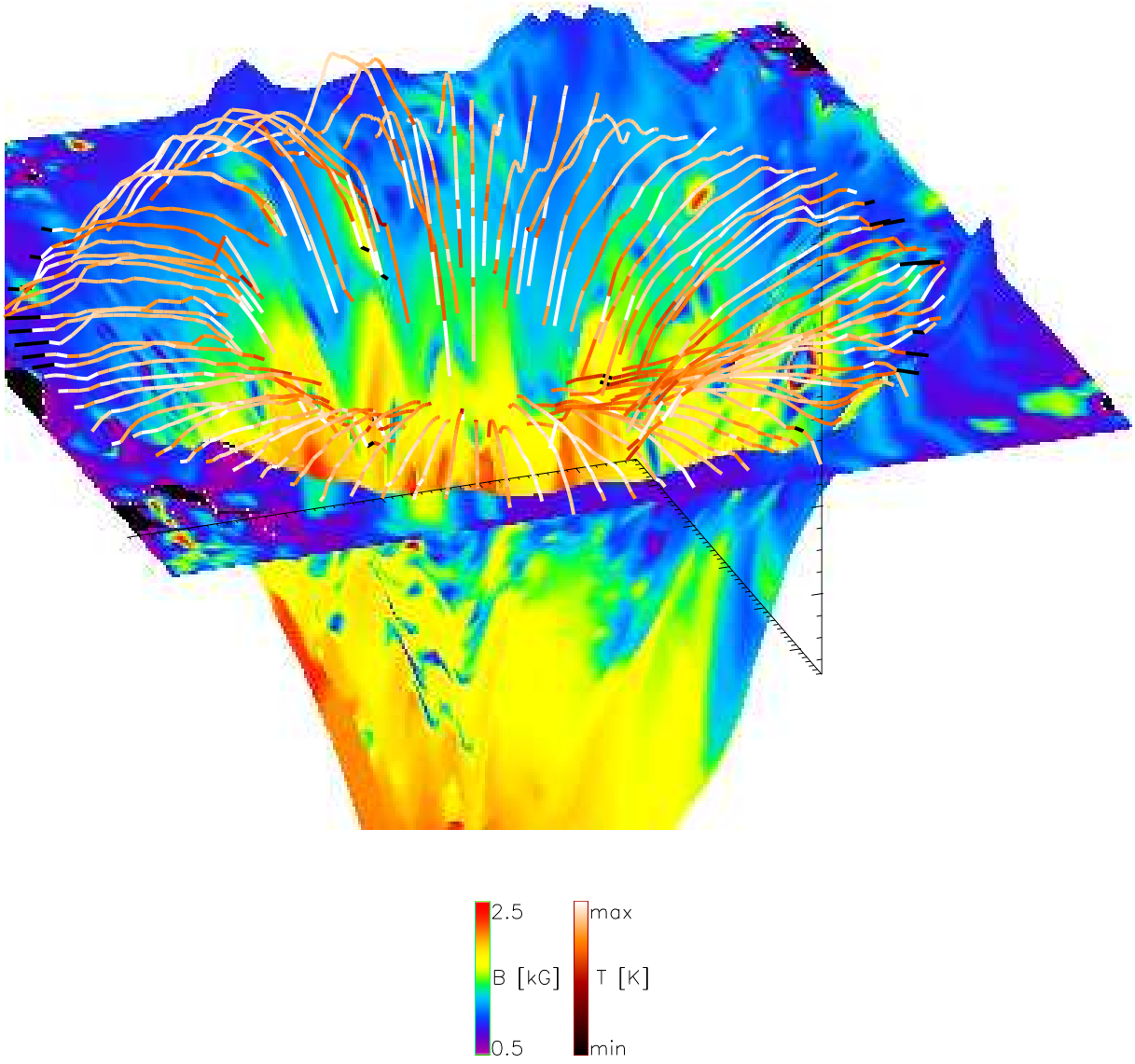
The 3-D topology of magnetic fields in and around sunspots



Christian Beck

Title page: *The 3-dimensional sunspot model which was constructed from the observations. The color code of the surface gives the field strength of the quasi-static background field; red corresponds to a field strength of 2500 Gauss, purple to 500 Gauss. The flux tube component that carries the Evershed flow is overplotted as lines at the derived geometrical height; the color of the lines indicates the temperature of the elevated loops.*

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Mathematik und Physik
der Albert-Ludwigs-Universität Freiburg im Breisgau

The three-dimensional topology of magnetic fields in and around sunspots

Christian Beck

Kiepenheuer Institut für Sonnenphysik, Freiburg
Dezember 2005

Dekan:	Prof. Dr. J. Honerkamp
Betreuer:	PD Dr. habil. W. Schmidt
Referent:	PD Dr. habil. W. Schmidt
Korreferent:	Prof. Dr. G. Strobl
Datum der mündlichen Prüfung:	26.1.2006

In the context of this thesis the following articles have been published:

Papers in Refereed Journals

1. Beck, C.A.R., Schmidt, W., Kentischer, T. & Elmore, D.
   Polarimetric Littrow Spectrograph - instrument calibration and first measurements, Astronomy & Astrophysics, 2005, V. 437, p. 1159-1167

2. Beck, C.A.R., Schlichenmaier, R., Collados, M., Bellot Rubio, L. R. & Kentischer, T.
   A polarization model for the German Vacuum Tower Telescope from in situ and laboratory measurements, Astronomy & Astrophysics, 2005, V. 443, p. 1047-1054

3. Bellot Rubio, L.R. & Beck, C.
   Magnetic flux cancellation in the moat of a sunspot: Results from simultaneous vector spectropolarimetry in the visible and infrared, Astrophysical Journal, 2005, V. 626, p. L125-L128

4. Khomenko, E. V., Martínez González, M. J., Collados, M., Vögler, A., Solanki, S. K., Ruiz Cobo, B. & Beck, C.
   Magnetic flux in the internetwork quiet Sun, Astronomy & Astrophysics, 2005, V. 436, L27-L30

5. Schmidt, W., Beck, C., Kentischer, T., Elmore, D. & Lites, B.
   POLIS: A spectropolarimeter for the VTT and for GREGOR, Astronomische Nachrichten, 2003, V. 324, p. 300-301

Conference Proceedings

1. Beck, C., Bellot Rubio, L. R. & Nagata, S.
   Chromospheric and coronal signatures of magnetic flux cancellation in a sunspot's moat, ESA, in press

2. Shimizu, T., Martinez-Pillet, V., Collados, M., Ruiz-Cobo, B., Centeno, R., Beck, C. & Katsukawa, Y.
   Supersonic downflows in the photosphere discovered in sunspot moat regions, 2005, ASP Conference Series, in prep.

Planned Publications

1. Beck, C., Bellot Rubio, L. R. , Collados, M. & Schlichenmaier, R.
   Calculation of differential refraction in multi-wavelength observations, A&A

2. Beck, C., Bellot Rubio, L. R. , Schlichenmaier, R. & Suetterlin, P.
   The magnetic properties of G-Band Bright Points in a sunspot moat, A&A

Contents

Abstract:

The most prominent features on the surface of the solar disc are the sunspots, which have been studied since their detection in the 17th century. Despite this years and centuries of scientific investigations, surprisingly many facets of sunspots are not well understood. In this thesis, the properties of a sunspots' penumbra are derived from the observed spectra by means of an inversion, a forward modeling technique. A method for the construction of a 3-dimensional model of the field topology from the inversion results is presented. It offers the possibility to build a toy-model of the spot. This allows to investigate the relationships between the geometry and other properties of the sunspot like flow velocities, intensity, temperature, or field strength. The application of the procedure to time series of observations will allow to study the temporal evolution of the geometry and the other properties, to determine the physical processes happening and their respective drivers.

A second part of the thesis is devoted to the analysis of isolated field concentrations in the surroundings of the sunspot. These isolated magnetic elements are assumed to be the elementary building blocks of the solar surface magnetism. In the turbulent surroundings of the solar atmosphere, only relatively strong concentrations with a field strength above 1 kG are assumed to be stable at all. Such field concentrations produce brightenings in the G band, where a lot of spectral lines of the CH molecule are present. Inside magnetic fields, the density is reduced, as the magnetic pressure contributes to the pressure balance equation. Thus, the reduced density allows to see to deeper layers of the atmosphere with higher temperature. If the CH-molecules dissociate due to the higher temperature, the spectral lines disappear and the intensity increases strongly. Due to the small size of the field concentration, the brightenings are strongly localized ("G-band Bright Points"). In this study, we find a much larger variability of the Bright Point properties than expected, which puts their elementary nature in some doubt.

In the conclusions of this work, the resulting 3-D topology of the sunspot is used in an attempt to develop a consistent picture of the development and the fine structure of sunspots.

Zusammenfassung:

Die hervorstechendsten Erscheinungen auf der Oberfläche der Sonne sind die Sonnenflecken, die seit ihrer Entdeckung im 17. Jahrhundert untersucht wurden. Trotz dieser Jahre und Jahrhunderte wissenschaftlicher Untersuchungen sind überraschend viele ihrer Facetten nicht wohl verstanden. In dieser Doktorarbeit werden die Eigenschaften der Penumbra eines Sonnenfleckes aus den beobachteten Spektren durch eine Inversion, eine Vorwärtsmodellierung der solaren Atmosphäre, abgeleitet. Eine Methode für die Konstruktion eines 3-dimensionalen Modells aus den Inversionsergebnissen wird beschrieben. Sie bietet die Möglichkeit, ein Spielzeug-Modell des Fleckes zu erstellen. Dieses kann dazu benutzt werden, die Zusammenhänge zwischen der Geometrie und anderen Eigenschaften des Sonnenfleckes wie Strömungsgeschwindigkeit, Intensität, Temperatur oder Feldstärke zu untersuchen. Die Anwendung des Verfahrens auf Zeitreihen von Beobachtungen erlaubt es, die zeitliche Entwicklung der Geometrie und der anderen Eigenschaften zu studieren, um die ablaufenden physikalischen Prozesse und ihre treibenden Kräfte zu finden.

Ein zweiter Teil der Arbeit ist der Analyse von Feldkonzentrationen in der Umgebung des Sonnenfleckes gewidmet. Diese magnetischen Elemente sollen die elementaren Bausteine des Magnetismus auf der Sonnenoberfläche sein. In der turbulenten Sonnenatmosphäre sind nur verhältnismäßig starke Konzentrationen mit einer Feldstärke von über einem kG stabil. Solche Feldkonzentrationen führen zu Aufhellungen im G-Band, in dem eine grosse Anzahl von Spektrallinien des CH-Moleküls vorhanden ist. Innerhalb der magnetischen Felder ist die Dichte verringert, da der magnetische Druck zum Druckgleichgewicht beiträgt. Die verringerte Dichte erlaubt es aber, bis zu tieferen Schichten der Atmosphäre mit höherer Temperatur zu sehen. Wenn die CH-Moleküle wegen der höheren Temperatur dissoziieren, verschwinden die Spektrallinien und die Intensität erhöht sich stark. Wegen der kleinen Ausmaße der Feldkonzentrationen werden die Aufhellungen "G-band Bright Points" genannt. In dieser Studie finden wir eine viel größere Variation in den Eigenschaften der Bright Points als erwartet, was ihre elementare Natur in Zweifel setzt.

In der Zusammenfassung dieser Arbeit wird die sich ergebende 3-D-Topologie des Sonnenflecks verwendet, um ein konsistentes Bild der Entwicklung und der Feinstruktur von Sonnenflecken zu entwerfen.

Chapter 1

Introduction

Preface The main aim of this work was the derivation of the 3-dimensional topology of the magnetic field in sunspots, as a prime ingredient to understand the properties of their fine structure and their evolution. The properties of the magnetic field can in the solar case only be derived by remote-sensing, i.e. from the polarization signal in spectral lines due to the Zeeman effect Thus, instrumentation for measuring the polarization signal with high spatial and spectral resolution is needed. An example of a present state-of-the-art vector spectro-polarimeter is the POlarimetric LIttrow Spectrograph (POLIS, cf. Sect. 2.3.1, or Schmidt et al. 2003; Beck et al. 2005b). As for all observational instruments, the measurement method introduces some errors, which in the present case mainly refer to changes of the polarization state of the light by the telescope or the instrument itself. These effects and methods for their removal have been described in the diploma thesis of the author (Beck 2002, available online under http://www.kis.uni-freiburg.de/cbeck_diplom.pdf), but on a purely theoretical base: the instrument was ready for observations after the end of the diploma thesis in May 2002. It has been found necessary to modify the approach in Beck (2002), to adjust it to the observational results. These modifications, finally leading to the present state of the data calibration and data reduction, will be touched only briefly in the present work. They are described in two publications (Beck et al. 2005a,b). However, achieving the goal of an efficient and robust data reduction code has consumed a considerable amount of time. This is the main reason why some topics of scientific interest, which could be addressed with the data described in the following, are indicated but not tackled.

In the present study, for the first time simultaneous data of two vector spectro-polarimeters in visible and infrared wavelength ranges have been combined. The procedure of the data analysis has been tested firstly on an observational target, which has a less complex internal structure than a sunspot: the isolated magnetic elements in its surroundings.

This leads to the organization of the present work: After the introduction on general properties of the solar atmosphere (Sect. 1.1) and the shape of magnetic fields on the solar surface (Sect. 1.2), the instrumentation (Sect. 2.3) and the data calibration will be described (Sect. 2.4) in some detail. The first part is ended with the description of the data analysis method in Sect. 3.3. The results from the analysis have been divided into two parts, one dealing with the properties of magnetic fields in the surroundings of a sunspot (Chapter 4), and another one describing the structure of the sunspot itself (Chapter 5). The results are combined into a hopefully consistent picture of sunspot fine structure, evolution, and decay in Chapter 6.

If not explicitly stated otherwise, all images of solar observations have been created from data taken by the author with the POLIS instrument.

A general comment I'm writing this thesis such that my father, who is a retired physics teacher, will be able to understand the text. Thus, the introduction is more elaborate than usually expected for a thesis. The reader is kindly asked to show leniency for that.

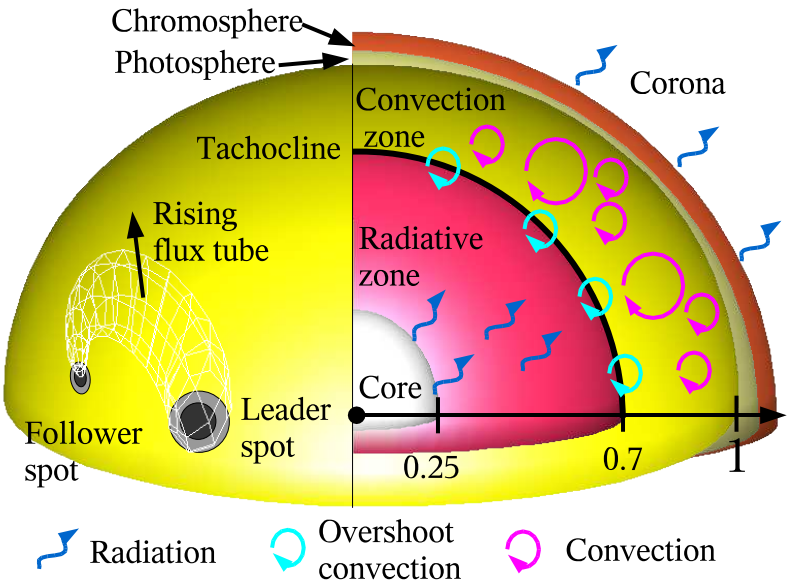


Figure 1.1: The global structure of the Sun. **Left half**: The rise of large-scale flux tubes above the solar surface leads to the appearance of sunspots at their cross-section with the surface. **Right half**: The internal structure of the Sun. The horizontal arrow gives the radial distance in fractions of the solar radius, $r_\odot$. The energy is generated by fusion in the solar core. The energy is transported outwards by radiation up to $r/r_\odot = 0.7$. Across the boundary layer between the rigidly rotating radiative zone and the differentially rotating convection zone, the tachocline, some convective motions can penetrate into the radiative zone (*overshoot convection*). The convection zone with energy transport by mass motions extends from $r/r_\odot = 0.7$ up to the surface. The solar atmosphere above the surface consists of the photosphere, the chromosphere and, finally, the corona, which extends to several solar radii.

1.1 The solar interior & The solar atmosphere

The internal structure of the Sun has been described and modeled in the work of the last century. The problem of maintaining its total energy output, given by the solar luminosity $L_\odot = (3.844 \pm 0.010) \cdot 10^{26}$ W, has been solved due to the discovery of the fusion process taking place inside the solar core. The distinction between solar interior and solar atmosphere is commonly given by the layer, where the solar plasma gets transparent for radiation. The generally accepted picture of the Sun can be summarized as follows:

- The Sun is an ordinary star, with a mass of $2{\cdot}10^{30}$ kg; it consists of a gas plasma that extends over a radius of $7{\cdot}10^{8}$ m from its center.
- In its core the temperature reaches about 16 million Kelvin, allowing for the fusion of hydrogen into helium. The mass deficit of the fusion process maintains the solar energy output. The solar core covers about 25 % of the solar radius (cf. Fig. 1.1).
- Energy transport by radiation is dominant up to 70 % of the solar radius, where it is exchanged to energy transport by mass motions, the convection. This *radiative zone* includes the solar core. The whole radiative zone is supposedly rotating as a rigid body with a constant angular velocity.
- The convection zone starts at 70 % of the solar radius and continues up to the solar surface, where the atmosphere becomes transparent and radiative energy transport takes over again. At the solar surface, the top of the convective cells can be seen in the *granulation* pattern. Inside the convection zone, the Sun shows a *differential rotation*. It is rotating faster at the solar equator than at the poles, with a period around 28 days at low latitudes.
- The magnetic fields appearing on the solar surface are presumably formed at the lower boundary of the convection zone. At the boundary layer between the rigid rotation of the radiative zone and the differential

rotation of the convection zone, the *tachocline*, the convective motions can penetrate into the radiative zone for some depth (*overshoot layer*). According to the current knowledge, mainly based on numerical simulations due to the non-accessibility of this depth by observations, the intrusion of the convection into the stably stratified radiative zone leads to a dynamo effect, where magnetic fields are generated.

- The generated magnetic fields are stored and intensified at the bottom of the convection zone. The differential rotation will drag field lines faster at the equator than at the poles, leading to the generation of flux bundles oriented parallel to the equator after a few revolutions regardless of their earlier orientation. These flux bundles have been found to be partly unstable against perturbations: if the flux is elevated from the equilibrium position at one point along the flux bundle it becomes buoyant there; it continues to rise to the solar surface and even further upwards. The shape of the so-called *active regions* of magnetic activity is due to the properties of the large-scale flux bundle or flux rope: the rope forms a large-scale Ω-shaped loop with an East-West orientation that bi-sects the solar surface in two points, if the top of the loop has risen above the surface (cf. the left part of Fig. 1.1). The cross-section of the flux rope with the solar surface leads to the appearance of bi-polar magnetic structures. In one footpoint the field lines leave the surface, while in the other they return. The flux at the intersection appears first in the form of small-scale fields, like *pores* and diffuse flux, that evolves into full-scale *sunspots*.

- The magnetic activity varies with a period of about 11 years from a solar maximum of activity to the next. The global solar magnetic field changes its polarity during a 22-year cycle. The magnetic flux, visible in spots, pores, and active regions, starts to appear around 30-40 degree heliographic latitude in the beginning of a new cycle, but then appears at decreasing latitudes over the cycle, until the last traces show up near the equator (*butterfly diagram*).

- In the outermost layers of the solar interior, the temperature has decreased to allow for the recombination of the protons and electrons to form hydrogen atoms. Due to the decrease of the number density of absorbers, photons emitted in that depth can escape the solar atmosphere. The light we observe almost completely originates from this layer of some 100 km extension, called the *photosphere*. It covers only the tiniest fraction of the solar radius. The photosphere mainly emits a black-body radiation with an effective temperature around 6000 K. Most absorption lines, which allow the study of the atmospheric properties, are formed within or slightly above the photosphere.

- Up to the photosphere, the temperature is found to decrease with increasing distance from the solar center. However, above the photosphere one of the most puzzling effects in solar physics takes place: the *chromosphere*, visible only at solar eclipses, and the even further out *corona* show an increase of temperature that eventually reaches again 1 million K in the corona. This dramatic number looses some of its terror, if one takes into account that the density of the gas plasma is lower than the best vacuum earthly techniques can produce. Still, one of the major issues in solar physics is an explanation how energy can be effectively transported to that height, if the medium is transparent and does not absorb radiation. The properties of the outer layers of the solar atmosphere, especially of the corona, are intimidately connected to the geo-solar interaction by the particle flow originating from the Sun (*solar wind*). Both the chromosphere – only partly – and the corona show *emission lines* instead of *absorption lines*.

The photosphere and the layers above it are most important for this study. For a more detail review the reader is referred to the textbook of Stix (2002)[1].

1.1.1 The photosphere

The photosphere is a layer of only some 100 km width, in which the solar atmosphere becomes transparent at a temperature of around 6000 K. The last layer of convective energy transport just below is visible in the pattern of the *granulation*.

Hot material rises in the center of granules and cools down by radiating energy into the transparent photosphere. The material then is swept towards the narrow *intergranular lanes*, where it descends again. The time scale of the granulation pattern is several minutes; the granules are continually replaced by new material from below in 10-15

[1]Which is – as usual – called "The Sun: An introduction" even if it has 490 pages, similar to the $>$ 1000 pages of Cohen-Tannoudji which also only give an "Introduction to quantum mechanics".

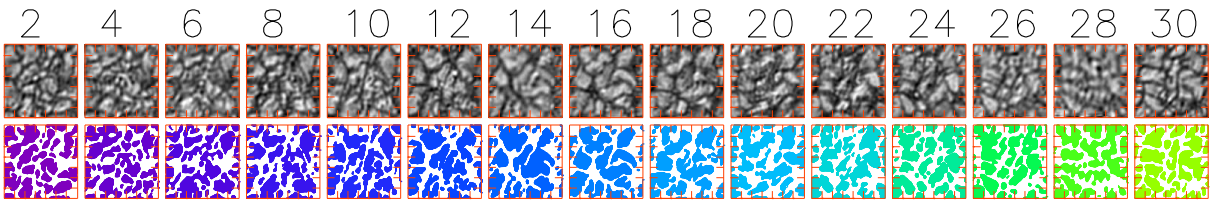


Figure 1.2: The temporal evolution of a small section of the solar surface of 7.5 $\times$ 7.5 arcsec^2 from a time series taken in the blue continuum with the Dutch Open Telescope on La Palma. On top the time in minutes is given. Tickmarks give 1 arcsec. The 2nd row displays a mask of the bright granules at each time. Individual granules are continually replaced by new ones on a time scale of around 10-15 minutes.

minutes (cf. Fig. 1.2). If granulation is observed in spectral lines, the respective velocities of the hot up- and cool downflows leads to a pattern like in Fig. 1.3: bright areas in intensity are correlated with upflows, dark areas, the intergranular lanes, are co-spatial with downflows. The typical size of granules is around 1-2" ($\sim$ 700-1500 km).

The intergranular lanes cover a smaller fraction of the surface than the bright granules. If spectroscopic data is spatially averaged – either the observed profiles before the derivation of the velocity, or the velocities from single spatial positions – a net upwards velocity results due to the different area fraction covered by up- and downflows. The effect is called the *convective blue shift*; its actual value depends on the spectral line and on the location of the observation area on the solar disc (cf. the definition of the wavelength scale in Sect. 3.3.5.2).

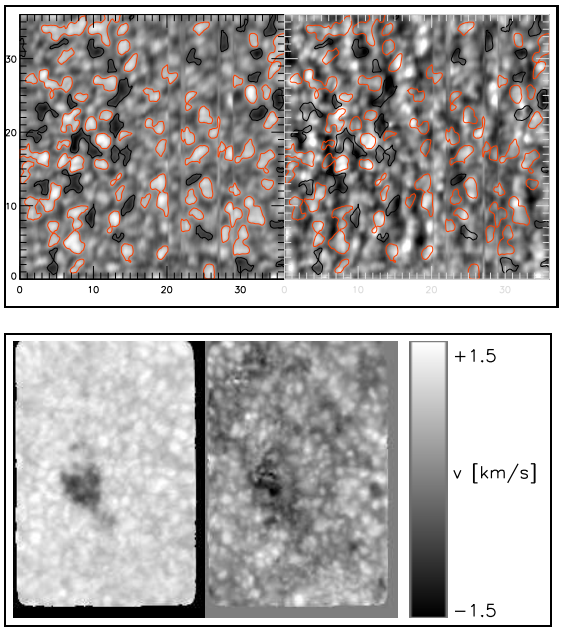


Figure 1.3: **Upper panel**: Intensity map (**left**) and Dopplergram (**right**) of granulation close to disc center as observed with POLIS. The Dopplergram shows velocities from -500 (in black) to +500 m/s (in white). A minus sign indicates motion away from the observer. The tickmarks give the size in arcsec. Black contour lines outline low intensity areas, red contour lines high intensity. There is a close correspondence in the velocity map: the same contours outline there predominantly downflows (dark) and upflows (bright), respectively. The **lower panel** gives an area around a pore as observed with the TESOS instrument at the VTT in October 2005. Due to its better spatial resolution the granular velocities and their relation to intensity are even better visible.

The random motion and pattern of the granulation, their splitting and merging, is assumed to generate oscillations and waves in the solar atmosphere, whose absorption may be driver for the chromospheric temperature increase. Because the shape of the magnetic field in the photosphere is important in the context of this work, it will be discussed separately in Sect. 1.2 below.

1.1.2 The chromosphere

The layer above the photosphere is called the chromosphere (gr. $\chi\rho o\mu o\sigma$ = color), because it appears as a red ring around the Sun in solar eclipses due to the red hydrogen Balmer-α line at 656.3 nm. This spectral line is usually called H_α in spectroscopy, and is widely used for the study of the chromosphere since the early days of solar observations (already Hale (1908) used it). The chromosphere extends for some 1000 km, but its boundaries are not well defined. In the chromosphere the temperature increases again, from a minimum value around 4000 K above the photosphere to 25000 K in the upper layers.

Observations of the chromosphere in H_α show a lot of large-scale structure (cf. Fig. 1.4). Long loop-like structures, the so-called *filaments*, can be seen in emission above the solar limb and in absorption on the disk. The

[2]NSO/Kitt Peak data used here are produced cooperatively by NSF/NOAO, NASA/GSFC, and NOAA/SEL.

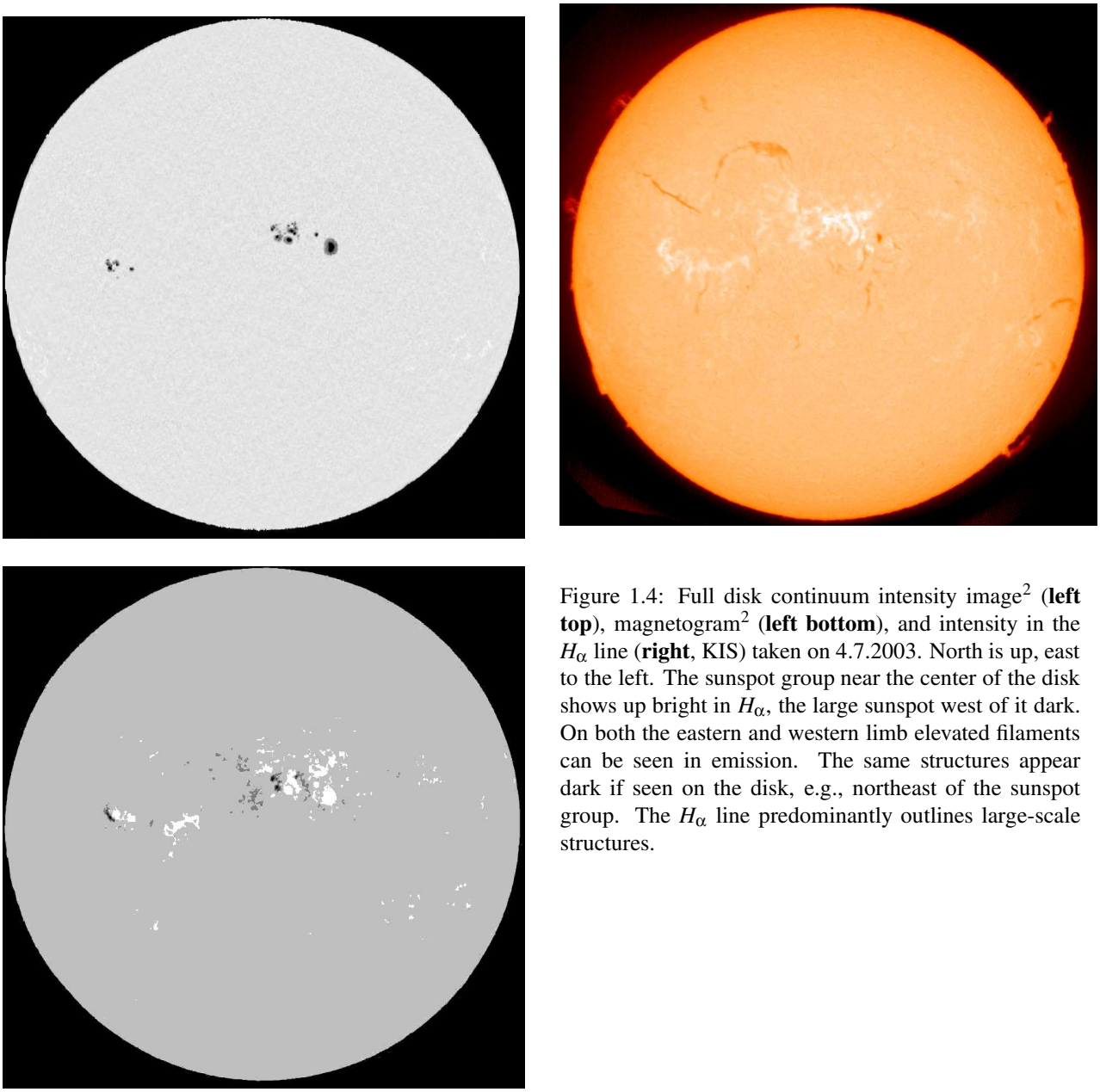


Figure 1.4: Full disk continuum intensity image[2] (**left top**), magnetogram[2] (**left bottom**), and intensity in the H_α line (**right**, KIS) taken on 4.7.2003. North is up, east to the left. The sunspot group near the center of the disk shows up bright in H_α, the large sunspot west of it dark. On both the eastern and western limb elevated filaments can be seen in emission. The same structures appear dark if seen on the disk, e.g., northeast of the sunspot group. The H_α line predominantly outlines large-scale structures.

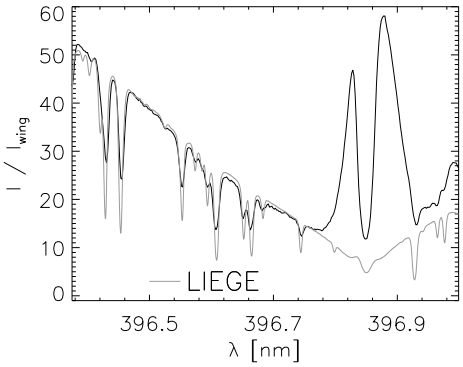


Figure 1.5: Intensity spectrum of the blue wing and the line core of the Ca II H line at 396.849 nm. The thin grey line is the average intensity from the Liege spectral atlas (Delbouille et al. 1973) in the same wavelength range. The profile shows two very prominent emission peaks close to the line core that indicate a temperature increase at a certain height. However, the profile is really an extreme case; the emission peaks usually do not exceed 20-30 % of the wing intensity.

filaments are found to be usually aligned parallel to neutral lines of the magnetic field, i.e. along the dividing line of fields with opposite polarity like between the leading and following spot of a bi-polar sunspot group. However, continuous observations in H_α or other "chromospheric" lines, i.e. spectral lines which are formed only at chromospheric heights due to the intrinsic properties of the respective transition, do not show the static picture which Fig. 1.4 suggests. The chromosphere has a very dynamical nature with transient brightenings lasting only for minutes. This behavior is presumably connected to the puzzling effect of the temperature increase with height. One of the explanations brought forward for this effect assumes that propagating and later absorbed waves transport the energy upwards.

An easy way of visualizing the increase of temperature is given by the Calcium II H and K lines. These broad spectral lines show distinct emission peaks close to the line core, where the intensity is minimum. This deviation from the shape of an absorption line can only be produced by a localized temperature increase in a certain height. An extreme example of this spectral "inversion" effect is displayed in Fig. 1.5. Thus, the "heating problem" of the solar atmosphere starts already in the chromosphere, but it gets worse in the transition region.

1.1.3 The transition region

The transition region is even less geometrically constrained than the chromosphere; it rather denotes a temperature domain. The chromospheric temperature of 25000 K increases to the 1.000.000 K or more of the corona in a short height range, but the actual location in height depends on the interaction between corona and chromosphere. In the transition region not only the temperature drastically changes, but also a quantity which is governing the dynamics, the ratio between magnetic and thermal pressure, the so-called *plasma*-β. It is given by:

$$\beta = \frac{2\mu_0 p}{B^2}\,, \tag{1.1}$$

where p denotes the hydrostatic pressure, $\mu_0 = 4\pi \cdot 10^7$ VsA^{-1}m^{-1} is the magnetic permeability of the vacuum, and B is the field strength.

$\beta \ll 1$ implies that the geometry of the field lines severely restricts the motions of the plasma, while $\beta \gg 1$ implies that the motion of the gas can push the field lines around and change their geometry, which is the case in the photosphere. The ionized plasma of the solar atmosphere cannot cross magnetic field lines in both cases, but plasma-β defines which part is dominating: the field lines or the flow fields.

The transition region is well accessible for observations due to the fact that with increasing temperature different spectral lines come into play by emission. The drawback of the emission lines in the transition region (and the corona) is their wavelength: the lines belong to (highly) ionized states of atomic species and lie in the near to far ultraviolet. They can only be observed with rocket and satellite missions, like the Solar Ultraviolet Measurements of Emitted Radiation (SUMER) spectrograph onboard the SOlar and Heliospheric Observatory (SOHO) satellite. Following different spectral lines upwards through the transition region it seems that areas with high emission are tracing the magnetic field lines (cf. Fig. 1.6).

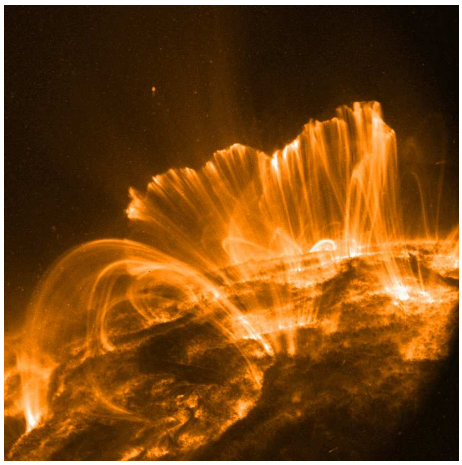


Figure 1.6: Image in a spectral line of the transition region taken with the TRACE satellite. The enhanced emission forms an arcade of loops which presumably follow the magnetic field lines. The footpoints of the loops are located in intense magnetic fields in the photosphere.

1.1.4 The corona

The corona is the final result of the solar "heating problem". Its enormous temperature of 1-2 million Kelvin has been deduced consistently from different approaches, e.g., the width of coronal emission lines or the simple existence of ions like Fe XXIII, i.e. twentytwofold (!) ionized iron. However, the temperature refers to the mean velocities of the particles, electrons or ions. Due to the very low density, the total emission is so small that the corona can only be observed during solar eclipses or with the use of a so-called *coronagraph*, a device that prevents all light from the solar disc to reach the focal plane of a telescope (cf. Stix 2002, p. 143f). The corona extends out to several solar radii. The corona was found to be intimately connected with magnetic fields; it shows a very high (structural) variability with the 11-year solar cycle. The magnetic field in the corona is indirectly visible in the intensity: areas of enhanced intensity and emission are found to form arcades and loops, which presumably trace closed field lines in the corona, whereas regions with open field lines appear darker.

The main importance of the corona for the geo-solar interaction is the solar wind. The Sun continually emits a stream of high velocity (several 100 kms^{-1}) particles that have been found to originate mainly in regions with open field lines. The interaction of the particles with the Earth's magnetosphere produces the *Aurora Borealis*: the particles follow the Earth's magnetic field towards the poles and lead to an emission of light from molecules in the Earth's atmosphere, which have been ionized by their passage. The large amount of particles eventually bombarding the Earth in the case of a *solar storm* can endanger satellites by inflating the outer layers of the Earth's atmosphere through heating, or directly threaten the inhabitants of a space station by a lethal dose of radiation. The solar wind is also the reason for the shape of comet tails, pointing away from the Sun.

However, the good news for further studies is that even with some success in modeling and theory still no generally accepted method has been proposed to deliver all the needed energy from the 6000 K photosphere to an 1.000.000 K corona.

1.2 Magnetic fields in the photosphere

The most prominent recurrent structure in the photosphere are the sunspots, or their smaller version, the pores. Both have been found to be due to magnetic fields. A closer inspection with higher spatial resolution revealed that magnetic flux is also present on smaller scales. The division between small-scale and large-scale refers to the organization of the structure: sunspots and pores are large-scale coherent structures, whereas the other types consist of individual (unrelated) patches of magnetic flux, even if they can cover the same or an even larger area than a sunspot. Only large-scale coherent structures appear darker than their surroundings due to the suppression of convection.

1.2.1 Small-scale magnetic structures

Isolated magnetic elements The same area of the solar surface shown in the upper panel of Fig. 1.3 exhibits a different picture in polarimetric observations (cf. Fig. 1.7). The presence of isolated magnetic flux concentrations

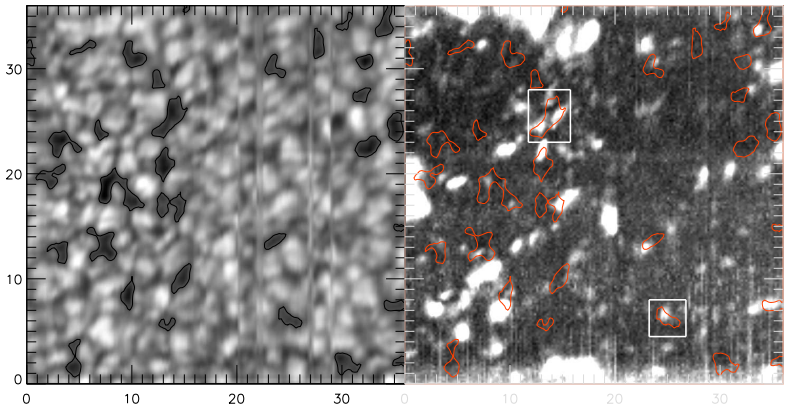


Figure 1.7: Intensity map (**left**) and integrated circular polarization (**right**) of a quiet Sun area close to disc center as observed with POLIS. The tickmarks give the size in arcsec. Black contour lines in the intensity map outline low intensity areas, red contour the same in the polarization map. The two white rectangles highlight the presence of magnetic flux inside a low intensity area given by the red contour.

is indicated by the increased circular polarization signal, which is proportional to the fraction of magnetic flux parallel to the line of sight (*longitudinal Zeeman effect*, cf. Sect. 3.1). In general, the continuum intensity gives no hint where to expect magnetic fields, but usually the field is found in areas of reduced intensity, the intergranular lanes. Due to the high plasma-β, the photospheric field lines are swept along with the gas motions. As the flows are directed towards the intergranular lanes, magnetic flux is accumulating there. The individual magnetic elements are not connected with each other. The magnetic elements are continually pushed around by the granulation pattern; some of the flux concentrations may be broken up again by the convective motions.

The field strength of magnetic flux patches ranges from some 100 G to 1.5 kG; only the stronger concentrations are commonly called *magnetic elements*. The accretion of flux leading to higher field strength is assumed to happen through two main effects, the converging granular flows and the so-called *convective collapse*. This effect has a long history, back to the initial work of Spruit (1979). He found that an already formed concentration of flux with a field strength below 1.3 kG at the surface is prone to a convective instability: if material inside the flux concentration is displaced downwards, this reduces the density, and hence, the gas pressure at its earlier location. The reduction can only be compensated by an increase of magnetic pressure ($\propto B^2$). The radius of the flux tube thus has to decrease, which yields a higher field strength, if the flux itself is conserved. Therefore, the flux concentration shrinks in radius, until a new equilibrium between outer and inner pressure is reached. The collapse is supposed to happen in some minutes only, making it an observational target difficult to catch. However, the results of the collapse, individual magnetic elements with a field strength above 1.3 kG, are observed.

Sect. 4.8.2 discusses a case where at least the signatures of the convective collapse may have been observed: the formation of an intense field concentration accompanied by strong downflows. The convective collapse has also been found to happen in magneto-hydrodynamic simulations of the solar atmosphere (Grossmann-Doerth et al. 1998).

The magnetic network Besides the granulation pattern due to the permanent convective energy transport on small spatial scales, also large-scale convection cells have been found to exist. The *supergranulation* cells, with typical scales around 16000 km and lifetimes of up to one day, lead to a further concentration and spatial structuring of the magnetic flux on the solar surface. Similar to the accumulation of flux at convergence points of granular flows, the large-scale granulation sweeps flux towards its respective boundaries, accreting flux now over a much longer time span. The boundaries of the supergranulation cells in the end are clearly outlined by magnetic flux concentrations (cf. Fig. 1.8). The spatial distribution of the flux on the surface, e.g., from polarimetric measurements or the emission in chromospheric lines, led to the name *chromospheric* or *magnetic network* around the almost field-free cell interior. The network elements have field strengths of up to 1.5 kG and are in general more stable against dissolution than isolated magnetic elements.

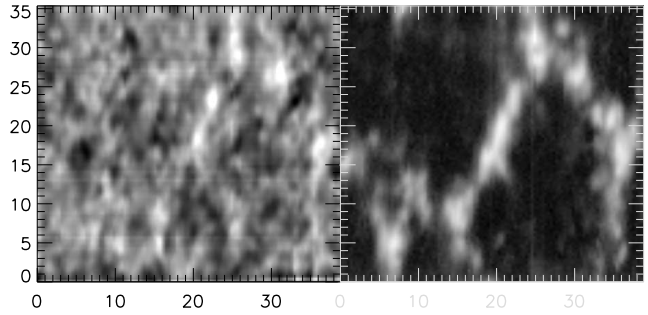


Figure 1.8: An example of magnetic network structure close to the solar limb. (**Left**): Continuum intensity map. (**Right**): Integrated circular polarization signal. Tickmarks are in arcsec. The magnetic flux is concentrated at the boundary of two supergranular cells, with about 10 arcsecs diameter each, in the upper left and lower right half of the image. The cell interior is almost field-free. Individual granules visible in the intensity map have a size around 1-2 arcsec only. The magnetic field appears bright in the continuum intensity.

Plage & Active regions The surroundings of growing or decaying sunspots are usually filled with magnetic flux. Almost the complete area shows polarization signal (cf. Fig. 1.9). However, no traces of this flux show up in the intensity, because it is diffuse and not concentrated enough to inhibit convection. Plage areas can persist for some time, even if the related sunspot or sunspot group has already disappeared. Both pores and sunspots have been found to grow by the accumulation of the diffuse flux in their surroundings.

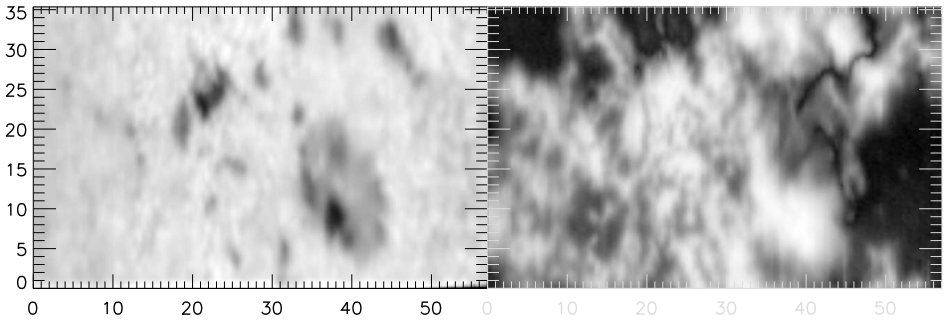


Figure 1.9: An example of *plage* near a decaying sunspot. (**Left**): Continuum intensity map. (**Right**): Integrated circular polarization signal. Tickmarks are in arcsec. Almost the whole area is covered with magnetic flux that shows no signature in intensity.

1.2.2 Large-scale magnetic structures: Pores and Sunspots

There are two types of magnetic structures in the solar photosphere that are coherent on larger scales and are supposedly not composed of individual elements: pores and sunspots. They commonly are distinguished by the fact that a pore consists only of a dark central part, the *umbra*, whereas a sunspot shows an umbra surrounded by a *penumbra*, magnetic flux outside of the umbral core with an intermediate intensity between the umbra and the granular surroundings. The intensity is reduced, because the ionized plasma of the solar atmosphere cannot cross the magnetic field lines, i.e. no overturning convection can form inside strong magnetic fields.

The detection of sunspots was claimed in Europe for the first time in the 17th century (e.g. Galilei et al. 1613; Galilei 1632), but they have been reported in China already 1700 years before (cf. Wittmann & Xu 1987). Hale (1908) was the first to prove that they are due to strong magnetic fields. A longer excursion on the structure of sunspots as found from studies in the last decades will be given here, whereas in Chapter 5 only the most recent works will be presented for comparison.

1.2.2.1 Pores

Pores are the smaller version of sunspots, which share some of their properties, but also differ markedly in others. Fig. 1.10 displays some examples of isolated and clustered pores. The term is attributed to magnetic flux concentrations of larger scale (2000 to 10000 km), which appear much darker than their surroundings due to the

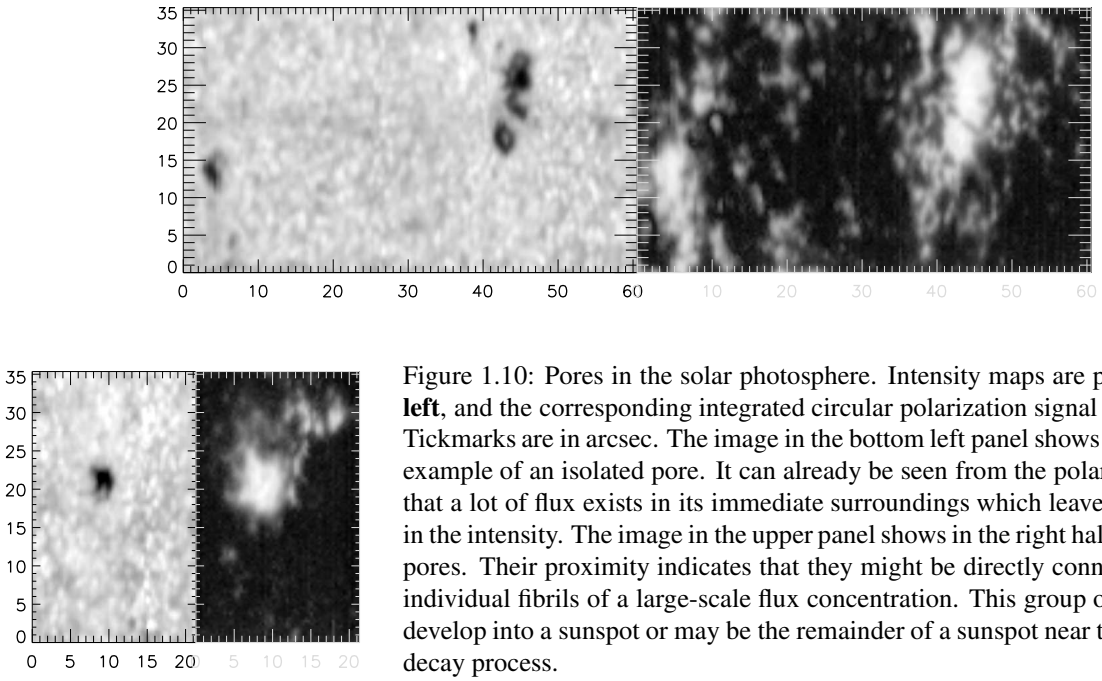


Figure 1.10: Pores in the solar photosphere. Intensity maps are plotted at the **left**, and the corresponding integrated circular polarization signal at the **right**. Tickmarks are in arcsec. The image in the bottom left panel shows the classical example of an isolated pore. It can already be seen from the polarization map that a lot of flux exists in its immediate surroundings which leaves no imprint in the intensity. The image in the upper panel shows in the right half a cluster of pores. Their proximity indicates that they might be directly connected, being individual fibrils of a large-scale flux concentration. This group of pores may develop into a sunspot or may be the remainder of a sunspot near the end of its decay process.

suppression of convection by the magnetic field. Pores have been found to form by the accumulation of flux swept towards them by converging horizontal flows (e.g. Leka & Skumanich 1998; Hirzberger 2003). Their field strength ranges from 600 to 2000 G (Sütterlin et al. 1996; Leka & Skumanich 1998).

After a pore has formed, it is surrounded by a ring of downflows, i.e. all along the boundary between the dark pore and the bright granulation the gas is streaming downwards (e.g. Keil et al. 1999; Tritschler et al. 2003; Sankarasubramanian & Rimmele 2003). The flow has been found to be organized into thin needle-like structures, with the pin pointing towards the pore (Sankarasubramanian & Rimmele 2003) and to be time-variable (Hirzberger 2003). The numerical simulations of Knoelker & Schuessler (1988) or Steiner et al. (1998), in which the interaction of magnetic elements with granulation were modeled, predict the development of persistent downflows at the boundary of flux concentrations. The flows are driven by the radiative cooling of the hot granular material, which can effectively radiate energy away into the less dense evacuated flux concentration. After cooling, the material condenses and starts to submerge.

Pores may develop into full-scale sunspots by forming a penumbra. The evolution presumably is triggered by the inclination of the magnetic field lines to the surface, which influences the ability to maintain overturning convection in the surroundings (cf. Chapter 6). The opposite effect may happen at the end of a sunspots' decay: the spot may loose its penumbra and only a dark umbra similar to a pore is left.

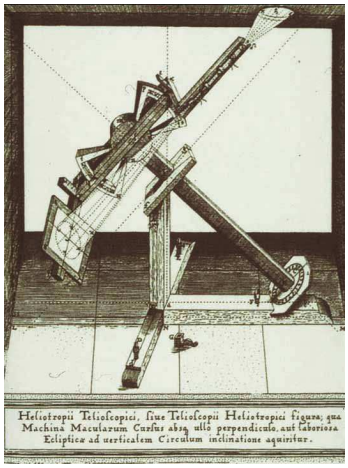


Figure 1.11: Helioscope of C. Scheiner with an equatorial mount (from the "Rosa Ursina", around 1630). The instrument was already designed primarily for the study of the Sun and was superior to Galileos' telescope.

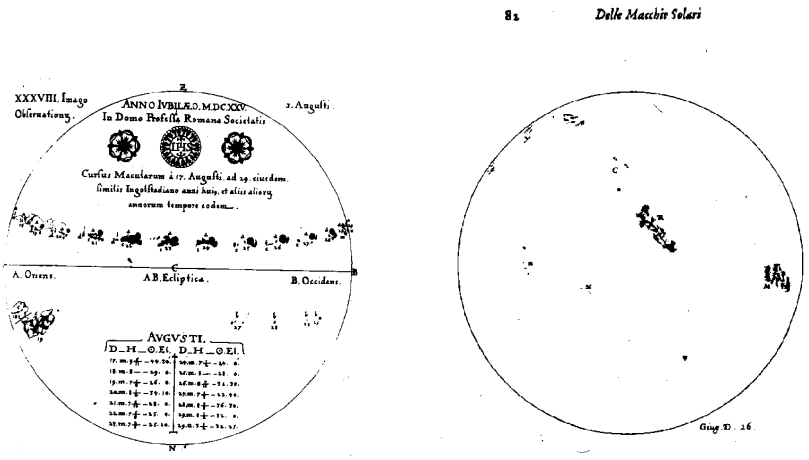


Figure 1.12: Drawing of the solar surface covered with sunspots from Scheiner (**left**, around 1630) and Galileo (**right**, Galilei et al. (1613)). The observations of Scheiner show the motion and evolution of a spot group.

1.2.2.2 Sunspots

Historical excursion Very soon after the invention of the telescope in the 17th century (cf. Fig. 1.11 for an example of an early type solar telescope)[3] the medieval world order – based on the Aristotelian image of "perfect" celestial bodies moving on circles around the Earth – was dealt another blow, besides the already heavy attack through the Copernican system. The first observers of the Sun found the surface marred with dark stains, which could not be traced down to the imagination of the observers or imperfections of their instruments. Even if the question, who first discovered the sunspots – and published the discovery – is still open, two works of the early 17th century describe them: the "*Lettere solari*" of Galilei et al. (1613), and the "*Rosa Ursina*" of Christoph Scheiner (around 1630).

Fig. 1.12 shows two drawings of the solar surface taken from the respective works. The observations are described also in Galileos' main work, "*Dialogue concerning the two chief world systems: The Ptolemaic and the Copernican*": "And on the face of the sun itself, with the aid of the telescope, they have seen produced and dissolved dense and dark matter, appearing much like the clouds upon the earth: and many of these are so vast as to exceed not only the Mediterranean Sea, but all of Africa, with Asia thrown in." (Galilei 1632, translation by Stillman Drake).

Galileo also noticed the formation and decay of the sunspots, which could not come only from a possible revolution of the Sun: "... many of these spots are seen to originate in the middle of the solar disc, and likewise many dissolve and vanish far from the edge of the Sun, a necessary argument that they must be generated and dissolved. For without generation and corruption, they could appear there only by way of local motion, and they all ought to enter and leave by the very edge." (ibidem)

Even recurrent sunspot and sunspots groups and the solar rotation period were known to him: "The truth about the same spots returning is merely what is written in the said Letters; namely, that some of them are occasionally of such long duration that they do not disappear in a single revolution around the Sun, which takes place in less than a month." (ibidem)

However, from this starting point the examination of sunspots was continued for centuries only by noting down their number and their location on the solar disc (cf. Hoyt & Schatten 1998). The closer study of sunspots was prompted by a list of technical and theoretical improvements much later. After the discovery of prism and grating spectrographs in the examination of light, F. Wollaston and J. Fraunhofer noted independently of each other "dark gaps" in the solar spectrum in 1802 and 1813, respectively. Their importance was only realized, after G. Kirchhoff and R.W. Bunsen discovered the unique connection between chemical elements and the spectral lines formed by them (1859). Spectroscopy was used in the study of prominences (Young 1874) and in the derivation of the rotation velocity of the Sun (Dunér 1890, 1905) from the displacements of spectral lines near the eastern or western solar limb due to the Doppler effect.

[3]The image and the following two were taken from the on-line description of an exhibition by the ETH Zürich focussed on Galileo Galilei. The main page can be found under http://www.ethbib.ethz.ch/exhibit/galilei/galileo.html.

Hale (1902) was about the first to propose a systematic study of sunspot properties at the Yerkes Observatory by means of photographic plates, including photographs of sunspot spectra. He was motivated by the idea of quantifying visual observations of what he called "widened lines" (ibidem, p. 216), which he assumed to be related to magnetic fields by the effect found by P. Zeeman in 1897. Erroneously he believed the magnetic fields to be produced by whirl flows of electrons around the sunspot, creating a steady current and, hence, an induced magnetic field (Hale 1908). Despite his wrong initial assumption, his publication "*On the probable existence of a magnetic field in sun-spots*" in 1908 was the founding stone of modern sunspot physics. By the usage of a quarter-wave plate for isolating the left- and right-circular polarization signal, Hale was able to demonstrate that the peculiarities of sunspot spectra were due to magnetic fields and the Zeeman effect: he found two oppositely circularly polarized components, displaced in wavelength to the red and to the blue of the line core position in the intensity spectrum. The field strength was derived by comparing the separation of the components in the observations to the spectrum of a spark lamp, which was placed inside the magnetic field of a electromagnet; the current through the magnet was adjusted, until the separation in the lamp's spectrum matched that of the solar observations. The known properties of the electromagnet then allowed for the calculation of the field strength, which Hale gives as around 2900 Gauss (Hale 1908, p. 327).

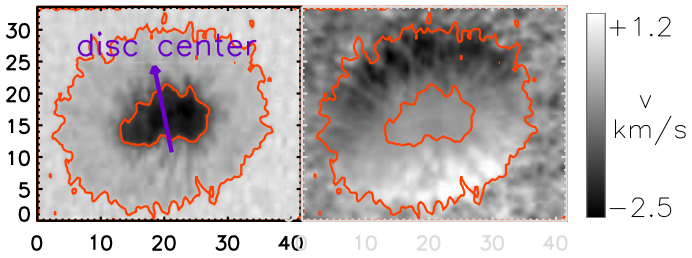


Figure 1.13: The signature of the Evershed effect in the line-of-sight velocity. (**Left**): continuum intensity map. (**Right**): line-of-sight velocity, derived from the line core position of the Ti I spectral line at 630.37 nm. The contour lines outline the boundaries of umbra and penumbra in the intensity image. Tickmarks are in arcsec. Negative velocities are towards the observer.

The Evershed effect Hale's description of the whirl flows, or "vortices", as he called them, as causes for the observed magnetic fields prompted J. Evershed to search for the signature of the motion in spectral lines in 1909. The results of his study unfortunately did not support Hale's prediction. Evershed (1909) found that all sunspots exhibit a flow pattern in the penumbra, which shows maximum positive and negative line-of-sight flow velocities on the symmetry line, which connects spot and disc center, whereas "no displacements are obtained when the slit bisects a spot in a direction at right angles to a line joining the spot and the centre of the Sun's disc" (ibidem, p. 455). This was interpreted as the signature of a radial outflow all around the spot, which is almost horizontal. Such a flow field at once yields the observed properties, if the projection on the line of sight is made. The vertical component of the flow had to be small, because also sunspots close to disc center showed the same pattern. The phenomenon conventionally is named *Evershed flow* or *Evershed effect* after its discoverer.

Evershed had to state that "it is somewhat disappointing, perhaps, that the hypothesis of a radial movement, which is so strongly supported by these observations, seems entirely out of harmony with the splendid discovery of the Zeeman effect in sun-spots, made by Prof. Hale. This seems to demand a vortex, or at any rate a circular movement, in sun-spots; and it was only after a considerable amount of evidence had accumulated that the preconceived conviction that the motion was circular was abandoned" (Evershed 1909, p. 457)[4].

Fig. 1.13 displays that the sunspots under examination here are no exceptions. It shows the line-of-sight velocity derived from the position of the line core of a spectral line for one of the two sunspots analyzed in this work. The symmetry properties and the systematic variation of the flow field can clearly be seen as the black-white pattern. Additionally, the smaller scale fine structure of the spot can be seen as slight variations in velocity, oriented in radial direction.

The studies of Evershed and Hale are the founding stones of sunspot physics. A consistent explanation for their findings is still under debate even almost a century later. In the remainder of this section the structure and morphology of sunspots will be described in no special historical order.

The formation of sunspots Sunspots appear in the activity belt from 30-40 degrees heliographic latitude down to the equator, but they do not pop up as already developed spots. Usually a pore or a group of pores can be seen

[4]Nowadays, one may have a hard time to find the words "harmony" and "splendid" with regard to another author's idea in a publication.

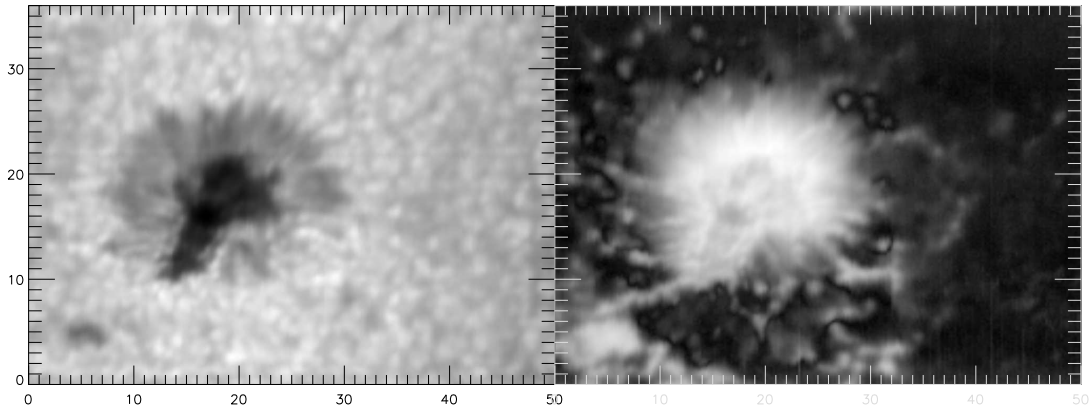


Figure 1.14: Decaying sunspot near the solar limb. (**Left**): Continuum intensity map. (**Right**): Integrated circular polarization signal. Tickmarks are in arcsec. The dark umbral core is surrounded by the brighter penumbra. The spot is not regular; at the lower part no penumbra exists. Instead, in the polarization map one can see a line connecting the spot to the small pore in the lower left corner. The filamentary structure of the penumbra is visible in both intensity and the polarization signal. As for pores, the polarization signal extends over the outer boundary of the spot in the intensity map.

first, surrounded by the more diffuse magnetic flux of an active region (cf. Figs. 1.9 and 1.10). The diffuse and concentrated magnetic flux is merging together and the field strength rises, where different mechanism may play a role: a large-scale convective collapse, draining of the magnetic flux rope from its enclosed material after rising above the surface, or a recoalescence driven by the fact that all the flux IS connected to a single structure below the surface (Solanki 2003). It is assumed that the development of a penumbra is triggered, if the accumulated flux amount reaches a critical value. The concentration of the field to the trigger value takes around 1 day or less. Interestingly, Leka & Skumanich (1998) found that the penumbra, which starts to get visible sector by sector around a pore, is already at once fully developed, i.e. it has the same general properties as the penumbra of a mature sunspot (e.g., in intensity or field inclination) and shows the Evershed effect from the very beginning (cf. Chapter 6). Mature sunspots can have sizes between some 1000 km and up to 60000 km.

Usually sunspots appear as part of bi-polar groups, i.e. one spot shows field lines oriented into the surface, while the other has the opposite orientation. The bi-polar spots and the accompanying active region are always more or less oriented in the East-West direction, reflecting the orientation of the large-scale flux ropes, whose cross-section with the solar surface produces the magnetic activity (cf. the beginning of Sect. 1.1). The more western spot is labeled the leader or leading spot, the more eastern spot the follower. Leader and follower spot show slightly different properties: the follower spot usually is less symmetric and tends to separate into a group of pores or sunspot fragments. The leader spot is more stable and decays more slowly.

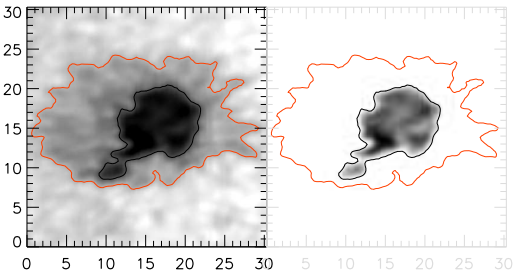


Figure 1.15: Umbral fine structure. (**Left**): Map of the continuum intensity of the whole spot. (**Right**): Same restricted to the umbra to enhance the contrast. The red and black contour lines outline the outer, respectively, the inner penumbral boundary. Tickmarks are in arcsec. Dark umbral cores are seen to be surrounded by slightly brighter areas.

The umbra The umbra, the dark central part, shows temperatures down to 4000 K, 2000 K less than the granulation in the surroundings. The peak magnetic field strength of sunspots is between 2000 G and up to 3700 G in the darkest parts of the umbra. The value seems to depend on the spot size and the state of the solar cycle (Livingston

2002). Even if the umbral intensity seems to be very uniform in Figs. 1.14 and 1.16, this is only due to the high contrast between umbra and granulation. Fig. 1.15 displays that the umbra itself has a substructure, consisting of very dark umbral cores and slightly brighter areas.

Another umbral peculiarity is not seen clearly in this map due to the spatial resolution: The so-called *umbral dots* (UDs). Observations with higher spatial resolution show individual bright points in the umbra, sometimes also inside the dark umbral cores visible in Fig. 1.15. UDs have been found to have a 10 % smaller field strength compared to their surroundings by a number of authors (e.g. Schmidt & Balthasar 1994, and the list of references in Solanki (2003), p. 186) and show a reduced magnetic filling fraction. A possible explanation of the origin of UDs has already been given by Parker (1979): if the sunspot does not consist of a single flux tube below the surface, but of a bundle of tubes, hot field-free material may intrude between the tubes. If it pushes its way up it will appear as a tiny bright point with locally reduced field strength and filling fraction. However, because the hot field-free material is not able to proceed all the way through the atmosphere, the strongest effects are only visible in spectral lines forming deep in the photosphere (Degenhardt & Lites 1993). The network-like brighter structure inside the umbra of Fig. 1.15 may consist of chains of unresolved UDs.

The field inclination in the umbra is almost vertical, it reaches inclinations to the local vertical of around 20-30 degrees at the umbral-penumbral boundary (e.g. Bellot Rubio et al. 2004; Borrero et al. 2005, or the present work).

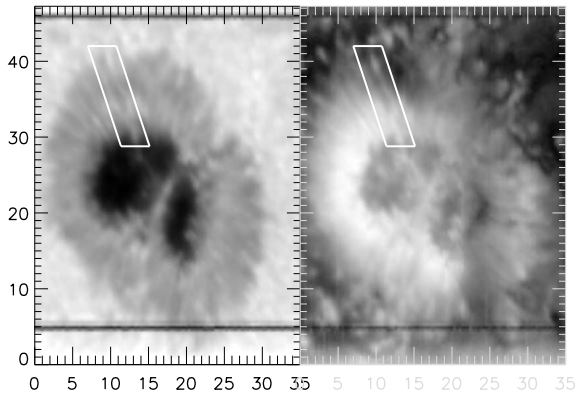


Figure 1.16: Example of a spot with two umbras divided by a light bridge. (**Left**): Continuum intensity map. (**Right**): Integrated circular polarization signal. Tick-marks are in arcsec. The white rectangle outlines a radially oriented penumbral filament. The polarization map reveals that also inside the uniform dark umbra the field has a fine structure.

Light bridges The umbra of spots is sometimes divided into two or more distinct umbras by so-called *light bridges*, as Fig. 1.16 displays. Similar to umbral dots, but more clearly seen in this case, a reduction of field strength is found inside the light bridges (Leka 1997). The magnetic filling fraction was even found to decrease to 0 by Lites et al. (1991). Light bridges are usually present in two cases: during the formation of a spot from the merging of pores and in the final decay phase of the spot, when it starts to split into pores again. Light bridges are most probably formed by the intrusion of field-free material from below and can reach the same intensity as the surrounding granulation.

The morphology of the penumbra The penumbra, the "halo" of magnetic flux that surrounds an umbral core, reaches around 70-90 % of the granular intensity level, implying a partial suppression of energy transport with respect to the granulation and an enhanced energy flux with respect to the umbra. With the improvement of the spatial resolution of observations, the penumbra has been found to be highly organized. The data of Figs. 1.14 and 1.16 with around 0.8 arcsec resolution show clearly the filamentary structure of the penumbra with alternating bright and dark radially aligned *penumbral filaments*. An especially prominent filament is outlined in Fig. 1.16: it is seen to start inside the umbra and continues outwards through the whole penumbra in both the intensity map and the map of circular polarization. It is the connection between a radially aligned brightening in the inner and a darkening in the outer penumbra. This shape is one of the signatures of the flow channels that presumably carry the Evershed flow: a hot upstream in the inner penumbra changes to a horizontal outflow, which cools down by radiation (cf. Schlichenmaier et al. 1998, 1999, 2005).

With increased spatial resolution even more details become visible. Fig. 1.17 displays a sunspot image in the G-Band around 430 nm (cf. Sect. 4.1) taken with the Dutch Open Telescope (DOT) on La Palma. The data has been treated with an image restoration code, the speckle-reconstruction method, to improve the spatial resolution (Rutten et al. 2004). The processed data reaches the diffraction limit of the DOT telescope in times of good seeing.

The minimum angular distance that can be resolved is then $1.22 \cdot \lambda/d$, where d is the telescope diameter of 45 cm, giving around 0.2" ($\sim$ 150 km on the solar surface) for the smallest discernible structures.

In the higher quality image the filamentary structure of the penumbra is clearly seen. The width of the filaments is around 150 to 400 km (Scharmer et al. 2002), their radial extension up to some 1000 km; single filaments may cross the whole penumbra. Additionally, the brightening of the filaments is seen to come from single individual bright points, the so-called *penumbral grains* (cf. the lower inlet of Fig. 1.17), which have been found in high resolution data already in the early 80's (e.g., Muller 1973). They are more prominent on the side of the spot which is closer to the solar limb for observations off disc center (*limb side*). This is connected to a line-of-sight effect discussed in more detail in Sect. 5.9.4.2. The penumbral grains have a slight resemblance to the shape of a comet, an isolated brightening with a tail-like extension.

However, even the well defined filaments and penumbral grains in Fig. 1.17 show a substructure. To the right of the sunspot image, a small subfield is shown in magnification. A filament, starting from a penumbral grain, seems to split up after a short distance into two bright stripes with a darker core between them. The diameter of the bright and dark structures together is the same as that of the penumbral grain. The first report on this was given by Rouppe van der Voort et al. (2004) using data of the new 1-m telescope of the Swedish Solar Tower on La Palma, but after their discovery the dark cores were found to be present also in other older data (Sütterlin et al. 2004). It is assumed that the dark core reflects an intensity variation across a single magnetic structure, the filament, rather than the development of three different structures from a single penumbral grain.

