-called *flux expulsion*).

The convection visible at the solar surface appears to exist on different scales; besides the normal *granules* there are also *mesogranules*, which are five times the "normal" granules, and the *supergranules* mentioned earlier which in turn are five times the mesogranular size. Recently (1998), even larger cells were discovered, so-called *giant cells.*
The topology of flow changes completely with depth. At the surface: strong downflows form closed network and gentle upflows in isolated patches. A few scale-heights below the surface strong downflows in isolated filaments and gentle upflows forming a more or less closed network are found. Physical reasons for the geometrical (and velocity) differences between the up and downflows of granules are sketched in Fig. 8.5.

Property		Granulation	Mesogranulation	Supergranulation
Sizes		1500 km	7000 km	25000–30000 km
Lifetimes		6–8 min	1 h	1 day (30–40 h)
$v_{\rm rms}$	vertical horizontal	1 km s^{-1} 1 km s^{-1}	60 m s^{-1} $\lesssim$ 1 km s^{-1}	$\lesssim$ 50 m s^{-1} 300 – 500 m s^{-1}
$(\delta I/I)_{\rm rms}$	continuum line core	0.1–0.2 > 0 < 0	≈ 0 ≈ 0	≈ 0 < 0 (network)
Surface topology		Isolated, bright upflowing granules, multiply connected, dark downflowing lanes	Similar to granules in velocity?	Downflows and bright network elements are isolated

Table 8.1: Observed properties of surface convective features. $(\delta I/I)_{\rm rms} > 0$ implies that upflows are bright, downflows are dark. $(\delta I/I)_{\rm rms} < 0$ means dark upflows, bright downflows.

8.2.1 Upper limit on granular size

It is possible to make a simple estimate of the largest size which granules can achieve based on the equation of continuity and buoyancy braking.
Consider a cylindrically symmetric convection cell. We assume stationairy flow. Then the equation of continuity reads:

$$0 = \nabla \cdot (\rho \vec{v}) = \frac{\partial}{\partial z}\,(\rho v_z) + \frac{1}{r}\frac{\partial}{\partial r}(r \rho v_r)\,, \tag{8.12}$$

where ρ is density, v_r is the radial velocity and v_z is the vertical velocity. Reasonable approximations are:

$$\begin{aligned} \rho &\sim \mathrm{e}^{-z/H_\rho} \ (H_\rho \text{ is the density scale height}) \\ v_z &\simeq \text{independent of } z \text{ an } r \\ v_r &\sim r \ (\text{1st order Taylor expansion around } r = 0) \end{aligned}$$

The equation of continuity then becomes:

$$\frac{\rho v_z}{H_\rho} = \frac{2 \rho v_r}{r}\,. \tag{8.13}$$

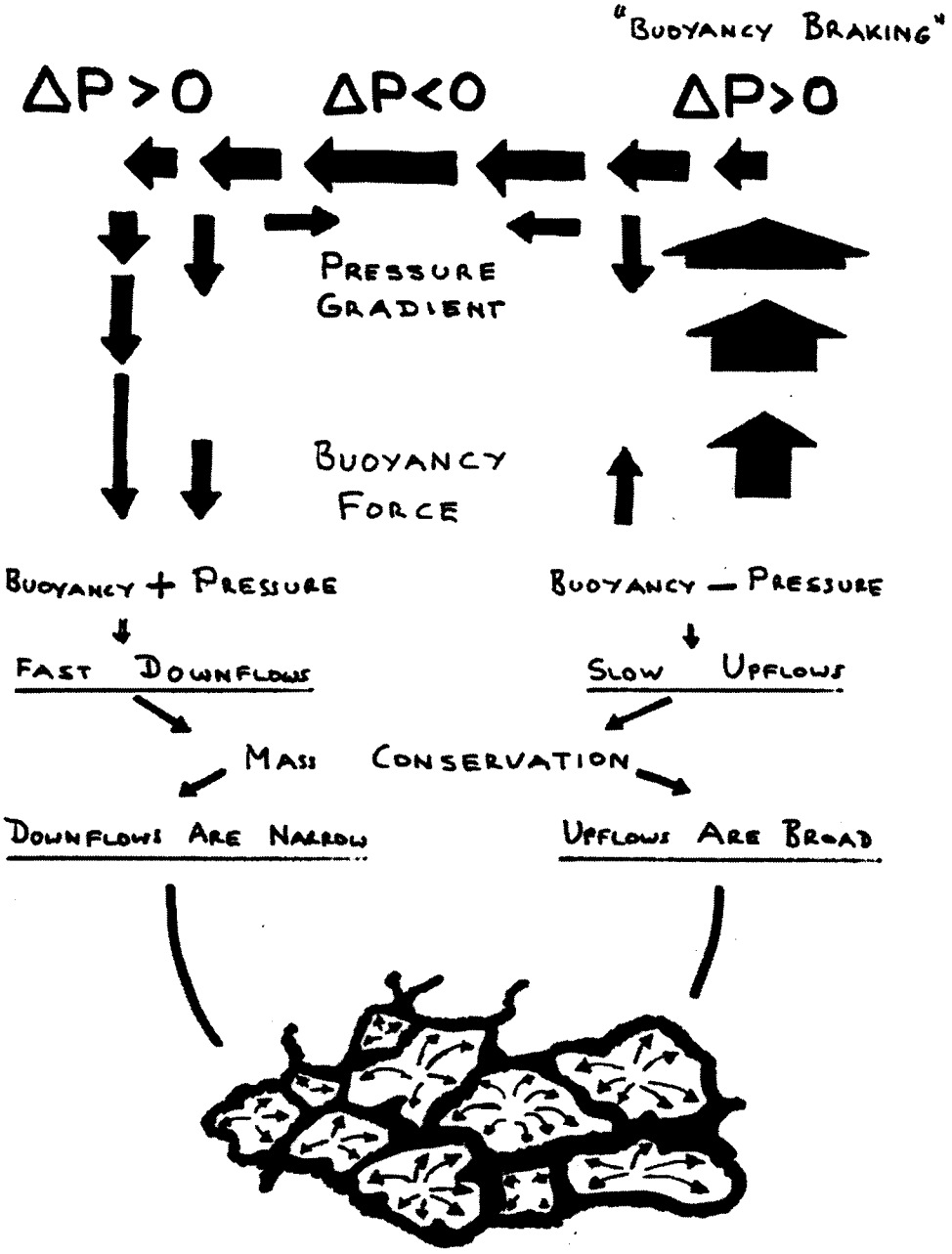


Figure 8.5: Topology of granulation.

If we set $r = R$ for the radius of the convection cell we obtain

$$R = 2H_\rho \frac{v_r(R)}{v_z} \,. \tag{8.14}$$

As R increases $v_r(R)$ must increase too, relative to v_z! This has to do with the stratification of the atmosphere, with scaleheight H_ρ. Only a thin layer with sufficient density is available in which to transport the matter horizontally.

As v_r increases, so must the magnitude of the excess pressure Δp over the upflow which drives v_r. But this Δp also acts as a downward directed force on the gas flowing up below it. It causes the upflow to slow and finally, if Δp becomes too large chokes it. Obviously the granules cannot be bigger than a certain size. The measured size distribution of granules is shown in Fig. 8.6.

Larger cells may exist but must have different properties, in particular a small v_z and a large v_r/v_z ratio (compare with the properties of mesogranules and supergranules in Table 8.2.

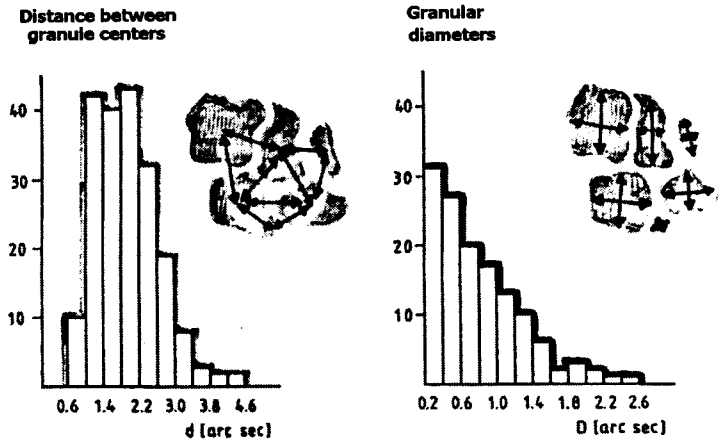


Figure 8.6: Distance d between granule centers (Bray and Loughead 1977) and granular diameters D (Roudier and Muller 1986). Maximum size is determined by mass conservation.

8.2.2 Supersonic convection

Granular convection is shocking. For large *Rayleigh numbers* the convective flow can become *supersonic* at some positions. The Rayleigh number Ra ($\sim$ superadiabatic gradient) measures the importance of the buoyancy force relative to the stabilizing effects of non-magnetic diffusion (Priest 1982). The presence of supersonic convection is in particular the case for the horizontal velocity of the largest granules

The shocks themselves have not yet been directly observed, because these require a very high spatial resolution to be resolved. Recently however, the very large line broadening due to shocks has been observed.

Chapter 9

Flux tube overview

Lecturer S.K. Solanki, notes by R. van Deelen

In the photospheric layers of it active regions we can see a whole zoo of magnetic structures, corresponding to fluxtubes of different sizes pointing out of the solar surface. An overview of the observed properties of these fluxtubes is given in Table 9. The dependence of the brightness and the field-strength of the fluxtubes as a function of their size are plotted in Figures 9.1 and 9.2, respectively.
It is difficult to obtain information about the so-called 'weak fields' and to resolve small magnetic features. In the next chapter we focus on the biggest structures in town: sunspots.

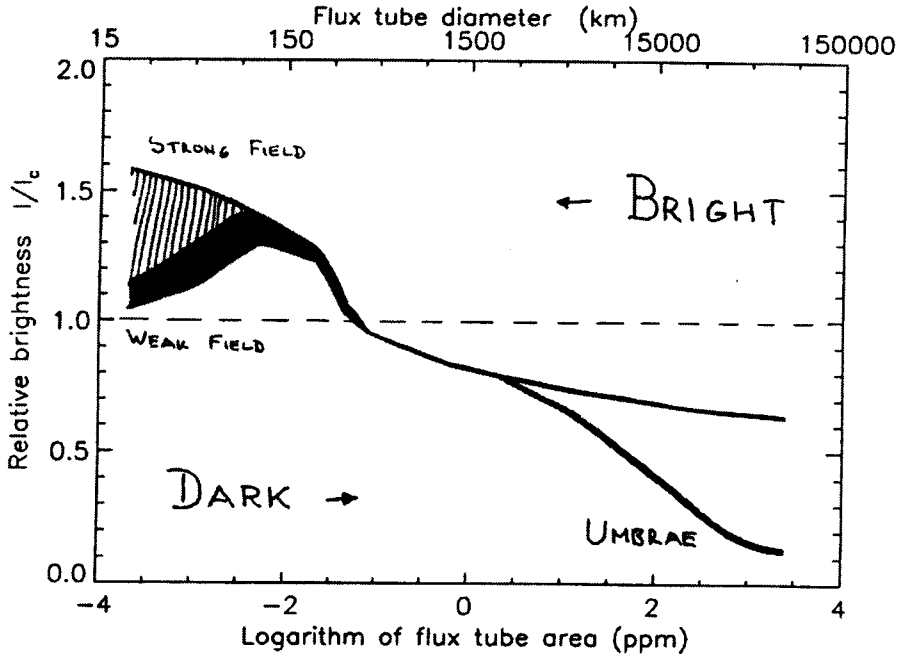


Figure 9.1: Because $I \sim T^4$, small differences in T give a good contrast. If a fluxtube is on the dotted line, this means the magnetic structure is equally bright as its surrounding granulae. The upper line in the right corner stands for the average relative brightness of sunspots, so we see umbra are very dark.

Properties	'Weak fields'	**Magnetic Elements**	**Knots**	**Pores**	**Spots**
Diameter	$\lesssim$ 100 km	$\lesssim$ 300 km	⟶		50000 km
Temperature	?	$T_{\rm eff} \approx 6000$ K (bright)	(neutral)	(dark)	$T_{\rm eff} \approx 4200$ K (umbra) $T_{\rm eff} \approx 5200$ K (penumbra) (dark)
Magnetic field-strength	200 – 800 G	1500 G 1500 G	⟶		2000 – 3000 G (umbra) 800 – 2000 G (penumbra)
Lifetime	min – h	min – h	⟶		days – months
Distribution	everywhere	active regions and network (whole disk)	⟶		active regions only (within 30° of equator)
Inner structure	?	perhaps	likely		lots

Table 9.1: An overview of the properties of photospheric magnetic structures. Information about the smaller features, especially about weak fields is difficult to obtain For example the lifetimes of the magnetic elements are not known well.

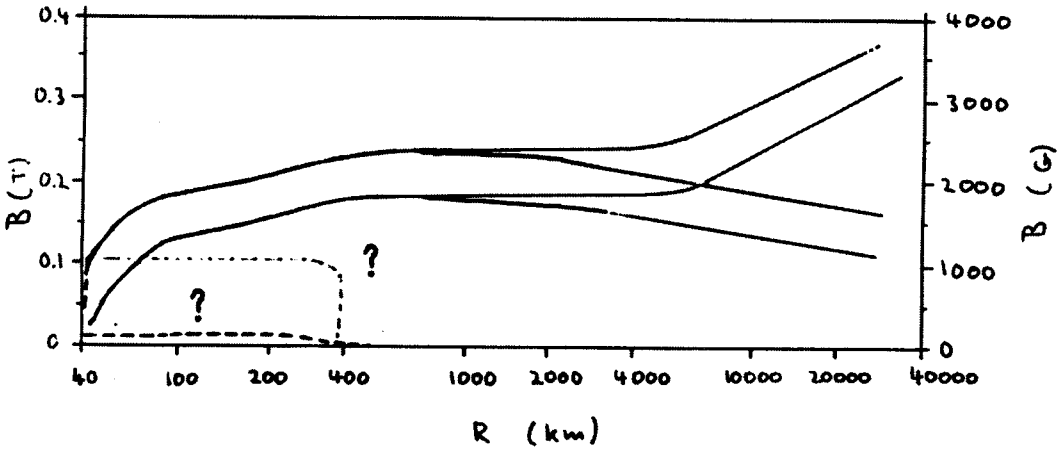


Figure 9.2: Field-strength B versus radius R of photospheric magnetic features. B averaged over the magnetic feature (lower set of curves in the right part of the figure) doesn't depend much on the size of the magnetic elements! The upper set of curves describes the peak field strengths.

Chapter 10

Slender flux tube physics

Lecturer S.K. Solanki, notes by N.L.J. Cox

10.1 Introduction

Most of the magnetic energy is stored within discrete features which are best described by the term flux tubes. Small scale solar magnetic features can be separated into two main types. These are the *small flux tubes* on the one hand and the *weak field features* on the other.

In flux tubes the fields are concentrated (up to several kG in the photosphere) and the magnetic energy density dominates over the thermal energy density (which is expressed by the pressure) and dynamic energy density at and above the visible solar surface. Whereas the weak-field features are much less concentrated (field strength below 1 kG) in the photosphere. In some cases the field strength is small enough for the thermal energy to be larger than the magnetic energy. Their magnetic flux is also smaller. It is important to note that the effect on the irradiance of the small flux tubes is significant, but too small to be measured for the weak field features. Nevertheless, at least 10% of the magnetic flux in the active region is stored in weak field form. In the quiet Sun this fraction reaches 50%.

Small flux tubes appear mainly as bright magnetic elements forming faculae (or plage in the chromospheric layers). Due to the measurement threshold of about 200 km a lot of elements are missed in the observations when only brightness is observed. Observations in net (*e.g.*, circular) polarisation are also able to reveal features far below the spatial resolution.

In the course of this chapter the features of flux tubes will be discussed in detail. First (Section 10.3) an approximation is introduced to provide a better understanding of the principles of flux tubes. Secondly, in Sections 10.4 and 10.5, we shall take a look at magnetic field and the theoretical explanation of field strength. Subsequently the two effects, flux expulsion and convective collapse, which determine the magnetic structures and field strengths in the interior of convection cells (or flux tubes) to a great extent are discussed (Section 10.6). The inclination of flux tubes is also briefly discussed at this point.

The remaining part of the chapter will concentrate mainly on flows and waves in and on flux tubes.

10.2 Flux tubes: What are they?

To get more in depth on the matter we first give a definition for a flux tube:

A flux tube is a bundle of magnetic field lines bounded by a topologically simple surface, generally modelled by a current sheet.

At first this seems an oversimplified definition, but later on it will become clear that even this view introduces extensive calculations in even the simpler models. Furthermore, some characteristics concerning the magnetic field and the active regions need to be discussed and this simple idea of a flux tube has to be even further simplified by use of the thin flux tube approximation to be able to draw conclusions without expanding with a large effort.

Figure 10.1: Schematic view of a simple flux tube according to the definition

Flux tubes can have a complex internal structure. This internal structure is described by a set of different physical quantities. Depending which characteristics are modelled, only a few quantities actually enter the model in order to ease the solving of all equations and draw clearer conclusions). Often used quantities are *size* (*i.e.*, the cross-sectional area or the radius), *magnetic field strength*, *brightness* (or *temperature*, see also Section 11) and the *magnetic filling factor*.

10.3 Thin flux tubes

First, we consider a vertical axially symmetric flux tube (see Fig. 10.2). In principle we now have a 3-dimensional problem. But if we introduce the cylindrical coordinates (r, ϕ, z), set $\partial/\partial\phi = 0$ and let $\vec{v} = 0$, as well as $\partial/\partial t = 0$ we can reduce it to a 2-dimensional problem.

Now let's have a look at any physical quantity f (*e.g.*, pressure P, $\vec{B}$-field) and expand it in a Taylor series:

$$f(r,z) = f(0,z) + r\frac{\partial f}{\partial r}\,|_{r=0} + r^2\frac{\partial^2 f}{\partial r^2} + \ldots \tag{10.1}$$

If the radius R of the flux tube is smaller than scale of vertical variation, or put differently, if $R \ll H_p$, we may assume that all quantities vary weakly with r (apart from a jump at $r = R$) and therefore the system may be truncated at zeroth order. This is what we call the *thin tube approximation*. So if we now choose the magnetic field $\vec{B}$ for the quantity f and insert (10.1) into the MHS equations.

$$\frac{{B_z}^2}{8\pi} + P_m = P_s \quad \text{(horizontal) pressure balance} \tag{10.2}$$

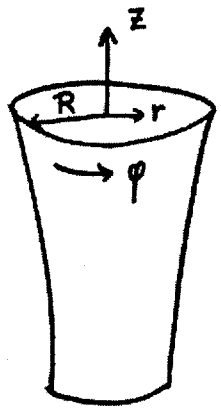


Figure 10.2: Axisymmetric flux tube representation

$$\frac{\partial p_i}{\partial z} = \rho_i g \quad \text{(vertical) hydrostatic balance} \tag{10.3}$$

$$P_i = \frac{\kappa}{m_p}\rho_i T_i \quad \text{ideal gas law} \tag{10.4}$$

$$\pi R^2 B = \Theta = \text{constant} \quad \text{magnetic flux conservation,} \tag{10.5}$$

where subscript m stands for the magnetic feature and the s for its surroundings. The subscript i can stand for both the m or s.

As with all approximations one should be careful when applying it. At this moment in the discussion a few remarks concerning the applicable range and usefulness are required:

- the radial and vertical components of the force balance equation decouple completely in this approximation (see the Equations 10.2 and 10.3);
- the force balance reduces to pure pressure balance, *i.e.*, the curvature forces are negligible (see eq. 10.2);
- allowing first order terms in equations (10.2) –(10.5) instead of only zeroth order terms, we may obtain a radial component of B: $B_r = -r/2\partial B_z/\partial z$. Higher order terms can be introduced to extend the domain of validity of this approximation, but these make the equations far more complex;
- if $\vec{v}$ and $\partial/\partial t$ terms are allowed into the equations a dynamic thin tube approximation is obtained (Roberts 1991). This is important for the description of flux tube waves. Flux tube waves are discussed in Section 10.13;
- the thin-tube approximation can describe flux tubes in the solar interior (even large flux tubes) due to the large $H_p \sim T$ dependence in the solar interior. The thin-tube approximation also describes magnetic elements in the solar photosphere, but cannot do so for large flux tubes such as sunspots. (see Fig. 10.3);
- because the gas pressure is lower in the magnetic feature, the opacity is generally also lower. And thus we see, via the Wilson depression, deeper layers.

In Fig. 10.3 the results from an expansion solution including second order terms, *i.e.*, including magnetic tension, and the thin tube approximation are compared. From which it can be concluded that for thin tubes (radius $r(z = 0) < 50$ km) the approximation does well till reasonable, but for thicker tubes the discrepancies increase, with the discrepancies starting at lower heights for larger radii.

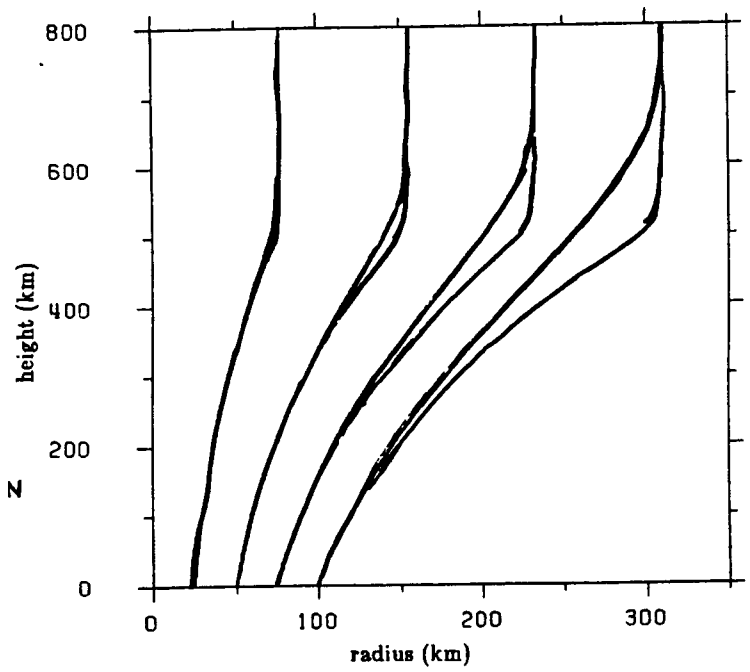


Figure 10.3: Comparison of the flux tube cross-section obtained with the thin tube approximation with the expansion solution (with the magnetic tension included) from Pneuman *et al.* (1986)

10.4 Magnetic fields in flux tubes

A theory of the magnetic field in flux tubes must satisfactorily describe the confinement of the field. Secondly, a process that is responsible for the concentration of the magnetic field to its high observed value after its emergence from the solar interior must exist. Both the predicted and the observed field strengths at emergence lie well below the values observed in mature stable flux tubes.

We see that most of the visible flux is in strong-field form (at least in the active regions and the network). In this region $B(z=0) \approx 1500-1700$ G which corresponds to a plasma of $\beta = 8\pi P/B^2 \approx 0.3$ (notice the peak in the right panel of Fig. 10.4), where β represents a measure of energy density in the gas relative to the magnetic energy density. For a $\langle\beta\rangle \approx 0.32$ the magnetic features are strongly evacuated and the internal magnetic energy density dominates over the internal gas energy density. How much of the magnetic flux is in the form of plasma with a certain β is shown in the right panel of Fig. 10.4. The histogram shows a peak at ≈ 0.3 with a long tail extending to larger β. The vertical dashed line gives the theoretical limit of convective collapse. Below this value of $\beta = 1.8$ flux tubes are stable against collapse, according to a linear stability analysis. It states that flux element with $\beta > 1.8$ which are embedded in a convectively unstable environment collapse to a strong-field state. The observed results from both the active (histogram) and quiet (solid line) regions of the Sun are plotted in the same panel. According to this figure in active regions most of the magnetic flux resides inside the strongly evacuated magnetic features and much less flux resides in features with a smaller magnetic field. In the quiet region the peak in the flux lies at a slightly larger β value and it also has a stronger and longer tail to still larger β. In the quiet regions of the Sun the magnetic elements are

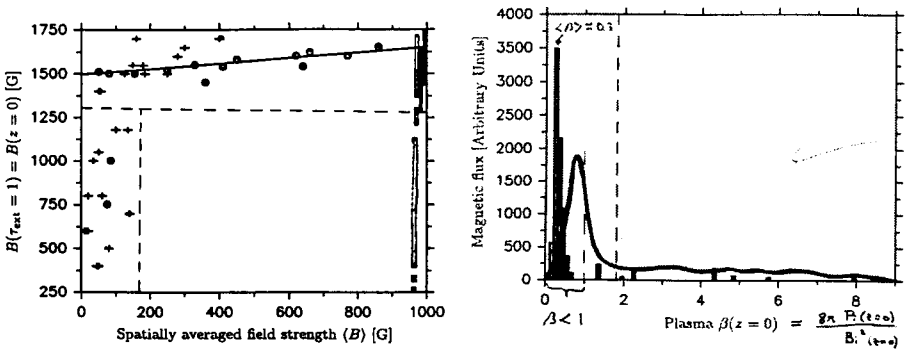


Figure 10.4: Left panel: B(z=0) vs. $\langle B \rangle$, the flux of a magnetic component divided by the area of the spatial resolution element. $\langle B \rangle$ is proportional to the magnetic filling factor. Right panel: Histogram of the magnetic flux contributed by all the magnetic elements (in an active region) vs. the plasma $\beta(z = 0)$ within each element. Overlaid in a solid line are the measured plasma β values in the quiet Sun. The vertical dashed line indicates the boundary for stability against convective collapse (according to linear stability theory). From Rüedi *et al.* (1992).

therefore not as much evacuated (due to a larger β) as the ones in the active region. Also, there is an almost equal amount of flux in weak-field features ($\beta > 1$) as in the strong-field ones. (*i.e.*, the two integrated areas are about the same).

A regression line has been drawn through the points above 1.3 kG (for which $\beta < 1$) in the left panel of Fig. 10.4. For the smaller filling factors, corresponding to the small features and small flux, a rapid drop of $B_{\text{intrinsic}}$ is shown. These weak-field features tend to have equipartition fields ($800\,\text{G} \lesssim B \lesssim 200\,\text{G}$). The observed variation of B with z is well described by standard flux tube models, including merging of neighbouring tubes (Solanki 1996).

10.5 Theoretical explanation of field strength

The confining force of the field in solar flux tubes is provided by a horizontal gas pressure gradient created by partial evacuation of the flux tubes. The Force balance $B^2/8\pi = p_e - p_i$ (see 10.2) must be satisfied at every height z. And because the gas pressure drops exponentially with z the field strength B must follow. The conservation of magnetic flux then forces the flux tube to expand with height, with its cross-section varying as $1/B(z)$, according to equation (10.5).

The lower gas pressure within flux tubes implies that the optical depth unity is reached at a deeper level than in a field-free gas, which makes you see deeper. In sunspots this height difference, called the *Wilson depression*, is directly measurable. The energy outside the magnetic element (*e.g.*, a flux tube) is mainly transported by convection. Inside a magnetic element the energy is transported by radiation from the hot walls of the element. For a optical thin element (*i.e.*, the radius is of the same order of magnitude as the mean-free path of the photons), the radiation is able to heat up its entire interior. Due to the presence of a magnetic field inside the element the photons more readily escape in the vertical direction. Together with the rapid temperature increase downwards and the fact that we see deeper into the magnetic element than into the non-magnetic atmosphere (due to the Wilson depression) we see why magnetic elements appears hotter and brighter, compared to their surroundings. This situation is illustrated in Fig. 10.5.

An interesting question that might arise is whether there is an upper limit for the B

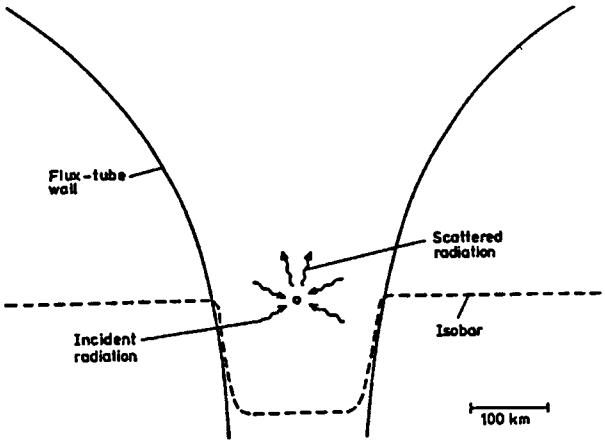


Figure 10.5: Radiative heating of the interior of a thin flux tube. From Stenflo 1994

at a given height z. This limit is obtained fairly simple by setting P_i equal to zero, or: $B^2/8\pi = P_e$. However, since we see ever deeper with decreasing $P_i(z_0)$ and B increases with depth there appears to be no limit on the observed field strength.

Can theory set a limit on the field strength, or, equivalently, on the decrease of P_i in the observed layers of fluxtubes? Perhaps surprisingly it can do exactly that. The relevant process is called convective collapse. The theory behind convective collapse will be discussed in Section 10.6.2.