Temporal evolution of the penumbral fine structure Filaments have lifetimes between 10 min and some hours for larger filaments (cf. Solanki 2003, p. 63, and the references therein). The penumbral grains at their end have been found to mainly migrate towards the penumbra by some authors (e.g. Muller 1976; Toenjes & Woehl 1982;

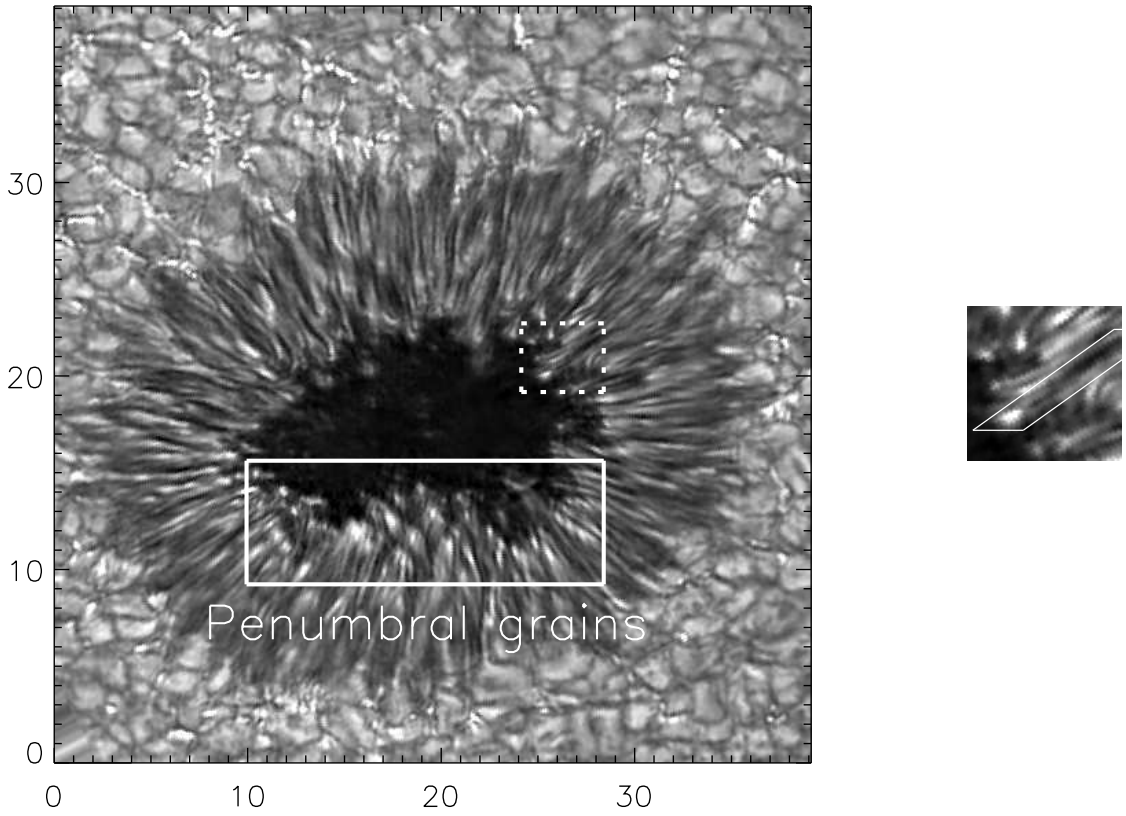


Figure 1.17: Speckle-reconstructed image of a sunspot in the G-band, taken with the Dutch Open Telescope. The direction to solar limb/disc center is the bottom/top of the image. Tickmarks are in arcsec. The lower inlet (solid lines) outlines the area, in which penumbral grains can be seen. The upper inlet (dashed lines) marks a prominent example of a dark core inside a penumbral filament. The same area is shown enlarged at the right.

Sobotka et al. 1995) with a speed of 0.3-1 kms^{-1}. In recent studies, a slightly more complex pattern has been found, where only penumbral grains in the inner penumbra preferentially move inwards, while those in the outer penumbra move outwards (Sobotka & Sütterlin 2001). The lifetime of penumbral grains ranges from 20 min to around 1 hour.

Sunspot decay Sunspots are starting to decay from the moment they are formed (McIntosh 1981). If no information on the amount of magnetic flux contained in the sunspot is available from polarimetric measurements, the spot decay is measured by the change of its area, A, with time, dA/dt. Depending on the type of the spot, i.e. being isolated or part of a sunspot group, recurrent, large or small, being the leader or follower spot of a bi-polar group, slightly different numbers have been found by some authors (cf. the references given in Solanki 2003, p. 7). The main findings were that there seemed to be either a linear ($dA/dt \sim$ const.) or a quadratic ($A(t) \propto f(t^2)$) decrease with time. Both cases imply a different physical meaning: If the decay is due to loss of flux at the perimeter of the spot, whose length in first approximation is given by $\sqrt{A(t)}$, the decay rate is given by $dA/dt \propto -\sqrt{A(t)}$ which leads to a quadratic decay law. If the decay law is linear, the flux loss is happening all throughout the spot with a constant rate. However, the change of the visible sunspot area has to be taken with care. Firstly, if the area decreases it does not necessarily imply that magnetic flux has disappeared; it may have been converted from a concentrated field to a more diffuse one with no signature in intensity. Secondly, one also has to keep in mind the origin of the spot: it is the intersection of a large-scale flux rope with the solar surface. The spot area can change drastically, if the flux rope, which presumably forms an Ω- or U-shaped loop, starts to submerge below or rise above the photosphere. The rapid disappearance of sunspots in less than one day thus most probably is not related to dissolution of the spot, but to the elevation of the flux rope it is formed from.

The surroundings of spots - The magnetic canopy and the moat flow A sunspot usually is well defined in the intensity image, because the penumbral intensity is smaller by 5-15 % than that of the granular surroundings. However, the magnetic field does not abruptly end at the outer boundary. In Figs. 1.10, 1.16, and 1.17 the polarization signal is seen to extent beyond the white light boundary. The magnetic field lines in this so-called *canopy* are almost parallel to the surface. Canopy fields have been found in observations using photospheric and chromospheric spectral lines (cf. the list of references in Solanki 2003). The field strength of the canopy field shows no discontinuity at the spot boundary to the field inside the spot. Thus, the canopy field is interpreted as the continuation of the inclined sunspot field. The field lines get horizontal after reaching a small elevation above the photosphere, because the outside gas pressure decreases exponentially and the magnetic flux can expand freely.

Another peculiarity in the surrounding of the spot is the presence of the so-called *moat flow*. Similar to the Evershed flow inside of a sunspots' penumbra, in the close proximity of a spot all material is found to move radially away from it. The velocity of this large-scale flow is around 0.5-1 kms^{-1}; it can be seen up to twice the spot radius (Brickhouse & Labonte 1988). The velocity is usually derived by two methods, spectroscopy of the surroundings of the spot and following the horizontal motions (*local-correlation tracking*) of prominent and uniquely identifiable features, like the *Moving Magnetic Features* (see next paragraph). The moat flow takes about one day to develop (Pardon et al. 1979a), after a sunspot has formed.

However, the moat flow is not necessarily a continuation of the Evershed flow: a closer inspection of the line core velocity displayed in Fig. 1.13 shows that the Evershed effect ends abruptly at the outer boundary of the sunspot. The presence of the moat flow is indicated there by the difference between the granular velocities in the bottom and top part of the image: in the lower part the average brightness in the velocity map – bright indicates motion away from the observer – is slightly larger than in the top part. The velocities due to the convective motions are much larger, but the effect can still be seen. Interestingly, pores do not show the same behavior, but rather an inflow towards the magnetic flux concentration (cf. the previous section). A possible explanation for the origin of the moat flow is discussed in Chapter 6.

Moving Magnetic Features Close to the spot, patches of magnetic flux are found, which are moving radially away from it, the so-called *Moving Magnetic Features* (MMF, cf. Sheeley 1969). They can been classified mainly into three groups: slow-moving unipolar MMFs with the same polarity as the spot, fast-moving MMFs with opposite polarity, and bi-polar MMFs (Shine et al. 2001). Fast-moving MMFs may show velocities, which are larger than the moat flow. The nature of the different types of MMFs is still not fully clear. Harvey & Harvey (1973) suggested that uni-polar MMFs with the same polarity come from field lines, which are detached from the main sunspot field and then can be carried away by the moat flow. The bi-polar MMFs are more difficult to

explain. Martínez Pillet (2002) argues that they cannot be due to detached field lines, because the total amount of flux they seem to carry away from the spot would be 3-4 times larger than the average flux loss of a sunspot. Sainz Dalda & Martínez Pillet (2005, in press) found bi-polar MMFs to originate already inside the penumbra, usually moving along an individual filament. Yurchyshyn et al. (2001) and Zhang et al. (2003) have drawn the picture of an U-shaped loop of an otherwise horizontal canopy field line to explain the bi-polar MMFs. If the field line dips below the photosphere and returns above it again afterwards, this produces a bi-polar structure, where the magnetic patch closer to the spot has the opposite polarity. Thus, even if the picture is still not fully clear, it seems that only uni-polar MMFs with the same polarity as the spot are related to sunspot decay, while the other types come from the canopy field lines and the field orientation at the outer penumbral boundary.

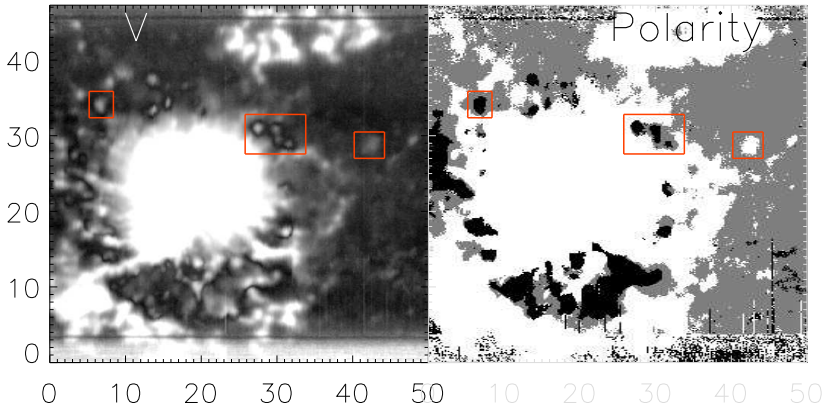


Figure 1.18: The Moving Magnetic Features in the surroundings of a sunspot. (**Left**): Integrated circular polarization. (**Right**): Polarity of the magnetic field. White (+) indicates field lines leaving the solar surface, black (-) ones entering it. The sunspot itself has positive polarity. The red rectangles highlight examples of MMFs with a polarity opposite to that of the spot. A MMF with the same polarity is marked in the right part of the image.

Their appearance in a single snap-shot of a sunspot is displayed in Fig. 1.18: The MMFs are isolated patches of magnetic flux; their polarity can be opposite or identical to that of the sunspot.

1.3 Motivation of this work: The space weather prediction for next week

"*As terrestrial magnetic storms are usually found to be associated with sun-spot groups in which eruptions are numerous, it seems probable that such eruptions may contribute largely to the stream of electrons which pass from the Sun's atmosphere to that of the earth.*" (Hale, 1908)

The geo-solar interaction has many interesting aspects. Nowadays, with a permanently manned space station and an ever increasing number of satellites in orbit, the so-called *space weather* is quite important, because it has economic consequences. Satellites are threatened directly by the immense radiation and particle bombardment in solar storms, i.e. rapidly and strongly increased solar activity in short-wave radiation (X-rays) and the solar particle stream. Satellites in low orbits are indirectly endangered by heating of the uppermost atmosphere layers of the Earth's atmosphere. The atmosphere expands slightly leading to an increased friction for the low-orbit satellites, which are slowed down and loose some height. The loss of height is critical, because the lower the orbit, the stronger the friction gets, a vicious circle. Thus, to protect satellites from damage, the status of the solar activity and the properties of the solar wind near the Earth have to monitored. However, monitoring may not be enough: if one detects that a solar storm has happened and satellites loose height, it may be already too late to counteract it by lifting them to a higher orbit or to turn down all electronic equipment, which may get easily damaged by particle hits.

Thus, it will be necessary to have a prediction of the space weather in advance. This goal was already proposed by Kiepenheuer during the 2nd World War, to be able to predict the disruption of the long-distance radio communication that relied on multiple reflections in the ionosphere of the Earth. Three major problems hamper the prediction of solar storms: firstly, no definite theory has been brought forth to explain, what is happening during a solar flare,

the explosion-like energy release on the Sun that leads to a solar storm. Secondly, on very few instances flares were directly observed with instruments suited to derive their structure, when they happen. And finally – the most critical issue – the solar area, where the storm originates, is usually not observed BEFORE the storm happens.

Thus, few things are generally accepted without questions:

- Solar flares lead to an intense emission in wavelengths from the near UV up to hard X-ray radiation and accelerate particles.
- They are definitely related to the magnetic field geometry from the photosphere up to the corona; flares only appear at places where magnetic flux is present (e.g. Dodson & Owren 1952; Sakurai et al. 1992; Hanaoka 1997).
- In some flares, material is ejected out of the solar atmosphere, the so-called *Coronal Mass Ejections* (CMEs). CMEs that hit the Earth are the reason for solar storms in the magnetosphere of the Earth.
- The flare events can last some hours, but the major energy release of around 10^{30-32} ergs (McLean & Dulk 1978) takes part in some minutes only, in the shape of single short bursts of some seconds (Melrose 1987).

The energy release from a small volume of the solar atmosphere is quite large and fast. Thus, it is assumed that the energy is not transported there during the flare event, but rather that energy stored at these locations is released. Presumably the energy is contained in the magnetic field geometry and the release due to re-configuration of the field lines or magnetic cancellation.

Some authors expect flares to be triggered by the interaction between two large-scale coronal loops (e.g. Falewicz & Rudawy 1999; Linton et al. 2002), whereas Melrose (1987) suggest a scenario of a single loop that is twisted by the motion of its footpoints in the photosphere. Regardless of the physical nature of the energy storage and release, for the prediction it is a prerequisite to know the magnetic field topology before, during and after a flare.

Large-scale loops of magnetic field in the corona are connected to the large-scale structures in the photosphere, sunspots or active regions. As the configuration of the magnetic field in the corona is not directly accessible to observations, the coronal magnetic field structure has to be found by extrapolation of the photospheric fields. Thus, if the 3-D topology of the magnetic fields in the photosphere before and after a flare is known, two things can be studied: firstly, how the field geometry has changed during the flare – yielding information on the energy storage and release processes – and secondly, which configuration of the field is prone to lead to flaring activity. If (all) the pre-flare configurations of magnetic fields are known, a prediction of the space weather is possible. Thus, the 3-D topology of magnetic fields in the photosphere is the basic ingredient for the derivation of the field geometry higher up by extrapolation; and presumably the structure of the photosphere will in the long run allow to predict the probability of flaring activity in the coronal layers.

Chapter 2

Polarimetry at the German Vacuum Tower Telescope

The properties of solar magnetic fields are derived by remote sensing from spectral lines affected by the *Zeeman effect*. Vector spectro-polarimeters are needed to derive the field vector, given by the field strength and the field orientation. These instruments measure the *Stokes vector* (cf. the following section), which describes the wavelength dependent polarization state of light. The German Vacuum Tower Telescope (VTT) and the instruments used in the observations are described in Sects. 2.2 and 2.3. The polarization state of the light is changed by the telescope and the instrument optics. Therefore, a data reduction procedure is necessary to compensate these effects (cf. Sect. 2.4).

2.1 The description of polarized light by the Stokes formalism

The Stokes formalism was formulated around 1850 by Sir G. E. Stokes. The derivation given here follows Collett (1992), Chapters 2-4. The starting point is the wave character of light. Stokes assumed the electric and magnetic field vectors to be solutions of a wave equation taken from mechanics, namely:

$$\Delta \mathbf{u}(\mathbf{r},\mathrm{t}) = \frac{1}{\nu^2}\,\frac{\mathrm{d}^2}{\mathrm{dt}^2}\,\mathbf{u}(\mathbf{r},\mathrm{t})\;, \tag{2.1}$$

where $\Delta = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}$, ν is the oscillation frequency, **r** = (x,y,z), and the components u_x, u_y and u_z are three displacements dependent on place and time.

The interference experiments of Fresnel and Arago (1816) showed that light can only perform *transversal* oscillations (Collett 1992, Chapter 12), which can also be derived from Maxwell's laws. If the z-axis always is parallel to the direction of the light propagation, only two field components are left. The initial orientation of the x-axis (or y-axis) is arbitrary; however, in all later calculations the orientation has to be either kept or taken into account, when the reference frame is changed.

A simple solution of Eq. (2.1) can be given by

$$\begin{aligned} E_x(z,t) &= E_{x,0}\cdot\cos\left(\,\omega t - kz + \delta_x\,\right) \\ E_y(z,t) &= E_{y,0}\cdot\cos\left(\,\omega t - kz + \delta_y\,\right)\,, \end{aligned} \tag{2.2}$$

where the displacement **u** has been substituted by the electrical field **E**. The longitudinal component, E_z, can be left out. The solution is characterized by the values of the angular velocity, $\omega = 2\pi\cdot\nu$, the wave number, $k = 2\pi/\lambda$, the field amplitudes, $E_{i,0}$, and an arbitrary phase, δ_i.

After squaring the equations in (2.2), the equation of the polarization ellipse can be derived:

$$\frac{E_x^2}{E_{x,0}^2} + \frac{E_y^2}{E_{y,0}^2} - 2\cdot\frac{E_x}{E_{x,0}}\cdot\frac{E_y}{E_{y,0}}\cdot\cos\delta = \sin^2\delta\,, \tag{2.3}$$

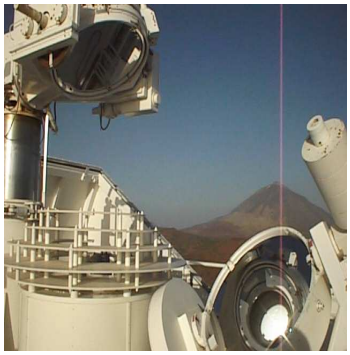


Figure 2.1: The building of the German Vacuum Tower Telescope in Izana, Tenerife (**left**). On the uppermost of the eleven floors the two-mirror coelostat system is located which feeds the light vertically downwards to the main mirror (**right**). The mountain in the background is the Pico del Teide, the highest mountain of Spain.

with the phase difference, $\delta = \delta_x - \delta_y$. The argument (z,t) has been omitted.

Eq. (2.3) gives the location of points (E_x, E_y), which are taken during one oscillation, forming a general ellipse. The specific shape depends on the amplitudes $E_{x,0}$ and $E_{y,0}$, as well as on the relative phase δ. These three parameters encode the polarization state of the light beam. For example, $\delta = \frac{\pi}{2}$ and $E_{x,0} = E_{y,0}$ leads to the equation of a circle, corresponding to circularly polarized light.

With the polarization ellipse of the instantaneous field vector it is only possible to describe fully polarized light. Furthermore, it cannot be used as measurement reading, since in the visible range the oscillation frequency amounts to about 10^{15} Hz, far beyond the time resolution of any instrument. The special merit of Stokes' work was the transformation of all information needed to characterize a polarization state into measurable quantities.

Without the ability to resolve the oscillation, it seems natural to imitate the measurement process possible, i.e. to average over a great number of oscillations. The time average is given by

$$\langle\, E_i(t), E_j(t)\,\rangle = \lim_{T\to 0} \frac{1}{T} \int_0^T E_i(t) E_j(t)\, dt \quad . \tag{2.4}$$

If one now substitutes in Eq. (2.3) the time dependent values ($E_x = E_x(z,t)$) with their time averages, one obtains after a short computation[1] the equation

$$\left(E_{x,0}^2 + E_{y,0}^2\right)^2 = \left(E_{x,0}^2 - E_{y,0}^2\right)^2 + \left(2 \cdot E_{x,0} \cdot E_{y,0} \cdot \cos\delta\right)^2 + \left(2 \cdot E_{x,0} \cdot E_{y,0} \cdot \sin\delta\right)^2 \quad . \tag{2.5}$$

With the terms in Eq. (2.5) the following definition of the Stokes parameters S_i, usually arranged in a vector[2], $\mathbf{S}$, for convenience, can be made:

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} E_{x,0}^2 + E_{y,0}^2 \\ E_{x,0}^2 - E_{y,0}^2 \\ 2 \cdot E_{x,0} \cdot E_{y,0} \cdot \cos\delta \\ 2 \cdot E_{x,0} \cdot E_{y,0} \cdot \sin\delta \end{pmatrix} \quad . \tag{2.6}$$

With different values for the phase difference ($\delta = 0,\ \pm\pi/2,\ \pi$) and the amplitudes $(E_{x,0}, E_{y,0})$ it can be seen that the Stokes parameters encompass the information on the polarization state of the light in the following way:

- S_0 is the total intensity of both polarized and unpolarized light.
- S_1 is the difference of the intensities of *linearly* polarized light at the angles 0° and 90°.
- S_2 is the difference of the intensities of *linearly* polarized light at the angles $+45^\circ$ and -45°.
- S_3 is the difference of the intensity of *right* and *left circularly* polarized light.

In contrast to the polarization ellipse, unpolarized as well as partially polarized light can be described. Note that "unpolarized" is only defined in a statistical sense (cf. the time average in Eq. (2.4)); it implies that all possible polarization states occur with the same relative frequency. Each individual photon is 100 % polarized.

[1] One can assume z=0. The calculation can be found in Collett, Collett (1992), p. 36.

[2] Even though addition and a norm are defined, the Stokes vectors form no vector space in a strict mathematical sense.

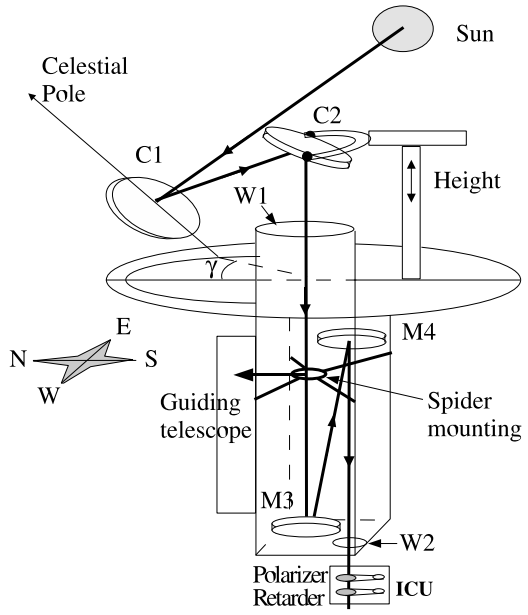


Figure 2.2: Beam geometry at the VTT; the drawing is not to scale. The coelostat consists of two flat aluminum coated mirrors (C1 & C2) which feed the light beam into the telescope. The telescope tube that contains the imaging main mirror (M3) and a folding flat mirror (M4) is evacuated. The light beam passes the entrance (W1) and exit window (W2) of the evacuated tube. Right behind the exit window the Instrumental Calibration Unit (ICU) is located. The spider mounting deflects a small part of the beam to the guiding telescope which allows for positioning on the solar disc and some image stabilization.

The requirements on a physical reasonable vector are:

$$S_0^2 \geq S_1^2 + S_2^2 + S_3^2 \,, \tag{2.7}$$

i.e. the polarization degree $P = \frac{\sqrt{S_1^2+S_2^2+S_3^2}}{S_0}$ cannot exceed 100 %, and the intensity S_0 must be larger than zero. A commonly used notation for the Stokes vector is $\mathbf{S} = (I,Q,U,V)$, instead of using the indices.

A change of the polarization state of the light by an optical element corresponds to a linear transformation of the Stokes vector inside the Stokes "vector space". It is described by the multiplication with a so-called *Mueller matrix*, $\mathcal{M}$: $\mathrm{S_{out}} = \mathcal{M} \cdot \mathrm{S_{in}}$. Mueller matrices are a sub-group of all 4×4-matrices, because they must never lead to a violation of the restrictions on the Stokes vector (negative intensity and overpolarization). For the description of the instrument and the telescope, the Mueller matrices for rotations, polarizers, waveplates, and reflections on mirrors will be used; their definition is given in Appendix A.

2.2 The German Vacuum Tower Telescope

The German Vacuum Tower Telescope (VTT, cf. Fig. 2.1) in Izana on Tenerife uses a coelostat to feed the light into the telescope (cf. Stix 2002). A sketch of the light beam geometry is displayed in Fig. 2.2. The coelostat consists of two flat mirrors (C1 & C2), which feed the beam into the telescope. The second mirror, C2, is mounted on a pillar. Its height depends on the Sun's declination and on the position of C1. C1 can be moved on a circle around C2, e.g., to avoid the shadow of C2 falling on C1, while its axis is parallel to the Earth's rotation axis. The displacement of C1 from the North-South direction is measured by the angle γ, which by convention is positive towards east. C1 is rotated to follow the Sun's track on the sky, i.e., only the angle of incidence on C1, i_{C1}, is time-dependent for a given position of C1 and C2. The small changes in solar declination during one day produce negligible variations of the angle of incidence on C2, i_{C2}.

The imaging main mirror, M3, and a folding flat mirror, M4, are contained in a vacuum tube with an entrance and an exit window made of BK7 glass. Just below the exit window, the Instrument Calibration Unit (ICU) for both spectro-polarimeters at the VTT, the Tenerife Infrared Polarimeter (TIP) and the POlarimetric LIttrow Spectrograph (POLIS), is installed to calibrate the polarization properties of the remaining light path: the optics of the Correlation Tracker (CT) system[3] and the instrument itself. The primary focus of the telescope is close to the 1st floor of the building, where the POLIS instrument is located. The TIP is integrated into the main spectrograph that extends over four floors below the ground.

[3]Since beginning of 2004, the CT is replaced by the Kiepenheuer Adaptive Optics System (KAOS, von der Lühe et al. 2003).

2.3 Post-focus instrumentation

The main ingredients of most vector-polarimeters, i.e. instruments which measure the whole Stokes vector (I,Q,U,V), are a modulator package and a specific method to convert the modulated polarization state into an intensity modulation. The modulator produces a time-variable Stokes vector, whose components and variation depend on the modulator properties and on the polarization state in front of it. The conversion to an intensity modulation is necessary, because no polarization-sensitive detectors exist. The conversion to intensity is usually achieved by placing a linear polarizer into the beam: only light parallel to its optical axis is transmitted, and hence, a variation of the polarization state leads to a variation of the transmitted intensity.

2.3.1 The Polarimetric Littrow Spectrograph

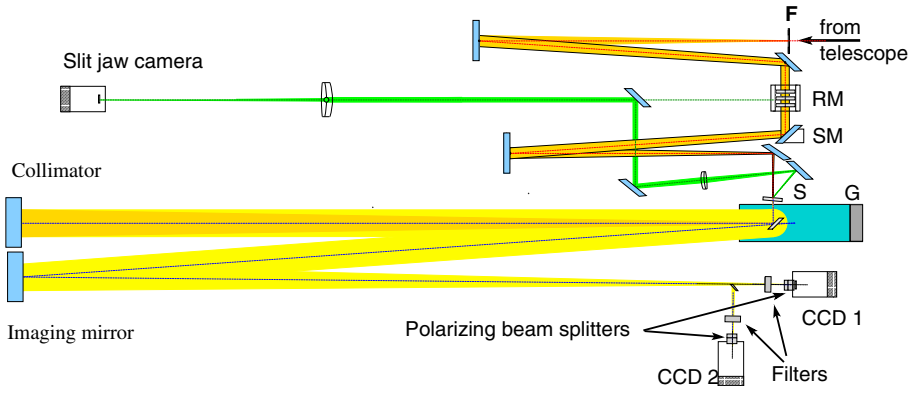


Figure 2.3: Optical layout of POLIS. Light enters the instrument at top right. F denotes the telescope focal plane. A collimator mirror forms an image of the entrance pupil near the rotating modulator, RM, and the scan mirror, SM. An imaging mirror re-images the solar image at F to the entrance slit, S, of the spectrograph. Slit jaw images are recorded from the reflective slit jaws of the entrance slit. The collimator re-images the entrance pupil to the grating, G. The dispersed beams are then re-imaged to the focal plane detectors, CCD 1 and CCD 2. A small pickoff mirror close the focal plane reflects the blue beam to CCD 2. Diffraction orders are isolated with narrow-band filters placed in front of each detector. Polarizing beam splitters placed directly in front of the CCDs separate the orthogonally polarized beams.

The POlarimetric LIttrow Spectrograph (POLIS) measures the Stokes vector (I,Q,U,V) simultaneously in two wavelength regimes on two separate CCD cameras. One CCD records the polarization state in a spectral range around 630 nm (in the following called *red channel*). The second camera records the line core and the blue wing of the Ca II H line at 396.8 nm (*blue channel*). The red and the blue branches of the instrument are dual beam systems with polarizing beam splitters (BS) in front of each CCD camera. The optical layout of POLIS is displayed in Fig. 2.3. The instrument consists of three major parts:

1. The scan unit, with the modulator and the scan mirror, which can be used to select area of the solar surface to be observed.

2. The slit jaw camera, which takes broad band images of the reflective front side of the slit.

3. The main disperser, with slit, grating, polarizing beam splitters, narrow-band pre-filters for order selection and the CCDs for the two channels.

The optical layout of the main disperser resembles that of a Littrow spectrograph: the light beams are in a plane perpendicular to the plane of dispersion. This allows for very small reflection angles at the collimating and imaging mirrors.

POLIS uses a single rotating retarder as polarization modulator that modulates the polarization signals in both wavelength bands with high efficiency. The modulator is a zero-order waveplate. Polarizing beam splitters in front of the cameras split the modulated light beam in two orthogonally polarized beams $I^{\pm}$, imaged on a single CCD in

each channel. The intensity of the individual beams $I^{\pm}$ as a function of the incoming Stokes vector, the modulator angular position, θ, and the retardance of the modulator, δ, is given by:

$$\begin{aligned} I^{\pm}(\theta,\delta) &= \frac{r^{\pm}}{2}\{I \pm Q\cdot(\cos^2 2\theta + \sin^2 2\theta\cdot\cos\delta) \\ &\pm U\cdot\sin 2\theta\cdot\cos 2\theta\cdot(1-\cos\delta) \mp V\cdot\sin 2\theta\cdot\sin\delta\}, \end{aligned} \tag{2.8}$$

where $r^{\pm}$ is the transmission of the respective beam. The ratio r^{+}/r^{-} is about 0.92, presumably due to the different path length in the polarizing beamsplitters (Schmidt et al. 2001; Beck 2002). Eq. (2.8) is an idealized equation, which is valid under the assumption that the optics between modulator and beamsplitter do not change the polarization state. The dual beam design decreases the amount of seeing induced cross-talk and increases the detection efficiency (Lites et al. 1993). In order to derive the values of Stokes I, Q, U, and V, the intensity signal is demodulated in real-time after 8 images, taken at the modulator positions $\theta = k \cdot 22.5^{\circ}$ (k=0...7), using a suitable summation scheme (Lites et al. 1993; Beck 2002). The instrument is located in the first floor of the VTT. More technical details on the instrument are given in Beck (2002); Beck et al. (2005b).

2.3.2 The Tenerife Infrared Polarimeter

The Tenerife Infrared Polarimeter (TIP) is integrated in the main spectrograph of the VTT (Martínez Pillet et al. 1999). It can be used inside a broad wavelength range in the near-infrared between 0.8 and 2 μm. The modulation of the incoming light is achieved with two consecutive tunable ferro-electric Liquid Crystal Retarders (LCRs); the orientation of their optical axis is a function of the applied voltage. The modulation unit is located just below the entrance slit of the spectrograph, directly followed by polarizing beamsplitters. The two beams proceed through the optics of the spectrograph, before they are imaged on a single CCD camera. The CCD camera is cooled down to around 77 K with liquid nitrogen, because otherwise the thermal radiation of the camera itself would prevent any observations. The TIP takes four images at different orientations of the two LCRs. The demodulation is part of the instrument calibration procedure.

2.3.3 Setup for the simultaneous usage of POLIS and TIP

The two spectro-polarimeters have been operated simultaneously using a 50-50-beamsplitter[4] in the first floor of the building: half of the light was reflected to POLIS, the other half transmitted to TIP. POLIS was remote-controlled by TIP to ensure simultaneous exposures. The two entrance slits to the respective spectrograph systems were aligned to have the same orientation by rotating the main spectrograph, which is used by TIP. The lateral displacement due to the beamsplitter was compensated with the scan mirror of POLIS.

The scanning of the solar surface, i.e. moving the image across the respective slits, was performed with the *Tip-Tilt* mirror (see below) of the Correlation Tracker (CT) Optics in front of both instruments (Ballesteros et al. 1996). The CT system is used to correct for image motion to stabilize the image. A reference image is taken and updated regularly; the current solar image is compared with the reference to find its displacement. The displacement is compensated with the Tip-Tilt mirror, a mirror which can be tilted in two orthogonal axes (*Tip* and *Tilt*) by applying a voltage to piezo-electric crystals fixed to its mounting. The same mirror can be used for scanning by applying a continuously increasing voltage to one of its axes.

2.4 Calibration of polarimetric data

The calibration of the polarimetric data can be divided into three steps: (i) the *flatfielding*, i.e. removal of persistent intensity patterns due to, e.g., dust particles on the slit. (ii) The polarimeter calibration, where the response of the instrument to different polarization states is determined. (iii) The telescope correction, which compensates the changes of the polarization state due to the telescope optics.

[4] In the meantime an infrared-visible BS is available that splits at a wavelength of 800 nm.

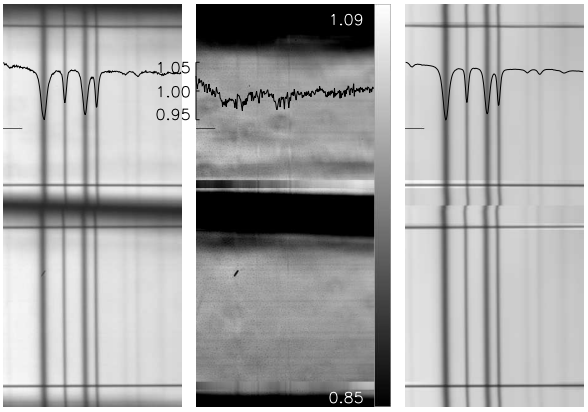


Figure 2.4: Example of the gain table correction for the red channel of POLIS, using data from August 2003. The image size is 288×652 pixel. **Left:** uncorrected image, I^+ and I^- are separated spatially in the vertical direction. **Middle:** Gain table. **Right:** Gain-corrected image. Profiles of each image from the row marked with the short horizontal bar at the left are overplotted.

2.4.1 Flatfield procedure

The so-called *flatfielding* takes the intensity response of individual camera pixels into account. The name indicates the final aim of the procedure, to have a "flat", i.e. constant, intensity all across the camera for a homogeneous illumination. Two main sources lead to intensity variations: the properties of individual camera pixels and dust and dirt in the optical system, e.g., on the slit, mirrors, or the CCD chip itself. The compensation for the intensity response is commonly called the *gain table*, a matrix with the same size as the images taken by the camera. The gain table contains the information on where the intensity observed has to be reduced or increased. In the following a general description of the procedure is given, which is valid for both the POLIS and the TIP instrument.

To derive the gain table, special data sets are taken. To have an almost homogeneous illumination without solar structure to start with, the telescope is (randomly) moved around while taking images. The flatfield images (around 15-50) are averaged, to reduce the visibility of spatial structures on the Sun even more. The observed wavelength is not changed, therefore, the flatfield data contains the same spectral lines as the science data (Fig. 2.4, left panel). The next step toward the construction of the gain table is the removal of the spectral line information from the flatfield images. The profiles, i.e. the rows of the image, are averaged in the spatial direction along the slit, giving a so-called *mean profile*, still containing the spectral lines. Isolated patches of reduced intensity are strongly suppressed in the mean profile due to the large-scale averaging. The gain table for the full image is then obtained by dividing the spectrum in each single row by the mean profile. Only individual camera pixels with reduced or enhanced intensity show up in the gain table, as the shape of the spectrum is the same in both mean and individual profiles. The gain table values are close to 1 (Fig. 2.4, middle panel). The application of the gain table to an observed image removes isolated structures (e.g., the scratch visible in the lower beam in Fig. 2.4), as well as large-scale structures like the intensity reduction at the borders of the CCD chip. The actual applied flatfield procedure is more complicated, at least for the POLIS instrument (cf. Beck et al. 2005b), but the principle is exactly as described above.

Without illumination, a residual intensity is observed due to electron and read-out noise, the *dark current*. The dark current is removed by simply subtracting a mean dark current image from each observation. For the TIP, the dark current correction is much more important, because of the infrared wavelengths observed and the liquid nitrogen cooling. The dark current is much more variable and has to be measured repeatedly with a much shorter cadence than for POLIS.

2.4.2 Polarimeter calibration

The polarimeter calibration determines the response of the instrument to different polarization states. It is based on a simple input-output comparison: the Instrumental Calibration Unit (ICU), located right behind the exit window of the evacuated telescope tube (cf. Fig. 2.2), is used to generate known polarization states (input). The actually measured Stokes vector (output) is compared to it and the deviations between them yield the information on the polarimeter response. Based on the properties of the Stokes "vector space", it is sufficient to generate only four linearly independent Stokes vectors as input to uniquely determine the instrument response to ALL possible Stokes vectors. However, for redundancy more Stokes vectors are generated to improve the accuracy of the determination. The information on the polarimeter response is stored in the so-called *response function*, $\mathbf{X}$, again a 4×4 Mueller matrix. High precision of the calibration requires good knowledge of the polarizing properties of the elements of

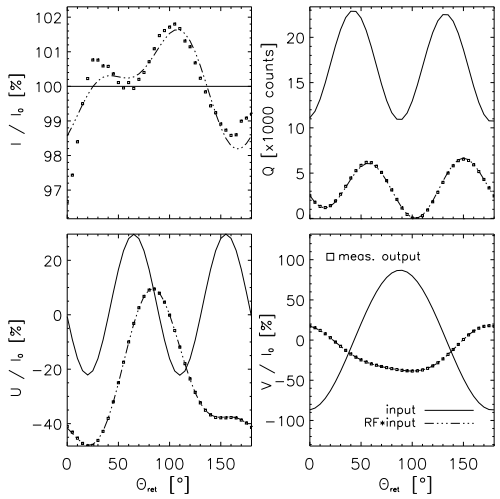


Figure 2.5: Stokes vector (I,Q,U,V) of a calibration measurement in the red channel. **Solid:** calibration input vector, calculated from the retarder and polarizer position angles. **Squares:** measured Stokes vector. **Dash-dotted:** Computed output vector $\mathbf{S}_{\mathrm{c}}^{\mathrm{out}}$ (Eq. (2.12)). Note the good agreement with the measured curve. Stokes Q is given in digital counts instead of the normalized intensity.

the ICU, a linear polarizer and a zero-order waveplate, to be able to calculate the generated input vectors. The following description is strictly valid only for the POLIS instrument. The TIP calibration principle is again the same, but the evaluation slightly differs and will be described at the end of this section.

From the angular orientation of the ICU optical elements the polarization state of the beam that passes it can be calculated. However, the light incident on the ICU is already partially polarized by oblique reflections in the telescope. Thus, the general equation between the solar unpolarized input, $(\mathrm{I}_0,0,0,0)^{\mathrm{T}}$, and measured output during the calibration, $\mathbf{S}^{\mathrm{out}}$, is given by:

$$\mathbf{S}^{\mathrm{out}} = \mathbf{X}\mathbf{M}_{\mathrm{r}}(\theta_{\mathrm{r}})\mathbf{M}_{\mathrm{p}}(\theta_{\mathrm{p}})\cdot\mathcal{T}\cdot\begin{pmatrix}\mathrm{I}_0 & 0 & 0 & 0\end{pmatrix}^{\mathrm{T}} = \mathrm{I}_0\cdot\mathbf{X}\mathbf{M}_{\mathrm{r}}(\theta_{\mathrm{r}})\mathbf{M}_{\mathrm{p}}(\theta_{\mathrm{p}})\cdot\begin{pmatrix}1 & \mathcal{T}_{2,1} & \mathcal{T}_{3,1} & \mathcal{T}_{4,1}\end{pmatrix}^{\mathrm{T}}, \quad (2.9)$$

where θ_{p} and θ_{r} are defined as the angles between terrestrial E-W and the transmission axis of the polarizer and the fast axis of the retarder. $\mathbf{M}_{\mathrm{i}}(\theta_{\mathrm{i}})$ are the Mueller matrices of the rotated optical elements, and $\mathcal{T}$ is the telescope Mueller matrix (see the following section 2.4.3). The total transmission of the telescope, $\mathcal{T}_{1,1}$, has been set to 1, because only relative photometry and polarimetry is done.

If the linear polarizer is kept fixed, e.g., at $\theta_{\mathrm{p}} \equiv 0^{\circ}$, Eq. (2.9) simplifies to:

$$\mathbf{S}^{\mathrm{out}} = \mathrm{I}_0\cdot\frac{1}{2}\left(1+\mathcal{T}_{2,1}\right)\cdot\mathbf{X}\mathbf{M}_{\mathrm{r}}(\theta_{\mathrm{r}})\cdot\begin{pmatrix}1 & 1 & 0 & 0\end{pmatrix}^{\mathrm{T}}. \quad (2.10)$$

The variation of $\mathcal{T}_{2,1}$ during the time needed for a calibration measurement (5-15 min, depending on the exposure time) is small; the term $\mathrm{I}_0\cdot\frac{1}{2}\left(1+\mathcal{T}_{2,1}\right)$ can be exchanged with a constant coefficient. Keeping the polarizer fixed completely decouples the polarimeter calibration from the telescope polarization, because a fixed polarization state with almost constant light level is created behind the polarizer.

A standard calibration data set for POLIS consists of one half rotation of the retarder in 5°-steps (for TIP: full rotation). The calibration input for $\theta_{\mathrm{p}} \equiv 0^{\circ}$ is thus given by:

$$\mathbf{S}_c^{\mathrm{in}} = M_{\mathrm{r}}(0^{\circ},5^{\circ},...,180^{\circ})\cdot\begin{pmatrix}1 & 1 & 0 & 0\end{pmatrix}^{\mathrm{T}}. \quad (2.11)$$

To derive the observed Stokes spectra, $\mathbf{S}_c^{\mathrm{out}}(\lambda)$, the flatfield procedure described in the previous section is applied to the images taken during the calibration. The two beams, I^{+} and I^{-}, are balanced in intensity, i.e. the slightly different global intensity response given by the ratio $\mathrm{r}^{+}/\mathrm{r}^{-}$ (cf. Eq. (2.8)) is compensated for. The two beams are "merged", i.e. added for I and subtracted for QUV (Beck 2002), and the resulting spectra are spatially and spectrally averaged.

The polarimeter response function is derived from the measurements after arranging the Stokes vectors into 37×4 matrices by a solution of the linear problem

$$\mathbf{S}_{\mathbf{c}}^{\mathrm{out}} = \mathbf{S}_{\mathrm{c}}^{\mathrm{in}}\cdot\mathbf{X}^{\mathrm{T}}. \quad (2.12)$$

The solution method by matrix inversion and the error estimate of the procedure can be found in the Appendix of Beck et al. (2005b). A comparison between generated input, measured Stokes vector, and the transformation

of input to output by the response function is displayed in Fig. 2.5. The largest deviations between measured and predicted values are found in Stokes I, which is less constrained than the polarization components due to its smaller variation during the calibration.

The response function is measured for typically 4-8 intervals along the slit, in order to account for a possible variation of the instrument properties in the spatial dimension. A typical result for the red channel of POLIS is:

$$\mathbf{X} = \begin{pmatrix} I\rightarrow I & Q\rightarrow I & U\rightarrow I & V\rightarrow I \\ I\rightarrow Q & Q\rightarrow Q & U\rightarrow Q & V\rightarrow Q \\ I\rightarrow U & Q\rightarrow U & U\rightarrow U & V\rightarrow U \\ I\rightarrow V & Q\rightarrow V & U\rightarrow V & V\rightarrow V \end{pmatrix} = \begin{pmatrix} 1.000 & 0.0504 & 0.0063 & -0.0258 \\ -0.0078 & 0.0993 & -0.3218 & 0.3891 \\ -0.0208 & -0.4448 & -0.2443 & -0.0943 \\ -0.0150 & -0.1833 & 0.2281 & 0.2914 \end{pmatrix} . \quad (2.13)$$

It shows significant cross-talk between Stokes V and Q ($V \rightarrow Q = 0.39$). The matrix entry $Q \rightarrow U$ is even larger, but cross-talk between the linear polarizations, $QU \leftrightarrow QU$, is less critical for several reasons (same shape of polarization signal, smaller amplitudes than in V, no influence on field inclination). For an ideal polarimeter, one would expect a diagonal matrix.

The total efficiency for measuring a Stokes parameter, e_i, is derived from the corresponding column of the response function by

$$e_{\mathrm{i}} = \sqrt{\sum_{\mathrm{j}=1..4} X_{\mathrm{j,i}}^2} , \quad (2.14)$$

which yields for the matrix in Eq. (2.13) a value of $(e_Q, e_U, e_V) = (0.494, 0.464, 0.496)$, i.e. about each 2nd polarized photon is actually counted.

TIP The calibration data set for TIP is very similar to that of POLIS (full rotation of the ICU retarder instead of a half revolution). The difference in the evaluation is that for TIP also the demodulation matrix, i.e. the conversion from the four images at different LCR orientations to the Stokes parameters $IQUV$, has to be found. However, the same input-output comparison as for POLIS is possible: the input Stokes vectors are known and the measured output intensity images are a linear combination of $IQUV$ with unknown coefficients. Again, with only four linearly independent Stokes vectors the system of equations can be uniquely solved.

2.4.3 The telescope model

*"The discovery of the Zeeman effect in sun-spots by Hale made it important to determine the action of the silver-on-glass mirrors of the tower telescope of the Mount Wilson Solar Observatory upon circularly polarized light. It was at once recognized that circularly polarized light would be changed to elliptically polarized by reflection from the silver surface and that its effect would vary with the angles of incidence and hence with the position of the sun." (*C.E. St. John, 1909*)*

The telescope model is described in detail in the diploma thesis (Beck 2002) and Beck et al. (2005a). Only a short summary will be given here.

The telescope (cf. Fig. 2.2) consists of in total 6 optical elements, which can influence the polarization state of the light: four mirrors, and the entrance and exit windows of the evacuated telescope tube. These optical elements are labeled C1 and C2 (the two coelostat mirrors), M3 and M4 (the main and the folding mirror), and W1 and W2 (entrance and exit window). The Mueller matrix of the telescope, $\mathcal{T}$, results from subsequently multiplying individual Mueller matrices to the incoming beam as the beam passes through the telescope. The Stokes vector, $\mathbf{S}_{\mathrm{ICU}}$, at the place of the Instrumental Calibration Unit, i.e. behind the exit window, is given by:

$$\begin{aligned} \mathbf{S}_{\mathrm{ICU}} &= \mathcal{T}\cdot\mathbf{S}_{\mathrm{in}} = \\ &= \mathcal{W}_{\mathrm{W2}}\cdot\mathcal{R}(\theta_4)\cdot\mathcal{M}_{\mathrm{M4}}\cdot\mathcal{M}_{\mathrm{M3}}\cdot\mathcal{W}_{\mathrm{W1}}\cdot\mathcal{R}(\theta_3)\cdot\cdot\mathcal{M}_{\mathrm{C2}}\cdot\mathcal{R}(\theta_2)\cdot\mathcal{M}_{\mathrm{C1}}\cdot\mathcal{R}(\theta_1(t))\cdot\mathbf{S}_{\mathrm{in}} . \end{aligned} \quad (2.15)$$

$\mathcal{R}(\theta_i)$ denotes rotations of the reference frames between subsequent optical elements, where $\mathcal{R}(\theta_1)$ rotates the fixed reference system on the Sun onto the first mirror. The reference frame on the Sun is defined by the great circle through Sun center and the celestial pole (celestial North-South, CNS). $+Q$ is parallel to the tangent to the circle at Sun center.

The Mueller matrix of the telescope depends on (1) the geometry of the beam, i.e. on the angles of inclination on the 4 mirrors or the rotation angles, $\theta_1(t)$, θ_2, θ_3, θ_4, between subsequent optical elements, and (2) on material

properties, i.e. the complex refractive index of the mirrors and the properties of the windows. The effect of the two windows can be included into one matrix, with an effective retardance and an effective angle of the fast optical axis.

The geometry of the beam path can be calculated using a modified version of the Eqs. given in Capitani et al. (1989). The material properties were determined by direct and indirect measurements (Beck et al. 2005a). With these parameters, the telescope Mueller matrix, $\mathcal{T}$, can be calculated at each time. As an example, the telescope matrix for March 21 at 10:00 UT ($\delta_\odot \sim 0^\circ$, $\gamma = 50^\circ$) at 630 nm is

$$\mathcal{T}_{630\mathrm{nm}} = \begin{pmatrix} 1 & 0.0232 & -0.0053 & -0.0007 \\ 0.0238 & 0.9653 & -0.2591 & -0.0311 \\ 0.0009 & 0.2605 & 0.9495 & 0.1731 \\ -0.0001 & -0.0154 & -0.1752 & 0.9841 \end{pmatrix} . \tag{2.16}$$

The matrix entries may exceed 20 % V→U cross-talk earlier in the day, but the diagonal elements stay close to 1.

2.4.4 Polarimetric accuracy

The absolute error of polarimetric measurements is hard to determine, because only relative measurements are done (the "true" solar intensity is not known, only its conversion to digital detector counts). Additionally, it is difficult to separate the error introduced by the telescope model from that due to the polarimeter response function.

Thus, it is better to give an estimate of the relative cross-talk error for the complete optical train, the combination of polarimeter and telescope. Using the estimates in Beck et al. (2005a,b) the combined error of the cross-talk corrections for telescope and polarimeter for POLIS is about $4-5\times10^{-3}$ and was found to depend on the light level in the telescope.

Note that the value given is the error in the determination of the *cross-talk coefficients* among the different polarization states. Especially the telescope polarization is a second order effect: the matrix entries are multiplied with the incoming Stokes vector, $\mathbf{S}_{\mathrm{in}}$, which has a polarization degree far from 100 %. Taking only the dominant $U \rightarrow V$ cross-talk from the matrix in Eq. (2.16) one has

$$V_{\mathrm{out}} = (0.984 \pm 0.005) \cdot V_{\mathrm{in}} + (-0.175 \pm 0.005) \cdot U_{\mathrm{in}} , \tag{2.17}$$

where usually U_{in} or V_{in} do not exceed $0.3 \cdot I_{\mathrm{in}}$.

Chapter 3

Theory: The interpretation of polarization measurements

3.1 Zeeman effect and line polarization

G.E. Hale used the splitting of spectral lines due of the Zeeman effect to prove the existence of magnetic fields on the Sun. This method works without taking into account any polarization effects, because in magnetic fields of sufficient strength the split components can be seen separately in the intensity spectrum. The required field strength depends on the magnetic sensitivity of the spectral line used. However, to obtain also the field direction, one has to use additional polarimetric information.

The Zeeman effect describes the change of the energy of atomic levels in the presence of a magnetic field. Its main feature is to break the energy degeneracy of atomic levels with different quantum number m. The actual fields on the Sun are 'weak' in the terminology of atomic physics. The so called Russel-Saunders coupling of the angular momentum, **L** , and the electron spin, **S** , to the total angular momentum, **J** , can be assumed. If the magnetic field **B** is taken to be (0, 0, B_z), the problem can be solved in the following way:

- **J = L + S**, where | L - S | $\leq$ J $\leq$ | L + S |
- the quantum number m_j can have the (2J+1) values -J,-J+1,...,J-1,J for a given value of J
- the energy level corresponding to m_J is corrected by a value $\Delta E_{B,m_j}$, for which perturbation theory yields

$$\Delta E_{B,m_j} = \mu_B \, g_j \, m_j \, B \, , \tag{3.1}$$

  where $\mu_B = \frac{e\hbar}{2m_e}$ is Bohr's magneton and the Landé factor g_j is calculated by

$$g_j = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \ . \tag{3.2}$$

This result can best be visualized with an example like Fig. 3.1. The lower level $^5P_1$ splits into three energetically different equidistant levels, while the upper level cannot split up ($\mathbf{J} = 0 \rightarrow m_J = 0,\ g_J = 0$). The three possible dipole transitions π, σ^+ and σ^- differ in wavelength[1] as well as in the polarization state. A transition with ΔM = 0 corresponds to linearly and $\Delta M = \pm 1$ to right-, respectively, left-circularly polarized light. The resulting spectrum consists of three differently polarized spectral lines instead of a single unpolarized one.

Of interest for the observations is also the calculation of the displacement of the σ-components in wavelength units, $\Delta\lambda$, instead of absolute energy units. The spectral displacement is given by

$$\Delta\lambda = \lambda_0 - \frac{hc}{E_0 \pm \Delta E_{B,m_j}} \ , \tag{3.3}$$

[1] $E_\gamma = h \cdot \nu = h \cdot c \ / \ \lambda$, and $E_\gamma = E_0$ ($\rightarrow \pi$) respectively $E_0 \pm \Delta E_{B,m_j}$ ($\rightarrow \sigma^\pm$).

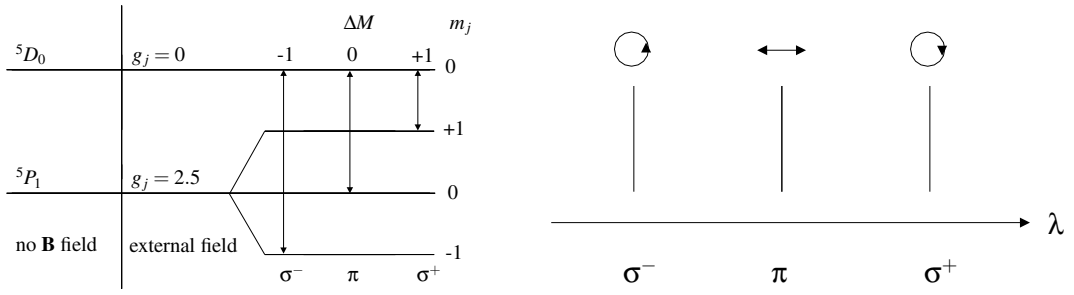


Figure 3.1: **Left**: Zeeman splitting of the Fe I line at 630.25 nm. The notation is $^{2S+1}L_J$. **Right**: Resulting spectrum an external magnetic field. The upper level remains unaffected by the magnetic field, as **J** = 0, and also the Landé factor g_j. The lower one splits into three different levels. The possible dipole transitions π, σ^+ and σ^- are coupled to a fixed polarization each, corresponding to linearly and left- or right-circularly polarized light, depending on the value of ΔM. The π-component lies at the wavelength of the undisturbed transition.

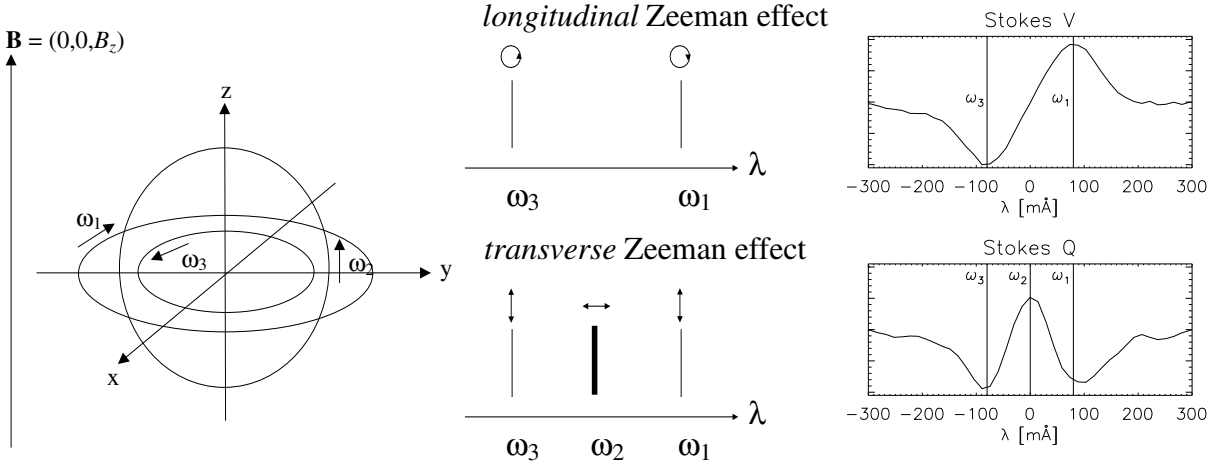


Figure 3.2: **Left**: A sketch of the classical treatment of the Zeeman effect. The rotation of an electron around the atomic nucleus in an external field results in different rotation frequencies ω_i, corresponding to three wavelengths. The frequencies depend on the sense of direction of the rotation and their difference is due to the Lorentz force **v** x **B** . **Middle**: Resulting spectra for observation along and perpendicular to the magnetic field. On observing parallel to the z-axis one can only see the two circularly components ω_1 and ω_3 (longitudinal Zeeman effect). On observing in the x-y-plane normal to the magnetic field there are three linearly polarized oscillations with the frequencies ω_i (transverse Zeeman effect). **Right**: Actually observed spectra in the Stokes parameters V (difference of left- and right-circular polarization) and Q (difference of linear polarization at 0° and 90°). From the relative strength of the components one can retrieve information on the direction of the magnetic field.

which after a short computation yields

$$\Delta\lambda = \pm \frac{g_j m_j e}{4\pi m_e c} \cdot \lambda_0^2 B \,, \tag{3.4}$$

i.e. the separation of the σ-components is directly proportional to the absolute field strength with a quadratic dependence on the rest wavelength of the transition, λ_0.

The case just discussed is the simplest that can happen (*normal Zeeman triplet*). If the Landé factors are not zero for both levels, one has to calculate all energy levels and use the rules for dipole transitions to obtain the allowed transitions. Usually, a spectral distribution results, in which lines of different polarization are mixed. Fortunately it is often possible to use an *effective* Landé factor to group lines with the same polarization together, to reach an analogon to the situation depicted.

The importance of the observation of the polarization does not lie in the calculation of the absolute field strength. This would for example be possible from the distance between the two circularly polarized components, if they are measured separately. The circular components would lie at different wavelengths in the spectrum, even if the splitting cannot be resolved in the intensity profile due to the line broadening. A classical treatment of the Zeeman effect (cf. Collett 1992, Chapter 18) reveals that in the strength of the different polarization components the information on the orientation of the magnetic field also is included.

The motion of an electron around an atomic nucleus is considered there as an oscillation in three dimensions, with a central force given by the coulomb attraction (cf. Fig. 3.2). This system is disturbed by the presence of the magnetic field, which leads to a Lorentz force $\propto \mathbf{v} \times \mathbf{B}$. This results in three different oscillation frequencies depending on the sense of revolution in the x-y-plane, while movement parallel to the z-axis is not influenced. On observing parallel to the magnetic field, i.e. looking down on the x-y-plane, only the two circularly polarized components ω_1 and ω_3 can be seen, as the radiation characteristic of an electric dipole forbids emission in the oscillation direction (*longitudinal* Zeeman effect). If one observes in the x-y-plane normal to the magnetic field lines, one sees three distinct linearly polarized lines at the frequencies ω_i with i=1,2,3 (*transversal* Zeeman effect).

From the strength of the single polarization components relative to each other and their spectral position, i.e. the displacement from the zero level of the undisturbed transition, one has enough information to obtain the strength as well as the direction of the magnetic field. The observed spectrum in the right panel of Fig. 3.2 shows how the mono-energetic Zeeman effect, with in principle indefinitely sharp energy levels, looks like in observations. The components are significantly broadened by the Doppler broadening of the line in the solar atmosphere due to the the thermal velocities of the absorbing atoms, and the limited spectral resolution of the spectrograph system.

The effects discussed here concerning an *emission* line can be also applied on the actual observed *absorption* lines.

3.2 Direct observables

The remote-sensing of the spectrally resolved polarization state of the light allows the retrieval of the conditions in the solar atmosphere, where the light originated. The information can be roughly divided into thermodynamic (temperature, flow velocity) and magnetic parameters (magnetic field vector with field strength and orientation). A number of these atmospheric parameters can be calculated directly from the Stokes profiles, the wavelength dependent polarization signal, without a complex analysis procedure, to avoid additional error sources. The direct observables allow for a comparison with the final result from the more sophisticated inversion procedure to test the consistency of the approach; they are calculated separately for the observed spectral lines with the highest magnetic sensitivity (1564.8/1565.2 nm in the infrared range and 630.15/630.25 nm in the visible range).

Line core velocity The line core velocity is derived from the position of the minimum intensity in the Stokes *I* profile (cf. Fig. 3.3). It gives an estimate of the flow fields in the field-free atmosphere outside the spot or in the magnetic field inside the spot, where each resolution element[2] is completely filled with magnetic flux. The line core velocity is also calculated for the weak Ti I line at 630.37 nm.

Polarization signal Photons emitted (or absorbed) inside of magnetic fields are polarized with a preferred polarization state, opposite to the random polarization in a field-free atmosphere. The amount of polarization is

[2]"resolution element" is used in the following as synonym for a single spatial position (x,y), not in the usual way as the size of the smallest discernible structure, which is 2×2 pixels.

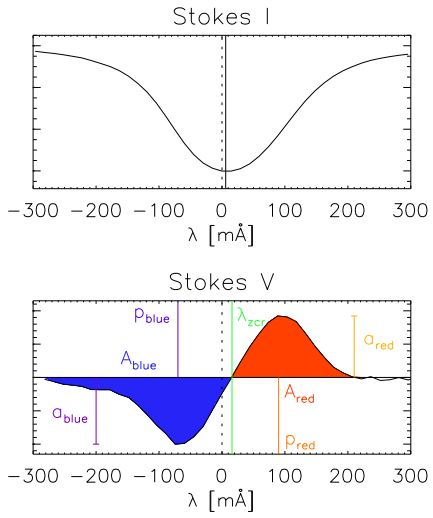


Figure 3.3: The direct observables of the Stokes parameters I and V. From the intensity profile (**top**) the displacement of the line core position relative to the rest wavelength is determined to estimate the velocity. For a regular, i.e. two-lobed, Stokes V signal (**bottom**) the amplitude, a, area, A, and the peak position, p, of the two lobes is determined. The field strength can be estimated from the distance p_{red} - p_{blue}, the velocity inside the magnetic field from the zero-crossing position, λ_{zcr}. The Net Circular Polarization is given by A_{red} - A_{blue}. The polarity is defined by the order of the lobes, the minimum or maximum V signal first.

thus proportional to the amount of magnetic flux present in an individual resolution element. To quantify the polarization amount the *total integrated polarization* (TP) is used. The TP is defined as

$$\mathrm{TP} = \int \frac{\sqrt{Q^2+U^2+V^2}(\lambda)}{I_c} d\lambda \,, \tag{3.5}$$

where I_c is the normalization coefficient described in section 3.3.5.1.

The TP can be used to determine, if a resolution element is field-free or contains a significant amount of magnetic field. However, also the uncorrelated noise in QUV leads to a non-zero TP. Therefore, a second quantity was used for the discrimination between magnetic and field-free resolution elements, the maximum polarization degree, p:

$$p = \max\left(\frac{\sqrt{Q^2+U^2+V^2}}{I}(\lambda)\right) . \tag{3.6}$$

The maximum polarization degree is less sensitive to noise, because it only increases if the polarization signal in QUV is higher at the same wavelength. To determine the presence of magnetic field in a resolution element, a threshold for p was set (0.2 % in infrared and 0.5 % in the visible). Only profiles with at least one spectral line above the respective threshold were assumed to indicate the presence of magnetic field. For profiles showing polarization signal above the threshold a number of additional quantities were calculated:

- The mean ratio of total linear and circular polarization, L/V:

$$L/V = < \frac{\sqrt{Q^2+U^2}}{|V|} > , \tag{3.7}$$

  in a small wavelength range around the maximum absolute Stokes V signal, i.e. the ratio at the position of the stronger σ-component. L/V allows an estimate of the field inclination.

- The relative area asymmetry of the Stokes V signal, dA. The relative area asymmetry is the difference of the area below the two lobes in the Stokes V signal (cf. Fig. 3.3), normalized by the absolute value:

$$\mathrm{dA} = \frac{|\mathrm{A_{blue}}| - |\mathrm{A_{red}}|}{|\mathrm{A_{blue}}| + |\mathrm{A_{red}}|} . \tag{3.8}$$

  It can be calculated easily by

$$\mathrm{dA} = \mathrm{polarity} \cdot \frac{\int \mathrm{V}(\lambda)\mathrm{d}\lambda}{\int |\mathrm{V}(\lambda)|\mathrm{d}\lambda} , \tag{3.9}$$

  where the polarity is ∓ 1 and denotes the orientation of the magnetic field vector to the line of sight. The polarity is given by the order of minimum and maximum Stokes V signal. As a by-product of the calculation, the *Net Circular Polarization*, $\int V(\lambda)d\lambda$, is derived.

- The number and position (with pixel accuracy) of "lobes" in Stokes Q, U, and V. "Lobes" are defined as local extrema above a threshold of 0.1 % in IR (0.25 % in the visible) from smoothed polarization profiles, to reduce the influence of the noise. The number of lobes is used to determine the regularity of the profile shape: two lobes in V and three in Q and U indicate a regular profile (cf. Sect. 3.1). The position of the lobes is used to determine the Doppler shift between circular and linear polarization signal, whose presence indicates multiple magnetic components inside the resolution element (cf. Sect. 5.1).

- For each regular Stokes V profile with exactly two distinct lobes (cf. Fig. 3.3):

  - The amplitude of the Stokes V lobes, a_{blue} and a_{red}, where blue (red) indicates the lobe at smaller (larger) wavelength.
  - The relative amplitude asymmetry, da. Analogously to the area asymmetry, da is given by:

$$\mathrm{da} = \frac{|a_{blue}| - |a_{red}|}{|a_{blue}| + |a_{red}|} \,. \tag{3.10}$$

  - The position of the V lobes with sub-pixel accuracy by a FFT method for a derivation of the field strength from their wavelength separation using Eq. (3.4).
  - The zero-crossing wavelength, λ_{zcr}, from a fit of a linear regression line to a small range around the minimum V signal between the lobes. The profile is implicitly assumed to be a regular V signal, where a_{blue} and a_{red} have opposite signs. λ_{zcr} allows an estimate of the flow velocity inside the magnetic field.

Non-zero amplitude or area asymmetry values indicate gradients or discontinuities in the field strength, the field inclination, and/or the flow velocity (cf. Sect. 5.5).

The corresponding results of the inversion (cf. the following section) can be directly compared; the comparison often shows which spectral line is dominating the determination of a specific parameter, e.g., the strongest split infrared line at 1564.8 nm has the largest influence on the field strength (cf. Sect. 4.6.2).

3.3 Inversion of spectra

In general, an inversion is based on minimizing the deviation between a set of synthetic and a set of observed profiles; the synthetic profiles, $p^{\mathrm{synth}}(\lambda)$, are generated from a model atmosphere, a prescribed stratification of the solar atmosphere that is iteratively modified to reproduce the observations. In other words, one creates an artificial model of the solar atmosphere and examines, if this atmosphere is able to generate the observed profiles and how it should be modified to do so.

The inversion of the spectra allows a more sophisticated determination of the solar atmospheric conditions than the direct observables, because it uses the whole observed spectral range, not only single wavelength points like the positions of the Stokes V extrema. Additionally, it attempts to reproduce all spectral lines at the same time, where the individual properties of the respective lines make some of them more sensitive to specific atmospheric parameters than the others in a complementary way. Inversion procedures were developed in close connection with the progress in instrumentation. After digitized measurements of the full Stokes vector were available, the need arose to automatically analyze a huge number of spectra.

The improvement of the initially assumed model atmosphere is achieved by minimizing a so-called *merit function*, χ^2, the squared deviation between observations and synthetic Stokes profiles given by

$$\chi^2 = \sum_i \left(p_i^{\mathrm{obs}}(\lambda) - p_i^{\mathrm{synth}}(\lambda) \right)^2 \,, \tag{3.11}$$

where i indicates the four Stokes parameters I, Q, U, and V. This approach commonly is referred to as *least-squares fit*.

Inversion procedures mainly differ in the number of degrees of freedom, i.e. how many parameters the inversion code can modify in the model atmosphere of the synthetic spectra. A serious limitation on this number is given by the needed computing power for one iteration: to generate the synthetic spectra, $p_i^{\mathrm{synth}}(\lambda)$, the radiative transfer through the model atmosphere has to be calculated; to minimize the χ^2 value, the atmospheric parameters have to be changed in the correct direction. If more parameters can be changed, the "reaction" of the synthetic profiles

to changes in all these parameters has to be determined. Earlier inversion codes thus tried to strongly restrict the number of free parameters, especially in connection with the available computer resources at that time.

The model atmosphere is iteratively modified, until a minimum of χ^2 is reached, where the synthetic and observed profiles match closely (*best-fit*). However, the final solution for the best-fit model atmosphere is by no means unique. For each set of observed Stokes parameters, an infinite number of possible atmospheres can be found, which generate exactly the same spectra. Also, several local minima with a very similar χ^2-value may exist in the parameter space of the model atmospheres. However, the amount of information from different spectral lines restricts the number of possibilities. Some stratifications can be excluded as highly improbable. A "reasonable" choice of the initial model atmospheres in most cases prevents that the inversion code runs into a physically improbable but possible solution of the inversion problem. However, one should keep in mind that it is formally incorrect to speak of a "measurement" of field properties; it always is an interpretation of the observed spectra with certain assumptions.

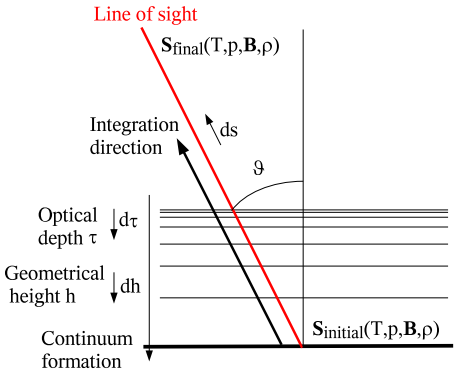


Figure 3.4: The geometrical situation for the calculation of the radiative transfer in the spectral synthesis. The line of sight, the connection between observer and light source, is inclined to the local surface normal by an angle θ. The geometrical height scale is measured in absolute length units along the surface normal, the optical depth in relative units along the same axis. The Stokes vector, $\mathbf{S}_{\rm initial}$, at the bottom of the solar atmosphere, where it is opaque (continuum formation layer), has to be propagated along the line of sight through the whole atmosphere to get the final results, $\mathbf{S}_{\rm final}$.