10.6 Formation of flux tubes

Inside the Sun the convective motion sweeps the frozen-in magnetic field to the boundaries of the convection cells (such as granules or super granules). But this effect, *convective flux expulsion*, is insufficient to produce the observed kG fields and the momentum of the convective flows is too small to confine these fields. It is only the gas pressure that can match the magnetic pressure of such fields. The mechanism leading to the further confinement of the field is the convective collapse (from Solanki 1996) To understand the principals behind these statements we now concentrate on the theory (and some modelling) behind these two phenomenon: *convective flux expulsion* and *convective collapse.*

10.6.1 Flux expulsion

Convection and the magnetic field try to avoid each other in a way so that the field is carried and concentrated into the downflow regions. Field strength is limited by the equipartition of magnetic and kinetic energy density, *i.e.*, $B^2/8\pi = \rho v^2/2$, where v is the velocity of the convective flows (basically of granulations). In this way $B \approx 200$ can be produced. Convection is not totally inhibited.

The results of a numerical experiments by Galloway and Weiss (1981) are represented in Fig. 10.6. In this experiment they started with a homogeneous magnetic field and imposed the constraint of flux conservation. An enhancement of the field strength and a spatial separation between the magnetic field and the flow are seen.

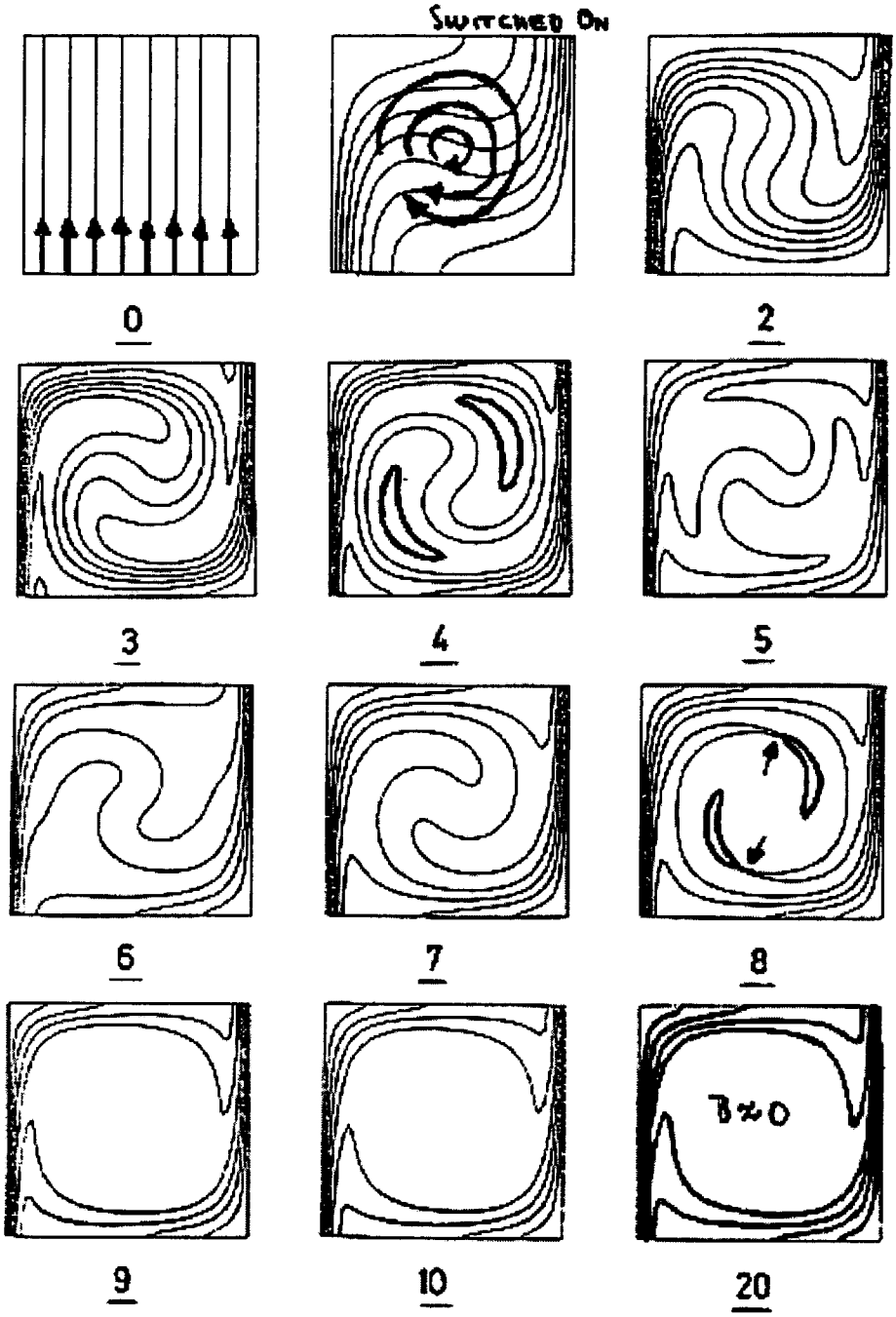


Figure 10.6: Flux expulsion: The initial condition is a homogeneous $\vec{B}$-field. At a certain time (time step 1) the convection is switched on, which imposes a circular motion. In time (time steps 2 and later) the frozen-in field lines are carried along by the flow. Slowly this results in an initial enhancement of the field. The production of strong field line curvature is initiated. Around time step 5 a cancellation of flux due to magnetic reconnection, near the regions with largest velocity, shows up in the model. From there on (time step 6) there's a steady expulsion of flux to the boundary of the convection cell. After a while a final (stationary) state is achieved. Strong fields ($B \ll B_0$) occur at the edges of the convection cell, while the interior is almost field-free. The final field strength near the convection cell boundary is given by the equipartition field: $B \approx (\sqrt{4\pi\rho})v$ obtained from Galloway and Weiss (1981).

10.6.2 Convective collapse

The mildly concentrated magnetic field resulting from flux expulsion is thermally insulated from the surroundings and is therefore also cooler. A small adiabatic downflow cools the diffuse magnetic tube further compared to its superadiabatic surroundings. Due to this cooling the gas pressure near the photosphere is lowered in the magnetic feature, since $H_p \sim T$ and also $P = P_0 e^{z/H_p}$.

In addition, a convective downflow within a weak field feature partially evacuates it, thus lowering the gas pressure even further. This creates a horizontal pressure gradient, *i.e.*, a force directed towards the flux concentration. In turn this causes gas to flow horizontally towards the feature. The converging gas flow carries the embedded magnetic field lines with it, leading to an enhancement of the field. Horizontal pressure balance is restored by compressing the field lines in the magnetic flux region. If the magnetic pressure of the amplified field has become strong enough further convective motions within the newly formed flux tube will be quenched and the flux tube will stabilize against further collapse. The effect described above is sketched in Fig. 10.7.

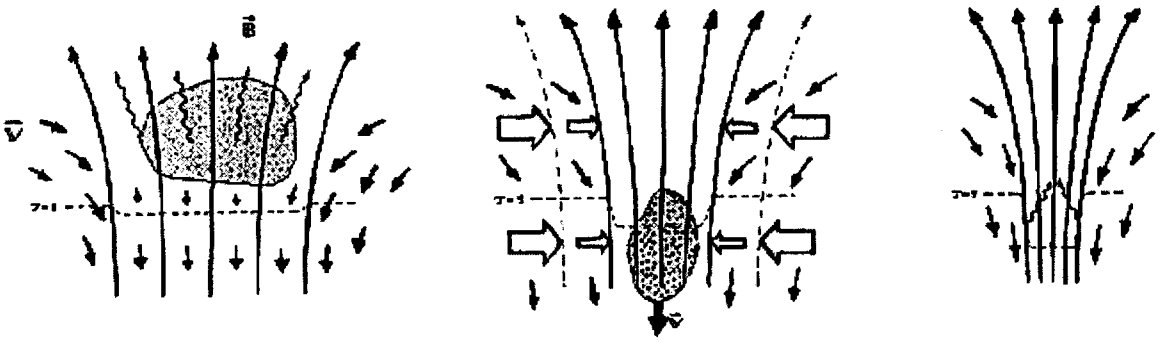


Figure 10.7: **Left**: the initial pressure equilibrium of material in the magnetic patch suffers radiative losses and consequentially cools. The downflow lanes are cool. The material inside cools down and becomes denser and starts to flow down. **Center**: the cool material sinks. This leads to a pressure deficit higher up. Surrounding material flows inwards and concentrates the field. The density in the magnetic patch is lower, so that the pressure is lower too. Inward flow pushes the field together and the field strength increases. **Right**: a new pressure equilibrium is reached when the magnetic pressure becomes sufficiently large. The time scale of the whole event is estimated to be a few minutes, which is confirmed by magnetohydrodynamic simulations. No direct observations of a convective collapse have yet been made. Figure courtesy of M. Bünte and O. Steiner.

The combination of flux expulsion and convective collapse can enhance the magnetic field strength, by a factor of 10 or even more.

10.7 Magnetic field inclines to a certain inclination

10.7.1 What the observations say

Magnetic elements are on average inclined by 10° to the vertical, with a tendency for a more inclined field at the edges of plage/network areas. No significant difference between plage and network has been observed. The distribution of inclination angles in plage is plotted in Fig. 10.8. Weak field features (surprisingly?) tend to be far more inclined. Can the observational result above be explained?

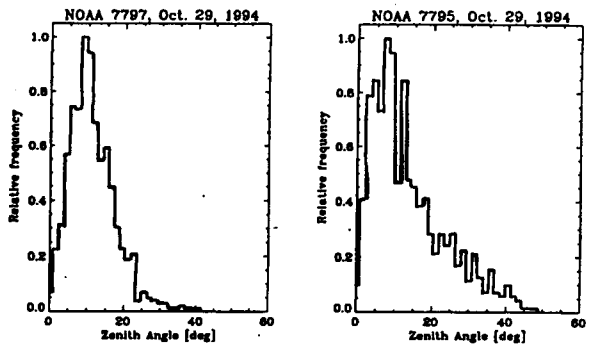


Figure 10.8: Inclination of tubes in two active regions (Figure from Martínez Pillet *et al.* 1997).

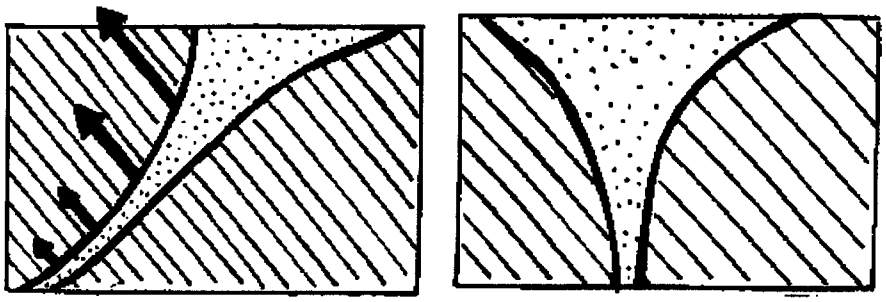


Figure 10.9: Buoyantly unstable situation (left): an upward force acts on the flux tube, which contains less dense material than its surroundings. Buoyantly stable situation (right): due the upward force the flux tube is vertically orientated. The tube is anchored at the bottom of the convection zone.

10.7.2 Is there a theory?

Indeed there is. The important physical process is magnetic buoyancy. The evacuation of material from inside the flux tube due to the pressure balance makes the flux tubes buoyant relative to the field-free gas in their surroundings. If we recall the fact that flux tubes are anchored at the bottom of the convection zone their alignment must be practically vertical at the solar surface.

In the left panel of Fig. 10.9 the buoyantly unstable, and in the right panel, the buoyantly stable situation are illustrated. In the unstable situation, denser gas overlies the less dense gas in the flux tube. In the stable situation the dense gas always lies below the less dense gas.

The cause of magnetic buoyancy is the horizontal pressure balance: $B_i^2/8\pi + p_i = p_e$ (the first term is the magnetic pressure). From the structure of the flux tube we have that the gas pressure is much lower inside magnetic feature than outside (*i.e.*, $\beta < 1$ (see also Section 4.4.2). We also note that $T_i \approx T_e$, so that by way of the ideal gas law the density is lower in the magnetic feature. Therefore a net force acts on the tube until its in an stable state with its surroundings, *i.e.*, when its aligned vertically.

How then is the observed inclination of 10° explained? There are various ways to increase the inclination:

- horizontal flows outside the flux tube, for example due to neighboring granules (see chapter 11) can cause flux tubes to have an inclination. If $\rho v^2/2 > B^2/8\pi$ the inclination can be significant. In this context the pressure of supersonic horizontal flows in

granules are important, since at the sound speed $\rho v^2/2 \approx p \approx B^2/8\pi$ (kinetic energy equals pressure equals magnetic energy);

- kink waves (see Section 10.13.3);
- the balloon-cluster effect (see Fig. 10.10) can occur due the divergence of fluxtubes with increasing height, causing the outermost flux tubes to be the most inclined, in agreement with the observations.

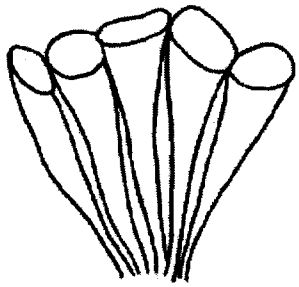


Figure 10.10: Illustration of the balloon cluster effect. The widening of the flux tube with height makes it impossible for the tubes to come closer together, due the lack of available space. Therefore the more outlying ones must have a small inclination.

10.8 Flows in small fluxtubes

Flows and waves in flux tubes cover a wide area of research. Three main aspects of this topic will be discussed in the next section. First there is the problem of observing stationary flows. So far none with a velocity over 250 km/s have been seen in small scale features, with the duly noted exception of Siphon flows (Section 10.12). The advantages and disadvantages of a high spectral resolution, in particular for the measurement of flow, are discussed in Section 10.9. The second issue that is to be treated briefly is how Stokes V profiles are affected by velocities and velocity gradients inside and outside magnetic flux tubes. Finally, a brief and simple introduction is given to waves in flux tubes (Section 10.13).

10.9 Spectral resolution

trying to obtain a high spectral resolution has many advantages. For example, a low spectral resolution means a low sensitivity to, *e.g.*, velocity (*i.e.*, shifts and broadening) and Zeeman splitting. When doing polarimetry, by observing the polarized Stokes parameters (Q, U, V), a higher resolution is needed. This is obvious, since these parameters change more rapidly with wavelength than the parameter I. Even more so the Q, U, V have positive and negative lobes and when these lobes get spectrally smeared it leads to cancellation. A significant loss of signal can be the result. This latter effect is clearly shown in Fig. 10.11. One can see that the observed values of V decreases rapidly with increasing instrumental broadening.

Three other consequences of too low spectral resolution are the enhancing of blending problems (nearby lines merge), the production of spurious line shifts (since the blue lobe of Stokes V is usually stronger than the red lobe this asymmetry leads to a spurious redshift

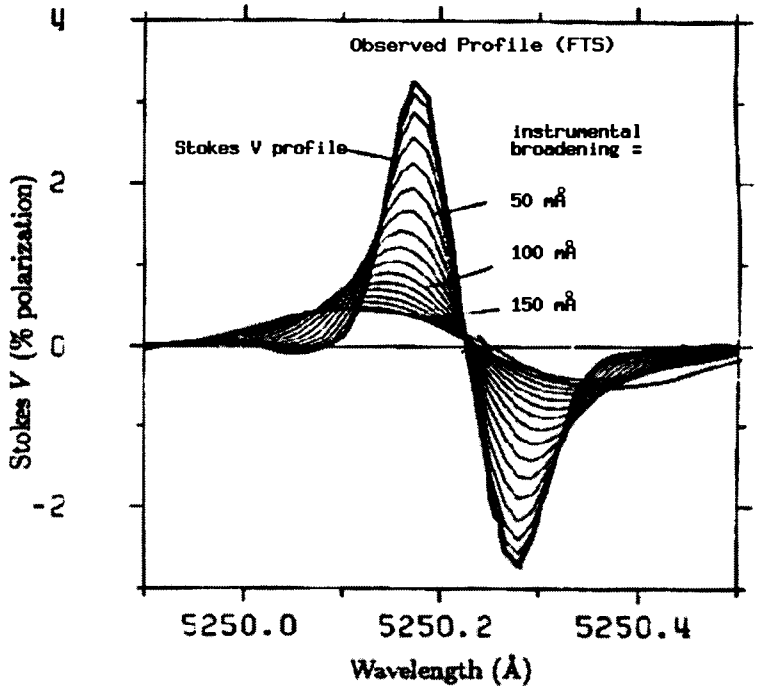


Figure 10.11: Observed Profile (FTS), where the different Stokes V profile lines indicate variations in the instrumental broadening. A higher and steeper profile is a result of a smaller broadening factor, *i.e.*, increased spectral resolution.

of the V profile) and the fact that a number of interesting features, such as complex profile shapes are missed. The major disadvantages of high spectral resolution are mainly practical, such as the increased photon noise or integration time, the date cube and loss of spectral coverage.

10.10 Stokes V profiles

10.10.1 Typical Stokes V profile

A typical V profile is shown in Fig. 5.5. One typical aspect of the profile is its asymmetric shape. In particular, the area of the blue wing is larger than the area of the red wing (*i.e.*, $A_{\rm blue} > A_{\rm red}$). Before having a closer look at the way these profiles are formed and how their shape is affected some definitions are recalled.

$$\begin{aligned} \text{absolute amplitude asymmetry:} \quad & \Delta a = a_{\rm b} - a_{\rm r} \\ \text{relative amplitude asymmetry:} \quad & \delta a = \Delta a/(a_{\rm b} + a_{\rm r}) \\ \\ \text{absolute area asymmetry:} \quad & \Delta A = A_{\rm b} - A_{\rm r} \\ \text{relative area asymmetry:} \quad & \delta A = \Delta A/(A_{\rm b} + A_{\rm r}) \end{aligned} \tag{10.6}$$

10.10.2 Area asymmetry

Consider the way area asymmetric V profiles are produced. We concentrate on this aspect since there are many ways of producing *amplitude* asymmetry so that its diagnostic

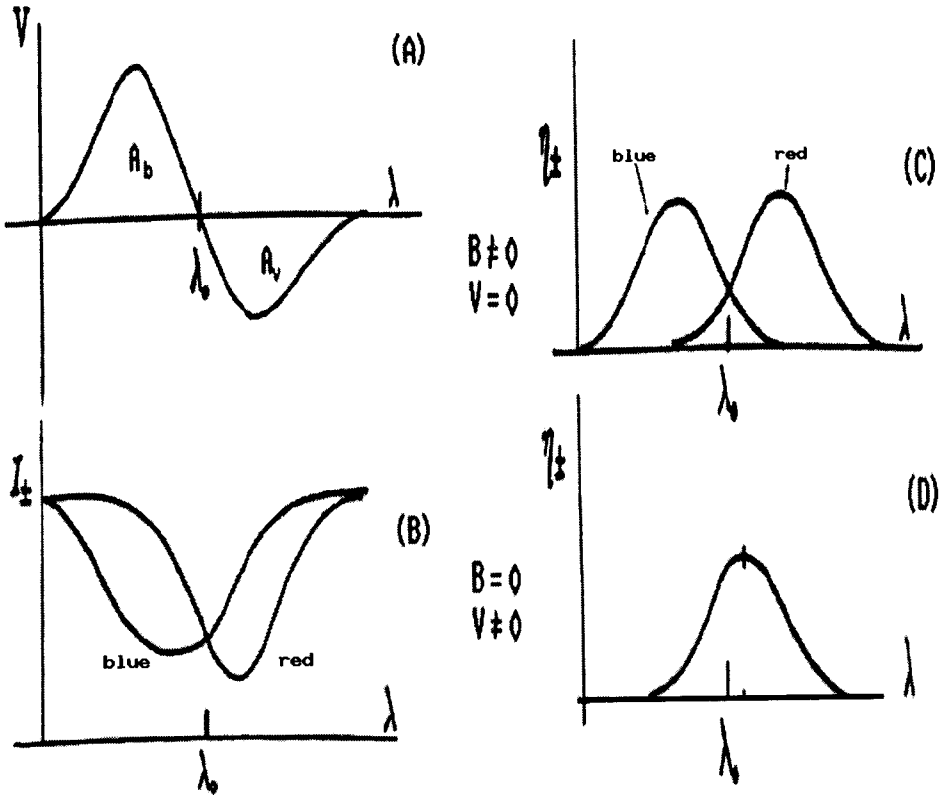


Figure 10.12: Production of Stokes V area asymmetry. The field is assumed to be taken parallel to the line of sight. (D (upper layer)): If $B = 0$, there's no polarization and no splitting of the spectral line into blue and red profiles. Because of the presence of a down flow velocity the profile is shifted towards the red. (C (lower layer)): Zeeman splitting due to the non-zero B-field, but because the velocity is zero the blue- and red-profile components of the Zeeman-split profile cross exactly at λ_0. (B): If $I_\pm$ is plotted the total absorption due to the blue-shifted component is larger than the red-shifted component. For increasing absorption the red profile overlaps more. (A): The Stokes V profile. We notice that $A_b > A_v$.

contents are limited. The ingredients for producing δA are $dv/d\tau$, $d|B|/d\tau$ and line saturation. The sign of the asymmetry is given by:

$$\delta A \propto \frac{d|B|}{d\tau}\frac{dv}{d\tau}\begin{cases} <0 & \rightarrow \quad \delta A > 0 \\ >0 & \rightarrow \quad \delta A < 0 \end{cases} \tag{10.7}$$

Models that try to reproduce the observed area asymmetries have to satisfy the constraint of no zero-crossing shifts ($\lambda_V - \lambda_I \approx 0$) (see Section 10.8). This condition can be fulfilled when there are no mass flows inside the flux tube. It is important to note that a velocity field outside the magnetic feature can also produce an area asymmetry without shifting the V profile. (when the line of sight crosses the boundary of the tube, see left panel of Fig. 10.13).

In the external, non-magnetic region a downward flow is present, due to the location of the flux tubes in downflowing intergranular lanes. They are probably responsible for the δA of Stokes V.

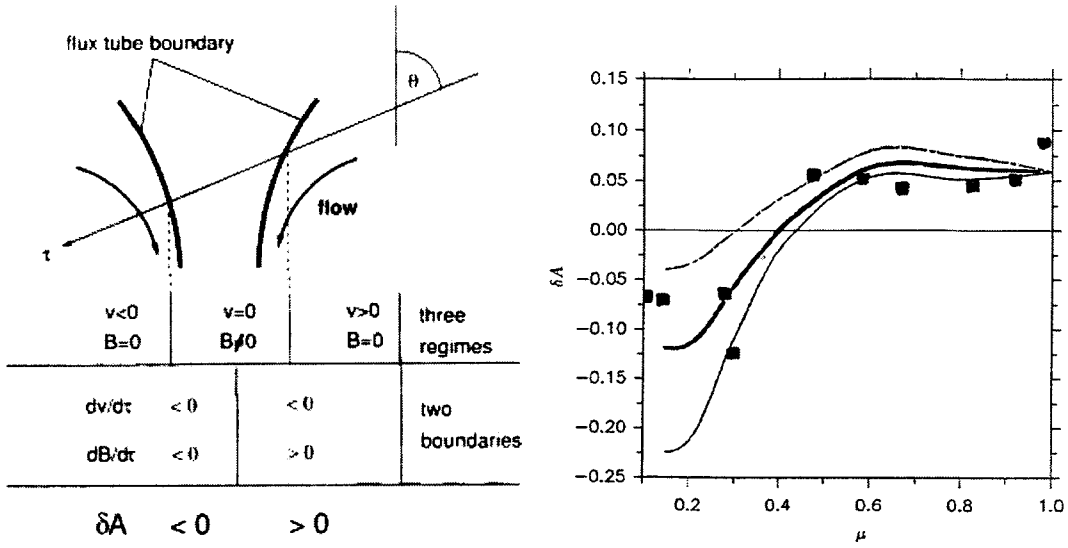


Figure 10.13: **Left** panel: Flipping of sign of δA. In region 1 $B = 0$ and $v_{||} < 0$. In region 2 $B \neq 0$ and $v = 0$. In region 3 $B = 0$ and $v_{||} > 0$. These are the three regimes a ray crosses. On the boundaries of these regimes we look at three terms: $\frac{dv}{d\tau}$, $\frac{d|B|}{d\tau}$ and δA. The first term is < 0 on both boundaries, but the second term is < 0 on the left and > 0 on the right boundary. And this latter sign change is propagated into the sign change of δA. From Bünte *et al.* (1993). **Right** panel: The center-to-limb variation of the relative area asymmetry δA. for the Fe I 5250.22Å line. The curves differ in assumed values for the flow speed in the external convection cell. (upper) $v_H > v_v$, (middle) $v_H \approx v_v$ and (lower)$v_H < v_v$. v_H represents the horizontal velocity, while v_v the vertical velocity. From Bünte *et al.* (1993).

10.10.3 Production near the limb

Production of an area asymmetry near the limb happens in a slightly different way than at disk centre. Now, in addition to the vertical component of the velocity of the granules, also their horizontal component plays a role (the horizontal flows are directed towards the flux tube). Sufficiently close to the limb most rays pass through the boundary of a flux tube twice (see left panel of Fig. 10.13). Thus the observed radiation travels through three magnetic field regimes. The $dv/d\tau$ factor has the same sign at the two boundaries, but $d|B|/d\tau$ changes sign. The δA of the V profile produced at the two boundaries of the flux tube thus have different signs (see eq. 10.7). The resulting sign of the final observed δA is caused by a balance of the conditions at the two boundaries.

10.11 Indirect evidence for non-stationary motions

10.11.1 Line width

Stokes V profiles exhibit non-thermal, non-magnetic line broadening of $1.5-2$ km/s RMS, due to some velocity dispersion. Non stationary motions in the unresolved flux tubes provide one. These velocity values are much larger than those observed directly through wavelength shifts.

10.11.2 Horizontal motions

Magnetic elements are moved around horizontally at velocities ranging up to 2 km/s. The motions can be random or quasiperiodic. It has been shown that at least the faster of these motions produces kink mode waves with sizable amplitudes. These carry enough energy to heat the upper solar atmosphere if the energy is efficiently dissipated.

10.12 Siphon flows

If the field strength at the footpoints of a loop is not the same, than due to the resulting gas pressure difference a flow from the footpoint with the weaker field to the footpoint with the stronger field is set up. Fig. 10.14 illustrates this effect. This theory holds when the footholds are thin enough (higher order effects can show up otherwise).

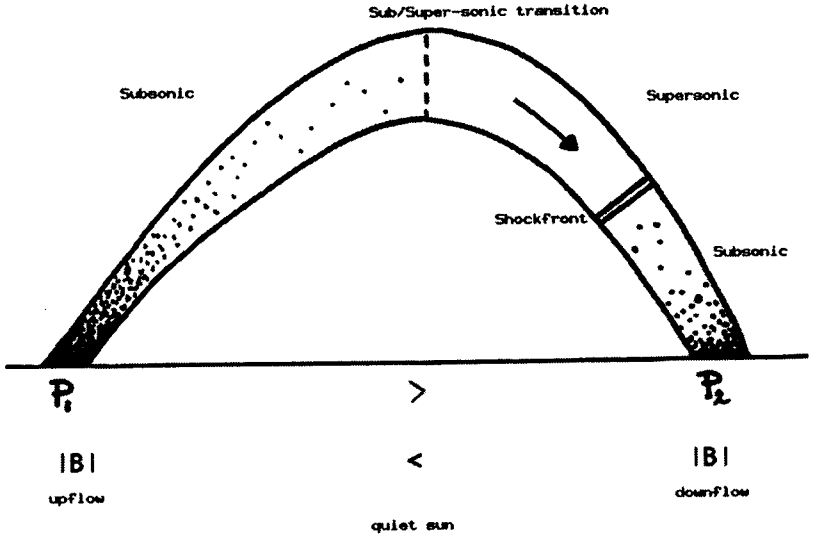


Figure 10.14: Illustration of the siphon flow

If the observations are compared with the theory it follows that the flow must be supersonic in the upper photosphere, and that it must shock back to subsonic velocities above the line formation height. Several question however remain unanswered, such as: like how often do siphon flows occur and how do they affect the fluxtubes?

10.13 Linear wave modes in thin fluxtubes

There are three wave modes that occur in flux tubes. These are the *longitudinal*, the *transverse* and the *torsional* waves. The transverse waves can be of different order, with the lowest order ones (the *kink* mode) being the only one present in thin flux tubes. Subsequently some characteristics of these waves are discussed.

10.13.1 Torsional Alfvén waves

If the flux tube is incompressible the waves are Alfvénic (see Fig. 10.15). The restoring force in this kind of flux tube is the magnetic curvature force. The propagation velocity

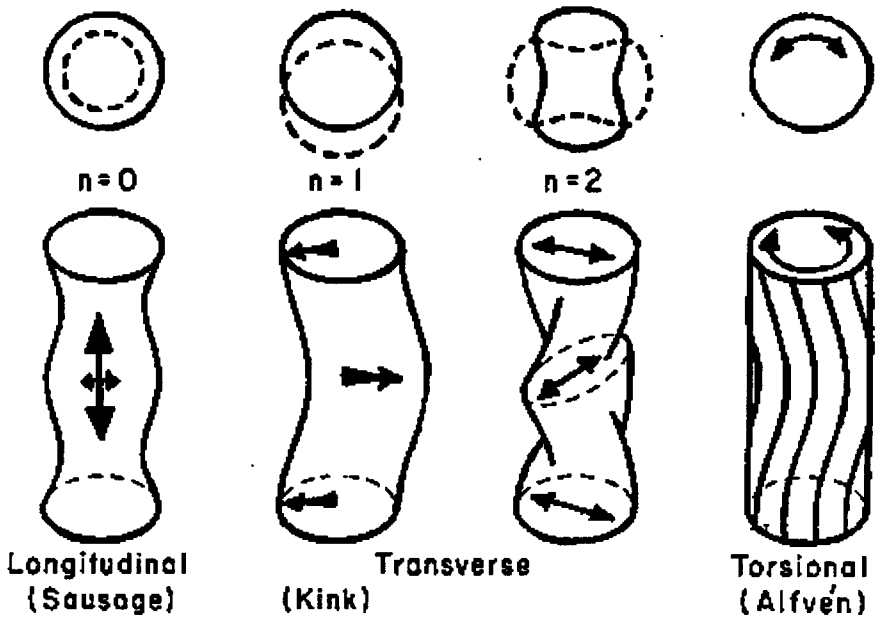


Figure 10.15: Linear wave modes of a flux tube

of these waves is exactly the Alfvén speed: $v_A = B/\sqrt{4\pi\rho_m}$, with ρ_m the gas density inside the flux tube. There is no cutoff frequency, *i.e.*, waves of all frequencies propagate (Roberts 1991).

10.13.2 Longitudinal or sausage mode

The matter inside the tube is compressible and thus supports magneto-acoustic waves. In this instance the restoring forces are the gas and magnetic pressure gradients. The propagation velocity is determined by the tube speed c_T, which is (Roberts 1991):

$$c_T = \frac{c_s v_a}{\sqrt{{c_s}^2 + {v_a}^2}} \tag{10.8}$$

with c_s = the sound speed. In this case a cutoff frequency does exist at $\omega_T \approx 1.2\omega_s \Leftrightarrow$ 160 s < 3 min for typical fluxtube parameters. Only waves above this frequency propagate along flux tubes. Those with longer periods are evanescent, *i.e.*, they decay exponentially with height.

10.13.3 Transverse or kink mode

There are an incompressible Alfvén-like waves. However, in this instance the restoring forces are a combination of the magnetic curvature and the buoyancy. The propagation velocity then becomes equal to the kink speed c_k:

$$c_k = \sqrt{\frac{\rho_m}{\rho_m + \rho_s}} v_A < v_A, \tag{10.9}$$

with ρ_s the density of the surroundings. Here the cutoff frequency is situated at $\omega_k \approx 0.4\omega_s \Leftrightarrow 480$ s = 8 min, which is the typical lifetime of granules (from Roberts 1991).

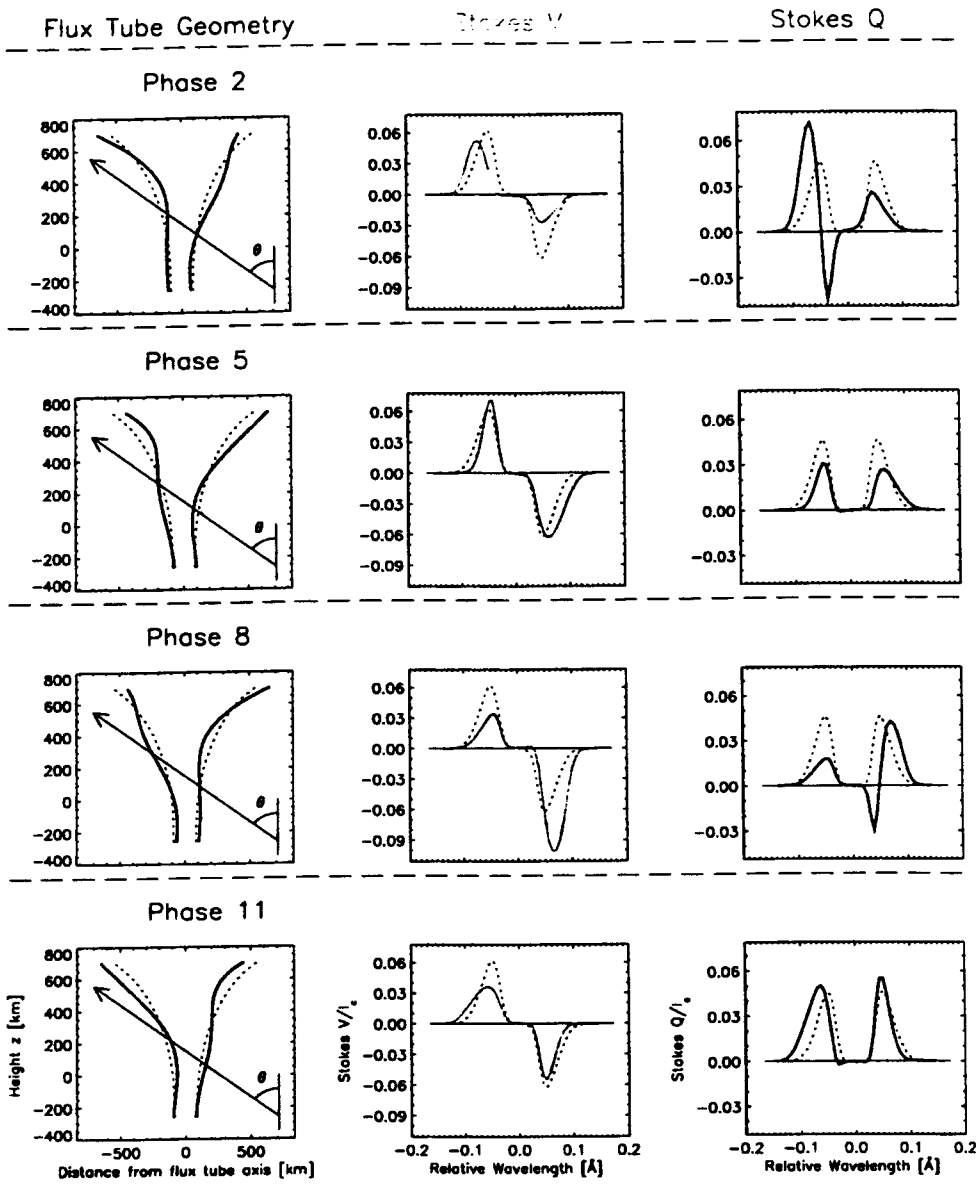


Figure 10.16: Flux tube geometry and their corresponding Stokes V and Q profiles. Note that all Stokes V profiles have asymmetry (Section 10.10.2 and Fig. 10.13). The dotted lines in the left panel show the flux tube in absence of a wave. From Ploner and Solanki (1997)

10.14 Tube geometry change and Stokes profiles

Consider now how flux tube waves influence Stokes profiles. An illustration of this is given in Fig. 10.16, in which in the left hand panels the change in flux tube geometry over a wave period is illustrated. A sample ray (along which Stokes profiles are formed) is also shown. In the central and right-handed panels the Stokes V and Q profiles generated at the four plotted phases are shown (solid lines), compared with the V and Q profiles formed in the absence of a wave (dotted profiles). Note the asymmetry of the V and Q profiles, and how this asymmetry changes over the wave period. Thus, if we could observe with sufficient

spatial resolution, such variable asymmetry would be a certain sign of the presence of a kink wave.

10.15 Spicules & flux tubes

In Section 2.3 the spicules were shortly described and shown. From these observations we know that spicules are elongated, narrow structures seen in emission at the solar limb in the H_α line. In devising models to fit the observations the spicules are viewed as outflows excited by waves propagating along magnetic field lines.

The photospheric source has been thought to be waves generated and propagated along flux tubes. Hence spicules are coupled to the theory of basic wave modes treated in the previous section and to the configuration of magnetic field lines of flux tubes.

Chapter 11

Sunspots

Lecturer S.K. Solanki, notes by R. van Deelen

In the western world sunspots were discovered in 1611 soon after the invention of the telescope. One popular notion was that the dark spots could be not yet molten rocks floating at the surface of the Sun, a giant sphere of lava. Could these rocks be drowned planets? They were wrong.

Wilson showed in 1769 that sunspots are saucer shaped depressions in the photosphere (Foukal 1990). Now we know that the Wilson depression must be caused by the magnetic field, specifically the pressure it exerts. Sunspots are the cross-sections at the photospheric level of giant fluxtubes pointing out through the surface.

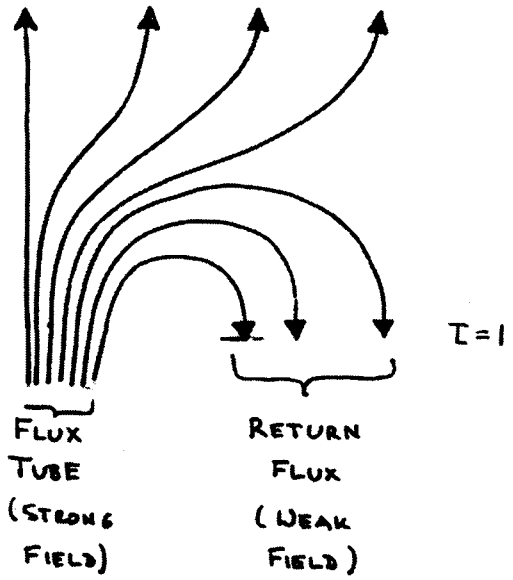


Figure 11.1: Return flux model of the sunspot magnetic field (Osherovich '82, Flå et al. '82).

11.1 Sunspot magnetic fields

The magnetic field is strongest in the central, darkest part of the umbra, where it is also vertical. This $B_0 = B(r = 0)$ is about 2000 – 3500 G. In the penumbra the $\vec{B}$-field is the weakest and nearly horizontal near the outer penumbral boundary. Here $B_p = B(r_p)$ is

700 – 1000 G. The average field-strength of the sunspot is therefore about 1000 – 1500 G. A very rough description of the radial dependence of the sunspot magnetic field throughout the spot is the following:

$$B(r) = B_0 \left(1 - \left(\frac{r}{r_p}\right)^2 \left(1 - \frac{B_p}{B_0}\right)\right) \tag{11.1}$$

and

$$\gamma(r) = 75\frac{r}{r_p} + O(r^2)\,, \tag{11.2}$$

where $\gamma(r)$ is given in degrees.

Outside the sunspot as seen in white light the magnetic field forms a *canopy* with its base in the photosphere. It is called a canopy because the bundles of magnetic field lines which start in the sunspot go up vertically and form a curved structure until at certain distance above the solar surface the field is horizontal and stretches out like a blanket over enormous distances. Just a minor part of the flux ($\sim$ 1%) returns to the solar interior at r_p. The field-strength decreases above the surface with height. At the axis of the spot it follows the approximate formula (for $z > 0$)

$$B_0 \sim \frac{1}{1+z^2}\,. \tag{11.3}$$

In the penumbra the field has a complex small-scale structure; the field along different filaments has different inclination with the inclination differences $\Delta\gamma \lesssim 40°$. The magnetic structure of sunspots is far more complex than originally thought.

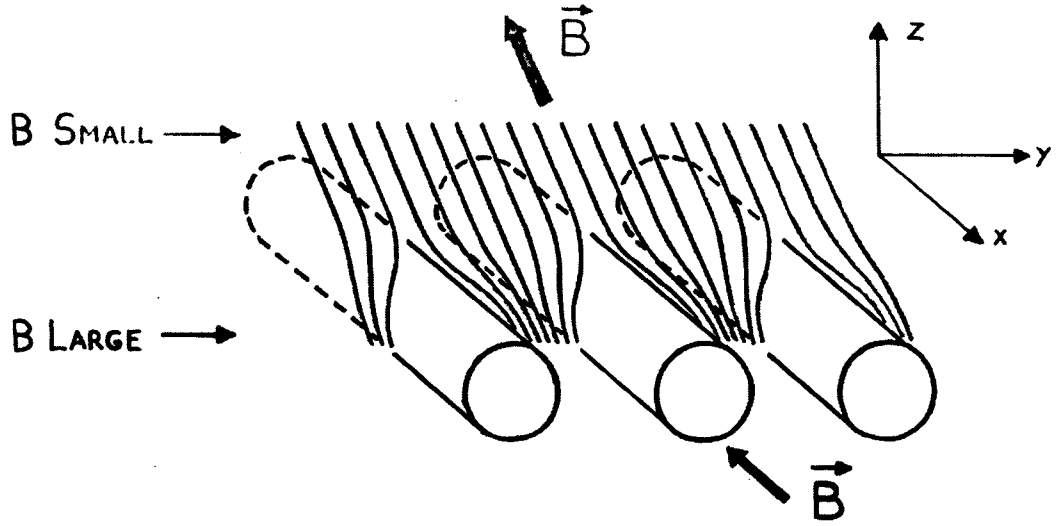


Figure 11.2: Small-scale magnetic structure of a sunspot with uncombed fields. Magnetograms only show large-scale horizontal B-gradient.

11.1.1 Is the divergence of the magnetic field non-zero in sunspots?

If we measure the the different gradients of the B-field components it seems like the $\nabla \cdot \vec{B}$ law is violated, particulary in the sunspot penumbras. If $\partial B_x/\partial x$ and $\partial B_y/\partial y$ are taken

from vector magnetograms and $\partial B_z/\partial z$ from B_z measured at different heights, it is seen that the gradient in the z-direction is an order of magnitude larger than in the horizontal direction:

$$\left|\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y}\right| \ll \frac{\partial B_z}{\partial z}\,. \tag{11.4}$$

A solution to the seeming violation is the idea of *uncombed fields*, where the $\vec{B}$-field has the following configuration as seen in Fig. 11.2. In this configuration obviously flux is conserved (all field lines entering the domain also leave it).

11.1.2 What is the subsurface structure of a sunspot?

Good question. There are two scenarios, the spaghetti or cluster model (b) or the more simpler monolith model (a) (see Fig. 11.3). Modelers tend to use the monolith model due to its simplicity. Reasons for the spaghetti model are that it is expected due to an instability (interchange instability) acting below the solar surface. It also allows energy to enter the umbra from below, since normal convection can penetrate between the spaghetti from below until almost the solar surface.

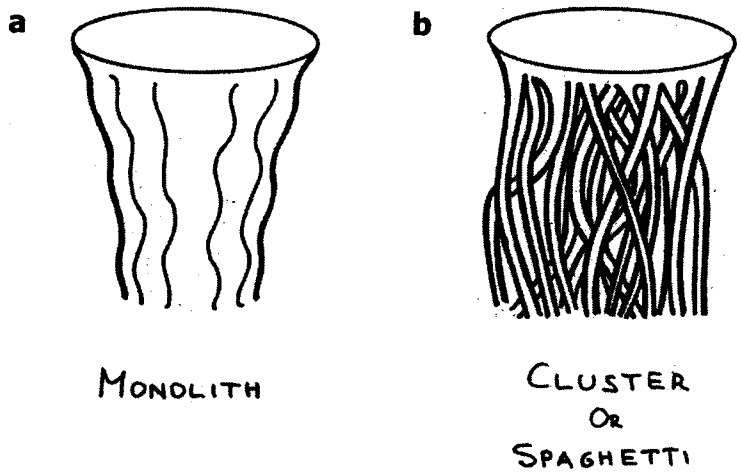


Figure 11.3: Monolith model a) versus spaghetti or cluster b) model of the subsurface structure of a sunspot.

11.2 Penumbral flows

In the penumbra and beyond observed spectral line profiles are distorted; in the diskward side of the penumbra spectral lines are blueshifted and asymmetric to the blue, limbwards the lines are redshifted and asymmetric to the red. This spectroscopic effect was first discovered by Evershed in 1909 and it is now called the *Evershed effect*. It is caused by a radial outflow in the sunspot, but the mechanism behind it (flows or waves?) is still being examined (see Rutten and Schrijver 1994), although currently flows are favoured.

This situation in a sunspot is illustrated in Fig. 11.4. The field and gas are in total pressure equilibrium; in the center of the sunspot the field is strong and the gas pressure

is low, but in the penumbra the gas pressure is large and the magnetic pressure ($B^2/8\pi$) is small. A flow starting in the penumbra can end in a stronger-field magnetic element. There is even more on flows in Section 10.8. Although only 1% of the magnetic flux returns (see Fig. 11.1), 80% of the mass returns into the solar interior.

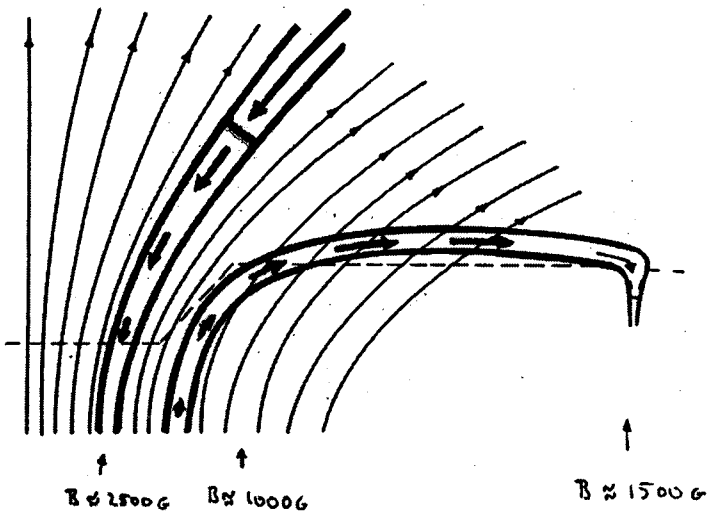


Figure 11.4: Schematic siphon flow model of Evershed effect. Think about sucking on a tube in a gas tank in order to lift the fluid through a higher potential to a lower one; if successful the fluid flows out of the tank until it is empty. Notice that a small difference in B is enough for a large effect. A similar siphon flow can take place on the Sun, along a magnetic loop if the gas pressure is different at the two footpoints of the loop (i.e. at equal geometrical potential). The gas flows from higher to lower gas pressure, i.e from lower to higher field strength.

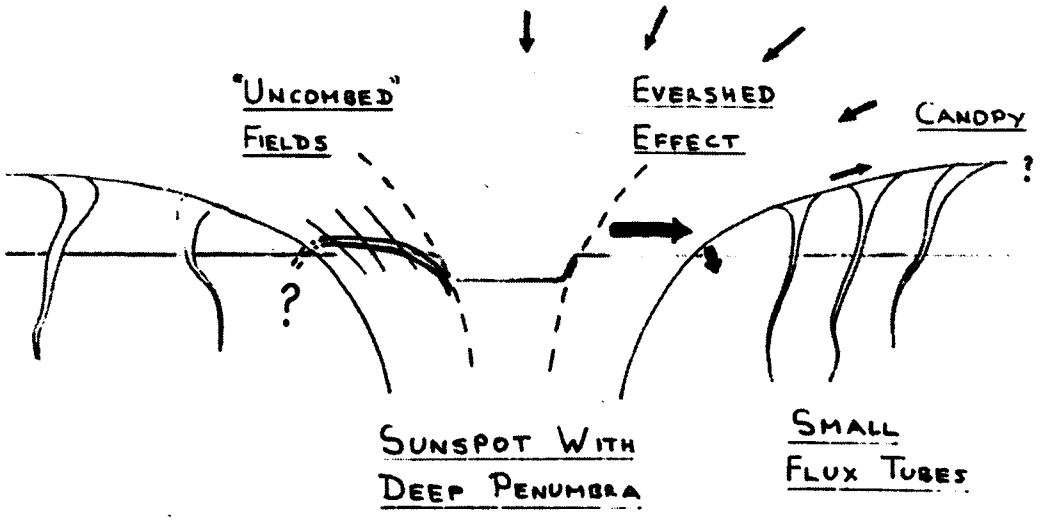


Figure 11.5: Magnetic structure of sunspots

11.3 Umbral brightness

Sunspots are not all equally bright. Early results suggested that larger spots are darker. Later this was thought to be an artifact of straylight: smaller spots are more strongly affected by it. Most recently evidence was found that even after removing the straylight there are still differences in brightness for different sunspots.

The umbral core brightness depends on the phase of the solar cycle. Umbras are cool during the early phase of the cycle by 15 – 20% compared to the late phase. The question is still not completely resolved whether this is a hidden latitude dependence of the umbral brightness. The brightness of sunspots is different in the center than in the limb, because the photons that leave the solar surface originate from areas of different temperatures. At different latitudes we see different temperatures throughout the sunspots that are on the same latitude. Sunspots appear on high latitudes early in the cycle (looking cool) and lower to the equator gradually ("butterfly"-diagram). This could be an explanation of the hidden latitude dependence.

11.4 Thermal structure of sunspots

Clearly there is a relationship between the sunspot's magnetic field and temperature, at a certain geometrical height. This is observed in the umbra, but further away from the center at the umbral boundary the relationship is broken, the field stays about the same, until in the penumbra an other relationship is reached. The temperature of different umbras varies significantly and therefore their brightness does too. Moreover, the temperature also varies within an umbra. On a large scale the temperature in penumbras is about constant, although there is considerable fine-scale structure.

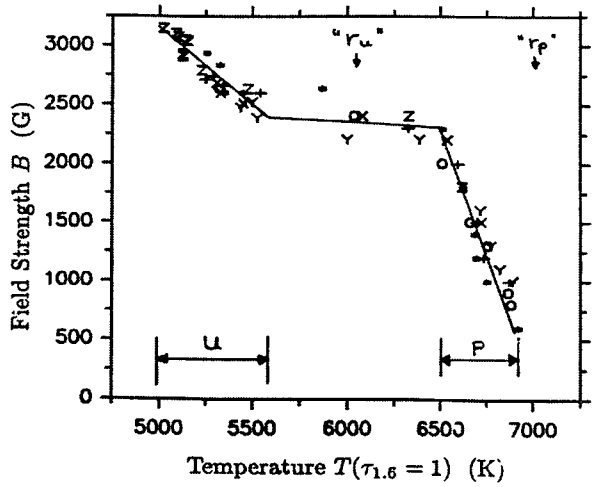


Figure 11.6: Magnetic field-strength versus temperature. Data is taken from infrared spectra of a large symmetric sunspot.

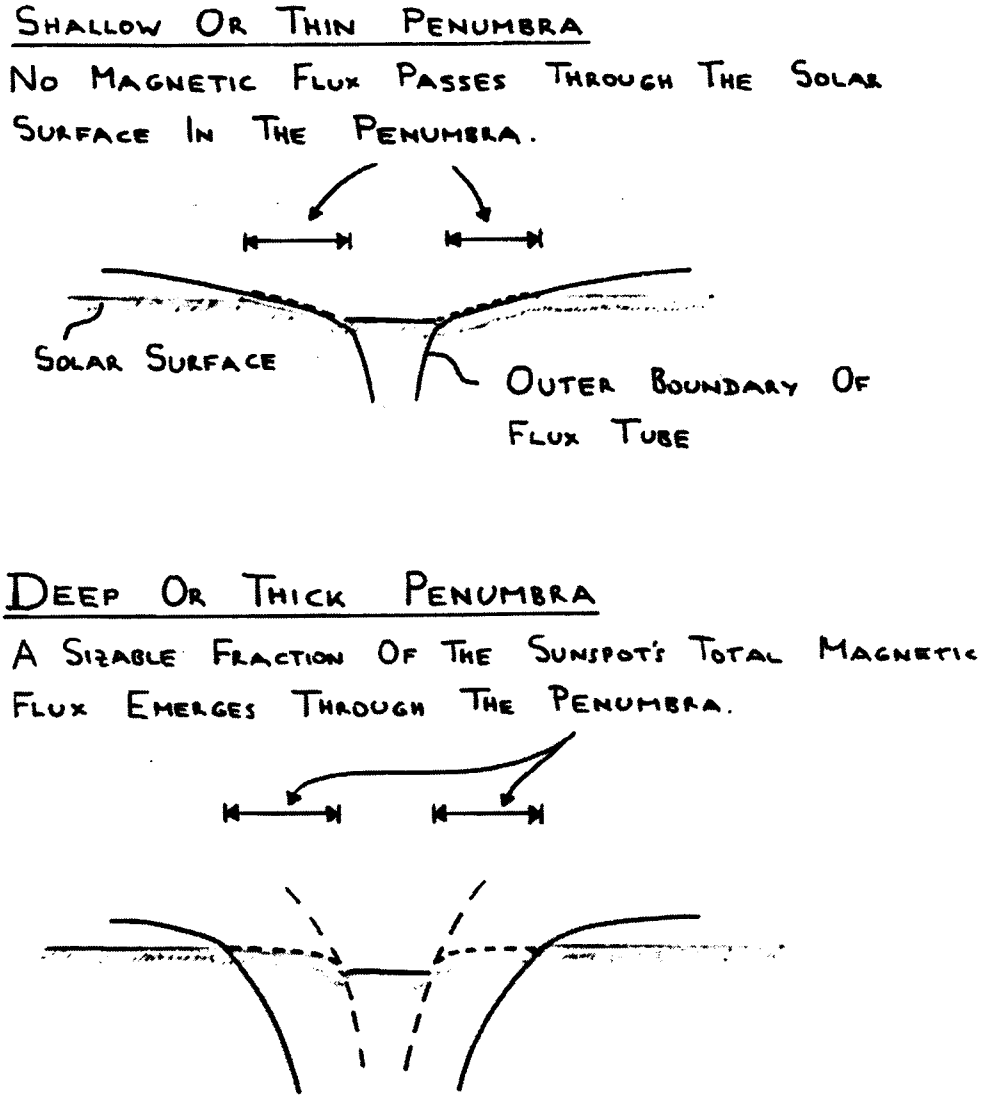


Figure 11.7: Shallow versus deep penumbras. The question is which one is correct? Observers say penumbras are deep.

11.5 Why are sunspots so dark?

Because magnetic pressure in the umbra (near the surface) is large ($\beta < 1$) and because the field lines are tied to the plasma, a so-called Wilson depression occurs. The convection sees the sunspot as an obstacle and diverts the energy flux around it. This causes the umbra to darken. Much of the blocked flux is stored in the convection zone so that it does not reappear as a bright ring around the sunspot.

Most photons escape into space, because the photon mean free path (both vertical as horizontal) is much smaller than the diameter of the sunspot which is about 30 times the altitude of the Wilson depression. Hence these photons do not contribute to the heating.

All these effects make sunspots look much cooler than the non-magnetic atmosphere

at equal geometrical height. It is remarkable that if all of the above darkening aspects of sunspots are taken into account, sunspots actually are brighter than expected. This is caused by the heating by radiation from the sides of the spot.

In order to understand the darkening this excess brightening shown by real sunspots, it is important which model is proposed for the energy transport below the umbra; the spaghetti or the monolith model (see Fig. 11.3).

In the spaghetti or cluster model "normal" convection between the individual fluxtubes takes place. In the other model, the monolith model *magnetoconvection* is proposed, with the determining parameter $\xi = \eta/\kappa$, where η is the magnetic diffusivity and κ is the thermal diffusivity. This ξ determines whether there is oscillatory ($\xi < 1$) or overturning ($\xi > 1$) convection.

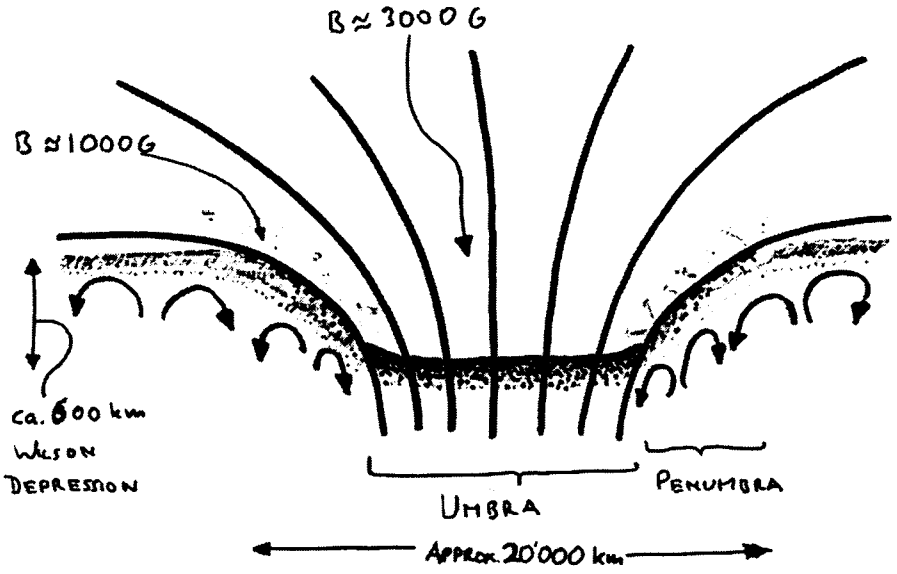


Figure 11.8: Picture of a sunspot (somewhat out of date). The sunspot is heated by the radiation of the penumbra.

We may conclude the following: just below the sunspot energy is mainly transported by convection. The strong magnetic field inhibits the convection, because the gas cannot flow across the field lines.

Chapter 12

MHD and spectral theory

Lecturer J.P. Goedbloed, notes by J. Wiersma

A large part of these lectures is about the surface of the Sun (or stars), by which we mean the photosphere at optical depth, $\tau \approx 1$. In that region the magnetic field is very well approximated by considering magnetic flux tubes that exist virtually as individual entities.

There are, however, many different regimes of the Sun's magnetic field: for example, the interior of the Sun (at $\tau \gg 1$). There, the magnetic field is generated, which means that the number of flux tubes is not conserved.

In this chapter the magnetic field outside the Sun ($\tau \ll 1$) is considered. Here, the dynamics are no longer dominated by the magnetic field but the magnetic field is frozen in a hydro-magnetic flow. The flux tubes that were tied firmly with both ends in the Sun have become undone at one end (or sometimes both) and interact with each other and with other magnetic fields generated by sources outside the Sun (*e.g.*, the Earth's magnetosphere).

First, a theory will be considered that describes the dynamics in this regime and the basic properties of this theory will be established. Then the results of this theory regarding the heliosphere will be treated.

12.1 The solar wind as an energy leak

Very deep inside the Sun, energy is produced by fusion reactions (at least that is what we think happens, it might be little gnomes riding bicycles instead...). This energy has to be transported out of the Sun and this is done in several ways.