3.3.1 Radiative transfer in the solar atmosphere

To generate synthetic spectra from a model atmosphere, the changes of the radiation propagating through the atmosphere have to be calculated. The geometrical situation depicted in Fig. 3.4 is commonly assumed: the line of sight is inclined by the angle θ to the local surface normal. Two height scales exists: the geometrical scale, h, given in absolute length units along the surface normal, and the optical depth scale, given in relative units in the same direction. The light propagates along the line of sight and the changes of the Stokes vector $\mathbf{S}$ on its passage through the atmosphere have to be calculated.

3.3.1.1 The optical depth scale

In the spectral synthesis it is sufficient[3] to use a relative height scale, the so-called *optical depth scale*, τ. It is defined by following the line of sight in the opposite direction, i.e. from the observer towards the light source, and calculating the absorption of a light ray with the intensity I_ν. With the absorption coefficient, κ_ν, the optical depth is defined (for normal incidence, $\theta = 0^\circ$) using

$$\frac{dI_\nu}{dh} = -\kappa_\nu I_\nu \tag{3.12}$$

by

$$d\tau_\nu = \kappa_\nu dh\,. \tag{3.13}$$

Integration of Eq. (3.13) yields the general expression

$$\tau_\nu = \int \kappa_\nu dh\,. \tag{3.14}$$

An easy interpretation of τ is given by looking at the intensity, which simply is $I_\nu(\tau) = I_\nu^0 \cdot \mathrm{e}^{-\tau_\nu}$, i.e. the larger τ, the less intensity is left from an initial value I_ν^0. The absorption coefficient, κ_ν, is a function of the physical properties of the atmosphere like temperature, density, pressure or chemical composition.

[3] In addition, the true geometrical scale is not known.

3.3.1.2 The radiative transfer equation

The radiative transfer equation for the Stokes vector, $\mathbf{S}$, has to be calculated along the line of sight going upwards in the atmosphere, using the path length, s, respectively, the infinitesimal change ds. As first step, the balance equation for the intensity, I_ν is

$$\frac{dI_\nu}{ds} = -\kappa_\nu I_\nu + \varepsilon_\nu \ , \tag{3.15}$$

where ε_ν is the emissivity in the atmosphere. From Fig. 3.4 it can be easily seen that

$$d\tau = -\kappa_\nu \cos\theta ds \ . \tag{3.16}$$

Thus, Eq. (3.15) can be changed to

$$\cos\theta \frac{dI_\nu}{d\tau} = I_\nu - \frac{\varepsilon_\nu}{\kappa_\nu} \tag{3.17}$$

in the optical depth scale. The fraction, $\frac{\varepsilon_\nu}{\kappa_\nu}$, is termed the *source function*, $\mathcal{S}_\nu$. In the case of *local thermal equilibrium* (LTE) – for practical purposes the most important one – it can be expressed at once by the Planck radiation law for black-body radiation:

$$\mathcal{S}_\nu(T) = B_\nu(T) = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{\mathrm{e}^{h\nu/kT} - 1} \ . \tag{3.18}$$

The LTE condition implies that all local temperature changes are at once transferred by collisions to all other hydrodynamic parameters like the pressure or the population number of atomic states; the temperature is then sufficient to calculate the local equation of state. The emitted radiation fulfills the definition of unpolarized light (cf. Sect. 2.1): the spontaneous emission has no phase relation of any kind and the emitted photons show all polarization states with the same relative frequency.

In the stratified solar atmosphere, the dependence of the temperature on geometrical height can also be expressed as dependence on optical depth, thus, Eq. (3.17) can be re-written into the *radiative transfer equation* (RTE)

$$\cos\theta \frac{dI_\nu(\tau)}{d\tau} = I_\nu(\tau) - B_\nu(\tau) \ . \tag{3.19}$$

Up to now the frequency dependence of κ_ν was not considered explicitly. In the case of an atomic transition, the absorption coefficient has a certain value, κ_0, at the resonance frequency (or rest wavelength) of the transition (*line absorption*). The absorption by other non-resonant processes (*continuum absorption*) is expressed by an only slightly wavelength dependent coefficient, κ_c. Additionally, for polarimetry not only the intensity is of interest but the full Stokes vector $\mathbf{S} = (I, Q, U, V)$. However, analogously to Eq. (3.19), the propagation of the Stokes vector $\mathbf{S}$ can be formulated in the generalized form of the RTE for polarized light by

$$\cos\theta \frac{d\mathbf{S}_\nu(\tau)}{d\tau} = \mathcal{M} \cdot \left(\mathbf{S}_\nu(\tau) - (B_\nu(\tau), 0, 0, 0)^{\mathrm{T}}\right) \ , \tag{3.20}$$

with the *total absorption matrix*, $\mathcal{M}$. It consists of two parts:

$$\kappa_c \cdot \mathcal{M} = \kappa_0 \cdot \eta + \kappa_c \cdot \mathbb{1} = \begin{pmatrix} \eta_I & \eta_Q & \eta_U & \eta_V \\ \eta_Q & \eta_I & \rho_V & -\rho_U \\ \eta_U & -\rho_V & \eta_I & \rho_Q \\ \eta_V & \rho_U & -\rho_Q & \eta_I \end{pmatrix} + \kappa_c \cdot \mathbb{1} \ , \tag{3.21}$$

where the continuum absorption is separated from effects due to the line transition in the Mueller matrix η.

The matrix entries of η depend on the properties of the respective transition, on the properties of the atmosphere, namely, the temperature, T, and on the presence (or absence) of magnetic field. A complete derivation following Landi degl'Innocenti & Landi degl'Innocenti (1972) is given in Appendix B. Here it is sufficient that the coefficients of η can be calculated in case the temperature and the magnetic field vector are given as function of τ. To generate a synthetic spectrum, all that is left is the integration of Eq. (3.20) with suitable boundary conditions.

3.3.2 The Milne-Eddington approximation

The problem of the radiative transfer can be simplified using the so-called *Milne-Eddington* (ME) approximation. The following assumptions are made (cf. Auer et al. 1977):

- The ratio κ_0/κ_c is constant throughout the atmosphere, and the source function $B_\nu(\tau)$ is linear in τ.
- The magnetic field vector is constant in τ.

With these assumptions, the radiative transfer equation can be solved analytically, and the emergent Stokes vector $\mathbf{S}_{\rm final}$ can be derived instantaneously for any given magnetic field. Starting from an initial field configuration, the atmosphere can be iterated to minimize the χ^2-value (cf. Eq. (3.11)), the deviation between the synthetic and the observed profiles.

For the minimization one has to calculate the gradient of χ^2 with respect to each free parameter, α_j:

$$\nabla_{\alpha_j}\left[\chi^2(\vec{\alpha})\right] = \nabla_{\alpha_j}\left[\left(\mathbf{S}^{\rm obs}(\lambda) - \mathbf{S}^{\rm synth}(\vec{\alpha},\lambda)\right)^2\right] \equiv \frac{1}{\Delta\alpha_j}\left[\chi^2(\vec{\alpha}+\Delta\alpha_j) - \chi^2(\vec{\alpha})\right] , \tag{3.22}$$

where the parameter set $\vec{\alpha}$ contains, e.g., the field strength, B, the field inclination, γ, the field azimuth, ϕ, or, in case of the ME-approximation, the slope and zero offset of the source function, $B_\nu(\tau) = B_0 + \beta\tau$. $\vec{\alpha}+\Delta\alpha_j$ is a shortcut for $(\alpha_1, \alpha_2, \ldots, \alpha_j + \Delta\alpha_j, \ldots, \alpha_n)$. The advantage of the ME-based inversion is its speed; the analytical solution of the radiative transfer allows at once the calculation of the gradients. Additionally, the ME-based inversions are usually very stable, i.e. the iteration converges very fast to a minimum, because of the limited number of degrees of freedom. The procedure suggested by Auer et al. (1977) has been refined by, e.g., Skumanich & Lites (1987), and has been applied to a lot of observational data.

However, the assumptions of the ME-atmosphere are quite strong. The limitation on the shape of the source function leads to a large possible error in the determination of the temperature. A ME-inversion is also not able to introduce gradients or discontinuities in physical parameters, leading to always perfectly symmetric or anti-symmetric Stokes profiles. An improved inversion code thus has to leave the ME-approximation in order to retrieve a better picture of the magnetic field geometry in the solar atmosphere from observations.

3.3.3 The SIR code

The SIR code (Stokes Inversion based on Response functions, Ruiz Cobo & del Toro Iniesta 1992) is one way to improve the realism of the inversion without enforcing a specific model of the solar atmosphere a priori and maintaining a large number of degrees of freedom at a reasonable computation time. It makes use of a side-effect in the calculation of the radiative transfer equation, where the so-called *response functions* are automatically determined.

As a reminder, the general inversion problem to be solved is the minimization of the squared deviation between synthetic and observed profiles, χ^2, by modifying the initial atmospheric stratification. Any gradient method for the minimization requires to calculate the derivatives of χ^2 with respect to each parameter to find the closest local minimum. However, without the ME-approximation (almost) all parameters are defined on a grid in optical depth, $\tau = (\tau_0, \tau_1, ..., \tau_k)$, and need not be constant with depth. Therefore, partial derivatives have to be calculated for a change of each parameter in each optical depth point. Eq. (3.22) has to be modified to

$$\nabla_{\alpha_j(\tau_i)}\left[\chi^2(\vec{\alpha}(\tau))\right] = \frac{1}{\Delta\alpha_j(\tau_i)}\left[\chi^2(\vec{\alpha}(\tau) + \Delta\alpha_j(\tau_i)) - \chi^2(\vec{\alpha}(\tau))\right] , \tag{3.23}$$

where $\alpha_j(\tau)$ is now modified only in the i-th entry, $\alpha_j(\tau_i) + \Delta\alpha_j(\tau_i)$. Thus, for the calculation of the gradient in one iteration of the model atmosphere, the radiative transfer equation has to be solved k$\times$ n times, where k denotes the points in optical depth and n the number of parameters in $\vec{\alpha}$. "This is certainly impracticable" (Ruiz Cobo & del Toro Iniesta 1992, p. 376).

However, Ruiz Cobo & del Toro Iniesta (1992) showed that the needed partial derivatives can be calculated during the integration of the radiative transfer equation (RTE) using the so-called *response functions* (RFs). The RFs give the resulting change of the synthetic Stokes profiles, $\Delta\mathbf{S}^{\rm synth}(\vec{\alpha},\lambda)$, on a perturbation of any physical parameter in a specific optical depth. The RFs are defined by

$$\Delta\mathbf{S}^{\rm synth}(\vec{\alpha},\lambda) = \int_0^\infty \mathbf{RF}(\lambda,\tau)\Delta\alpha_j(\tau)d\tau , \tag{3.24}$$

where a linear response of $\mathbf{S}^{\rm synth}(\vec{\alpha},\lambda)$ to the perturbation $\Delta\alpha_j(\tau)$ is assumed (Landi Deglinnocenti & Landi Deglinnocenti 1977). $\mathbf{RF}(\lambda,\tau)$ is a vector with four entries, which contains the response of the Stokes parameters *IQUV* separately. The right-hand side of Eq. (3.23) can then be modified to

$$\chi^2(\vec{\alpha}(\tau)+\Delta\alpha_j(\tau_i))-\chi^2(\vec{\alpha}(\tau))=\Delta\chi^2\equiv\chi^2(\mathbf{S}+\Delta\mathbf{S})-\chi^2(\mathbf{S})=\sum_m f(\mathbf{S})\cdot\Delta S_m\,, \tag{3.25}$$

where $m = IQUV$ and $\mathbf{S} = (\mathbf{S}^{\rm synth}(\vec{\alpha},\lambda),\mathbf{S}^{\rm obs}(\lambda))$. Using Eq. (3.24) and introducing the discrete spectral sampling of the observations with λ_l ($l=0,\ldots,q$) one obtains

$$\Delta\chi^2=\left[\sum_l\sum_m f(\mathbf{S}(\lambda_l))\cdot RF_m(\lambda_l,\tau_i)\right]\cdot\Delta\alpha_j(\tau_i)\equiv\nabla_{\alpha_j(\tau_i)}\left[\chi^2(\vec{\alpha}(\tau))\right]\cdot\Delta\alpha_j(\tau_i)\,, \tag{3.26}$$

with $m = IQUV$. The $RF_m(\lambda,\tau_i)$ are thus identical to the derivatives of χ^2, except for some factors (Ruiz Cobo & del Toro Iniesta 1992).

For a fast integration of the RTE, the so-called *Diagonal Lambda Operator Method* (DELO, Rees et al. 1989) can be used. In this algorithm, some quantities have to be calculated on each depth point τ_i during the integration of the RTE. From exactly the same quantities the $\mathbf{RF}(\lambda,\tau_i)$ can be determined (Sanchez Almeida 1992). Thus, all response functions $\mathbf{RF}(\lambda,\tau)$, and with them the gradient of χ^2, can be calculated during a single integration of the radiative transfer equation. It is not necessary to integrate the RTE separately after each change by $\Delta\alpha_j(\tau_i)$. All necessary calculations can thus be performed during a single integration of the RTE, yielding a) the synthetic spectra and b) the partial derivatives of the χ^2-value with respect to all free parameters on all points τ_j of the depth grid.

The only remaining item for the inversion code that needs consideration is the stepwidth. Each gradient-based minimization method first determines the "direction", in which to proceed with a specific parameter (reverse to the gradient). The step size in this direction has to be chosen. A common procedure also in other numerical applications is the so-called *Marquardt algorithm*, which uses second derivatives, the Hessian matrix, to determine the stepwidth (cf. Press et al. 1986).

Due to the derivation of all RFs during the – unavoidable – integration of the radiative transfer equation, the SIR code allows to keep a large number of degrees of freedom. The parameters can in principle be used with no restrictions on the shape of their curves with optical depth. However, some limitations are implemented in the SIR code to save at least some computing time.

3.3.4 The SIR structure – Model building in the inversion code

With the exception of the penumbra of sunspots (cf. Sect. 5.1) or the canopy field around spots (cf. Sect.1.2.2.2), one expects that the atmospheric parameters in the solar atmosphere show a smooth variation without discontinuities in optical depth. Thus, it is not necessary to change single parameters separately on each optical depth point τ_i. The SIR code therefore has been set up to use polynomial curves in τ for the iteration of the atmospheric stratification; the order of the polynomial can be chosen. A polynomial of 0th order allows the code to change a specific parameter by the same constant amount on all depth points, 1st order allows for linear trends with τ like in the ME-approximation, 2nd order for a parabolic change reducing/increasing a parameter in the lower and upper τ-levels at the same time, and so on.

However, except for the temperature curve, $T(\tau)$, the option of using gradients in physical parameters has not been used in the present work. Gradients in atmospheric quantities correspond to an increased number of degrees of freedom, which increases the number of existing local minima of the χ^2-value. As any gradient-based minimization method will converge in the nearest local minimum, the probability for "unreasonable" solar model stratifications as inversion result increases. Even if in the present case an uniquely large amount of information from 7 spectral lines from visible to infrared spectral ranges was available, the allowance for gradients was found in many cases to lead to possible, but "unreasonable" solar model stratifications. We thus have restricted the inversion freedom to zeroth-order polynomials ($\equiv$ constant in τ). This forces the inversion code to use the simplest possible model for the reproduction of the observed profiles.

Even if the restriction to parameters constant in optical depth leads to little improvement compared to the ME-approximation, the freedom in the temperature dependence makes the approach superior to the ME-inversion. The SIR code also allows for the fit of multiple spectral lines, a requirement for the analysis of the simultaneous infrared

and visible spectra used in the present work. Finally, the model of the solar atmosphere can be refined by using two independent atmospheric components.

Firstly, the parameter set, $\vec{\alpha}$, for each model atmosphere is given by the following list[4]:

- The temperature stratification, $T(\tau)$.
- The electron pressure, $e(\tau)$.
- The line-of-sight velocity, $v_{\rm LOS}(\tau)$.
- The magnetic field vector, given by field strength, $B(\tau)$, the inclination angle between field vector and line of sight, $\gamma(\tau)$, and the field azimuth, $\Phi(\tau)$.
- The microturbulent velocity, $v_{mic}(\tau)$.
- The macroturbulent velocity, v_{mac}.

The electron pressure, $e(\tau)$, is not used as a free parameter, because the profiles observed do not contain enough information to restrict its value. It is forced to have a boundary condition at the uppermost point of the optical depth grid, where it is assumed to have the corresponding value of the Harvard Smithsonian Reference Atmosphere (HSRA, Gingerich et al. 1971). In the other layers it is calculated consistently in the LTE-approximation by the requirement of hydrostatic lateral pressure equilibrium.

Each model can be easily specified to be field-free: If the initial field strength is set to 0, the synthetic spectra will be calculated without any magnetic field; also during the iteration no magnetic fields will be introduced.

The two parameters $v_{mic}(\tau)$ and v_{mac} are somewhat artificial quantities. Originally, they have been introduced to account for unresolved velocity fields in a resolution element. The label macro- or microturbulent indicates, if the characteristic length scale of the line-of-sight velocity component is much larger or smaller than the mean free photon path (cf. de Jager & Vermue 1977). The effect of such spatially unresolved velocity fields is a broadening of the spectral lines, due to the velocity difference between parts of the resolution element.

In the inversion of spectra, $v_{mic}(\tau)$ and v_{mac} can additionally be used to correct for the finite spectral resolution (*Spectral Point Spread Function*, SPSF) of the observation instrument, being it a spectrograph or any other system. The SIR code in principle allows to take the line broadening due to SPSF into account, if the SPSF is explicitly given; however, for both spectro-polarimeters at the VTT no measurement of the SPSF is available due to the technical difficulties involved in deriving them[5]. Thus, the two parameters $v_{mic}(\tau)$ and v_{mac} are used to resemble the effect of the SPSF and to account for the presumably existing velocity dispersion inside a resolution element. In the profile synthesis, the broadening corresponding to a non-zero microturbulent velocity is added to the natural line width, that of the macroturbulent velocity is added to the thermal Doppler width (cf. Appendix B), before the profiles are convolved with the respective broadening function. All synthetic spectral lines calculated by the SIR code are thus broadened by a given amount to yield a better reproduction of the observed spectra.

Secondly, the SIR code allows to estimate the amount of straylight in a resolution element. To this extent a so-called *straylight profile*, $p_{\rm stray}$, has to be supplied, whose relative contribution, δ, to the synthetic profiles is fitted. The straylight is intended to compensate for the contamination of the observations on a specific spatial position by unwanted reflections into the light path happening inside the Earth's atmosphere and the telescope or instrument optics. Dealing with straylight is in principle a very complicated task; one has to consider both *local* straylight from the close surroundings of an individual spatial point, and *global* straylight, which may come from the complete field of view of the instrument and/or the telescope. We assume that the straylight is mainly coming from the whole field of view of the instruments. The straylight profile thus is derived by averaging intensity spectra of granulation dominated areas outside of sunspots, because of their high intensity level. The straylight profile does not contain any polarization signal.

Finally, the SIR code allows to choose up to two independent model atmospheres. The model atmospheres can be chosen individually to be field-free or to contain magnetic fields; their respective set of parameter values $\vec{\alpha}_1$ and $\vec{\alpha}_2$ is determined independently of each other. The relative filling fraction of each component, ε, is also a free parameter. In the present work the combination of one field-free and one magnetic atmosphere has been used

[4]The argument τ has been kept for completeness, although most parameter values are kept constant with optical depth.

[5]A requirement for the measurement would be an almost mono-chromatic light source like a mode-stabilized laser at the corresponding wavelength.

for the magnetic elements outside of the sunspot, whereas inside the penumbra two magnetic components were assumed.

The general structure of the solar model atmosphere used by the inversion code in the generation of the synthetic spectra is then given by the following equation:

$$p_i^{\text{synth}}(\lambda, \vec{\alpha}_1, \vec{\alpha}_2) = \delta \cdot p_{\text{stray}} + (1-\delta) \cdot [\varepsilon \cdot p_i^1(\lambda, \vec{\alpha}_1) + (1-\varepsilon) \cdot p_i^2(\lambda, \vec{\alpha}_2)] \,, \tag{3.27}$$

where $i = IQUV$ and p_i^j ($j = 1,2$) indicates the profiles, which are generated in the atmospheric component j with the atmospheric structure given by $\vec{\alpha}_j$.

"Fine tuning" of inversion setup The inversion setup is usually adjusted to the properties of the profiles to be inverted to increase the speed of convergence:

- The orientation of the initial magnetic field is chosen to be in agreement with the polarity of Stokes V.
- The type of the profile is determined from its continuum intensity. Three different types are discerned: a) inside the dark umbra of a sunspot, b) inside the penumbra, c) outside of the spot. A mask of the respective type of each resolution element is created before the inversion.
- The number and the properties of atmospheric components to be used by the inversion code is adjusted to the type.

However, some parameter values and setup options were not changed in all inversions:

- The optical depth grid was always set to span the range from $\log\tau$ = -4 to $\log\tau = 1.4$ with a 0.1 stepwidth in $\log\tau$, i.e. 55 equidistant grid points in the logarithm of the optical depth were considered.
- The initial temperature and electron pressure stratification of all model atmospheres was set to be that of the Harvard Smithsonian Reference Atmosphere (HSRA, Gingerich et al. 1971).
- The SIR codes allows to give a different weight, w_i, to the contribution of a specific Stokes parameter to the χ^2:

$$\chi^2 = \sum_i w_i \left(p_i^{\text{obs}}(\lambda) - p_i^{\text{synth}}(\lambda) \right)^2 \,, \tag{3.28}$$

  where $i = IQUV$. The weights have always been set to put most emphasis on the circular polarization signal and least on intensity, i.e. usually they were set to $(w_I, w_Q, w_U, w_V) = (1,5,5,10)$.
- For a two-component inversion, the SIR code was forced to use the same value of macroturbulent and microturbulent velocity for both atmospheric components ($v_{mic}^1 = v_{mic}^2$ and $v_{mac}^1 = v_{mac}^2$), because the unresolved velocities and the SPSF should be identical.
- The wavelength range to be fitted was restricted to the spectral lines and their surroundings. The telluric spectral lines in the visible spectral range were excluded. The same applies to two molecular blends of OH^-, which appear near the 1565.2 nm line as soon as the temperature is low enough for the formation of this molecule, mainly in the umbra of a sunspot (cf. Mathew et al. 2003). A telluric line blend in the red wing of the 1565.2 nm line is removed: a fit to the average straylight profile p_{stray} is performed; the best fit profiles do not contain the telluric absorption line. The deviation between fit and the observed profile is used to correct for the line blend in $IQUV$.
- The heliocentric angle, θ, in Eq. (3.20) was set to be 0° in the inversion ($\equiv$ observation assumed to be at disc center). Effects due to the actual off-center position of the observation targets were included in the intensity normalization.
- The intensity normalization and the wavelength scale were derived from the respective straylight profile, p_{stray} (cf. the next two sections).

Additional differences in the inversion setup for magnetic elements outside the spot and the magnetic fields in the spot itself are given at the beginning of the respective chapter.

3.3.5 Preparation of the data for inversion

3.3.5.1 Intensity normalization

SIR internally needs the Stokes parameters, $IQUV$, to be expressed in an absolute energy scale. However, the digitized observations yield only a relative intensity measurement in (arbitrary) detector counts. To allow for an easy conversion between the different ways to express the intensity, the HSRA model is taken as a reference. The SIR code converts the absolute energy scale in the synthetic spectra to a relative scale by division with the spectral energy density, which is retrieved from integrating the HSRA model. As the HSRA model is an average temperature stratification for *quiet Sun* (QS) intensity at disc center, this implies that the observations should also be normalized with the QS intensity in detector counts at disc center. However, to avoid having to take another data set at disc center only for normalization purposes, usually the mean intensity in a presumably quiet Sun area somewhere inside the field of view (FOV) is used.

For observations off disc center the normalization thus has to be corrected for the center-to-limb variation of the intensity (*limb darkening*). The limb darkening is a consequence of the thermally stratified solar atmosphere, $T(h)$, where T decreases with the height above the surface, h. If the line of sight is inclined to the solar surface, it has stronger contributions from the cooler upper layers due to the increased path length in them. The effect is most pronounced near the solar limb, where the LOS is almost parallel to the surface and essentially samples a constant temperature, $T(h = h_0)$. The limb darkening as function of the heliocentric angle is available for the visible spectrum in tabulated form. However, for the data sets examined in this work an additional complication arises, because of the simultaneous usage of infrared (IR) and visible (VIS) spectral lines. IR and VIS have a different slope in the center-to-limb intensity variation due to the wavelength dependence of the source function, $B_\nu(T(h))$ (cf. Eq. (3.18)), and no limb darkening tables are available for the IR.

To compensate the change of intensity due to the off-center position consistently in both wavelength ranges, wavelength dependent limb-darkening coefficients, $c_{\rm ld}$, for the heliocentric angle of the observations were derived with the SIR code. The $c_{\rm ld}$ were computed using the HSRA reference atmosphere in the spectral synthesis mode of SIR: synthetic profiles were created, setting the heliocentric angle, θ, in Eq. (3.20) to that of the observations; the continuum intensity was compared to a profile generated with $\theta = 0°$. The intensity ratio $c_{\rm ld}$ was used for scaling the intensities of the observed spectra.

Global intensity trends due to the change of the solar elevation above the horizon during the longer observations have also been removed. To this end, the mean intensity in a quiet Sun area was derived for each scan step, separately for IR and VIS. A parabola was fitted to the intensity curves with time. For each scan step k the corresponding value of the parabola, $pb(k)$, was used as normalization coefficient in the respective wavelength.

The procedure for the normalization of a single profile in scan step number k, $p_s(\lambda, k)$, thus consists of the following steps:

- Calculation of an average quiet Sun profile, $p_{\rm stray}$, from a part of the field of view.
- Derivation of the continuum intensity in $p_{\rm stray}$, I_c, outside the spectral lines .
- Division of $p_s(\lambda, k)$ with I_c ($\rightarrow$ normalization to disc center).
- Multiplication with the limb-darkening coefficient, $c_{\rm ld}$ ($\rightarrow$ normalization to off-center position).
- If necessary, multiplication with the value of the correction curve for intensity trends during the observation, $pb(k)$.

3.3.5.2 Wavelength scale

The wavelength scale depends on two factors: the spectral dispersion and the rest wavelength of the spectral lines, λ_0^i, where i cycles through the spectral lines observed (cf. Table 4.1). The dispersion affects the relation between magnetic field strength and corresponding separation of the Stokes V lobes and any velocity derived from the relative displacement to $\lambda_0^{\rm i}$. To have a common reference in both IR and VIS spectra, it is assumed that the convective blue shift (cf. Sect. 1.1.1) in the average quiet Sun profile, $p_{\rm stray}$, is given by the model calculations described in Borrero & Bellot Rubio (2002).

To set up the scale, λ_0 is set for a first spectral line in the IR spectrum (1564.8 nm) and visible spectrum (630.15 nm) such that the observed position of the line core in the straylight profile is at the predicted convective blue shift

spectral line, λ_0 in nm	630.15	630.25	630.37	1564.8	1565.2
blue shift velocity in m/s	-185	-262	-424	-445	-469
spectral dispersion in pm / pixel		1.486		2.965	
velocity dispersion in m/s per pixel		700		570	

Table 3.1: 1st row: The convective blue shift of the respective spectral lines from the 2-component quiet Sun model of Borrero & Bellot Rubio (2002), expressed by the corresponding velocity in m/s. The 2nd row gives the average spectral dispersion of the respective instrument, POLIS and TIP; the 3rd the corresponding velocity dispersion, the conversion coefficient from pixel to velocity. For POLIS the blue shift velocity of the 630.15 nm line corresponds to a spectral shift of only 0.257 pixel (cf. Fig. 3.5).

value given in Table 3.1. This means, the rest wavelength λ_0 of the line has exactly the distance from the line center position required for the respective blue shift (cf. Figs. 3.5 and 3.6). The dispersion is adjusted until a second line, 1565.2 (630.25) nm, shows the correct convective blue shift. The dispersion found for each instrument is around 29.65 mÅ/pixel for the IR spectrum from TIP and around 14.86 mÅ/pixel for the visible spectrum from POLIS.

In the case of the time series studied in the following chapter, a new straylight profile is calculated for each repetition of the scan using only profiles from the respective repetition; the procedure just described is applied individually to each straylight profile, with the mean dispersion values given in Table 3.1. For the sunspot data analyzed in Chapter 5, only one straylight profile was used, with dispersion values of 2.958 pm/pixel for TIP and 1.484 pm/pixel for POLIS, respectively.

For the VIS data from POLIS, the telluric line at 630.20 nm is used to ensure that all individual profiles are relative to the wavelength scale defined from the straylight profile. The line position of 630.20 nm in the straylight profile and in each individual profile is compared. If they deviate, the individual profiles are shifted by the position difference. This takes into account the spectral curvature along the slit due to the short focal length of the compact POLIS spectrograph design (cf. Sect. 2.3.1, and Beck et al. 2005b). For TIP, the curvature along the slit is negligible due to the larger focal length and the smaller CCD size and no correction has to be applied.

The reader may question why this approach to set up a relative wavelength scale was chosen, although for the visible spectrum two telluric spectral lines with known rest wavelengths allow to retrieve an absolute wavelength scale. The main argument against relying on the telluric lines is that in the infrared spectrum NO telluric lines can be used. The telluric blend of 1564.8 nm is too weak and its rest wavelength is unknown. Especially due to the necessity of a common wavelength scale in the infrared and visible spectral ranges, we felt compelled to use a method, which treats both spectral windows in the same way, to ensure that at least a comparable error will be introduced. The advantage of the method also is that all relative motions between the Sun and the observer are effectively removed. However, the question on the accuracy of the wavelength scale remains. Three major items have to be considered:

1. The accuracy in the determination of line core positions or the corresponding velocities from displacements to the rest wavelength: using the spectra displayed in Fig. 3.5 and the velocity dispersion of > 500 m/s per

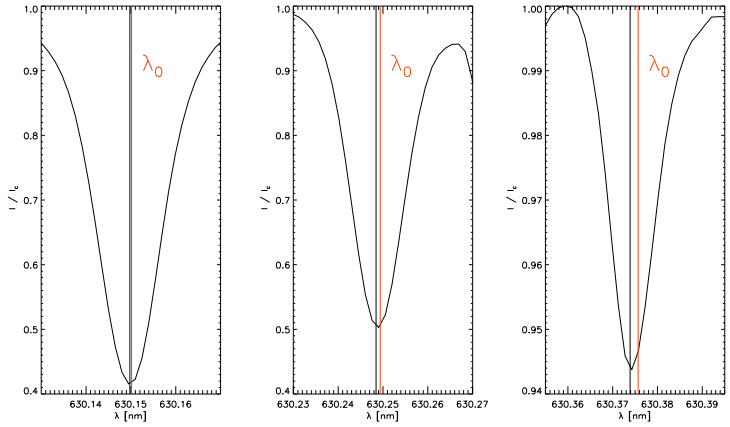


Figure 3.5: Wavelength scale for the visible spectrum from POLIS. For each spectral line the observed line core position (**black**) is displaced from the rest wavelength λ_0 (**red**) by the amount predicted by the 2-component quiet Sun model of Borrero & Bellot Rubio (2002).

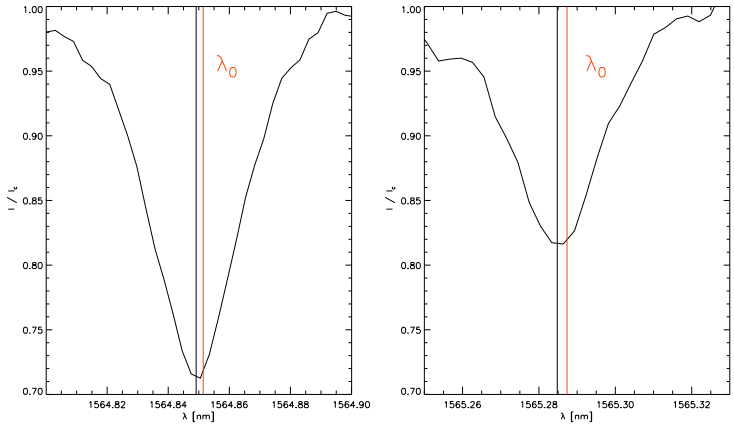


Figure 3.6: Same as Fig. 3.5 for the infrared spectral lines observed with TIP.

pixel, it can be estimated that velocities can be determined only up to $\pm$ 0.2 pixel $\cdot$ 500 m/s = $\pm$ 100 m/s. This limitation already applies to the setup of the wavelength scale.

2. The accuracy of the convective blue shift values from the 2-component quiet Sun model: Borrero & Bellot Rubio (2002) give an upper limit of 150 m/s at 500 nm.

3. The presence of the moat flow in the surroundings of sunspots (cf. Sect. 1.2.2.2): the straylight profiles are sampled in a granular area still partly inside the moat flow of the spot. A moat flow of 0.5 kms^{-1} may lead to an additional red or blue shift of around 200 m/s at an heliographic angle around 30°. However, this effect will be a global shift of all velocities in the same direction. It does not affect an observed velocity difference between field-free and magnetic atmosphere in one resolution element or between velocities in different resolution elements.

In total, the accuracy of any velocity determination will be limited to around $\pm$ 100-150 m/s. The inversion results will be more precise than any direct observable (cf. Sect. 3.2), because they rely on the full spectral range, not only single-valued parameter like line core positions.

3.4 Error discussion

In the present study, two types of error sources influence the results. Firstly, the observed Stokes profiles are influenced by several effects:

- Atmospheric disturbances during the observations lead to spatial smearing, which has strong effects, if close-by solar structures differ markedly in their properties.
- The observed spectra show a white noise pattern due to the thermal and read out noise of the CCD cameras.
- Calibration errors can change the polarization state.
- The finite spectral sampling limits the determination of parameters from the spectra.

Thus, already the observations set some limits on the accuracy that can be achieved. Secondly, the later interpretation of the polarization signal, using the direct observables or the inversion, introduces additional uncertainties. For the inversion, mainly the ambiguity of the final best-fit model atmosphere is a problem. In the following an estimate on the order of magnitude of the main error sources will be given.

The SIR code The SIR program code has been tested extensively already by several authors (e.g. Westendorp Plaza et al. 1998; Ruiz Cobo 1998; Bellot Rubio et al. 2000b,a; Bellot Rubio & Collados 2003; Bellot Rubio et al. 2004). Commonly, synthetic profiles from a given atmospheric stratification with gradients or even discontinuities in physical parameters were calculated. Some noise was added and the resulting profiles were inverted with the SIR code. The final best-fit model atmosphere was compared to the known initial values. In some cases, even

step functions in physical parameters were correctly retrieved by the inversion code. The general conclusion of the authors mentioned above was that the limit of the inversion procedure is given by the noise level in the profiles and the complexity of the initial atmospheric stratification.

In the present work, the model assumptions were set to be the simplest possible case of parameters constant in optical depth. Due to the exclusion of any gradients, the fit results give a robust average value of physical parameters over the height range, in which the spectral lines are sensitive. The final best-fit atmospheric stratification can therefore be assumed to be the simplest way to reproduce the observed profiles in visible and infrared spectral lines. This does NOT exclude that they can be also produced by an infinite number of other more complex and less likely stratifications.

Polarimetric sensitivity The limiting factor for the detection of magnetic fields is the noise level in the observations. The noise level can be estimated from the *root-mean-squared value* (rms)[6] in continuum windows: outside of the Zeeman sensitive spectral lines the polarization in *QUV* shows only random fluctuations. The rms-noise level in the observations was around 0.4-1 $\cdot 10^{-3}$ of the continuum intensity, I_c. Stokes profiles where classified as showing significant signal if the maximum polarization amplitude exceeded 2.5 $\cdot 10^{-3}$ in infrared spectral lines (7.5 $\cdot 10^{-3}$ in visible), thus, at least 1-σ above the noise level.

Polarimetric accuracy An estimate of the accuracy in the cross-talk determination of around 4-5$\cdot 10^{-3}$ of I_c has already been given in Sect. 2.4.4. The combined effect of noise and cross-talk on all inversion results is hard to determine. The inversion setup with constant parameters in optical depth has an advantage: the inversion code will "ignore" the random noise because of the limited number of degrees of freedom. Cross-talk effects on the derived inclination angle are discussed below.

Field strength The relation between the absolute field strength, B, and the separation of the two lobes in circular polarization is given by (cf. Eq. (3.4))

$$\Delta\lambda = \pm 4.67 \cdot 10^{-12} \cdot g_{eff} \lambda_0^2 B \,, \tag{3.29}$$

where B is given in Gauss and λ_0 in nm.

For the infrared line at 1564.8 nm with g_{eff} = 3, the spectral dispersion of 2.97 pm/pixel yields a peak separation coefficient of 0.01 pixel per Gauss. The spectral line broadening, due to the spectrograph and the thermal Doppler width, leads to a line width at half the line depth of around 8 pixel. A field of 500 G leads to a separation of the two V lobes of around 10 pixels, which can be clearly determined. Assuming the same accuracy of 0.2 pixel for the determination of the V lobe position as used in the section on the wavelength scale (cf. Sect. 3.3.5.2), a difference in field strength of 20 Gauss can be detected. For fields below about 400 G the 1564.8 nm is in the weak field limit, where the line broadening and the peak separation are comparable, and Eq. (3.29) is no longer valid. Thus, to be on the safe side, the accuracy in field strength is taken to be $\pm$ 50 G.

Field inclination The field inclination to the line of sight, γ, is roughly proportional to the ratio of total linear polarization, $L = \sqrt{Q^2 + U^2}$, and circular polarization, V, with

$$L/V \sim 0.5 \cdot \frac{\sin^2 \gamma}{\cos \gamma} \tag{3.30}$$

(e.g., Landi degl'Innocenti & Landi degl'Innocenti 1972; Jefferies et al. 1989; Landi Degl'Innocenti 2003).

Magnetic fields in the solar photosphere outside of sunspots are predominantly vertical with respect to the surface due to buoyancy forces, which put the field lines upright. Thus, for the heliocentric angles of the observations in this study of 6° and 30°, respectively, the field lines will generate a stronger polarization signal in Stokes V (cf. Sect. 3.1) than in linear polarization. For small inclination angles between LOS and field orientation, the linear polarization signal will be small and thus may be lost in the noise; the inversion code will underestimate the inclination in this case.

[6] rms $\equiv \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1..N} (x_i - \bar{x})^2}$.

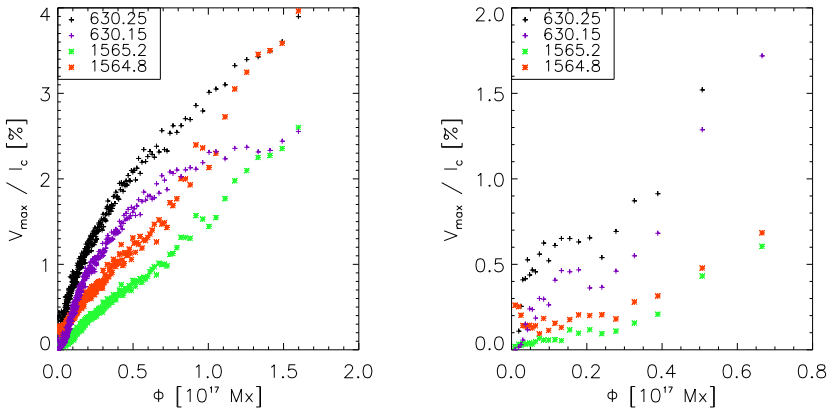


Figure 3.7: **Left**: The maximum Stokes V amplitude of the observed spectral lines as function of the magnetic flux, Φ. Only spatial positions with $B < 1400$ G were considered. **Right**: Same for field strengths above 1400 G.

In all other cases the inclination retrieved will be influenced by the relative polarimetric accuracy: cross-talk from circular to linear polarization may lead the inversion code to introduce an inclination, where there should be none (or vice versa); the relative amplitudes of linear and circular polarization are strongly influenced by the respective coefficients in the polarimeter response function (cf. Sect. 2.4.2), which scales the polarization of Q,U, and V independently of each other. A solution of Eq. (3.30) in terms of γ yields

$$\cos\gamma = -L/V \pm \sqrt{(L/V)^2+1}\,. \tag{3.31}$$

As numerical example, values comparable to actual observations are chosen: $L = 0.01$ and $V = 0.10$ in fractions of the continuum intensity, I_c. Taking twice the relative polarimetric accuracy of $5{\cdot}10^{-3}$, the worst possible case is given by $L_\pm = L \mp 0.01 \cdot V$ and $V_\pm = V \pm 0.01 \cdot L$. This leads to a variation of γ by less than 2°:

$$\gamma_\pm = \mathrm{acos}\left(-L_\pm/V_\pm \pm \sqrt{(L_\pm/V_\pm)^2+1}\right) = 25^\circ \pm 1.2^\circ\,. \tag{3.32}$$

Eq. (3.30) is strictly valid only for fully split (separation of σ-components $>$ line broadening) lines and constant magnetic field parameters. As upper limit therefore an error of 5° in inclination is estimated.

Magnetic filling fraction The maximum amplitude of the Stokes V signal, A_V, can be estimated in the weak-field limit from the field strength, B, the line-of-sight inclination, γ, and the *magnetic filling fraction*, β, by

$$A_{V,i} \propto a(B,\lambda_i)\,\beta B\cos\gamma\,, \tag{3.33}$$

where i indicates the spectral lines observed, and $a(B,\lambda_i)$ is a coefficient, which depends, e.g., on the field strength, the intrinsic properties of the respective transition or the spectral resolution of the instrument. Eq. (3.33) is commonly called the *magnetograph equation*, because earlier type instruments for measuring solar magnetic fields, the *magnetographs*, made use of it (e.g. Babcock 1953). Eq. (3.33) is valid if the splitting of the σ-components is still smaller than the line broadening (cf. Sect. 3.1). The V amplitude increases with field strength, because the overlap of the two σ-components of opposite sign in V is reduced, if they separate further.

For the inversion results, the magnetic filling fraction β can be calculated from the straylight contribution, δ, and the relative filling fraction, ε, by $\beta = (1-\delta)\cdot\varepsilon$ (cf. Eq. (3.27)). As field strength and inclination are also known, the flux $\Phi = \beta B\cos\gamma\cdot A_{\rm res}$, with the area of one resolution element, $A_{\rm res}$, can be derived. Φ can be substituted as first approximation on the right-hand side of Eq. (3.33).

As illustration of the effect, the Stokes V amplitudes of the different observed spectral lines at the same flux value are displayed in Fig. 3.7. In the left panel only fields below 1400 G were considered, to stay in the weak-field limit. The V amplitude increases with the flux, but the slope is decreasing for high flux values. The visible lines

B	γ	v	$\frac{\Delta\Phi}{\Phi}$	$\frac{\Delta\beta}{\beta}$
50 G	5°	150 ms^{-1}	14 %	10 %

Table 3.2: Error estimates for physical atmospheric parameters. For field strength, B, inclination, γ, and velocity, v, the absolute error range is given. For magnetic flux, Φ, and magnetic filling fraction, β, a relative error is stated.

(630 nm) show a stronger signal in V, because their line depth is much larger: the stronger absorption leads to a larger amount of induced polarization signal, if magnetic fields are present. The right panel of Fig. 3.7 displays that the relation is different for strong fields above 1400 G: the V amplitude is approximately constant regardless of the flux, which indicates that the lines are completely split. In the fully split case, the V amplitude is only directly proportional to the magnetic filling fraction.

The filling fraction is only used in the inversion of the spectra, where the inversion code derives $a(B,\lambda_i)$ from the atmospheric properties for each line. All spectral lines have to be reproduced at the same time with one value for β. Thus, re-writing Eq. (3.33), the problem to be solved in first order is:

$$\beta = \frac{< A_{V,i} >}{B\cos\gamma} , \tag{3.34}$$

where $< A_{V,i} >$ indicates that information of mainly the four most sensitive spectral lines is used. B is known accurately from the splitting of the 1564.8 nm line and γ from the relative fraction of linear and circular polarization. Standard error propagation for Eq. (3.34) yields

$$\frac{\Delta\beta}{\beta} = \sqrt{\left(\frac{\Delta < A_{V,i} >}{< A_{V,i} >}\right)^2 + \left(\frac{\Delta B}{B}\right)^2 + \left(\frac{\Delta\cos\gamma}{\cos\gamma}\right)^2} . \tag{3.35}$$

To estimate the relative error of the filling fraction, the data presented in the following chapter is used. For simplicity, it is assumed that the error of $< A_{V,i} >$ is half the error of a single line, which has an 1 %-error due to cross-talk ($\Delta < A_{V,i} >= 0.01 \cdot A_V/\sqrt{4}$). For the magnetic field, ΔB = 50 G was found above. The mean inclination to the LOS, γ, is around 30° for the data set; using the inclination error of 5° derived above the relative error, $\frac{\Delta\cos\gamma}{\cos\gamma}$, is around 5 %. Calculation of Eq. (3.35) for the about 22500 data points (x,y) with $B < 1400$ G, using the inversion result of $B(x,y)$, and the amplitude $A_V(x,y)$ of the 630.25 nm line, yielded an average value of

$$\frac{\Delta\beta}{\beta} \sim 10\% . \tag{3.36}$$

Magnetic flux The magnetic flux, Φ, is calculated by

$$\Phi = \beta B\cos\gamma \cdot A_{\mathrm{res}} . \tag{3.37}$$

Using the same approach and the same estimates as for the filling fraction, the error propagation yields an average relative error of

$$\frac{\Delta\Phi}{\Phi} \sim 14\% . \tag{3.38}$$

Velocity As discussed in the previous section, the determination of velocities from direct observables or the inversion is limited to about $\pm$ 100-150 ms^{-1}.

Spatial resolution Even if the spatial resolution of the observations is not an "error" by itself, it is an error source for the interpretation of the results. The polarimetric data has been taken with image stabilization, but without an Adaptive Optics system to correct for image degradation due to the Earth's atmosphere. The spatial resolution of the data is limited to around 0.8" ($\sim$ 600 km). The solar structures examined, magnetic elements outside of the sunspot or the penumbral fine structure, are in fact clearly below the resolution limit.

However, with polarimetry one has an advantage: if there is only one source of polarized radiation in a small area, one is able to "see" below the resolution limit. Straylight and seeing reduce the fraction of polarized light, but they

do not introduce additional polarization signals. Thus, the shape of the polarization signal, e.g., the position of the Stokes V lobes, of a single isolated source will be kept regardless of spatial smearing. In the penumbra of the sunspot the situation is worse, because all points in the surroundings emit polarized light. In this case, the additional information contained in the spectrally resolved polarization signal helps: the presumably geometrically confined filaments are present in a certain height range only. Thus, they leave a signature only in part of the spectrum, where information on their properties can again be retrieved from the shape of the polarization signal. However, in general one has to keep in mind for a direct interpretation of the results that the observations already are a spatial average over a certain area.

Table 3.2 summarizes the error estimates of the various atmospheric parameters.

Chapter 4

Magnetic elements in the sunspot moat

4.1 Introduction: The enigmatic G-Band Bright Points

Magnetic flux in the solar atmosphere can appear in various forms. Large concentrations, like pores and sunspots, are visible as dark features in the photosphere, because of the suppression of convection by the magnetic field (cf. Sect. 1.2.2.2). Paradoxically, the opposite is found for many small-scale flux concentrations. Their description as Bright Points or "filigree" dates back to the mid 70s (Dunn & Zirker 1973; Muller 1983). The brightenings are visible in many photospheric lines and in continuum windows. The visibility depends strongly on the heliocentric angle, θ: on disc center, no brightenings are seen in the intensity (cf. Fig. 1.7, p. 10), only close to the solar limb (cf. Fig. 1.8, p. 11 or Fig. 4.1). These brightenings are therefore also called *limb faculae*.

Interestingly, brightenings also appear in the spectral wavelength range termed *G band* by Fraunhofer, where a lot of spectral lines of the CH molecule can be found (cf. Berger & Title 2001). In the G band, the intensity strongly increases, when the CH molecules are thermally dissociated and their spectral lines disappear. These brightenings are usually very localized and were therefore called G-band Bright Points (BPs). Opposite to the continuum brightenings, BPs can be found at all heliocentric angles. It is still not clear if the two features – G-band BPs ($\neq f(\theta)$) and continuum brightenings ($= f(\theta)$) – are maybe the signature of the same physical effect, an enhanced temperature in the presence of magnetic fields.

BPs have been found to be related to (isolated) patches of magnetic flux both in observations (Berger & Title 2001) and in magneto-hydrodynamical simulations (Steiner et al. 2001; Carlsson et al. 2004), but the details are still under debate. BPs are used as tracers of magnetic field concentrations, if no polarimetric observations are available. As described in Sect. 1.2.1, it is generally assumed that most of the flux leading to BPs is in the form of stable kG flux tubes, which have formed by convective collapse of previously weaker and diffuse flux. The BPs are thus assumed to trace the elementary building blocks of the small-scale solar surface magnetism.

Some previous studies concentrated on their spectral signature (Langhans et al. 2002, 2004) or their dynamics by tracing their horizontal motions (van Ballegooijen et al. 1998; Bovelet & Wiehr 2003). However, up to now the properties of the magnetic flux connected to BPs could not be studied in detail from observations. Most times only co-spatial and simultaneous magnetograms were available, giving only the magnetic flux (cf. the previous section). In the present study, we have obtained simultaneous data from two spectro-polarimeters at the German Vacuum Tower Telescope, which measure the Stokes vector in visible (VIS) and infrared (IR) lines and allow a derivation of the vector magnetic field, together with G-band filtergrams[1] from the Dutch Open Telescope on La Palma.

[1] An image of the Sun through a narrow-band interference filter is called *filtergram*.

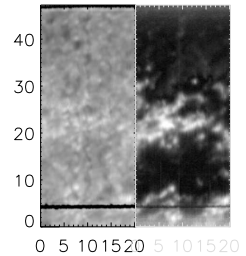


Figure 4.1: An example of faculae near the solar limb. **Left**: Continuum intensity map. **Right**: Integrated circular polarization. Tickmarks are in arcsec. The black horizontal stripes at the upper boundary of the image, respectively, around y = 4" are the hairlines on the POLIS slit that allow for spatial alignment in the data reduction procedure. The locations with stronger magnetic fields are brighter in intensity, not darker.

From the results of our observations, we want to caution that the terminology used for BPs may be inadequate. Brightenings in the G band are commonly addressed as individual separate features, which have "lifetimes", which "merge" or "split". The terminology suggests that the brightening itself is a physical entity capable of performing actions. We suggest a different view: BPs are due to a special configuration of magnetic fields. A flux concentration of sufficient field strength and total flux leads to a brightening in the G band. Both total flux and field strength of BPs cover an extended range and are not necessarily restricted to single flux tubes or small magnetic elements of kG field strength. The disappearance of a BP does not imply the disappearance of the flux by cancellation or submerging. In our data set, we find patches of magnetic flux, which persist during one hour, but show up as a BP only part of the time. Splitting and merging should also be seen as a reconfiguration of the magnetic field, not an "action" of the BPs. In general, we find that BPs not necessarily indicate a stable flux tube with kG fields. Only the reverse relation is true: a stable kG flux tube has an associated G-band Bright Point. This casts some doubt on the use of G-band BPs as tracers of magnetic field: BPs only indicate the presence of magnetic field, but give little information on the properties of the fields.

The chapter is organized as follows: After a short description of the observations (Sect. 4.2), the data alignment procedure is discussed in some detail. The Bright Points are identified in the DOT G-band data (Sect. 4.4). The inversion setup for this data set is described in Sect. 4.5. A statistical analysis of the properties of BPs in both direct observables and the inversion result is given in Sect. 4.6. Sect. 4.8 discusses selected examples of BP structure, BP formation, and flux disappearance. Section 4.9 summarizes the results.

element ion. state	λ [nm]	transition $^{2S+1}L_J$	g_j
POLIS			
solar absorption lines[c], CCD 2			
Ca II H	396.849	$^2S_{0.5} - ^2P_{0.5}$	1.33
Fe I	396.452	$^3P_1 - ^5P_2$	2.00
Fe I[f]	396.551	$^5D_3 - ^5D_4$	
Fe I	396.607	$^3F_2 - ^3D_3$	2.00
Fe I[f]	396.665	$^5D_0 - ^5F_1$	
Fe I[f]	396.681	$^3D_2 - ^5F_1$	
Fe I	396.745	$^3H_4 - ^1G_3$	0.75
Fe I	396.792	$^5D_3 - ^3G_4$	1.00
Fe I	396.918	$^3F_4 - ^3F_3$	1.50
Fe I[f]	396.963	$^3D_3 - ^3H_4$	
solar absorption lines, CCD 1			
Fe I	630.151[a]	$^5P_2 - ^5D_2$	1.67[b]
Fe I	630.250[a]	$^5P_1 - ^5D_0$	2.50
Fe I	630.346	$^5G_6 - ^5G_5$	
Ti I	630.375	$^3F_3 - ^3G_3$	
telluric absorption lines, CCD 1			
O_2	630.200[a]	-	-
O_2	630.276[a]	-	-

element ion. state	λ [nm]	transition $^{2S+1}L_J$	g_j
TIP			
Fe I	1564.7410	$^7D_2 - ^5P_2$	1.25
Fe I	1564.8515	$^7D_1 - ^7D_1$	3
Fe I	1565.2874	$^7D_5 - ^7D_4$	1.45

Table 4.1: Observed spectral lines in the blue and the red channel of POLIS (**left**); the same for TIP (**top**). The respective element and its ionization state are given in the first column. In the spectroscopic notation Fe I indicates neutral iron. a: rest wavelengths as given by Pierce & Breckenridge (1974); b: effective Landé factor from Solanki et al. (1987); c: all values in this section from J.Bruls (personal note) if not indicated otherwise. f: Nave et al. (1994). All spectral lines with a Landé factor $g_j \neq 0$ are sensitive to the presence of magnetic fields.

4.2 Coordinated observations on Tenerife and La Palma

We have observed the sunspot NOAA 04025 on the 9th of August 2003 during UT 9:36-10:34 with two spectropolarimeters at the German Vacuum Tower Telescope (VTT) at the Observatorio del Teide (Tenerife, Spain), the Polarimetric Littrow Spectrograph (POLIS), and the Tenerife Infrared Polarimeter (TIP, cf. Sect. 2.3). Both instruments were used simultaneously. The same sunspot was observed between UT 8:25-11:58 with the Dutch Open Telescope (DOT) at the Observatorio del Roque de los Muchachos (La Palma, Spain). The sunspot was located at an heliocentric angle of 27 deg ($\mu = 0.89$).

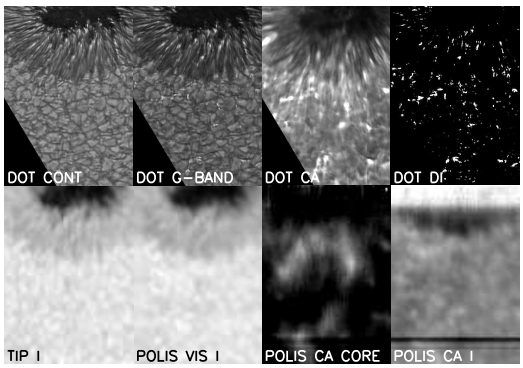


Figure 4.2: Maps of the first scan after coarse alignment of DOT and VTT data. **Top row, left to right**: blue continuum, G-band, and Calcium filtergram from DOT, difference image of G-band and continuum. **Bottom row, left to right**: continuum intensity in the IR from TIP, same in the visible from POLIS, intensity of the K_2V emission peak of the Calcium II H line, intensity in the Calcium II H line wing.

The image quality at the VTT was improved using the Correlation Tracker (CT, cf. Sect. 2.3.3) system. The CT compensates image motion, but it is not effective against the *seeing*, i.e. the image degradation due to disturbances of the refractive index in the Earth's atmosphere, which leads to a partial and time-dependent de-focussing and stretching of the solar image. The estimated spatial resolution of the VTT data is about 1 arsec.

The stepwidth of the scanning at the VTT was 0.35", with a total integration time of 6 seconds per scan step. The observations presented here consist of a scan of 70 steps (=24.5 arcsec) over the limb side of the sunspot and part of its surroundings, which was repeated 8 times. The cadence for the VTT data is about 7 minutes.

Table 4.1 summarizes the properties of the spectral lines observed at the VTT. For the infrared data and the visible channel of POLIS, the full Stokes vector of all lines is available. From the blue channel of POLIS (Ca II H at 396.85 nm, termed UV in the following) only the intensity profile could be used due to the low light level.

The data from the DOT consists of speckle-reconstructed filtergrams in the G-band at 430.5±0.5 nm, Ca II H at 396.8±0.06 nm, and blue continuum at 432.0±0.3 nm, with a cadence of 1 minute. The total field of view (FOV) is 92×73 arcsec^2, with a spatial sampling of 0.071" per pixel. More detailed information on the image processing and data reconstruction can be found in Sütterlin et al. (2004). After the speckle-reconstruction, the resolution reaches the diffraction limit of the telescope of 0.2" at 430.5 nm. Fig. 4.2 displays the maps of the co-spatial data of the first scan after a coarse alignment.

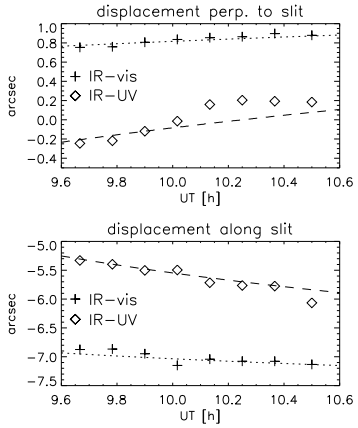


Figure 4.3: Displacement in arcseconds between infrared and visible (+), respectively, infrared and UV (⋄) intensity maps. For each repetition of the scan across the spot displayed in Fig. 4.2, the shifts perpendicular to the slit (**top**) and along the slit (**bottom**) have been calculated through an image correlation. The dotted (dashed) line are the shift values used for the alignment.

4.3 Spatial alignment of the data sets

For the investigation of the magnetic properties of the G-band BPs, which are identified using the DOT data, it is crucial to have a good alignment of the data sets from the two telescopes. Additionally, the simultaneous inversion of infrared and visible spectral lines requires the polarization signal in both wavelength ranges to be co-spatial and co-temporal.

The alignment uses the TIP resolution elements[2] as the reference system. It proceeds from a position $(x,y)_{TIP}$ in the TIP map, where x corresponds to the scan step and y to the position along the slit, to the co-spatial data points in the visible and UV channel of POLIS and the DOT filtergrams, respectively. The TIP data have a spatial sampling along the slit of $0.35''$ per pixel row. The data from the visible channel of POLIS have a finer spatial sampling of $0.145''$, which is degraded to the TIP resolution by the linear interpolation scheme described in the next section. The data from the VTT show displacements due to the actual position of each camera and an additional variable offset due to differential refraction in the Earth's atmosphere.

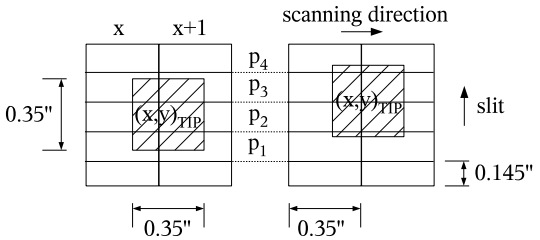


Figure 4.4: Calculation of the profile in the visible channel that is co-spatial to the infrared resolution element $(x,y)_{TIP}$. Two cases are possible: the IR resolution element can correspond to 3 (**left**) or 4 (**right**) visible profiles along the slit. The profiles p_i from both scan steps, x_{vis} and $x_{vis}+1$, are added with weights that correspond to the fraction of the POLIS resolution element covered.

4.3.1 Alignment between IR and VIS/UV data

Calculation of shifts. To find the co-spatial points of the TIP and the POLIS data, the continuum intensity maps of each scan across the spot have been correlated separately. For the UV channel, a map of the intensity in the wing of the CaII H line is used, which shows photospheric structure (cf. Fig. 4.2). The resulting shift values in the direction along and perpendicular to the slit for the 8 scans allow to fit a curve to the displacements (cf. Fig. 4.3). The best-fit curve is used to determine the shift in x and y for each scan step.

Application of shifts. For the UV channel of POLIS the spatial sampling along the slit of $0.29\ ''$ per pixel is similar to TIP. The shift in pixel, $(\Delta x_{UV}, \Delta y_{UV})$, is simply added to the position $(x,y)_{TIP}$, and the closest resolution element is taken:

$$(x,y)_{UV} = \text{round}\left[(x,y)_{TIP} + (\Delta x_{UV}, \Delta y_{UV})\right] , \tag{4.1}$$

where the rounding automatically selects the closest co-spatial resolution element.

For the visible channel of POLIS, the spatial sampling is almost three times better than for TIP. The co-spatial information is retrieved by a weighted average along and perpendicular to the slit. First, the co-spatial position $(x,y)_{vis}$ is given by:

$$(x,y)_{vis} = (x,y)_{TIP} + (\text{floor}(\Delta x_{vis}), \Delta y_{vis}) , \tag{4.2}$$

where "floor(Δx_{vis})" rounds to the largest integer smaller then x_{vis}.

The TIP resolution element will then cover a fractional part of visible profiles from two scan steps, x_{vis} and $x_{vis}+1$ (cf. Fig. 4.4). The co-spatial Stokes profile $p(\lambda,x,y)_{vis}$ is calculated from the weighted average

$$p(\lambda,x,y)_{vis} = \frac{\sum_i a_i p(\lambda,x,y_i)_{vis}}{\sum_i a_i} + \frac{\sum_i b_i p(\lambda,x+1,y_i)_{vis}}{\sum_i b_i} , \tag{4.3}$$

[2]As a reminder, "resolution element" is used as synonym for a single spatial position (x,y). The usage of "pixel" is restricted to spatial or spectral properties of individual CCD cameras and the DOT data, where only 2-D imaging without spectral information is available.

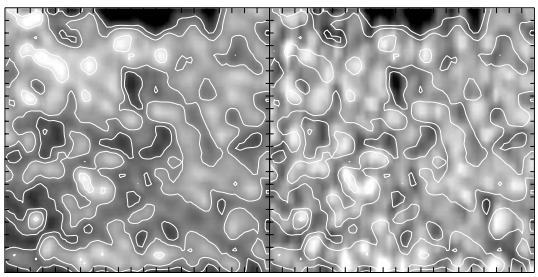


Figure 4.5: The lower part of the continuum intensity map of the first repetition, **Left** : POLIS visible channel @630 nm. **Right**: IR @1.5 μm. The contour lines outline bright granules and dark intergranular lanes as observed in the visible channel. Tickmarks are separated by 1 arcsec.

where i= 1...3 (or 4). The weights, a_i and b_i, are set corresponding to the fraction of the POLIS resolution element covered by the TIP resolution element.

Fig. 4.5 displays the co-aligned infrared and visible continuum intensities of the lower part of the first scan. The agreement of the spatial structures is remarkably good. The IR data has a slightly better resolution, because this wavelength is less affected by atmospheric disturbances.

The accuracy of the final alignment of TIP and the visible POLIS channel should be similar to the scatter of the observed shifts in Fig. 4.3 around the assumed shift curve, i.e. about 0.1 arcsec or roughly a third of a TIP pixel. The maximum displacement perpendicular to the slit in the observations is 1 arcsec, corresponding to a maximum time gap between co-spatial profiles of 18 seconds (3 scan steps of 0.35" with 6 sec integration time each).

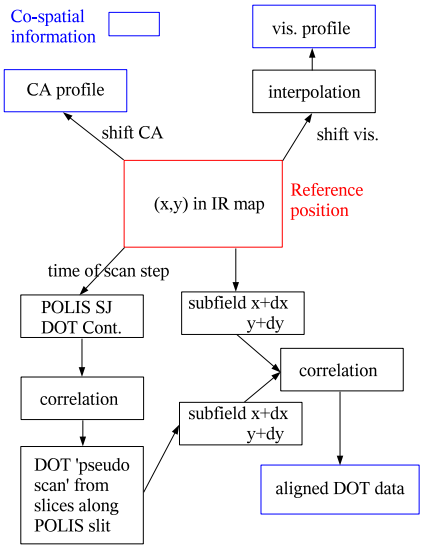


Figure 4.6: Flowchart of the alignment procedure: From the reference position $(x,y)_{TIP}$ in the TIP data set (**red rectangle**), the co-spatial information is derived (**blue rectangles**). For the POLIS profiles, the shift values for the visible (UV) channel are added to $(x,y)_{TIP}$. For the visible channel of POLIS, the co-spatial spectrum is derived by a linear interpolation along the slit, because of the finer spatial sampling. For the UV data, the closest resolution element is taken. For the alignment with the DOT data, the POLIS slit jaw image and the DOT continuum image closest in time are correlated and used for the creation of a so-called *pseudo scan*. The final alignment is made by taking a subfield of the TIP map around $(x,y)_{TIP}$, the corresponding subfield from the DOT pseudo scan, and then correlating the subfields to find the residual displacements.

4.3.2 Alignment between DOT and VTT data

In this case co-spatial *and* co-temporal positions have to be found. First a co-temporal map from the DOT time series is constructed (*pseudo-scan*), which subsequently is spatially aligned to the VTT data.

4.3.2.1 Co-temporal DOT map

In order to create the pseudo-scan image, i.e. the image that would result when stepping the POLIS/TIP slit across the DOT field of view (FOV), the slit-jaw images of POLIS are used. They display a 100×100 arcsec^2 FOV of the solar surface centered on the slit. Each slit-jaw image is correlated with the DOT continuum image closest in time, which is degraded to the same spatial resolution before. From the correlation one finds the shift in x and y for the co-alignment of these images from the two telescopes. The POLIS slit from the slit-jaw image is superimposed on the co-aligned DOT data. A slice of 5 pixel width (= 0.355 arcsec) is taken from the DOT images at the slit position, and placed accordingly in the pseudo-scan map. Examples of the resulting pseudo-scan maps for the three DOT channels are displayed in Fig. 4.2.

4.3.2.2 Co-spatial DOT map

The final alignment of the DOT maps is achieved by taking a 25×25 pixel subfield of the TIP map around each resolution element, $(x,y)_{TIP}$, the corresponding area from the DOT pseudo-scan continuum map, and again correlating these two subfields. This procedure yields the position of a 5×5 pixel area in the DOT maps that is co-spatial to the TIP resolution element. The area is cut out from the 3 DOT pseudo-scan maps (G band, blue continuum, Ca II H) and placed in new maps accordingly. The flowchart in Fig. 4.6 summarizes the steps for the alignment of both VTT and DOT data. Fig. 4.7 displays the co-aligned maps of the first repetition of the scan. Only the lower part of the FOV without the sunspot is shown. The intergranular lanes visible in the DOT map can clearly be traced

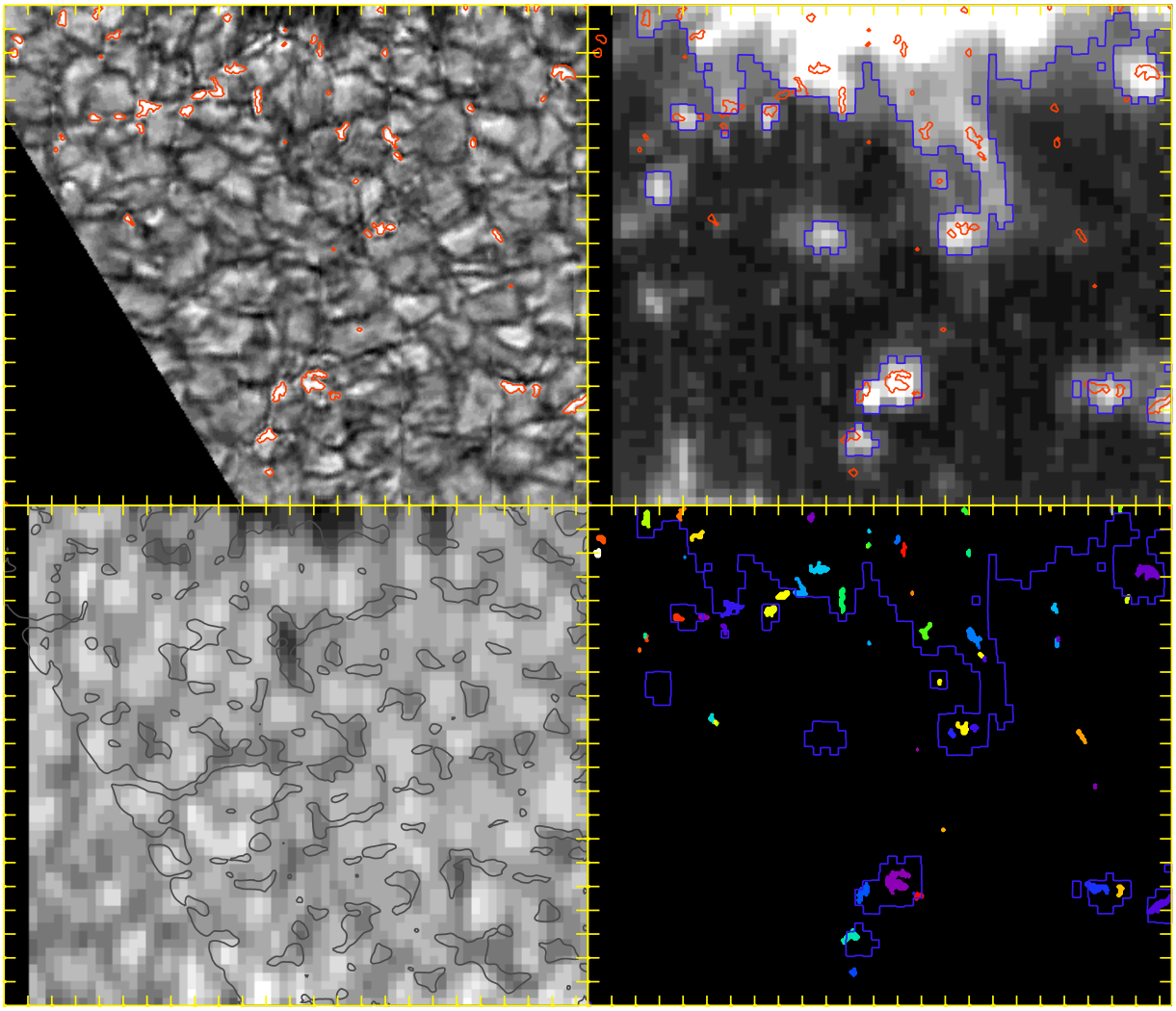


Figure 4.7: **Left column, top:** G-band intensity from DOT. **Left column, bottom:** IR continuum intensity. Tick-marks are separated by 1 arcsec. Despite the worse spatial resolution, the intergranular lanes visible in the DOT map can clearly be traced in the IR intensity map (black contours). **Right column, top:** Total integrated polarization. **Right column, bottom:** DOT Bright Point (BP) mask. The individual BP areas are color coded according to their consecutive number. Blue contour lines outline strong polarization signals in the canopy of the spot, and of individual patches in the moat. Most of the identified BPs outlined by the red contours are co-spatial with increased polarization signal.

in the IR continuum map. In general, the time gap between the different data sets is always less than 1 minute, and the spatial mis-alignment between VTT and DOT data should be at most one TIP pixel, i.e. 0.35".