For example, neutrinos transport the energy almost directly from the core of the Sun to the outside. Another part is transported by radiation via a long process of absorption and emission in which the emitted gamma radiation is converted into the solar spectrum that the Earth receives every day.

The interesting process here is the solar dynamo that transforms energy into magnetic field which comes up to the surface of the Sun and then flies out in the form of a solar wind transporting magnetic energy away from the Sun. This magnetic field makes up the heliosphere.

But, whereas the Earth's magnetosphere is simply made up of the magnetic field that comes directly from the Earth's dynamo, the magnetic field in the heliosphere is embedded in and transported by a plasma.

It can be concluded that the heliosphere cannot be simply described by flux tubes, nor

can it be described by electro-magnetism alone. Instead, a theory is needed that describes the behavior of plasmas that carry a magnetic field. This theory is called Magneto-Hydro-Dynamics (MHD for short). This section will start with the basic properties of MHD, focusing on those parts which are relevant to the heliosphere.

12.2 A "construction" of magnetohydrodynamics

Plasmas are ionised gases and we study them via their macroscopic properties (such as density, pressure, flow velocity, *etc.*). A theory that deals with such macroscopic quantities is called a fluid theory and the fluid theory of plasmas is MHD. To put it simply MHD consists of the sum of gas dynamics (hydrodynamics), Maxwell theory and the interaction between the two.

To obtain the equations the following method can be used:

From kinetic theory...

Since the plasma is a gas it can be described by kinetic theory, which gives us a distribution function $f_\alpha(\vec{r},\vec{v},t)$ for each particle species α telling us what fraction of those particles have a position $\vec{r}$ and velocity $\vec{v}$ at time t. The fundamental equation for the dynamics is Boltzmann's equation:

$$\frac{\partial f_\alpha}{\partial t}+\vec{v}\cdot\frac{\partial f_\alpha}{\partial \vec{r}}+\vec{a}\cdot\frac{\partial f_\alpha}{\partial \vec{v}}=\left.\frac{\partial f_\alpha}{\partial t}\right|^{\text{coll}} \tag{12.1}$$

where $\vec{a}$ is the sum of all the accelerations the particles are submitted to as a result of the different macroscopic forces working on them and $\partial f_\alpha/\partial t|^{\text{coll}}$ is the change in the distribution function due to collisions. This equation expresses the fact that the number of particles with a certain position and velocity can only change in time (the first term) due to the movements of the particles (the second term), the forces working on them (the third term) and the collisions they undergo (the right-hand-side). Of course, collisions are also accompanied by forces, but they are treated differently in kinetic theory because they depend on the microscopic properties of the gas.

...via two fluid equations...

However, the variables $\vec{r}$ and $\vec{v}$ are not useful for studying the behaviour of the plasmas, because they are microscopic properties of the particles instead of macroscopic quantities that can be measured. To get around this problem Boltzmann's equation has to be rewritten in terms of macroscopic quantities such as density, flow velocity, pressure, *etc.* The trick is to take different moments of Boltzmann's equation, *i.e.*, multiplying the equation with different powers of $\vec{v}$ and integrating over all possible velocities. The macroscopic quantities are then defined as the integrals you are left with (*e.g.*, the particle density of particles of species α is defined as $n_\alpha \equiv \int f_\alpha \mathrm{d}^3v$, *i.e.*, the number of particles at a certain position irrespective of their velocities). This way fluid equations are obtained for each species of particles. Assuming that there are only electrons and one kind of ions this gives us two species of particles; hence the name "two fluid equations".

...to the one fluid equations of MHD

In MHD, however, the plasma is considered as *one* fluid. This means the equations have to be rewritten again so that they only contain "one fluid variables" (such as total

density $\rho \equiv \sum_\alpha m_\alpha n_\alpha$, where $\sum_\alpha$ sums over all particle species and m_α denotes the mass of particles of species α). The two-fluid variables that are left are removed by making several approximations. The whole exercise results in *the non-relativistic, resistive MHD equations* (see Boyd and Sanderson (1969) chapter 3: equations (3-77)–(3-82); their entropy equation really includes the entropy):

Conservation of mass:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \vec{v}), \tag{12.2}$$

Conservation of momentum:

$$\rho \left(\frac{\partial}{\partial t} + \vec{v} \cdot \nabla \right) \vec{v} = -\nabla p + \frac{1}{\mu_0} (\nabla \times \vec{B}) \times \vec{B} + \rho \vec{g}, \tag{12.3}$$

(Near) conservation of entropy:

$$\left(\frac{\partial}{\partial t} + \vec{v} \cdot \nabla \right) p = -\gamma p \nabla \cdot \vec{v} + \frac{\gamma - 1}{\mu_0^2} \eta (\nabla \times \vec{B})^2, \tag{12.4}$$

(Near) conservation of magnetic flux:

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times (\vec{v} \times \vec{B}) - \frac{1}{\mu_0} \nabla \times (\eta \nabla \times \vec{B}), \tag{12.5}$$

The equation of magnetic fields:

$$\nabla \cdot \vec{B} = 0. \tag{12.6}$$

These are eight partial differential equations (two scalar equations and two vector equations) for the eight macroscopic variables: the density ρ, the three components of the flow velocity $\vec{v}$, the pressure p and the three components of the magnetic field $\vec{B}$ with an extra restriction ($\nabla \cdot \vec{B} = 0$).
Their meanings are:

- **Conservation of mass.** This equation says that the density can only change at a point if this is due to a flow of mass to or from a point. Hence there is conservation of mass.

- **Conservation of momentum.** The operator between brackets on the left-hand-side is the Lagrangian derivative. This means that the left-hand-side is the change in velocity as experienced when moving *along* with the flow multiplied with the density. In other words it is the change in momentum (density) as experienced by the particles in the flow.

  This should be equal to the total force the particles feel. Therefore the right-hand-side contains those forces: the first term is the force due to pressure gradients, the second term is the electro-magnetic force (μ_0 is the magnetic permeability) and the last term is gravity (with $\vec{g}$ the gravitational acceleration).

  As you can see the electric field $\vec{E}$ is absent. This reflects that, in non-relativistic plasmas, the physics of the electric field can be left out. (See Sect. 4.3.1 for an explanation of what this means).

- **(Near) conservation of entropy.** In this equation γ is the ratio of specific heat at constant pressure to the specific heat at constant volume. It is just a constant here ($\gamma = 5/3$). The symbol η stands for the resistivity of the plasma. It is the factor of

proportionality in Ohm's law ($\vec{E} + \vec{v} \times \vec{B} = \eta \vec{j}$, in which $\vec{j}$ is the current density which can be eliminated everywhere by using Ampère's law $\vec{j} = \nabla \times \vec{B}/\mu_0$ since the electric field can be left out; see Eq. (4.5) in Sect. 4.3.1).

As you may have noticed there is no entropy in this equation. It can actually be rewritten in terms of the entropy because entropy is a function of density and pressure for ideal gases ($s = C_V \ln(p\rho^{-\gamma}) + \text{const.}$, s is identified with the entropy per unit mass and C_V is the specific heat of the plasma at constant volume; the plasma is assumed to behave as an ideal gas). If you do this the equation will turn out to be a conservation law for the entropy if it wasn't for the last term in this equation: the one containing the resistivity. Hence the name. (See Boyd and Sanderson page 105)

This means that entropy is not conserved, if the resistivity of the plasma is not zero. In other words the processes in the plasma become irreversible: something gets lost. This expresses the fact that if the plasma is resistive, the energy in the currents running through the plasma gets transformed into heat, which is an irreversible process.

- **(Near) conservation of magnetic flux.** This equation is Faraday's law $\partial \vec{B}/\partial t = -\nabla \times \vec{E}$ with $\vec{E}$ eliminated using Ohm's law. If the resistivity is zero it can be rewritten in a conservation equation for the flux going through a surface element that moves with the flow (see Sect. 4.5.1). To make this more visual, take Fig. 12.1. It depicts a flux tube that is being pinched by the flow. The directions of the magnetic field and the flow velocity have been indicated with **B** and **v** respectively. The direction of their cross product has also been indicated (the dashed arrows). As you can see, in this case, the dashed arrows have a vorticity (indicated by curl($\mathbf{v} \times \mathbf{B}$)) that is directed along the direction of the magnetic field. According to the flux conservation equation this vorticity is equal to the change in the magnetic field. So the magnetic field gets stronger which should happen if there is flux conservation since the area of the cross-section of the flux tube gets smaller.

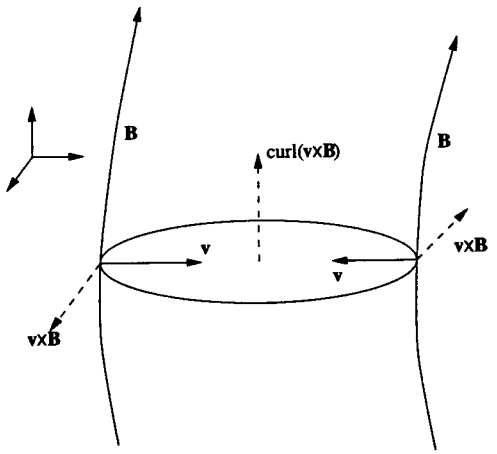


Figure 12.1: A flux tube that is being pinched by the flow. Because of flux conservation the magnetic field has to become stronger since the surface area of the cross-section of the tube becomes smaller.

The term containing the resistivity expresses the fact that if currents get dissipated (see previous paragraph) the magnetic field also gets dissipated (*cf.* Ampère's law $\nabla \times \vec{B} = \mu_0 \vec{j}$).

- **The equation of magnetic fields** states that there are no magnetic monopoles. In other words field lines can not end anywhere. This fact is the reason why magnetic field configurations can never be spherically symmetric.

In most plasmas of interest the resistivity is very low and therefore one usually sets $\eta = 0$. The theory that results from this last assumption is called **Ideal MHD**.

Notes

The following remarks have to be made:

- The velocity $\vec{v}$ in the MHD equations is not the same velocity as the velocity in kinetic theory. In kinetic theory the velocity is the velocity of individual particles and the theory only gives an estimate of how many of the particles have that velocity. In the MHD equations the velocity is the center of mass velocity of a plasma element and the theory gives the evolution of such elements and their velocities. (In Boyd and Sanderson the distinction is made by denoting the microscopic velocity by v and the macroscopic velocity by U.)
- The assumption of ideal MHD, namely that the resistivity is zero, is not always valid. If you look at the terms in the MHD equations containing the resistivity you will see that those terms contain derivative operators as well (curls of the magnetic field). This means that if the gradients get very large (large currents) those terms may not be small enough to be neglected.
- In the heliosphere the density is so low that electrons and ions do not interact with each other very much and they no longer have the same temperature. This means that their behavior can differ considerably. In the process of going from two fluid to one fluid equations these differences have been averaged out. Therefore MHD gives no information about such differences. To consider these differences two fluid theory is needed, but two fluid theory is not very different from one fluid theory (except for the two temperatures).

12.3 Waves in MHD

To get a better feel for the dynamics, consider first the 'simple' case of a static homogeneous plasma. Perturbations are used to study the dynamics (the plasma is given a small kick).

Let's first look at the unperturbed state of the plasma and denote all quantities in this state with subscripts 0. Because the plasma is static the velocities will be zero ($\vec{v}_0 = 0$) and all the other equilibrium quantities (ρ_0, p_0 and $\vec{B}_0$) will be constant in time and, since it is homogeneous, they will also be constant in space.

For such a situation the MHD equations (12.2)–(12.6) are only satisfied if gravity is not present ($\vec{g} = 0$). However, in the case without gravity, the MHD equations are trivially satisfied since they only contain derivatives.

12.3.1 The linearised MHD equations

Now consider the perturbations on this background field. The different quantities of the plasma will differ from their equilibrium values:

$$\begin{aligned} \rho &= \rho_0 + \rho_1, & \vec{v} &= \vec{v}_0 + \vec{v}_1 = \vec{v}_1, \\ p &= p_0 + p_1, & \vec{B} &= \vec{B}_0 + \vec{B}_1, \end{aligned} \tag{12.7}$$

here the perturbations have been indicated with subscripts 1. Inserting these expressions in the MHD equations (12.2)–(12.6) and assuming that these perturbations are very small, all terms that are of second order or higher in the perturbations can be thrown away. This results in the *linearised MHD equations*:

$$\frac{\partial \rho_1}{\partial t} = -\nabla \cdot (\rho_0 \vec{v}_1), \tag{12.8}$$

$$\rho_0 \frac{\partial \vec{v}_1}{\partial t} = -\nabla p_1 + \frac{1}{\mu_0}(\nabla \times \vec{B}_1) \times \vec{B}_0, \tag{12.9}$$

$$\frac{\partial p_1}{\partial t} = -\gamma p_0 \nabla \cdot \vec{v}_1, \tag{12.10}$$

$$\frac{\partial \vec{B}_1}{\partial t} = \nabla \times (\vec{v}_1 \times \vec{B}_0), \tag{12.11}$$

$$\nabla \cdot \vec{B}_1 = 0. \tag{12.12}$$

To solve these equations introduce $\vec{\xi}(\vec{r}, t)$: the displacement of a plasma element at position $\vec{r}$ from its equilibrium position due to the perturbations. The velocity perturbation $\vec{v}_1$ is the change in position with time of the plasma elements:

$$\vec{v}_1(\vec{r}, t) = \frac{\partial \vec{\xi}}{\partial t}. \tag{12.13}$$

Using this expression to eliminate $\vec{v}_1$ from the linearised MHD equations the mass, entropy and flux equations can be turned into algebraic equations relating ρ_1, p_1 and $\vec{B}_1$, respectively, to $\vec{\xi}$ by integrating them over time (remember that all quantities with subscripts 0 are just constants):

$$\begin{aligned} \rho_1 = -\rho_0 \nabla \cdot \vec{\xi}, \qquad p_1 = -\gamma p_0 \nabla \cdot \xi, \\ \vec{B}_1 = \nabla \times (\vec{\xi} \times \vec{B}_0), \end{aligned} \tag{12.14}$$

(note that $\nabla \cdot \vec{B}_1 = 0$ is always satisfied in this case). Inserting these expressions for p_1 and $\vec{B}_1$ into the momentum equation (12.9) gives us a second order differential equation for $\vec{\xi}$:

$$\rho_0 \frac{\partial^2 \vec{\xi}}{\partial t^2} = \gamma p_0 \nabla \nabla \cdot \vec{\xi} - \frac{1}{\mu_0} \vec{B}_0 \times \Big(\nabla \times (\nabla \times (\vec{\xi} \times \vec{B}_0)) \Big), \tag{12.15}$$

which has the form of a wave equation: second order derivatives in time related to second order spatial derivatives.

12.3.2 Alfvén dynamics

In 1942 H. Alfvén showed that in conducting liquids, such as incompressible plasmas, waves can occur (See Alfvén 1942). This discovery got him the Nobel Price. He called

these waves electromagnetic-hydrodynamic waves, known today as *Alfvén waves*. They propagate along the magnetic field with a velocity equal to $B_0/\sqrt{\mu_0\rho_0}$, the Alfvén velocity.

These waves are one of the solutions of Eq. (12.15). To represent them mathematically use plane waves:

$$\vec{\xi}(\vec{r},t) \propto e^{i(\vec{k}\cdot\vec{r}-\omega t)}, \tag{12.16}$$

in which $\vec{k}$ is the wave vector and ω is the wave frequency. Solutions of this type are typical for wave equations and they represent perturbations that have a size that is homogeneous in the plane perpendicular to $\vec{k}$ and varies as a sine function of the distance in the direction of $\vec{k}$. This whole configuration then shifts in the direction of $\vec{k}$ with time.

Note that such a solution is possible because the equilibrium situation is homogeneous and constant in time.

After substituting this solution for $\vec{\xi}$ in the wave equation (12.15), the equation stays the same but with time derivatives replaced by multiplication with $-i\omega$ and spatial derivatives replaced by multiplication with $i\vec{k}$ which yields an algebraic equation:

$$-\rho_0\omega^2\vec{\xi} = -\gamma p_0(\vec{k}\cdot\vec{\xi})\vec{k} + \frac{1}{\mu_0}\vec{B}_0 \times \Big(\vec{k}\times(\vec{k}\times(\vec{\xi}\times\vec{B}_0))\Big), \tag{12.17}$$

which is called the dispersion equation. Alfvén's solution is the case for which the wave vector is perpendicular to the displacement $\vec{\xi}$ (this condition is always satisfied in incompressible liquids which is the reason why Alfvén spoke of conducting liquids instead of gases). In this case the waves are called transverse waves. The dispersion equation then simplifies to:

$$\rho_0\omega^2 = (\vec{B}_0\cdot\vec{k})^2/\mu_0, \tag{12.18}$$

which is the dispersion equation for waves with a phase speed $v_{\rm ph} \stackrel{\rm def}{=} \omega/|\vec{k}| = B_0\cos\theta/\sqrt{\rho_0\mu_0}$, in which θ is the angle between the wave vector and the magnetic field. In other words if the wave propagates parallel to the magnetic field its speed is equal to the Alfvén speed $b \equiv B_0/\sqrt{\rho_0\mu_0}$.

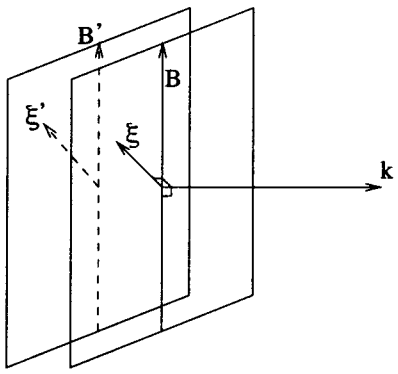


Figure 12.2: Marginal stability of Alfvén dynamics: two planes have been drawn that are perpendicular to the wave vector $\vec{k}$. The magnetic field and the displacement both lie in those planes. Due to the difference in displacement the planes are shifted with regard to each other, but since the magnetic field is homogeneous nothing changes.

The Alfvén dispersion equation (12.18) also shows the velocity of Alfvén waves is zero when the wave vector is perpendicular to the magnetic field. In other words, (incompressible) perturbations which are homogeneous along the field lines won't propagate and the

plasma stays static ($\partial\vec{\xi}/\partial t = 0$). This is to be expected after seeing what actually happens. The situation is shown in Fig. 12.2. The planes that are drawn in that picture are perpendicular to the wave vector $\vec{k}$. In other words, the size of the displacement is the same everywhere in those planes (*cf.* the discussion of plane waves after Eq. (12.16)) and varies perpendicular to them. The magnetic field lines and the displacement vector both lie in those planes since we have chosen the wave vector perpendicular to both ($\vec{k} \perp \vec{\xi}$ for transverse waves, $\vec{k} \perp \vec{B}$ for marginal stability). Because the displacement lies in the planes and is of the same size all over the planes it can also be said that the planes as a whole are displaced with respect to each other. But since the magnetic field is homogeneous and also lies in those planes the perturbed situation is just the same as the unperturbed situation and is thus in equilibrium. (Note that in Eq. (12.14) all perturbations are zero if $\vec{k} \perp \vec{B}$ and $\vec{k} \perp \vec{\xi}$.)

Another important feature is that Alfvén waves can be used to diagnose magnetic fields: they *sample* the magnetic field lines. In addition Alfvén wave dynamics are often dominant in plasmas and determine the behavior for a large part.

12.3.3 Other MHD waves

Alfvén waves are not the only solutions of the dispersion equation (12.17). There are two other solutions: the fast and the slow magneto-acoustic waves. Their dispersion equations are more complicated than Eq. (12.18). To visualise them look at the Friedrichs diagrams which are shown in Fig. 12.3. They schematically show the phase and group speed dependencies on the angle θ between the direction of propagation and the magnetic field (the group speed is defined as $v_{\text{group}} \stackrel{\text{def}}{=} \partial\omega/\partial|\vec{k}|$ and is the speed with which a superposition of plane waves propagates through the plasma). In the left diagram the phase velocities are shown. To simplify the explanation a direction is indicated by a vector labeled **k** which makes an angle labeled θ with the direction of the magnetic field. The distance between the origin and a point where this vector crosses one of the lines indicates the size of the speed of propagation in that direction for that kind of waves. So for the direction indicated, the fast waves roughly go one and a half times as fast as the Alfvén waves and somewhat less than three times as fast as the slow waves.

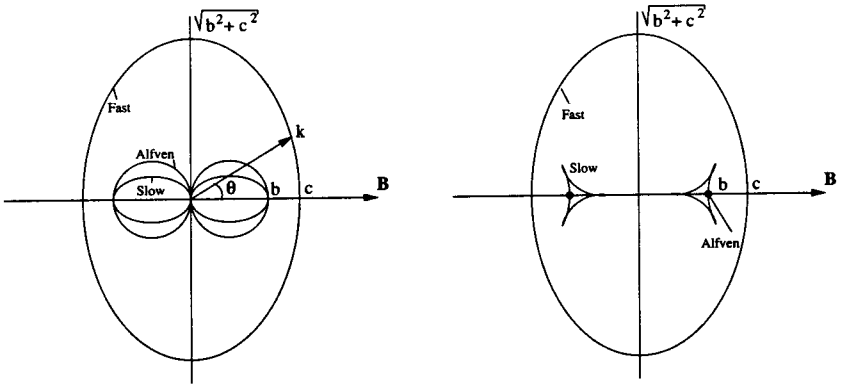


Figure 12.3: Phase and group diagrams of the three MHD waves. The x-axis is the direction of the magnetic field. The vector labeled **k** is a direction of propagation which makes an angle θ with the magnetic field. The values on the axes are $b \equiv B_0/\sqrt{\rho_0\mu_0}$, the Alfvén speed and $c \equiv \sqrt{\gamma p_0/\rho_0}$, the sound speed in the absence of a magnetic field. Here, b has been chosen smaller than c but this need not always be the case.

The phase diagram is only valid for homogeneous plasmas, however, and is therefore

not very useful to apply to the solar wind. The group diagram, on the other hand, indicates the speed of 'wave packets' and as a special case thereof point-like disturbances (like a rock that is thrown in the water). So if the plasma is no longer homogeneous the group diagram can still be applied for very localised disturbances since they do not feel the rest of the plasma (whereas the plane waves of the phase diagram are always infinitely broad).

What the group diagram tells us is that an 'Alfvén wave packet' (an undefined entity as of yet) will *only* propagate along the field lines since their group diagram only consists of two dots on the field line axis. As will be shown later, in inhomogeneous plasmas waves can only have certain frequencies ω and in particular the Alfvén waves are restricted to a certain frequency range so that they can be uniquely identified and can indeed be used to sample the magnetic field as stated at the end of the previous section.

Note: the direction of propagation for plane waves is always equal to the direction of the wave vector $\vec{k}$ which is why **k** could be used to indicate that direction in the phase diagram. The direction of propagation of a wave packet is, however, *not* equal to the direction of its collective $\vec{k}$. *In the group diagram the group speed is shown with respect to the direction of propagation.* If you would plot the group speed against the direction of $\vec{k}$ you would get very different results.

12.4 Spectral theory

So far, a very unphysical case has been considered: infinite, homogeneous, constant plasmas do not exist in real life. A more general result has to be obtained.

Actually, already one general result has been obtained: the Friedrichs group diagram which is also valid locally in an inhomogeneous plasma. However, what happens with the plasma as a whole also needs to be known: is a given configuration stable for example. The theory that attempts to give an answer to that question is spectral theory.

In the previous section an expression for the frequency ω was obtained for the different MHD waves. There, the frequency could go through a whole continuum of values because the wave vector could be chosen freely in a homogeneous configuration. In an inhomogeneous configuration, maybe even confined, plane waves are no longer solutions of the problem. The solutions can still be characterised by a frequency ω, however, but this frequency can no longer have all possible values. It might on the other hand be imaginary, corresponding to an exponentially growing or decaying solution: an unstable situation. Spectral theory tries to find all possible values of this frequency for given situations.

12.4.1 Stability and instability

One of the aims of spectral theory is to find out which plasma configurations are stable and which are not. Such stability analysis is important since it tells us a lot about the dynamics of a configuration: unstable situations will usually lead to short time-scale dynamics if they are perturbed so that unstable configurations might be used to explain observations of, for example, active regions on the Sun. Stable configurations lead to stationary states with long time-scales and are important in the solar wind which is basically stationary.

Intuitively, the stability is determined in the way sketched in Fig. 12.4: if a system is in a state 0 then this state is stable if a perturbation from that state to a state 1 leads to the system having a higher energy. In that case the urge of the system to lower energy will pull it back to state 0. If, however, the energy of the system is lower in state 1, the

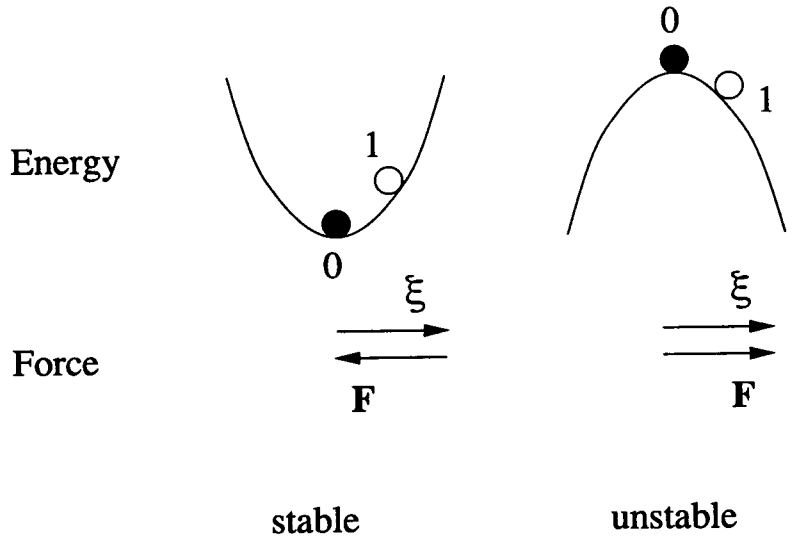


Figure 12.4: Intuitive approach of stability using energy principle or forces.

perturbation will grow because the urge for lower energy will pull the system *away* from state 0 and thus state 0 is unstable.

Another way of looking at it is using forces and displacements as sketched at the bottom of Fig. 12.4. If a perturbation causes a displacement ξ and this induces a force $\mathbf{F}$ that is directed in the opposite direction of ξ, this force will push the system back to the original state and that state is stable (at least, for perturbations in this particular direction of ξ). If the induced force points in the same direction as ξ the original state is unstable because the system will evolve away from it.

As was indicated in the previous paragraph the direction of the displacement might be important for determining the stability. If a configuration is found to be stable for perturbations in one particular direction this does not mean that the system is stable under all conditions. If, on the other hand, one direction is found for which it is unstable it is certainly not stable.

To illustrate this take the example of a system that is dominated by Alfvén wave dynamics. As seen in the previous section such a system will begin to oscillate if it is perturbed by a plane wave that has a component of the wave vector along the magnetic field. So if the Alfvén waves are really dominating, the system is stable for such perturbations. But if there is a wave with its wave vector perpendicular to the magnetic field the oscillating frequency is zero, *i.e.*, nothing happens if it is up to the Alfvén waves. This is called *marginal stability*: the perturbations do not grow but they do not go away either. This means that if there are any other factors in the dynamics *they* will determine if the state is stable for such perturbations. This makes the stability analysis a very complex problem.

12.4.2 Force operator formalism

In Sect. 12.3.1 a second order differential equation was obtained for the displacements in a homogeneous plasma due to perturbations via a simple method of linearising the MHD equations (see Eq. (12.15)). There the equilibrium situation was trivial because the background plasma was homogeneous. The same can be done for inhomogeneous plasmas:

start with an equilibrium situation and then study perturbations on this equilibrium by linearising the equations in these perturbations.

The equilibrium situation becomes non-trivial now, however, since the MHD equations are not satisfied for all inhomogeneous plasmas.

Take for example a static inhomogeneous ideal plasma. This means that again all derivatives in time are zero and that the equilibrium velocity is zero ($\vec{v}_0 = 0$). Using this in the ideal MHD equations (Eqs. (12.2)–(12.6) with $\eta = 0$) all are trivially satisfied except the momentum equation (12.3) and $\nabla \cdot \vec{B}_0 = 0$. This means there has to be force balance in equilibrium:

$$\nabla p_0 = \frac{1}{\mu_0}(\nabla \times \vec{B}_0) \times \vec{B}_0 + \rho_0 \vec{g}, \tag{12.19}$$

which restricts the equilibrium situation.

Introduce perturbations on this equilibrium again (as in Eq. (12.7) in the previous section), linearise the equations in these perturbations and again introduce the displacement $\vec{\xi}(\vec{r}, t)$ and use it to get algebraic equations for ρ_1, p_1 and $\vec{B}_1$. The momentum equation is left, which again becomes a second order differential equation of the form:

$$\rho_0 \frac{\partial^2 \vec{\xi}}{\partial t^2} = \mathbf{F}[\vec{\xi}], \tag{12.20}$$

$\mathbf{F}$ is called the force operator and symbolises the right-hand-side of the momentum equation. It is an operator working on the displacement $\vec{\xi}$ containing only spatial derivatives.