Appendix C contains all aligned maps of the DOT G-band, infrared intensity, total polarization degree and the BP mask for the analyzed section of the FOV (Figs. C.1 and C.2).

4.4 Identification of Bright Points

To identify BPs in the DOT data, the mean intensity of the maps in G-Band and blue continuum is equalized by multiplying the blue continuum map with the ratio of the mean intensities. The two maps are then subtracted. Only strong relative enhancements of the G-band intensity are left over. The map after subtraction will be referred to as *difference image* (DI) in the following. An example can be found in Fig. 4.2.

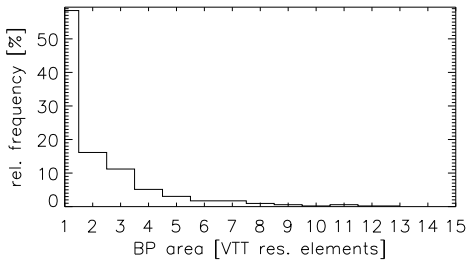


Figure 4.8: Histogram of the BP area in the binary mask for the polarimetric data from the VTT. Most bright structures in the G band do not extend over more than one VTT resolution element (0.35 arcsec^2).

The BPs proper are taken to be all closed contours in the DI above a threshold of 1450 counts. Some very short contour lines have been excluded, to avoid spurious signals of 1-5 pixels extent, which can be created during the alignment procedure. No further criterion on contrast or shape of the areas is applied. The closed contours are consecutively numbered, which allows to address individual BPs (cf. the color code in Fig. 4.7). In total, 538 separate BP-structures were identified in the full data set. Our selection of bright areas by sharp thresholding leads to extended patches, which would perhaps be split up by a more sophisticated procedure like the one proposed by Bovelet & Wiehr (2003).

A re-scaled mask for the polarimetric data from the VTT is constructed by marking all VTT resolution elements, which contain one or more bright DI pixels. A control sample of Non-Bright Points (NBPs) is taken from all resolution elements, which are at least three pixels (=1.05 arcsec) apart from any BP and not inside the canopy of the spot (cf. Sect. 1.2.2.2) at the upper boundary of the FOV. The NPB mask of the VTT is displayed in Fig. C.1 for the first repetition of the scan. The BP masks of DOT and VTT data thus allow the discrimination between spatial positions with and without significant G band enhancements.

Due to the five times coarser spatial sampling of the VTT data, even extended BPs in the DOT data correspond in most cases to only a single resolution element of 0.35 arcsec squared (cf. Fig. 4.8).

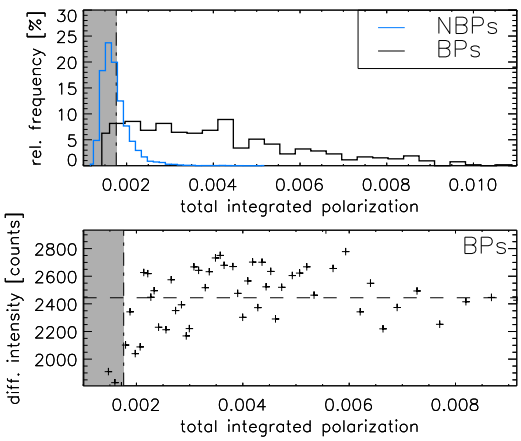
4.5 Inversion setup: magnetic elements

The targets in this study consist of presumably unresolved isolated magnetic elements, found predominantly in also unresolved intergranular lanes. For spatial positions, where the profiles showed significant polarization above the threshold (cf. Sect. 3.2), the SIR inversion code was therefore set to use a 2-component inversion with one field-free atmosphere, one magnetic atmosphere, and an additional straylight contribution. The field-free component is supposed to represent the intergranular lane with mainly redshifts, the straylight mimics the contribution the granular surroundings with convective blueshifts, and the magnetic atmosphere represents the magnetic flux concentration. All points without a polarization signature were also inverted. At these positions, only one field-free atmosphere component without straylight contribution was used. Table 4.2 summarizes all free parameters.

parameter	unpol.	pol., 1st comp.	pol., 2nd comp.
temperature, T	x	x	x
velocity, v	x	x	x
field strength, B	-	-	x
inclination, γ	-	-	x
azimuth, χ	-	-	x
microturb., v_{mic}	x	x	x
macroturb., v_{mac}	x	0	0
straylight, β	-	0	0
fill. fraction, δ	-	0	0

Table 4.2: Parameter settings for the inversion of the unpolarized (one-component model, 2nd column) and the polarized profiles (two-component model, 3rd and 4th column). x: free parameter, - : not used, 0 : same parameter value for both model atmospheres.

The inversion determines the set of free parameters as discussed in Sect. 3.3.4: the temperature stratification with optical depth for each component, $T(\tau)$, the magnetic field vector, given by field strength, B, the inclination angle to the line of sight, γ, and the field azimuth, ϕ, and the respective line-of-sight (LOS) velocities of the atmospheric components, $v_{\rm LOS}$.



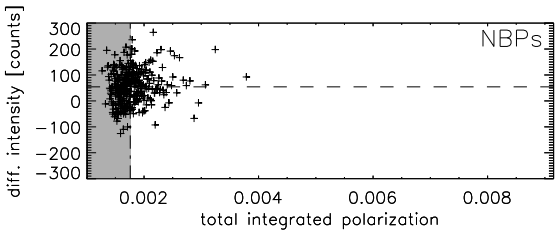


Figure 4.9: **Left top**: Histogram of the relative frequency of total integrated polarization (TP) for BPs and NBPs. **Left bottom:** Scatterplot of TP vs. the difference intensity for BPs and NBPs (**right top**). Dashed horizontal lines are the mean intensities. The grey shaded area denotes the noise level.

4.6 Results of the statistical evaluation

Our results include observables derived directly from the profiles (cf. Sect. 3.2) as well as the atmospheric properties from the inversion. In some cases, we can use values from both approaches for cross-checking the consistency, while in other cases they are complementary. The results for the Bright Points are compared with the corresponding values for all Non-Bright Points (cf. Sect. 4.4 for the respective definition).

4.6.1 Polarization signature of Bright Points

A visual inspection of Figs. 4.7, C.1 or C.2 suggests that G-band Bright Points are co-spatial with polarization signal. To quantify this impression, the total integrated polarization (TP, cf. Sect. 3.2) can be used. The TP is calculated separately for the four spectral lines at 603.15/25 nm and at 1564.8/1565.2 nm by

$$\mathrm{TP} = \int \frac{\sqrt{Q^2+U^2+V^2}(\lambda)}{I_c} d\lambda \,, \tag{4.4}$$

and then averaged. The TP is in first order proportional to the magnetic flux. All magnetic field lines in the resolution element contribute to the TP, either in linear or circular polarization.

The noise level of the TP is defined from all profiles, which were unpolarized and were not inverted with a magnetic atmospheric component. The histogram of the relative frequency of TP values (top right panel of Fig. 4.9) displays that most BPs (94 %) have a TP above this range, whereas 2/3 of the NBPs have a polarization signature within the noise limits.

The difference intensity between G-Band filtergram and the blue continuum filtergram is used for the identification of the BPs. Fig. 4.9 also displays the difference intensity of BPs and NBPs as a function of the total integrated polarization. We find that the difference intensity of BPs depends on the total polarization only for small polarization levels, which are of the order of integrated noise. For a TP higher than 0.003 the BP difference intensity is approximately constant with the TP.

The graph for the NPBs shows no systematic trend at all and stays close to a DI value of zero, which suggests that our control sample resembles a different kind of solar structures, even if polarization signal is also present.

From the maximum polarization degree of the 1454 profiles co-spatial to BPs we again find the same fraction of polarized profiles, with only 6 % being unpolarized. This suggest that the chosen polarization threshold for single profiles was a good criterion to separate between noise and signal.

The examination of the shape of the Stokes V profiles of BPs and NBPs gives additionally information on the structure of the magnetic fields. We find that 80 % of the polarized BP profiles are of a regular type with two lobes in Stokes V, and 20 % irregular with only one or three lobes. For the NPBs the fraction of unpolarized and irregular profiles is higher. Table 4.3 summarizes the spectral signature for BPs and NBPs.

From the high correlation of BPs and polarization signal above noise level we conclude that all G-Band brightenings are co-spatial to magnetic fields to less than 0.35". The profiles are preferentially of regular type, which indicates single-type structures without multiple magnetic components.

spectral line	1564.8	1565.2	630.15	630.25
BP profiles, 1454 in total				
unpol.	2.3	9.1	10.3	4.3
polarized	97.7	90.9	89.7	95.7
regular	32.3	75	86.9	85.3
blue only	0.7	2.7	0.7	2.5
red only	0.5	0.5	1.0	2.6
irregular	66.5	21.8	11.4	9.6
NPB profiles, 9497 in total				
unpol.	33.7	72.8	74.5	48.6
polarized	66.3	27.2	25.5	51.4
regular	49.4	53.7	77.7	42.4
blue only	9.4	25.0	13.3	39.3
red only	2.4	5.5	5.9	14.6
irregular	38.8	15.8	3.1	3.7

Table 4.3: Relative fraction of polarized and unpolarized profiles for BPs and NBPs. The polarized profiles are additionally classified depending on the number of lobes in Stokes V: a profile is called regular if it has 2 lobes, irregular (3 or more lobes), or one lobed (blue/red lobe at shorter/larger wavelength than the intensity minimum). A profile is unpolarized if the maximum polarization degree is less than the threshold for the respective spectral line. The IR line at 1564.8 nm is blended with a magnetically sensitive line, which leads to the high number of irregular profiles.

4.6.2 Magnetic field strength of Bright Points

The magnetic field strength, $|\mathbf{B}|$, can be calculated from the separation of the Stokes V lobes for fully split spectral lines as

$$|\mathbf{B}| = \frac{\Delta\lambda}{4.67 \cdot 10^{-12} \cdot g_{eff} \cdot \lambda_0^2}, \tag{4.5}$$

where $\Delta\lambda$ is half the wavelength separation of the two lobes in Stokes V (cf. Fig. 3.3).

Fig. 4.10 displays a scatterplot of the BP field strength from the inversion and the field strength derived from the splitting of different spectral lines. The splitting of the IR line at 1564.8 nm (+) serves as a good approximation for magnetic fields down to 500 G. For 1565.2 nm (□), and especially the visible lines, there is a strong deviation for fields below at least 1200 G. The same result with a larger statistical base is derived from the NBP control sample. We conclude that in the inversion the field strength in the magnetic atmosphere is determined mainly from the splitting of 1564.8 nm, in agreement with the expectations: this line has the strongest sensitivity to magnetic fields and is fully split around 500 G. The inversion code has to adjust the field strength to reproduce the position of the Stokes V lobes of the 1564.8 nm, in order to reduce the deviation between observed and synthetic profiles.

The relative frequency of the field strength values from both approaches is displayed in Fig. 4.11 for both BPs and NPBs. Again, the inversion and 1564.8 nm give essentially the same result. We find that BPs range from 500 G to about 1500 G with an essentially "flat" distribution. The field strength is not limited to kG values, but covers moderately strong to strong fields with no preferred value.

For the field strength in our NBP sample we find a distribution from the inversion (or 1564.8 nm), which points at an exponentially decreasing probability for higher field strength. The same was found in the study of Khomenko et al. (2003) for intra-network fields in quiet Sun from IR data alone. It is noticeable that the spot does not seem to influence the distribution of the weak fields in the close-by area.

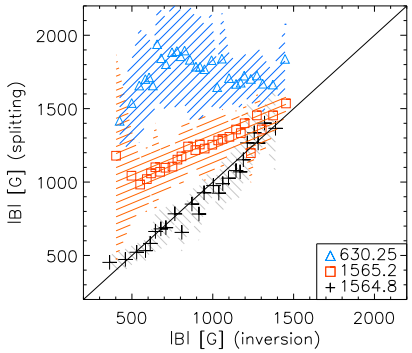


Figure 4.10: The magnetic field strength for the BPs is derived by two methods, from the peak separation of the Stokes V lobes, and the inversion. The correlation is close for the inversion result against the splitting of 1564.8 nm (+), slightly deviating for inversion against 1565.2 nm (□), and strongly deviating for inversion against 630.25 nm (△). The colored areas give the scatter inside the bins.

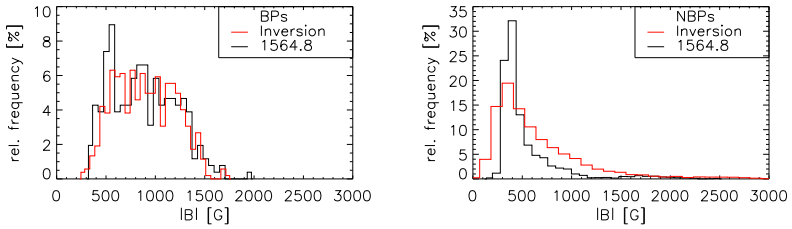


Figure 4.11: **Left:** Relative frequency of field strength for BPs from the inversion and the splitting of the IR line at 1564.8 nm. **Right:** same for the NBP sample.

4.6.3 Line-of-sight velocities of Bright Points

An estimate for the line-of-sight velocity inside the magnetic atmosphere is given by the zero-crossing velocity, v_{zcro} (cf. Sect. 3.2). The corresponding parameter from the inversion is the velocity of the magnetic component. For the velocities in the surroundings of the magnetic elements and for unpolarized profiles, the line core velocity or the velocity of the field-free inversion component can be used. By convention, positive velocities correspond to downflows here.

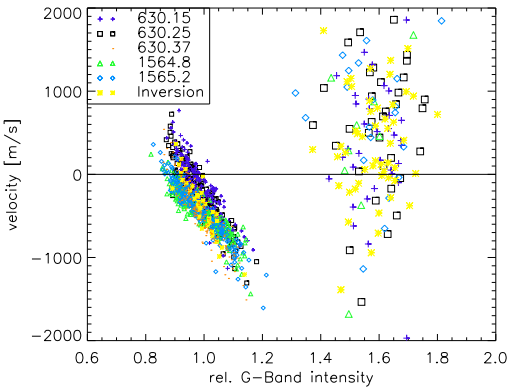


Figure 4.12: LOS velocities against G-Band intensity, positive velocities correspond to downflows. Small symbols: LOS velocity of line core and non-magnetic inversion component for the NBP control sample. Large symbols: zero-crossing velocity and LOS velocity of magnetic inversion component for the BPs. The spectral lines are represented with symbols of different shape.

Fig. 4.12 displays a scatterplot of velocities, which were derived from the spectra of each resolution element from the VTT, vs. the corresponding intensity in the G band from the DOT. For the NBP sample (small symbols), we find that both the velocity derived from the position of the core of the spectral lines and that of the field-free inversion component are in good agreement with the expectation for an area dominated by granulation. The bright granules are correlated with blue shifts, while the intergranular lanes have reduced intensity and show downflows.

Contrary to that, no trend of the velocity inside the magnetic field with intensity is found for the BPs. The velocity range of BPs spans about ± 2 kms^{-1} for the binned data. The extreme velocities, when considering only points with a clear polarization signal above 1 %, reach ± 4 kms^{-1}. In general, there is a preference for downflows in the BP magnetic atmosphere.

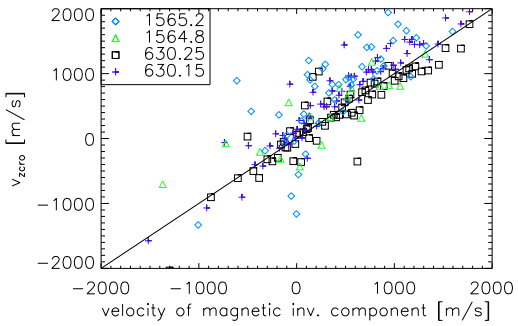


Figure 4.13: The zero-crossing velocity, v_{zcro}, of the spectral lines as a function of the velocity of the magnetic inversion component for BPs. The closest correlation is seen between the inversion and the 630.25 nm line (black squares) because of the steep slope, which this line has around the zero-crossing in Stokes V.

The velocity of the magnetic inversion component is usually in closer agreement with v_{zcro} of the visible lines than

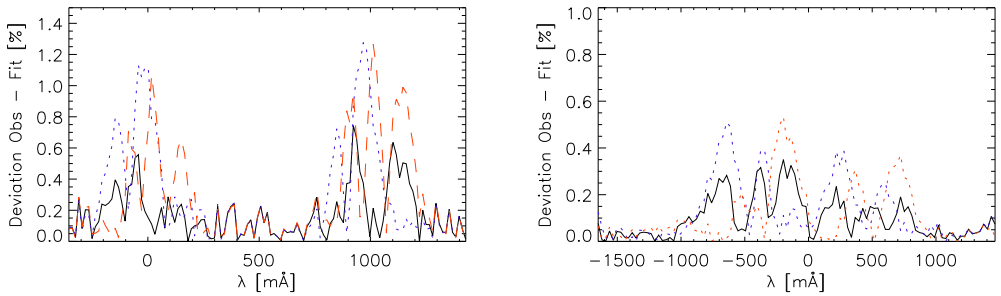


Figure 4.14: The deviation between observed and best-fit profiles with varied line-of-sight velocity for the visible lines (**left**) and the IR line at 1564.8 nm (**right**). **Black**: difference between observed and original best-fit profile. **Red/blue**: difference between the observed and the best-fit profile, which was shifted by $\pm 1\,\mathrm{kms}^{-1}$. The deviation increases much stronger for the visible lines than for the IR.

that of the IR lines (cf. Fig. 4.13). The close agreement with the visible lines is due to their steep slope in Stokes V around the zero-crossing wavelength. For visualization, Fig. 4.14 displays the absolute deviation between an observed profile and the best-fit profile for three cases: the deviation between observed and original best-fit profile with the "correct" velocity and the same after the best-fit profiles have been shifted by ± 1 kms^{-1}. For the visible lines, the shift approximately doubles the deviation, whereas for the IR line at 1564.8 nm the increase and the absolute value of the deviation is smaller. Therefore, the least-squares minimization technique of the SIR code will yield a close match with the visible lines. Contrary to the field strength, where the IR lines are more important, the line-of-sight velocity is mainly given by the visible lines – despite the fact that the velocity resolution of the visible lines is worse (cf. Tab. 3.1).

In Fig. 4.13 it can also be seen that both IR lines shows stronger downflows in v_{zcro} than the visible lines, where the IR line at 1565.2 nm usually shows the largest velocities. Note that the IR line at 1564.8 nm has a much worse statistics, as v_{zcro} is only calculated for regular profiles (cp. Table 4.3, only 32 % regular). To investigate the difference between IR and visible lines, Fig. 4.15 shows the average v_{zcro} from both IR lines as function of the average velocity of the visible lines. The IR lines show a tendency for higher flow velocities, which suggests the presence of velocity gradients in the atmosphere.

Table 4.4 summarizes the mean velocities of BPs and NBPs. For BPs we find an average red-shift between 400 m/s and 800 m/s in the magnetic part of the atmosphere. The number is in agreement with, e.g., the results of Amer & Kneer (1993) or Sigwarth et al. (1999). Even if the mean v_{zcro} of all spectral lines indicates redshifts, the mean line core velocity of the BPs partly shows blueshifts. The magnetic elements and intergranular lanes are not resolved in the VTT data and thus are contaminated with the unpolarized straylight of the immediate surroundings, which show a convective blueshift. For the NBP sample, we find average velocities very close to the "default" velocity for the spectral lines used in the determination of the wavelength scale. The default velocity of each line is the predicted convective blueshift from the 2-component quiet Sun model of Borrero & Bellot Rubio (2002). The NBP sample is thus very similar to a quiet Sun area, which is dominated by granular dynamics.

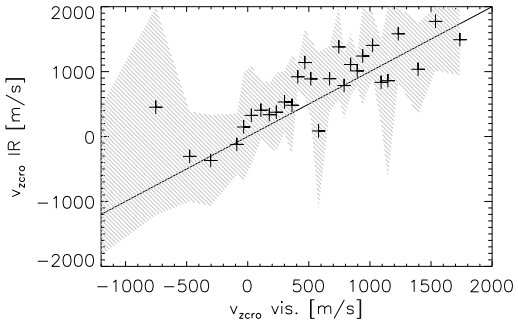


Figure 4.15: The zero-crossing velocity, v_{zcro}, of the IR lines as a function of v_{zcro} of the visible lines. The IR lines tend to show a larger redshift, indicative of a velocity gradient with depth. The grey shaded area denotes the scatter.

line	630.15	630.25	630.37	1564.8	1565.2
BPs, magn.	504	661	-	404	821
BPs, line core	31	22	-107	-266	-231
NBP, line core	-172	-250	-451	-448	-447
QS, default	-185	-262	-424	-445	-469

Table 4.4: LOS velocities for BPs and NPBs in m/s, positive values correspond to downflows. 1st row: mean zero-crossing velocity of BPs for the respective lines. 2nd row: mean line core velocity of BPs. 3rd row: mean line core velocity of NPBs. 4th row: the convective blueshift from the 2-comp. QS model of Borrero & Bellot Rubio (2002) used to set up the wavelength scale.

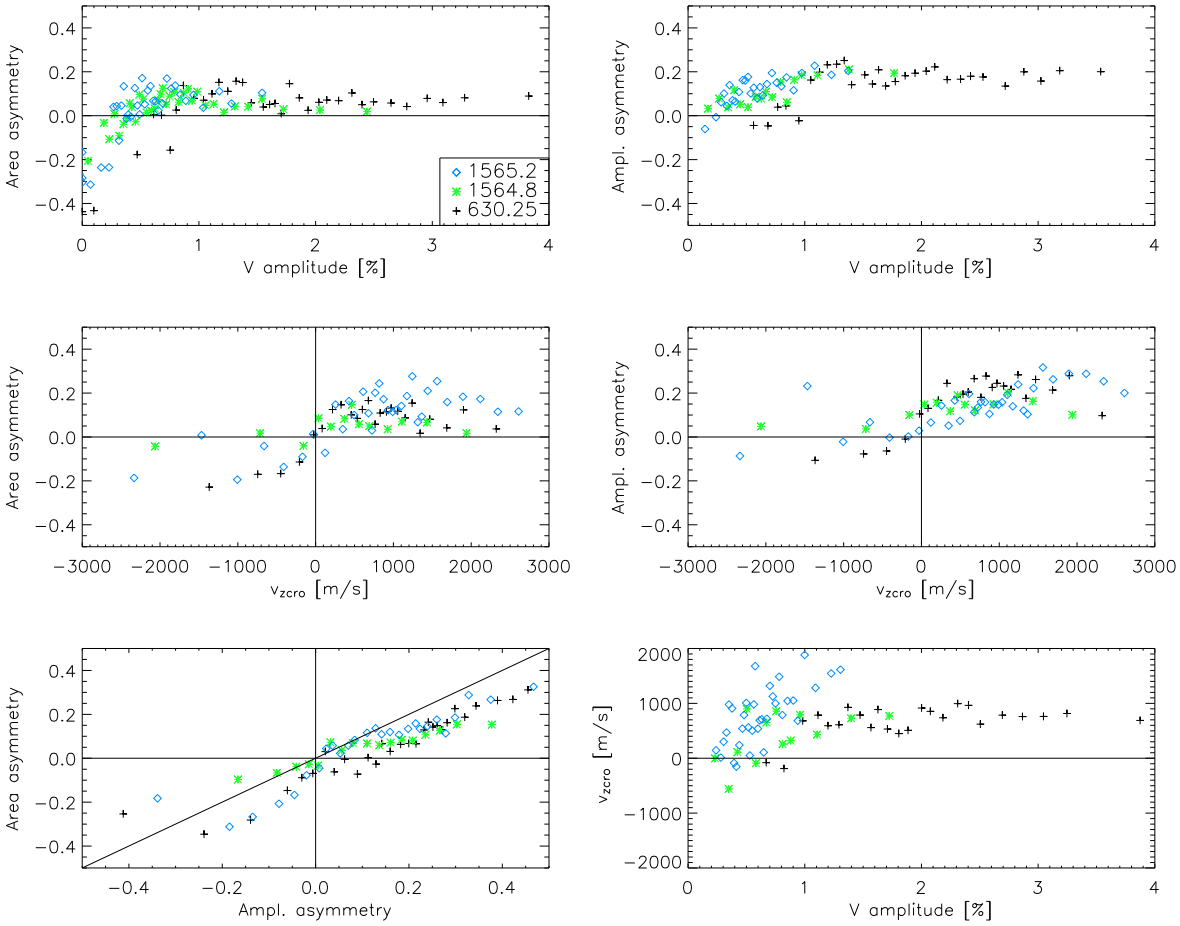


Figure 4.16: **Left to right, top to bottom :** dA and da vs. the Stokes *V* amplitude, dA and da vs. v_{zcro}, scatterplot of dA vs. da, and scatterplot of V amplitude vs. v_{zcro}. The data is taken from the BP sample only.

4.6.4 Area and amplitude asymmetries of Bright Points

A second approach to examine the flow velocity inside the magnetic elements is given by the amplitude asymmetry, da, and the area asymmetry, dA, of the Stokes V signal (cf. Sect. 3.2, p. 33). The area asymmetry is connected to gradients along the LOS in velocity and field strength by (Solanki & Pahlke 1988):

$$\text{signum}(\text{dA}) \propto \text{signum}\left(-1\frac{\text{d}|\text{B}|(\tau)}{\text{d}\tau}\frac{\text{dv}(\tau)}{\text{d}\tau}\right). \tag{4.6}$$

Magnetic flux concentrations are assumed to be in horizontal pressure balance with their surroundings. In the stratified solar atmosphere, there is a strong reduction of density and temperature with height above the surface, leading to a decrease of gas pressure. The field strength thus should decrease with height as the flux is spread out over a larger area. It has also be shown that velocity gradients alone are sufficient to produce asymmetries (López Ariste 2002). Additionally, discontinuities along the LOS can produce asymmetries, e.g., a jump of velocity and field strength at the border of a flux concentration. The asymmetries thus are a sensitive tool to derive information on the stratification of atmospheric properties.

In Fig. 4.16 we display the correlation between the asymmetries and other parameters for the BPs. We find that most BPs exhibit a positive area and amplitude asymmetry.

Negative asymmetries are only encountered for weak polarization signals. The zero-crossing velocity, v_{zcro} changes sign with both da and dA, i.e. there are very few examples of, e.g., blueshifted profiles with positive asymmetries. We find zero-crossing velocities of 0 to +2 kms^{-1} with almost constant asymmetry values. The zero-crossing velocity does also not seem to depend on the strength of the polarization signal, as similar velocities are found for most BPs with stronger signal. Area and amplitude asymmetry are correlated to some degree but not totally. The non-zero asymmetry values support the presence of velocity gradients in the atmosphere.

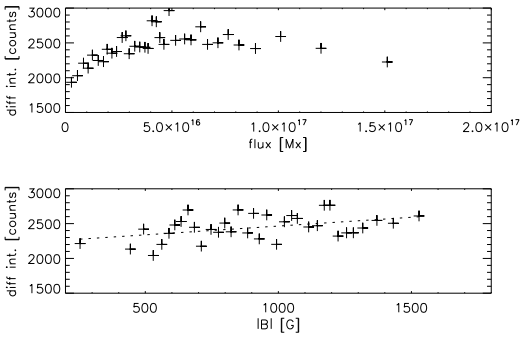


Figure 4.17: **Top panel:** The intensity difference between G-band and blue continuum (DI) vs. the total unsigned magnetic flux, Φ. The DI increases with flux up to around $1.5{\cdot}10^{17}$ Mx. For higher flux values it decreases again, similar to the shape in the total integrated polarization (cf. Fig. 4.9). **2nd panel:** Difference intensity against field strength, B. Dashed line: linear fit to the data points.

4.6.5 Relations between temperature, flux, field strength and difference intensity

The dependence of the difference intensity on magnetic flux and field strength In the upper panel of Fig. 4.17 the relation between the G-band difference intensity and the total unsigned flux is displayed. The DI increases only slightly with magnetic flux for values below $5{\cdot}10^{16}$ Mx. It stays almost constant over a large range and drops again slightly for fluxes above $1.5 \cdot 10^{17}$ Mx. The range of flux values for BPs we find is comparable with Berger & Title (2001). The relation of DI and flux is very similar to that of DI and the total integrated polarization (cf. Fig. 4.9, p. 56). This is a good sign of consistency of the inversion results: the TP is proportional to the flux and is derived directly from the observed profiles. The flux however is calculated using three different inversion parameters – the field strength, the magnetic filling fraction and the line-of-sight inclination – but shows the same shape in the relation to DI.

For the field strength (lower panel), we find that there is a general but slight increase of DI with higher field strength.

The temperature dependence of the difference intensity The relation between the temperature of the inversion components and the DI are plotted in the top left panel of Fig. 4.18. The temperature of the respective atmospheric component has been averaged in the optical depth range of $\log\tau$ = 0 to -2. The DI shows no clear correlation

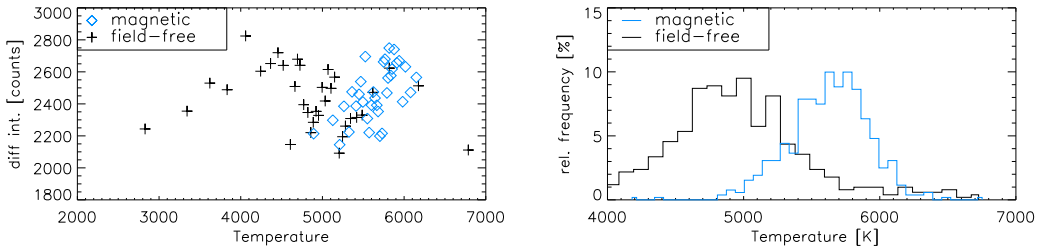


Figure 4.18: **Left:** Difference intensity vs. temperature for both atmospheric components in the inversion, the non-magnetic surroundings (+) and the magnetic atmosphere (◊). **Right:** Histogram of the temperatures. The magnetic component is usually hotter than the surroundings by about 1000 K at the same optical depth.

with the temperature of the magnetic or the field-free component. However, the two inversion components show a systematic difference in their temperature value. The histograms of the relative frequency of magnetic and field-free atmosphere temperature in the top right panel reveal that the magnetic atmosphere is hotter by about 1000 K at equal optical depth. The mean temperature is 5800 K for the magnetic component and around 4800 K for the field-free surroundings.

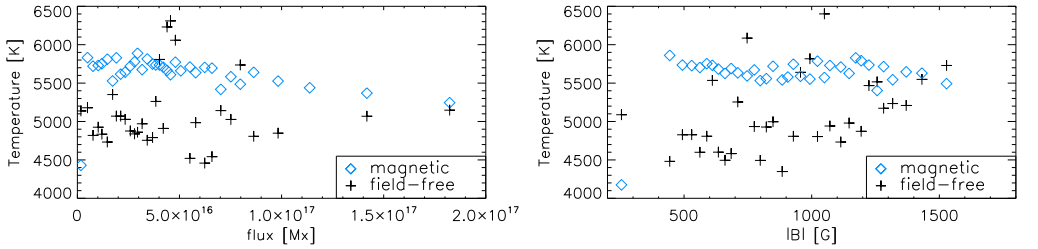


Figure 4.19: **Left:** Absolute magnetic flux vs. temperature for both inversion components. The temperature is decreasing linearly with the flux for $\Phi > 10^{17}$ Mx. **Right:** Magnetic field strength against temperature. A slight decrease is visible. For the field-free component, no clear trend with neither Φ nor B can be seen.

The influence of field strength and flux on temperature The mutual influence of hydrodynamic parameters on the magnetic field is also of general interest outside the question of the BP properties. A strong correlation is found between total magnetic flux and temperature (cf. lower left panel of Fig. 4.18). For the small structures under examination, a reduction of the temperature of the magnetic component with increasing flux is found. The correlation between field strength and temperature shows a similar behavior, but the reduction of temperature for stronger fields is marginal.

4.6.6 Line-of-sight inclination of Bright Point magnetic fields

To examine the dependence of BP intensity on the inclination of the magnetic field to the line of sight, the inclination from the inversion, γ, and the ratio of linear and circular polarization, L/V, can be used. The L/V ratio is more affected by the noise in the data, especially for small amplitudes of either L or V. For fully split lines these two variables are related by (Landi Degl'Innocenti 2003):

$$L/V \propto 0.5 \cdot \frac{\sin^2\gamma}{\cos\gamma}. \tag{4.7}$$

The upper left panel of Fig. 4.20 shows a scatterplot of the L/V ratio of 1564.8 nm line vs. that of the 630.25 nm line. The values in both spectral ranges are consistent; the only significant deviations are seen for more inclined

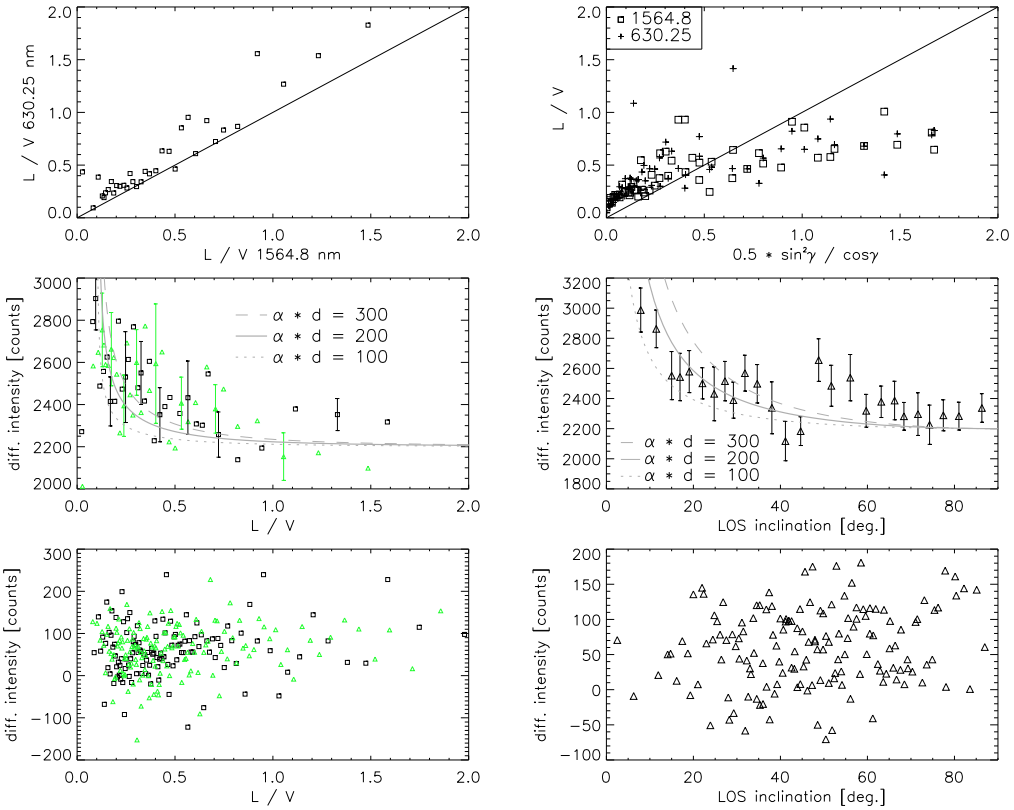


Figure 4.20: **top, left**: L/V of 1564.8 nm vs. that of 630.25 nm. **top, right:** L/V of 1564.8 nm (□) and 630.25 nm (+) vs. the corresponding inversion parameter, $0.5 \cdot \sin^2\gamma/\cos\gamma$. **middle, left**: DI vs. L/V for 1564.8 nm (△), and L/V averaged over both visible lines (□) for BPs. **middle, right**: LOS inclination vs. DI for BPs. Overplotted are curves calculated from Eq. (4.11) for 3 values of α·d. Error bars correspond to the error of the mean value of the binned data. **Bottom left**: Same as above for NBP and all spectral lines, 1565.2 nm (◊), 630.15 nm (+). **Bottom, right:** Same as above for NBP.

fields. The same is true for the validity of Eq. (4.7): we find that the approximation is valid even for the visible lines in the weak field limit for small inclinations between LOS and magnetic field (top right). The L/V values are slightly larger than those predicted from the inclination due to the noise, which prevents the quantity $L = \sqrt{Q^2+U^2}$ to reach zero.

Interestingly, the brightness in the G band seems to depend stronger on the inclination than on any other parameter discussed up to now. In the scatterplot of DI vs. L/V or DI vs. inclination from the inversion in the middle panel of Fig. 4.20, we find a strong correlation of the largest G band enhancement with magnetic field parallel to the LOS. BPs with a large inclination between LOS and the field lines have a systematically reduced intensity. We believe this trend to be significant, even if the DI shows additional scatter due to the intrinsic properties of the different BPs like field strength or flux, which also have some effect on DI. The overplotted curves are discussed in section 4.7 below. For the reference sample we do not find any clear trend of DI in neither L/V nor γ (bottom panels).

4.6.7 The size of Brigth Points in DOT and VTT data

The filling fraction of the magnetic atmosphere component denotes the fraction inside a single VTT resolution element occupied by magnetic flux. It can be converted to an effective diameter, d, of a circular structure with the same area. The upper limit of d for a single resolution element is 292 km. However, some BPs extend over more than one resolution element. The size of the BPs in the VTT data is thus calculated from the average filling fraction of the respective BP in all resolution elements covered, multiplied by the number of resolution elements.

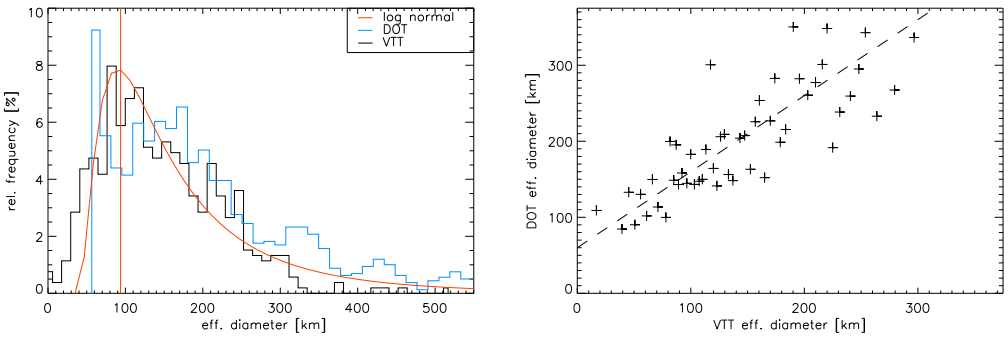


Figure 4.21: **Left**: Histogram of the effective diameter of BPs from the VTT (black) and DOT data (blue). The red line gives the maximum of the log-normal distribution (red) at d=93 km. **Right**: The effective diameter from the VTT vs. that from the DOT. The dashed line has a unity slope and 60 km offset.

For the DOT, the number of bright pixels belonging to a single BP structure is converted to an effective diameter. Note that for the DOT data, a BP has to have a minimum of 60 km diameter ($\equiv$ 1 bright DOT pixel), otherwise it would not have been identified as BP. For the VTT, no lower limit exists, because the magnetic filling fraction can go smoothly towards zero.

Fig. 4.21 displays a histogram of the BP diameter in the both data sets. We find a distribution with a mean value of 150 km for the VTT and 210 km for the DOT and few large-scale structures above 300 km diameter. The maximum diameter found is 530 km for VTT data and 700 km for DOT data. For the DOT data, there is an increase of the relative frequency for small structures corresponding to 1 or 2 bright pixels (60-100 km). These will most probably be artifacts of the data processing. The decrease of the number of BPs with a diameter below 100 km, which is seen in both DOT and VTT data, therefore seems to be significant (cf. Wiehr et al. 2004). The found size of BPs is comparable to the width of intergranular lanes.

The right panel of Fig. 4.21 displays the effective diameter from the VTT data vs. the corresponding value from the DOT for the same BP. Taking into account that a) the data is from two different telescope and that b) the effective diameter for the VTT data is derived indirectly from the filling fraction of the magnetic inversion component, the agreement is good. Unfortunately, there is an offset of 60 km, which corresponds exactly to the effective diameter of a single DOT pixel. We were unable to trace down the offset to an error in the calculation of the effective diameter and can offer no explanation for it. Besides the offset, for each BP to first order the same size is found in both data sets.

4.7 Discussion of the statistical evaluation

It is difficult to merge all the findings into a single model of BP structure. Nevertheless, the amount of parameters available from our multi-wavelength observations allows at least to restrict the possible explanations for the origin of G-band BPs. A G-band BP primarily outlines a deficiency in the abundance of the CH molecule. The amount of CH in the solar atmosphere depends on both density and temperature (Steiner et al. 2001; Langhans et al. 2002). The hotter the atmosphere, the more CH molecules dissociate. The G-band intensity then increases due to the disappearance of the molecular spectral lines.

Some of the statistical results can be predicted from the following theoretical considerations. Firstly, the presence of magnetic fields leads to a decrease in density as the lateral pressure balance requires that

$$p_{ext}(\rho_{ext}, T_{ext}) = p_{int}(\rho_{int}, T_{int}) + \frac{B^2}{8\pi}, \tag{4.8}$$

where B is given in Gauss, p denotes the gas pressure, and ext (int) denotes the exterior (interior) of a flux concentration. Radiation from outside into the flux concentration leads to a lateral temperature equilibrium for each geometrical height h, $T_{ext}(h) = T_{int}(h)$. The only option to ensure pressure balance then is a reduction of ρ_{int}.

Due to the reduced density in the magnetic atmosphere, a point of same optical depth, τ, in the magnetic and field-free atmosphere corresponds to different geometrical heights. Thus, for a given value of τ, deeper layers with

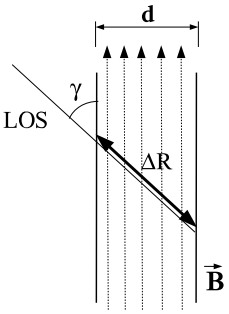


Figure 4.22: Sketch of the geometry of an isolated flux concentration in field-free surroundings. The line of sight (LOS) is at an angle, γ, with the boundary of the vertical flux. The path length of the ray inside the flux, ΔR, depends on the diameter, d, and γ, by $\Delta R = d/\sin\gamma$.

higher temperature are seen in the magnetic part of the atmosphere (Bellot Rubio et al. 2000b; Schüssler et al. 2003).

The temperature of the interior of the flux concentration is determined by the radiative equilibrium with the surrounding granulation, which participates in the convective energy transport. T_{int} should depend on the radius, r, of the flux concentration: for an infinitesimal height interval, dh, the area of the hot granulation seen from inside is given by $A_{hot} = \pi \cdot r \cdot dh$, whereas the volume to be heated is $V_{int} = \pi \cdot r^2 \cdot dh$.

This is in agreement with the findings of section 4.6.5: the temperature decreases with increasing flux. As the field strength of BPs is limited to below 1.4 kG (cf. Fig. 4.11), a larger total flux will occupy a larger area, leading to an increase of the radius r, and hence, a decrease of T. On the other hand, the magnetic atmosphere component at equal optical depth is found to be hotter than the field-free surroundings, which is in agreement with the change of the optical depth scale due to the reduced density in the flux concentration.

The G-band intensity or the relative G-band enhancement given by the difference intensity shows two trends: A negative correlation with flux and a positive correlation with field strength. The first point is the same as the temperature-flux relation. For the G-band, the reduced temperature leads to less dissociation of the CH molecules, and thus a lower G-band intensity. For field strength, it is the opposite: stronger fields lead to a reduced density and shift the optical depth scale downwards. Due to the lateral temperature equilibrium and the temperature stratification, deeper layers with a higher temperature and hence, a stronger dissociation of the CH molecules contribute to the spectrum and lead to an increase of the G-band intensity and the DI.

A natural explanation for the dependence of the G-band intensity on inclination between LOS and field lines is given by assuming the geometry depicted in Fig. 4.22: a vertical flux tube of diameter d is embedded in field-free surroundings. A light ray, which hits the outer boundary of the flux at an inclination angle, γ, has a path length, ΔR, inside the flux tube, which is

$$\Delta R = \frac{d}{\sin\gamma}. \tag{4.9}$$

The contrast enhancement (CE) of the G-band with respect to the nearby continuum wavelength will be proportional to the path length in the interior of the flux concentration, where the density is reduced. Two effects may play a role: less absorption, because of the reduced abundance – together with the total number density – of CH in the flux concentration, or less absorption, because deeper layers with higher temperature, and hence less CH, are seen:

$$CE = \alpha \cdot \Delta R(\gamma) = \frac{\alpha \cdot d}{\sin\gamma}\,. \tag{4.10}$$

α is a conversion coefficient with the unit counts/length of an unknown numeric value; it contains, e.g., the number of photons, which originate from this layer of the solar atmosphere and the efficiency of their detection at the DOT telescope.

After crossing the flux concentration, the line of sight will finally hit the boundary layer between flux and the granular surroundings or reaches a depth, where the atmosphere inside the magnetic flux concentration becomes opaque. In both cases, almost no CH is present there due to the increased temperature. This yields an almost constant contribution, I_0. The difference intensity between G band and blue continuum then can be written as:

$$DI(\gamma) = I_0 + CE = I_0 + \frac{\alpha \cdot d}{\sin\gamma}, \tag{4.11}$$

The maximum contrast enhancement is limited by the maximum path length, ΔR_{max}[3] that the line of sight can

[3]It may be larger than d.

travel, before it encounters an opaque atmospheric layer at some depth due to the increase of density even for normal incidence.

Using Eqs. (4.7) and (4.11) one can calculate the G-Band intensity as a function of either γ or L/V, $\alpha \cdot \mathrm{d}$, and I_0. I_0 is adjusted to the observed DI value for large inclinations, where ΔR is smallest and its contribution can presumably be neglected.

The middle row of Fig. 4.20 displays the observed relationship between L/V and γ and the difference intensity DI. With a value for I_0 of 2200 counts, as suggested by the intensity for large inclinations, curves of the G-band intensity for $\alpha \cdot \mathrm{d} = 100$, 200 and 300 have been calculated and overplotted. Even with the large scatter in the data and the simple modeling, we find that we can reproduce the dependence on inclination sufficiently. If the conversion coefficient, α, would be known, a direct measurement of the diameter would be possible.

The model leads to a natural explanation of enhanced G-band intensity for both granular edges and isolated brightenings inside intergranular lanes, when magnetic field is in the line of sight – if the amount of flux and the field strength are sufficient to see layers of significant reduced CH abundance. If the field lines are parallel to the line of sight, the deepest layers with the highest temperature and the smallest CH density are seen. Outside of the flux concentration, the opaque atmospheric layers are reached higher up in the atmosphere, because of the larger density and stronger absorption, and hence, lower temperatures without a dissociation of CH are seen.

4.8 Single case studies

4.8.1 The spatial structure of an isolated Bright Point

Some of the G-band BPs are transient and disappear during the observations. However, some are persistent for more than one hour, presumably because they are connected to larger concentrations of magnetic flux. Fig. 4.23 displays an example of the latter type. The G-band brightening, and, even more, the emission in the chromospheric Ca II H line is co-spatial to an increased polarization signal. The whole structure has a diameter of about 1000 km in the chromospheric emission and the VTT data sets, while in the G-band it shows an internal substructure down to less than 200 km extension. The BP is located at the end of the canopy field of the sunspot and moves away from the sunspot. The shape of the structure in G-band and Calcium data from the DOT changes, while in the VTT data the patch shows no spatial evolution besides horizontal motion.

The Stokes profiles of a cut through approximately the center of the BP in the first repetition of the scan are displayed in Fig. 4.24; the full set of profiles are plotted in Fig. D.1 and Fig. D.2 in the Appendix. It can clearly be seen that the canopy signal, characterized by linear polarization with a weak field strength and small splitting, extends up to the location of the BP. The BP itself then shows magnetic fields with a much higher field strength, which are nearly parallel to the LOS.

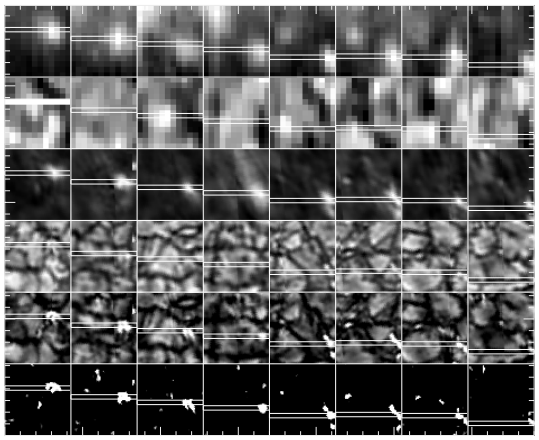


Figure 4.23: Example of a BP persistent during the observation time of 1 hr. **Top to bottom**: total integrated polarization, IR continuum intensity, DOT Ca II H filtergram, DOT blue continuum, DOT G-Band, difference intensity. Tickmarks correspond to 1 arsec, full FOV is 4.2x4.7 arcsec^2. The scanned area unfortunately ends at the right border of the subfields. The horizontal white bar in the first column in the IR intensity map denotes the orientation of the cut through the BP displayed in Figs. 4.24 and 4.25. The other white lines denote the position of the BP at later times.

The various observation and inversion parameters along the cut are displayed in Fig. 4.25. They reflect the statistical results rather well:

- The maximum integrated polarization matches the shape of the co-spatial Calcium emission.
- The relative area and amplitude asymmetries of the spectral lines are positive around 20 %.

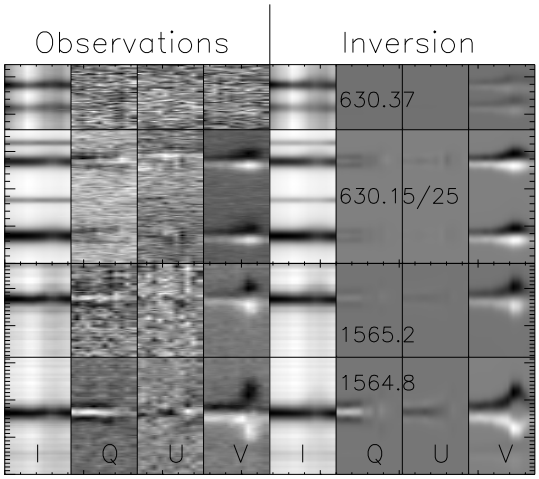


Figure 4.24: **Left**: Observed Stokes profiles along the horizontal cut marked in Fig. 4.23. **Right**: The corresponding best fit profiles from the inversion. **Top to bottom:** visible lines, IR lines. Horizontal tick marks correspond to 1 arcsec. At the location of the BP, the more horizontal field indicated by the linear polarization, mainly Stokes Q, vanishes. The field strength is more than doubled, and the field is almost parallel to the LOS. The profiles of the cut are displayed individually in Fig. D.1 and D.2.

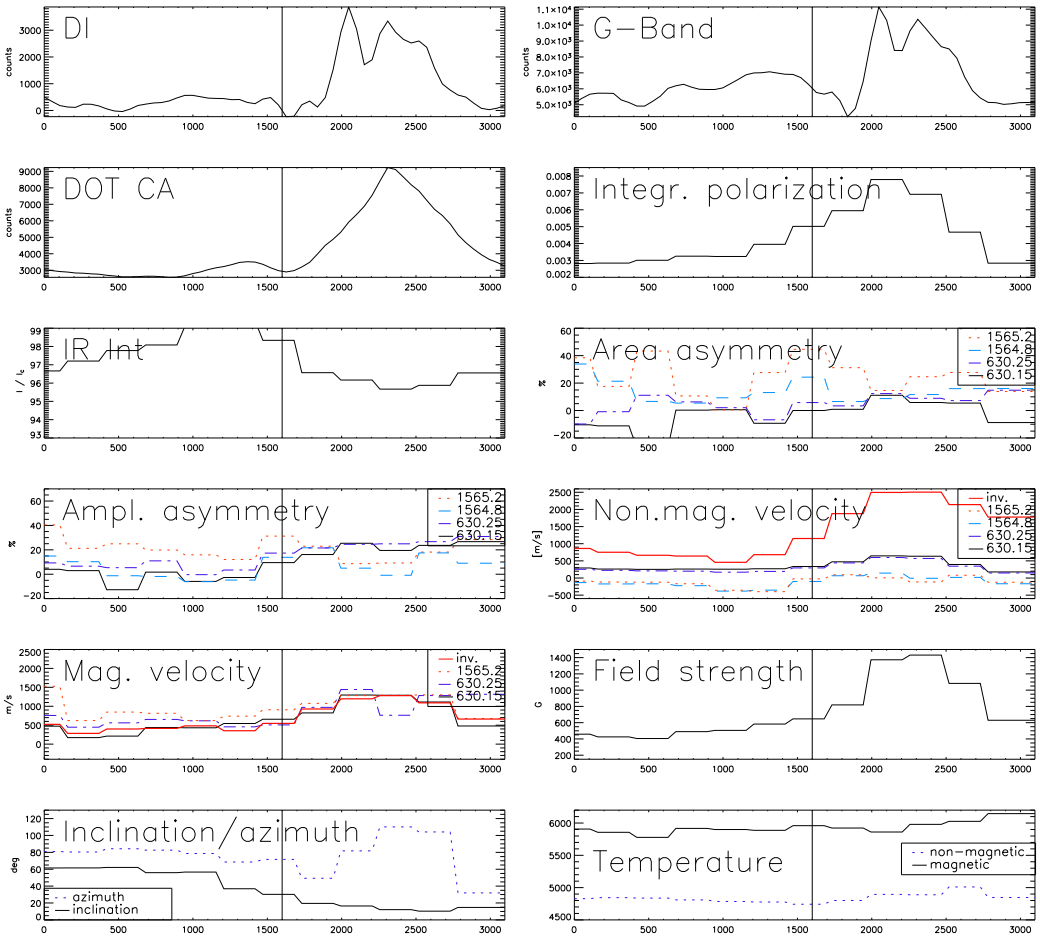


Figure 4.25: Observation parameters along the cut through the BP marked in Fig. 4.23. The horizontal axis is in km. The BP is located from x = 1600 km to 2800 km. The maximum intensity in G-band or Calcium is co-spatial to the largest polarization signal. The BP is inside an intergranular as is evident from the intensity in IR and the non-magnetic velocities, which are red-shifted. The minimum inclination is slightly displaced from the polarization of field strength maximum, but co-spatial to higher temperature.

- The velocity inside the BP magnetic field, as derived from v_{zcro} or the magnetic inversion component, is far more redshifted than the immediate surroundings. The canopy field shows a redshift as well.

- The temperature of the magnetic atmosphere component is about 1000 K hotter than the non-magnetic part.

The shape of the curve for the LOS inclination is smooth without discontinuities at the border of the BP. The BP has the opposite polarity as the umbra of the sunspot and appeared directly at the sunspot boundary in the DOT data, some time before the maps shown here. Because of its close proximity to the spot, the BP could be the endpoint of a sunspot filament, which dips down into the photosphere again. In this case, the observed redshifts may correspond to a steady downflow of material inside the BP, where the Evershed flow along the filament supplies the necessary mass. However, most other BPs are farther away from the spot and most probably not connected to any filament, excluding this explanation for most of the observed downflows.

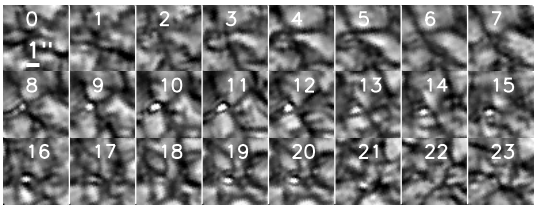


Figure 4.26: Development of a G-band Bright Point in the DOT time series. The area shown is 5".3 squared. The time is given in minutes. The BP appears at t = 8 min and persists for about 15 min.

4.8.2 The formation of a Bright Point

In the previous sections, all G-band Bright points were found to be co-spatial to magnetic field. However, the opposite relation does not hold: Not all magnetic fields lead to brightenings in the G-band. A prominent example can be found in the first scan. A large patch of magnetic flux near the center of the field of view in Fig. 4.7, p. 54, shows no associated BP in the co-spatial DOT map. Some minutes later, a BP appears at this location. An interesting question is, how this magnetic flux differs in the beginning from other patches and how it evolves with time to finally generate a BP.

Fig. 4.26 displays the temporal evolution in the DOT G-band time series with 1 min cadence. Up to t = 7 min nothing special can be seen in the G-band, but at t = 8 min a BP appears, which persists for about 15 min afterwards. The same area in the polarimetric data from the VTT in Fig. 4.27 clearly shows the difference of the magnetic field parameters before and after the formation of the BP. At t = 0 and 7 min the patch of magnetic flux, outlined by its strong polarization signal, shows only relatively weak diffuse field with field strengths around B = 400 G. At t = 7 min at the lower border of the patch an increase of B can be seen. It is co-spatial with a strong downflow in the map of the line-of-sight velocity of the magnetic inversion component. The velocity of up to 2 kms^{-1} is not only the signature of the intergranular lane at the border of the magnetic patch, but indicates a significantly increased downflow inside the magnetic field. At t = 8 min the BP appears in the DOT data (cf. Fig. 4.26). However, due to the lower cadence the magnetic field map is only available again at t = 14 min. In the polarimetric data it can be seen that the now existing BP has a field strength around 900 G. Its mean flux per resolution element is above $2 \cdot 10^{16}$ Mx. The area of the magnetic patch is greatly reduced. The BP is co-spatial to enhanced emission in the Ca II H filtergram.

The difference of the magnetic field properties before (t = 7 min) and after (t = 14 min) the BP formation can best be seen by a comparison of values along the two cuts marked in Fig. 4.27. Before the BP formation at t = 7 min, the magnetic flux along the cut is around $2 \cdot 10^{16}$ Mx throughout the patch (cf. Fig. 4.28). The concentration process can be seen to start at the border of the patch, where the field strength and the line-of-sight velocity strongly increase, but outside the patch at a low flux value. At t = 14 min, the G-band contrast is nearly doubled at the location of the BP (around x = 4"). In both the field strength and the magnetic flux a co-spatial strong increase is found. The field strength reaches around 900 G and the mean flux per resolution element exceeds $2.5 \cdot 10^{16}$ Mx. Thus, the BP appears after the diffuse magnetic flux has been concentrated above a threshold value.

The concentration process may have been the *convective collapse* of an already existing flux concentration with B below 1.3 kG (cf. Sect. 1.2). Two time scales are of interest in the process: the generation of a flux concentration and the later field amplification by the collapse, which leads to the appearance of the Bright Point.

Simulations of convective collapse In the 2-D magneto-hydrodynamical simulations of Grossmann-Doerth et al. (1998), a homogeneous initial seed field of 400 G was introduced in a simulation box (their model C1). They found that after about 3 minutes a flux concentration was formed by the convective motions, which developed in the simulation area. However, the flux concentration was instable. A slight downward displacement of material inside the flux concentration led to an accelerated downflow, evacuating the flux concentration. Such an evacuation leads to a localized drop of the total pressure inside the field. The only option to regain a horizontal pressure equilibrium

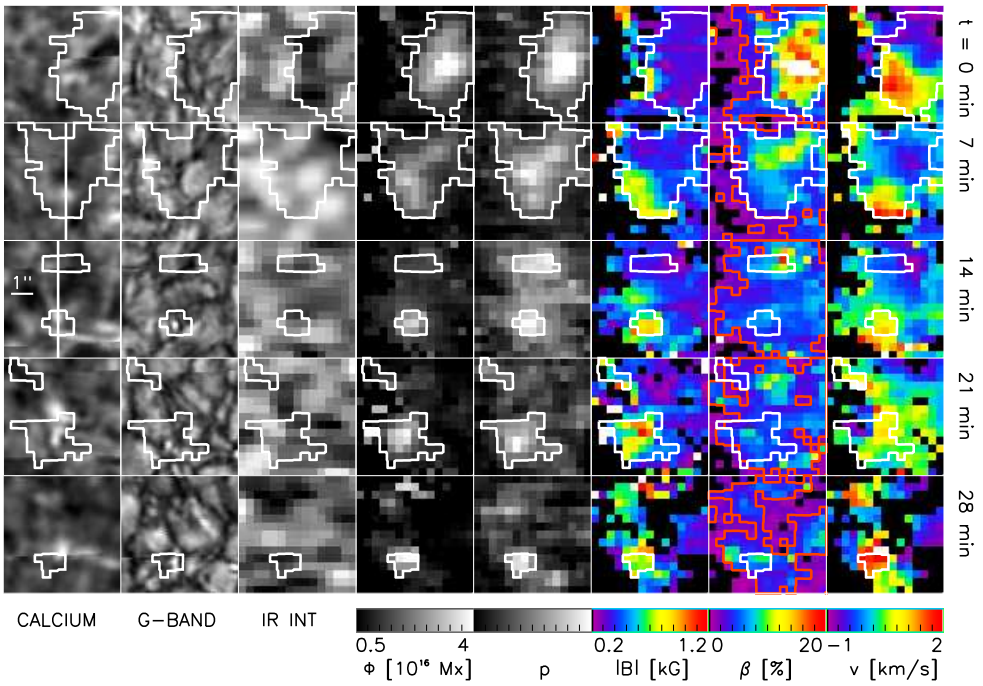


Figure 4.27: Formation of a G-band Bright Point from a patch of diffuse flux. **Left to right**: DOT Ca II H filtergram, G-band filtergram, infrared continuum intensity, magnetic flux (Φ), total integrated polarization (TP), field strength (B), magnetic filling fraction (β), line-of-sight velocity (v). The temporal evolution is displayed from **top to bottom** with 7 min cadence. White contours outline areas of increased TP. Red contours (column 7) denote a magnetic filling fraction of 2.5 %. All inversion parameters from resolution elements with less magnetic filling fraction have been set to zero. The IR intensity map at t = 7 min has been smoothed to allow for a direct comparison with the DOT G-band data. The two vertical white lines in the 1st column at t = 7 and 14 min indicate the location of the cuts shown in Fig. 4.28.

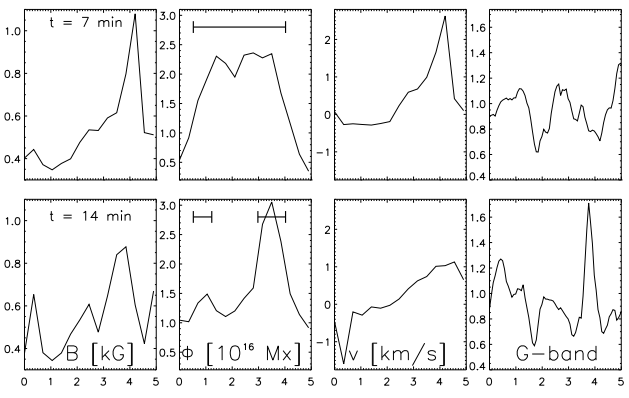


Figure 4.28: Cut through the magnetic flux patch before (**top row**) and after (**bottom row**) the BP formation. From **left to right** the field strength (B), the flux (Φ), the line-of-sight velocity (v), and the relative G-band contrast are shown along the lines marked in Fig. 4.27. The black horizontal bars in the 2nd column outline regions of enhanced flux. At t = 7 min the concentration process can already be seen to start at the right boundary of the magnetic patch around x = 4". At t = 14 min the BP is clearly visible in the G-band data, co-spatial to a strong (900 G) magnetic field with a mean flux per resolution element of more than $2.5{\cdot}10^{16}$ Mx.

with the surroundings is to increase the magnetic pressure, which is proportional to the field strength by $B^2/4\pi$. The increase in B is reached by a reduction of the area of the flux concentration. The evacuation and the following field amplification were found to happen in less than 1 minute in the simulations. After the collapse, the continuum intensity was found to increase at the location of the flux concentration due to the change of the optical depth scale (less absorption in the evacuated flux concentration). The final flux concentration in the simulations had a field strength around 1 kG.

Observational findings In the case of the BP formation just discussed, the situation is similar. The initial field in the patch of diffuse flux is around 400 G. At t = 7 min a field intensification is seen at the border of the patch, accompanied by strong downflows of up to 2 kms^{-1} inside the magnetic field. In the DOT data with 1 min cadence the BP appears abruptly at t = 8 min. The field strength in the BP in the next polarimetric map at t = 14 min is around 900 G. The time scales in observations and simulations agree. The formation of the initial flux concentration takes some minutes, whereas the final field intensification and the increase in the intensity is very fast. Thus, the formation of the BP in the observations at least has the signatures of the convective collapse, although the temporal evolution cannot be traced very well due to the 7 min cadence of the polarimetric data.

4.8.3 The removal of photospheric flux through reconnection

During the 1 hour of observations not only the formation of Bright Points through the concentration of flux was observed, but also the opposite case, the disappearance of flux and the BPs related to it. The observations are described in detail in Bellot Rubio & Beck (2005), only a summary of the findings will be given here.

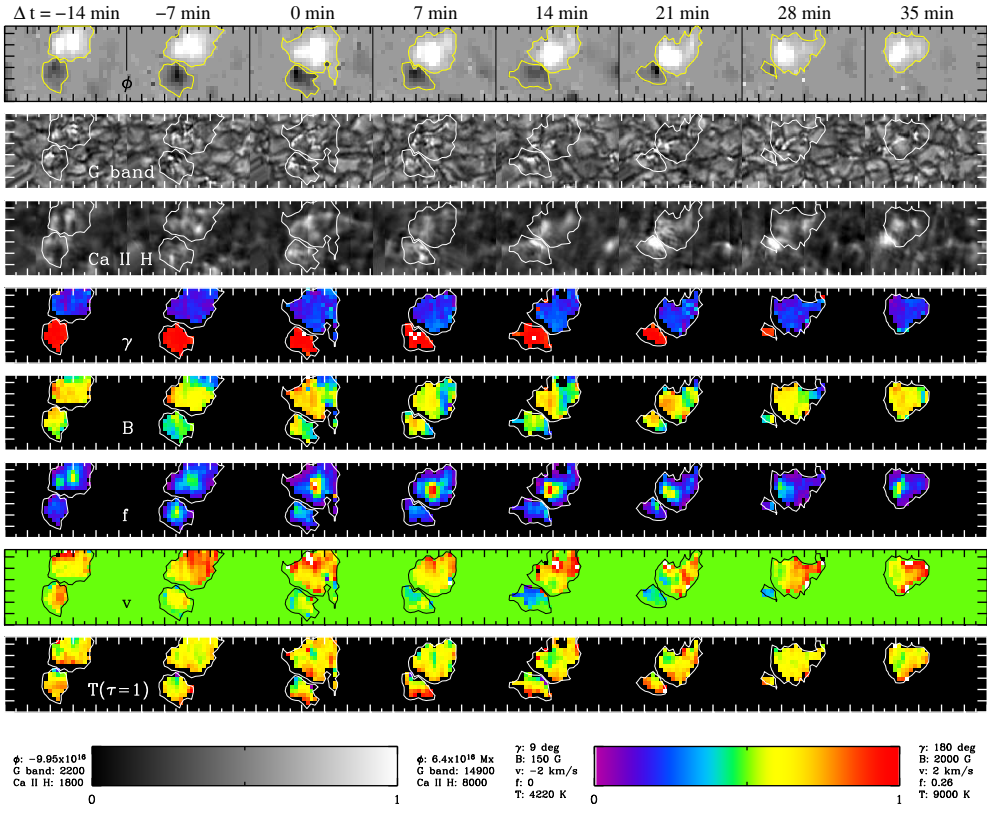


Figure 4.29: Temporal evolution of physical quantities during the flux cancellation event. **Top to bottom** magnetic flux (Φ), G-band intensity, Ca II H filtergram, magnetic field zenith angle in the local reference frame ($\gamma_{\rm LRF}$), magnetic filling factor (β), line-of-sight velocity (v), and temperature (T) at $\tau = 1$. Contours outline the two flux concentrations of opposite polarity. Tickmarks are in arcsec. Δt = 0 min denotes the start of the cancellation.

The flux cancellation takes part between a Moving Magnetic Feature (MMF) and a plage element located at the

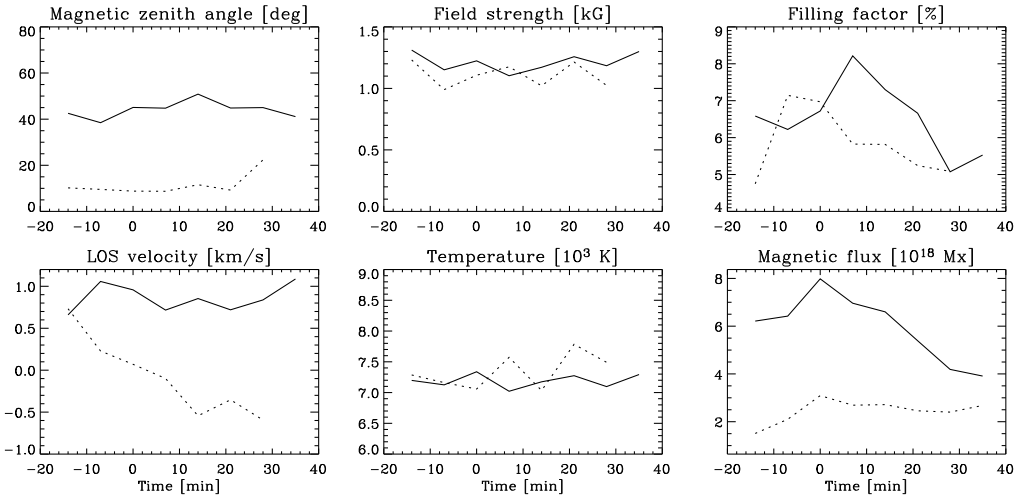


Figure 4.30: Evolution of average physical quantities in the Moving Magnetic Feature (**solid**) and the plage element (**dashed**). Negative velocities correspond to upflows. In the last panel, solid and dashed lines represent the unsigned and net magnetic flux, respectively. The zero point of time has been set to when the flux is first seen to decrease.

boundary of the moat flow. The MMF is swept towards the plage element of opposite polarity by the moat flow. The magnetic flux is found to reduce until the smaller plage element has completely disappeared. Fig. 4.29 displays the temporal evolution of the physical quantities during the cancellation event. The emission in the chromospheric Ca II H line is found to shift from the center of the respective magnetic patch towards the neutral line separating them. Most physical quantities of the photospheric field stay constant during the cancellation. The only significant change is seen in the line-of-sight velocity, where blue shifts appear at the location of the plage element.

Fig. 4.30 displays the evolution of the most important physical quantities, averaged over the area of the respective patches. It can be seen that the total magnetic flux decreases by about $2{\cdot}10^{18}$ Mx, whereas the *net* flux, the difference of positive and negative flux, stays constant. This implies that the same amount of positive and negative flux disappears by cancelling each other. The G-band BPs at the location of the plage element are found to disappear together with the flux, whereas the chromospheric emission persists.

4.9 Summary

We have presented the derivation of the magnetic properties of G-band Bright Points from infrared and visible spectral lines observed at the German VTT on Tenerife. The BPs were identified in co-spatial and simultaneous data taken with the DOT telescope on La Palma. The co-spatial Stokes profiles of infrared and visible spectral lines were inverted simultaneously with the SIR code, using a simplified model of the solar atmosphere, where magnetic field and velocity were constant with depth in the solar atmosphere.

For the BPs, we find the following properties from a statistical analysis of 538 identified BP structures:

- More than 90 % of the BPs are co-spatial to increased polarization signal and, hence, magnetic fields.
- The magnetic field strength of BPs ranges from 500 to 1500 Gauss. Most BPs show fields around 1000 G.
- The magnetic part of the atmosphere preferentially exhibits downflows with a different velocity in infrared and visible spectral lines. The velocity in the infrared lines is around 800 m/s and around 500 m/s in the visible lines. The profiles of both visible and IR lines show amplitude and area asymmetries, which also indicate velocity gradients.
- The BP intensity depends only slightly on the magnetic flux. It increases for flux values below $5{\cdot}10^{16}$ Mx and decreases for flux values above $1{\cdot}10^{17}$ Mx . This can be explained by the fact that the flux must lead to a significant reduction of the density to produce a BP at all, whereas for larger flux values the size of

the magnetic concentration increases, leading to a lower temperature, and hence less dissociation of the CH molecules.

- In the inversion model, two atmospheric components are used, a field-free component to describe the granular surroundings of the flux tube, and a magnetic atmospheric component to resemble the flux concentration. The two components each have an independent optical depth scale. In comparison, the magnetic atmosphere is about 1000 K hotter than the field-free component at identical optical depth. The temperature of the magnetic component is reduced for large flux values, which again is in agreement with the assumption that these structures have an increased diameter with reduced energy input by radiation from the side.

- The intensity of the BPs scales with the inclination of the magnetic field to the line of sight. This seems to be due to a purely geometrical effect of the path length inside the magnetic atmosphere. In the evacuated flux concentration less absorption happens, due to the reduced CH density, and deeper atmospheric layers with higher temperature, and hence less CH, are seen.

- From the filling fraction of the magnetic atmosphere inside the VTT resolution elements and the number of bright pixels in the DOT data, we derive an effective diameter of the bright structures. We find a value around 160-200 km which is comparable to, e.g., the values given by Berger & Title (2001) and Wiehr et al. (2004) or the width of intergranular lanes.

From our reference sample of quiet Sun data points, outside the canopy and away from any BP, we find a distribution of field strength with an exponential decay of the probability for higher field strengths, which has been found in Khomenko et al. (2003) for a quiet internetwork region from infrared lines alone.

The examination of a single isolated BP yields a close agreement with the BP properties derived from the statistical analysis: the BP is co-spatial to magnetic field with a field strength around 1 kG; it shows downflows; and the field lines are more or less parallel to the line of sight, which implies rather vertical fields with respect to the solar surface.

We have observed the formation of a G-band Bright Point from a patch of diffuse weak flux. In agreement with the statistical results, the BP appears only after the field strength exceeds 500 G. On the other hand, the diffuse weak flux has no signature in the G band.

At the outer boundary of the moat flow, we have observed a cancellation event between a Moving Magnetic Feature and a plage element of opposite polarity. In this case, magnetic flux is removed from the photosphere. The study has been included here, because it was derived from the same data set used for the statistical analysis. It will be discussed later in Sect. 6.3, because it is connected to the question of sunspot decay.

4.10 Conclusion

G-band Bright Points We conclude from the statistical analysis that G-band brightenings are caused by concentrated magnetic fields. The minimum requirements for a BP can be stated to be:

- A magnetic field strength above 500 G.

- A total flux above $2{\cdot}10^{16}$ Mx.

Apart from that, G-band Bright Points show a large intrinsic variability in the magnetic field parameters. The inclination to the line of sight, and hence to the solar surface, indicates that not all fields are vertical. The relative frequency of BP field strengths (cf. Fig. 4.11) shows a broad distribution rather than a single-peaked value.

This is in some contrast to the present image of BPs. It is commonly assumed that BPs mark *stable* flux concentrations with a field strength of 1.3 kG or larger, which are vertical to the solar surface. These elementary building blocks of the photospheric magnetic flux are pushed around by the granulation, merge, or split, but the fragments are supposed to still be kG flux tubes in most cases. However, from our observations we find a much more dynamical behavior. Magnetic flux can evolve from a concentrated form into diffuse flux or vice versa. Only if the requirements above are met, the magnetic field has an associated G-band BP. From the 1 hour time-series, it seems that the photospheric magnetic flux itself has a high variability with few stable (strong) field concentrations.

In recent high resolution observations with the new 1-m Swedish Solar Tower (Berger et al. 2004a,b), G-band brightenings at disc center were found to resemble more elongated filamentary structures than isolated roundish

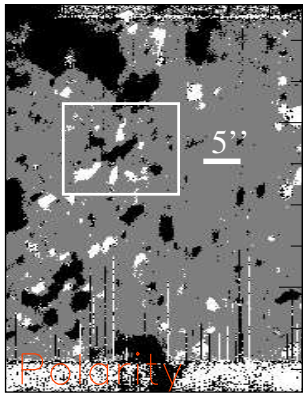


Figure 4.31: Polarity mask of a quiet Sun area on disc center (cf. the intensity image in Fig. 1.7, p. 10). White (black) correspond to fields oriented towards the observer (away from the observer). The white bar in the upper right part denotes a length of 5 arcsec. Polarity changes may occur on spatial scales of 1 arcsec or less, e.g., inside the white rectangle.

bright points. Hirzberger & Wiehr (2005) found both geometrical shapes, round BPs and elongated brightenings, in observations at an heliocentric angle of 38°. The existence of both geometrical shapes can be explained with small modifications to the geometry depicted in Fig. 4.22. Assuming a circular flux concentration with field lines parallel to the line of sight, a round BP with a radius d would be seen. However, if the magnetic flux is concentrated in an intergranular lane, it would rather form a flux sheet of thickness d. In this case, its boundaries would appear bright in the G band, because the hot granular wall would be seen, where CH is dissociated due to the higher temperature. The variability of the BP field properties found in the present work makes the simultaneous appearance of both cases probable, regardless of the heliocentric angle of the observations.

Probability density of field strength or The seemingly VIS-IR discrepancy Another conclusion to be drawn from the present study deals with the probability density of field strength in the solar atmosphere. The relative frequency of field strengths in Fig. 4.11 can be assumed to reflect the probability density function (PDF) for the occurrence of a certain field strength. The analysis of the quiet Sun reference sample yields a peculiarly shaped PDF: the higher the field strength, the less probable the occurrence. The PDF has a very steep gradient at weak fields, similar to an exponential decay. Khomenko et al. (2003) found a similar PDF from an examination of only infrared (IR) data in the same lines at 1565 nm as used in the present work.

However, many previous studies based on visible lines (VIS) yielded a different PDF with almost no weak and only strong (>1 kG) fields (e.g. Frazier & Stenflo 1972; Sigwarth et al. 1999; Domínguez Cerdeña et al. 2003). To reconcile the different results, Socas-Navarro & Sánchez Almeida (2003) proposed a distribution of field strengths already inside each resolution element. The respective Zeeman sensitivity of VIS and IR lines would then lead to different PDFs, because of the different efficiency of VIS and IR for the detection of weak or strong fields. From simultaneous IR and VIS observations at two telescopes, Sánchez Almeida et al. (2003) claimed that infrared and visible lines may even show opposite polarities in the same resolution element. This seemingly discrepancy of the IR and VIS results has created an intense debate about its origin.

Some points have to be considered in this discussion:

- The magnetic field configuration may change on very small scales. Opposite-polarity structures can be found on spatial scales of around 1 arcsec (cf. Fig. 4.31).

- The visible spectral lines are in the *weak field* limit for almost all magnetic fields outside of sunspots and pores. The two lobes of left- and right-circularly polarized light are spectrally broadened and influence each other. An increase of field strength mainly changes the amplitude of the Stokes V signal, not the separation of the lobes. In contrast, the IR lines are fully split for fields larger than 400 G.

- For the visible lines, the derivation of the field strength, B, is difficult. It is almost impossible to distinguish between an increase of B or of the magnetic filling fraction, because in first order both effects only increase the amplitude of Stokes V. In contrast, for the IR the field strength changes the peak separation and only the filling fraction changes the amplitude.

- Many earlier observations used magnetograms of VIS spectral lines with low spectral resolution, which are based only on the Stokes V amplitude.

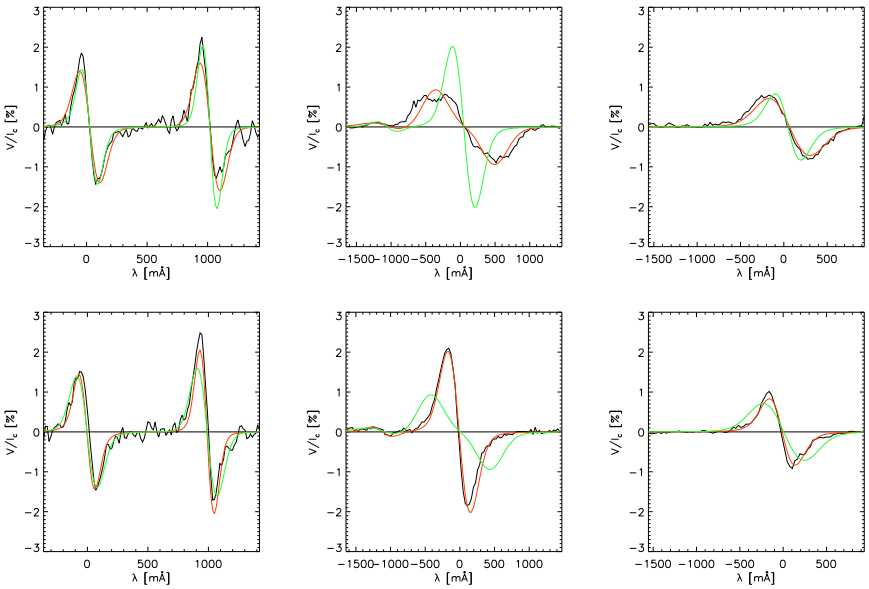


Figure 4.32: **Left to right:** Co-spatial profiles of the visible lines at 630.15/25 nm, IR line at 1564.8/1565.2 nm. The observations are in black, the best-fit profiles in **red**. In the upper row, the inversion yielded a magnetic filling fraction of 18 % and a field strength of 1240 G. In the lower row, the filling fraction was 21 % and the field strength only 360 G. The **green** lines in the 1st row are the best-fit profiles of the 2nd row, and vice versa. For the visible lines, the fit profiles can be exchanged, whereas for the IR the "wrong" fit is a complete mismatch.

- The data is influenced by noise to some extent. The same level of noise influences inversion results from IR and VIS in a different way.

The last point has been studied by Bellot Rubio & Collados (2003) in detail. They found that an original PDF with a peak at weak fields was not retrieved from the inversion of synthetic spectra of VIS spectral lines, if noise was added to the profile. The VIS tended to result in higher field strength with increasing noise level, whereas from the IR lines they were still able to derive the original PDF.

Khomenko et al. (2005) calculated IR and VIS spectra from a snapshot of a numerical magneto-hydrodynamic simulation, which included magnetic fields in a granular surrounding. They showed that the different magnetic sensitivity of IR and VIS leads to significantly different Stokes V amplitudes. In some cases, the signal of IR or VIS was lost in the added noise. However, the polarity of IR and VIS was found to be always the same.

In the present work, co-spatial IR and VIS spectra with high spatial and spectral resolution, which have been taken at the same telescope with a time lag of at maximum 18 sec, have been simultaneously inverted. Firstly, we do not see any case of opposite polarities in IR and VIS. The profile shape of IR and VIS is in general identical. The cases of opposite polarity in IR and VIS in the same resolution element found by Sánchez Almeida et al. (2003) are due to the combination of close-by opposite polarities and the fact that the THEMIS solar telescope, which they used for the VIS observations, has no image stabilization of any kind. Telescope excursions of >1 arcsec during the exposure time may well generate a "wrong" polarity signal in the data.

Secondly, the field strength determination in the simultaneous inversion has been found to be dominated by the IR spectral line with the highest magnetic sensitivity (cf. Sect. 4.6.2). Hence, it is not surprising that the derived PDF is in full agreement with Khomenko et al. (2003). However, the VIS spectral lines had been included in the fit with the same weight and were *equally well* reproduced. For a better visualization of the strong limitations imposed by the weak field limit for all VIS spectral lines, Fig. 4.32 displays two examples of co-spatial IR and VIS profiles, where the magnetic field strength differed by 800 G. The 1st row shows a strong field example. The field strength derived by the inversion code was 1240 G and the magnetic filling fraction inside the resolution element was 18 %. The 2nd row shows a weak field example, with $B = 360$ G and a filling fraction of 21 %. The best-fit profiles of the visible lines on the two positions can be mutually exchanged and still give a reasonable fit to both sets of observed profiles. The visible lines thus can be fitted by either a strong or weak field with a slightly modified filling

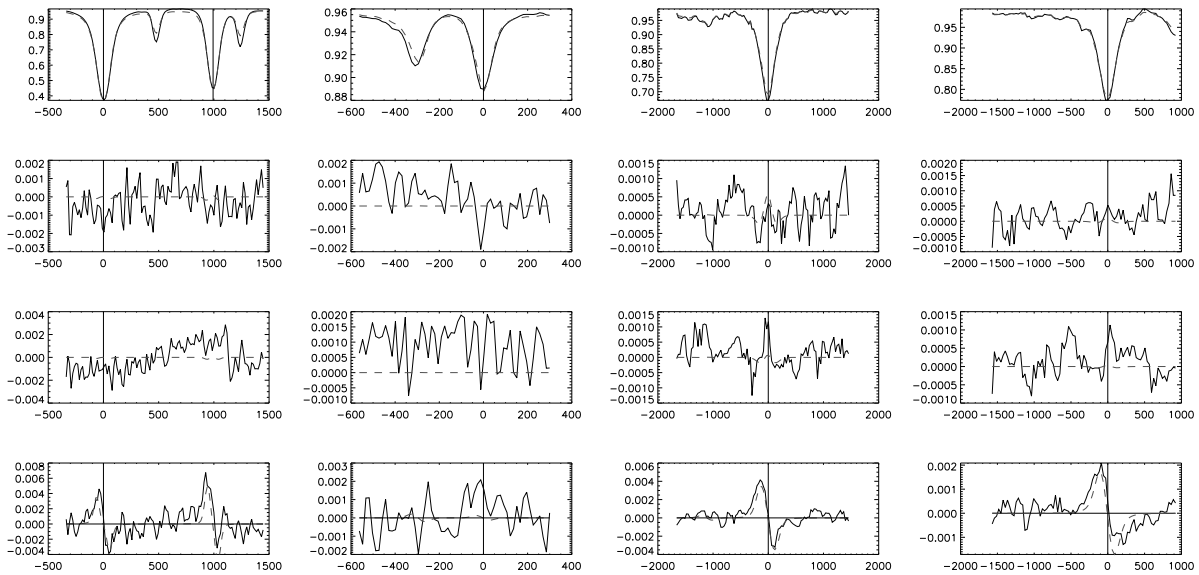


Figure 4.33: Profile example of a spatial position with a weak magnetic field of only 316 G and 3.5 % magnetic filling fraction. **Left to right**: visible spectral lines at 630.15/25 nm, 630.37 nm, infrared lines at 1564.8 nm and at 1565.2 nm. **Top to bottom:** Stokes $IQUV$.

fraction. For the IR lines, the fit profiles in the case of the kG field are fully incompatible with the observations of the weak field and vice versa.

In our simultaneous observations, we find that the VIS and IR spectral lines *both* react to the presence of either weak or strong field. A magnetic field of sufficient filling fraction generates a significant Stokes V signal in both wavelength ranges, regardless of the field strength. As example, profiles from a spatial position with weak magnetic field and a low filling fraction of 3 % are displayed in Fig. 4.33. It can be clearly seen that the best-fit profile reproduces the IR and VIS profiles at the same time. However, any suitable method for the derivation of the correct field strength from the VIS spectra alone can hardly be conceived in this case.

Thus, no discrepancy at all between IR and VIS is seen in the present work. The seemingly discrepancy between IR and VIS can be ascribed to a problem of the interpretation of the polarization signature. The information contained in the polarization of *any* (visible) spectral line in the weak field limit simply does not allow to derive the field strength, because the respective contributions from field strength or filling fraction can not be isolated. Unfortunately, this has far reaching consequences. For example, looking at the corresponding flux of the weak and strong field case displayed in Fig. 4.32, the values of the flux are $0.5{\cdot}10^{17}$ Mx and $1.5{\cdot}10^{17}$ Mx, respectively. If the profiles of the visible lines were observed with a magnetograph, where mainly the amplitude of V determines the derived flux, the magnetograph would deliver the same flux value for both, even if the "true" value differs by a factor of 3.

Until recently, no observations of IR spectral lines were possible due to the technical difficulties in the production of high-quality polarization optics and CCDs for IR wavelength. Many results and conclusion on magnetic fields in the solar atmosphere have thus been based on the accessible visible spectrum. However, the simultaneous IR and visible spectra taken during the present work give convincing evidence that spectral lines in the weak field limit do not allow to retrieve reliably the magnetic field strength. This is clearly visible already in the co-spatial profiles without any kind of evaluation. Thus, many earlier results, which rely on (visible) lines in the weak field limit, have to be taken with care, when field strength or flux values are of importance for the conclusions drawn.

Chapter 5

Sunspots

5.1 Introduction

Sunspots have been observed and studied already for centuries now, with an ever increasing amount of sophistication in both measurements and theory. The structure of the *umbra* and the morphology of the *penumbra* have been described in detail in sect. 1.2.2. As a short repetition, the dark umbra of a sunspot consists of almost vertical magnetic field with a field strength around 2000-3700 G. It is surrounded by the brighter penumbra, which is organized in radially aligned *filaments*. The penumbral intensity is still smaller than that of the granular surroundings. At the inner end of the filaments, close to the umbra, usually bright points can be found, the *penumbral grains*. The penumbral grains move along the filaments, preferentially inwards towards the umbra. Outside of the white light boundary of the sunspot magnetic field lines are found to continue into the magnetic *canopy*, supposedly consisting of almost horizontal field lines above unimpeded granulation. Isolated magnetic elements, the *Moving Magnetic Features*, are seen to appear at the outer penumbral boundary and are swept away from the spot by the *moat flow*, a radial flow field directed away from the sunspot.

From spectroscopic observations already Evershed (1909) found that sunspots exhibit a peculiar pattern inside the penumbra (*Evershed effect*), which was explained as the signature of a horizontal radial outflow, similar to the moat flow outside the spot.

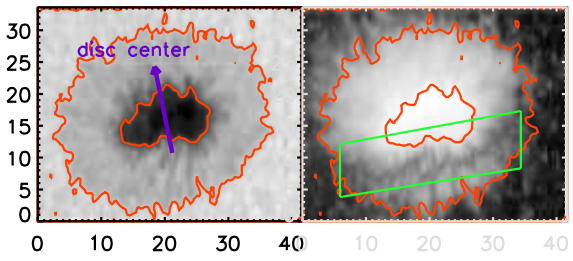


Figure 5.1: The location of the neutral line of Stokes V in a sunspot. **Left**: continuum intensity. **Right**: integrated circular polarization. Tickmarks are in arcsec. The red contour outline the umbra and penumbra, respectively. The green rectangle indicates the location of the neutral line, where the Stokes V signal is strongly reduced.

Evidence of multiple components in polarimetric observations With polarimetric observations the situation got unfortunately more complex. The horizontal Evershed flow is present everywhere in the penumbra (cf. Fig. 1.13, p. 14). However, the magnetic field inclination *on average* is not even close to horizontal inside a sunspot (cf. Title et al. 1993; Westendorp Plaza et al. 2001a,b; Mathew et al. 2003; Borrero et al. 2004, or the extensive reviews by Bellot Rubio (2002), Solanki (2003) and Thomas & Weiss (2004)). In the solar atmosphere, this difference poses a serious problem. The ionized plasma can only move along field lines. Thus, the observed horizontal Evershed flow requires horizontal fields. The mismatch of flow and average field inclination suggested that at least two different types of structures are present in the penumbra of sunspots, of which only one is involved in the Evershed effect.

Another more direct proof for the inherent complexity of the penumbra are the profiles of the circular polarization, Stokes V, in the so-called *neutral-line*. The amplitude of the circular polarization signal is proportional to the fraction of magnetic field lines parallel to the line of sight (cf. Sect. 3.1). If a sunspot is observed off disc center, it exhibits a line, where the V signal becomes very weak, because most of the field lines are perpendicular to the line

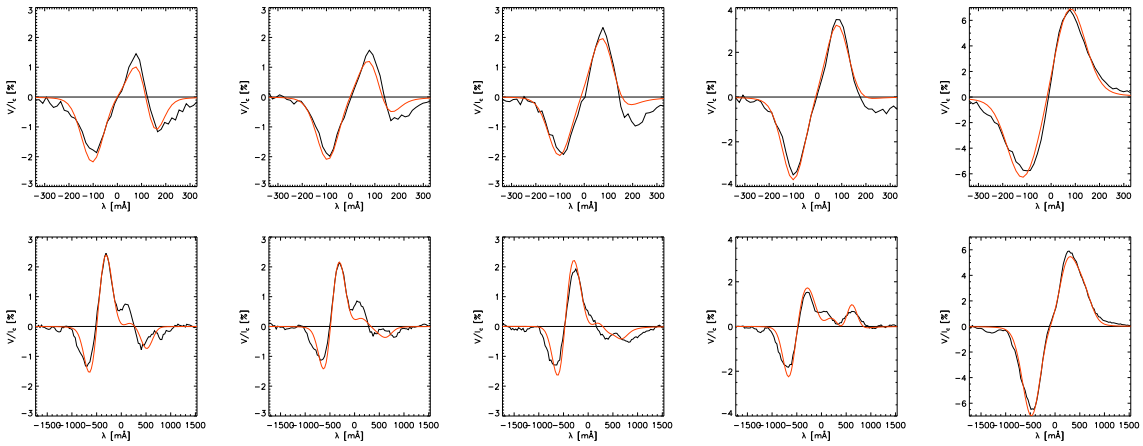


Figure 5.2: Examples of Stokes V profiles in the neutral line. **Top row**: Stokes V of the Fe I line at 630.15 nm. **Bottom row**: The co-spatial profiles of Fe I at 1564.8 nm. The red line gives the corresponding best-fit profile of the 2-component inversion. The last column shows a regular profile outside the neutral line for comparison.

of sight. The distance of this neutral line to the spot center depends on the heliocentric angle of the spot; it appears on the limb side of the spot for small heliocentric angles and moves up to inside the umbra for large heliocentric angles (cf. Fig. 5.1). The polarity of the field reverses across the neutral line.

The profiles of Stokes V in a neutral line inside the penumbra are very peculiar. They usually do not show no signal at all, but have clear signs of multiple components, i.e. more than 2 lobes in the polarization signal (cf. Figs. 5.2 or 5.27). To produce V profiles with multiple lobes, it is necessary to have at least two components with different line-of-sight velocities and opposite polarities inside one resolution element. Fig. 5.3 shows by a simple addition of profiles that two field components with opposite polarities will result in a regular profile. Only if the two components have opposite polarity *and* different LOS velocities, a multiple-lobed profile is created. Note that the polarity of a V profile (cf. Sect. 3.2) is defined relative to the line of sight. Opposite polarity implies a different angle between field and line of sight for the two components, not anti-parallel field lines. Furthermore, no conclusion can be drawn on the orientation of the field relative to the surface: the appearance of opposite polarity for observations at an heliographic angle θ only indicates that the sum of the field inclination to the vertical and

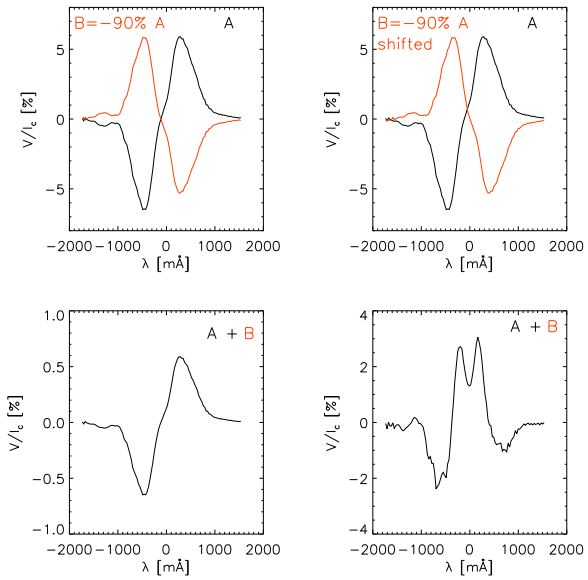


Figure 5.3: A simple method to produce peculiar V profiles. The addition of the two profiles A and B in the top row yields the profiles in the bottom row. **Left column**: profile B equals -90 % of A and is located at the same wavelength, yielding a regular profile with reduced amplitude after addition. **Right column**: B now equals -90 % of A, but has been shifted corresponding to a flow velocity of 2 kms^{-1}. The addition yields a multi-lobed profile in this case.

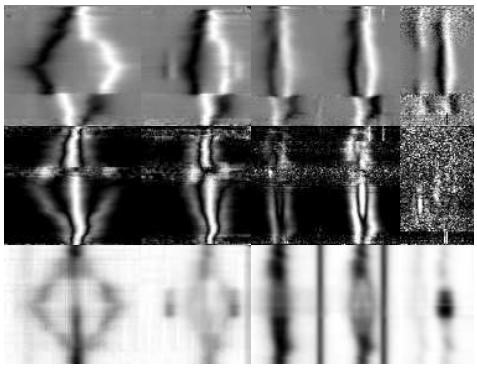


Figure 5.4: Stokes profiles along the symmetry line connecting spot center and disc center. The limb side is down, the center side up. **Bottom row**: intensity. **Middle row**: Total linear polarization, $\sqrt{Q^2+U^2}$. **Top row**: Circular polarization. **Left to right**: IR spectral lines at 1564.8 nm, 1565.2 nm, VIS lines at 630.15 nm, 630.25 nm, and Ti I at 630.37 nm. The profiles have been normalized independently in each row for better visibility. Both circular and linear polarization show a neutral line on the limb side and center side, respectively, where the profile shape changes abruptly.

θ is larger than 90°. Therefore, multiple-lobed profiles in Stokes V indicate two magnetic field components with different inclination and flow velocities.

Similar evidence for multiple components is found from a comparison of linear and circular polarization profiles. A regular Stokes V profile is defined to have two lobes, whose separation is proportional to the field strength, while the linear polarizations show three components (cf. Sect. 3.1). The zero-wavelength, λ_0, is in both cases exactly in the "middle", i.e. where the middle component of the linear polarization signal or the middle between the two lobes in Stokes V is located. λ_0 gives the position, where the corresponding spectral line in the intensity profile would appear. A close examination of the zero-wavelength of linear and circular polarization signals in the penumbra reveals that the two values do not coincide in many places.

These effects of different behavior in linear and circular polarization signal can also be visualized by following the profiles on a cut along the symmetry axis of a sunspot, the line connecting disc center and spot center. Fig. 5.4 displays, how the circular and total linear polarization, $\sqrt{Q^2+U^2}$, change along the symmetry axis. The circular polarization shows a sign reversal across the neutral line on the limb side. The profiles in the inner limb side penumbra indicate stronger field with smaller flow velocities. Across the neutral line this changes to a smaller field strength and higher flow velocities. A similar neutral line can be seen in the linear polarization on the center side. Again, in the inner penumbra stronger fields with a smaller velocity change to weaker fields with a higher velocity. The Stokes profiles are dominated by the component, whose signature is favored by the respective geometry on limb or center side (cf. Bellot Rubio et al. 2004). In the neutral line of V, both components can be seen at the same time. In the image it also can be seen that the Ti I line at 630.37 nm, which was also included in the present study, gets strong only in the umbra of the spot. At the same locations, the two OH^- blends left and right of the 1565.2 nm appear in the spectrum. They are sensitive to both temperature and magnetic fields (Mathew et al. 2003), but were not used due to the technical difficulties involved in calculating molecular spectral lines.

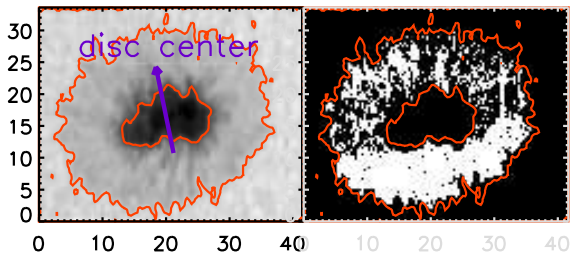


Figure 5.5: The locations of Stokes profiles with indications of multiple components. **Left**: Continuum intensity. **Right**: Mask of profiles indicating multiple components. Tickmarks are in arcsec. The red contour outline the umbra and penumbra, respectively. On each white point of the mask the Stokes profiles either showed more than 2 lobes in Stokes V, or the velocity of linear and circular polarization signal differed by more than 0.5 kms^{-1}.

The multiple components are present all throughout the penumbra, not only at the neutral lines of linear, respectively, circular polarization. In a first step of the data analysis, the observed profiles of the sunspot data were controlled for the signature of multiple components by two criteria:

1. The number of lobes in the Stokes V signal.

2. The velocity difference between λ_0 of the linear and circular polarization signal.

If the number of lobes in V was larger than 2 or the velocity difference larger than 0.5 kms^{-1}, the corresponding profiles were taken to indicate multiple components. A map of the locations, where one or both of the two criteria were met, is shown in Fig. 5.5. On the limb side of the spot, almost all profiles around the neutral line showed multiple components, whereas on the center side some larger areas are not marked. The difference is due to the visibility of the two components. On the limb side, the stronger and more vertical field is perpendicular to the line of sight. The more horizontal component thus stands out clearly in Stokes V. On the center side, the stronger and more vertical field is less inclined to the line of sight than the weaker and more horizontal component, making the detection of the weaker field more difficult. This systematic difference between limb and center side due to purely geometrical reasons persists to some extent also in the final results. However, from the almost complete coverage of the limb side penumbra it seems satisfied to assume the presence of the two components everywhere inside the penumbra.

Another indicator of the inhomogeneity of the penumbral atmosphere is the observed *Net* or *Broadband Circular Polarization* (NCP, respectively, BBCP). For an atmosphere with constant magnetic field in optical depth, the Stokes V signal is perfectly anti-symmetric (cf. Sect. 3.1). If the integral of V, $\int V(\lambda)d\lambda$, does not vanish, gradients or discontinuities in the field strength and/or the flow velocity have to be present. The NCP will be discussed separately in Sect. 5.5. At this point, it suffices to state that a non-zero NCP is observed throughout, again, the whole penumbra.

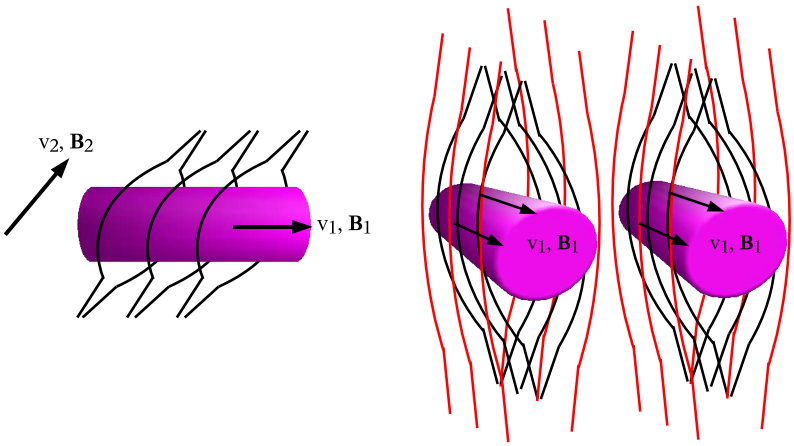


Figure 5.6: The uncombed penumbra. The horizontal flux tubes, which carry the Evershed flow (v_1, $\mathbf{B}_1$), are enclosed by more vertical field lines ($v_2 \sim 0$, $\mathbf{B}_2$). The line of sight will pass through the more vertical field above and below the flux tube. At the boundary of the tube it encounters discontinuities in all physical parameters.

These various indications of multiple structures inside the penumbra have induced a lot of interest. Any model of the penumbral fine structure has to include a list of observational findings:

- The Evershed flow is seen everywhere in the penumbra. The Evershed effect in the line core velocity ends rather abruptly at the outer white light boundary of the spot. The flow must be very close to horizontal.
- The magnetic field changes polarity on the limb side for observations of sunspots off disc center. In the penumbra, both circular and linear polarization show a neutral line, where the properties of the field change. In the neutral line of Stokes V, profiles directly show multiple atmospheric components.
- The average magnetic field is not horizontal inside the penumbra.
- The NCP has a non-zero value in the penumbra.
- The penumbral grains are found predominantly on the limb side of the spot at the umbral boundary.

The basic concept to explain the peculiarities of the observed penumbral Stokes profiles today is the *uncombed* or *fluted penumbra*, proposed by, e.g., Solanki & Montavon (1993) or Title et al. (1993). Fig. 5.6 shows a sketch of the assumed geometry: both horizontal flow channels and more vertical field lines are present in each resolution

element. The more vertical field lines wind around the horizontal flow channels. In the spaces between neighboring flow channels the vertical field lines are undisturbed.

The geometry sketched in Fig. 5.6 is in agreement with most requirements in the list above. A line of sight, which passes through the the atmosphere, will encounter the more vertical field lines above and below the flow channel. At the boundary of the flow channel, it will meet a discontinuity in the magnetic field quantities, a jump in the field strength, field direction and in the line-of-sight velocity. These jumps are the necessary ingredients to create a Net Circular Polarization signal. The average inclination will not be horizontal, because both flow channels and more vertical fields contribute. Depending on the heliocentric angle and the location in the penumbra, the horizontal flow channels or the more vertical field will be more prominent in the observed profiles, leading to a significant difference between the limb side and the center side of the spot. If the line of sight is almost perpendicular to the more vertical field, the necessary requirements for the generation of multiple-lobed Stokes V profiles are fulfilled: two magnetic components of opposite polarity with different flow velocities are seen.

However, even if the geometry fits to the observational findings, an important question remains open: Why should this geometry develop in a sunspot penumbra ? Also some other points are not explained by the geometry itself:

a) What is the size of the horizontal flow channels ? Where are they located in the atmosphere ?

b) What drives the Evershed flow in them ?

c) Where do the channels end at the outer spot boundary ? Where does the mass moved by the Evershed flow go to ? Where does it come from ?

d) What is the temporal evolution of an individual flow channel ?

e) What is the relation between the flow channels and filaments ? Are filaments single flow channels, or do they consists of a number of individual flow channels ?

This list surely is not complete, but it indicates the direction of the present scientific investigations. Some of these points will be addressed in the following chapter from the observational side, while others already have been tackled on a theoretical base.

Based on the picture of the fluted penumbra, Schlichenmaier et al. (1998) studied the evolution of a flow channel (or *flux tube*) in the penumbra of a sunspot in a numerical 2-D simulation. They placed an isolated flux tube inside the static background field of a spot. Initially, the flux tube was located at the outer boundary between magnetic field and the field-free surroundings. The flux tube was found to become buoyant due to radiative heating and started to rise. The final steady-state result of the simulation was a horizontal flow channel, slightly elevated about the solar surface. Inside the flux tube, a radial outflow of up to 14 kms^{-1} was found. The simulation of this *Moving Tube Model* (MTM) gave for the first time a consistent explanation to how the fluted penumbra forms and what properties it has.

To summarize, the penumbral fine structure shows clear indications of multiple magnetic field components in each resolution element. In the shape of the MTM model a possible explanation for the formation and the properties of the flow channels is proposed from the theoretical side. From the observational side, this work is aimed at determining the properties of observed flux tubes for a comparison with the numerical model. In first order, the multi-wavelength spectro-polarimetric observations examined in this work clearly favor the final geometry of the MTM model.

The final aim, the derivation of the properties of the flux tubes carrying the Evershed flow for a comparison with theory or simulation, leads to the organization of this chapter. The observations are described in Sect. 5.2. To be able to isolate the properties of the more horizontal flow channels from the expected quasi-static background field at rest, the data are first analyzed in terms of a 2-component inversion (Sect. 5.3.1). The conversion of the inversion results from the line-of-sight reference frame to the local reference frame is described in Sect. 5.3.2. The two components will be distinguished by their inclination to the surface (Sect. 5.3.3). The atmospheric parameters after the distinction into background field and flux tube component are analyzed, where also the properties of the static background field are of interest for a model of sunspot geometry. Sect. 5.4 describes quantities, whose derivation from the inversion results is more or less straightforward. Among other things, also the question of field-aligned flows is addressed, the seemingly difference between field inclination and the horizontal Evershed flow in previous studies.

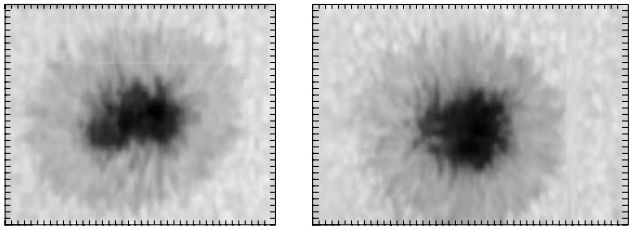


Figure 5.7: Maps of the continuum intensity in the near infrared of the two spots examined. The distance between tick marks is 1 arcsec ($\sim$ 725 km). The left spot was at a heliocentric angle around 30° and is foreshortened, the right spot was at around 7° and appears to be almost circular.

Additionally, the results of the 2-component inversion are used as initial values for a so-called *uncombed "Gaussian" inversion* (cf. Sect. 5.6.1), where the horizontal flow channel is embedded in one of the atmospheric components, to derive the width and the location of the flux tube in the atmosphere. To derive the 3-D topology from the results of the Gaussian inversion, the optical depth scale has to be converted to geometrical height (Sect. 5.8.1). The resulting geometry is described in Sect. 5.8.2. The integration of the inclination is introduced as a visualization tool in Sect. 5.9. The final 3-D spot model, combining Gaussian and 2-component inversion, is given in Sect. 5.9.5. Sections 5.12 and 5.13 summarize and discuss the results in the frame of current studies. The conclusions in Sect. 5.14 end the chapter.

5.2 Observations

"*The other observation, for those not in the rankest ignorance of perspective, is that from the changes of shape observed in the spots, and from their apparent changes in velocity, one must infer that the spots are in contact with the Sun's body, and that, touching its surface, they are moved either with it or upon it .... Their motion proves this by appearing to be very slow around the edge of the solar disc, and quite fast toward its center; the shapes of the spots prove the same by appearing very narrow around the Sun's edge in comparison with how they look in the vicinity of the center. For around the center they are seen in their majesty and as they really are; but around the edge, because of the curvature of the spherical surface, they show themselves foreshortened.*" (Galilei 1632)

The data analyzed in this work consists of two sunspots, which were observed on the 8th and 9th of August in 2003 with the two vector spectro-polarimeters at the German Vacuum Tower Telescope (VTT) in Izana on Tenerife: the Polarimetric Littrow Spectrograph (POLIS, cf. Sect. 2.3.1 or Beck et al. 2005b) and the Tenerife Infrared Polarimeter (TIP, cf. Sect. 2.3.2 or Martínez Pillet et al. 1999). The two instruments were used simultaneously using a 50-50-beamsplitter to feed both at the same time. During the observations POLIS was remote-controlled by the TIP control computer to ensure strictly simultaneous exposures. The image motion due to the fluctuations of the refractive index in the Earth's atmosphere ("seeing") and wind-induced motions of the telescope structure was reduced using the Correlation Tracker System (CT, Schmidt & Kentischer 1995; Ballesteros et al. 1996, or Sect. 2.3.3) that stabilizes the Field-of-View (FOV). From the instruments the full Stokes vector (cf. Sect. 2.1) in visible, respectively, infrared wavelengths is retrieved. The observed spectral lines are listed in Table 4.1, p. 50.

Fig. 5.7 display maps of the continuum intensity in the infrared of the two spots. The 2-D intensity maps are constructed from the intensity along the slit for each scan step, the scanning direction is left to right. The different location of the sunspots on the solar disc can already be seen from their shape: the first spot is foreshortened towards the solar limb by projection effects; it was at an heliocentric angle of about 30°. The second spot was close to the center of the solar disc at only 7° heliocentric angle.

The steps of the analysis are identical for both observations; thus, examples of the analysis procedure will be shown only for one of the spots, the observation off disc center. The differences and similarities of the two spots will be discussed in a separate section (Sect. 5.11) before the summary.

5.3 Evaluation of the observed spectra by a 2-component inversion

5.3.1 Inversion setup: 2-component inversion

The SIR inversion code tries to reproduce the observed spectra from an assumed stratification of the solar atmosphere (cf. Sect. 3.3.3). An initial atmospheric stratification is iteratively modified, until the synthetic spectra from the model match the observations. SIR allows to characterize the properties of the solar atmosphere by one or

two different atmospheric components (*model atmospheres*). For the investigation of the sunspot data, the number and the properties of the model atmospheres were adjusted to the type of the observed spectra at a given spatial position. To this end, a mask was created, which distinguishes between the *umbra* of the spot, the *penumbra* (cf. Sect. 1.2.2.2), and the profiles with significant polarization signal outside the spot. The continuum intensity was used as criterion for separating the different areas. Spatial positions without magnetic signature were not inverted.

The inversion can be repeated after the first convergence to a minimum χ^2-value with more degrees of freedom or changing of the respective weights, (w_I, w_Q, w_U, w_V), to improve the fit quality (cf. Eq. (3.28), p. 41). A 2nd iteration was only done for penumbral profiles.

The following setup was used for the three different profile types:

- Outside the spot: one field-free model atmosphere, one magnetic model atmosphere, $(w_I, w_Q, w_U, w_V) = (1, 10, 10, 10)$.
- Penumbra: two magnetic model atmospheres, one repetition. In the 1st iteration the weights were set to $(w_I, w_Q, w_U, w_V) = (1, 2, 2, 15)$, for the repetition to $(w_I, w_Q, w_U, w_V) = (1, 5, 5, 10)$.
- Umbra: one magnetic model atmosphere, $(w_I, w_Q, w_U, w_V) = (1, 5, 5, 10)$.

All parameters of the model atmospheres besides the temperature were always set to be constant in optical depth; only variation by adding or subtracting constant offsets was allowed for (no gradients). For temperature, linear increase or decrease with optical depth was allowed for (cf. Sect. 3.3.3). In the fit of penumbral profiles, the 1st iteration was restricted to use only constant offsets in temperature and linear changes in the 2nd iteration.

The initial model atmospheres were adjusted to have the field orientation parallel or anti-parallel to the LOS, depending on the polarity of the Stokes V signal (cf. Sect. 3.2). For the penumbral profiles, the two initial model atmospheres differ in the inclination by 30° and in the line-of-sight velocity by 1 kms^{-1}, to allow the code a more clear distinction between two alternatives. The field azimuth and the field strength is identical; the same initial values are used throughout the spot. The parameters, which specify the physical atmospheric properties in the inversion, are derived in the line-of-sight reference frame. They are given with respect to the ray path that connects observer and the observation target, and the plane perpendicular to it. To establish the properties with respect to the solar surface, the magnetic field vector has to be converted to the Local Reference Frame (LRF). The LRF consists of the solar surface plane and the local vertical (Sect. 5.3.2). After the conversion to the LRF, the two inversion components can be separated by their inclination to the surface.

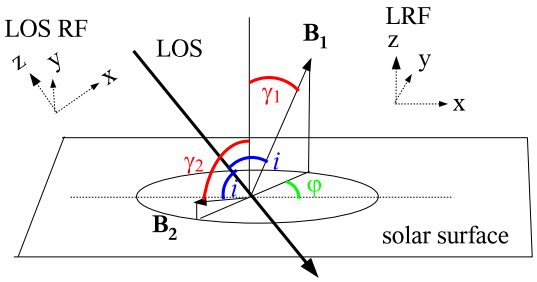


Figure 5.8: The magnetic field inclination and azimuth angle in the line-of-sight reference frame (LOS RF) and the Local Reference Frame (LRF). The inclination of the magnetic field vector to the LOS, i, can be determined exactly, but its azimuth, ϕ, is ambiguous by 180°. The two possible orientations of the field for ϕ and $\phi + 180°$ correspond to the two field vectors $\mathbf{B}_1$ and $\mathbf{B}_2$, which have the same field strength and the same inclination to the LOS, but a different inclination γ_1, respectively, γ_2, in the LRF.

5.3.2 Conversion from the Line-of-sight to the Local Reference Frame

5.3.2.1 Line-of-sight and local coordinate systems

To specify the magnetic field vector, $\mathbf{B} = (B_x, B_y, B_z)$, in a spherical coordinate system, which is best suited for the purpose of changing reference frames, three parameters are necessary: the field strength, B, and two angles, the inclination, i, and the field azimuth, ϕ. The inclination measures the angle between field vector and the z-axis, while the field azimuth indicates the orientation of the projection of the field vector in the x-y-plane. The azimuth is counted by convention positive from the x-axis towards the y-axis. Two sets of reference frames are important: the line-of-sight (LOS) reference frame, where the z-axis is given by the direct connection between observer and target, and the Local Reference Frame (LRF), where z is perpendicular to the solar surface. These two coordinate systems are identical only for observations on the center of the solar disc (cf. Fig. 5.8).

The parameters retrieved by the inversion are defined with respect to the line-of-sight reference frame. Zero inclination of the field means that it is parallel to the LOS and points towards the observer. The azimuth angle is measured in the x-y-plane perpendicular to the LOS. The zero azimuth position, i.e. the orientation of the x-axis in space, is a free parameter. In the case of the data analyzed in this work, it is defined by the model of the telescope polarization properties: the reference frame on the solar surface is chosen there to have positive Stokes Q parallel to celestial North-South (CNS, cf. Sect. 2.4.3)). Thus, a zero azimuth angle in the LOS system indicates field lines along the CNS-direction. However, the LOS azimuth angle can be determined only with an ambiguity of 180 degree. From the observed spectra no distinction is possible between an azimuth of ϕ and ϕ+180°, which leads to exactly the same profiles in Stokes $IQUV$.

In the Local Reference Frame, the inclination of the field is defined by the angle of field and local vertical, i.e. the vector that is perpendicular to the solar surface. To clearly distinguish between LOS and LRF inclination, γ will be used exclusively for the inclination in the LRF. The zero position of the azimuth can again be chosen; for a sunspot one can make use of the symmetry properties of a spot in the choice. A number of physical parameters, like the flow velocity or the Net Circular Polarization (NCP, cf. Sect. 5.5), show distinct symmetries with respect to the line that connects disc center and the center of the sunspot. The symmetries are mainly due to the geometric projection effects for a spot located off disc center. Additionally, choosing this orientation makes the conversion to the LRF easier. Thus, the zero point of the LRF azimuth has been put along the line connecting spot center and disc center.

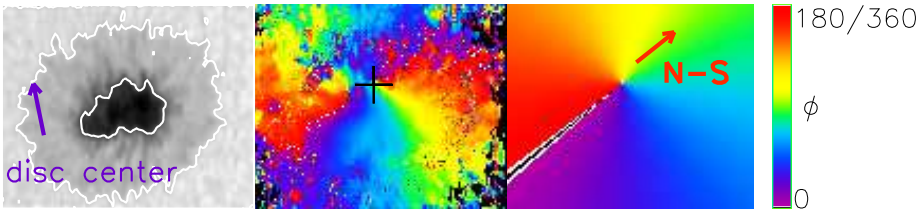


Figure 5.9: **Left to right**: Map of the continuum intensity of a sunspot observed in the near infrared, LOS azimuth of the 1st inversion component without correction (range 0...180°), theoretical LOS azimuth (range 0...360°). The purple arrow points to disc center. The black cross denotes the azimuth center of the spot, the red arrow points in the celestial N-S-direction.

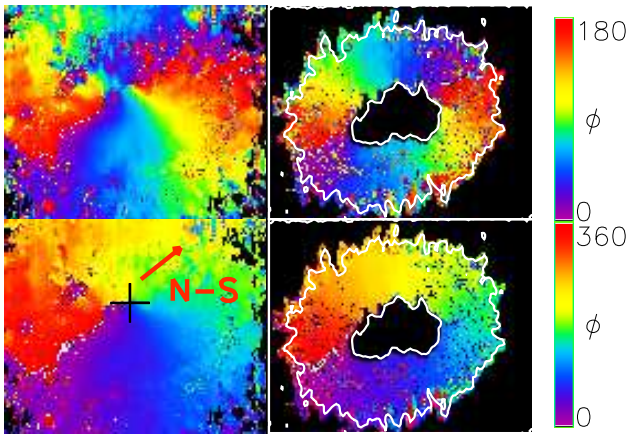


Figure 5.10: **Top row, left to right**: LOS azimuth of the 1st inversion component without correction (range 0...180°), same for 2nd component. **Bottom row, left to right**: LOS azimuth of the 1st component after correction (range 0...360°), same for 2nd component. The effect of the correction is mainly the addition of 180° to the azimuth values above and left of the celestial N-S direction in the upper part of the sunspot.

5.3.2.2 Resolving the 180° azimuth ambiguity

The 180°-ambiguity of the azimuth in the LOS system presents a problem in the conversion to the LRF. Depending on the heliocentric angle of the observations, the inclination to the local surface may be quite different in the two possible azimuth cases, as the vectors $\mathbf{B}_1$ and $\mathbf{B}_2$ in Fig. 5.8 display. Three approaches are possible for resolving the azimuth ambiguity:

- The requirement of more vertical fields: it is known from simulations and theoretical calculations that isolated magnetic flux concentrations tend to be more or less vertical with respect to the solar surface due to buoyancy forces. Solutions in hydrostatic equilibrium are in most cases only possible for almost vertical field lines. This requirement is a good approach outside of sunspots; it is met by choosing the solution of azimuth that yields the more vertical fields.

- The requirement of minimum currents: the inversion is usually performed on a pixel-to-pixel base; the field configuration in neighboring pixels is not taken into account during the inversion. The 180°-ambiguity then may lead to a strong variation of the field orientation on small scales, leading to strong gradients in the spatial dimensions. As the current density, $\mathbf{j}$, is given by $\nabla \times \mathbf{B}$, the gradients lead to strong currents. However, currents should dissipate due to the finite conductivity of the solar atmosphere and thus constantly deposit energy into the atmosphere, converting magnetic field energy into thermal energy. Magnetic field thus should vanish rapidly, which is not what is observed. The minimum current solution then requires to select the configuration, which shows the least spatial gradients in the vector field. The solution is found by choosing one of the two possible azimuth angles for a single pixel, and then always choosing the field configuration in the neighboring pixels that is closer to the orientation of the first one. It ensures a smooth variation of the field orientation inside a patch of magnetic flux, but one initial starting orientation has to be selected.

- The assumption of radial fields: for a sunspot, to first order, the field can be assumed to be oriented radially outward from the spot center. The deviation from radiality, the magnetic twist, has been found to be less than 10° for regular sunspots. The ambiguity can be corrected for by creating an azimuth map that would correspond to a purely radial field starting from the spot center, and then always choosing the azimuth solution that is closest to the purely radial case.

For the observations presented here the third approach has been chosen.

Fig. 5.9 displays the steps to calculate the theoretical LOS azimuth that corresponds to a purely radial field. First, the azimuth center of the spot is determined from the LOS azimuth of the 1st inversion component (black cross). The LOS azimuth without correction ranges twice from 0° (white) to 180° (red) if one moves a full circle around the azimuth center. This clearly indicates that in one half of the spot the azimuth should be increased by 180°. The theoretical azimuth can be constructed, because the orientation of the POLIS (and likewise the TIP) slit to celestial North-South (CNS) is known: from a drift scan and the orientation of the spectrograph during the observations it is known that the slit(s) are at an angle of 51.6° with CNS. Thus, the zero value of the theoretical azimuth has to be from the azimuth center along the direction of CNS in the data. The spot is in fact of negative polarity, i.e. the field lines are pointing down into the solar surface, which reverses the orientation in the azimuth. Note that the observed

LOS azimuth is in good agreement with that: the white pixels with zero LOS azimuth values form a line that is more or less in the CNS direction, but opposite to it. This location of the zero azimuth line is a sign of consistency that the spot has negligible twist and the telescope model is working properly.

Fig. 5.10 displays the LOS azimuth values of both inversion components before (top) and after the ambiguity correction. The correction only changes the azimuth by $\pm 180°$, it does not force the azimuth to be identical to the theoretical radial azimuth. Again, the the zero values of the observed LOS azimuth are in good agreement with CNS for both inversion components.

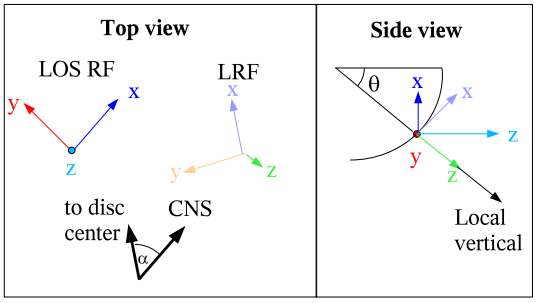


Figure 5.11: The conversion from the line-of-sight reference frame (LOS RF) to the Local Reference Frame (LRF). LOS RF: The x-axis is given by the telescope model and points along CNS. The z-axis is parallel to the LOS and perpendicular to the page. LRF: The z-axis is parallel to the local vertical and the x-axis is parallel to the line that connects spot center and disc center. If the LOS RF is rotated by α around its z-axis, the side view on the right is valid. The y-axes of both RFs are identical, thus, a rotation by θ around the y-axis transforms vectors in the LOS RF into the LRF.

5.3.2.3 Conversion to the Local Reference Frame

To convert the ambiguity-corrected LOS field vector to the Local Reference Frame, the position of the spot on the solar disc has to be known. It is defined by two angles: the angle, α, between CNS and the line that connects spot center and disc center and the heliocentric angle of the spot, θ. With these numbers, the conversion from LOS to LRF can be achieved in two steps. Firstly, the LOS RF is rotated by α around the LOS (cf. Fig. 5.11). This means that the x and z-axis of both the rotated LOS RF and the LRF are in the same plane. The two y-axes are identical, because the respective y-axis has to be orthogonal to the x-z-plane. Secondly, the LOS RF is rotated around the new y-axis by θ to move vectors from the LOS RF into the LRF.

For the observations, the respective values of α and θ were derived from a full disc image of the Sun in white light (cf. Fig. 5.12) taken with the Michelson Doppler Imager (MDI) instrument onboard the SOlar & Heliospheric Observatory (SOHO) satellite. The spot is south and east of the disc center. It was located at an heliocentric angle of $\theta = 27.8°$, while the angle between CNS and the symmetry line of the spot was $\alpha = 62.7°$. The two steps of the conversion are then executed in the following way:

- Calculation of the magnetic field vector with respect to the LOS:

$$\mathbf{B}_{\mathrm{LOS}} = |\mathbf{B}| \cdot \begin{pmatrix} \cos(\phi_{\mathrm{LOS}} - \alpha) \cdot \sin i \\ \sin(\phi_{\mathrm{LOS}} - \alpha) \cdot \sin i \\ \cos i \end{pmatrix} , \tag{5.1}$$

  where the subtraction of $\alpha = 62.7°$ from the field azimuth, ϕ_{LOS}, corresponds to moving the LOS x-axis onto the symmetry line.

- Rotation of the field vector in the LOS RF by $\theta = 27.8°$ around the y-axis:

$$\mathbf{B}_{\mathrm{LRF}} = \cdot \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \cdot \mathbf{B}_{\mathrm{LOS}} . \tag{5.2}$$

After the conversion to the LRF the field vector can again be expressed by field strength, the inclination, γ, and the field azimuth, ϕ, in the LRF. The values can be derived from the vector in the cartesian coordinate system, $\mathbf{B}_{\mathrm{LRF}}$, via:

$$\mathbf{B}_{\mathrm{LRF}} = |\mathbf{B}| \cdot \begin{pmatrix} \cos\phi_{\mathrm{LRF}} \cdot \sin\gamma \\ \sin\phi_{\mathrm{LRF}} \cdot \sin\gamma \\ \cos\gamma \end{pmatrix} . \tag{5.3}$$

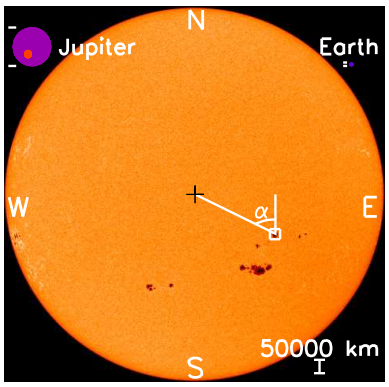


Figure 5.12: Full disc image of the Sun in white light from the Michelson Doppler Imager (MDI) onboard the SOlar & Heliospheric Observatory (SOHO) satellite. The sunspot observed was the active region NOAA 0425, south and east of the disc center, marked by the white rectangle. The angle α between celestial North-South and the line connecting spot center and disc center is indicated. For comparison of the size, some astronomical objects are plotted in the corners of the figure.

Fig. 5.13 displays the field azimuth and the inclination in the LRF after the conversion. The azimuth center has now moved to the center of the spot. The white pixels that indicate zero azimuth values now form a line that is in good agreement with the direction to disc center, which can be seen more clearly for the 2nd component (top right). Again, this implies a radial field, and the consistency of the conversion. The transition from the one-component inversion inside the umbra and outside the spot does not lead to large discontinuities of the field azimuth. The maps display the two components after the separation into background field and flux tube component described in the following section.

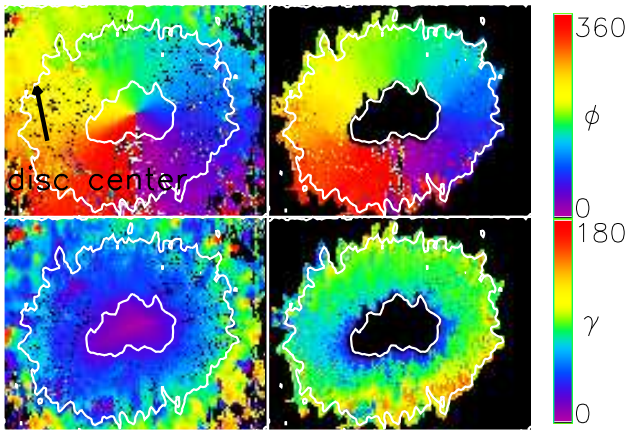


Figure 5.13: **Top**: field azimuth of the two inversion components in the Local Reference Frame (range 0..360°). **Bottom**: field inclination (range 0..180°). **Left column**: background component, **right column**: flux tube component. The azimuth center has moved to the center of the spot (cp. with Fig. 5.10). The components have already been separated into background field and flux tube component by their inclination (cf. Sect. 5.3.3).

5.3.3 Classification into background field and flux tube component

At this point the two inversion components can now be "classified" as belonging to two different types of structures, the "background field" (bg) and the "flux tube" (ft) component. Until now, the components were treated to be equivalent: the SIR code does not contain assumptions on which component should have the higher flow velocity, the larger inclination, or the weaker field. From the observed profiles it can be seen that different properties are required for the two model components to reproduce the observations. The code attributes certain atmospheric properties to one *or* the other component, simply led by the initial atmospheric values and the need to minimize the χ^2-value. Now, in the LRF a distinction can be made by the inclination to the surface. Reasons for making this distinction have already been given in Sect. 5.1.

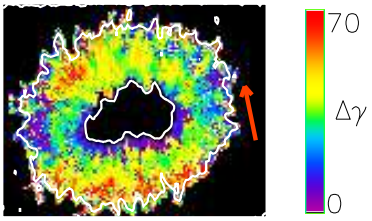


Figure 5.14: The difference in the inclination to the solar surface of the flux tube and background component. The red arrow points to disc center. On the disc and the center side, the two components have markedly different inclinations. On the line perpendicular to the symmetry line the inclination values are very similar, which makes the classification by inclination ambiguous.

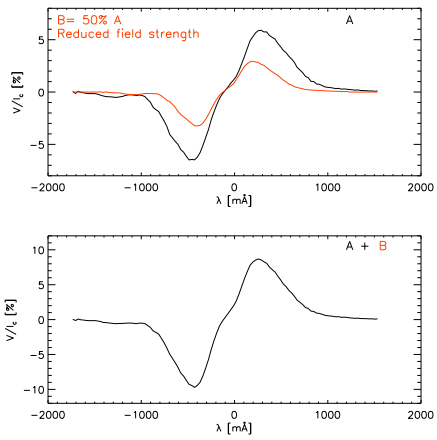


Figure 5.15: Resulting profile for zero velocity difference between two magnetic components with the same polarity. Along the line that is perpendicular to the symmetry axis of the sunspot the Evershed flow velocity passes through zero. The more vertical stronger background field (component A) and the inclined weaker flux tube component (component B) produce a Stokes V signal at the same wavelength. The addition of the two profiles leads to a regular V signal, where the properties of the ft component cannot be retrieved.

From now on, the two inversion components are characterized by their inclination in the Local Reference Frame. At each pixel, the component with more vertical magnetic field is termed background field; the atmospheric parameters of this component are collected in one map. The component with more horizontal field is termed flux tube component, and put into a second map. The shortcuts bg field and ft component will be used frequently.

Fig. 5.14 outlines a limitation of the selection procedure, which basically stems from the sunspot structure. The LOS velocity due to the Evershed effect is passing through zero on the line that is perpendicular to the symmetry axis. In the profiles observed there, the contributions of bg and ft component are exactly at the same spectral wavelength position (no Doppler shift). They differ mainly in their inclination and field strength, which changes the amplitude and the position of the Stokes V lobes. Fig. 5.15 displays, how the resulting Stokes V profiles look like in this case. The signature of the ft component is lost, opposite to the neutral line in V (cf. Figs. 5.2 and 5.3). On these locations, the spectra do not have enough information to clearly separate the properties of two components. The selection by inclination on these places is at least doubtful. On the limb and the center side of the spot, the two components can be distinguished in the profiles, and the inclination criterion is valid due to a sufficiently large difference of the inclination values.

Consequences of classification The classification is made by the value of one parameter only, the inclination to the solar surface. An investigation of the resulting maps of the other atmospheric parameters will show whether or not the classification makes sense: the two components should differ significantly in, e.g., the line-of-sight velocities or the field strength, if the classification really selects two different types of structures.

Fig. 5.16 displays the velocities of the background field and flux tube component. It can clearly be seen that the ft component carries the largest fraction of the Evershed-flow, while in the bg component only in the outer part of the limb side of the spot velocities above 0.5 kms^{-1} can be found; otherwise the material in the bg field is at rest.

An equally clear distinction between the two components can be seen in Fig. 5.17, which displays the field strength and the temperature of the two components together with their difference. The bg field is generally stronger in the complete penumbra, only at the outer penumbral boundary the field strength of bg and ft becomes identical. On the center side of the spot, a pattern of radially aligned structures with enhanced field strength can be seen in the bg field. These so-called "spines" (Westendorp Plaza et al. 2001a,b; Mathew et al. 2003) will be of interest in the later discussion. For the temperature, the ft component is markedly hotter than the bg at the inner penumbral boundary on the limb side but cooler on the center side. This point is of much interest in the comparison with the Moving Tube model of Schlichenmaier et al. (1998).

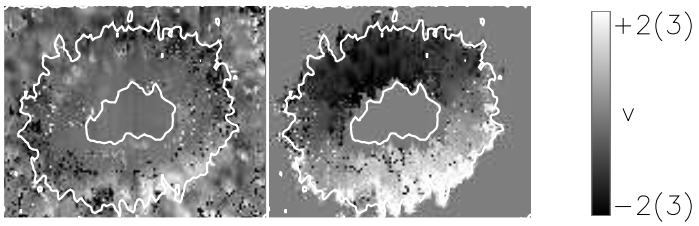


Figure 5.16: LOS velocity of the background (**left**) and the flux tube component (**right**) after the separation of the components. The range is $\pm 2\,(3)$ kms^{-1} for the bg (ft). Negative velocities are towards the observer.

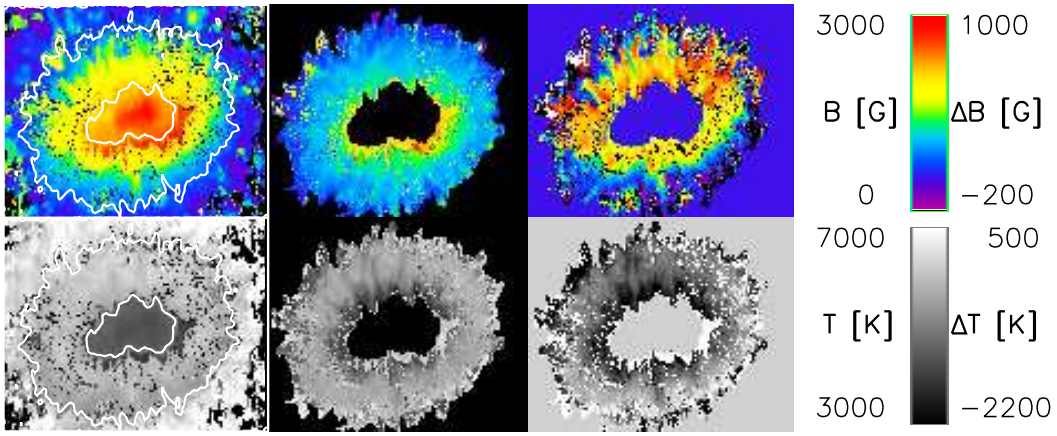


Figure 5.17: **Top row, left to right**: Magnetic field strength of the background field, same for the flux tube component, difference bg-ft. Red in the difference indicates that the bg field is stronger, black that it is weaker. **Bottom row**: temperature of the bg field, same for the ft component, difference ft-bg. White in the difference indicates that the ft is hotter, black that it is cooler.

Fig. 5.18 displays the amount of straylight and the relative filling fraction of bg and ft component in the inversion. The straylight contribution stays more or less constant throughout the umbra and the penumbra, respectively. Only at the outer penumbral boundary, its value increases strongly on a short spatial scale. However, the increase at the outer penumbral boundary does not indicate the presence of field-free material in the resolution element. For the inversion code, the straylight contribution is mainly an efficient tool to create a higher intensity level in the synthetic profiles inside the sunspot. In the filling fraction, bg and ft behave different: in the inner penumbra the bg field contribution to the best-fit profile is stronger, in the middle to outer penumbra the ft component dominates. In the filling fraction of the bg component (middle image), the spine structure on the center side is again clearly visible.

Concluding this section, the separation by the inclination to the surface selects two types of structures, which markedly differ in all atmospheric parameters: the ft component has weaker field, shows larger flow velocities, and is hotter on the inner limb side penumbra. This behavior justifies the approach chosen.

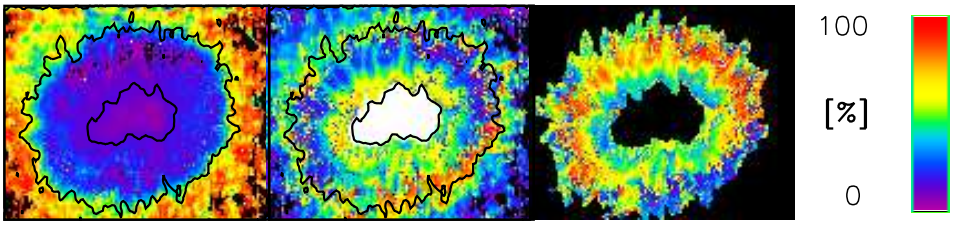


Figure 5.18: **Left to right**: straylight coefficient, filling fraction of background field, same for flux tube component.

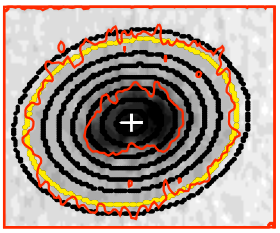


Figure 5.19: Ellipses for the calculation of azimuthal and radial variation of various atmospheric parameters. **Red**: intensity contours outlining the outer borders of umbra and penumbra. **Yellow**: the ellipse which defines the spot radius $r/r_{spot} = 1$. The white cross marks the center for the elliptical paths. The innermost "penumbral" path cuts partly through umbra and penumbra because the spot is not very regular.

5.4 Direct analysis

The term *direct analysis* refers to information that can be retrieved from the 2-component inversion results with few additional assumptions, similar to the *direct observables* of Sect. 3.2. To support the visual impression of the 2-d maps in the preceeding section, Sect. 5.4.1 quantifies the respective properties of the two components inside the sunspot penumbra, the background field and the flux tube component, from azimuthal averages. Sect. 5.4.2 describes the derivation of the direction of the Evershed flow from the observed velocities. Sect. 5.4.3 compares the flow orientation with that of the magnetic field vector to address the question of field-aligned flows. Sect. 5.4.4 summarizes the direct analysis.