This is completely analogous to the derivation of the wave equation (12.15) for homogeneous plasmas. Only, the operator $\mathbf{F}$ is more complicated than the right-hand-side of Eq. (12.15). This does mean that plane waves are no longer solutions of Eq. (12.20). But because the only time derivative is still the second order derivative on the left-hand-side, solutions can be used that have a time behavior of the simple form

$$\vec{\xi}(\vec{r}, t) \propto \mathrm{e}^{-i\omega t}.$$

This transforms Eq. (12.20) into:

$$-\rho_0 \omega^2 \vec{\xi} = \mathbf{F}[\vec{\xi}]. \tag{12.21}$$

As has been said earlier, solutions of this form will oscillate if ω is a real number and grow exponentially if ω is an imaginary number. This means that if, for a given configuration, Eq. (12.21) only has solutions for real ω that configuration is stable. If, however, there is even just one solution that has an imaginary ω the configuration is unstable.

You can also see this if you look at what happens to Eq. (12.21) for real and imaginary ω. Because of the construction of the solutions the force is always directed parallel to the displacement, so, using the force approach of stability, if ω is real then ω^2 is positive and the force $\mathbf{F}$ points in the opposite direction from $\vec{\xi}$. This means that that particular solution is stable. If ω is purely imaginary the force points in the same direction as the displacement and the plasma is unstable. Q.E.D.

Eq. (12.21) is an eigenvalue equation for the operator $\mathbf{F}/\rho_0$ with eigenvalues ω^2. It can be solved with methods similar to those used in quantum mechanics to solve the Schrödinger equation. A Hilbert space can be constructed from the displacements $\vec{\xi}$ (provided that

the kinetic energy of the plasma stays finite) and all mathematical tools provided by the Hilbert space formalism can be used.

This method of solving the linearised MHD equations for inhomogeneous plasmas is called the *force operator formalism*.

In particular it can be proven that the force operator $\mathbf{F}$ is self-adjoint, which means that its eigenvalues are real numbers. This means that the eigenvalues of $\mathbf{F}/\rho_0$ are real as well since ρ_0 is just a real (positive) number. And the eigenvalues of $\mathbf{F}/\rho_0$ are exactly those values of ω^2 for which the linearised MHD equations have a solution.

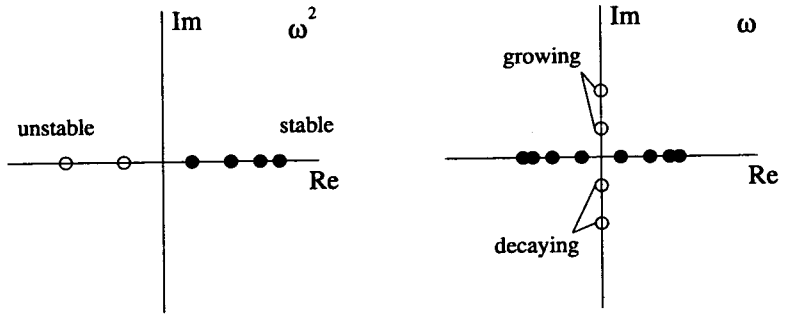


Figure 12.5: Example of a set of ω's for which the eigenvalue equation (12.21) has a solution. The eigenvalue equation yields values for ω^2 (left picture), the corresponding values for ω are indicated on the right. Eigenvalues corresponding to stable solutions are indicated by filled dots, those corresponding to unstable solutions are indicated by open circles.

In Fig. 12.5 an example is given of a set of eigenvalues. They are plotted in the complex plane. Since all values of ω^2 are real all the dots in the left picture lie on the real axis. When ω^2 is positive the two corresponding roots ω lie on the real axis as well and $\exp(-i\omega t)$ is an oscillating function of time representing a stable solution. But if ω^2 is negative the corresponding roots are purely imaginary and $\exp(-i\omega t)$ becomes an exponential function of t representing an unstable solution.

12.4.3 The spectrum of ideal MHD

The set of eigenvalues for which the MHD equations have solutions is called the spectrum. If this spectrum can be found for a given configuration, it can be established whether that configuration is stable. But the problem can also be turned around: try to use the spectrum as an observational tool. The oscillations in plasmas may be observed to make an experimental spectrum of the occuring frequencies. If these spectra can be matched with theoretically derived spectra a lot can be learned about the structure of these plasmas.

This has worked very well in helioseismology. There the spectral problem is solved for the hydrodynamic equations with gravity but without the magnetic field (which does not play a big role *inside* the Sun). The experimental spectrum is obtained by looking at oscillations in Doppler observations of light integrated over the disk of the Sun. In geophysical seismology even bigger successes have been achieved.

The eigenvalue problem (12.21) of MHD can, however, not be as easily solved. The problem in this case is much more complex due to the anisotropy introduced by the magnetic field. As has been stated earlier spherical symmetry is not possible which really reduces the possibilities of solving the problem.

Consider the spectrum of what seems to be still a simple case. But it will turn out that even this spectrum is non-trivial. Take a plasma that is homogeneous in two directions

(call them y and z) and inhomogeneous in a third (the x direction) and take the magnetic field in the yz-plane and gravity along x. Then the solution can be represented as a plane wave in the homogeneous directions:

$$\vec{\xi}(x,y,z,t) \propto f(x)\mathrm{e}^{i(k_y y + k_z z - \omega t)}, \tag{12.22}$$

in which k_y and k_z are the wave numbers in the y and z direction respectively. Note that this is the same as the plane wave solution in the homogeneous case, except that the behavior as a function of x is unspecified as of yet.

All the equilibrium quantities will now be functions of x and to correspond to an equilibrium situation they have to satisfy the force balance equation (12.19). If the equations are linearised and expression (12.22) is substituted for $\vec{\xi}$ in the result, a differential equation for $f(x)$ is obtained which has the following form (see Goedbloed 1971):

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{N(x;\omega)}{D(x;\omega)}\frac{\mathrm{d}f}{\mathrm{d}x}\right) - g(x;\omega)f(x) = 0, \tag{12.23}$$

where $N(x;\omega)$, $D(x;\omega)$ and $g(x;\omega)$ are functions of x and ω (ω is a parameter). A differential equation is obtained since only derivatives to y, z and t could be eliminated using solutions of the form (12.22) and the derivatives to x remain. The factor N/D is proportional to a function of ω:

$$\frac{N(x;\omega)}{D(x;\omega)} \sim \frac{(\omega^2 - \omega_{\mathrm{A}}^2(x))(\omega^2 - \omega_{\mathrm{S}}^2(x))}{(\omega^2 - \omega_{\mathrm{s0}}^2(x))(\omega^2 - \omega_{\mathrm{f0}}^2(x))},$$

where we denote the numerator of this expression with N and the denominator with D; all the $\omega_{\ldots}(x)$ with subscripts are functions of x. For example $\omega_{\mathrm{A}}(x)$ is the local Alfvén frequency at position x:

$$\omega_{\mathrm{A}}(x) = \frac{B_{0x}(x)k_x + B_{0y}(x)k_y}{\sqrt{\rho_0(x)\mu_0}},$$

which is a function of x since all equilibrium quantities depend on x.

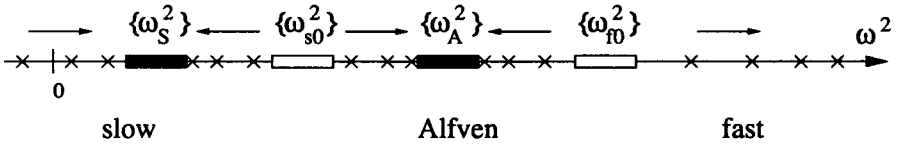


Figure 12.6: The spectrum of a slab of plasma that is inhomogeneous in one direction. The problem has solutions for given k_y and k_z only for those values of ω that have been indicated by crosses and the boxed regions. The directions of increasing degree of spatial oscillation are indicated by arrows. When this degree increases with increasing ω the behavior is called Sturmian. The opposite case is called anti-Sturmian.

Now choose values for k_y and k_z and try to solve Eq. (12.23) for those values. The spectrum is shown in Fig. 12.6. It shows that this can not be done for all values of ω. There are only solutions for certain discrete values of ω and for the ranges of values indicated with boxes. The black boxed continua occur because the differential equation (12.23) has solutions for those ω for which $N(x)$ is zero for some value of x. If $B_0(x)$ and $\rho_0(x)$ span some continuous region of values there will also be a continuous region of values for ω for which $N(x)$ is zero for some value of x. The solutions for these ω can be very localised (delta functions). If the plasma is incited from the outside with a frequency that lies in one of those ranges you get resonant behavior: the amplitude of the disturbances can get

infinite in a very localised region around the position x where $N = 0$; the solution permits this. This means that you get high gradients and currents and in practice this means that the resistive terms in the MHD equations start to count and dissipation will occur heating the plasma; this is called resonant absorption because energy gets absorbed from the exciting process to heat the plasma.

The unfilled boxes indicate the regions where $D = 0$ for some value of x. They have caused quite some controversy (*e.g.*, see Goedbloed 1998). Because, although they span a continuous region, they are not *mathematical* continua. The difference is, to put it simply, that *mathematical* continua correspond to solutions that can be delta functions. The unfilled boxes correspond to global solutions only, however. In homogeneous plasmas they are the solutions that have $k_x = 0$. In inhomogeneous plasmas they are the solutions that don't oscillate spatially.

How much the solutions oscillate spatially is given by how many times the solution changes sign along the x direction. In homogeneous plasmas this is given by the wave number in that direction. But here there are no longer any wave numbers in the x direction. Here, it is measured by how many times the solution $f(x)$ of the Eq. (12.23) goes through zero. This is usually different for different values of ω. In the spectrum in Fig. 12.6 the directions in which the degree of oscillation increases have been indicated with arrows.

For most waves ω increases if this degree increases. Take for example the dispersion equation for Alfvén waves in a homogeneous plasma (12.18). If ω increases k has to increase as well. This is logical. If the perturbations oscillate faster spatially the gradients will be bigger and this will cause the forces to be bigger so that the plasma will move faster in reaction to the perturbations and hence the frequency of oscillation will be bigger. Such behavior is called Sturmian behavior. It is valid for almost all waves: sound waves, light waves, *etc.* In the MHD spectrum, however, the opposite behavior is also seen (anti-Sturmian). This shows how different MHD is from normal gas dynamics.

Note: in the limit $B \to 0$ the Alfvén and slow continua collapse into the origin so that only the fast spectrum survives. The fast waves actually transform into normal sound waves so that normal gas dynamics is recovered which has no continua.

12.5 Flow

So far, static plasmas have been considered and they have already demanded a lot of work to gain even a little insight. The plasmas that have to be modeled (the heliosphere) are not static at all. One of the most important aspects of the heliosphere is that it has a stationary flow. As stated at the start of this chapter, energy is transported away in the form of magnetic fields and that can not be done in a static plasma. There, energy can only be transported by waves but that is not what is seen.

12.5.1 Spectral theory with flow

So, the theory needs to be expanded to stationary plasmas with flow. To be able to do a stability analysis an equilibrium situation has to be constructed again. For a stationary state all quantities have to be constant at a point ($\partial/\partial t = 0$) but apart from that the equilibrium quantities are not restricted. This means quite a lot of terms are left in the ideal MHD equations:

$$\nabla \cdot (\rho_0 \vec{v}_0) = 0, \tag{12.24}$$

$$\rho_0 \vec{v}_0 \cdot \nabla \vec{v}_0 + \nabla p_0 = \frac{1}{\mu_0}(\nabla \times \vec{B}_0) \times \vec{B}_0 + \rho_0 \vec{g}, \tag{12.25}$$

$$\vec{v}_0 \cdot \nabla p_0 + \gamma p_0 \nabla \cdot \vec{v} = 0, \tag{12.26}$$

$$\nabla \times (\vec{v} \times \vec{B}) = 0, \tag{12.27}$$

$$\nabla \cdot \vec{B}_0 = 0. \tag{12.28}$$

So now even the equilibrium is described by a full set of non-linear equations. Again, the equations can be linearised but because $\vec{v}_0 \neq 0$ the resulting equations do not collapse to the simple form $\rho_0 \partial^2 \vec{\xi}/\partial t^2 = \mathbf{F}[\vec{\xi}]$ because an extra time derivative is left in the momentum equation (12.25): the second term on the left-hand-side becomes $\vec{v}_0 \cdot \nabla(\partial \vec{\xi}/\partial t)$, whereas it disappeared in the case without flow. The equation for $\vec{\xi}$ becomes:

$$\rho_0 \frac{\partial^2 \vec{\xi}}{\partial t^2} + 2\rho_0 \vec{v}_0 \cdot \nabla \left(\frac{\partial \vec{\xi}}{\partial t} \right) = \mathbf{F}[\vec{\xi}], \tag{12.29}$$

here the operator $\mathbf{F}$ is different from the one in the static case:

$$\mathbf{F} = \mathbf{F}^{\text{static}} + \nabla \cdot \left[\rho_0 (\vec{v}_0 \cdot \nabla \vec{v}_0) \vec{\xi} - \rho_0 \vec{v}_0 (\vec{v}_0 \cdot \nabla \vec{\xi}) \right]. \tag{12.30}$$

Solutions of the form $\exp(-i\omega t)$ can still be used to get an eigenvalue equation:

$$\rho_0 \omega^2 \vec{\xi} + 2i\rho_0 \omega \vec{v}_0 \cdot \nabla \vec{\xi} + \mathbf{F}[\vec{\xi}] = 0, \tag{12.31}$$

but as you can see the problem has now become a quadratic eigenvalue problem which has to be solved for ω instead of a linear eigenvalue equation for ω^2. This also means that the values of ω^2 are no longer restricted to real values and ω will in general have *both* a real and an imaginary component. Thus the solutions still oscillate in time but they could have an exponentially growing or decaying amplitude:

$$\vec{\xi}(\vec{r}, t) \propto e^{\Im[\omega]t} \cos(\Re[\omega]t),$$

these are called over-stable or under-stable solutions for $\Im[\omega] > 0$ or $\Im[\omega] < 0$ respectively.

Take the configuration with inhomogeneities in one direction of the previous section again and add flow in the yz plane to it. What essentially happens is that the frequencies in the spectrum get a Doppler shift. This is derived to get a good feel of what happens.

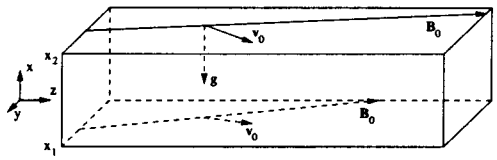


Figure 12.7: The geometry of a gravitating slab with inhomogeneities in one direction: the x-direction. The magnetic field and the velocity lie in the yz plane. Gravitation works in the x-direction. The plasma extends to infinity in the y and the z direction but is bounded between x_1 and x_2 in the x-direction.

The situation is sketched in Fig. 12.7. There are only inhomogeneities in the x-direction so all gradients of equilibrium quantities point along x. Since the plasma is

infinite and homogeneous in the y and z direction plane wave solutions can be used in those directions:

$$\vec{\xi}(x,y,z,t) \propto f(x)\mathrm{e}^{i(k_y y+k_z z-\omega t)}. \tag{12.32}$$

Now look at the eigenvalue problem (12.31). Using solutions of the above form the second term can be transformed: $\vec{v}_0\cdot\nabla\vec{\xi} = i(v_{0y}k_y+v_{0z}k_z)\vec{\xi}$ because $v_{0x}=0$. Define $v_{0y}k_y+v_{0z}k_z \equiv \vec{v}_0\cdot\vec{k}$, although k_x does not exist (this does not matter since $v_{0x}=0$). Now look at expression (12.30) for $\mathbf{F}$. The first term in the divergence term contains $\vec{v}_0\cdot\nabla\vec{v}_0$ and this is zero because $\vec{v}_0$ lies in the yz-plane and $\nabla\vec{v}_0$ along the x-direction so that they are perpendicular. The remaining equation is:

$$\rho_0\omega^2\vec{\xi} - 2\rho_0\omega(\vec{v}_0\cdot\vec{k})\vec{\xi} + \mathbf{F}^{\mathrm{static}}[\vec{\xi}] - \nabla\cdot\left[\rho_0\vec{v}_0(\vec{v}_0\cdot\nabla\vec{\xi})\right] = 0.$$

The last term contains $\vec{v}_0\cdot\nabla\vec{\xi} = i(\vec{v}_0\cdot\vec{k})\vec{\xi}$. Expand the divergence using the product rule:

$$\nabla\cdot\left[\rho_0\vec{v}_0(\vec{v}_0\cdot\vec{k})\vec{\xi}\right] = \\ (\vec{v}_0\cdot\vec{k})(\vec{v}_0\cdot\nabla\rho_0)\vec{\xi} + \rho_0(\vec{v}_0\cdot\vec{k})\vec{v}_0\cdot\nabla\vec{\xi} + \rho_0(\vec{v}_0\cdot\vec{k})(\nabla\cdot\vec{v}_0)\vec{\xi} + \rho_0(\vec{v}_0\cdot\nabla(\vec{v}_0\cdot\vec{k}))\vec{\xi}.$$

The first term is zero because $\nabla\rho_0$ is in the x-direction so that its inner product with $\vec{v}_0$ is zero. The second term can be rewritten as $i\rho_0(\vec{v}_0\cdot\vec{k})^2\vec{\xi}$. The third is zero because $\nabla\cdot\vec{v}_0=0$ since $\vec{v}_0$ is only a function of x, but has no component in the x-direction. The fourth term is zero because $\vec{v}_0\cdot\vec{k}$ is only a function of x so that its gradient points in the x-direction and the inner product with $\vec{v}_0$ is zero. The equation is now reduced to:

$$\rho_0\omega^2\vec{\xi} - 2\rho_0\omega(\vec{v}_0\cdot\vec{k})\vec{\xi} + \rho_0(\vec{v}_0\cdot\vec{k})^2\vec{\xi} + \mathbf{F}^{\mathrm{static}}[\vec{\xi}] = 0,$$

which can be rewritten:

$$-\rho_0\tilde{\omega}^2(x)\vec{\xi} = \mathbf{F}^{\mathrm{static}}[\vec{\xi}], \tag{12.33}$$

in which $\tilde{\omega}(x) \equiv \omega - \vec{v}_0(x)\cdot\vec{k}$ is a function of x. This equation looks almost the same as the eigenvalue equation (12.21) of static MHD. If the plasma is homogeneous (with flow) $\tilde{\omega}$ will be a constant and the solutions will be the same so that the ω for which the problem with flow has solutions will be the ω of the static case with a 'Doppler shift' $\vec{v}_0\cdot\vec{k}$. However, if the plasma is not homogeneous Eq. (12.33) is not an eigenvalue equation any more because $\tilde{\omega}$ depends on x. So, in general, the solutions of Eq. (12.33) will not have the same form as in the static case. Only the continuous parts which corresponded to delta function solutions can be treated this way: they only exist at *one* value of x so that their value of ω will be the static one with the local Doppler shift at that position x. So there are again continua in the spectrum but this time for the following values of ω:

$$\begin{aligned} \omega &\in \{\Omega_{\mathrm{A}}^{\pm} \equiv \Omega_0(x) \pm \omega_{\mathrm{A}}(x) \mid x_1 \le x \le x_2\}, \\ \omega &\in \{\Omega_{\mathrm{S}}^{\pm} \equiv \Omega_0(x) \pm \omega_{\mathrm{S}}(x) \mid x_1 \le x \le x_2\}, \end{aligned} \tag{12.34}$$

in which $\Omega_0(x) \equiv (\vec{v}_0(x)\cdot\vec{k})$. So, the delta function perturbations of the continuum solutions simply move along with the flow. The rest of the spectrum, however, will change in a more complicated matter.

Only the stable part of the spectrum (ω is real) will be treated. First the static spectrum: if the square root of Fig. 12.6 is taken the top part of Fig. 12.8 is obtained. The spectrum in the case with flow is drawn schematically in the bottom part. The new

continua are indicated with $\{\Omega_A^{\pm}\}$ and $\{\Omega_S^{\pm}\}$. The equivalents of $\{\omega_{s0}\}$ and $\{\omega_{f0}\}$ have been indicated with $\{\Omega_{s0}\}$ and $\{\Omega_{f0}\}$ respectively. The spectrum always satisfies:

$$-\infty \leq \Omega_{f0}^- \leq \Omega_A^- \leq \Omega_{s0}^- \leq \Omega_S^- \leq \Omega_0 \leq \Omega_S^+ \leq \Omega_{s0}^+ \leq \Omega_A^+ \leq \Omega_{f0}^+ \leq \infty. \tag{12.35}$$

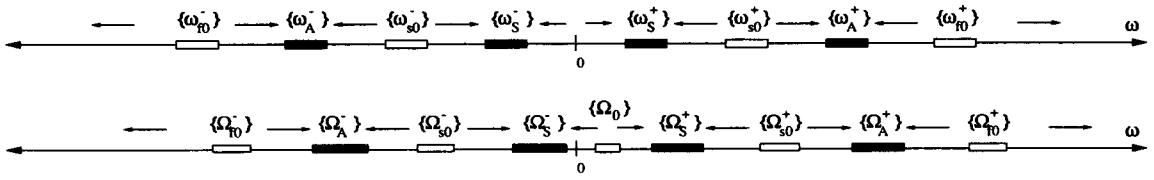


Figure 12.8: The spectrum of real ω's for which the MHD equations have solutions for given k_y and k_z in the case of a plasma with inhomogeneities in one direction only. The top spectrum is for the static case of the previous section (it is the square root of Fig. 12.6). The bottom spectrum is for the stationary case with flow in the xy plane.

12.5.2 Spatial characteristics

The last subject that has to be treated regarding flows is what is known as *wave-like equilibrium*. This is something which influences the equilibrium solutions when the flow becomes 'supersonic' (this word is not really adequate since there are no sound waves, only MHD waves).

The perturbations that have been studied in the previous sections propagate through the plasma. Think of the plane waves in Sect. 12.3.2. There the perturbations moved through the plasma as waves. In the other (inhomogeneous) cases most of the solutions no longer look like plane waves any more but a perturbation in the plasma still propagates through the plasma. Usually they propagate as surfaces that move with time. For example a perturbation of a normal gas spreads as a sphere that gets bigger and bigger. These 'paths' along which the perturbation propagates are called characteristics.

The characteristics in MHD can become very complicated. In a static homogeneous plasma the characteristics are given by the Friedrichs group diagram (right side of Fig. 12.3). This diagram shows a cross-section of the characteristics of a perturbation that has occured some moments ago in the origin. The same thing for a normal gas would be a circle. As you can see this is already a complex thing.

If the plasma flows it gets more complicated of course. Now if a perturbation occurs in the origin the generated waves get dragged along with the flow so that the result is almost the same but with the Friedrichs diagram shifted in the direction of the flow.

It gets more complicated if the perturbation persists. Then characteristics will be formed continually. Let us look at the situation in a normal gas first: the situation is depicted in the left side of Fig. 12.9. Two of the characteristics are drawn that were formed at times t and $2t$. In reality a whole continuum of waves has been sent out. Through interference they will annihilate each other except on the two lines that are drawn enveloping them. This means that a whole cone will form in the gas where information about the perturbation is present, independent of time. Such a cone is called a spatial characteristic.

In a plasma something similar happens. There, too, spatial characteristics are formed. This is shown in the right side of Fig. 12.9. Depending on the speed of the flow and the direction with respect to the magnetic field even several spatial characteristics can occur.

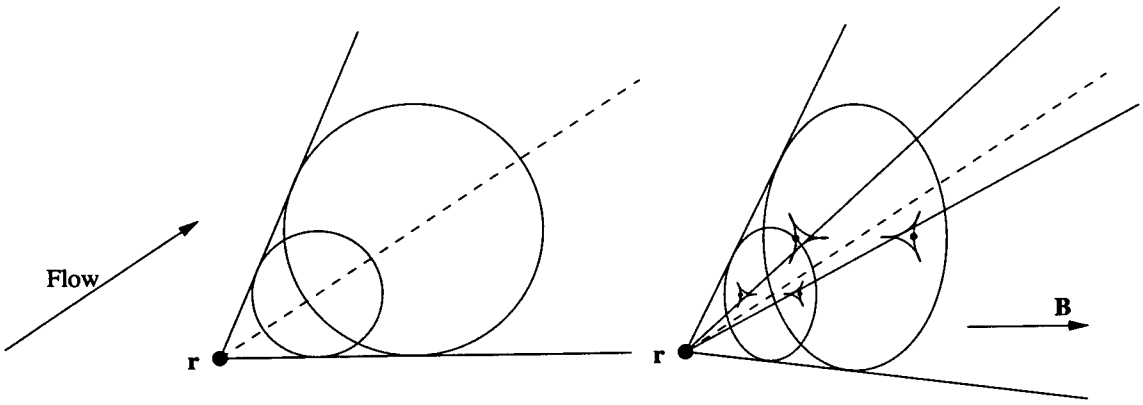


Figure 12.9: 'Supersonic' flow in normal gases (left) and in plasmas (right). The temporal characteristics in a gas are circles (spheres), in plasmas they are given by the Friedrichs group diagram. In the normal gas two spatial characteristics develop, in the plasma four: two 'fast' and two 'slow' ones.

The point of all this is that such a 'persisting perturbation' is no longer a perturbation because it can be constant in time. In a flowing plasma such characteristics can occur because of the shape of the configuration or because there is something in the way (a planet, for example). Then the characteristics are a part of the equilibrium situation instead of only being caused by perturbations on this equilibrium. So in supersonically flowing plasmas (and the solar wind is such a plasma) waves and equilibrium are no longer separate issues.

Chapter 13

Wind theory

Lecturer J.P. Goedbloed, notes by S. Van Loo

13.1 Introduction

In the preceding chapter we have studied heliospheric (or astrophysical) problems from a magnetohydrodynamics point of view. It treated MHD modeling by conservation laws and also generic magnetic configurations. Then we have deducted the flux conservation law and the Alfvén waves, both essential in MHD. Analogous to quantummechanics we could create a spectral theory for static equilibria and in practice this means MHD spectroscopy, like helioseismology, is possible. At the end we have seen a spectral theory for equilibria with background flow, along with the determination and the properties of stationary equilibrium states with flow. Important is that the equilibrium becomes wave-like.

In this section we are going to take a look at the heliosphere. The solar wind confines the planetary magnetic fields into magnetospheres and by impinging on the magnetic fields of the planets it creates a bow shock on the day side and drags the magnetotail on the night side of the magnetosphere. But the solar wind also produces a cavity surrounding the Sun in the local interstellar medium. The sun has its own magnetosphere called the **heliosphere**; it is the region within the galactic medium where the plasma digs in, and fills a cavity. A hot corona cannot be in static equilibrium with the interstellar medium and must expand. Once the solar wind is introduced there is a surface surrounding the sun at which the pressure of the solar wind balances the pressure of the interstellar gas which is the heliospheric boundary. It is in fact the solar wind termination by a shock transition from super-to sub-sonic flow, that creates the heliospheric boundary. The position of the boundary is not known with great accuracy but is probably of order of 100 AU. The heliosphere contains most of the solar system but not the most distant comets. The global solar magnetic field and the local interstellar medium organizes the heliosphere.

13.2 Shocks

The question we have to ask ourselves is how we have to connect flow from one state to another through the shocks? Which instruments are needed to take on this problem? We will replace time and space derivatives in the MHD equations by

$$\frac{\partial f}{\partial t} \to -u[f], \qquad \nabla f \to \mathbf{n}[f]$$

where $[f] = f_1 - f_2$.

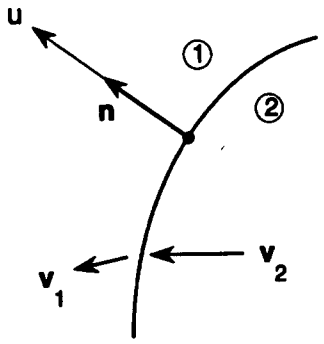


Figure 13.1:

This will lead to the jump conditions:

$$
\begin{aligned}
&-u[\rho] + \mathbf{n} \cdot [\rho \mathbf{v}] = 0, \\
&-u[\rho \mathbf{v}] + \mathbf{n} \cdot [\rho \mathbf{v}\mathbf{v} + (p + \frac{1}{2}B^2)\,\mathbf{I} - \mathbf{B}\mathbf{B}] = 0, \\
&-u[\frac{1}{2}\rho v^2 + \rho e + \frac{1}{2}B^2] + \mathbf{n} \cdot [(\frac{1}{2}\rho v^2 + \rho e + \frac{1}{2}B^2) \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{B}\mathbf{B}] = 0, \\
&-u[\mathbf{B}] + \mathbf{n} \cdot [\mathbf{v}\mathbf{B} - \mathbf{B}\mathbf{v}] = 0, \qquad \mathbf{n} \cdot [\mathbf{B}] = 0,
\end{aligned}
$$

whereas the entropy should increase across the shock:

$$[\rho^{-\gamma} p] \leq 0.$$

This is an extra condition we have to add. The jump conditions connect the variables across the shock front. Jumps can only be permitted when the partial differential equations (PDE) are hyperbolic. The peculiarity of the MHD (and gas dynamics) equations is that they change from elliptic to hyperbolic in the middle of the domain when a critical speed is crossed. In gas dynamics this is the sound speed so that the flow is called *transsonic*. We use the same word in MHD, although different critical speed occur. Consequently there is a variety of elliptic and hyperbolic flow regimes in MHD. We will illustrate this by means of examples taken from Goedbloed and Lifschitz (1997) and Keppens *et al.* (1999).