5.4.1 Radial variation of parameters

The radial variation of a quantity is determined from the average over the azimuthal paths displayed in Fig. 5.19. *Azimuthal* now refers to the spatial position inside the spot, not the orientation of the magnetic field vector. The azimuth inside the spot measures the angle between disc center direction and any other location. The limb side has 180° azimuth. The spot is foreshortened along the symmetry line by the off-center projection effects. In the innermost penumbra the ellipses cut partly through umbral and penumbral areas. This has been taken into account by restricting the average of the flux tube component to pixels, where two inversion components were present. For the background field, in the umbra the results of the one-component fit are used.

Fig. 5.20 displays the radial variation of six parameters. Starting in the lower right, the intensity in visible and infrared continuum wavelengths has been used to place the transition from umbra to penumbra at $r/r_{spot} = 0.4$, where r is the distance from the spot center, and r_{spot} the spot radius of around 13000 km (= 13 Mm). The middle right panel displays that the straylight level is constant at around 10 % in the umbra, and around 15 % in the penumbra. The filling fraction of the bg field drops from 70 % at the inner penumbral boundary to a minimum value of 40 % at $r/r_{spot} = 0.8$, and levels off at 50 % at the outer penumbral boundary. The ft component naturally does the opposite, as its filling fraction is 100 % minus that of the bg. The uppermost panel in the right column displays the average value of the absolute line-of-sight velocity, $|v_{LOS}|$, of both components. The ft component shows a value around 1.5 kms^{-1} as soon as it is present in the elliptical paths, while the background only shows significant velocities above 0.5 kms^{-1} for $r/r_{spot} > 0.8$.

The temperature plot in the lower left panel clearly shows the ft component to be hotter than the background field in the innermost penumbra, interestingly already before the umbra-penumbra boundary at $r/r_{spot} = 0.4$ is reached. The shape of the bg field temperature curve follows the observed radial variation of the intensity. The ft component has a similar shape, but it is displaced further out towards the outer penumbral boundary. The counterintuitive decrease of the bg field temperature towards the outer penumbral boundary is connected to the increase of the straylight contribution: the straylight profile added to the synthetic spectra of the atmospheric components resembles a hot (averaged) granular spectrum. The increase of the straylight contribution will increase the continuum intensity level of the final synthetic profiles more efficiently than an increase of the temperature of the model components. The inversion results suffer here from a trade-off between the two options to increase the intensity, which makes the temperature of the model atmospheres less reliable in the outer penumbra.

From the plot of the inclination to the local surface it can be clearly seen that on average the ft component is pointing downwards for $r/r_{spot} > 0.8$. The background field never reaches 90°, the maximum inclination at the outer penumbral boundary is close to 60°. Finally, the plot of field strength shows the ft component to be weaker than the background by about 800 Gauss in the innermost penumbra, but the field strength at the outer penumbral boundary is similar. The field strength is decreasing monotonically for both components. For the bg field, the slope of B with radius changes at the umbra-penumbra boundary, being steeper in the penumbra. For the ft component,

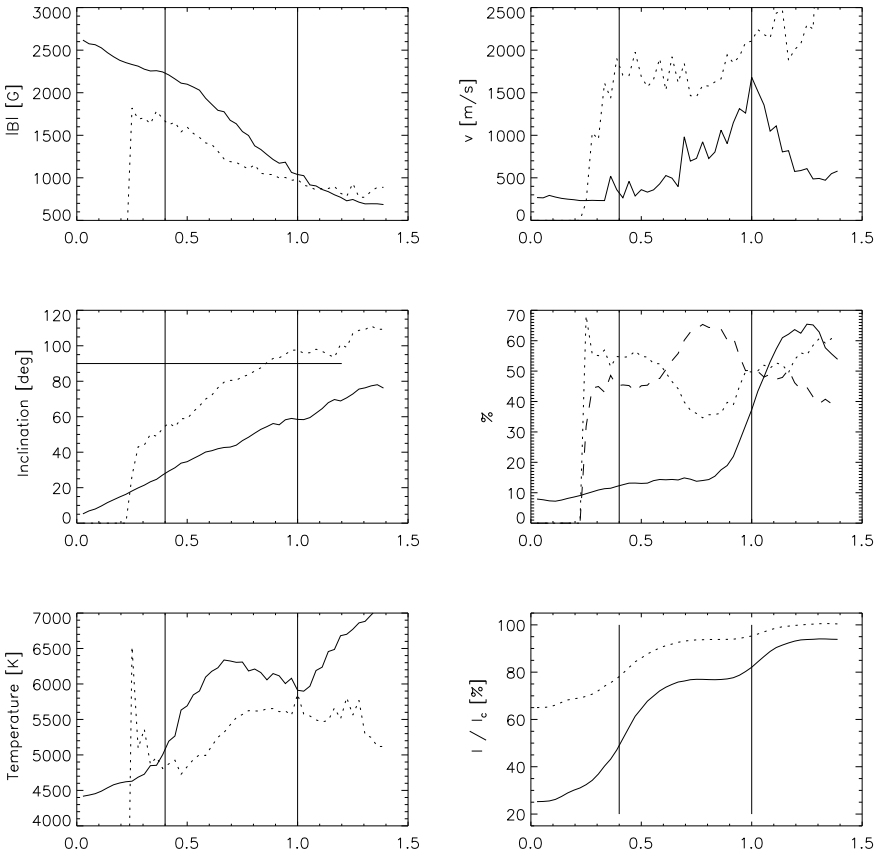


Figure 5.20: **Left column, top to bottom:** Radial variation of $|B|$, LRF inclination, temperature at log τ = 0. **Solid**: background field, **dotted**: flux tube component. **Right column, top to bottom**: averaged absolute LOS velocity (**solid**: bg, **dotted**: ft), straylight amount (**solid**) with overplotted filling fractions (**dotted**: bg, **dashed**: ft), intensity of visible (**solid**) and infrared continuum (**dotted**) normalized to the quiet Sun intensity. The x-axis gives the fractional radial distance r/r_{spot}, the black horizontal lines mark the inner and outer boundary of the penumbra. The horizontal black line in the inclination plot marks the 90° level, values above the line correspond to field lines going down.

the slope changes in the mid penumbra around $r/r_{spot} \sim 0.7$, where the ft filling fraction starts to increase, to a less steep decrease in the outer penumbra.

5.4.2 Horizontal and vertical velocities

If the observed Evershed flow is a radial outflow, the flow field for all points (x,y) inside the penumbra can be described by two parameters, the inclination to the surface, $\alpha(x,y)$, and the velocity, $v(x,y,)$. If one examines scales, which are larger than individual flow filaments, one expects that the dependence on the spatial position (x,y) reduces to a dependence on the radial distance from the spot center, $r = \sqrt{x^2+y^2}$. Then, going azimuthally around the spot center at a given radius r, the projection of the flow field $\mathbf{v}$ on the line of sight leads to a sinusoidal variation of the line-of-sight velocity:

$$v_{\mathrm{LOS}}(r,\phi) = \mathbf{v}(r,\phi)\cdot\mathbf{LOS} = v_h(r)\cdot\sin\theta\sin\phi + v_z(r)\cdot\cos\theta\,, \tag{5.4}$$

where ϕ measures the azimuthal angle around the spot center, v_h and v_z denote the horizontal, respectively, vertical component of the velocity field, **LOS** denotes the direction of the line of sight expressed in the local reference frame, and θ is the heliocentric angle of the spot (e.g. Schröter 1967; Title et al. 1993; Schlichenmaier & Schmidt

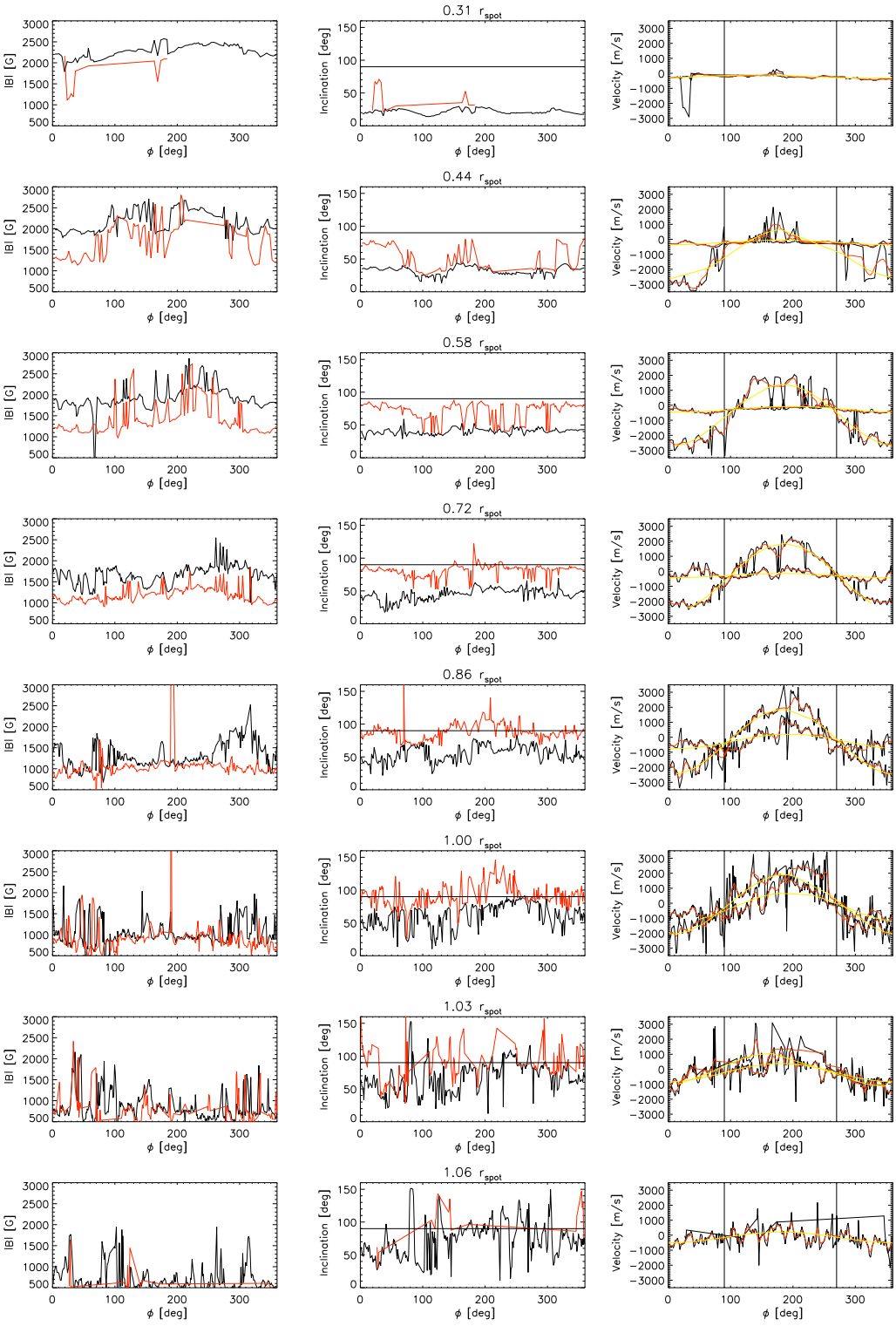


Figure 5.21: Parameter values of |B|, inclination in LRF, and LOS velocity along the azimuthal curves for different radii. **Black**: bg field, **red**: ft component. In the velocity plot the black curve is the LOS velocity, red is the smoothed curve used as input for the fit, and yellow is the best-fit curve, for both bg and ft respectively.

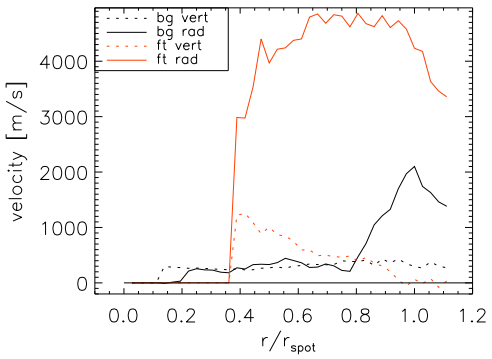


Figure 5.22: Horizontal and vertical velocities for both components from the fit of a sinusoidal to the azimuthal variation of v_{LOS}. **Black:** background field. **Red:** flux tube component. **Dashed:** vertical velocity. **Solid:** horizontal velocity. The ft component shows downward oriented flows ($v_z < 0$) in the outermost penumbra. The background field carries a significant fraction of Evershed flow ($v_h \sim 1\text{kms}^{-1}$ for $r/r_{\text{spot}} > 0.8$).

2000; Bellot Rubio et al. 2003, 2004). The inclination of the flow direction – termed *flow angle* in the following, for the sake of brevity and to make a distinction to the field inclination – is then given by $\alpha(r) = \arctan(v_z/v_h)(r)$.

To derive v_h and v_z from the observations, a fit of a sinusoidal, $a + b \cdot \sin\phi$, is made to the LOS velocities of the inversion components on the azimuthal paths of Fig. 5.19. For each radial distance from the spot center, v_h and v_z are then given by

$$v_h = \frac{b}{\sin\theta} \quad \text{and} \quad v_z = \frac{a}{\cos\theta}\,, \tag{5.5}$$

For illustration, Fig. 5.21 shows the azimuthal variation of field strength, inclination, and LOS velocity for the background field and flux tube at some radial positions. The curve of the ft component velocities shows a clear sinusoidal for $r/r_{\text{spot}} > 0.4$; the bg field only for $r/r_{\text{spot}} > 0.8$. On small scales, the curves are dominated by the properties of the individual radially aligned filaments, which may have different velocities. To reduce the influence of these small scale variations on the fit, the data are smoothed in the azimuthal direction. The fit is iterated once without all azimuthal positions with a large deviation to the first fit curve. The values used later in the comparison of flow angle and field inclination have been taken only from the same azimuthal points as in the last step of the fit for consistency.

The fit result is only significant in the mid to outer penumbra, because the flow velocity has to be large enough to produce a detectable systematic azimuthal variation through the term $v_h \cdot \sin\theta \sin\phi$. Fig. 5.22 displays the radial and vertical velocities found from the sinusoidal fit. The bg field shows almost zero horizontal velocity up to r/r_{spot} = 0.8. From there outwards, it participates in the Evershed effect; the velocity at the outer penumbral boundary is about 2 kms^{-1} in the horizontal direction. The vertical velocity component is always positive, which is in agreement with the fact that the field inclination of the bg field never reaches 90° (cf. Fig. 5.20). The ft component has an almost constant horizontal velocity of 4 kms^{-1}; its vertical velocity is smaller than zero, i.e. the flow angle points downwards, at the outer penumbral boundary.

The flow angle and especially the question, if it is above or below 90°, critically depends on the offset from zero, a, of the fit curve: the sign of v_z is equal to the sign of a. $v_z < 0$ implies downwards oriented flows regardless of the value of v_h. However, the value of a is mainly determined by the average value of v_{LOS} along the azimuthal path and thus depends also on the setup of the wavelength scale: an error in the wavelength scale of ± 100 m/s will change v_z by a slightly larger amount, which may be enough to change from upwards to downwards oriented flows. The flow angle thus has to be regarded with some caution, taking additionally the assumption of a perfectly radial flow field with only a radial dependence of parameters into account.

5.4.3 Field-aligned flows

After the derivation of the flow angle from the horizontal and vertical component of the flow field, one can compare it with the inclination of the magnetic field. For the background field the two inclination values do not match (cf. Fig. 5.23). The flow angle is about 20° larger almost everywhere. However, the slope of the curve of the flow angle is similar to that of the field inclination. For the flux tube component, the flow angle and field inclination match quite well for $r/r_{\text{spot}} > 0.7$. The difference between the two values is less than 3° in the outer penumbra. In the inner penumbra the flow angle is larger than the field inclination.

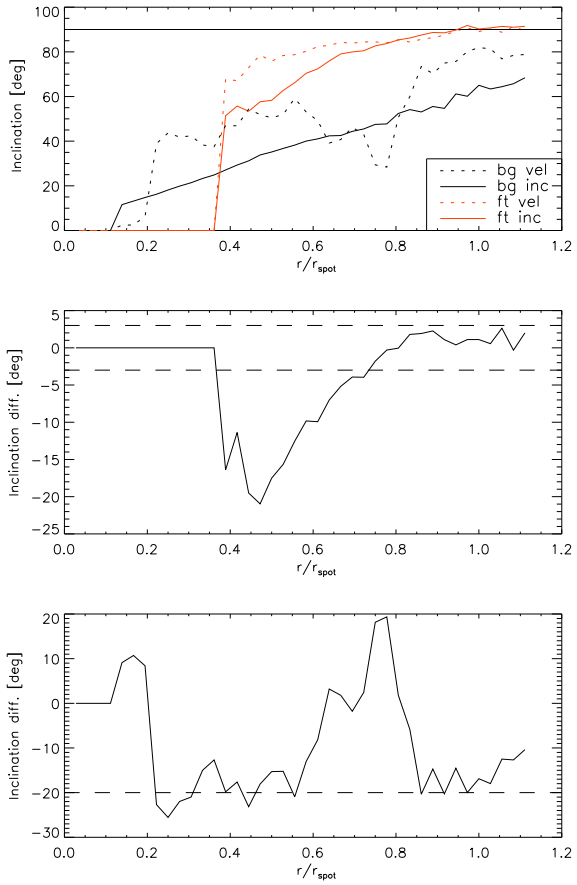


Figure 5.23: **Top row**: Inclination of magnetic field from the inversion (**solid**), and the flow angle derived from the azimuthal variation of the LOS velocity (**dashed**). The background field is in black, the ft component in red. **Middle row**: Difference of flow angle and field inclination for the ft component, dashed lines at $\pm 3^\circ$. **Bottom row**: Same difference for the background field, dashed horizontal line at -20°.

The derivation of the flow angle depends on parameters that are sensitive to both the noise and the evaluation procedure, e.g., the mean velocity on an azimuthal path. The method is based on the assumption that the spot has no azimuthal variation of structure at all, but only a radial dependence in all parameters, which will not be fully true given the penumbral fine structure. Hence, it is not surprising that a good agreement between flow angle and field inclination is only found in the flux tube component in the mid to outer penumbra. Only there the amplitude of the azimuthal variation of the line-of-sight velocity is large enough to make a satisfactory fit of a sinusoidal, and the error in the vertical velocity v_z can be neglected compared to the much larger radial velocity v_h. For small radial velocities the flow angle is dominated by the systematically more inaccurate vertical velocity v_z. The rather constant offset of 20° between flow angle and field inclination of the bg field in the 2-component inversion may well stem from a problem with the zero level of velocities due to the setup of the wavelength scale (cf. Sect. 3.3.5.2).

From the horizontal and vertical velocity, and the reasonable agreement – taking into account the limitations just cited – between flow angle and field inclination for the ft component it seems that the bg field takes part in the Evershed flow in the outer penumbra, starting at $r/r_{spot} = 0.8$. As the bg inclination is never above 90°, a significant outflow of material into the magnetic canopy above the photospheric layers must exist. The flow field is found to be parallel to the field lines for the ft component in the outer penumbra.

5.4.4 Summary of direct analysis

The radial variation (cf. Fig. 5.20) of the field strength, B, and the inclination, γ, agrees well with the findings of Bellot Rubio et al. (2004) and Borrero et al. (2004). The flux tube component is weaker by 1000 Gauss in the innermost penumbra, but reaches the same strength as the background field at the outer spot boundary due to its smaller radial decrease. The azimuthally averaged inclination of the flux tube component is above 90° with respect to the local vertical in the outermost penumbra, i.e. on average the flux tube component returns to the solar surface

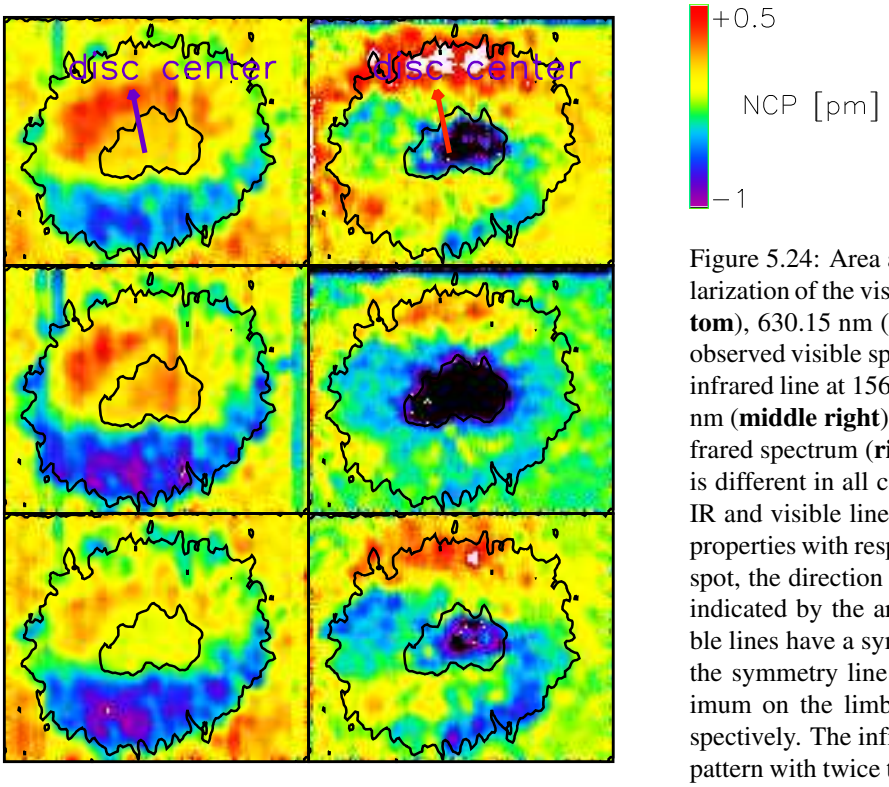


Figure 5.24: Area asymmetry or Net Circular Polarization of the visible line at 630.25 nm (**left bottom**), 630.15 nm (**middle left**), and of the whole observed visible spectrum (**left top**). Same for the infrared line at 1564.8 nm (**right bottom**), 1565.2 nm (**middle right**), and of the whole observed infrared spectrum (**right top**). The displayed range is different in all cases (~ -1 to ~ 0.5 pm). The IR and visible lines show very distinct symmetry properties with respect to the symmetry line of the spot, the direction from spot center to disc center indicated by the arrow in the top row. The visible lines have a symmetric pattern with respect to the symmetry line with one maximum and minimum on the limb side and the center side, respectively. The infrared lines have an asymmetric pattern with twice the frequency, two maxima and two minima.

close to or even inside the spot boundary. The background field never increases beyond about 60° in inclination, i.e. the field lines continue upwards.

The filling fraction does slightly differ from Bellot Rubio et al. (2004): it levels off at about 50 % in the inner penumbra and does not drop smoothly to zero as in Bellot Rubio et al. (2004). This may be due to an artificial cutoff in the inversion: umbral profiles below an intensity threshold were forced to be 1-component fits; the 2-component inversion ends abruptly at the inner umbra-penumbra boundary defined from the whitelight intensity.

The temperature of the flux tube component is higher than the background temperature in the innermost penumbra, and lower in the mid to outer penumbra.

The amplitude of the sinusoidal variation of the line-of-sight velocity on azimuthal paths reaches 2 kms^{-1} for the flux tube component; the projection to horizontal and vertical velocity yields horizontal velocities up to 6 kms^{-1}. The horizontal velocity is roughly constant throughout the penumbra. The background field participates in the Evershed effect in the outer penumbra; its horizontal velocity reaches 2 kms^{-1} at the outer spot boundary.

The flow angle, derived from the horizontal and vertical velocity, is not in good agreement with the field inclination for the background field. However, field inclination and flow angle of the flux tube component match quite well in the outer penumbra for $r/r_{\mathrm{spot}} > 0.7$, where the error in the vertical velocity, v_z, can be neglected in comparison with the large horizontal velocity, v_h.

5.5 The Net Circular Polarization

The *Net Circular Polarization* (NCP or area asymmetry) will be discussed here after the direct analysis of the previous section, because it is a direct observable, whose reproduction requires to improve the 2-component inversion. The NCP is defined as the Stokes V signal, integrated in wavelength, $\int V(\lambda) d\lambda$ (cf. Sect. 3.2). If the parameters of the magnetic field vector – field strength, inclination, and azimuth – and the flow velocity inside the magnetic field are constant with depth, the resulting Stokes V signal is fully anti-symmetric with respect to the zero wavelength given by the flow velocity. The integration has to yield a zero NCP value in this case. If the the area below the two lobes – of opposite sign – of the Stokes V signal is not identical, a non-zero NCP will remain after the integration. The NCP is indicative of two possible effects: firstly, gradients with optical depth in the physical parameters (other

than the temperature) can lead to a non-zero NCP, and secondly, discontinuities along the line of sight, which can be interpreted as very strong localized gradients, can produce a non-zero NCP.

An analytical derivation yields the following relation (cf. Illing et al. 1975; Solanki & Pahlke 1988; Sanchez Almeida et al. 1989) between gradients in field strength, $\frac{d|B|(\tau)}{d\tau}$, and velocity, $\frac{dv(\tau)}{d\tau}$, and the sign of the NCP:

$$\text{signum}(\text{NCP}) \propto \text{signum}\left(-1\frac{\mathrm{d}|\mathrm{B}|(\tau)}{\mathrm{d}\tau}\frac{\mathrm{dv}(\tau)}{\mathrm{d}\tau}\right), \tag{5.6}$$

whereas the absolute value can only be calculated for a given atmospheric stratification.

The NCP is a very sensitive parameter in a twofold sense: it is sensitive to gradients of the atmospheric quantitites and, unfortunately, to the quality of the observations. The NCP "measures" the difference in the area of the two opposite signed lobes in the Stokes V signal – which may be a small value due to the subtraction – and, hence, is easily influenced by the noise level or intensity gradients in the spectral dimension.

The actual NCP value of a profile thus shows a larger error than most other direct observables. However, already the shape of the NCP values in large-scale 2-D maps reveals qualitative information, without relying on the actual NCP values on single positions. Fig. 5.24 displays NCP maps of the spot, measured in the magnetic sensitive visible, respectively, infrared spectral lines. The differences between visible and IR are very prominent: the NCP of the visible spectral lines is symmetrical with respect to the symmetry line of the spot, the line connecting spot center and disc center. The NCP of the visible spectral lines has a maximum on the limb side of the spot, and a minimum on the center side[1]. Contrary to that, the infrared lines show an anti-symmetric pattern with respect to the symmetry line, with two maxima and minima on a circle around the spot center.

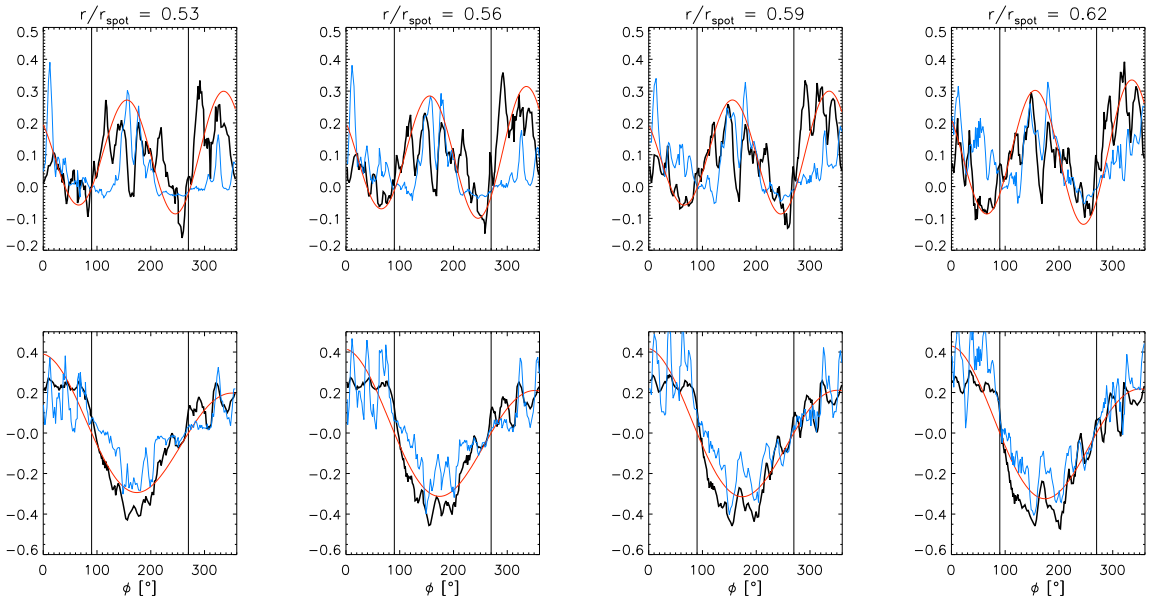


Figure 5.25: Observed Net Circular Polarization in pm along azimuthal paths for the infrared spectrum (**top row**) and the visible spectrum (**bottom row**) (**thick black**). Overplotted in (**red**) is a sinusoidal with a double period for the infrared and a single period for the visible spectrum. The amplitude of the sinusoidal depends on the azimuth angle Φ. The (**blue**) curve is the NCP of the best-fit profiles of the Gaussian inversion. The radial position in units of the spot radius is 0.53 to 0.62. The black vertical lines are at 90° and 270°.

For an easier visualization of the different behavior of the NCP in the two spectral ranges, Fig. 5.25 displays the NCP values of the observed infrared and visible spectrum, i.e. integrated over all spectral lines in the respective wavelength range, on some of the azimuthal paths around the spot center. The paths are located in the mid penumbra, between $r/r_{\text{spot}} = 0.53$ to 0.62. The frequency of the variation agrees well with a dependence on sin ϕ for the

[1]This depends on the polarity of the spot. If the field lines are oriented into the surface, the positions of maximum and minimum NCP are exchanged.

visible lines and sin 2 ϕ for the infrared lines. The amplitude of the sinusoidal is not constant, which is clearly visible in the infrared NCP. For the visible NCP the shape deviates from a pure harmonic function; it resembles more a step function, especially at $\phi = 90°$.

Thus, the NCP is a parameter, which shows a markedly different signature in visible and infrared spectral lines. However, the 2-component inversion without gradients does not use this information at all: the perfectly anti-symmetric best-fit profiles of the 2-C inversion do always lead to a zero NCP value for all spectral lines. To make use of also the NCP of the spectral lines in the derivation of the solar atmospheric stratification, the 2-C inversion was exchanged by a more sophisticated Gaussian inversion (cf. the following section), which is able to produce a NCP in the best-fit profiles.

For comparison with the observed NCP, the blue line in Fig. 5.25 gives the corresponding NCP of the best-fit profiles from the Gaussian inversion. At least for the visible spectra, the fit NCP curve follows the observed NCP quite close. Even if at some places the absolute value is not the same, the shape of the two curves is almost identical: at $\phi = 180°$ in the plot of the visible NCP at r/r_{spot} = 0.5, the local extrema of the observed NCP are exactly repeated in the fit.

The NCP of the infrared spectral lines is not as well reproduced as that of the visible lines. The NCP of the best-fit profiles in the infrared matches the observed NCP only on the limb side of the spot around $\phi = 180°$. The difference between center and limb side is again related to line-of-sight effects: the limb side contains the neutral line of Stokes V, where the background field is almost perpendicular to the line of sight and the properties of the flux tube component can be derived more accurately. Further discussion of the NCP will be postponed to the comparison with the theoretical predictions in Sect. 5.13.

5.6 Evaluation of the observed spectra by an uncombed "Gaussian" inversion

In the previous section it was shown that the 2-component inversion without gradients does not use all information available in the observed spectra, because it cannot reproduce the NCP. There is a second argument for the use of a more sophisticated model in the inversion: in the 2-C inversion, the two components are assumed to be unrelated and *spatially* separated in each resolution element. However, in the picture of the uncombed penumbra (cf. Fig. 5.6, p. 80), the background field and the horizontal flow channels are *vertically* interlaced, with the background field winding around the flow channels. An improved inversion setup thus should take both points into account.

5.6.1 Inversion setup: uncombed "Gaussian" inversion

The uncombed Gaussian inversion is a modified version of the SIR code (cf. Sect. 3.3.3), which uses the same basic structure as the standard code but has some added program modules. It is "tailor-made" for the problem under investigation in this work, the study of the penumbral fine structure. Similar to the 2-component inversion, it contains two magnetic model atmospheres, whose parameters are constant with optical depth. However, to one of the components a perturbation with a Gaussian shape is added (Bellot Rubio 2003). The effect can best be visualized with the example shown in Fig. 5.26: to each parameter of the atmospheric stratification connected with the magnetic field (field strength, field inclination, field azimuth) or the thermodynamic properties (temperature, line-of-sight velocity) a Gaussian is added, with an identical central position in the log τ scale of the stratification and with an identical width in all parameters. The central position and the width of the Gaussian can be changed by the code to minimize the χ^2-value. The amplitude of the Gaussian is a free parameter for each physical property, which is fitted independently; it can be positive or negative. The resulting stratification resembles an embedded flux tube as close as possible without including "real" discontinuities, which are very difficult to treat in the profile synthesis. The transition between background and flux tube properties thus may have strong gradients, but still is always a smooth curve.

In principle, there are several options to include the presumably uncombed structure of the penumbra into an inversion code. The jump in field properties from the background field to the flow channels has to be quantified and put into a geometrically constrained model. The implementation of the Gaussian perturbation was chosen for mainly two reasons: firstly, it is relatively easy to add to the inversion code. It introduces only a limited number of new free parameters, which however allow to generate a lot of different geometrical situations. A second argument against using a step function like in Borrero et al. (2005) – which would be the first idea of how to include the

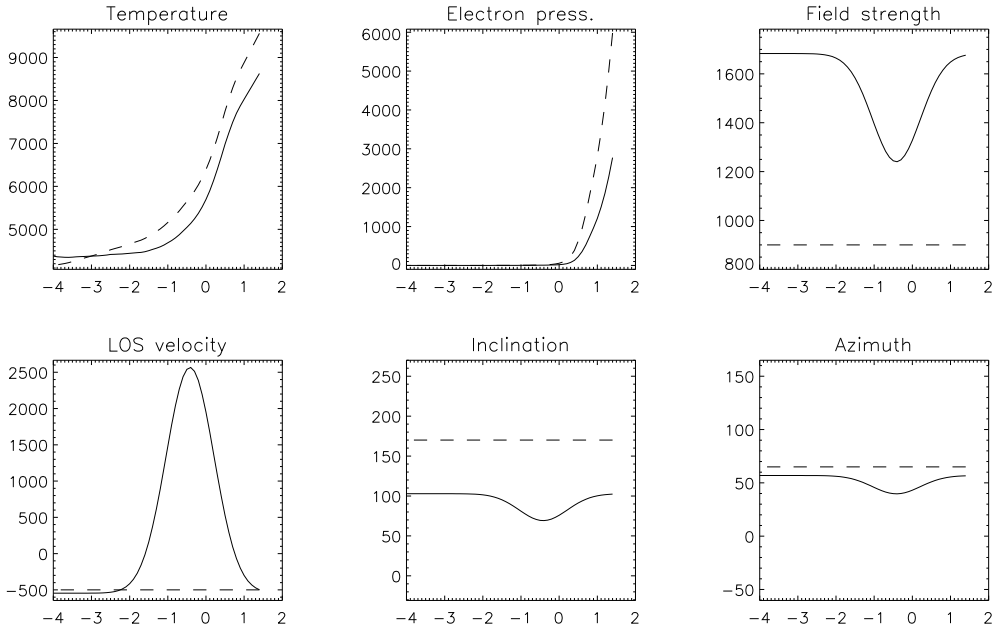


Figure 5.26: Example of the best-fit model atmosphere from the Gaussian inversion. **Top row, left to right**: Temperature in Kelvin, electron pressure, field strength in Gauss. **Bottom row**: line-of-sight velocity in m/s, field inclination to the LOS in degree, field azimuth in degree. **Solid**: Final best-fit model atmosphere, **dashed**: Initial model atmosphere of the background field.

discontinuities – is the horizontal extension of the flow channels. If the flow channel is not resolved, the light rays from one resolution have encountered different geometries. Some of them will have passed the center of the flow channel, whereas others will have only grazed the outer part. This leads to a distribution of the encountered jumps in magnetic properties, which the Gaussian shape automatically includes. In the following, this inversion setup will be referred to as "Gaussian inversion", to remind the reader of the atmospheric geometry depicted in Fig. 5.26. However, it should also be kept in mind that the Gaussian perturbation is only one way to model vertically interlaced magnetic components.

5.6.2 Derivation of the initial model atmosphere

The drawback of this inversion setup is a rather strong sensitivity on the values of the initial model. Opposite to the 2-component inversion, the initial model cannot be chosen to be the same for all spatial positions; in fact, it has to be chosen very carefully for each spatial position. However, a good first guess is available from the 2-C inversion, after the two model atmospheres have been sorted by their respective inclination to the surface:

- The best-fit model corresponding to the less inclined component in the 2-C inversion was substituted as initial model for the background field in the Gaussian inversion.
- The amplitude of the Gaussian perturbation was estimated by subtracting the respective values (T,B,v, γ,ϕ) of the two model atmospheres in the 2-C inversion from each other.

This procedure for the derivation of the initial model values was found to be able to yield an initial model, which the code was able to iterate to an improved best-fit solution in most cases.

The central position and the width of the Gaussian perturbation were set to an initial value of -0.5 in log τ, respectively, a width of 0.5 in log τ units. These initial values were constant throughout the spot.

A comparison of the fit quality between the 2-component and the Gaussian inversion is displayed in Fig. 5.27. It displays the best-fit profiles of the two inversion setups for the same spatial position, together with the actual observed spectra (black). Already the 2-component inversion (blue) traces the shape of the profiles quite well, even

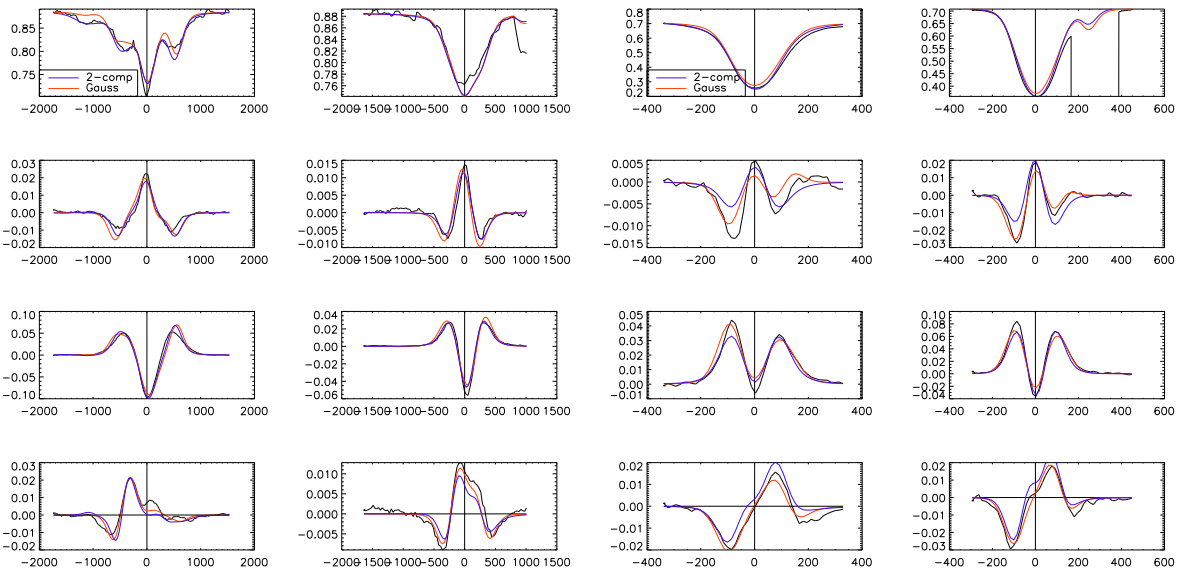


Figure 5.27: Comparison of the best-fit quality of the 2-component inversion (blue), the Gaussian inversion (red), and the observed profiles (black) for one spatial position. **Left to right**: the infrared lines at 1564.8 nm and at 1565.2 nm, the visible lines at 630.15 nm and 630.25 nm. **Top to bottom**: Stokes *IQUV* as fraction of the continuum intensity in quiet Sun. The black vertical lines mark the zero-wavelength of the respective wavelength scales. The dispersion on the x-axis is in mÅ for all spectra.

if it does not match the amplitudes at all wavelengths. The Gaussian inversion (red) yields a better fit; it reproduces the linear polarization signal of the visible spectral lines (2nd row, 3rd and 4th column) much closer. However, the main argument for the use of the Gaussian inversion is not necessarily only a better fit quality, but the more realistic structure of the solar atmosphere in the model. The best-fit profile from the Gaussian inversion in this case demonstrates that all observed spectral lines can be reproduced simultaneously by an embedded flow channel in a more vertical background field.

5.6.3 The formation height of spectral lines

Despite the claim just made, an important question has not been addressed up to now: why should the Gaussian inversion be able at all to place the perturbation at some specific height in the atmosphere ? Where does the necessary information come from ? A hint to the answer is given by the observed NCP: the NCP is produced by gradients in atmospheric properties, thus, each spectral line has to be sensitive to at least some range in the atmosphere, which commonly is called its *formation height*. Each spectral line corresponds to a transition of an electron between two energy levels, an excited state and a corresponding ground state. The electron transitions have specific properties due to the electron orbitals involved in the transition that lead to a slightly different behavior with respect to changes of temperature (photon density) and gas pressure (absorber density and collision probability). In the stratified solar atmosphere, the physical height, where most of the absorption takes part, is thus not identical for all transitions. Commonly, this is expressed by the so-called *contribution function*, $C_\lambda(h)$, which gives the amount of absorption at a given wavelength as a function of height in the atmosphere, h. The contribution function has a single maximum in the solar atmosphere: at the solar surface, the high density and temperature leads to an equilibrium between absorption and emission through collisional induced excitation ($C_\lambda(0) = 0$), whereas in the upper atmospheric layers the density is so reduced that almost no absorption happens ($C_\lambda(h \gg 1000 \text{ km}) \sim 0$). The range where the contribution function exceeds a certain threshold, usually 1/e of the maximum value, is called the formation height of the spectral line.

The formation height of the infrared lines is deep in the photosphere with a very restricted height range, in which the contribution function is large. The formation height of the visible lines extends to higher layers, and includes that of the IR lines (see for example Muglach & Solanki 1992; Bellot Rubio et al. 1996; Cabrera Solana et al.

2005). Thus, the information of the IR lines is suited to fix the thermodynamic and magnetic field parameters in a certain height range, while the information on height variations is contained in the visible lines.

Consequences for the uncombed "Gaussian" inversion The presence of embedded (unresolved) flow channels inside the background field leads to discontinuities or equivalently strong localized gradients in physical parameters along the line of sight. The LOS will encounter the flow channel at a certain optical depth and will either pass completely through it or only pass the upper boundary between flow channel and background field, depending on the integrated optical thickness of the flow channel. The jumps in parameter values are able to break the perfect (anti-)symmetry of the polarization signal. Additionally, the observed infrared and visible spectral lines form in slightly different height ranges of the solar atmosphere.

Taken together, this yields enough information to place the Gaussian perturbation correctly in the optical depth scale: a strong NCP in the visible lines but none in IR implies that the flow channel is high in the solar atmosphere, whereas a strong NCP in IR and a small one in the visible implies a low lying flow channel. The NCP value depends also on the question, if the boundary between flow channel and background field is passed once (upper boundary) or twice (upper and lower boundary). Thus, the observed spectra contain enough information to locate the flow channel in the optical depth scale and to determine its width, in addition to the derivation of the magnetic field parameters, which the 2-C inversion already was able to retrieve.

However, one should not forget that the inversion problem cannot be solved uniquely. The Gaussian inversion is *systematically* better than the 2-C inversion, because it uses the information on the NCP in the fit. Unfortunately, the geometry of the inversion setup is in principle designed with some a priori assumptions: the Evershed flow should happen in more or less horizontal embedded flow channels. Even with a perfect reproduction of a set of observed profiles, in principle the only proof gained is that the model of embedded flow channels is *not in contradiction* with the multi-wavelength observations in several spectral lines.

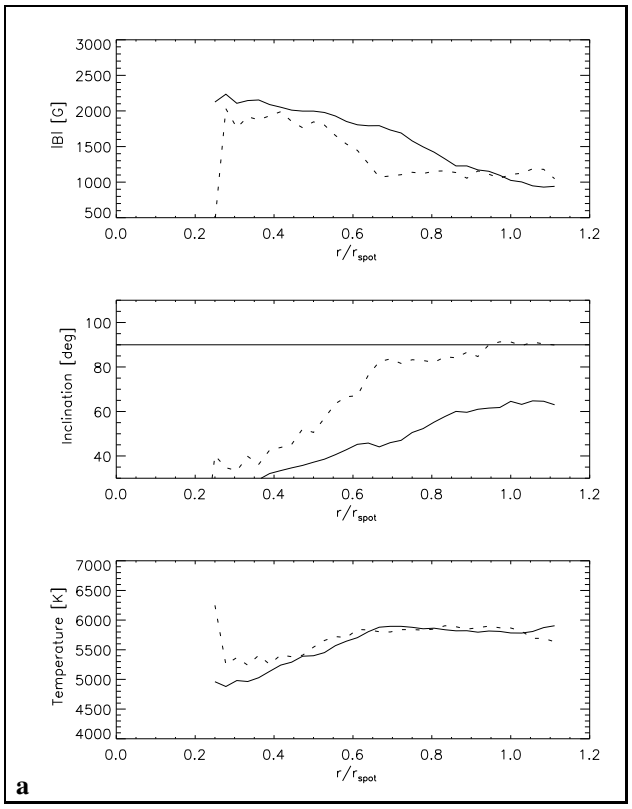
5.7 Results of the uncombed "Gaussian" inversion

The results of the Gaussian inversion will mainly be analyzed with respect to the parameters, which the 2-C inversion cannot retrieve. There will be a short comparison with the corresponding results of the 2-C inversion in the following section. The agreement between the observed NCP and that of the best-fit profiles will be examined in Sect. 5.7.2. The properties of the flow channel, given by the location and the width of the Gaussian perturbation, are discussed in Sect. 5.7.3. The conversion from optical depth to geometrical height in Sect. 5.8.1 leads finally to the 3-dimensional structure of the flow channels described in Sect. 5.8.2.

5.7.1 Comparison with the 2-component inversion

A direct comparison of Gaussian and 2-C inversion is difficult, because the atmospheric properties are constant with optical depth τ in the 2-C inversion, whereas in the Gaussian inversion they change in the seconnd model atmosphere, depending on the amplitude, location and width of the Gaussian perturbation. To have comparable quantities, the atmospheric properties of the 2nd component in the Gaussian inversion were taken at a level of $\tau = 1$, i.e. the values close to the solar surface were sampled. The inversion components will be termed as before: the less inclined component without the Gaussian perturbation will be called background (bg) field and the component with the added Gaussian will be called flux tube (ft) component in the following. The same analysis as described in Sect. 5.4 was then performed: derivation of the radial variation by azimuthal averages on ellipses and derivation of the horizontal and vertical velocities from a fit of a sinusoidal to the line-of-sight velocity (cf. Sect. 5.4.2).

Radial variation of parameters Fig. 5.28 displays the radial variation of field strength, inclination and temperature of both components in the Gaussian inversion. In general, the picture is similar to that of the 2-C inversion in Fig. 5.20: the ft component is weaker than the bg field, hotter in the inner penumbra, and exceeds 90° inclination inside the spot. The bg field reaches a maximum inclination of around 60° at the outer spot boundary. However, there is a markedly difference in the ft component around $r/r_{\mathrm{spot}} \sim 0.7$: both field strength and inclination show a distinct jump in their values, followed by an almost constant value for larger r. The radial position coincides with an abrupt increase of the filling fraction of the 2nd inversion component. A possible explanation why this radial position should be outstanding in some way will be given in Sect. 5.7.3.



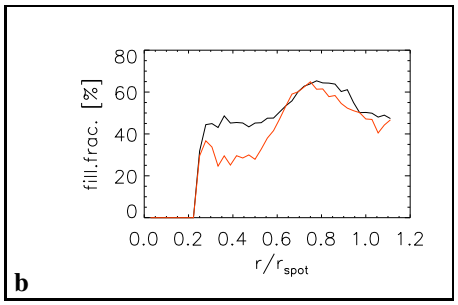


Figure 5.28: **a**: The radial variation of field strength (**top**), inclination (**middle panel**), and temperature (**bottom**) in the Gaussian inversion. **Solid**: bg field. **Dashed**: ft component. The x-axis gives the fractional radial distance $r/r_{\rm spot}$. The horizontal black line in the inclination plot marks the 90° level, values above the line correspond to field lines going down. The ft component shows a distinct jump in its parameters at $r/r_{\rm spot} \sim 0.7$. **b**: Radial variation of the filling fraction of the flux tube component for the 2-component inversion (**black**) and the Gaussian inversion (**red**). The filling fraction levels off at 50 (30) % filling fraction at the umbral-penumbral boundary, and shows a distinct peak around $r/r_{\rm spot} = 0.8$.

Horizontal and vertical velocities For the horizontal and vertical flow velocities, and the flow angle derived from them later, the Gaussian inversion again shows a very similar picture as the 2-C inversion. The background field participates in the Evershed flow in the outer penumbra (cf. Fig. 5.29), the flux tube components shows downflows in the outermost penumbra, and the flow angle and the magnetic field inclination (not shown in the figure) are only in reasonable agreement for the ft component in the outer penumbra.

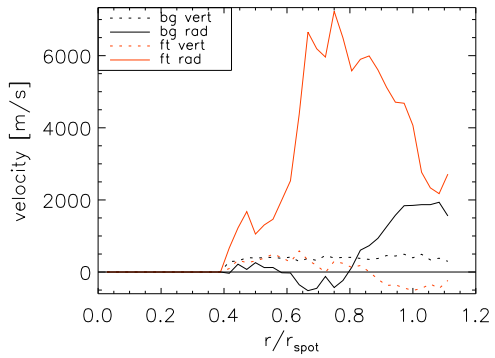


Figure 5.29: Horizontal and vertical velocities for bg field and flux tube component from the fit of a sinusoidal to the azimuthal variation of $v_{\rm LOS}$. **Black:** background field, **red:** flux tube component. **Dashed:** vertical velocity, **solid:** horizontal velocity.

Conclusion of comparison The close similarity between the quantities of the 2-component inversion and the Gaussian inversion, which were examined up to now, is not surprising. The best-fit model atmospheres of the 2-component inversion *give* a good reproduction of the observed profiles (cf. Fig. 5.27); the same values were used as initial model in the Gaussian inversion. Especially for the bg field, which is constant in optical depth in both inversion setups, atmospheric properties were in most cases only slightly modified. The amplitude of the Gaussian perturbation was also found to stay usually close to the difference of the corresponding parameters in the 2-C inversion. This close agreement is to be seen positive: the 2-C inversion yielded a list of properties of the

flow channels (weaker field, higher flow velocity, hotter in inner penumbra); the Gaussian inversion with the added option of locating the flow channels in optical depth does not change this general picture. The similarity suggests that the 2-C inversion already gave a close match of the physical properties.

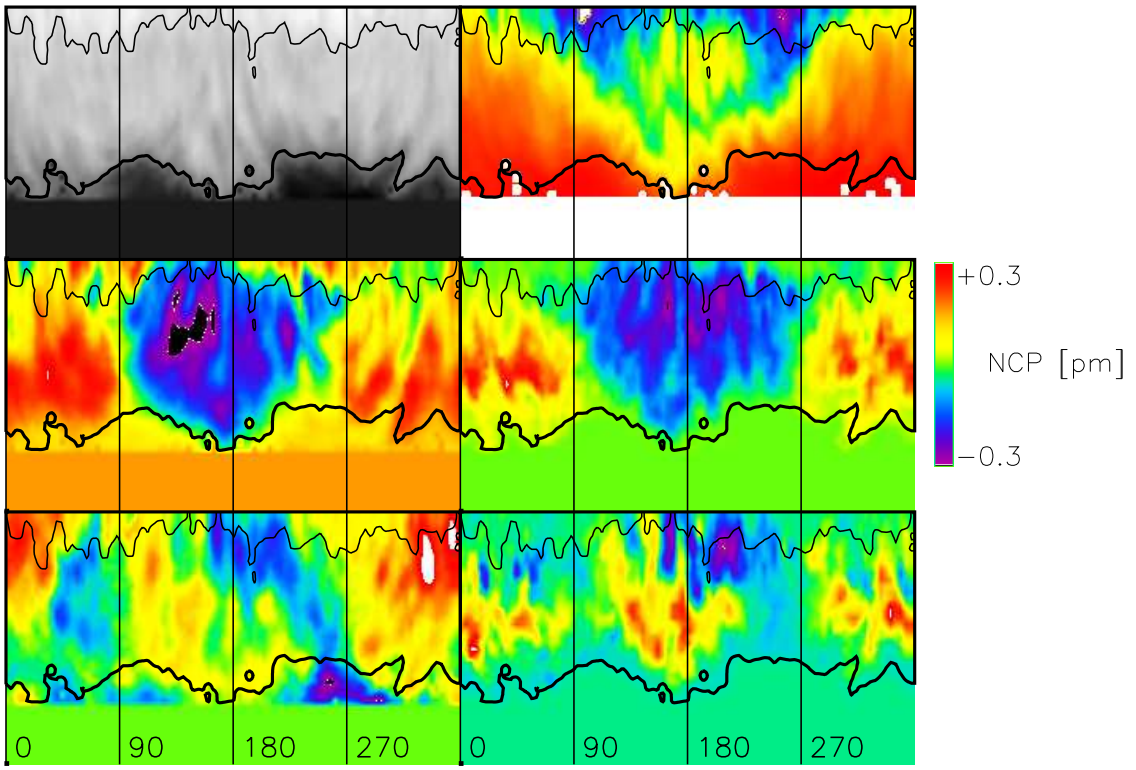


Figure 5.30: Comparison between the observed NCP and that of the best-fit profiles of the Gaussian inversion. The abscissa gives the azimuth angle around the spot center. 0°/180° corresponds to center side, respectively, limb side. The spot center is at the lower border of the images. **Top row, left to right**: continuum intensity in the infrared, integrated absolute Stokes V. The neutral line is visible on the limb side at 180° as the green area enclosed by the stronger signal in yellow. **Middle row, left to right**: observed and fit NCP of the visible line at 630.25 nm. **Bottom row**: same for the infrared line at 1564.8 nm. All NCP maps are displayed in the range from -0.3 pm (black) to 0.3 pm (white). The black contours outline the inner and outer penumbral boundary.

5.7.2 The Net Circular Polarization of the best-fit profiles

The NCP is a source of information in the observed spectra that is used by the Gaussian inversion, but not the 2-component inversion. Naturally, one should have a close look if the Gaussian inversion setup is able to reproduce this quantity in the observations. Fig. 5.30 displays the NCP of both observations and the best-fit profiles of the Gaussian inversion. The respective NCP values have been taken on radially aligned slices of 4° angular width. The limb side is at 180°, the center side at 0°.

For the visible line at 630.25 nm, the agreement is close on both center and limb side. The radially aligned filamentary structure is reproduced by the fit NCP. In comparison with the intensity map, it can be seen that the NCP values are smaller close to dark filaments, but enhanced in the dark filament itself. The best example of this behavior is located at an azimuth angle of around 320°, where a dark filament extends through the whole penumbra, but the NCP is reduced just left of it.

For the infrared line at 1564.8 nm, the agreement between observed and fit NCP is only close on the limb side, where the signature of the more horizontal flow channels can be detected more clearly, because the line of sight is perpendicular to the more vertical background field. For azimuth angles between 0 and 90° and from 270 to 360°, the sign of the fit NCP may be opposite to the observations on some locations.

The difference between the fit NCP of the infrared and visible lines is due to their respective formation height

(cf. Sect. 5.6.3). For the NCP of the IR lines with a restricted formation height – comparable in extent to the diameter of the flow channel – it is crucial to place the flow channel at the correct optical depth, or the IR lines may "miss" the upper or lower boundary of the flow channel. A missing discontinuity can lead to a change of sign in the NCP. The visible lines are sensitive in a larger range; as long as the flow channel is placed anywhere inside this range, they will encounter both the upper and the lower boundary. This may explain, why on the center side even the sign of NCP is wrong for the IR, whereas on the same spatial position the visible NCP is reproduced well enough. On the limb side, the signature of the flow channel can be detected in the Stokes V signal and the channel can be placed precisely by the inversion code, hence, the IR NCP is also reproduced. To conclude, the Gaussian inversion setup reproduces the NCP well enough that the retrieved location of the flow channel is a well determined quantity on most locations. As discussed in the previous section, other quantities like field strength or line-of-sight velocities already were well determined by the 2-component inversion and only slightly modified in the Gaussian inversion.

Note that this is the first time that the NCP map of a whole spot has been reproduced from an inversion of observed spectra. The SIR code is a relatively "fast" inversion code; the 2-component inversion of the full map needed about a week. The Gaussian inversion, restricted to around 6000 pixels in the penumbra and using the final results of the 2-C inversion as initial model, added about the same amount of time. Many previous works were limited to analyze only single radial cuts, due to the time involved in retrieving the solar atmospheric stratification from the observations (e.g., the 750 pixels examined by Borrero et al. 2004).

5.7.3 Location and width of the flow channel

The Gaussian inversion retrieves information about the geometrical vertical structure of the solar atmosphere, in addition to the quantities of the 2-C inversion. The two parameters that can be derived from the final best-fit model atmosphere of the Gaussian inversion are the location and the width of the flow channels, as given by the location and width of the Gaussian perturbation added to the background atmosphere. Both parameters are specified in optical depth, because the inversion code needs an optical depth scale in the synthesis of the profiles. This optical depth scale has to be converted into geometrical height, which is in principle only possible if the gas density or pressure are exactly known. Unfortunately, these quantities cannot be derived directly from the observed profiles. In this case, some assumptions have to be made about the value of the gas pressure inside a sunspot (Sect. 5.8.1).

Fig. 5.31 displays the azimuthally averaged values of location and width as function of the radial distance to the spot center, analogously to Fig. 5.20. Some things are very prominent about the location of the tube:

- For $r < 8$ Mm the tube always stays close to the $\log \tau = 0$ level.
- There is a distinct maximum in the location value at $r \sim 8$ Mm. A strong increase from a value around 0 in the inner penumbra to about 2 in $\log \tau$ is immediately followed by about the same change back to 0 or less for $r > 8$ Mm.
- The location in optical depth stays roughly constant at -0.5 in $\log \tau$ for $r > 10$ Mm.
- A not very pronounced minimum in optical depth – corresponding to the maximum in geometrical height – is reached at around $r = 11$ Mm.

Contrary to this variation of the location of the flow channel, its width stays more or less constant throughout the whole penumbra. It shows a distinct maximum in the innermost penumbra at the umbra-penumbral boundary. A similar maximum as in the location value is also present around 8 Mm.

For an interpretation of the behavior one has to take into account that the values of both parameters are given in the optical depth scale. If the flow channel is located at $\log \tau = 0$, it can be interpreted in two ways: it may indicate a low-lying flux tube or indicate a tube that is optically thick. The isolated maximum at $r = 8$ Mm – with the counterintuitive effect of the flow channel moving to smaller values in $\log \tau$ for *both* larger and smaller radius – suggests that for $r < 8$ Mm the flow channel is either so deep in the atmosphere that only its upper boundary is encountered along the line of sight, or optically thick. In both cases, the inversion code would place the Gaussian perturbation at $\log \tau = 0$, the lower end of the optical depth scale. The following decrease of the location from $\log \tau = 2$ to around -0.5 could then be interpreted as the rise of a flow channel through the surface, where the line of sight is able to penetrate through the tube. The profiles then contain information on where the tube has to be placed. This interpretation of the local maximum at 8 Mm is supported by the relative filling fraction of the bg

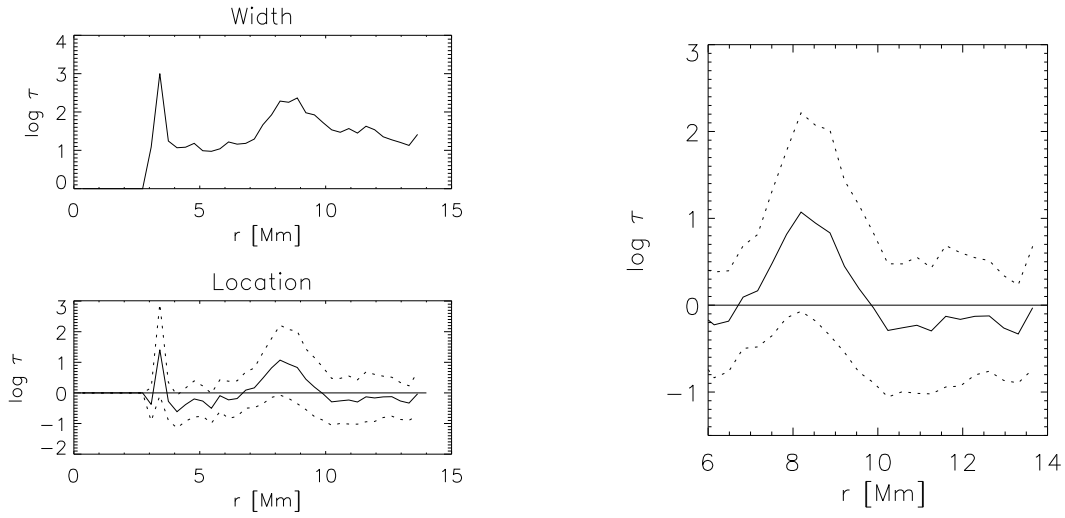


Figure 5.31: **Left panel**: Radial variation of width and position (both in log τ) of the gaussian perturbation in the inversion. Dashed in the position plot is the upper and lower boundary of the tube, position $\pm$ width/2. **Right panel**: Close up of the outer penumbra. The maximum in log τ is reached around 11 Mm, with a slight decrease for larger r.

field and ft component. The 8 Mm correspond to 0.6 $r_{\rm spot}$, where the filling fraction of the ft component strongly increases (cf. Fig. 5.28). This seems to indicate that around this radius the flow channels reach the surface defined by the log $\tau = 0$ level and start to have a stronger contribution to the observed profiles. It also indicates that for the inner penumbra the location of the Gaussian perturbation derived by the inversion code is relatively meaningless. In the inner penumbra, the Gaussian inversion does detect the presence of discontinuities and reproduces the NCP created by them, but it cannot locate them in optical depth. The line of sight does not penetrate through the tube and only the upper discontinuity is seen.

5.8 Derivation of the 3-D topology from the uncombed "Gaussian" inversion

To derive the 3-D structure of the solar atmosphere, the optical depth scale has to be converted to a geometrical height scale. The location and width of the flow channels retrieved by the Gaussian inversion can be converted to the physical location and vertical extension of the flow channels that carry the Evershed flow. The geometry and the additional information on the line-of-sight velocity yield the final model of the horizontal flow channels in Sect. 5.8.2, which was derived from the observed profiles with the least number of assumptions as seemed possible.

5.8.1 Conversion from the optical depth scale to the geometrical height scale

The conversion from the optical depth scale to the actual physical height is not easy, because the optical depth scale – in principle – is unique on each spatial position; it depends on the atmospheric properties of the solar atmosphere on each point. To have nevertheless a first impression, some simplifications are made in the following.

There are two items one can use in the conversion of the optical depth scale to geometrical height. Firstly, solar atmosphere models for quiet Sun have been studied extensively in both observations and simulations in the past decades. This is already used in the inversion: the initial thermodynamic model parameters, i.e. temperature and electron density, are identical to the Harvard Smithsonian Reference Atmosphere (HSRA Gingerich et al. 1971). For the HSRA model atmosphere, gas density and temperature are exactly known, and, hence, the relation between geometrical height and optical depth can be calculated from Eq. (3.14), p. 36. The HSRA has been found to be a good approximation for the solar atmosphere outside spots.

Secondly, from global spot models and observations the so-called "Wilson depression" of spots is roughly known. The Wilson depression describes the effect that due to the reduced density inside the magnetic field the same level of absorption is reached at larger geometrical depths for a light traveling towards the solar surface. The effect can be derived from analytical calculations requiring horizontal temperature equilibrium and horizontal pressure balance between regions inside (int) and outside (ext) magnetic fields:

$$p_{ext}(\rho_{ext}, T_{ext}) = p_{int}(\rho_{int}, T_{int}) + \frac{B^2}{2\mu_0}, \tag{5.7}$$

where B≠0 leads to a reduction of ρ_{int}, if $T_{int} = T_{ext}$. In a sunspot, no lateral temperature equilibrium can be assumed throughout the whole spot, which complicates things; but there still will be a reduction of density in umbra and penumbra.

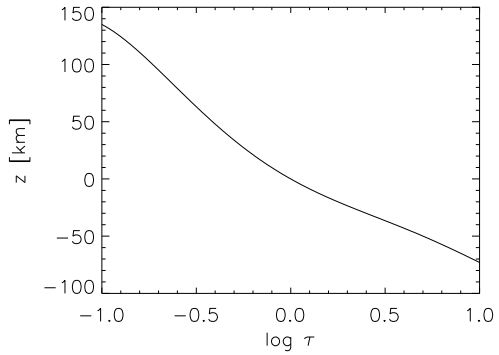


Figure 5.32: Conversion curve between optical depth, τ and geometrical height, z. For a reference atmosphere with known density and temperature, the amount of absorption can be calculated for any distance traveled in the atmosphere. In the opposite direction, the height corresponding to any given optical depth can be read off from the ordinate.

The Wilson effect can be imagined as shifting the log τ = 0 level downward in the atmosphere due to reduced density. However, the precise value of the Wilson depression and a possible dependence on properties of the spot like total flux or age is still under debate. Depending on the sources used (cf. Solanki 2003; Mathew et al. 2003) numbers between 100 km to 800 km are given for the penumbra. In this work, the Wilson depression is assumed to increase linearly throughout the penumbra and to reach about 400 km at the umbral-penumbral boundary. Taking the penumbral radius into account, this results in a slope of the log τ = 0 level of 3° throughout the penumbra (Schlichenmaier & Schmidt 2000).

For the conversion of the parameters from the Gaussian inversion, all numbers in optical depth are converted to geometrical height above/below the zero level with the small slope, using the conversion curve shown in Fig. 5.32. It displays the conversion between geometrical height and optical depth for the HSRA model atmosphere. For each optical depth value the geometrical height can be read off from the curve. As indicated from the existence of the maximum in the location value discussed in the previous section, for $r < 8$ Mm the Gaussian results are not reliable. However, with the integration of the line-of-sight inclination discussed in Sect. 5.9 a suitable substitute has been found.

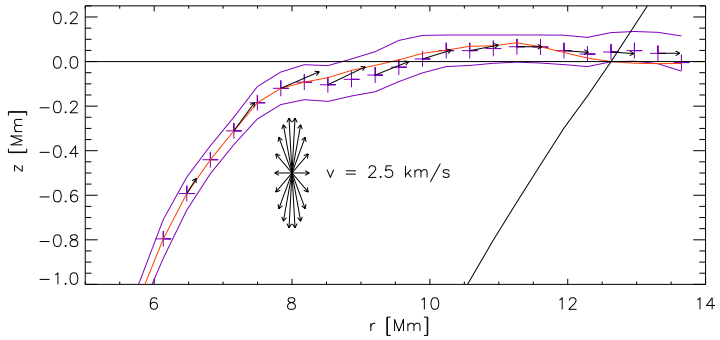


Figure 5.33: The average flow channel from the Gaussian inversion in the geometrical height scale. The inclined black line gives the integrated inclination of the background field. The arrows give the flow direction in the flux tube component, the length is proportional to the velocity. The purple crosses mark the location of the center of the flux tube.

5.8.2 The final flux tube model from the uncombed "Gaussian" inversion

The integration of the field inclination to the surface as another method to derive the vertical structure of the solar atmosphere will be described in the next section. However, the results of this method will be already used in the next two figures to complement the Gaussian inversion results. The location of the flow channel for $r < 8$ Mm and the structure of the background field are derived from the 2-component inversion. Otherwise the construction of the 3-D geometry of the average flow channel properties is straightforward: the location of the Gaussian perturbation gives the central height of the flow channel; the width of the Gaussian yields the geometrical extension.

The calculation leads to the picture for the 3-D geometry displayed in Fig. 5.33: the average flow channel is a flux tube of around 200 km diameter that lies almost horizontally along the solar surface in the outer penumbra. The tube starts to sink below the surface around $r = 8$ Mm; its inclination steepens very fast for smaller r. Inside the outer penumbra, the line-of-sight flow speed in the tube is almost constant around 4-5 kms^{-1}.

Fig. 5.34 displays the values of field strength, filling fraction, velocity and temperature for both the ft component and the background field in the context of the derived geometry of the ft component. The figure gives a nice visualization of the properties of the ft component: the ft component shows weaker fields and higher flow velocities in an almost horizontal flow channel. All quantities of the ft component show a marked jump in their values at the radius, where the flow channel reaches the surface.

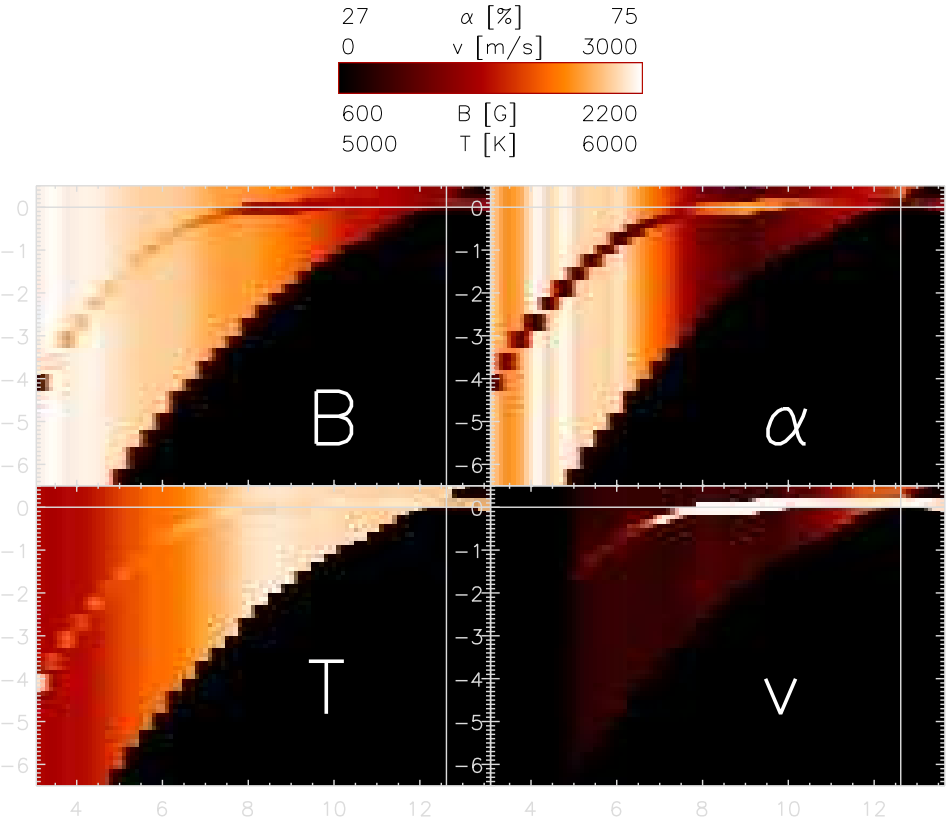


Figure 5.34: **Clockwise, starting left top**: field strength, filling fraction, horizontal velocity, and temperature of the background field and the flux tube component. The height scale along the vertical dimension and the radius along the horizontal dimension are given in Mm. For the background field, the parameter value is assumed to be constant in height with only a radial variation. The ft component is placed at the geometrical height derived, its diameter is increased by a factor of 1.6 for better visibility. The color bar on top gives the displayed ranges.

5.9 Integration of the Local Reference Frame inclination – A powerful visualization tool or even more ?

As has been discussed in Sect. 5.7.3, the results on the location of the flow channel in the Gaussian inversion are not reliable for $r < 8$ Mm. Additionally, for the background field the inversion assumes constant field properties with optical depth. It thus has become necessary to use another method for the derivation of the 3-D topology. Opposite to the geometry derived from the Gaussian inversion, some more (critical) assumptions have to be made. The reader is cautioned that the approach chosen has some severe limitations, which will be discussed in Sect. 5.9.3 in more detail.

A possible way for the retrieval of the height dependent properties of the magnetic field, its three-dimensional configuration, from inversion results *without* any intrinsic height variation is given through the integration of the inclination to the solar surface in radial direction. For each radial position r_i the azimuthally averaged inclination, $\gamma(r_i)$, of background field and flux tube component are known (cf. Fig. 5.20, p. 91). With the assumption that the inclination does not strongly change with depth below the surface (or height above it), the physical height, h, of a field line in the solar atmosphere can be calculated from an integration of the inclination in the Local Reference Frame (LRF) by

$$h(r_i) = \sum_{j=0\ldots i} \frac{1}{\tan\gamma(r_j)} \cdot \Delta r \,, \tag{5.8}$$

where $\Delta r = 340$ km is the difference of the respective semimajor axes of two subsequent ellipses (cf. Fig. 5.19).

Eq. (5.8) can be applied not only to the average values of inclination with radius, but also to the inclination values along radial cuts. This allows to take the spatial variation of the inclination inside the spot into account. With the spatially resolved inclination values, the geometry of the spot can be re-constructed from the inversion results.

The derived height, $h(r)$, has to be adjusted to have a common reference point. For the ft component no inclination values are available in the umbra, but at the outer penumbral boundary both components are observed. All geometrical height curves of the integrated LRF inclination are thus fixed to have the same value at the outer spot boundary. The reference value is taken from the curve of the integrated bg field inclination in the 2-component inversion.

5.9.1 Azimuthally averaged geometry

Fig. 5.35 displays the resulting height curves of the integrated LRF inclination of bg field and ft component in the 2-component inversion. For comparison, the boundary layers between penumbra and umbra, respectively, penumbra and the field-free surroundings of the magneto-static sunspot model of Jahn & Schmidt (1994) have been overplotted. In their spot model in hydrostatic equilibrium, the so-called *peripatopause* (umbra-penumbra boundary) and the *magnetopause* (outer spot boundary) are the current sheets that separate the penumbra from the umbra and the quiet Sun outside the spot, respectively. The current sheets in the plot have been shifted to the reference value of z = 25 Mm at the outer spot boundary like all other curves. The magnetopause has been moved radially closer to the umbra by 0.8 Mm to agree with the observed spot radius of 12.6 Mm. The ratio of penumbral to umbral radius was taken to be 2 in the spot model, while for the observed spot it is close to 2.5. Moving the magnetopause by 5 % radially without changing its slope should still be close to a stable solution.

Two results are eminent from a look at the integrated LRF inclination of the background field: firstly, the integrated LRF curve is in close agreement with the magnetopause of the theoretical spot model in the whole penumbra from $r \sim 5$ Mm to 12.6 Mm. In the umbra and at the umbra-penumbra boundary, the integrated LRF curve differs significantly from the magnetopause; it is less steep. This deviation still is a good sign: for an expanding flux tube the inclination at the surface should be larger than in the deeper layers. The integration of the surface inclination – where the observations were made – thus *should* result in a less deep structure. The effect should get stronger the deeper the structure actually is. It is on the other hand rather astonishing that the magnetopause and the integrated LRF curve are in close agreement even at depths of up to 5 Mm – keeping in mind that the field inclination has been measured 5 Mm higher up in the solar atmosphere !

Secondly, the integration at once yields a "thick" penumbra with a depth of some Mm, not only a shallow surface layer. At the outer spot boundary, the field lines of the background field continue upward.

For the flux tube component, the integration gives a different curve. The integration leads to an arched loop, which starts at the umbra-penumbra boundary, rises above the reference height in the mid penumbra at around $r =$

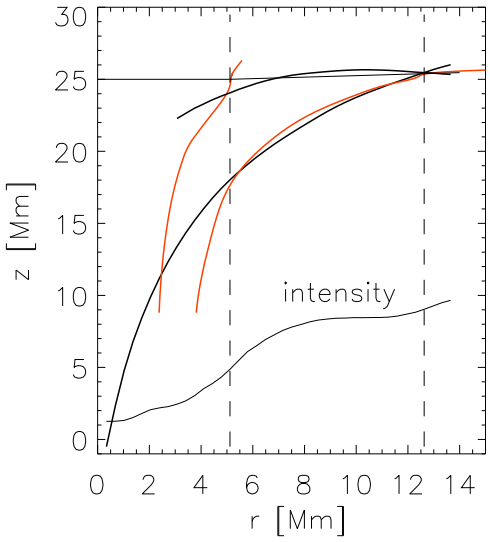


Figure 5.35: Integrated LRF inclination of bg field (**lower thick black line**) and ft component (**upper thick black line**). **Red**: The transition layer between umbra and penumbra, respectively, penumbra and the field-free surroundings of the spot model of Jahn & Schmidt (1994). The **thin black line** in the lower part of the figure gives the radial variation of continuum intensity. The **vertical dashed lines** mark the inner and outer boundary of the penumbra. The **horizontal thin black line** at z = 25 Mm gives the log $\tau = 0$ level.

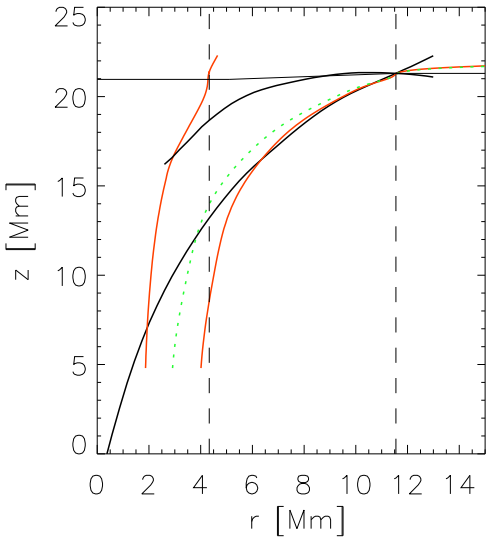


Figure 5.36: Same as Fig. 5.35 for the 2nd spot. The green dotted line is the transition layer between penumbra and field-free surroundings of the spot model of Jahn & Schmidt (1994), when it is only shifted to fit to the smaller spot radius without changing its slope. The red line is the transition layer, when it is compressed in radius.

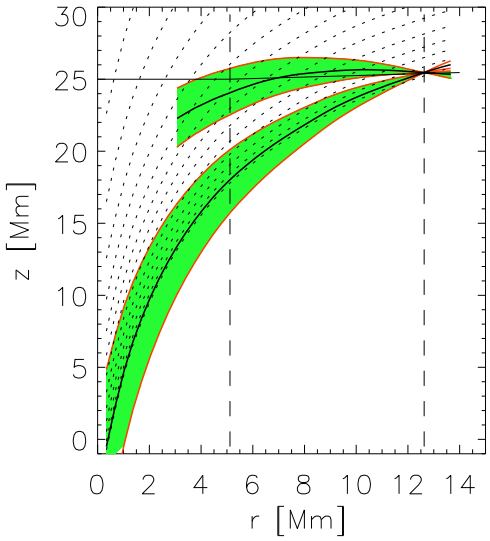


Figure 5.37: Same as Fig. 5.35, integrated LRF inclination of bg and ft component. The green area enclosed by the red lines is the result of integrating $\gamma\pm$ 20 % for the bg component ($\pm$ 15 % for the ft component). **Dashed**: field lines that have the same inclination at z = 25 Mm as that observed for the background field.

7 Mm, returns to the surface at the outer boundary (reference point !) and continues downward outside the spot. Fig. 5.36 display the resulting curves, if the approach is applied to the 2-C inversion results of the second spot analyzed in this work. The general picture is identical: the bg field is in agreement with the magnetopause of the spot model of Jahn & Schmidt (1994), if the different spot radius is taken into account. The ft component gives an arched loop, slightly elevated above the reference height.

Fig. 5.37 displays how the resulting curves of geometrical height are affected by small changes in the inclination, keeping in mind that the field inclination at the surface is expected to be be larger than in deeper layers. The inclination value is in-/decreased by 20 % for the background field (15 % for the ft component) and integrated again. The effect of the changed inclination is found to be small: the discussion given for Fig. 5.35 is still valid.

Additionally, in Fig. 5.37 a band of field lines has been overplotted as dashed lines. They have been constructed by shifting – only vertically – the curve of the integrated LRF inclination of the bg field in such a way that their inclination around z = 25 Mm is identical to the bg inclination actually observed. They thus correspond to a possible model of the spot structure that would result in exactly the observed inclination values. They give a reasonable picture how the field lines of the sunspot will be spreading with height, i.e. the expansion of the radius of the flux tube due to the drop of the outside gas pressure.

5.9.2 Close-up: penumbra

The structure of the integrated LRF inclination of the flux tube component in the mid to outer penumbra is displayed magnified in Fig. 5.38. Again, the band of field lines that would have exactly the inclination value observed on the log $\tau = 0$ level are overplotted, this time for the ft component.

The integrated LRF inclination of the flux tube component forms a loop with a maximum elevation of around 400 km above the log $\tau = 0$ level. Most interesting are the dotted lines below the loop. They represent all field lines, whose return to the solar surface is actually observed. The field lines above the loop not necessarily have to do this, they most probably continue upward into the canopy. The return of these lines to the surface would be far outside the spot. The integrated LRF inclination is in this context best interpreted as being the envelope of returning field lines. The intersection point of the arched loop with the inclined log $\tau = 0$ level is at $r = 7$ Mm. Interestingly, the intersection of the loop with the reference height value, which corresponds to the location of the solar surface outside of the spot, is around $r = 8$ Mm, similar to the position, where the flux tube in the Gaussian inversion made contact with the surface.

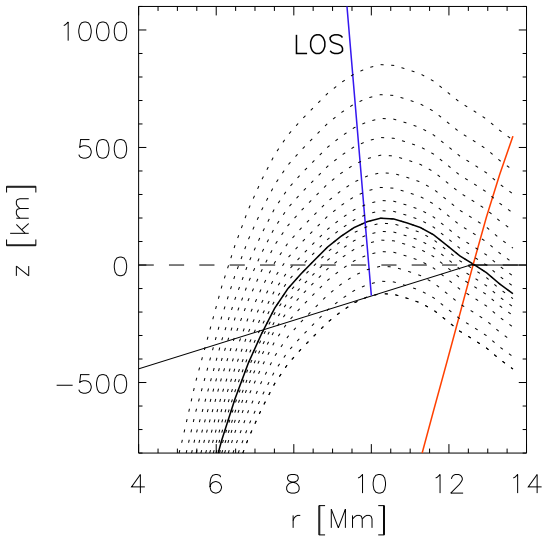


Figure 5.38: Zoom of the outer penumbra, height now in km. The dashed horizontal line marks the zero level of the height scale, the reference value at the outer penumbral boundary. **Thick black**: integrated LRF inclination of the flux tube component. **Thin black line**: the log $\tau = 0$ level with a 3° slope (cf. Sect. 5.8.1). **Red line**: integrated LRF inclination of the background. The dotted lines are field lines that would have the inclination value observed for the ft at the log $\tau = 0$ line. **Blue**: line of sight on the limb side with 27° inclination to the surface.

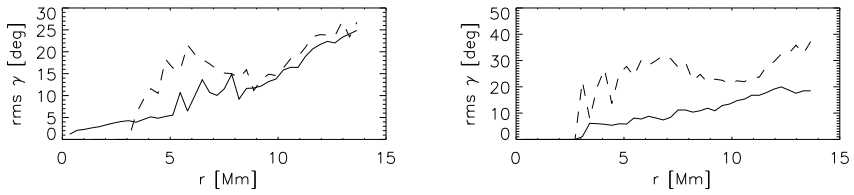


Figure 5.39: The intrinsic scatter of the inclination values along the azimuthal paths, given for each radial position by the root-mean-squared (rms) value. **Left**: 2-component inversion. **Right**: Gaussian inversion. The flux tube inclination (**dashed**) has a rms value around 15° in the middle penumbra, where the ft component can best be traced. The bg component (dashed) has less than 5° rms in the umbra, but the rms is steadily increasing with radius.

5.9.3 Validity of the integrated LRF inclination

The main problem of using the integration of the LRF inclination to derive the 3-D geometry is the extrapolation involved. The inclination values derived from the observed spectra are only valid in the formation height of the spectral lines. These formation heights are close to the solar surface, given by the log $\tau = 0$ level. If z = 0 km is placed at the log $\tau = 0$ level, the spectral lines have contributions from -50 km to less than 400 km. If the integration yields a height value outside of this range, any property attributed to that location *was not* directly observed there.

The question of the validity can also be put the other way around: if the magnetic field inclination on a spatial location is constant with height/depth in the solar atmosphere, the integrated LRF inclination yields the exact geometry. This was assumed for the dotted lines in Figs. 5.37 and 5.38: the field lines at any height and radial position are given by shifting the integrated curve up or down, until it intersects with the log $\tau = 0$ level at the specified radial position. Thus, to address the question of validity of the integrated LRF inclination, the depth/height dependence of the field inclination has to be examined.

A first item to be examined is the variation of the inclination in the horizontal dimension. As most structures in a penumbra are radially aligned, the variation of the inclination value measured on the azimuthal paths reflects the small-scale fine structure of the sunspot. The amount of variability gives a hint, what level of small scale variation to expect in the third dimension, the height direction.

Intrinsic variation on azimuthal paths The variance, σ^2, of a discrete parameter, x_i, is commonly defined by

$$\sigma^2 = \frac{1}{N-1} \cdot \sum_{i=0\ldots N} (x_i - \overline{x})^2 \ , \tag{5.9}$$

where $\overline{x} = 1/N \cdot \sum_{i=0\ldots N} x_i$ is the mean value of the parameter. σ^2 measures the squared deviation between individual values and the mean, i.e. the scatter around the average value. The square-root of σ^2 is termed the root-mean-squared (rms) value[2].

Fig. 5.39 displays the scatter of the inclination values on each azimuthal path with given radius. The rms-value for the bg component is below 15° throughout the spot in both the 2-component and the Gaussian inversion. For the ft component the scatter is considerably larger and the effect of inclination changes will be more drastic there: the maximum loop height strongly depends on the inclination values, especially on the values at the outer penumbral boundary. Here one has to keep in mind that the result of integrating the azimuthally averaged inclination may well be different from individual radial cuts or a single isolated filament picked out by hand. On the other hand, the difference in rms value is also a sign of consistency with the image of dynamic flux tubes (larger rms) embedded in a static background.

Theoretical considerations on height/depth dependence The band of field lines in Fig. 5.37 or 5.38 were constructed by shifting the integrated LRF inclination curve up or down in the vertical dimension. This corresponds to the assumption that the inclination does not change with depth or height in the atmosphere. However, there are

[2]For a parameter whose true value is unknown $\overline{x} \pm 1/\sqrt{N} \cdot \sigma$ defines the range in which the result with 68 % probability will lie on the base of the available data.

theoretical arguments against a constant continuation of field lines with height. The density of the solar atmosphere and the temperature drop rapidly with increasing geometrical height. Thus, the gas pressure outside the magnetic field has to decrease as well, as soon as hydrogen atoms start to recombinate with the free electrons, whereas the corresponding magnetic pressure term is primarily coupled to the field strength, $p_{\mathrm{mag}} \propto B^2$. To be in pressure equilibrium, the magnetic flux – which is a conserved quantity due to $\nabla \cdot \mathbf{B} = 0$ – has to spread over a larger area, because B measures the density of field lines per area. This leads to a rapid expansion of the area of a flux concentration with height, until the field is almost horizontal. Thus, a field inclination measured at the solar surface will be larger than the actual inclination values below the surface and smaller than the actual values above it.

Conclusion The scatter of the inclination values in the horizontal dimension is about 10° to 15°. The theoretical considerations predict an additional systematical trend with height. However, the general picture of the field geometry from the integrated LRF inclination does not change, if the inclination in the height range concerned changes by 15° (cf. Fig. 5.37). The bg field inclination will yield a thick penumbra of some Mm and tend to rather underestimate the true inclination and depth. From the close agreement between the integrated LRF inclination of the bg field and the magnetopause of the theoretical spot model of Jahn & Schmidt (1994), it seems that the variation of inclination with depth can be neglected from the outer to mid penumbra. For the ft component, the loop elevation would be reduced, if the field inclination turns to horizontal fields above the surface.

To conclude, the integrated LRF inclination can be assumed to give a reliable picture of the three-dimensional topology as long as "small" depths (< 4 Mm) and even smaller heights above the photosphere are involved. It gives a powerful tool for visualization purposes, as the next sections will show. And with some care, it may also help to pinpoint the processes governing the dynamical evolution inside the penumbra, because it allows to derive at least a first guess of the spatially resolved subsurface structure of a sunspot.

5.9.4 Integration in azimuthal bins

The parameters used in the previous sections were azimuthally averaged over the elliptical paths. Thus, the results reflect the statistical mean over an unknown number of flow channels, the idealized version of the flux tube that is assumed to be present all around the spot. The spatial fine structure of the sunspot penumbra gets lost in the process. However, the approach of the derivation of the 3-d topology can be also applied to radial cuts or values averaged in small azimuthal bins. This allows to investigate the spatial variation in the properties of background field and flux tube component and the possible relationships between them.

5.9.4.1 Parameter values in azimuthal bins

Instead of averaging azimuthally over the ellipses, all physical quantities given by the inversion were binned in 4° bins, azimuthally around the spot center. The size of the bins is limited by the available unique spatial points, especially for small distances from the spot center. Fig. 5.40 displays the resulting values for the LRF inclination and field strength for the bg field and the ft component in the 2-component inversion. The graph is already interesting by itself without further processing.

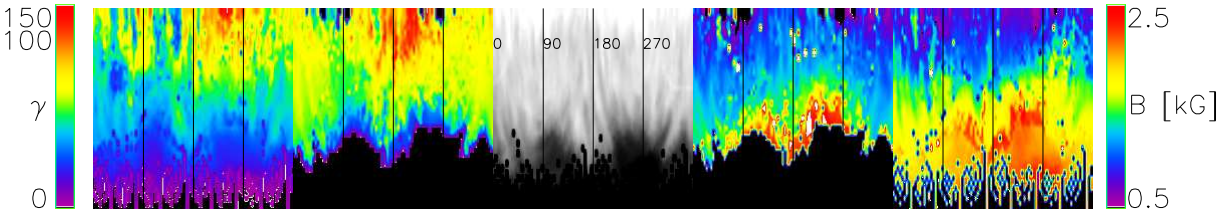


Figure 5.40: **Left to right**: LRF inclination of the bg field (0..100°), same for ft component (0..150°), intensity, field strength of the ft component, same for the bg field. In the intensity map the four vertical lines marking the azimuthal angle around the spot are labeled. 0° corresponds to the center side, 180° to the limb side direction.

The filamentary structure of the sunspot penumbra is clearly visible in the intensity map. The outer boundary of the spot is more or less at the same distance from the center of the spot at the lower border of the image, but the transition between umbra and penumbra shows a much larger variation in its location. Especially on the limb

side at 180°, a penumbral intrusion into the umbra is present. Interestingly, the maximum inclination of the ft component (2nd image from the left) can be found at the outer spot boundary at 180°. Also the opposite case can be seen: at places where the umbra-penumbra boundary is further away from the spot center (near 90°), the LRF inclination of the ft component is reduced. This implies that the structures, which develop in the penumbra, are influenced by the shape and properties in the umbra.

A prominent feature in the field strength of the background field can be seen around 0° to 20° (last image on the right). The field strength shows subsequently stripes of enhanced and reduced field strength. These so-called "spines" have already been found in earlier observations (Westendorp Plaza et al. 2001b).

5.9.4.2 Projection effects due to the inclined line of sight

If one uses the spatially resolved inclination values without azimuthal averaging, one has to keep in mind a limitation, which is due to the off-center location of the sunspot. The heliocentric angle of the spot was 27.8°; hence, the line of sight makes the same angle with the solar surface. The limb side and the center side of the spot are not equivalent, because the structures examined have a variable inclination to the surface, which may be smaller or larger than 27.8°. Fig. 5.41 displays the effect using the curves of the integrated LRF inclination. On the limb side, all structures can be traced everywhere in the umbra and penumbra. However, on the center side a limiting angle exists: if the inclination of a coherent structure drops below 27.8°, the continuation of the structure is hidden from the LOS. In the graph, the LOS lines for the center side have been set to intercept the integrated LRF curves at the places, where the field inclination is $\sim 28°$. For the bg field this is somewhere inside the umbra, but the integrated curve continues. For the flux tube component it is almost exactly at the first radial position, where the ft component appears at all. The azimuthally averaged inclination of the ft component – which incorporates the values on both the limb and the center side – is about 26° at the first non-zero value. This suggests that on the center side the inner footpoints of the flux tubes are hidden by their own continuation as arched loops. The larger the elevation of a loop is, the stronger is this shading effect.

Another effect, which has not been included in the graph, may happen at the outer penumbral boundary on the center side: the Wilson depression inside the penumbra is always treated as a linear increasing function with 0 km at the outer spot boundary and about 400 km at the umbra-penumbra transition (cf. Sect. 5.8.1). However, if the Wilson depression is more like a step function, at the outer penumbral boundary on the center side another shading effect appears. Some 200-300 km of the outermost penumbra cannot be seen on the center side, because they are "below" the surface, i.e. the log $\tau = 0$ surface of the quiet Sun blocks the line of sight.

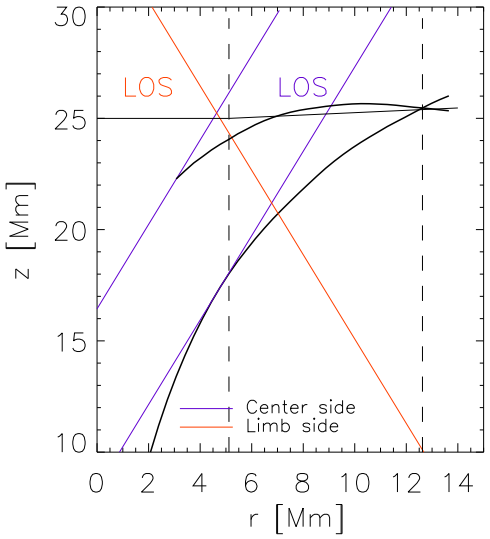


Figure 5.41: The effects of the inclined line of sight on the observations on limb or center side of the spot. **Thick black**: Integrated LRF inclination of the 2-component inversion as in Fig. 5.35, flux tube component (upper line) and background field (lower line). The LOS with a 27.8° inclination to the surface is overplotted in purple for the center side and in red for the limb side. On the center side a limiting angle exists, where a structure cannot be traced any longer: if the inclination to the vertical changes to below 27.8° the structure will be hidden from the LOS. The two LOS lines on the center side cut the bg and ft curves at where their inclination is $\sim 28°$.

Note that the line of sight effects affect already the observed profiles, independently of any evaluation method employed later on. If a significant difference between the limb side and the center side of the sunspot remains in the final results, the evaluation method does not necessarily have to be the culprit. For example, the hot inner

footpoints of the flux tubes are not seen on the center side. No method, which uses the observed profiles for the derivation of the temperature, will be able to retrieve them. However, one can be quite sure that in fact they *will have been there* in the solar atmosphere.

5.9.5 The final spot model

To derive the spatially resolved geometry, the LRF inclination of background field and flux tube component in the azimuthal bins is integrated in radial direction using Eq. (5.8). The reference point of 19.07 Mm for the resulting geometrical height scale is the average height value of all bg field bins at the outer penumbral boundary. As intermediate step to the final radially aligned spot model, the results for the azimuthal bins are presented in a linear alignment. Figs. 5.42 and 5.43 display the value of the field strength, $|B|$, and the line-of-sight velocity, v, covering the resulting 3-D surface of the integrated bg field inclination. The ft component is overplotted as lines at its derived height, where the line color gives the parameter value.

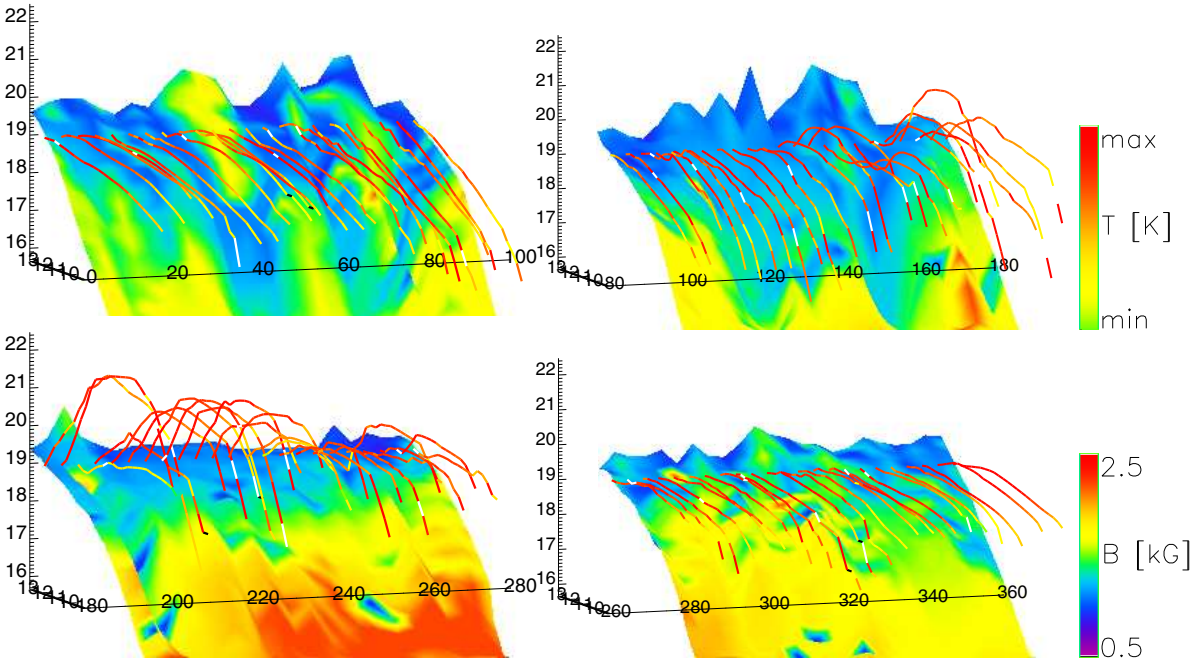


Figure 5.42: Resulting 3-D topology after integration of the inclination maps of background field and flux tube component in Fig. 5.40 in each 4°-bin. The z-axis and the y-axis going into the paper plane are in Mm. The x-axis going left to right gives the azimuthal angle around the spot center. 0°/180° is the center/limb side. Upper row displays the azimuth 0°...180°, bottom row 180°...360°. The color code on the surface of the bg field gives its field strength. The ft component is placed at the value of its height after the integration, the color code gives the temperature of the ft component in the corresponding bin (scaled individually).

Fig. 5.42 can be used to visualize the resulting 3-D topology of the background field and the flux tube component. The color of the bg surface displays its field strength, corresponding to the 2-D map of field strength in Fig. 5.40. The ft component is overplotted as lines at the geometrical height, which the integration yields for the ft in each bin; the color code gives the temperature of the ft component in the corresponding bin. The ft components' temperature is maximum in most cases at the inner penumbral boundary where the flux tubes start.

The "spines" of enhanced field strength around x=25° and 55° (yellow color) or also at 170° (red color) are seen to be elevated with respect to their surroundings. They are more horizontal in the outer penumbra, but have a similar inclination as their surroundings in the umbra.

The deep gap at x=190° coincides with a dark filament in the intensity map in the middle of Fig. 5.40. The dark filament is seen to start in the umbra, and continues through the complete penumbra. The ft component at the gap is lying very "low", opposite to the high loop and the more elevated background field next to it on either side. This

indicates an influence of bg field structure on the ft behavior, and further of ft behavior on average temperature and intensity.

Another prominent feature is the markedly difference between limb and center side penumbra. This is also already visible in the inclination maps in Fig. 5.40, but gets even more clear after the integration: the flux tube loops are much more elevated on the limb side of the spot (180° to 260°). Even taking into account that the spot is foreshortened due to the off-center position on the sun, the sunspot unfortunately is also not very regular. The penumbral intrusion into the umbra at x = 180° makes it difficult to decide, if the differences between limb and center side are due to the spot structure, or due to projection effects because of the inclined line of sight discussed in the previous section. The reduced elevation of the ft component on the center side may partly be due to the fact that the outermost penumbra, where the flow channels dive back to the surface, cannot be seen due to the shading effect of the Wilson depression. Thus, on the center side the highest inclination values ($\gg 90°$) are missing which results in less elevated loops after the integration.

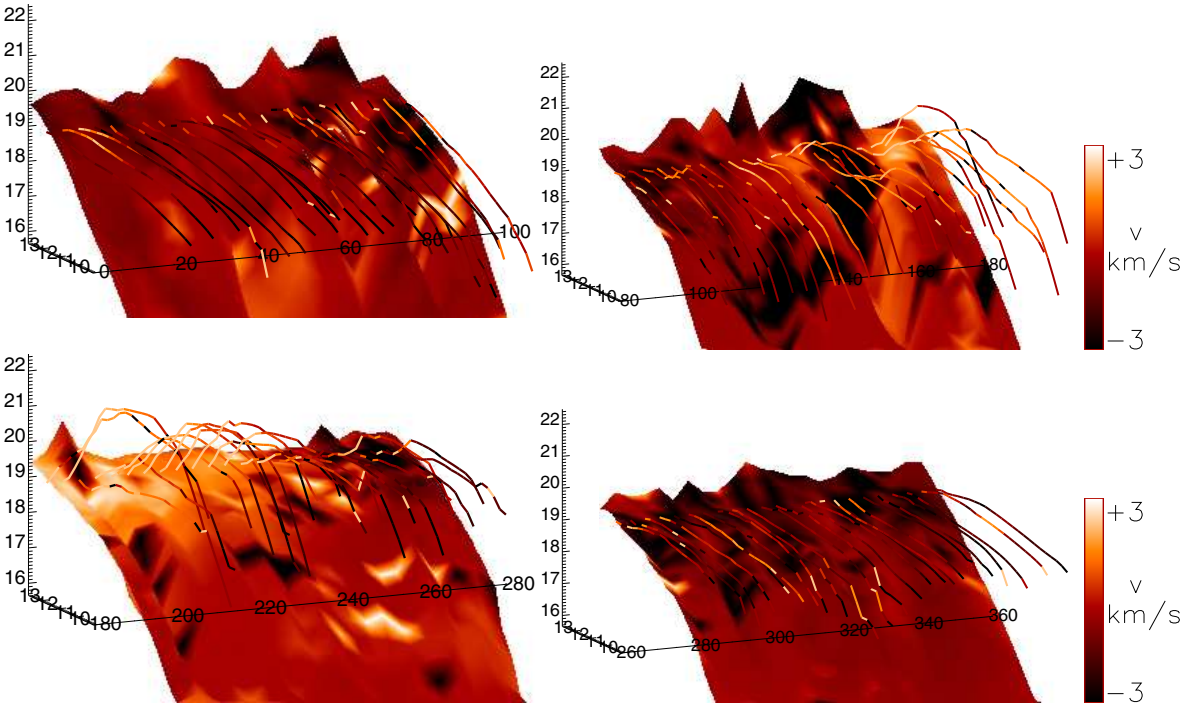


Figure 5.43: Same as Fig. 5.42, but now the both the surface and the flux tube component are color coded according to their line-of-sight velocity.

Fig. 5.43 displays the same topology, but the color code of the bg field surface and of the flux tube component now gives the line-of-sight velocities. On the limb side of the spot around x = 180°, it can clearly be seen that the LOS velocities of the bg field and the ft component strongly increase in the outer penumbra. If one follows one of the elevated loops around 180° to 220° from the umbra outwards towards the outer spot boundary, one sees the LOS velocity increase as soon as the maximum elevation is reached. However, this is not the signature of flows that speed up downward from the maximum height, like material in free fall along the field lines; it rather reflects the fact that the field lines of the ft component are going back to the surface at this radius. The flow velocity was found to be almost constant with radius in the azimuthal averages (cf. Figs. 5.22 and 5.29). As the flow should be along the field lines in the ionized plasma of the solar atmosphere, the projection of the flow velocity onto the LOS has to increase, if the field lines bend downwards. The increase of the LOS velocity at the point of maximum loop elevation is an indirect proof that the field lines are diving down towards the surface from this radius outwards – as the LRF inclination value also claims.

Another interesting feature on the limb side of the spot are the LOS velocities on the inner penumbral boundary. At the inner footpoint of the loops, the velocity is negative (black color), i.e. the flow is directed towards the observer; for comparison, on the center side around 40° the flux tube are colored black along their full length. For a better visualization, Fig. 5.44 displays the LOS velocity map of the flux tube component in the azimuthal 4°-bins.

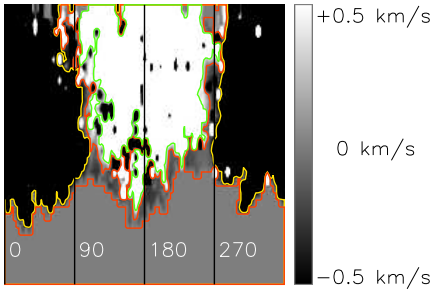


Figure 5.44: LOS Velocity of the ft component. Contour lines give the following velocity levels: -500 m/s in yellow, 0 m/s in red, +500 m/s in green. Positive velocities are along the line of sight, i.e. downflows in the solar atmosphere. In the map white corresponds to +3 kms^{-1}, black to -3 kms^{-1}.

The areas of positive and negative LOS velocities are traced by contour lines, where the red contour line is most interesting: it marks the zero velocity level. The green contour outlines +500 m/s, i.e. material motion along the LOS away from the observer.

On the limb side around x = 180°, it can clearly be seen that at the inner penumbral boundary a small area with negative velocities is present. Going upward from the spot center at the lower border of the image at exactly x = 180°, the lower red contour marks the transition to the umbra where the ft component has zero velocity. The color then turns to a slightly darker shade than in the umbra, which means negative LOS velocities, i.e. motion towards the observer. The next red contour marks again the zero level. Continuing outward the velocity is always positive and increases in amplitude. The observation has a natural interpretation in terms of localized upflows at the innermost umbra-penumbra boundary, which is a main prerequisite to maintain the radial Evershed outflow in the penumbra: the material motion observed in the outer penumbra has to have a source of material somewhere.

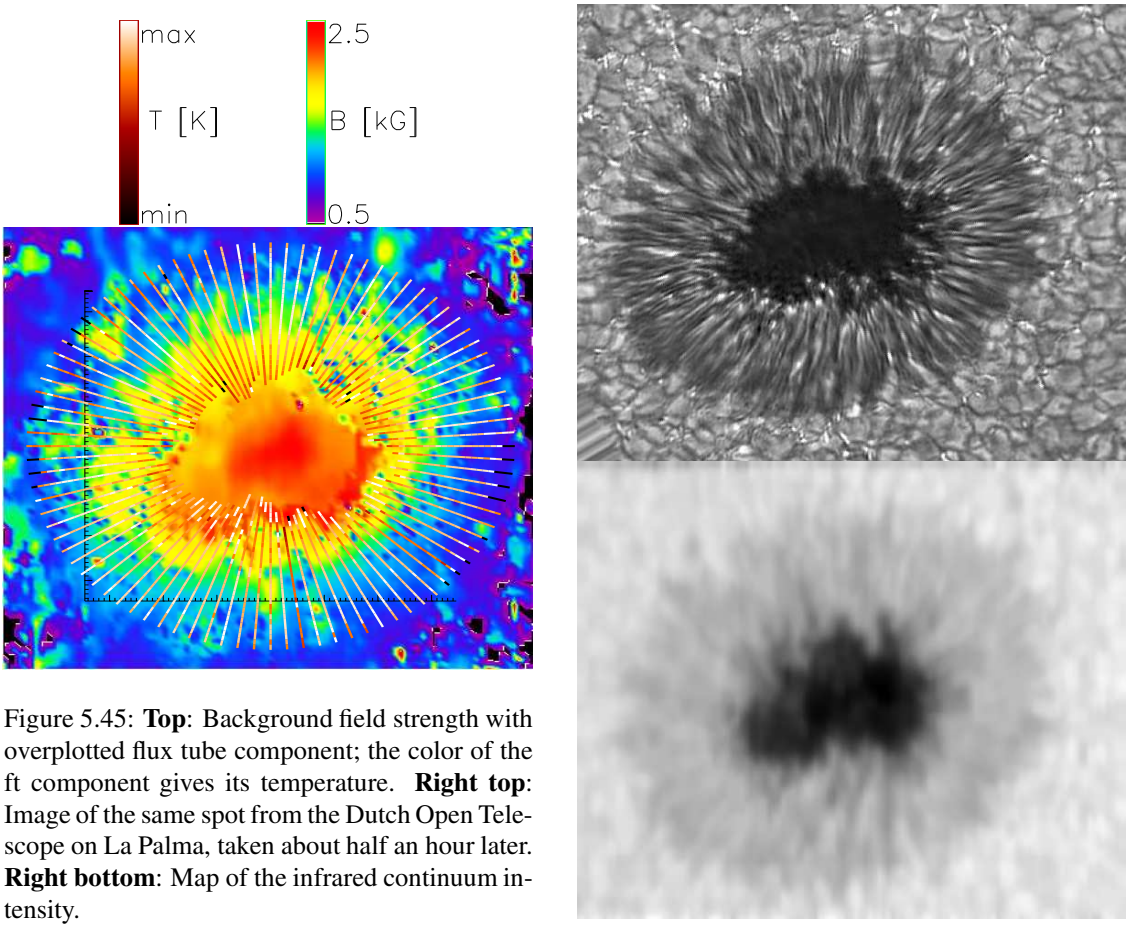


Figure 5.45: **Top**: Background field strength with overplotted flux tube component; the color of the ft component gives its temperature. **Right top**: Image of the same spot from the Dutch Open Telescope on La Palma, taken about half an hour later. **Right bottom**: Map of the infrared continuum intensity.

The final spot model is constructed from the azimuthal bins by assuming a perfect round spot. The binned values can be interpolated to yield a smooth surface like in the previous graphs, but now they are arranged azimuthally around a center, not linearly next to each other. The first figure of the full spot model in Fig. 5.45 presents a top view on the model. The background fields is color coded with its field strength, and the flux tube component is color coded according to its temperature as in Fig. 5.42. For comparison, two intensity images of the same sunspot are also displayed: the map of the infrared continuum intensity of the TIP data set evaluated and an image of the same sunspot taken with the Dutch Open Telescope (DOT) on La Palma about half an hour later. The DOT image is a speckle-reconstructed filtergram in the G-band and has been aligned only roughly to the other map. A very peculiar agreement between the DOT image and the temperature of the ft component can be seen: On the limb side of the spot (lower part) at the umbra-penumbra transition, one can see a number of the so-called *penumbral grains*, i.e. local brightenings at the beginning of penumbral filaments (cf. Sect. 1.2.2.2). On the particular DOT image many penumbral filament on the limb side start from such a grain, while on the center side not a single penumbral grain is visible. In the IR intensity map the effect cannot be seen due to the much worse spatial resolution. However, the ft temperature in the top view of the spot model shows exactly the same structure as the DOT image: the temperature maxima of the fts are located at the umbral-penumbral boundary, and they are most prominent on the limb side of the spot. On the center side – as for the DOT image – no temperature enhancement can be seen. No trace of this behavior is visible in the IR intensity map; the information that the temperature of the ft component is significantly larger at the inner footpoints comes from the polarization signal, where the hot upstreams of the footpoints have left their otherwise invisible signature.

Fig. 5.46 displays the spot model from four different viewing angles: the top view, from the side, from below and from 40° above the solar surface. The general impression from the azimuthally averaged values in Sects. 5.9 and the linearly arranged ones in Sect. 5.9.5 is confirmed by the individual tubes in the 3-D display: the ft component forms more or less elevated loops that bend downwards in the outer penumbra. The elevation is different on limb and center side, maybe due to the projection effects on the line of sight. The subsurface structure of the bg field is seen to be ragged; it consists of stripes of more and less inclined field lines. A detailed comparison with the observations would be possible by calculating the spectra, which would originate from the geometrical arrangement, but this is beyond the scope of this work.

Fig. 5.47 displays the 3-D model for the second spot analyzed in this work at a heliocentric angle of 8°. It shows a more uniform structure of the ft component all around the spot with little difference between center and limb side, presumably due to the less inclined line of sight. The elevation of the loops is strongly reduced. The hot upstreams are still only seen on the limb side of the spot. The more regular shape of this spot indicates the sensitivity of the integrated LRF inclination to line-of-sight effects, which are missing close to disc center.

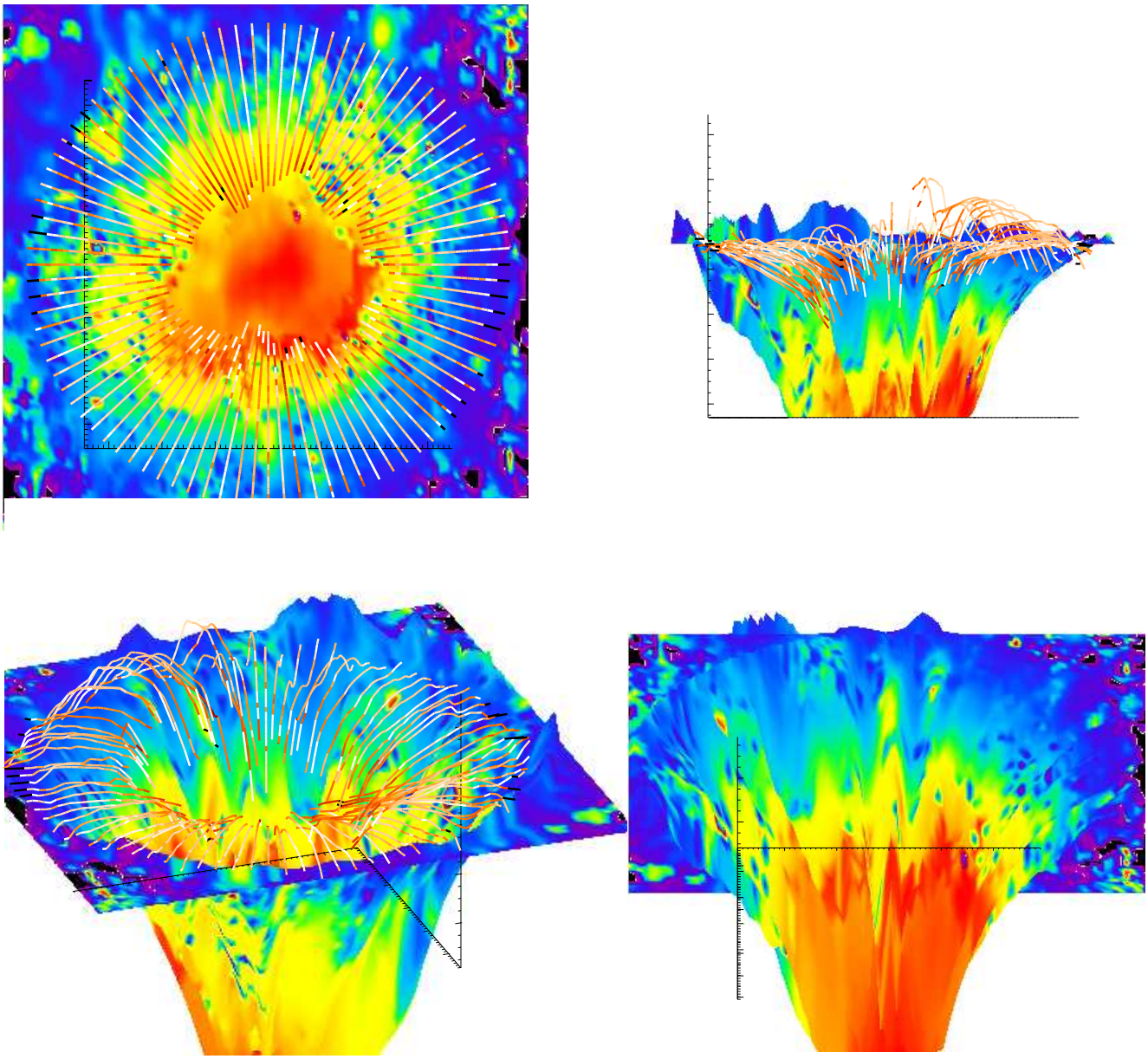


Figure 5.46: The full spot 3-D model in **clockwise, starting left top**: Top view (same as Fig. 5.45, for completeness), side view in the surface plane, from below (ft component not plotted), from above at 40°. The color coding is the same as in Fig. 5.45.

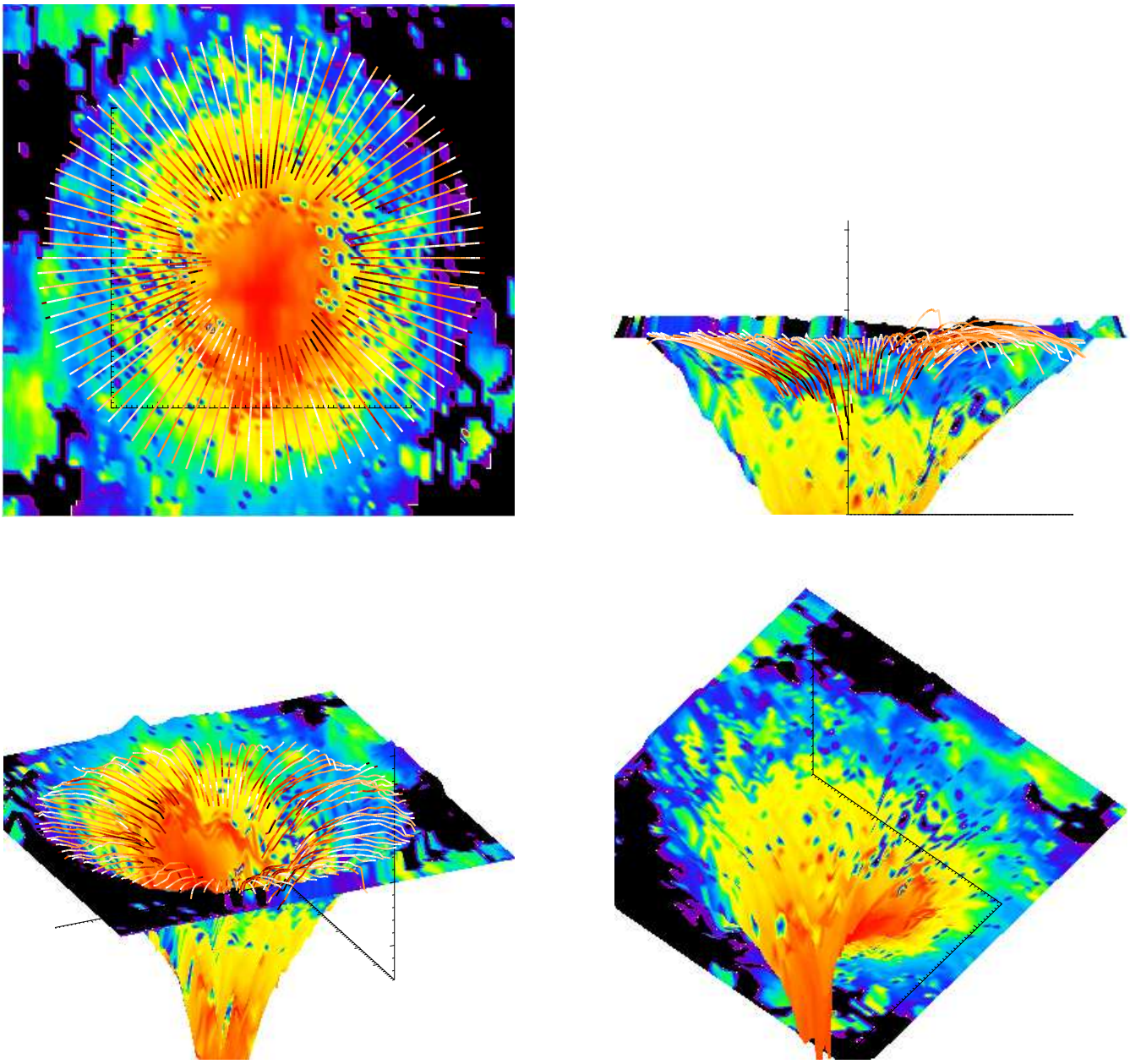


Figure 5.47: Same as Fig. 5.46 for the second spot analyzed in this work. **Clockwise, starting left top**: Top view, side view in the surface plane, from below (ft component not plotted), from above at 40°. The heliocentric angle of the observation was $\theta = 8°$ in this case. The color coding is the same as in Fig. 5.45.

5.10 Final comparison: Integrated LRF inclination vs. location of Gaussian perturbation

For the Gaussian inversion, two methods are available to derive the 3-D topology: The integrated LRF inclination of the flux tube component can be calculated in the same way as for the 2-component inversion, and the location of the Gaussian perturbation in the log τ scale can be converted to geometrical height. Fig. 5.48 summarizes in some way the result of the Gaussian and also the 2-component inversion.

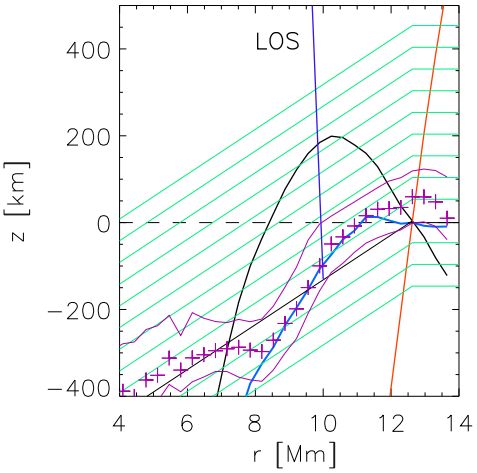


Figure 5.48: **Black**: Integrated LRF inclination of the ft component in the 2-component-inversion. **Blue**: Same for the ft component in the Gaussian inversion. **Red**: Integrated LRF of the bg field in the 2-component-inversion. **Purple crosses**: Location of Gaussian perturbation converted to geometrical height above the log τ = 0 level (**thin black**, 3° slope). **Purple solid lines**: Upper and lower boundary of the tube converted to geometrical height. **Green**: lines parallel to the log τ = 0 level, in 50 km steps. The LOS on the limb side is overplotted in the upper part. The integrated LRF of the Gaussian inversion is in very good agreement with the location of the tube from the Gaussian inversion.

The most prominent feature is that the location of the tube in geometrical height (purple crosses) from the location of the Gaussian perturbation is in very good agreement with the integrated LRF inclination of the ft component of the Gaussian inversion (blue line). The range in radius, in which the LRF inclination stays close to 90° (cf. Fig. 5.28, p. 101), is reflected in the Gaussian location, which is almost constant for 10 Mm $<$ r $<$ 12 Mm (cf. Fig. 5.31, p. 104). The localized maximum in the optical depth scale around 8 Mm leads to a minimum in geometrical height at this radius. The comparison with the curve of the integrated LRF inclination creates the impression that the inversion code has insufficient information to determine the location of the tube for smaller *r*. Both height curves for the Gaussian are well below the overplotted integrated LRF inclination of the 2-component inversion. This still agrees with the interpretation that the integrated LRF of the 2-component inversion is the envelope of all returning field lines.

To conclude, both approaches to derive the 3-D topology for the Gaussian inversion lead to a horizontal flux tube component with only a small elevation around 100 to 150 km above the log τ = 0 level. How these results fit together with earlier observations and theoretical or numerical models will be discussed separately in Sect. 5.13.

5.11 Comparison of the two spots observed

The second sunspot analyzed for this study was only inverted with the 2-component inversion setup. The general properties of most quantities were very similar. For the radial variation of magnetic field properties, also the 2nd spot showed the more horizontal inversion component to be weaker and to have higher flow velocities. The integration of the inclination to the surface yielded a very similar picture of horizontal flow channels that bend down in the outer penumbra (cf. Fig. 5.47, p. 118). The most significant difference with respect to the 3-D topology derived from the observations is the uniformity of the flow channel geometry around the spot. The strongly elevated loops and the pronounced center-limb side differences are absent. This clearly indicates that for the first spot at an heliocentric angle of around 30° already the observed profiles were systematically influenced by projection effect due to the inclined line of sight, which are not (or much less) present at an heliocentric angle of 8°. Especially for the integration of the LRF inclination, a systematic trend in the values will add up. To conclude, the list of properties in the following section is valid for both spots analyzed.

5.12 Summary of observational results

The following list samples the results of the investigation in the order they have been presented up to here:

- Classification into background field and flux tube component: The separation into bg field and ft component, using only the inclination to the solar surface as criterium, leads to two different types of structures with markedly different properties in all other atmospheric parameters.

- Radial variation: The azimuthally averaged values for background field and flux tube component differ in the absolute numbers and the trend with radius.
  Field strength: The bg field strength drops from 2500 Gauss in the umbra to 1000 Gauss at the outer spot boundary, with a different slope in umbra and penumbra. The ft component is weaker by about 1000 Gauss in the inner penumbra, and decreases slower in the penumbra. At the outer spot boundary both components have equal field strength.
  Inclination to the surface: The background field inclination is always below 90°; it starts with 0° in the umbra and reaches around 60° at the outer spot radius with a uniform slope. The ft component is first visible with an inclination around 26° at the inner penumbral border, reaches 90° in the mid to outer penumbra, and increases further. The field lines are bending down again already inside the spot.
  Temperature: The background field is hotter than the ft component everywhere but for the innermost penumbral boundary. The ft component is hotter by about 1000-1500 K at the first appearance of the ft component.
  Velocity: The background field is almost at rest in the umbra, and shows significant velocities only at the outer spot boundary. Opposite to that, the ft component shows line-of-sight velocities around 2 kms^{-1} as soon as it is visible. The velocity in the ft component increases to around 6 kms^{-1}, if one removes the line-of-sight effects.

- Field aligned flows: The agreement between the field inclination to the surface and the flow angle is only good for the flux tube component in the outer penumbra, where the flow angle can be determined with some accuracy, because of the higher flow velocity. The two values are close enough to support the concept of field aligned flows, i.e. the material motion is parallel to the field lines.

- Net Circular Polarization (NCP): The NCP of infrared and visible spectral lines has different symmetry properties inside the spot. The visible NCP shows one maximum/minimum on an azimuthal path around the spot center, while the infrared NCP has twice that number of extrema. The Gaussian inversion is able to reproduce the NCP of single observed profiles on the limb side of the spot very well. This indicates that the atmospheric stratification derived by the inversion code has correctly positioned discontinuities along the line of sight, i.e. the profiles contained enough information to place the flux tube component at the correct height in the atmosphere.

- The integration of the azimuthally averaged inclination to the solar surface (LRF inclination) in radial direction – based on the assumption of either a single coherent field line or negligible depth variation of the inclination – leads to a reasonable curve in a geometrical height scale. The curve for the background field fits very well to the magnetopause of the tripartite spot model of Jahn & Schmidt (1994) from the outer penumbral boundary up to a radius of r = 6 Mm. For $r < 6$ Mm their spot model is much more vertical than the integrated LRF inclination. This implies that the LRF inclination values measured at the surface (formation height of the spectral lines !) is not appropriate for the deeper layers. Note that around r = 6 Mm a depth of > 5 Mm **below** the surface is reached.

  The height of the ft component, derived from the integrated LRF inclination, is in surprisingly close agreement with the height of the Gaussian perturbation after conversion from optical depth to a geometrical height scale for r > 8 Mm. The ft component shows an elevation of around 100 to 150 km above the solar surface in the outermost penumbra.

- Location of the Gaussian perturbation: For $r < 8$ Mm the Gaussian perturbation is located close to the solar surface. At r = 8 Mm it moves to lower layers, but immediately rises again. It reaches the surface defined by the log τ = 0 level around r = 9.5 Mm. For $r > 9.5$ Mm the ft component is above the surface, and continues slightly elevated until the outer penumbral boundary. The counterintuitive behavior for $r < 8$ Mm can be

traced back to the inversion procedure. If the line of sight cannot penetrate through the embedded flux tube, only its upper boundary is seen. The code then has no information, where to place the ft component along the line of sight, besides that it should be near the log $\tau = 0$ level, where the atmosphere becomes opaque.

- Diameter of the ft component: The width of the Gaussian perturbation yields a structure with a diameter of around 150 km, after conversion from optical depth to the geometrical height scale.
- Integration of the LRF inclination in azimuthal bins – instead of using the azimuthally averaged values – yields a "toy-model" of the 3-D topology of the sunspot. The 3-D model indicates that some properties of the spot, like the "spines" of enhanced field strength in the bg field, are a consequence of the subsurface structure of the spot. Positioning also the flux tube component inside the model, with its temperature as color code, leads to a strikingly resemblance with high-resolution sunspot images for a top view of the model: the bright penumbral grains at the umbra-penumbra transition on the limb side of the spot correspond very well to the inner footpoints of the ft component, where the temperature is increased. The shape of the ft component shows the influence of the bg field structure: low/high elevation of the bg field appears in the ft component in the same way.
- The line-of-sight velocity indicates upstreams in the limb side of the spot at the inner penumbral boundary, where the ft component is hotter than the bg field. The highest LOS velocities are reached on the inner penumbral boundary for the center side and at the outer penumbral boundary on the limb side. This supports the assumption of field-aligned flows, because the field inclination given by the inversion is parallel to the LOS on these locations.

To summarize the results in a few sentences, the observations and their evaluation in the present work give the following picture: the Evershed flow is confined to slightly elevated flow channels with 100-200 km diameter that show hot upstreams in the innermost penumbra. Some of the flow channels return to the surface inside the penumbra. Their respective shape around the spot is influenced by the structure of the background field. The flow velocity in the channels reaches 5-6 kms^{-1}; the flow angle is in good agreement with the field inclination in the outer penumbra. The background field continues upward everywhere in the spot and at the spot boundary.

The integration of the field inclination to the surface allows – within limits – to derive the field structure of both fields above and below the solar surface given by the log $\tau = 0$ level. To emphasize it, the final geometrical model has been derived (more or less) directly from the observed profiles.

5.13 The results in the framework of current knowledge: Comparison with observations and numerical models

The main sources for the comparison with earlier observations and the theoretical and numerical works will be described briefly in the following list to provide the reader with the necessary information without having to read the respective papers. For the observational publications, the conclusions drawn depend to some extent on the approach chosen in the evaluation of the data.

Observations The following more recent publications on sunspot observations will be used as main items for the comparison: Two papers of the series by Westendorp Plaza et al. (2001b,a), the studies of Mathew et al. (2003); Borrero et al. (2004), and the work by Bellot Rubio et al. (2004). These studies differ in the observational data used and the evaluation methods, but they can partly be compared to each other and the work presented here. In connection they can give an impression of the continuous evolution of data evaluation and interpretation that led to the present state of sunspot physics – keeping in mind the large amount of information all these works have from the last decades of scientific inquiries on solar physics. Other publications used in the discussion will be cited directly on their appearance as they usually refer only to some details.

Westendorp Plaza et al. (2001b,a): The series of papers present observations of the visible lines at 630.15 nm and 630.25 nm, taken with the Advanced Stokes Polarimeter (Tomczyk et al. 1992; Skumanich et al. 1997). The data were analyzed with the SIR code in terms of a 1-component atmosphere with gradients in physical parameters. The code was allowed to use 5 nodes in the optical depth scale to modify the atmospheric stratifications

(cf. Sect. 3.3.3). The discussion of the results is given in the optical depth scale, no conversion to geometrical height is attempted. The data contain 4214 spatial positions that were inverted individually.

Mathew et al. (2003): The data consists of infrared observations at 1.5 μm taken with the TIP. They are analyzed with the SPINOR code (Frutiger et al. 1999; Frutiger 2000) using a 2-component model, but only one of the components contains a magnetic field. The inversion is allowed 6 nodes in optical depth for the variation of physical properties, which makes the approach similar to the one of Westendorp Plaza et al. (2001b). The inversion takes into account a pair of OH-lines close to the FE I-line at 1565.2 nm that were also observed, but not used in the present work. A fit of these lines requires a more complicated procedure, where for example the abundance of the molecule with temperature has to be calculated as well.

Borrero et al. (2004): Again, these work presents observations of the infrared lines at 1.5 μm taken with the TIP, which were analyzed with the SPINOR code. However, in this work inversions using 1-component (1C, following their notation) with gradients are compared to 2-component (2C) atmospheres with parameters constant in optical depth.

Bellot Rubio et al. (2004): Observations of the infrared lines at 1.5 μm, taken with the TIP. These data were analyzed with the SIR code in terms of a 2-component atmosphere throughout the penumbra and umbra. This publication can be seen as the direct predecessor of the present work.

The approach chosen in the evaluation of the data in the present work is the continuation of the methods in the observational publications. The usage of 1-component inversions with gradients in Westendorp Plaza et al. (2001b) and Mathew et al. (2003) is followed by the comparison between 1C and 2C inversion by Borrero et al. (2004).

Theory and numerical models On the theoretical and numerical side, the main sources for the comparison are the works of Jahn & Schmidt (1994), Schlichenmaier et al. (1998) and Müller et al. (2002). The last publication is especially important, because it gives predictions on the observational findings that match with the current observations presented in this work.

Jahn & Schmidt (1994): The paper presents the calculation of a magnetostatic sunspot model based on the assumption of two flux sheets, which separate the umbra from the penumbra, respectively, the penumbra from the field-free surroundings. An initial model is iterated, until a static configuration for the magnetic field is reached. The model incorporates a "thick" penumbra, i.e. the field in the penumbra is not only a surface phenomenon, but extends to a depth of some Mm.

Schlichenmaier et al. (1998): In this work the temporal evolution of a flux tube is studied, which is placed in one of the sunspot models of the previous publication. In the beginning, the flux tube is located along the outer flux sheet which separates penumbra and field-free surroundings. The iteration of the magneto-hydrodynamic equations leads to a rise of the flux tube through the penumbra, which lead to the nickname "Moving Tube Model" (MTM) for these simulations. The final stationary result of the model is a slightly elevated flux channel, in which a steady outflow is generated.

Müller et al. (2002): The paper is based on the previous publication. The final result of the MTM was placed in a global sunspot model. Profiles of the infrared line at 1564.8 nm and the visible line at 630.25 nm are calculated using the radiative transfer equation. The calculations for different line-of-sight inclinations provide a prediction of the Net Circular Polarization in the respective wavelengths.

An important result of Westendorp Plaza et al. (2001b) refers to the field geometry outside the white-light penumbral boundary. Just outside the penumbral boundary, they found inclined fields with $\gamma \sim 60^\circ$ in the upper layers of the atmosphere at $\log \tau = -2.8$, together with horizontal or downwards oriented field lines in the lower layers at $\log \tau = 0$. This has a striking resemblance to the results given in, e.g., Fig. 5.35: the background field continues upward at the outer penumbral boundary with an inclination to the surface around 60°, whereas the flux tube component is horizontal or diving back below the surface. The same structure for the outer penumbral is suggested by Weiss et al. (2004) from numerical simulations of the so-called "flux pumping": field lines outside the spot are pushed below the surface by the convective motions of the granulation.

Westendorp Plaza et al. (2001b) found also downflows inside the penumbra, which were co-spatial to fields with an inclination above 90°. The results of Mathew et al. (2003) are in general very similar to Westendorp Plaza et al. (2001b). The major disagreement is with the found magnetic field gradients with optical depth or height in the atmosphere: Mathew et al. (2003) found fields decreasing with height, whereas Westendorp Plaza et al. (2001b) found increasing field strength on many of the inverted pixels. This contradictory behavior is most likely not due to the respective inversion code or the usage of infrared or visible spectral lines. It outlines the weak point in both

their approaches: the number of nodes in the inversion. With 5 nodes, the number of degrees of freedom gets too large. The resulting atmospheric stratifications will correspond to the actual solar atmosphere only in a restricted range, but may go to rather arbitrary values at all height levels that contribute only weakly to the synthetic profiles.

For this work, simultaneous data from the same visible and infrared spectral lines as used by the previous authors was available. Nonetheless, we have refrained from allowing for gradients in the atmospheric stratification. On test runs, we found that the atmospheric stratifications of the best-fit profiles with gradients showed a large pixel-to-pixel scatter, i.e. the results on adjacent pixels were very different and of random character. Quantities derived from the gradients – which essentially is what the comparison of layers with different optical depths is based on – seemed to be unreliable. We believe that the usage of the 2-component inversion, with parameters constant in optical depth, is better suited to resemble the physical conditions in a sunspots' penumbra.

The study of Borrero et al. (2004) concludes the two earlier works in some way. Their comparison of the 1-component (1C) and the 2-component (2C) inversion showed that the two different model approaches are consistent. They found that the parameter values of the 1C model are mirrored in the values of the 2C model: a field strength, decreasing from 2 to 1.2 kG from log τ = 0 to -1.0 in the 1C inversion on a spatial position, led to 1.9 kG, respectively, 1.3 kG in the 2C inversion results of the same profiles (cf. their Fig. 3, the same holds for other parameters like velocity or inclination). Their curve of radial variation of field strength is in close agreement with the curve given in Fig. 5.4.1: the 2nd (flux tube) component decreases radially with a significantly smaller slope than the stronger background field. Borrero et al. (2004) gave already a good argumentation for the work presented here: firstly, they noted that they had to include the gradients to fit the Net Circular Polarization, but that the gradients without further restriction led to unreasonable results as mentioned above. Secondly, the 2C inversion can be interpreted straightforwardly, and yielded already a very good fit to the infrared lines.

Their findings thus support the approach of using a 2C inversion, which is in first order equivalent to an 1C inversion with gradients, but lacks the problem in the interpretation of the gradients. This leads to the remaining observational publication, the study of Bellot Rubio et al. (2004). Their results agree on a list of points with the work presented here:

- The radial variation of parameters (their Fig. 6) is almost identical to Fig. 5.4.1. The field strengths of the two inversion components is identical at the outer penumbral boundary, with the ft component being weaker elsewhere. The ft component reaches and exceeds 90° around r/r_{spot} = 0.8, i.e. inside the penumbra, and the background field is around 50° inclination at the outer penumbral boundary.
- The flow velocities are around 6 kms^{-1} for the ft component in the middle penumbra, and around 2 kms^{-1} for the bg field at the outer spot boundary (cf. Fig. 5.22: 5-6 kms^{-1} and 1.5-2 kms^{-1}, respectively).
- The flow angle and the magnetic field inclination are in very close agreement; in their case throughout the whole spot.

In general, the earlier observations yield the following picture: one component inversions are more unreliable due to the (unphysical) gradients, but they have been shown to be – within limits – equivalent to a 2-component inversion without gradients (Borrero et al. 2004). The 2-component inversions support the confinement of the Evershed flow to (more) horizontal channels with weaker fields. Part of the more inclined component dives back below the surface already inside the spot. However, it has to be noted that the 2C inversion have a serious drawback: their assumption of two fully independent atmospheric components in each resolution element looses all height information on the respective structures represented by the components. The final proof of any model of a sunspot penumbra thus will have to use the Net Circular Polarization (NCP), the only criterion, where the 1C inversion with gradients and 2C-inversion with constant parameters are not equivalent. The Gaussian inversion in this work (cf. Sect. 5.6.1) is a way to model the penumbral fine structure, because it uses the NCP to retrieve the height information.

In the comparison with the numerical simulations, the close agreement with the magnetostatic model of Jahn & Schmidt (1994) has to be noted. The integration of the azimuthally averaged inclination to the solar surface – measured close to the surface – yields a curve for the background field in the geometrical height scale that is very close to the theoretical calculation (cf. Fig. 5.35). This item is more important than it may seem on first look, because the magnetostatic model is the basic ingredient to the later numerical works. In a sense, the observations agree with the theoretical considerations *ab initio*.