13.3 Stationary MHD flows

13.3.1 The equations

In this paragraph we are going to work with self-similar translation-symmetric MHD flow. But before we do that we are taking a closer look at the equations of ideal MHD:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{13.1}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \nabla(\frac{1}{2}v^2 + \frac{\gamma}{\gamma - 1}\rho^{\gamma-1}S) - \frac{1}{\gamma - 1}\rho^{\gamma-1}\nabla S - \mathbf{v} \times \omega + \frac{1}{\rho}\mathbf{B} \times \mathbf{j} = 0, \tag{13.2}$$

$$\frac{\partial S}{\partial t} + \mathbf{v} \cdot \nabla S = 0, \tag{13.3}$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) = 0, \qquad \nabla \cdot \mathbf{B} = 0. \tag{13.4}$$

where we have defined the entropy as $S \equiv p\rho^{-\gamma}$, the vorticity as $\omega \equiv \nabla \times \mathbf{v}$ and the *current density* as $\mathbf{j} \equiv \nabla \times \mathbf{B}$.

We can now treat the stationary, translation-symmetric MHD flows. This means we have to take the partial time derivatives and the partial derivative in the z-direction equal to zero. ($\partial_t = 0$ and $\partial_z = 0$) We can also define a flux function ψ and a stream function as a function of ψ, $\chi(\psi)$:

$$\mathbf{B}_p = \mathbf{e}_z \times \nabla\psi,$$
$$\mathbf{v}_p = \mathbf{e}_z \times \nabla\chi.$$

where $\mathbf{B}_p$ is the poloidal magnetic field and $\mathbf{v}_p$ the poloidal flow. This makes that five arbitrary flux functions χ', H, S, I, Ω collapse into three flux functions, namely $\Pi_{1,2,3}(\psi)$. Before we now rewrite the ideal MHD equations, we define the poloidal Alfvén Mach number as $M^2 \equiv \rho v_p^2 / B_p^2 = \chi'^2/\rho$.

So with this background we can rewrite the ideal MHD equations and we obtain the core problem:

– a *nonlinear PDE* for the flux $\psi(x, y)$:

$$\nabla \cdot [(1 - M^2)\nabla\psi] + \frac{\Pi_1'}{M^2} - \frac{\Pi_2'}{\gamma M^{2\gamma}} + \frac{\Pi_3'}{1 - M^2} = 0, \tag{13.5}$$

– the *Bernoulli equation* for the determination of $M^2(x, y)$:

$$\frac{1}{2}|\nabla\psi|^2 - \frac{\Pi_1}{M^4} + \frac{\Pi_2}{M^{2(\gamma+1)}} + \frac{\Pi_3}{(1 - M^2)^2} = 0. \tag{13.6}$$

The two equations (13.5) and (13.6) exhibit a striking parallelism when we consider them as functions of two variables, ψ and M^2, with derivatives of ψ in the first equation, and of M^2 in the second. We can also derive that the transition from ellipticity to hyperbolicity of the PDE (13.5) for ψ is governed by the coefficients in front of the second order derivatives, where the implicit dependence of M^2 on $|\nabla\psi|^2$ in the Bernoulli equation is taken into account. Another remark that we can make, is that, although it is an algebraic equation, the Bernoulli equation is more difficult to solve than the PDE.

As a restriction on the choice of the arbitrary profiles, we will consider the general profiles $\Pi_{1,2,3}$ to have the same shape. If we now consider scale-invariant or self-similar flows, further reduction is possible. Assume now one master profile with two constants, A (entropy) and B (longitudinal field), so that the three remaining flux functions become:

$$\Pi_1 = \pi(\psi), \qquad \Pi_2 = A\pi(\psi), \qquad \Pi_3 = B\pi(\psi).$$

with the power profile $\pi = \psi^{2-2/\lambda}$. A simplification that than creates self-similar solutions is:

$$M^{-2} = X(\theta),$$
$$\psi = r^\lambda Y(\theta).$$

where we can associate M^{-2} with the inverse speed and θ is the coordinate along the magnetic/flow surface perpendicular to the direction of the symmetry (z).

The two equations (13.5) and (13.6) can be written as

$$\frac{dX}{d\theta} = \frac{H}{(1-1/X)J}Z,$$
$$\frac{dY}{d\theta} = \frac{1}{1-1/X}Z, \qquad Z = \pm\frac{\sqrt{-2F_0}}{1-1/X}.$$

where $H = H(X;Y)$ and $J = J(X,Y)$ are functions which we use to simplify the equations, and $F_0 = F(Y' = 0)$, with F the Bernoulli equation.

What are now the qualitative features of the $X - Y$ phase space?

- We see that we have in these equations some singularities: if $J = 0$, then θ stops to progress with X. And at $X = 1$ we have a singularity where the Alfvén speed is reached (This is a forbidden area.).
- The trajectories that must be followed in order to have 'transsonic' flow, can be calculated from $dY/dX = J/H$.
- For solutions to exist, F_0 has to be smaller or equal to zero, so that we have a boundary, the so called *Bernoulli boundary*, defined by $F_0 = 0$. We can calculate the trajectories everywhere in the phase space, but it only makes sense inside the Bernoulli Boundary.
- Finally we have also an indication for the hyperbolicity of the ordinary differential equation (ODE) by the inequality $\Delta(X) > 0$, where Δ is a determinant coming from the PDE (13.5).

13.3.2 Some examples

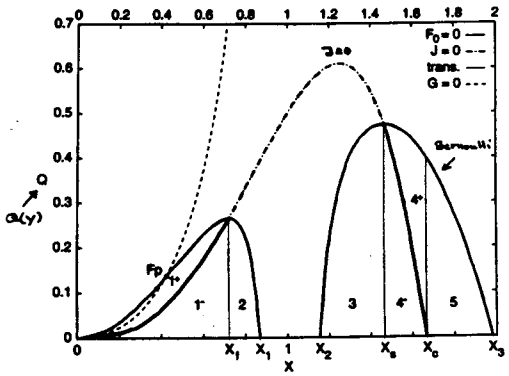


Figure 13.2:

In this figure (13.2) we see a phase diagram of the flow regimes and the limiting lines. Also around $X = 1$ we see the Alfvén gap. On the picture some numbers are printed, indicating the regimes, wherefore we can determine whether they are hyperbolic or elliptic by Δ: 1^+: hyperbolic super fast; 1^-: hyperbolic fast; 2: elliptical fast; 3: elliptical slow; 4^-: hyperbolic slow; 4^+: hyperbolic sub slow; 5: elliptical sub slow.

We notice that when a critical speed is crossed the equations flip from elliptical to hyperbolic, and vice versa.

If we now put the trajectories on the phase diagram (see figure 13.3), it appears that they cross the gap and the singular line $J = 0$ without problems (!?). According to the

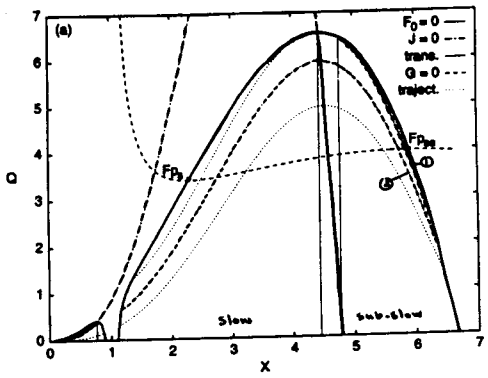


Figure 13.3:

formula for the trajectories we can calculate these trajectories everywhere in the phase space, but the method of finding the solution expires at these singularities, and that is something that we haven't incorporated.

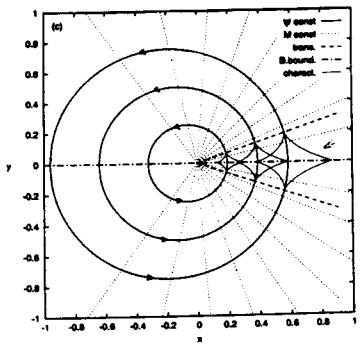


Figure 13.4:

Figure (13.4) shows the transition from hyperbolic flow to elliptical flow, in particular from the 4^+-hyperbolic sub slow region to the 5-elliptical sub slow region. The sub slow flow starts in the hyperbolic region, with two sets of characteristics bisected by the streamlines, and converts to the elliptical regime at the dashed line. At $\theta = \pi$ the flow will encounter the Bernoulli boundary, where it smoothly joins to the other branch of the Bernoulli function on the bottom half plane. This is what we call a periodic solution.

It is now interesting to take a look at transitions across the limiting line. Because, as already is said that jumps can only be permitted when the PDE's are hyperbolic, we must take the transition between two hyperbolic regions, e.g. 4^- and 4^+ from the slow to the sub slow.

A first look at the figure (13.5) tells us that the slow flow pattern is 'reflected' by the limiting line. If we would enlarge the area where the approach of the limiting line is, like in figure (13.6), we see the same reflection of flow lines and characteristics even better. In reality the trajectory really corresponds to two different flow patterns and its smooth crossing of the singular line is misleading, since the two flow patterns have nothing to do with each other. We have just obtained a second solution occupying the same space. The continuous connection of characteristics is just a convenient way of presenting the two flows in one picture.Thus, we have found that the crossing of the limiting line by

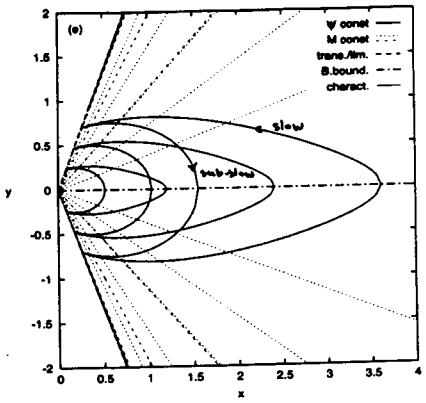


Figure 13.5:

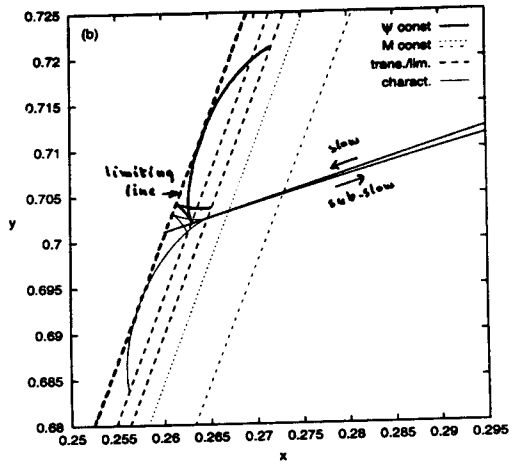


Figure 13.6:

continuous flows is impossible.

13.3.3 Stationary shocks

Because continuous solutions don't exist, we are trying to find discontinuous solutions, so we come back to the shock conditions. We will take Y and B continuous, so $[Y] = 0$ and $[B] = 0$, and X and A discontinuous. We get the shock conditions:

$$[(1 - 1/X)Y'] = 0,$$
$$[1/X]\lambda^2 Y^{2/\lambda} + [X^2 + (1 - 1/\gamma - X)AX^\gamma] = 0.$$

and the entropy condition $[A] \leq 0$.

At the shock position we have 5 variables, namely:

$$X_1 \neq X_2, \qquad Y \equiv Y_1 = Y_2, \qquad A_1 \neq A_2.$$

If we eliminate A_2, we have distilled a jump condition: $f_1(Y, X_1, X_2, A_1) = 0$, and if we substitute A_2, we have distilled an entropy condition: $f_2(Y, X_1, X_2, A_1) \geq 0$. We now

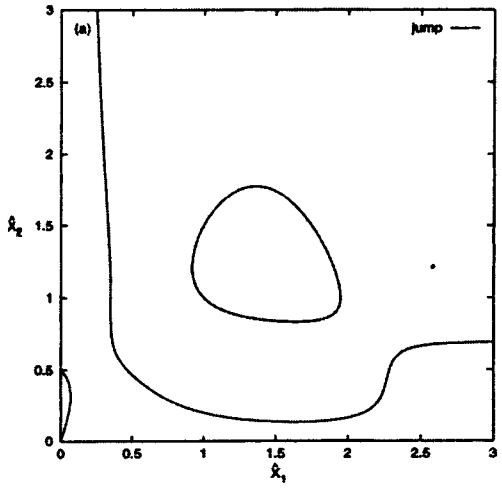
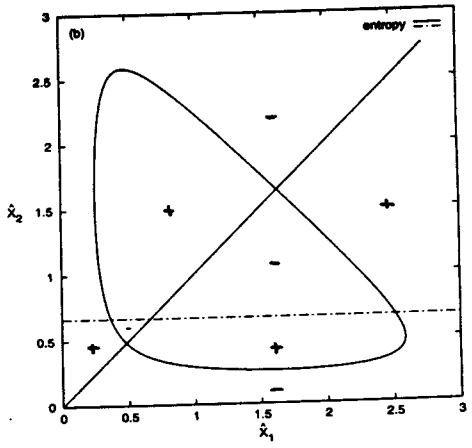


Figure 13.7:



Allowed entropy regions (+)

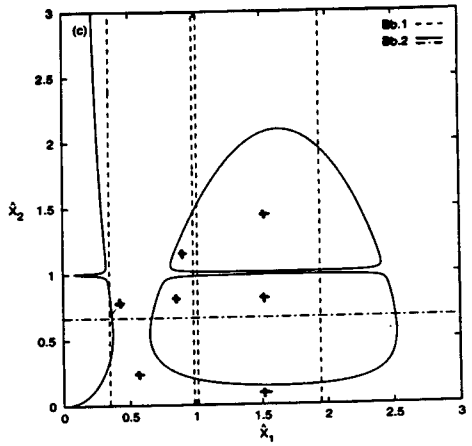
Figure 13.8:

plot f_1 and f_2 in the $X_1 - X_2$ plane for a given (Y, A_1). We will solve these equations graphically.

This first picture (13.7) shows the distilled jump condition. But with this picture alone, we are nowhere, because there are other conditions for the shock waves, like the entropy condition. In figure (13.8) the allowed regions are indicated by a +.

Another condition is the Bernoulli equation, and we must also cut out the unphysical pieces. This gives us figure (13.9). To get the solution of the shock waves, we have to put the three previous figures together. We see that there remain three shock waves, namely the fast, Alfvén and slow shock.

Figure (13.11) is a blow up of (13.10) and on this figure is already indicated between which regions the shocks appear. Because this not very handy to read, we have put the information of figure (13.11) on a schematic figure: (13.12). It is obvious to see that there are four flow regions and three types of shocks.



Allowed Bernoulli regions (+)

Figure 13.9:

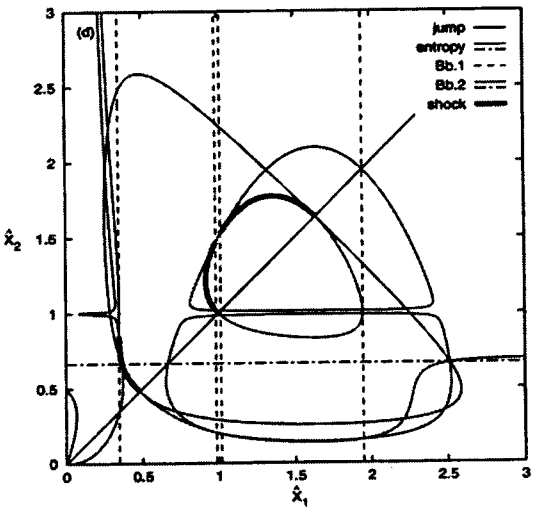


Figure 13.10:

13.3.4 Conclusions

- There are *four flow regions*, which are separated by the limit lines and the Alfvén gap and cannot be connected by continuous flows.
- The limiting line signals the expiration of continuous solutions and guarantees the existence of discontinuous flows jumping across.
- Fast shocks jump across the fast limiting line Lf, intermediate shocks jump across the Alfvén gap A and slow shocks jump across the slow limiting line Ls.
- The three obstacles just create the right conditions to produce *three kinds of strongly discontinuous flow* which may be considered as the non linear counterpart of the weak discontinuities of linear MHD.

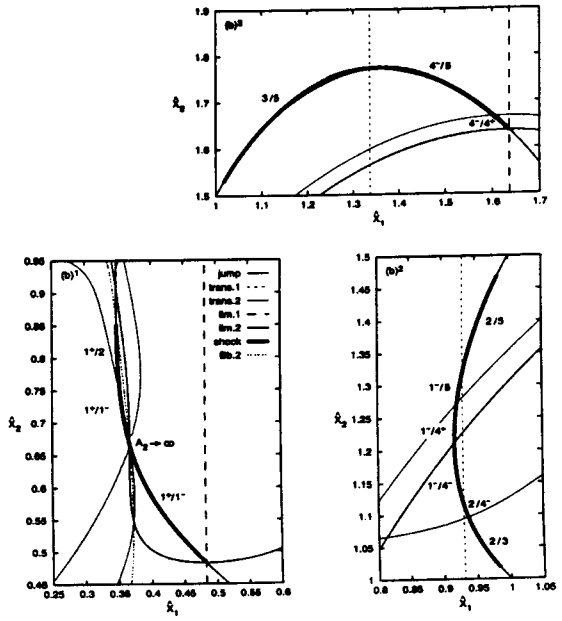


Figure 13.11:

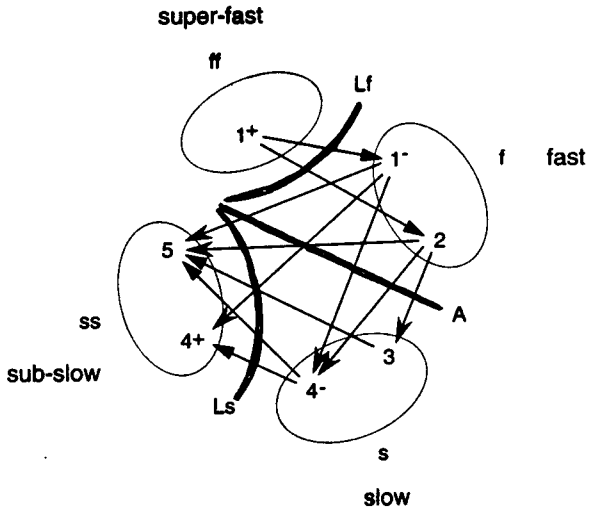


Figure 13.12:

13.4 Simulations of stellar winds and accretion flows

In this paragraph we are going to construct stationary flow patterns, which will be guided by the analytical solutions, in spherical and toroidal geometry by means of the Versatile Advection Code (Toth, Keppens). The specifications of this code are versatile with respect to:

– the applications: Shallow water, HD, MHD

- geometry, it is dimension independent: 1D, 2D, 2.5D, 3D
- computer platforms: PC's to IBM SP
- spatial discretisations: 6 shocks-capturing schemes
- temporal discretisations: explicit, semi-implicit, implicit

Also there is shock-capturing in the second order and finite volume on structured grids. On the basis of 1D wind models, that are used as a test bed for numerical modeling, 2D transsonic outflows are realized. Also accretion flows in the equatorial plane are modulated. The realizations of the VAC project are illustrated in the next paragraphs, beginning from 1D to 2D and accretion.

13.4.1 The Parker model of the solar wind

The model of Parker from 1958, was a model for a hot expanding corona. This model was hydrodynamic, isothermal, spherically symmetric and stationary. We take the r-component of the momentum:

$$\rho v_r \frac{\partial v_r}{\partial r} = -\frac{\partial p}{\partial r} - \rho \frac{GM}{r^2}.$$

where we take a base temperature T_0 and density ρ_0. We can manipulate this equation by using the isothermal assumption or $p = a^2\rho$, with a the speed of sound, and mass conservation: $\rho r^2 v_r = Cst$. This equation becomes:

$$\frac{\partial v_r}{\partial r} = \frac{v_r}{r}\left(\frac{2a^2 - GM/r}{v_r^2 - a^2}\right).$$

We see that this last equation has a singular point at $v_r^2 = a^2$. This is nothing else than the sonic point. Integration and re-ordering leads then to:

$$\frac{v_r^2}{2} + a^2 ln\left(\frac{\rho}{\rho_0}\right) - \frac{GM}{r} = E.$$

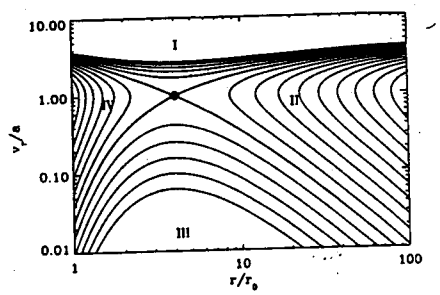


Figure 13.13:

Figure (13.13) shows the topology of the isothermal solution of Parker. Each curve is a locus of constant energy, and represents an acceptable solution from the standpoint of global energy conservation. You have only two solutions that are physically important,visually the solutions that start at the surface of the sun, here indicated by $r/r_s = 1$, and that go away from the suns surface passing through the sonic point.

The Parker model is the starting point for more complicated models. If we stay with the 1D hydrodynamic solutions, we could change for example the isothermal condition into a polytropic.

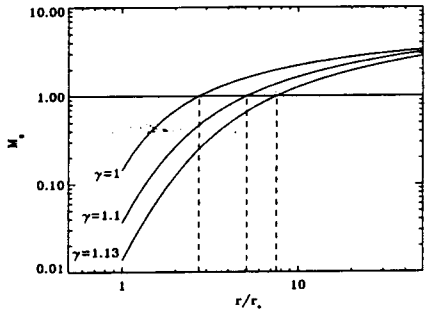


Figure 13.14:

We see on figure (13.14) that for different γ we obtain different critical points. There is even an equation that gives the connection between the critical point and γ :

$$\frac{r_s}{r_*} = \left(\frac{v_{esc}}{2c_{s*}}\right)^{\frac{2(\gamma+1)}{5-3\gamma}} \left(\frac{c_{s*}}{v_{r*}}\right)^{\frac{2(\gamma-1)}{5-3\gamma}}.$$

If we change from non-rotating to a rotating system at the equator, we get figures like (13.15).

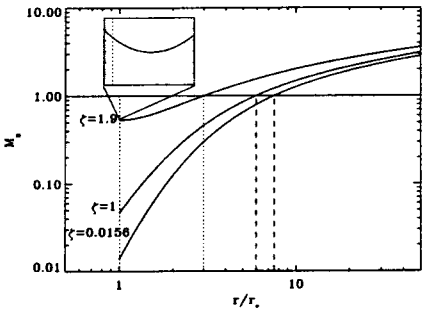


Figure 13.15:

Now we find that it is possible to have two critical points. We can understand this by looking at the equation that connects the critical point with γ and the rotation speed η:

$$\eta^2 \left(\frac{r_*}{r_s}\right)^2 - \frac{v_{esc}^2 r_*}{2c_{s*}^2 r_s} + 2\left[\left(\frac{v_{r*}}{c_{s*}}\right)\left(\frac{r_*}{r_s}\right)^2\right]^{\frac{2(\gamma-1)}{\gamma+1}} = 0.$$

which has two solutions, depending on the coefficients of the polynomial. The next thing is to introduce a magnetic field and for this model we have the Weber-Davis solution.

As we see on figure (13.16), there are two critical points and one Alfvén point. Another conclusion is that $\mathbf{v} \parallel \mathbf{B}$ in a corotating frame.

13.4.2 2D wind solutions

In the next figures we are going to show a 2D polytropic axisymmetric HD wind flow.

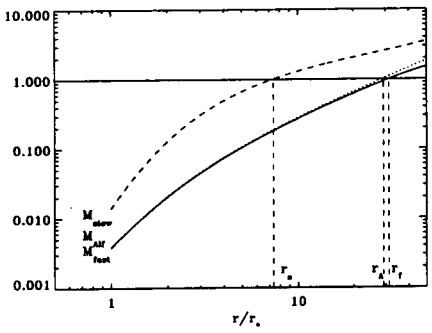


Figure 13.16:

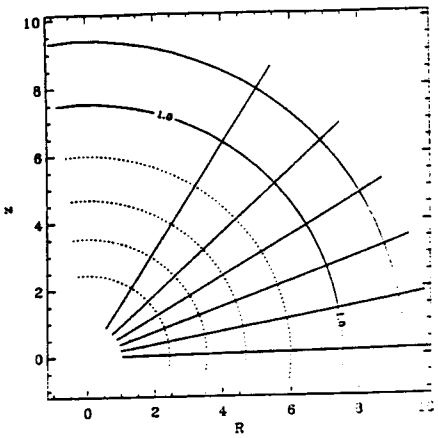


Figure 13.17:

In the first figure (13.17) we have a solution for a slow rotation rate and we see that the flow practically streams radially. Figure (13.18) shows a solution for a high rotation rate, and as we can see, the flow is dragged by the rotating object. There are other possibilities, for example: axisymmetric magnetized wind containing a 'wind' and 'dead' zone.

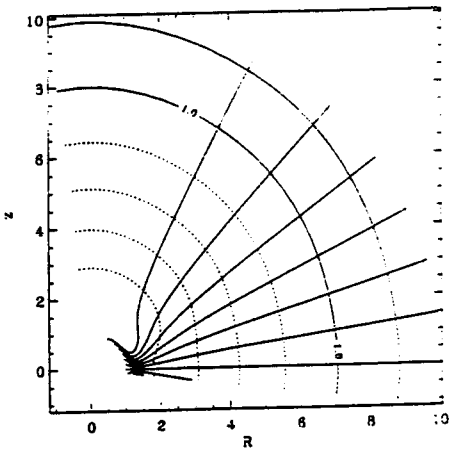


Figure 13.18:

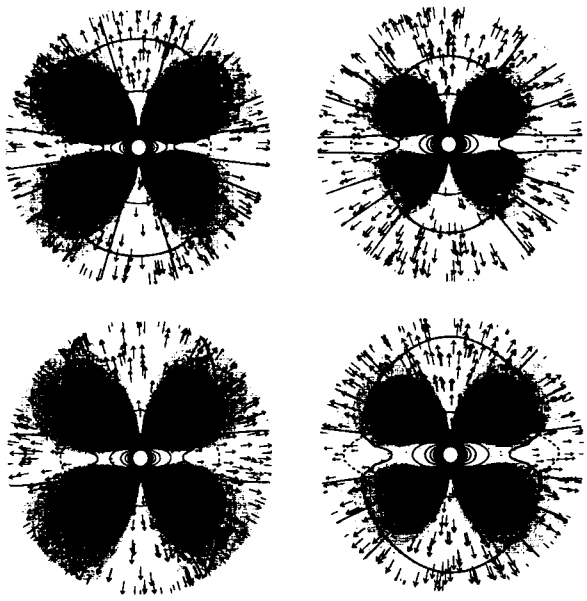


Figure 13.19: 2.5D wind with variable field strength and dead zone extent

In the figures (13.19) we see that there is no flow out of material at the dead zone, but after a certain distance, we can detect a flow. (In the figure the flow is indicated with arrows.)

13.4.3 Accretion flow

First we treat the 1D HD problem and as a model we use accretion onto a black hole. This problem is solved fully implicitly, using a first order accurate total variation diminishing (TVD1) discretisation, because of two reasons. First, the convergence behavior is superior to an explicit scheme, and second we can speedup by a factor of 40.

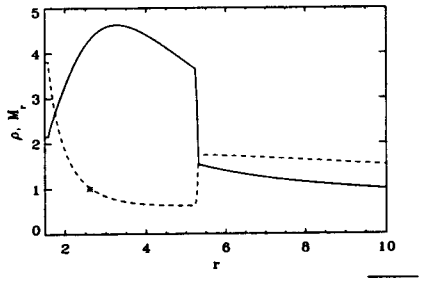


Figure 13.20: accretion onto a black hole

Figure (13.20) shows the result.

With the use of the 1D solution we can create a 2D solution in the equatorial plane. This research by Molteni, Tóth and Kuznetsov is at this moment in ApJ press. In figure (13.20) a calculation is shown of stability against non-axisymmetric perturbations in 2D accretion. We see that the system is not stable.

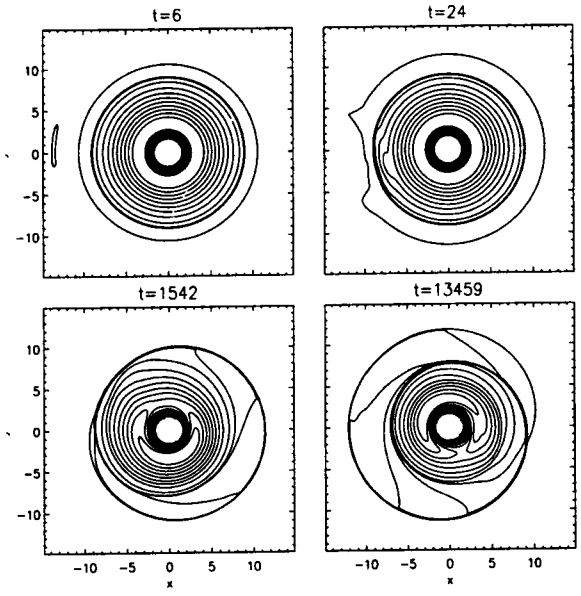


Figure 13.21:

13.4.4 Conclusions

The realizations of this project are immense, but the work isn't finished: the combination of several insights have to open the road to 3D models and also a realistic energy equation has to be incorporated in the future.

Chapter 14

The Sun and our climate

Minnaert Lecture by S.K. Solanki, notes by D.B. Roy

> *(...) the overwhelming balance of evidence and scientific opinion is that it is no longer a theory, but now a fact that global warming is for real.*
>
> (Clinton 1997)

> *Satellite data has shown no global warming whatsoever in the past twenty years. Most climatologists believe that global warming is a hoax.*
>
> (Reagan 1997)

14.1 Introduction

The question of the Sun's effect on the Earth's climate is not only of scientific interest, but also carries significant social relevance. Although it is generally agreed these days that one of the greatest influences on our climate comes from mankind, it is still very much unclear to what extent our actions have been responsible for the rising temperatures witnessed over the last century.

At the moment, this issue is most prevalent in the discussion surrounding global warming, where the observed temperature increase may be caused by the greenhouse effect, or equally well fully or in part by solar influence. A better understanding of the Sun-Earth connection could therefore help determine how much damage, if any, has been done. Furthermore, it would provide information on the degree of natural variability due to solar causes present in the Earth's climate, something which must find its way into models predicting global change for the next century.

In this chapter, an overview will be presented of the timescales on which a solar influence is present. For each timescale, an overview of the corresponding solar effects will be presented, followed by a discussion of some experimental data.

14.2 Long timescales: solar evolution

On a time scale of 10^8 to 10^9 years, solar evolution as driven by chemical changes in the Sun's core starts having a noticeable influence on the Sun's brightness, which is the primary parameter when considering influences on the Earth. We will first examine the evolution of the Sun's brightness, followed by a comparison of this evolution with what we know about the evolution of the Earth's temperature.

14.2.1 The evolution of the solar brightness

Like any other star, the Sun began its life as a collapsing gas cloud. The release of gravitational energy due to the collapse caused a temperature increase which ignited nuclear proton-proton fusion $4.5 \cdot 10^9$ years ago. Within a relatively short period the gas disk around the new-born star coalesced into several planets and approximately 4 billion years ago, the stage was set for the evolution of life as we know it.

As this life evolved on Earth, the Sun was also gradually changing. The burning of solar hydrogen via the proton-proton chain is essentially a process transforming four H nuclei into one He^4 nucleus. This means that during the Sun's life, its average mass per particle has increased.

From models of stellar composition and evolution, it follows that the brightness L of a light star like the Sun varies proportionally to its average mass per particle μ as

$$L \propto \mu^{7.5} \tag{14.1}$$

(see for example Lamers (1997), or Bowers and Deeming (1984)). From this proportionality, it follows that as the Sun aged, it must have become brighter.

If we make the estimate that the Sun has currently fused 40% of its hydrogen store, *i.e.* μ has risen from 0.5 to 0.62, we find that the Sun has increased in brightness by approximately 30% in the 4.5 billion years since its birth.

14.2.2 The early evolution of the Earth's temperature: the "faint Sun" paradox

If we want to determine the evolution of the Earth's temperature over the last several billion years, it is obviously impossible to rely on direct measurements. We are thus forced to examine indirect pointers, such as fossil records. Limited as our information on this period is, even a cursory evaluation produces a serious contradiction.

Changes in the Sun's brightness will eventually cause a change in the Earth's average equilibrium temperature. Current climatological models estimate that a variation in solar brightness of 0.1% could produce a change in the global equilibrium temperature of 0.2 degrees (Lang (1995)). Combining this estimate with the knowledge that when the Sun was born, its brightness was 30% lower, leads to the conclusion that 4 billion years ago, the Earth was a barren wasteland, covered with ice.

However, sedimentary rocks, which must have been deposited in liquid water, have been found dating back $3.8 \cdot 10^9$ years ago. Furthermore, there is fossil evidence of life going back as far as $3.5 \cdot 10^9$ years. All this evidence points towards a warm, liquid environment. How can this be reconciled with the previously calculated below-zero temperature? This problem is called the "faint Sun" paradox.

The key to this enigma lies in considering the changes in the Earth's atmosphere. After Earth's creation, it's atmosphere was formed by the gaseous releases of volcanic eruptions and enriched by cosmic debris. It is currently thought that this atmosphere contained much more CO_2 and water vapor than it does now, causing a large greenhouse effect and keeping the temperature above freezing (see Fig. 14.1).

As the Sun evolved and got brighter, two processes occured that countered the heating effect from the Sun. The net effect of both processes was to reduce the amount of CO_2 in the atmosphere, thus reducing the greenhouse effect.

Firstly, the increase in temperature caused more water to evaporate and increased the average rainfall. This led to more carbon dioxide dissolving in rain as carbonic acid, which

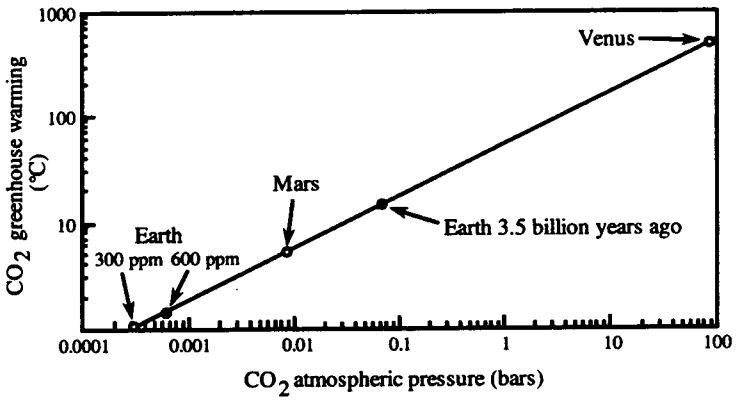


Figure 14.1: The relationship between the partial pressure of carbon dioxide and rise in surface temperature for the inner planets. The current CO_2 pressure of 300 ppm and the pressure 3.5 billion years ago of the Earth are both shown. Also, the effect of a doubling of the Earth's CO_2 pressure is plotted. From Bryant (1997).

reacted with calcium present in rocks. Sea organisms used the calcium-rich water to form skeletal calcium carbonate ($CaCO_3$), which after their death was trapped as sediment on the ocean floor.

A second counter-effect came from the evolution of strains of bacteria and early plants which absorbed CO_2 directly from the air, releasing oxygen. As these lifeforms multiplied, they slowly began transforming the carbon dioxide in the atmosphere into oxygen.

Unfortunately, the continuing increase of solar brightness will eventually cause the oceans to evaporate, leading to a large increase in the level of water vapor in the atmosphere. This will cause a greenhouse effect severe enough to make the Earth uninhabitable.

Hence, although the greenhouse effect made possible the development of life, it will also make that life untenable within approximately one billion years, if Man does not speed up the process.

14.3 Intermediate timescales: orbital effects

Geometrically, our solar system is all but ideal: instead of perfect spheres describing unvarying circular orbits, we have wobbling unspherical planets moving in ever-changing ellipsoidal orbits. The combined effect of these geometrical imperfections causes climatological variations on a time scale of 10^5—10^6 years.

We will first examine the deviations from ideal geometry that exist in the Earth-Sun system, neglecting the effect of the moon. Following that, we will attempt to reconcile these deviations with observed phenomena in the Earth's temperature record.

14.3.1 Eccentricity, obliquity and precession

The Earth-Sun system has three fundamental orbital attributes: the eccentricity of the ellipsoidal orbit around the Sun, the tilt of the Earth's spin axis, termed the angle of obliquity, and the times of closest approach to the Sun, called the equinoxes (see Fig. 14.2).

Each of these three attributes varies over periods of thousands of years, as shown in Figs 14.3 and 14.4.

The eccentricity of the Earth's orbit varies between 0.0 and 0.07, and is currently

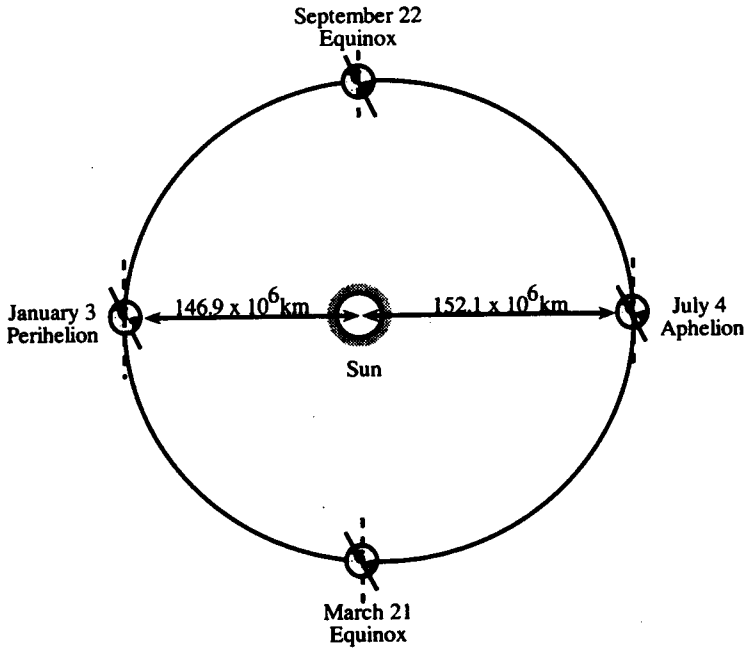


Figure 14.2: An overview of orbital attributes of the Earth about the Sun. From Bryant (1997).

equal to 0.017. The cycle of eccentricity varies between 90,000 and 100,000 years. The angle of obliquity presently has a value of 23.5° and varies between 21.4° and 24.4° with a period of approximately 40,000 years.

Finally, each time the Earth revolves about the Sun, it does not come back to its original location, but tends to move forward slightly in its orbit. This causes a precession of the equinoxes. The current rate of precession is 50.2554 arc-seconds per year, on longer timescales the length of the precession cycle lies between 19,000 and 23,000 years.

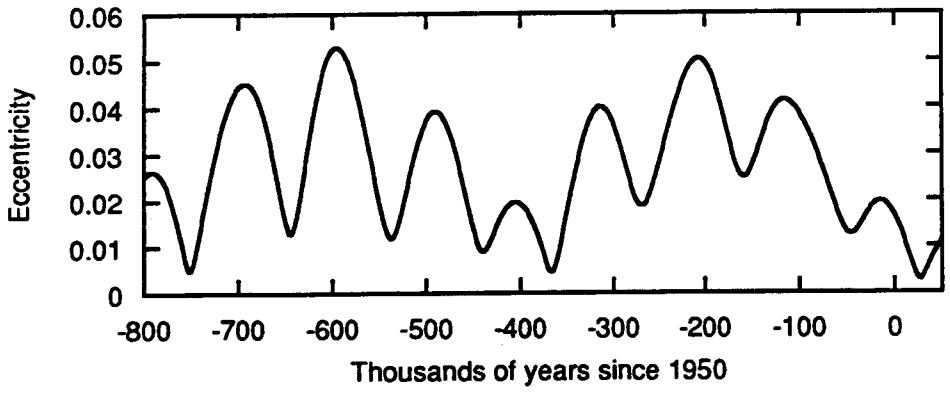


Figure 14.3: A model for the eccentricity of the Earth's orbit as a function of time from 800,000 years before present to 50,000 years in the future. From Hartmann (1994).

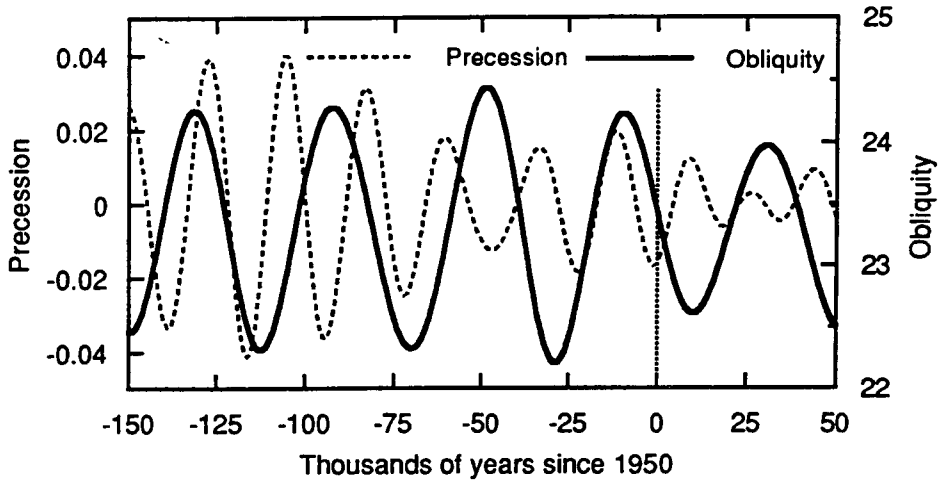


Figure 14.4: Models for the obliquity of the Earth and precession of the Earth-Sun equinoxes as a function of time from 800,000 years before present to 50,000 years in the future. From Hartmann (1994).

14.3.2 The Earth's temperature record: glacial cycles

As was the case on timescales of billions of years, the evolution of the Earth's average temperature over the last million years can only be determined by using indirect data. However, on timescales of the order 10^6 years these indirect measurements are accurate enough to construct a relatively precise temperature record, see Fig. 14.5.

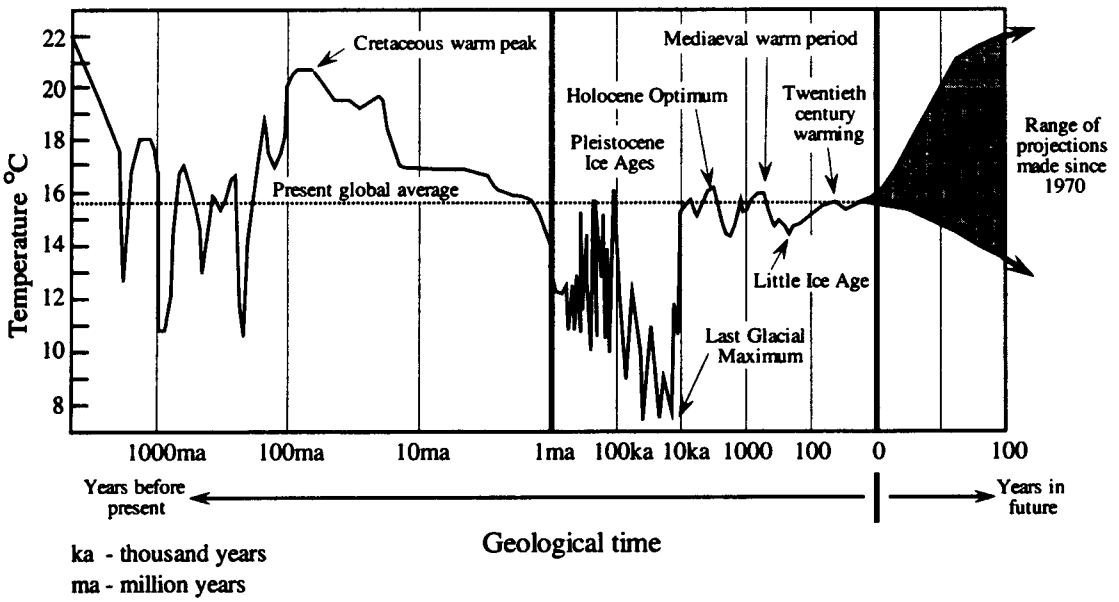


Figure 14.5: Temperature record of the Earth over the past billion years. The glacial cycles during the last million years are clearly visible. The timescale is logarithmic, with breaks at one million years and the present. From Bryant (1997).

When this temperature record is examined, it is found that there have been cyclic

variations of 4—10 degrees in the Earth's temperature during the last million years, which have led to the repeated onset of ice-ages. This is called cyclic glaciation. At present, at least twenty distinct ice-ages have been identified.

The climate tends to drift slowly into ice-ages, becoming continuously colder, until the trend is rapidly terminated by sudden warming. The last ice-age, termed the Last Glacial, peaked 22,000 years ago, after which temperatures rose abruptly to their present values in as little as three to five years.

Currently, it is thought that a significant contribution to cyclic glaciation comes from a combination of eccentricity, obliquity and precession. This so-called Milankovitch hypothesis states that the three orbital effects lead to variations in the degree of radiation reaching various latitudes on the Earth's surface.

The variability in received radiation, combined with the unequal distribution of land masses across the hemispheres causes the cycles of glaciation. The distribution of land masses contributes to the Milankovitch effect, because land has much less heat capacity than sea. If land and sea were equally distributed on the Earth's surface, orbital variations would have little effect on global climate. Currently, most of the land masses are on the Northern hemisphere, so that changes in orbital parameters do have significant effects.

Although variation in eccentricity does not change the total annual amount of received radiation by the Earth, in combination with the obliquity, it does alter the ratio of radiation received by the two hemispheres (see Fig. 14.6).

When the eccentricity equals zero, the Earth's orbit is circular, and both hemispheres receive the same amount of radiation over the year, regardless of the obliquity. However, when the orbit becomes eccentric, the Earth's rotation axis is tilted differently when the Earth is closest to the Sun than when it is furthest from the Sun, and thus one hemisphere receives more radiation than the other.

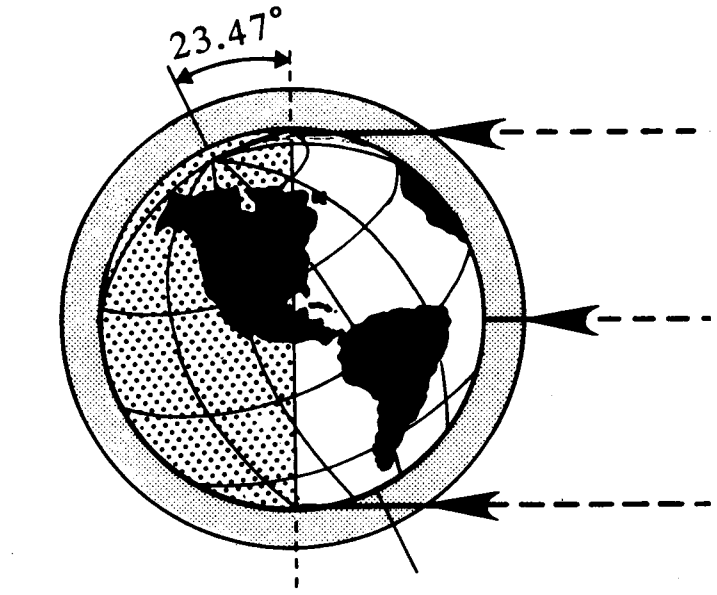
Due to the current eccentricity, the southern hemisphere receives 6.7% more radiation than the northern hemisphere, and this difference can increase to 28%. The largest effect is obtained when the rotation axis is tilted towards the Sun at the same time that the orbital distance is either largest or smallest.

The precession of the equinoxes changes the interrelationship between obliquity and eccentricity in the sense that in the course of time the moment at which the Earth's rotation axis is tilted towards the Sun gradually changes from the time when the Earth is closest to the Sun to the time when it is furthest away.

The cycles of these three orbital effects, as well as several of their harmonics and mixed harmonics of interaction, can be found in the temperature records of the ice-ages. However, although eccentricity, obliquity and precession control the timing of the ice-ages, their impact is not large enough to explain the magnitude of the glacial variations, especially the rapid temperature fluctuations during and at the end of ice ages. It is therefore thought that factors amplifying the Milankovitch effect can be found in the Earth-ocean-atmosphere system and in geological influences such as volcanism and geomagnetism.

14.4 Short timescales: the not-so-quiet Sun

On time scales of thousands of years to months or days, the solar evolution is unnoticable and we may assume that the geometry of the Sun-Earth system does not vary. The main solar influence on the Earth on these time-scales is generated by the rich variety of magnetic structures at the solar surface and in the solar atmosphere, which are driven by



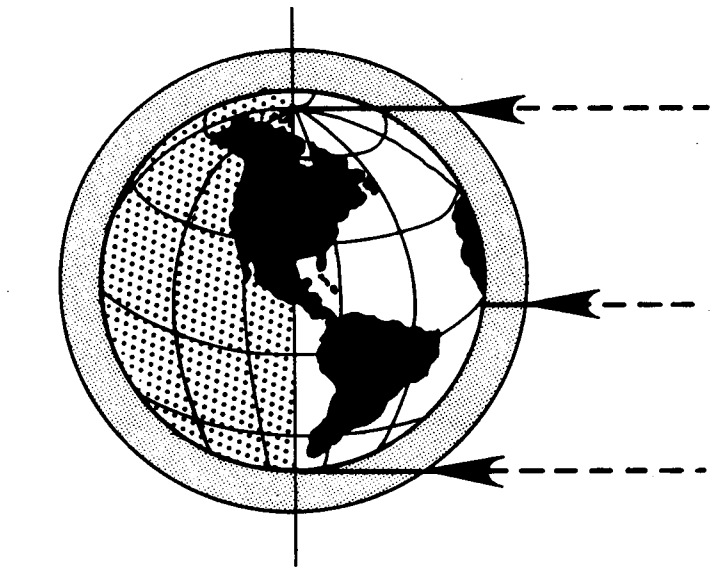


Figure 14.6: The effect of obliquity upon the input of solar radiation at the Earth's surface. From Bryant (1997).

the solar dynamo.

Since these structures have already been covered in these lecture notes, we will restrict treatment of them to a simple enumeration of the different phenomena, with Figs 14.7—

14.13 giving a graphical overview. However, the evidence (or lack of it) for the effect of these phenomena on the Earth's climate will be extensively discussed.

14.4.1 Magnetic structures on the Sun

Sunspots (Fig. 14.7) are regions on the Sun's surface where magnetism increases 100—1,000 times above average. The intense magnetic field inhibits convection, and as a result, a sunspot appears darker than the adjacent surface, so a sunspot causes the solar flux integrated over the entire disc, a quantity called the irradiance, to decrease.

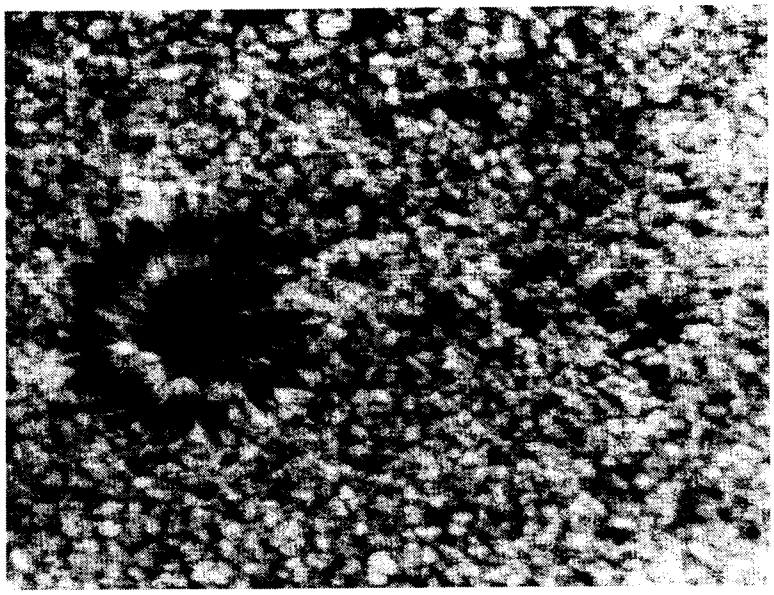


Figure 14.7: A picture of a sunspot taken in August 1985 from Spacelab 2 on the Space Shuttle *Challenger*. The photo resolves features as small as 400 kms. The sunspot structure of a central, dark, *umbra* surrounded by a brighter, filamentary, *penumbra* is clearly visible. From Wentzel (1989).

Bright areas known as **faculae** or **plage** surround sunspots. They are formed by a concentration of small magnetic features, called magnetic elements. Although each magnetic element causes the Sun to brighten by many orders of magnitude less than a sunspot causes it to dim, they are present in sufficient numbers to cause the total irradiance of the Sun to increase when sunspots are numerous (see Fig. 14.8).

Solar flares (Fig. 14.9) are explosions leading to ejections of protons and electrons from the Sun's atmosphere at speeds in excess of 1 500 $\mathrm{km\,s^{-1}}$ and often develop in active regions, which also harbor sunspots. A typical flare takes about an hour to fade away. Flares enhance the solar wind, and are accompanied by a pulse of electro-magnetic radiation in the X-ray and radio spectrum. Very large solar flares are called proton events, because they lead to the ejection of highly energetic protons, also called **solar cosmic rays**.

Prominences (Fig. 14.10) are magnetic structures of relatively cool plasma, which are held up in the, much hotter, million degree corona by magnetic fields. They usually have a slab-like structure, surrounded by loop-like magnetic field lines. After a period which can range from hours to months, prominences destabilize. The method with which this occurs can vary from a slow dimming and dissapearing *in situ*, to a violent disruptive eruption called a **disparition brusque** (Fig. 14.11).

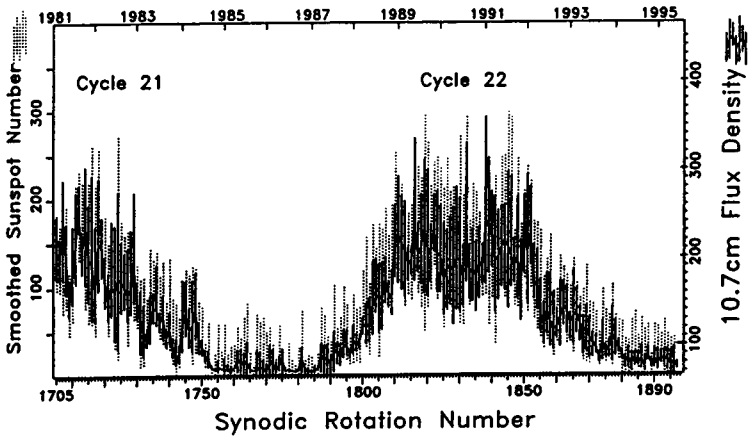


Figure 14.8: The irradiance of the Sun plotted against the sunspot number. From Bishop (1996).

Figure 14.9: A solar flare observed in the light of hydrogen on August 7, 1972. From Wentzel (1989).

Under certain circumstances, these explosions can trigger a violent ejection of matter from the corona, called a **coronal mass ejection** (Fig. 14.12). Both disparition brusques

and coronal mass ejections enhance the solar wind and cause a strong X-ray pulse to be emitted.

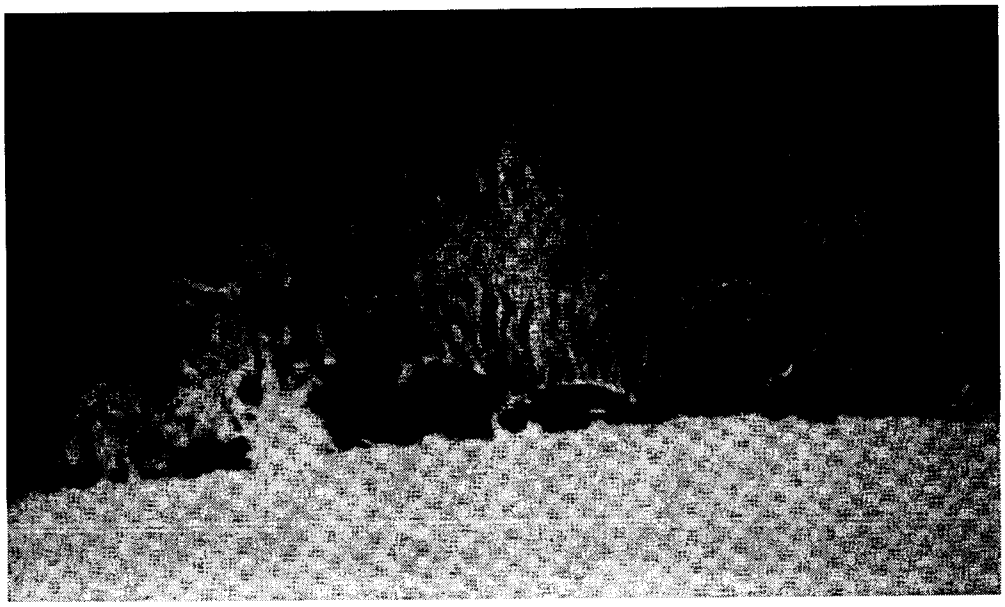


Figure 14.10: A filamentary prominence, suspended above the solar surface. From Tandberg-Hanssen (1995).

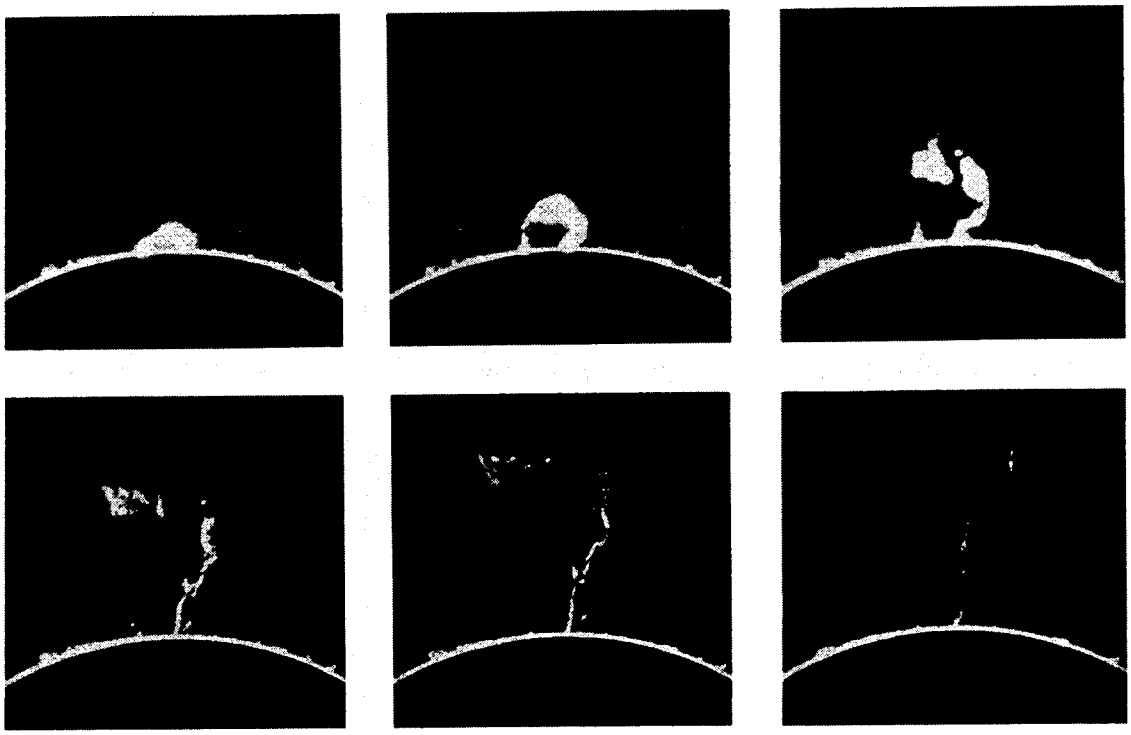


Figure 14.11: A disparition brusque event where one end of the prominence remains attached to the chromosphere. From Tandberg-Hanssen (1995).

Within the solar corona, the gas pressure is sufficiently larger than in the interstellar

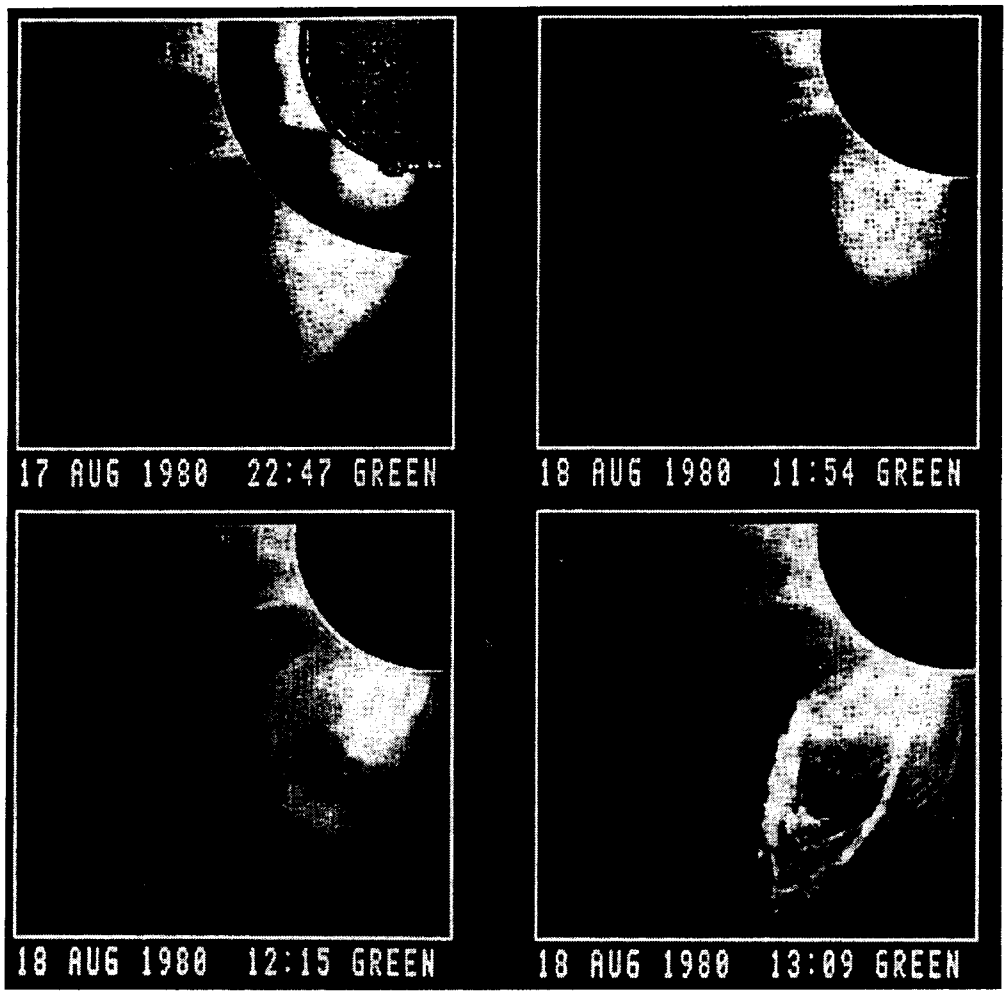


Figure 14.12: Development of a coronal mass ejection. The associated prominence is visible in the composite picture at 2247 UT. Following the ejection, the death of the prominence leads to the typical thread-like structure of a disparition brusque (1309 UT). From Tandberg-Hanssen (1995).

medium to drive gas away from the Sun at supersonic speeds, despite the pull of gravity. This outward flow of coronal gas is called the **solar wind**. Under normal circumstances, most of the solar wind doesn't reach the Earth's surface, but is redirected to flow past the Earth along the so-called magnetopause, shown in Fig. 14.13.

The Sun's **magnetic field** is highly complex and dynamic in the solar interior and atmosphere. Further from the Sun, it is strongly affected by the solar wind, and forms a large-scale spiral pattern.

In the absence of forcing by the solar wind and magnetic field, the Earth's magnetic field would be close to that of a dipole. However, due to the strong solar magnetic influence, the Earth's field is compressed on one side and highly expanded on the other, with a tail extending in the anti-solar direction for hundreds of Earth radii (see Fig. 14.13).

14.4.2 The solar cycle

When the number of sunspots is plotted as a function of time, it is found that sunspots occur in cycles, lasting an average 11.2 years (see Fig. 14.14). Within a cycle, it takes

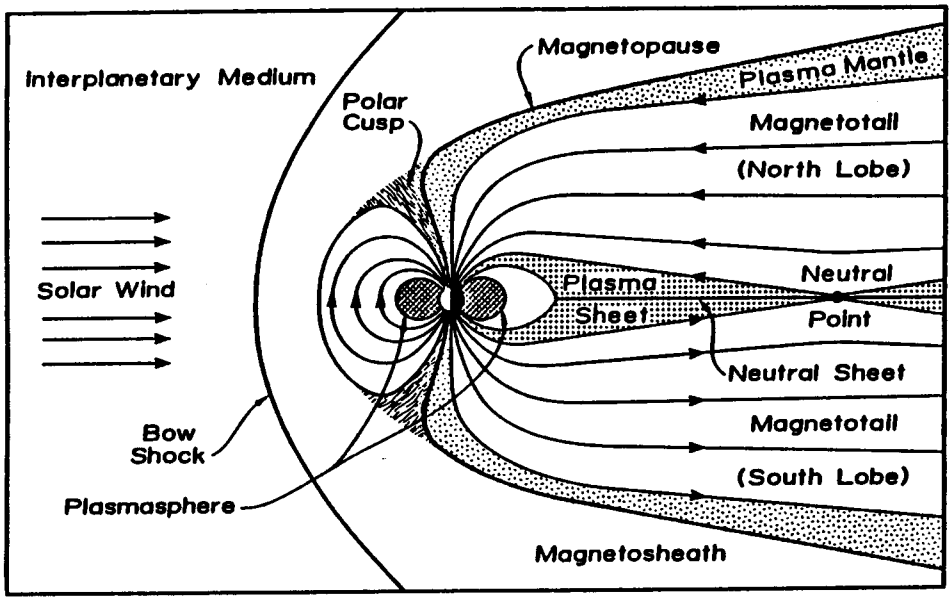


Figure 14.13: A schematic representation of the Earth's magnetosphere. The magnetosphere is bordered on the outside by the magnetopause, the boundary between the solar and terrestrial magnetic field and on the inside by the ionosphere, the upper ionized layer of the Earth's atmosphere. From Akasofu and Kamide (1987).

approximately four years to rise to a maximum number of sunspots, and roughly seven years to fall to a minimum. Between alternate cycles, the magnetic polarity of sunspots is reversed and thus the Sun has a fundamental twenty-two year magnetic cycle, termed the Hale cycle, consisting of two 11.2 year sunspot cycles with reversed polarity.

This Hale cycle is not only found in sunspots, but turns out to apply to all solar magnetic activity. The magnetic field varies in strength and shape from being directed along the poles in a dipole shape at minimum, called a poloidal configuration, to a more circular configuration along the equatorial plane at maximum, called a toroidal configuration.

Furthermore, other features and processes of magnetic origin, such as solar flares and prominences, also depend heavily on the solar cycle in the sense that they are far more common at activity maximum than at minimum. This cycle is therefore said to control solar activity as a whole.

The explanation of the Hale cycle is thought to lie in the so-called dynamo theory. In this theory convection and solar rotation convert kinetic energy into thermal and magnetic energy to yield cyclic variations in the magnetic field, which, in turn, give rise to the chromospheric magnetic structures. Dynamo theories are currently poorly understood, and there is still much work to be done before they can fully explain and predict the solar cycle. Chapters 6) and 7 give more information on the current state of dynamo theories.

14.4.3 Influences on the Earth

On timescales of up to centuries, there is usually quite reliable data available and the problem becomes separating the solar influence from other climatic effects or random variations. Furthermore, suggestions that certain climatic systems may have a chaotic character due to the non-linear equations that drive them have not helped matters.

We consider here the influence of the solar magnetic activity on the magnetosphere,

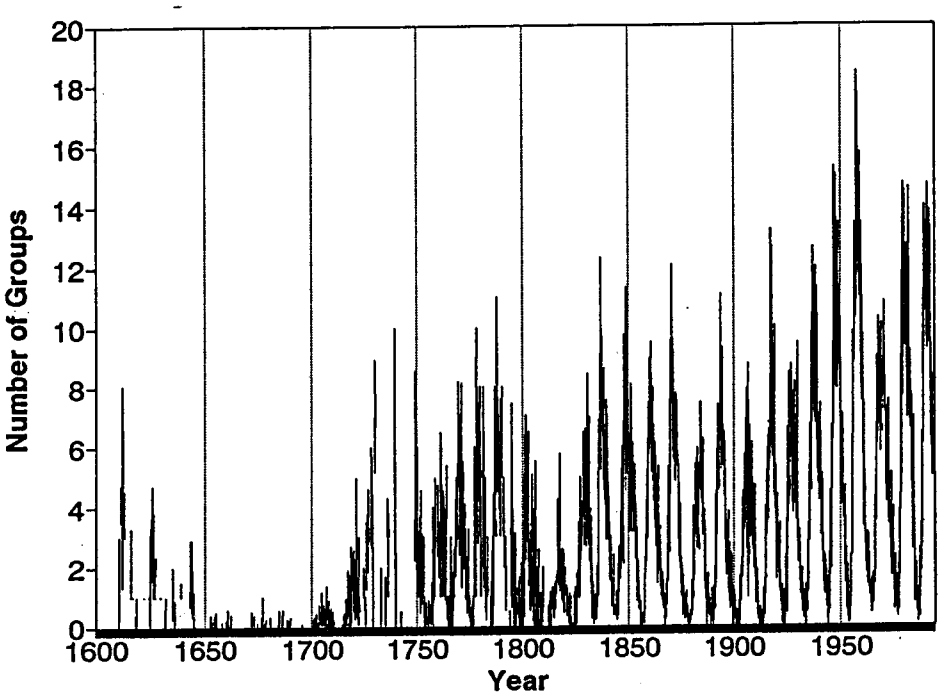


Figure 14.14: Sunspot numbers between the years 1600 and 1993. From Hoyt *et al.* (1993).

ionosphere, and on the temperature in the troposphere, the prime parameter describing the Earth's climate.

14.4.3.1 Magnetospheric and ionospheric influences

Events that enhance the solar wind, such as coronal mass ejections and, to a lesser extent, flares, can dramatically increase the influx of solar matter and photons in the upper ionosphere of the Earth. This leads to effects on both the Earth's magnetosphere and it's ionosphere.

As the solar wind, which acts as a conductor, flows into the magnetic field of the Earth, a dynamo effect takes place, leading to energy production. Most of this energy is dissipated in the lower ionosphere, leading to effects like the *Aurorae Borealis*, or Northern lights, see Fig. 14.15.

A magnetic effect is caused by the leakage of solar wind particles through the magnetopause, via the regions known as the polar cusps. These particles become trapped in the Earth's magnetic field, resulting in a depression of the magnetic field on the surface of the Earth, causing a so-called geomagnetic storm.

Although some of the details are not completely understood, both the above effects have been confirmed experimentally.

14.4.3.2 Influences on climate

To determine the influence of solar activity on the Earth's temperature, data on both the solar activity and the Earth's temperature are required. There are a number of proxies available to measure both parameters up to several thousand years ago, although with steadily decreasing accuracy at increasingly earlier times.

Figure 14.15: A photograph of the aurora borealis. Taken from Akasofu and Kamide (1987).

Solar magnetic activity can be followed, although somewhat qualitatively and with a time resolution lower than 10—20 years, by measuring C^{14} activity up to 7000 years back (see Stuiver (1993)). Indirect temperature measurements can be made by means of data on the sizes of glaciers.

The only directly measured data comes from the last three centuries, during which both the solar activity (mainly the number of sunspots) and the temperature have been actively measured. Initially, these measurements were of low accuracy, but the quality of the data has steadily improved. Direct measurements of solar irradiance have only been made since 1978.

When the available data are compared, it is found that there is a very strong correlation between average land mass temperature and solar activity as measured by the inverse length of the solar cycle, see Fig. 14.16. The length of this cycle varies roughly inversely with sunspot activity: a shorter cycle is generally associated with more solar activity.

However, the solar cycle length is an arbitrary choice from a wide range of available variables. When other indicators for solar activity, such as the sunspot number, are used, the correlation is of far lower quality, to the point were temperature variations precede sunspot variations!

To clarify this situation, a physical mechanism is required. There are currently two suggested mechanisms which may influence the temperature.

The first is caused by the solar variation in brightness. Due to faculae, the Sun is brighter during active periods. On the long-wavelength end of the spectrum, it is proposed that this increase may influence the Earth's radiation budget, leading to a rise in the equilibrium temperature.

The increase on the short-wavelength end, where the rise in irradiance is largest, leads

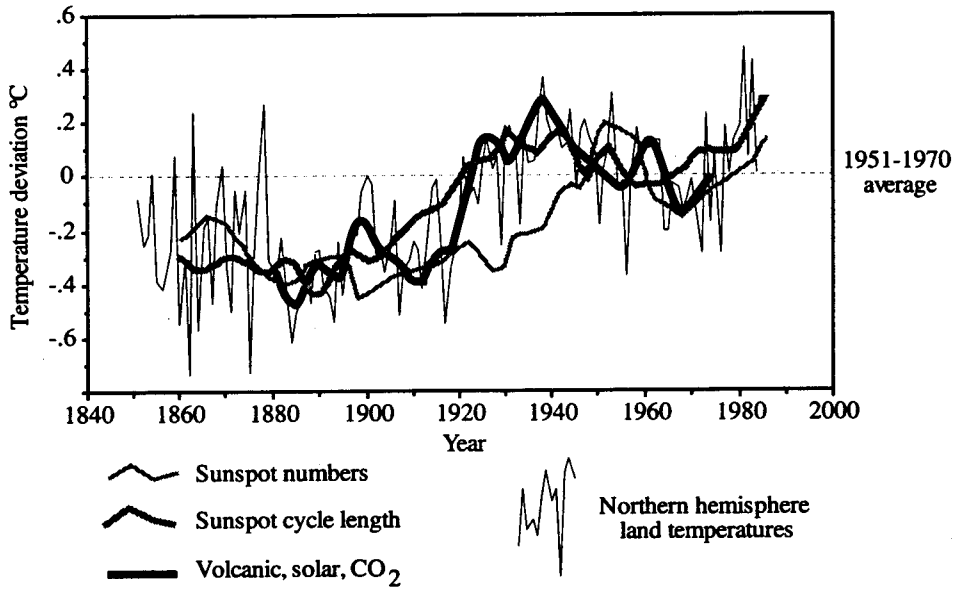


Figure 14.16: The temperature record of the Northern hemisphere, with the results from various models for temperature fluctuations superimposed. From Bryant (1997)

to a decrease in the ozone concentration in the stratosphere due to the reaction

$$O_3 + h\nu \longrightarrow O_2 + O \tag{14.2}$$

for photons with wavelengths below 4000 Å. The reaction rate increases as the influx of solar ultraviolet photons does.

Less ozone means that ultraviolet solar radiation can reach the Earth's surface unhindered. Hence the amount of ultraviolet radiation reaching the Earth's surface changes far more than its solar output does.

A second proposed mechanism deals with cosmic rays. The amount of cosmic rays that reaches the Earth is strongly dependent on the Sun's magnetic field, which in turn is correlated with solar activity. When the Sun is more active, and its magnetic field is stronger, less cosmic rays reach the Earth.

It is currently thought that a rise in cosmic rays causes an increase in cloud formation. This leads to a cooler Earth, since the main effect of clouds is to reflect sunlight, thus keeping it from reaching the Earth's surface, where it would otherwise be absorbed.

Because the oceans of the Earth have a huge heat capacity and large seasonal temperature variations need to be averaged out, we require data on solar irradiance and magnetic field strength or cosmic flux over a period of several hundred years to accurately verify if these models do indeed describe how the Sun influences the Earth's climate.

There are currently two main ways forward in the solution of this problem. On the one hand, data on solar irradiance may be obtained from solar models, which are becoming increasingly accurate.

Alternatively, there may be proxies for these parameters that have not yet been examined. It is currently thought that Beryllium-10 (Be^{10}) may be a proxy for the cosmic ray flux (see Beer *et al.* (1993)), but this is still somewhat uncertain.

14.5 Conclusion

The influence of solar variations on the Earth's climate has been examined on three different timescales. On each timescale we have found evidence for solar influence, and we have discussed possible physical models to explain this influence.

For each timescale, to further elucidate the solar influence there are two main avenues of research. Firstly, more accurate data is required on several solar parameters, which can be obtained by using more accurate models and new or improved proxies. Secondly, the proposed physical models must be further refined, using the newly obtained data.

To perfect our understanding of solar-terrestrial forcing and increase our ability to separate anthropogenic influences from natural ones, it is imperative that both avenues are fully explored.

References

Akasofu, S. and Kamide, Y. (Eds.): 1987, *The solar wind and the Earth*, Terra Scientific Publishing Company, Kluwer, Tokyo

Alfvén, H.: 1942, *Nature* **405**, 3805

Beer, J., Joos, F., Lukasczyk, C., Mende, W., Rodriguez, J., Siegenthaler, U., and Stellmacher, R.: 1993, in E. Nesme-Ribes (Ed.), *The solar engine and its influence on terrestrial atmosphere and climate*, NATO ASI Series I 25, Springer-Verlag, Berlin, p. 221

Bishop, R. L.: 1996, *Observers handbook 1996*, RASC, Canada

Böhm-Vitense, E.: 1989, *Introduction to stellar astrophysics. I. Basic Stellar Observations and Data*, Cambridge Univ. Press, Cambridge UK

Bowers, R. L. and Deeming, T.: 1984, *Astrophysics I. Stars*, Jones and Bartlett, Boston

Boyd, T. and Sanderson, J.: 1969, *Plasma dynamics*, Nelson, London

Bransden, B. H. and Joachain, C. J.: 1989, *Introduction to quantum mechanics*, Longman Scientific & Technical, Harlow, U.K.

Bruzek, A. and Durrant, C. J.: 1977, *Illustrated glossary for solar and solar-terrestrial physics*, Reidel, Dordrecht

Bryant, E.: 1997, *Climate process & change*, Cambridge Univ. Press, Cambridge UK

Bullard, E. C.: 1955, *Proc. Cambr. Phil. Soc.* **51**, 744

Bünte, M., Solanki, S. K., and Steiner, O.: 1993, *A&A* **268**, 736

Clinton, W. J.: 1997, *US President Clinton and Vice President Gore open climate discussion*, `http://www.globalchange.org/moderall/97sep33f.htm`

Cowling, T. G.: 1980, *Magnetohydrodynamics*, Hilger, Bristol

Cox, A. N., Livingstone, W. C., and Matthews, M. S. (Eds.): 1991, *Solar Interior and Atmosphere*, The University of Arizona Press

Donati, J.: 1999, *Astron. Soc. Pac. Conf. Ser.* 27–37

Foukal, P.: 1990, *Solar Astrophysics*, Wiley and Sons, New York

Galloway, D. J. and Weiss, N. O.: 1981, *ApJ* **243**, 945

Goedbloed, J.: 1971, *Physica* **53(3)**, 412

Goedbloed, J.: 1998, *Phys. Plasmas* **5(9)**, 3143

Goedbloed, J. and Poedts, S.: 1997, *Magnetohydrodynamics of Laberatory and Astrophysical Plasmas*, Lecture notes, Utrecht University

Goedbloed, J. P. and Lifschitz, A.: 1997, *Phys. Plasmas* **4**, 3544

Golub, L. and Pasachoff, J. M.: 1997, *The Solar Corona*, Cambridge Univ. Press, Cambridge UK

Hartmann, D. L.: 1994, *Global physical climatology*, Academic Press, Inc., San Diego

Harvey-Angle, K. L.: 1993, *Magnetic Bipoles on the Sun*, PhD Thesis, Utrecht University

Hatzes, A. P.: 1996, in K. G. Strassmeier and J. L. Linsky (Eds.), *IAU Symp. 176, Stellar Surface Structure*, Dordrecht, Kluwer, p. 305

Holman, G. and Benedict, S.: 1996a, *Coronal Mass Ejections, Solar Flares, and the Sun-Earth Connection*, `http://hesperia.gsfc.nasa.gov/sftheory/flare.htm`

Holman, G. and Benedict, S.: 1996b, *What is a Solar Flare*, `http://hesperia.gsfc.nasa.gov/sftheory/flare.htm`

Hoyng, P.: 1992, in J. T. Schmeltz and J. C. Brown (Eds.), *The Sun, a Laboratory for Astrophysics*, Kluwer Academic Publishers, p. 99

Hoyng, P.: 1994, *Plasma-astrophysics*, Lecture notes, Utrecht University

Hoyt, D. V., Schatten, K. H., and Nesme-Ribes, E.: 1993, in E. Nesme-Ribes (Ed.), *The solar engine and its influence on terrestrial atmosphere and climate*, NATO ASI Series I 25, Springer-Verlag, Berlin, p. 57

Jackson, J. D.: 1962, *Classical Electrodynamics*, Wiley & Sons, New York

Keppens, R., Goedbloed, J. P., and Tóth, G.: 1999, *A&A* in press

Kivelson, M. G. and Russell, C. T.: 1995, *Introduction to Space Physics*, Cambridge Univ. Press, Cambridge UK

Kuperus, M.: 1998, *Gasdynamica en magnetohydrodynamica*, Lecture notes, Utrecht University

Lamers, H. J.: 1997, *Sterinwendigen en evolutie*, Lecture Notes, Sterrekundig Instituut Utrecht (in Dutch)

Landi Degl'Innocenti, E.: 1992, in F. Sánchez, M. Collados, and M. Vázquez (Eds.), *Solar Observations: Techniques and Interpretation*, First Canary Islands Winter School, Cambridge Univ. Press, Cambridge UK, p. 71

Landi Degl'Innocenti, E. and Landi Degl"Innocenti, M.: 1985, *Solar Phys.* **97**, 239

Landstreet, J. D.: 1992, *A&AR* **4**, 35

Lang, K. R.: 1995, *Evolution of life: Sun, Earth and sky*, Springer-Verlag, Berlin

Linsky, J. L.: 1985, *Solar Phys.* **100**, 333

Lites, B. W., Rutten, R. J., and Berger, T. E.: 1999, *ApJ* **517**, , in press

Löfdahl, M. G., Berger, T. E., Shine, R. S., and Title, A. M.: 1998, *ApJ* **495**, 965

MacGregor, K. and Charbonneau, P.: 1997, *ApJ* **486**, 484

Martínez Pillet, V., Lites, B. W., and Skumanich, A.: 1997, *ApJ* **474**, 810

Mestel, L. and Weiss, N.: 1974, *Magnetohydrodynamics*, Fourth Saas Fee Course, Geneva Observatory

Middelkoop, F.: 1982, *Ca II H and K emission from late-type stars*, PhD thesis, Utrecht University

Oranje, B. J.: 1985, *Solar-type Stellar Chromospheres*, PhD thesis, Utrecht University

Parker, E. N.: 1979, *Cosmical magnetic fields*, Clarendon Press, Oxford

Parker, E. N.: 1993, *ApJ* **408**, 707

Parker, E. N.: 1996, in T. R. Corés and F. Sánchez (Eds.), *The Structure of the Sun*, 6th Canary Islands Winter School, Cambridge Univ. Press, Cambridge UK, p. 301

Phillips, K. J. H.: 1992, *Guide to the Sun*, Cambridge Univ. Press, Cambridge UK

Piskunov, N.: 1996, in K. G. Strassmeier and J. L. Linsky (Eds.), *IAU Symp. 176, Stellar Surface Structure*, Dordrecht, Kluwer, p. 45

Ploner, S. R. O. and Solanki, S. K.: 1997, *A&A* **325**, 1199

Pneuman, G. W., Solanki, S. K., and Stenflo, J. O.: 1986, *A&A* **154**, 231

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

Priest, E. R. and Hood, A. (Eds.): 1991, *Advances in solar system magnetohydrodynamics*, Cambridge Univ. Press, Cambridge UK

Reagan, M.: 1997, *Global warming is a hoax*, `http://www.reagan.com/ HotTopics.main/ HotMike/document-7.28.1997.3.html`

Rice, J. B.: 1996, in K. G. Strassmeier and J. L. Linsky (Eds.), *IAU Symp. 176, Stellar Surface Structure*, Dordrecht, Kluwer, p. 19

Roberts, B.: 1991, in E. R. Priest and A. Hood (Eds.), *Advances in solar system magnetohydrodynamics*, Cambridge Univ. Press, p. 117

Roberts, P. H.: 1967, *An Introduction to Magnetohydrodynamics*, Longmans, London

Rodono, M., Cutispoto, G., Pazzani, V., Catalano, S., Byrne, P. B., Doyle, J. G., Butler, C. J., Andrews, A. D., Blanco, C., Marilli, E., Linsky, J. L., Scaltriti, F., Busso, M., Cellino, A., Hopkins, J. L., Okazaki, A., Hayashi, S. S., Zeilik, M., Helston, R., Henson, G., Smith, P., and Simon, T.: 1986, *A&A* **165**, 135

Rüedi, I., Solanki, S. K., Livingston, W., and Stenflo, J. O.: 1992, *A&A* **263**, 323

Rutten, R. G. M.: 1987, *Magnetic activity of cool stars and its dependence on rotation and evolution*, PhD thesis, Utrecht University

Rutten, R. J.: 1999a, in B. Schmieder, A. Hofmann, and J. Staude (Eds.), *Solar Magnetic Fields and Oscillations*, Procs. Third Adv. Solar Physics Euroconf., Astron. Soc. Pac. Conf. Series, in press

Rutten, R. J.: 1999b, *Radiative Transfer in Stellar Atmospheres*, Lecture Notes Utrecht University, 5th WWW Edition

Rutten, R. J. and Schrijver, C. J. (Eds.): 1994, *Solar Surface Magnetism*, NATO ASI Series C 433, Kluwer, Dordrecht

Schatzman, E. L. and Praderie, F.: 1993, *The Stars*, Springer, Berlin

Schrijver, C. J.: 1986, *Stellar magnetic activity. Complementing conclusions based on solar and stellar observations*, PhD thesis, Utrecht University

Schrijver, C. J. and Zwaan, C.: 1999, *Solar and stellar magnetic activity*, Cambridge Univ. Press, in press

Schutgens, N. A. J.: 1998, *Oscillating Prominences*, PhD Thesis, Utrecht University

SOHO: 1998, `http://sohowww.estec.esa.nl/`

Solanki, S.: 1993, *Space Sc. Rev.* **63**, 1

Solanki, S.: 1996, in K. G. S. . J. L. Linsky (Ed.), *IAU Symp. 176, Stellar Surface Structure*, Dordrecht, Kluwer, 1996, p. 201

Stenflo, J. O.: 1984, *Appl. Optics* **23**, 1267

Stenflo, J. O.: 1994, *Solar Magnetic Fields, Polarized Radiation Diagnostics*, Kluwer, Dordrecht

Stix, M.: 1976, in V. Bumba and J. Kleczek (Eds.), *Basic Mechanisms of Solar Activity*, Reidel, Dordrecht, p. 367

Stix, M.: 1989, *The Sun. An Introduction*, Springer, Berlin

Stuiver, M.: 1993, in E. Nesme-Ribes (Ed.), *The solar engine and its influence on terrestrial atmosphere and climate*, NATO ASI Series I 25, Springer-Verlag, Berlin, p. 203

Tandberg-Hanssen, E.: 1995, *The nature of solar prominances*, Kluwer, Dordrecht

Tayler, R. J.: 1997, *The Sun as a Star*, Cambridge Univ. Press, Cambridge UK

Vaiana, G. S., e.: 1981, *ApJ* **244**, 163

van Ballegooijen, A.: 1982, *ApJ* **113**, 99

Vaughan, A. H., Preston, G. W., Baliunas, S. L., Hartmann, L. W., Noyes, R. W., Middelkoop, F., and Mihalas, D.: 1981, *ApJ* **250**, 276

Vogt, S. S., Penrod, G. D., and Hatzes, A. P.: 1987, *ApJ* **321**, 496

Wang, W. and Sheeley, N. R.: 1994, *ApJ* **430**, 399

Wentzel, D. G.: 1989, *The restless Sun*, Smithsonian institution Press, Washington, D.C.

Wilson, P. R.: 1994, *Solar and Stellar Acitivity Cycles*, Cambridge Univ. Press, Cambridge UK

Yoshimura, H.: 1975, *ApJ* **201**, 740

Zwaan, C., van Ballegooijen, A. A., Brants, J. J., Middelkoop, F., Oranje, B. J., and Savonije, G. J.: 1982, *Magnetische aktiviteit in zon en sterren*, Lecture notes, Utrecht University

Index