The results of the Moving Tube Model in the work of Schlichenmaier et al. (1998) can be summarized as follows: the final stationary configuration of the flux tube shows a hot (up to 14000 K) upstream point, where the flux

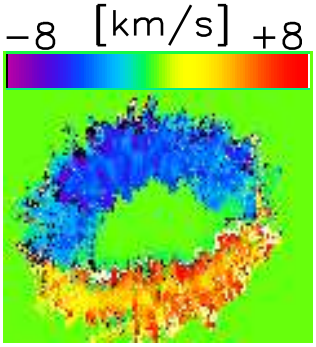


Figure 5.49: The "true" velocity of the flux tube component after the de-projection of the line-of-sight effects. It has been calculated with the assumption of field-aligned flows from the field inclination to the LOS, γ, by $v_{\rm true} = v_{\rm LOS}/\cos\gamma$. The true velocity is very uniform all around the spot. The sign of the velocity is positive/negative, if the flow is oriented away from/towards the observer.

tube reaches the solar surface around r = 8 Mm[3]. The flow velocity inside the elevated (around 100 km above the surface) channel starts with 3 kms^{-1} at the inner footpoint, and reaches values up to 13 kms^{-1} in the outer penumbra. To first order, the numbers are in good agreement with the values of the horizontal velocity for the flux tube component derived from the observations, around 6 kms^{-1} (cf. Fig. 5.22).

However, there is an additional way to derive the flow velocity from the measured line-of-sight velocities: with the assumption that the flow is perfectly parallel to the field lines (cf. Sect. 5.4.3 for a partial proof), the true velocity, $v_{\rm true}$, can be derived from the observed line-of-sight velocity, $v_{\rm LOS}$, and the field inclination to the LOS, γ, by $v_{\rm true} = v_{\rm LOS}/\cos\gamma$.

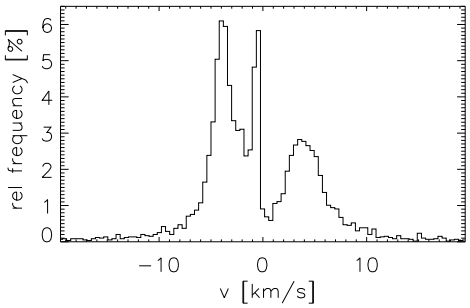


Figure 5.50: Histogram of the relative frequency of the "true" velocity after the de-projection of the line-of-sight effects. The two distinct peaks at $\pm$ 5-6 kms^{-1} give the average velocity of the Evershed flow on the limb side, respectively, the center side. The third peak around -0.5 kms^{-1} comes from the pixels on the inner penumbral boundary on the limb side, where the hot upstreams are located.

Fig. 5.49 displays the resulting velocity map of $v_{\rm true}$ for the flux tube component: all around the spot the flux tube component shows an almost uniform velocity with a mean value around 4 kms^{-1}. The histogram of the relative frequency of velocities in Fig. 5.50 reveals two distinct peaks around $\pm$ 5-6 kms^{-1}, with in both cases an extended tail to velocities up to $\pm$ 10 kms^{-1}. This result is compatible with the findings of Bellot Rubio et al. (2004), who found similar numbers applying the same method. The higher velocities are closer to the prediction of the MTM simulations. The 3rd sharp peak in the histogram around -0.5 kms^{-1} – negative velocities denote motion towards the observer – has been found to come from the hot upstreams at the inner penumbral boundary on the limb side. A mask showing only pixels with velocities between -1 and 0 kms^{-1} (not displayed) had the main contribution on the inner limb side penumbra (cf. Fig. 5.44, p. 115). The deprojected velocity does support the high flow velocities predicited by the MTM model.

Thanks to the conversion to geometrical height, the final result for the flux tube component from the observations can be directly compared to a snapshot of the MTM. The agreement between the the theoretical and observational results is surprisingly close. Fig. 5.51 displays in the top panel the geometrical shape of the flux tube in the MTM, with the field strength of background and flux tube as color coding. The middle panel displays the same from the observations, with an identical range for B. The characteristic reduction of B at the position of the hot upstream footpoint in the simulations is present as well in the observations. The slight dip in the flux tube elevation around r = 8.7 Mm in the MTM is discernible in the observations, at least as a change of slope at the same radial position. The trend of flux tube and background field to an identical field strength at the outer penumbral boundary also appears in both images. The lowermost panel displays the flow velocity in the observations, taken to be only the

[3]The spot radius in the simulation is 13.5 Mm, comparable to the radius of the observed spots.

horizontal speed. Even if the parameter does not show the trend to reach the supersonic velocities in the outer penumbra as in the simulations, the values at and close to the upstream point are comparable.

Another on the first look serious disagreement between simulations and observations are the actual numbers of physical parameters: in the simulations, the temperature in the hot upstream point and the flow velocity in the outer penumbra are a factor of 2 larger than the corresponding values in the observations. However, this has an easy explanation in the spatial resolution of the observations. The nominal spatial resolution is 2 pixels, equal to 0.7" (= 500 km). Taking into account the seeing due to the disturbances in the Earth's atmosphere worsens to around 1" (= 725 km). Even if the additional information of the polarization degree allows to see – in some sense – below the resolution limit, i.e. one has information on the fraction in the resolution element covered with a certain type of structure, the observations do not directly resolve the flux tubes with 100-200 km diameter (or radius). At least for the flow velocity one could take into account the filling fraction of the flux tube component which is around 50-60 % in the penumbra to argue that the true velocity in the flux tubes will be 1.6-2 times larger if the structure were resolved. In general, the observations are in close agreement with the MTM as far as the geometry of the penumbral flux tubes is concerned (cf. Fig. 5.51).

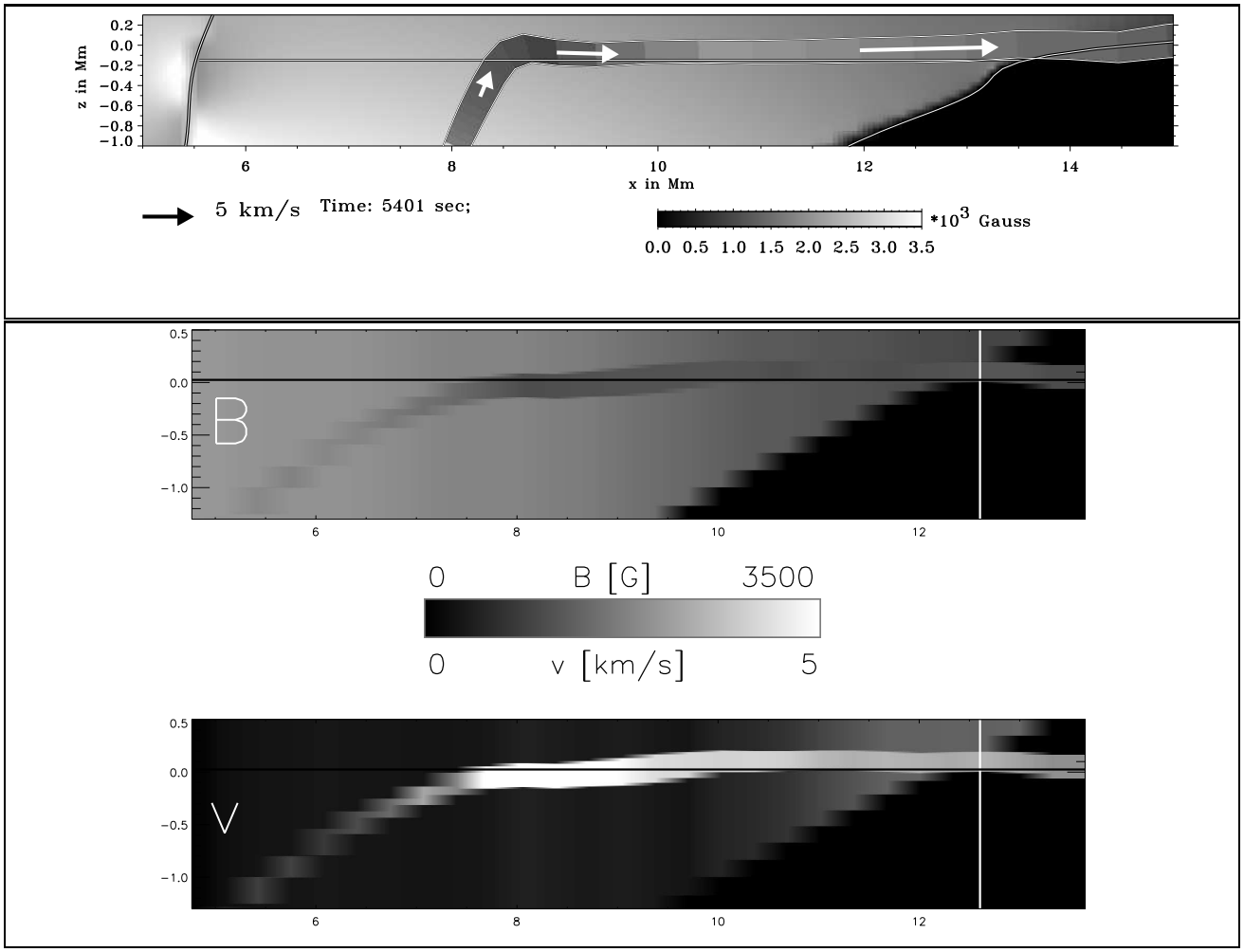


Figure 5.51: Comparison of a snapshot of the Moving Tube Model (Schlichenmaier et al. 1998) (**top panel**) with the final result for the flux tube component in this work (**lower two panels**). The plots have been arranged to have approximately the same scaling and position in the horizontal axis which gives the distance from the spot center in Mm. Both **top** and **middle** panel display the flux tube component and the background field with their respective field strength as color coding. The bottom panel uses the horizontal velocity of the ft component as color scale.

Keeping in mind the complexity of the data evaluation procedure and the following interpretation of the inversion results, one might wonder if an easier way exists to proof the picture drawn by the final state of the Moving Tube

Model. Actually, there is[4]: the Net Circular Polarization. As described in Sect. 5.5, the NCP is connected to gradients or discontinuities in physical parameters of the magnetic field and the velocity. The MTM yields an embedded flux tube, which HAS discontinuities between the background field and the flux tube. In the work of Müller et al. (2002), NCP maps of a complete sunspot were calculated on the base of the MTM results. The final MTM model was put inside a static background field similar to the one used in the simulations. The spot was assumed to be perfectly circular; all azimuthal directions were identical, i.e. for each angle around the spot center the same radial structure, given by the final MTM result, was used. The heliocentric angle, and hence, the viewing angle on the sunspot model, could be freely chosen. On all spatial positions inside the spot synthetic profiles were then calculated from the known atmospheric stratification. Moving azimuthally around the spot center at a fixed radial distance, the geometry encountered along the line-of-sight ray is very different, even if the tube and the background field have the same physical properties. Thus, the NCP changes strongly, depending on the azimuthal position around the spot center.

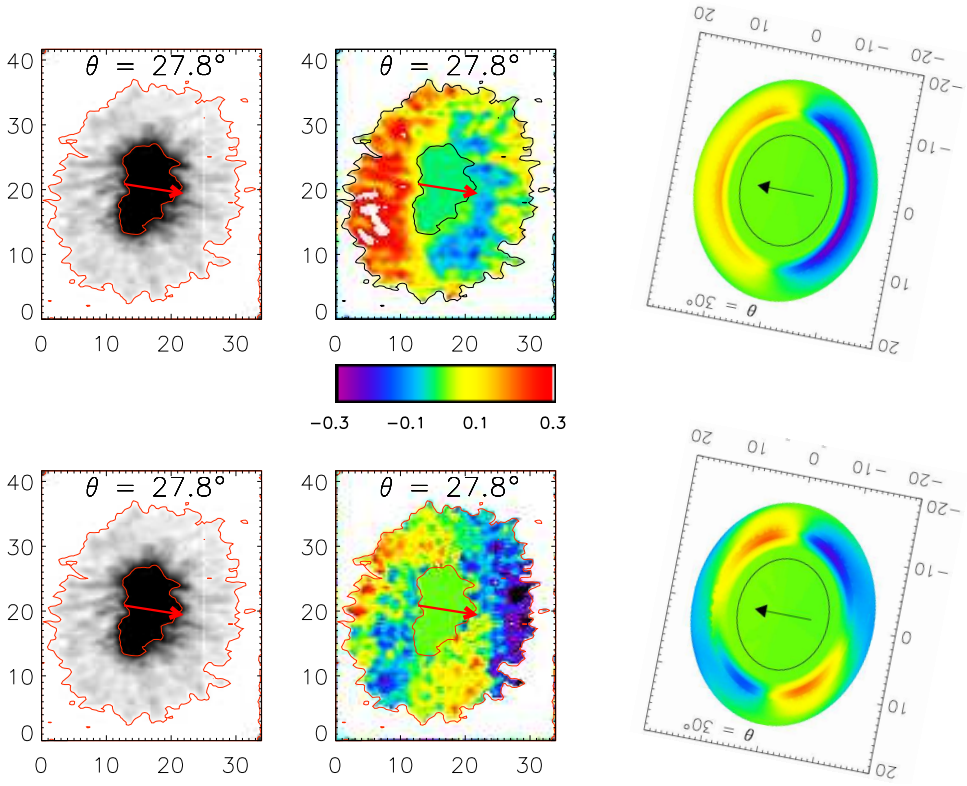


Figure 5.52: Comparison of observed and predicted Net Circular Polarizations. **Left to right**: observed intensity, observed NCP, and theoretical prediction (Müller et al. 2002). **Top row**: visible spectral line at 630.25 nm, **bottom row**: infrared spectral line at 1564.8 nm. The red, respectively, black arrow points towards disc center. The observed spot had negative polarity (field lines going into the surface) which flips the symmetry axes by 180°.

Additionally, the NCP of (magnetically sensitive) spectral lines depends on the ratio of their Zeeman splitting to the their thermal Doppler width or velocity induced Doppler shifts (Müller et al. 2002). The Zeeman splitting is proportional to λ^2, whereas the Doppler broadening and Doppler shifts are linear in λ. This is very important in the comparison of the infrared lines at 1.5 μm and the visible lines at 630 nm, separated by a factor of around 2.5

[4]How surprising...

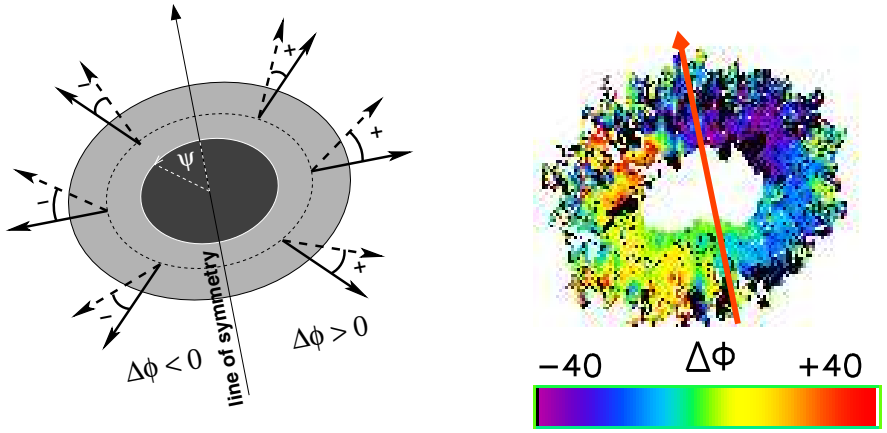


Figure 5.53: The projection to the line of sight leads to an azimuth difference between the horizontal tube (solid black arrows in left panel) and the less inclined background field (dashed arrows). The thin black arrow points towards disc center (image courtesy of D.Müller). For comparison, the observed azimuth difference between the two inversion components is plotted in the right panel. Here the red arrow indicates the orientation of the symmetry line.

in λ. Fig. 5.52 compares the observed NCP maps of the visible spectral line at 630.25 nm and the infrared line at 1564.8 nm for the sunspot with its heliocentric angle of 27° with one of the artificial NCP maps of the same lines for an heliocentric angle of 30°, taken from Müller et al. (2002). The different symmetry properties of the visible and infrared (IR) spectral lines are clearly visible in both observations and theoretical prediction: Going once azimuthally around the spot center, the visible line shows a single period, whereas the IR line has twice that frequency. Note that the observed sunspot had opposite polarity, i.e. the field lines are pointing into the surface. The theoretical map had thus to be rotated by 169°. The absolute value of the NCP is off by a factor of 3-5 which most probably is again due to the spatial resolution of the observations, where the flux tubes are not resolved, leading to a filling fraction of the flux tube far from the 100 % assumed in the theoretical model.

Müller et al. (2002) derived that the prominent difference between visible and infrared lines is due to a different sensitivity of the lines to the discontinuities in inclination, γ, and field azimuth, Φ: Δγ leads to a symmetric shape with respect to the symmetry axis, the line connecting disc center and spot center, whereas ΔΦ is anti-symmetric, and enters the NCP with double frequency (Landolfi & Landi degl'Innocenti 1996). Müller et al. (2002) concluded that "the ΔΦ-effect is by far the dominating source of asymmetry for the infrared line, while for the FE I 630.25 nm, the ΔΦ- and Δγ-effects are of the same strength" (ibidem, p. 314). The observations are here in very close agreement with the theoretical prediction in a direct observable, which is only based on the observed profiles without the complex inversion procedure.

The sign of the discontinuity in field azimuth, ΔΦ, can be derived easily from the configuration of an horizontal flux tube in a less inclined background field (cf. the left panel of Fig. 5.53). The corresponding observation value, the difference of the field azimuth of the two inversion components in the line-of-sight reference frame, is plotted in the right panel. Again, the theoretical predictions is matched by the observations. The observed azimuth difference has the predicted symmetry properties, with a positive/negative sign on one/the other side of the symmetry line.

5.14 Conclusion

From the observations the following picture of a sunspot penumbra can be drawn: Inside the penumbra the sunspot exhibits two different types of structures in EACH resolution element, a more inclined component with weaker fields, which harbors the Evershed flow (*flow channel* or *flux tube*), and a less inclined component with stronger field, which is almost at rest (*background field*). The derivation of the 3-D topology from the uncombed "Gaussian" inversion, which uses the information contained in the Net Circular Polarization (NCP), leads to slightly elevated flow channels with a diameter of around 200 km. The flow channels start from hot upstreams in the innermost

penumbra, rise above the solar surface defined by the continuum forming layer (log τ = 0) around a radius of 7-8 Mm and continue almost horizontally outwards. The flow channels return to the surface in the outermost penumbra, whereas the background field continues upwards all around the spot.

The observations and their final result for the background field structure and the flow channel geometry given in Figs. 5.34 or 5.33 support the simulations of the Moving Tube Model (MTM Schlichenmaier et al. 1998). They agree in the initial (e.g, the shape of the background field) and the final geometry of the simulations, and also – within the limits set by the spatial resolution of the observations – with some physical properties of the flux channels, like temperature and velocity. The observed NCP and azimuth difference follows the prediction of Müller et al. (2002) derived from the MTM; the predicted behavior of both visible and infrared lines is matched by the observations.

The derivation of the field geometry in 3-D all around the spot by integrating the inclination to the surface in radial direction offers the opportunity to relate the behavior and evolution of the flow channels to the subsurface structure of the background field at the same spatial position. The appearance of the bright penumbral grains on the limb side of the spot can be shown to happen naturally in the 3-D spot model, if viewed from above, even if no trace of the grains is visible in the corresponding intensity maps of the observations due to the limited spatial resolution. The integration of the inclination suffers from a severe drawback due to the extrapolation involved. Nonetheless, it is a powerful visualization tool and yields – to first order – the same geometry of the penumbra as the more complex Gaussian inversion.

An important conclusion to be drawn from the present work is the fact that the Evershed flow is not driven by a siphon flow mechanism as suggested by Thomas & Montesinos (1993). A siphon flow model is based on the total equilibrium of magnetic and gas pressure, p, of the two footpoints of an arched loop:

$$p_i + \frac{\mathrm{B_i^2}}{8\pi} = p_o + \frac{\mathrm{B_o^2}}{8\pi}\,, \tag{5.10}$$

where the subscripts i and o denote the inner and outer footpoint, with the assumption that the loop is located in a sunspot and oriented radially. With $\mathrm{B}_o > \mathrm{B}_i$, the gas pressure on the outer footpoint will be smaller, leading to a flow of material from the inner footpoint outwards. The problem in adopting such a model as driver of the Evershed effect in a sunspot penumbra is the lack of stronger fields at the outer footpoints. The radial variation of the field strength of the two inversion components given in Fig. 5.4.1, p. 90, shows no indication for an increase of B with radius, even if the slope of the flux tube component is smaller than that of the background field.

It has been argued that the eventually existing stronger outer footpoints have not been observed due to the spatial resolution of the observations (e.g., in Borrero et al. 2005) or that the flow channels from inside the penumbra connect to magnetic elements outside the spot (Weiss et al. 2004). To the first point it can be answered that neither the present nor other observations cited in the previous section show at a spatial resolution of 1" any traces of localized field concentrations with higher field strength in the outer penumbra. Especially using the infrared lines with their large Zeeman splitting, it is hardly conceivable that localized increases of B can be missed, even if the structures are not fully resolved. The second argument for stronger outer footpoints can be shown to be even weaker: Weiss et al. (2004) state that "the magnetic field strength at the inner footpoint is around 1000 G (a typical penumbral field strength) but at the outer footpoint is of order 1500 G, corresponding to a typical intense, collapsed photospheric magnetic element" (ibidem, p. 1086). The first number is not in agreement with the results found in the analysis of the sunspot data in this work: The penumbral field strength of the flow channel is reaching 1 kG at the outer penumbral boundary, and definitely not anywhere inside the penumbra. The second number is in serious contradiction with the results for the properties of magnetic elements given in chapter 4: 50 % of the G-band bright points – which indicate strong magnetic elements – have field strengths below 1 kG, and virtually none exceed 1.5 kG. Thus, it seems very improbable that the Evershed flow is driven by a stationary pressure gradient between the weaker inner and stronger outer footpoint, because the field strength is larger for the inner footpoint.

A second conclusion refers to the structure of the penumbra: From the good agreement of the MTM simulations and the observations, or assuming the integration of the inclination to the surface to be valid at least for some depth range, the penumbra is not only a shallow surface layer. It extends to a depth of some Mm, which implies that there is sufficient space for the dynamical evolution of structures inside the penumbra.

Chapter 6

Conclusion

This section is aimed at placing the results on the sunspots' global and penumbral fine structure and the properties of the magnetic elements in the close surroundings of a sunspot in a coherent picture of formation, evolution, and decay of a sunspot. The first section describes the "origin" of a sunspot using the differences between sunspots, with umbra and penumbra, and pores. The fine structure of the penumbra and its relation to the energy transport are discussed in the subsequent section, while the final section of the conclusion draws a picture of the disappearance of a sunspot from the solar surface.

6.1 Large-scale magnetic fields in the solar atmosphere – The difference between sunspots and pores

Two types of large-scale magnetic fields can be observed in the solar atmosphere: sunspots and pores. They are both due to the presence of intense magnetic fields on large-scales, yet, their properties are markedly different. Sunspots have been found to develop from pores, and decaying sunspots can revert in the latest stage to something similar to pores. Sunspots show the separation into umbra and penumbra, whereas pores resemble an isolated umbra. Thus, two questions arise naturally: What is the difference between the two types of structures, and what do they have in common ?

To answer the question, one has to turn back to the origin of the magnetic fields visible in the photosphere. It is generally assumed that the solar magnetic flux is generated in the boundary layer between the rigidly rotating radiative core and the differentially rotating convection zone. Existing fields will be dragged around the Sun, where again the differential rotation, which is faster at the equator than at the poles, will produce field lines parallel to the equator in a short time regardless of the earlier field orientation. Schussler et al. (1994) and Caligari et al. (1998) studied the evolution of these flux concentrations at the bottom of the convection zone in 3-D magneto-hydrodynamical simulations. A bundle of field lines (a "flux rope") was slightly displaced upwards from its equilibrium position at some point (cf. Fig. 6.1, panel a). The flux rope was found to become buoyant at the disturbed position and started to rise through the convection zone, until it reached the solar surface defined by the $\log \tau = 0$ level of the optical depth scale. The rest of the flux rope remained anchored in the convection zone, leading to a large loop structure with two footpoints. Large-scale magnetic fields on the sun are in this picture to be seen as the cross-sections of the flux ropes with the surface. Note that the model at once yields an explanation for the existence of bi-polar sunspot groups with two spots of opposite polarity and an East-West orientation. However, the loops formed by these flux ropes are orders of magnitude above the size of a single sunspot. For the present discussion, it thus is sufficient to take only one of the footpoints and neglect the fact that the field lines have to be closed again to somewhere else.

Being left with a single magnetic concentration of still large scale, the question arises as to which form this field will take, when it breakes through the solar "surface". The fate of the field is intimately connected to the definition of the surface – taking into account that the Sun consists of a plasma. The usual definition of the surface is based on the optical depth scale. Light rays going towards the Sun will be absorbed by its gas and penetrate only to a certain depth. The absorption depends on the number density of free absorbers, (ionized) atoms and electrons. However, the number density experiences a drastic change at some point in the atmosphere, when the temperature

is low enough for the re-combination of protons and electrons to hydrogen atoms, or, going the other way, for the ionization of hydrogen. Thus, the usual definition of the solar surface as the height, where the amplitude of an incident light ray with a wavelength of 500 nm is reduced to 1/e, puts the surface layer close to where the ionization (or re-combination) of hydrogen happens.

For the flux concentration this implies that its boundary conditions change drastically around that height: The outside gas pressure starts to drop exponentially, because the free electrons disappear. For the sake of clarity in the following some quantities like magnetic curvature forces will be neglected and some assumptions will be made, e.g., lateral temperature equilibrium. With this simplifications, the remaining equation for the evolution of the flux concentration is the horizontal pressure balance between its surroundings, $p_{\rm ext}$, and its interior, $p_{\rm int}$:

$$p_{\rm ext}(\rho_{\rm ext}, T_{\rm ext}, B_{\rm ext}) = p_{\rm int}(\rho_{\rm int}, T_{\rm int}, B_{\rm int}) \tag{6.1}$$

which gives, introducing the gas pressure, $p_{\rm gas}$, and using the assumptions

$$p_{\rm gas}(\rho_{\rm ext}, T) = p_{\rm gas}(\rho_{\rm int}, T) + \frac{B^2}{8\pi} \,. \tag{6.2}$$

Note that B is given by the total amount of magnetic flux, Φ, and the area, A, covered by the flux: $B = \Phi/A$. If the particle density now is reduced, the outside gas pressure (and the interior as well) will start to drop, whereas the magnetic pressure is unaffected. To establish a pressure balance the field strength has to decrease, which means that the area A has to increase. The field lines will thus be pushed outwards more and more, leading to an ever growing inclination to the surface. Assuming the outside pressure to be identical in granulation everywhere on the Sun's surface, the final field inclination will depend to first order only on the amount of flux.

The distinction between a sunspot and pore results from the effect of the inclined field lines on their surroundings (cf. Fig. 6.1, panels b and c). The surroundings can be characterized by the following short list:

- In the field-free gas outside of the flux concentration, the energy transport is achieved by convective mass motions, because the gas is opaque.
- At the surface as defined above, the energy can be radiated away.
- The energy flux is not isotropically, but oriented radially outwards away from the solar core, where the energy is produced in the fusion process. The net energy flux will thus be vertically and directed towards the surface on longer time scales.
- The solar atmosphere below the surface is a plasma. No mass motion across field lines is possible.

For the pores with a lower total flux, the field inclination still is so close to vertical that the convection outside of the flux concentration is virtually unimpeded by the field lines. The energy can be transported upwards by convective motions. At the outer boundary of the flux concentration, the hot plasma can cool effectively at and close to the surface by radiating energy directly into the interior of the flux concentration with its reduced density. The material cools off, condenses, and finally starts to sink down along the boundary layer, leading to a persistent downflow around the pore, which has been observed to be the case by, e.g., Tritschler et al. (2002), or Sankarasubramanian & Rimmele (2003).

For sunspots with a higher total magnetic flux, the field inclination gets much larger in some depth below the surface, where the atmosphere still is opaque. The material outside of the spot at the boundary is kept hot by convection from below, but it cannot cool by radiation like in the case of a pore and cannot form convection rolls across the field lines. The energy transport inside the spot is definitely not as efficient as granulation: at the solar surface, the penumbra has only about 80-85 % of the intensity of undisturbed granulation. Thus, the inclined fields of a sunspot reduce the efficiency of the vertical energy transport, while the energy input from below is still the same. The possible consequences of this local excess of energy input over the energy transported away will be discussed in the next section.

To conclude the discussion of the large-scale magnetic fields, the main difference between pores and sunspots will be the critical flux value, Φ_c, or more precisely, the critical inclination value, γ_c: if the critical inclination is exceeded, the energy transport in the surroundings of the flux concentration is seriously affected by the magnetic field lines (cf. Rucklidge et al. 1995; Yang et al. 2003). The critical inclination has to be exceeded in opaque atmospheric layers below the surface, where the vertical energy transport by convective motions will be impeded

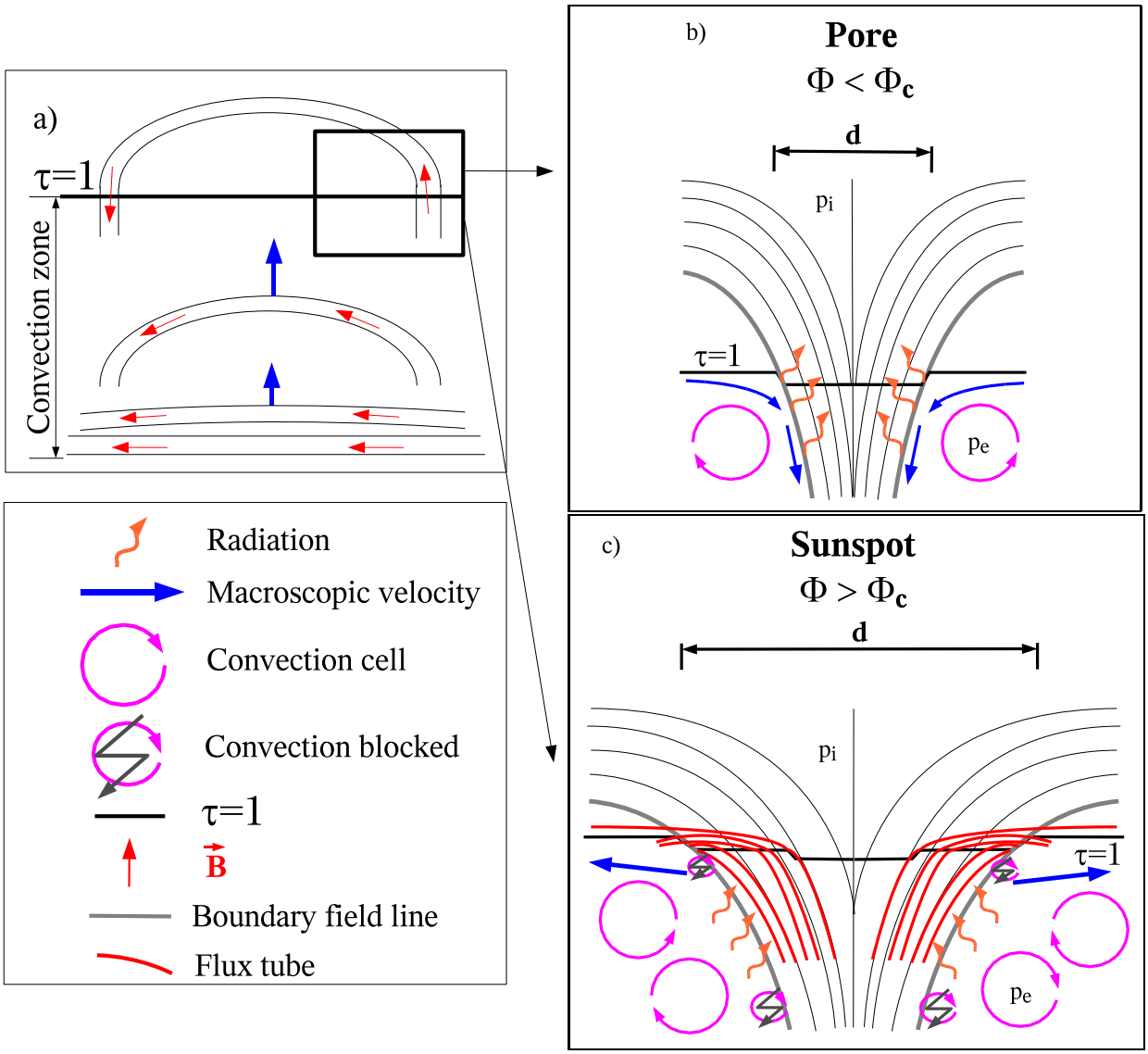


Figure 6.1: A flux rope stored at the bottom of the convection zone, which is slightly displaced from its equilibrium position, gets buoyant and starts to rise through the convection zone (panel a). The evolution after reaching the solar surface defined by the $\tau = 1$ level of optical thickness can lead to two different structures depending on the total magnetic flux of the flux rope, Φ : A pore (panel b) or a sunspot (panel c) can be formed. The difference between the two types of large-scale field concentrations is due to the inclination of the boundary field line. The value of inclination is governed by the pressure balance of internal and external pressure, p_i and p_e, respectively. As p_e is identical in the surrounding granulation, a larger amount of flux has to be spread over a larger cross-section, d, to yield the same field strength, B, that enters the internal pressure by the term $B^2/8\pi$. Thus, sunspots with their higher total flux will have a larger diameter, and hence, larger field inclinations than pores. However, the larger field inclination has further effects: energy has to be transported upwards, but the formation of overturning convection by the plasma close to the spot is suppressed by the inclined field lines already in some depth. The energy thus has to be transported by radiation, which strongly heats up the boundary layers between the sunspot and the field-free surroundings. This lateral heating leads to the rise of buoyant flux tubes in the sunspot penumbra. The macroscopic velocity fields around the large-scale structures have a straightforward explanation in the case of the pore: The plasma close to the pore can cool efficiently by radiation through the evacuated atmosphere inside the field concentration. The cooling leads to a density increase, and finally the plasma starts to sink producing a persistent downflow around the pore – which has been observed. For sunspots the opposite effect is the case: the moat flow, a radial outflow away from the spot, is an observational fact. A possible driver for the moat flow might be the heat trapped below the field lines of the sunspot. The local energy excess due to the reduced efficiency of the vertical energy transport could lead to a pressure gradient, which drives material away from the spot.

and radiative cooling is not possible. For pores, the surroundings in first order remain unchanged, because the flux amount is too low to enforce sufficiently inclined fields below the surface. For sunspots, the blocking of the vertical energy transport by convective motions – because the solar plasma cannot cross magnetic field lines – will lead to an energy excess below the outer sunspot boundary.

6.2 The penumbral energy flux problem & The penumbral fine structure

The penumbral intensity is only about 80 % of the intensity of undisturbed granulation. However, to achieve even this energy level in the absence of convective motions is not easy. In the solar convection zone, radiative transfer is inefficient (Stix 2002). The idea of the penumbra as a shallow surface layer, through which the energy can be transported by radiation alone, requires its depth to be very small. The penumbral field lines then would have to be nearly horizontal everywhere, which is in contradiction to the observations. Indeed, the penumbra was found to extend over a depth of some Mm (cf. Solanki & Schmidt 1993; Jahn & Schmidt 1994; Schlichenmaier & Solanki 2003; Solanki 2003; Balthasar & Collados 2005, or the present work). This, on the other hand, rules out energy transport by convection, as several convective rolls in the vertical direction would be needed to transport energy up to the surface – besides the suppression of any kind of overturning convection by the magnetic field lines. Overturning convection would also be in contradiction with the velocity field observed in sunspots. The primary velocity is the horizontal Evershed flow, which is oriented radially outwards, whereas convection would result in up- and downflows like in granulation.

One idea to solve the energy flux problem is the assumption that field-free material is enclosed between the field lines (Spruit & Scharmer 2005, subm.). Inside the field-free material, overturning convection would be possible. However, even if this geometry is able to explain the observed intensity level of the penumbra, it has some severe drawbacks. The existence of the field-free material should reduce the polarization level in observations. From various observations, the fraction of field-free material inside the penumbra was found to be between 0 % and at maximum 5 % (cf. Solanki 2003, p. 180). In the present work, the straylight coefficient of the inversion in the penumbra is around 14 % (cf. Fig. 5.20, p. 91). If one corrects this for the straylight level in the umbra of around 9 %, field-free material can cover again only a maximum fraction of 5 % of each resolution element. Spruit & Scharmer (2005) claim that their model "will be able to reproduce the observed NCP" (ibidem, p. 11), because it contains strong localized perturbations, but without any proof or calculation. In view of the difficulties in the reproduction of the infrared and visible NCP values at the same time, this unspecific claim seems rather doubtful. And finally, their model gives no explanation for neither the existence of the Evershed channels nor their observed properties.

Another suggestion to supply the energy needed to sustain the penumbral intensity is the so-called *interchange convection* (Jahn & Schmidt 1994). In this picture, the energy transport is due to the rise and descent of individual flux tubes. The flux tubes are first located at the lower boundary layer separating the field-free surroundings and the sunspot. They are heated up by radiation, until they become buoyant and rise to the surface. After they have deposited their energy by radiative cooling, they submerge again. This scenario has prompted Schlichenmaier et al. (1998) to study the evolution of such flux tubes (Moving Tube Model, MTM). The final result of the MTM simulation was an slightly elevated flow channel with a stationary flow. The MTM model is in good agreement with many observed properties of sunspot penumbrae. To name only some of them:

- The flux tube harbors an outflow of up to 14 kms^{-1}. It stays nearly parallel with respect to the surface throughout the outer penumbra ($\rightarrow$ the horizontal Evershed effect).
- A hot upstream is seen at the location, where the flux tubes reaches the surface. During the simulation, this intersection point moves towards the umbra ($\rightarrow$ penumbral grains in the inner penumbra).
- The flux tube is optically thick near the intersection point with the surface ($\rightarrow$ location and width of the perturbation in the Gaussian inversion).
- The calculation of the Net Circular Polarization for the geometry of the MTM by Müller et al. (2002) is in agreement with the observed NCP in visible and infrared spectral lines.

However, the simulations of Schlichenmaier et al. (1998) end in a stationary solution of an elevated flow channel, which does not submerge again: the inertia of the flow inside the flux tube stabilizes it. In its final state, the flow channel is only able to heat a restricted volume of the penumbra. The hot upstream point with more than 10000 K

cools very fast through radiation (Schlichenmaier et al. 1999). Even taking into account the flow along the tube, it is not possible to heat the penumbra along the full length of the flow channel (ibidem). A single stationary flow channel is thus not sufficient to deliver the amount of energy needed. To overcome this shortcome, Schlichenmaier & Solanki (2003) thus proposed that the penumbra should consist of two or more flow channels, arranged in radial direction one after the other. However, this would require that the flow channels have to bend downward already inside the penumbra, with observable consequences:

- The line-of-sight velocity should change sign.
- The inclination to the surface should exceed 90° for the submerging flow channel and be close to vertical for the next channel in the radial direction.

Another strong argument against the interchange convection as main energy supply in the penumbra is the long lifetime of penumbral filaments. Solanki & Rüedi (2003) and also later Langhans et al. (2005) found lifetimes of some 30 min or even hours. Solanki & Rüedi (2003) claimed that magnetic structures are even more longlived than intensity structures and remain without changes for up to two hours. Based on this result, Schlichenmaier & Solanki (2003) argued that "interchange convection as the dominant energy transport mechanism within the penumbra can be ruled out".

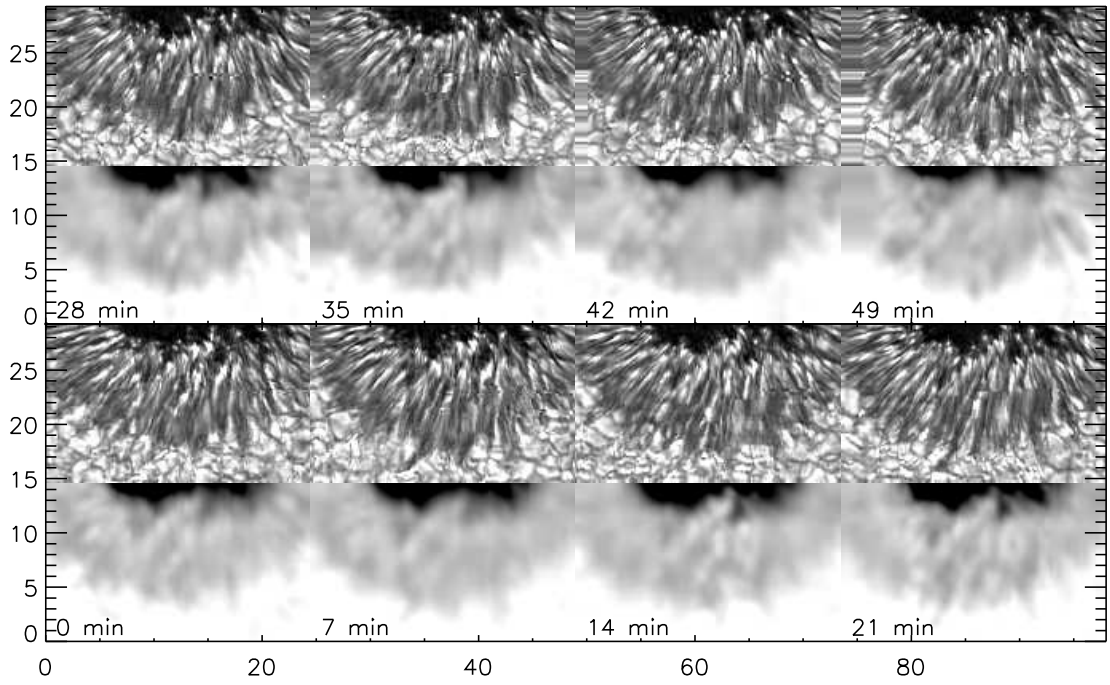


Figure 6.2: The temporal evolution of a sunspot's penumbra with 7 min cadence. The lower row gives the continuum intensity map in IR from the VTT, the upper row the co-temporal and co-spatial map in G-band from the DOT. The white rectangle outlines an area, where the shape of the penumbral intrusion stays more or less the same, whereas in the same time several penumbral grains can be seen to migrate into the area. Tick marks are arcsec.

However, there are some serious arguments against the scenario just described and in favor of interchange convection. Firstly, the scenario of a series of radially aligned flow channels is not supported by the present observations or the corresponding results of Bellot Rubio et al. (2004) (their Fig. 4). Inside the penumbra, neither the field inclination nor the line-of-sight velocity of the flow channel indicate a downward bending of one and the appearance of another flow channel. The field inclination of the flow channels exceeds 90° in the outer penumbra, but this seems to depend only on the radial distance to the spot center. Secondly, as far as the variability of the intensity and magnetic field structures is concerned, the type of data used by Solanki & Rüedi (2003) and Langhans et al. (2005) has to be taken into account. In both cases, line-of-sight magnetograms were used, which are based on

the measurement of only the circular polarization on a few wavelength points. Line-of-sight magnetograms cannot measure flux that is perpendicular to the line of sight, which shows up only in the linear polarization signal. Especially on the center side, a magnetogram signal will be dominated by the properties of the background field, whereas the horizontal flow channels do not show up at all in the signal. Magnetograms may also be misleading in the neutral line of Stokes V with its abnormal profiles, because they are taken at a fixed wavelength. Thus, they are intrinsically unsuited to detect the dynamic evolution of the horizontal flow channels. Additionally, taking into account the existence of recurrent sunspots with lifetimes of more than one month, the structure of the background field should not change drastically in a short time span.

As a demonstration of the variability of the penumbra, Fig. 6.2 displays the upper part of the field of view of the observations used in Chapter 4. The limb side of the spot was scanned repeatedly with a cadence of 7 min. The white rectangle marks a penumbral intrusion into the umbra, whose shape changes little during the almost 1 hour of the observations. The same penumbral intrusion shows up also in the 2 1/2 hour time average of the DOT data (cf. Fig. 6.3). However, in the same time several penumbral grains can be seen to migrate into the area, change their shape or disappear completely. The figure suggests that the time scale of small-scale dynamical changes – indicated by the evolution of the penumbral grains – is around 15 min or less. All longer-lived intensity structures are of larger spatial extent and most probably are connected to the more stable background field that defines the geometry, but does not take part in the energy transport.

Thirdly, the simulations of the MTM model predict that flux along the boundary of the spot should get heated up by radiation and get buoyant. If this scenario is correct, the process should repeat. As soon as part of the flux gets buoyant and rises upward, the next layer should experience the same. The infinite continuation of the process leads of course to one major problem of the interchange convection: the flux at the sunspot boundary has to be replaced, otherwise the whole penumbra would have to transform to rising flux tubes and should disappear from the photosphere.

The final piece of information for the construction of a consistent model of penumbral fine structure and a possible solution of the penumbral heat flux problem comes from the study of Leka & Skumanich (1998), which was already cited in the introduction of this work. These authors found that a penumbra is already fully developed, when it first appears on the solar surface. The newborn penumbra has the same inclination as a mature spot and shows the Evershed effect from the very beginning.

Collecting all this pieces of information, a possible scenario for the origin and the structure of the penumbral fine structure may look like this:

- If a large-scale flux rope reaches the solar surface, it will initially have almost vertical fields, because below the solar surface the magnetic pressure term is much smaller than the outside gas pressure. Inside of the flux concentration, overturning convective is impossible for the ionized plasma of the solar atmosphere. The flux concentration will thus have a reduced intensity and would be termed a pore or an umbra.

- If the flux rope breakes through the surface, the field lines will bend outward due to the reduction of the outside gas pressure (recombination of hydrogen, no free electrons). The inclination of its boundary field to the field-free surroundings depends on the total flux contained in the flux rope. The larger the amount of flux, the larger will be the inclination in subsurface layers, where the atmosphere is opaque. At the solar surface, the radius of the flux concentration would increase, but it should still show a uniform intensity.

- If the inclination exceeds a critical value, the vertical energy transport by convection below the surface is so reduced that a local energy excess is created just below the boundary field line. The energy excess heats up the outermost layer of the flux concentration, which gets buoyant and rises. It is able to push its way up, because the other field lines are aligned radially. An azimuthally aligned horizontal field would strongly impede the rise of the flux tube.

- The buoyant flux tube reaches the solar surface with the properties as described by the MTM simulations of Schlichenmaier et al. (1998). A hot upstream point is seen at the intersection of the tube with the surface, which migrates inward towards the umbral core. The intensity of the flux tube would be above the umbral intensity, because it comes from deeper layers and is hotter, and below the granular intensity, because below the surface the heat transport by radiation is less effective than convective motions.

- The field lines of the flux tube are gradually pulled away from the outer spot boundary during the rise (cf. Fig. 6.1). The part of the flux tube, which below the surface still is close to or at the boundary layer, still

is heated. If the distance to the boundary field line is large enough, the energy supply from outside is cut off and the flux tube cools by radiation. For the MTM simulations, this point is never reached, as the flux tube is *kept fixed* to the outer spot boundary at the lower border of the simulation domain.

- After the rise of the outermost layer, the next layer of the flux concentration gets immediately heated, until it also rises and follows the previous flux tube. Thus, flux is constantly moved towards the umbra. The field strength, and hence, the magnetic pressure increase and push field lines outwards, until horizontal pressure balance is reached again. The flux at the outer spot boundary is thus replaced from the background field and the process can continue indefinitely. It is not necessary that this replacement of flux has the form of submerging flux tubes.

- The rate of flux exchange will depend on the amount of heat deposited from outside. The larger the inclination of the outer boundary field line, the stronger is the reduction of convective energy transport and the more heat is transferred by radiation into the outermost flux layer of the spot.

In this picture, the dynamical evolution of the penumbra and its small-scale fine structure will be due to the rising flux tubes. The inclination of the background field will determine the rate and evolution of the flux tubes. As the background field structure presumably is more stable, long-lived filaments will exist as observed in magnetograms. The derivation of the 3-D topology not for a single snapshot, but for a time series of observations, may be able to answer the question, if the penumbral fine structure is only due to the flux tube properties and how far the process is influenced by the background field structure.

The rise of the flux tubes inside the sunspot is triggered by the local heat excess at the sunspot boundary due to the reduction of the vertical energy transport by the inclined field lines. Another consequence of the excess energy below the inclined field lines of the sunspot deals with the moat flow, a radial outflow away from the spot, but outside the outer penumbral boundary. Opposite to the fast rise of the flux tubes inside the penumbra, the large-scale flow pattern of the moat develops slower, during around one day after the sunspot appearance (Pardon et al. 1979b). Gizon et al. (2000) found the moat flow to be present some Mm below the surface using helioseismology. This would be in agreement with a gradual build-up of a pressure excess in subsurface layers. The vertical energy flux cannot be completely transferred into the sunspot: its surface brightness is smaller than that of granulation. The excess energy may be the driver of the moat flow: the trapped energy below the inclined fields leads to a pressure excess below the penumbra, which will be able to drive material away from the spot.

However, also another explanation is possible for the moat flow: the moat flow may be the continuation of the Evershed effect outside the spot. The observations presented in this work, Westendorp Plaza et al. (2001b), or the numerical simulations of Thomas & Weiss (2004) show that part of the field lines, which harbor the Evershed flow, are bending downwards in the mid to outer penumbra. What happens to them outside of the spot ? This interesting question is in the moment studied on the base of the data, which were also used in the present work.

To conclude, it seems that the true nature of the penumbral fine structure can only be determined, if the properties of the horizontal Evershed channels can isolated from the more static background field. At present, only the analysis of spectro-polarimetric observations by multi-component ($\geq$2-C) inversions allows to separate the actual flux tube properties from the background field structure. The derivation of the 3-D topology from the Gaussian inversion or the integration of the inclination looks like a promising tool to follow the dynamical evolution of both, the physical quantities and the geometry. Relating the dynamical changes to the properties of the background field, maybe also the parameters and properties, which control the dynamical evolution, can be determined.

6.3 The surroundings of a sunspot: The traces of decay

If the picture of a sunspot and its internal dynamics drawn in the last section is taken as consistent and reasonable, it implies that the amount of magnetic flux is not reduced by the interchange convection as the rising flux tubes are "recycled". The final question left is then: "Why does a sunspot disappear ?" The answer may be contained in Sect. 4, which describes the properties of magnetic flux in the surroundings of a sunspot. The surroundings are only field-free to first order; a large number of isolated patches containing magnetic flux are present. On large scale, they are swept away from the spot by the moat flow, while following a course enforced by the surrounding granulation on small scales. Some of them are presumably related to penumbral extensions into the quiet Sun.

The time series of polarimetric measurements described in Sect. 4 shows one event of magnetic reconnection (cf. Sect. 4.8.3). This reconnection event takes place between a patch of magnetic flux, which has originated from

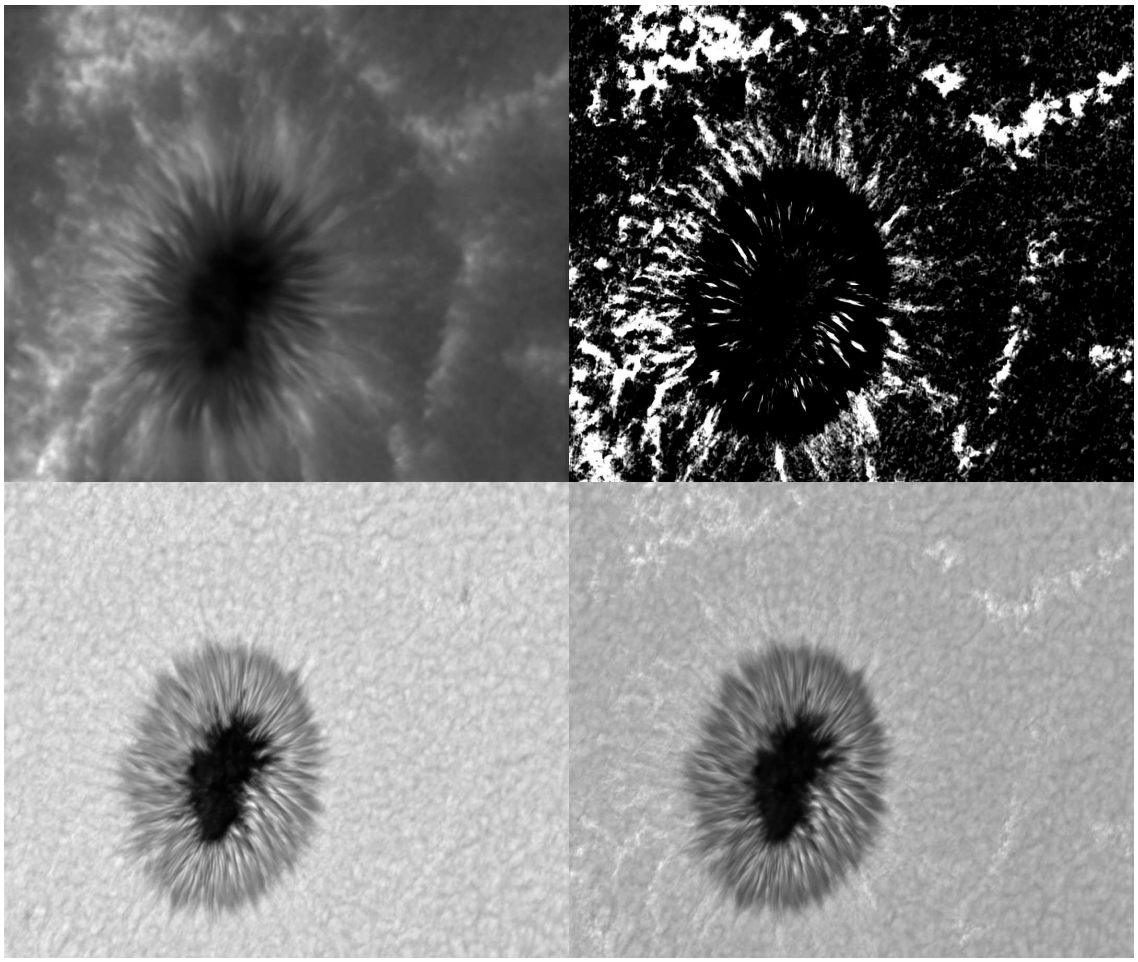


Figure 6.3: Average intensity of the 2 1/2 hours time series taken with the Dutch Open Telescope on 9.8.2003 between 8:25 and 11:58 UT. Clockwise, starting left top: Ca II H line core intensity, difference image of G-band and blue continuum, G-band intensity, blue continuum intensity. The outer boundary of the moat flow is marked by intense network elements, built from the flux swept to that limit, forming a circle of twice the spot radius. The traces of the horizontal motions of the flux with time, visible as enhanced intensity outside the spot in all but the blue continuum image, show a very distinct pattern around the sunspot. They are oriented radially, but appear only at some azimuthal angles around the spot center, building a (persistent) filamentary structure.

close to the outer penumbral boundary, and a network element at the boundary of the moat flow. During the event, a flux amount of $2{\cdot}10^{18}$ Mx is removed from the solar atmosphere, i.e. it is de facto removed from the sunspot. The observed flux loss rate of a sunspot can be achieved by 50 of such events per day (Martínez Pillet 2002; Bellot Rubio & Beck 2005). This flux removal from the spot is a common feature: in the surroundings of sunspots the so-called "Moving Magnetic Features" (MMFs) have been regularly observed (e.g., Harvey & Harvey 1973; Lee 1992; Yurchyshyn et al. 2001; Zhang et al. 2003), especially in the case of decaying sunspots.

They usually consist of one or two patches (with opposite polarity) of magnetic flux that appear at the outer penumbral boundary and are swept away from the spot by the moat flow. Most of the models proposed for the generation of MMFs assume that they are due to field lines which are still connected to the spot (e.g., the three proposed models in Lee 1992). For the decay of sunspots, only the uni-polar MMFs with the same polarity as the spot are of relevance. They are presumably created by field lines at the outer penumbral boundary, which can detach from the spot – or rather be detached by the granular motions. A preference for following the same path as earlier loosened flux can be seen in Fig. 6.3, which displays the time-averaged intensity of the observations taken at the Dutch Open Telescope. It can clearly be seen that in the average images the brightenings in G-band or the chromospheric Ca II H line core, which indicate flux concentrations, are not found everywhere around the spot,

but only at some azimuth angles. Regardless of the question, if these paths were created by flux still connected with the spot or not, they indicate that the sunspot field geometry, especially of the less dynamic background field, is not identical all around the spot. The same was also seen in the 3-D spot model from the observations in the azimuthal variation of the background field geometry.

Again, it would be worthwhile to compare the 3-D structure with the location of the radial paths of the flux parting from the spot, but this also has to be postponed to further studies.

Bibliography

Amer, M. A. & Kneer, F. 1993, A&A, 273, 304

Auer, L. H., House, L. L., & Heasley, J. N. 1977, Sol. Phys., 55, 47

Babcock, H. W. 1953, ApJ, 118, 387

Ballesteros, E., Collados, M., Bonet, J. A., et al. 1996, A&AS, 115, 353

Balthasar, H. & Collados, M. 2005, A&A, 429, 705

Beck, C. 2002, Diploma thesis, Univ. Freiburg: The calibration of the vectorpolarimeter POLIS, http://www.kis.uni-freiburg.de/cbeck_diplom.pdf

Beck, C., Schlichenmaier, R., Collados, M., Bellot Rubio, L., & Kentischer, T. 2005a, A&A, 443, 1047

Beck, C., Schmidt, W., Kentischer, T., & Elmore, D. 2005b, A&A, 437, 1159

Bellot Rubio, L. & Beck, C. 2005, ApJ, 626, L125

Bellot Rubio, L. R. 2002, in ESA SP-505: SOLMAG 2002. Proceedings of the Magnetic Coupling of the Solar Atmosphere Euroconference, 3

Bellot Rubio, L. R. 2003, in Astronomical Society of the Pacific Conference Series, 301–+

Bellot Rubio, L. R., Balthasar, H., & Collados, M. 2004, A&A, 427, 319

Bellot Rubio, L. R., Balthasar, H., Collados, M., & Schlichenmaier, R. 2003, A&A, 403, L47

Bellot Rubio, L. R. & Collados, M. 2003, A&A, 406, 357

Bellot Rubio, L. R., Ruiz Cobo, B., & Collados, M. 1996, A&A, 306, 960

—. 2000a, ApJ, 535, 475

—. 2000b, ApJ, 535, 489

Berger, T. E., Rouppe van der Voort, L. H. M., Löfdahl, M. G., et al. 2004a, A&A, 428, 613

Berger, T. E., Rouppe van der Voort, L. H. M., Lofdahl, M. G., et al. 2004b, American Astronomical Society Meeting, 204

Berger, T. E. & Title, A. M. 2001, ApJ, 553, 449

Borrero, J. M. & Bellot Rubio, L. R. 2002, A&A, 385, 1056

Borrero, J. M., Lagg, A., Solanki, S. K., & Collados, M. 2005, A&A, 436, 333

Borrero, J. M., Solanki, S. K., Bellot Rubio, L. R., Lagg, A., & Mathew, S. K. 2004, A&A, 422, 1093

Bovelet, B. & Wiehr, E. 2003, A&A, 412, 249

Brickhouse, N. S. & Labonte, B. J. 1988, Sol. Phys., 115, 43

Cabrera Solana, D., Bellot Rubio, L. R., & del Toro Iniesta, J. C. 2005, A&A, 439, 687

Caligari, P., Schuessler, M., & Moreno-Insertis, F. 1998, ApJ, 502, 481

Capitani, C., Cavallini, F., Ceppatelli, G., et al. 1989, Sol. Phys., 120, 173

Carlsson, M., Stein, R. F., Nordlund, Å., & Scharmer, G. B. 2004, ApJ, 610, L137

Collett, E. 1992, Polarized light. Fundamentals and applications (Optical Engineering, New York: Dekker)

de Jager, C. & Vermue, J. 1977, Sol. Phys., 54, 313

Degenhardt, D. & Lites, B. W. 1993, ApJ, 416, 875

Delbouille, L., Roland, G., & Neven, L. 1973, Atlas photometrique DU spectre solaire de [lambda] 3000 a [lambda] 10000 (Liege: Universite de Liege, Institut d'Astrophysique, 1973)

Dodson, H. W. & Owren, L. 1952, AJ, 57, 10

Domínguez Cerdeña, I., Sánchez Almeida, J., & Kneer, F. 2003, A&A, 407, 741

Dunér, N. C. 1890, Astronomische Nachrichten, 124, 267

—. 1905, Astronomische Nachrichten, 167, 167

Dunn, R. B. & Zirker, J. B. 1973, Sol. Phys., 33, 281

Evershed, J. 1909, MNRAS, 69, 454

Falewicz, R. & Rudawy, P. 1999, A&A, 344, 981

Frazier, E. N. & Stenflo, J. O. 1972, Sol. Phys., 27, 330

Frutiger, C. 2000, Ph.D. Thesis

Frutiger, C., Solanki, S. K., Fligge, M., & Bruls, J. H. M. J. 1999, in ASSL Vol. 243: Polarization, 281–290

Galilei, G. 1632, Dialogo DI Galileo Galilei Linceo matematico spraordinario dello stvdio DI Pisa. (Fiorenza, Per Gio: Batista Landini, 1632)

Galilei, G., Welser, M., & de Filiis, A. 1613, Istoria E dimostrazioni intorno alle macchie solari E loro accidenti comprese in tre lettere scritte all'illvstrissimo signor Marco Velseri ... dal signor Galileo Galilei ... SI aggiungono nel fine le le ttere, E disquisizioni del finto Apelle [Christoph Scheiner] (Roma, G. Mascadi, 1613)

Gingerich, O., Noyes, R. W., Kalkofen, W., & Cuny, Y. 1971, Sol. Phys., 18, 347

Gizon, L., Duvall, T. L., & Larsen, R. M. 2000, Journal of Astrophysics and Astronomy, 21, 339

Grossmann-Doerth, U., Schuessler, M., & Steiner, O. 1998, A&A, 337, 928

Hale, G. E. 1902, ApJ, 16, 211

—. 1908, ApJ, 28, 315

Hanaoka, Y. 1997, Sol. Phys., 173, 319

Harvey, K. & Harvey, J. 1973, Sol. Phys., 28, 61

Hirzberger, J. 2003, A&A, 405, 331

Hirzberger, J. & Wiehr, E. 2005, A&A, 438, 1059

Hoyt, D. V. & Schatten, K. H. 1998, Sol. Phys., 181, 491

Illing, R. M. E., Landman, D. A., & Mickey, D. L. 1975, A&A, 41, 183

Jahn, K. & Schmidt, H. U. 1994, A&A, 290, 295

Jefferies, J., Lites, B. W., & Skumanich, A. 1989, ApJ, 343, 920

Keil, S. L., Balasubramaniam, K. S., Smaldone, L. A., & Reger, B. 1999, ApJ, 510, 422

Khomenko, E. V., Collados, M., Solanki, S. K., Lagg, A., & Trujillo Bueno, J. 2003, A&A, 408, 1115

Khomenko, E. V., Martínez González, M. J., Collados, M., et al. 2005, A&A, 436, L27

Knoelker, M. & Schuessler, M. 1988, A&A, 202, 275

López Ariste, A. 2002, ApJ, 564, 379

Landi Degl'Innocenti, E. 2003, Astronomische Nachrichten, 324, 393

Landi degl'Innocenti, E. & Landi degl'Innocenti, M. 1972, Sol. Phys., 27, 319

Landi Deglinnocenti, E. & Landi Deglinnocenti, M. 1977, A&A, 56, 111

Landolfi, M. & Landi degl'Innocenti, E. 1996, Sol. Phys., 164, 191

Langhans, K., Scharmer, G. B., Kiselman, D., Löfdahl, M. G., & Berger, T. E. 2005, A&A, 436, 1087

Langhans, K., Schmidt, W., & Rimmele, T. 2004, A&A, 423, 1147

Langhans, K., Schmidt, W., & Tritschler, A. 2002, A&A, 394, 1069

Lee, J. W. 1992, Sol. Phys., 139, 267

Leka, K. D. 1997, ApJ, 484, 900

Leka, K. D. & Skumanich, A. 1998, ApJ, 507, 454

Linton, M. G., Dahlburg, R. B., & Antiochos, S. K. 2002, Bulletin of the American Astronomical Society, 34, 789

Lites, B. W., Bida, T. A., Johannesson, A., & Scharmer, G. B. 1991, ApJ, 373, 683

Lites, B. W., Elmore, D. F., Seagraves, P., & Skumanich, A. P. 1993, ApJ, 418, 928

Livingston, W. 2002, Sol. Phys., 207, 41

Müller, D. A. N., Schlichenmaier, R., Steiner, O., & Stix, M. 2002, A&A, 393, 305

Martínez Pillet, V., Collados, M., Sánchez Almeida, J., et al. 1999, in ASP Conf. Ser. 183: High Resolution Solar Physics: Theory, Observations, and Techniques, 264

Martínez Pillet, V. 2002, Astronomische Nachrichten, 323, 342

Mathew, S. K., Lagg, A., Solanki, S. K., et al. 2003, A&A, 410, 695

McIntosh, P. S. 1981, in The Physics of Sunspots, 7–54

McLean, D. J. & Dulk, G. A. 1978, Proceedings of the Astronomical Society of Australia, 3, 251

Melrose, D. B. 1987, Proceedings of the Astronomical Society of Australia, 7, 6

Muglach, K. & Solanki, S. K. 1992, A&A, 263, 301

Muller, R. 1973, Sol. Phys., 29, 55

—. 1976, Sol. Phys., 48, 101

—. 1983, Sol. Phys., 85, 113

Nave, G., Johansson, S., Learner, R. C. M., Thorne, A. P., & Brault, J. W. 1994, ApJS, 94, 221

Pardon, L., Worden, S. P., & Schneeberger, T. J. 1979a, Sol. Phys., 63, 247

—. 1979b, Sol. Phys., 63, 247

Parker, E. N. 1979, ApJ, 234, 333

Pierce, A. K. & Breckenridge, J. B. 1974, The Kitt Peak table of photographic solar spectrum wavelengths (Kitt Peak National Observatory Contribution, Tucson: Kitt Peak National Observatory, 1973-1974)

Press, W. H., Flannery, B. P., & Teukolsky, S. A. 1986, Numerical recipes. The art of scientific computing (Cambridge: University Press, 1986)

Rees, D. E., Durrant, C. J., & Murphy, G. A. 1989, ApJ, 339, 1093

Rouppe van der Voort, L. H. M., Löfdahl, M. G., Kiselman, D., & Scharmer, G. B. 2004, A&A, 414, 717

Rucklidge, A. M., Schmidt, H. U., & Weiss, N. O. 1995, MNRAS, 273, 491

Ruiz Cobo, B. 1998, Ap&SS, 263, 331

Ruiz Cobo, B. & del Toro Iniesta, J. C. 1992, ApJ, 398, 375

Rutten, R. J., Hammerschlag, R. H., Bettonvil, F. C. M., Sütterlin, P., & de Wijn, A. G. 2004, A&A, 413, 1183

Sánchez Almeida, J., Domínguez Cerdeña, I., & Kneer, F. 2003, ApJ, 597, L177

Sütterlin, P., Bellot Rubio, L. R., & Schlichenmaier, R. 2004, A&A, 424, 1049

Sainz Dalda, A. & Martínez Pillet, V. 2005, ApJ, nn

Sakurai, T., Shibata, K., Ichimoto, K., Tsuneta, S., & Acton, L. W. 1992, PASJ, 44, L123

Sanchez Almeida, J. 1992, Sol. Phys., 137, 1

Sanchez Almeida, J., Collados, M., & del Toro Iniesta, J. C. 1989, A&A, 222, 311

Sankarasubramanian, K. & Rimmele, T. 2003, ApJ, 598, 689

Schüssler, M., Shelyag, S., Berdyugina, S., Vögler, A., & Solanki, S. K. 2003, ApJ, 597, L173

Scharmer, G. B., Gudiksen, B. V., Kiselman, D., Löfdahl, M. G., & Rouppe van der Voort, L. H. M. 2002, Nature, 420, 151

Schlichenmaier, R., Bellot Rubio, L. R., & Tritschler, A. 2005, Astronomische Nachrichten, 326, 301

Schlichenmaier, R., Bruls, J. H. M. J., & Schüssler, M. 1999, A&A, 349, 961

Schlichenmaier, R., Jahn, K., & Schmidt, H. U. 1998, A&A, 337, 897

Schlichenmaier, R. & Schmidt, W. 2000, A&A, 358, 1122

Schlichenmaier, R. & Solanki, S. K. 2003, A&A, 411, 257

Schmidt, W. & Balthasar, H. 1994, A&A, 283, 241

Schmidt, W., Beck, C., Kentischer, T., Elmore, D., & Lites, B. 2003, Astronomische Nachrichten, 324, 300

Schmidt, W. & Kentischer, T. 1995, A&AS, 113, 363

Schmidt, W., Kentischer, T. J., Bruls, J., & Lites, B. W. 2001, in ASP Conf. Ser. 236: Advanced Solar Polarimetry – Theory, Observation, and Instrumentation, 49

Schröter, E. H. 1967, in Solar Physics, 325–+

Schussler, M., Caligari, P., Ferriz-Mas, A., & Moreno-Insertis, F. 1994, A&A, 281, L69

Sheeley, N. R. 1969, Sol. Phys., 9, 347

Shine, R., Title, A., & Murdin, P. 2001, Encyclopedia of Astronomy and Astrophysics

Shurcliff, W. A. 1966, Polarized light. Production and use (Cambridge, Mass.: Harvard University Press)

Sigwarth, M., Balasubramaniam, K. S., Knölker, M., & Schmidt, W. 1999, A&A, 349, 941

Skumanich, A. & Lites, B. W. 1987, ApJ, 322, 473

Skumanich, A., Lites, B. W., Martínez Pillet, V., & Seagraves, P. 1997, ApJS, 110, 357

Sobotka, M., Bonet, J. A., Vazquez, M., & Hanslmeier, A. 1995, ApJ, 447, L133+

Sobotka, M. & Sütterlin, P. 2001, A&A, 380, 714

Socas-Navarro, H. & Sánchez Almeida, J. 2003, ApJ, 593, 581

Solanki, S. K. 2003, A&A Rev., 11, 153

Solanki, S. K. & Montavon, C. A. P. 1993, A&A, 275, 283

Solanki, S. K. & Pahlke, K. D. 1988, A&A, 201, 143

Solanki, S. K., Pantellini, F. G. E., & Stenflo, J. O. 1987, Sol. Phys., 107, 57

Solanki, S. K. & Rüedi, I. 2003, A&A, 411, 249

Solanki, S. K. & Schmidt, H. U. 1993, A&A, 267, 287

Spruit, H. & Scharmer, G. 2005, A&A, tbd

Spruit, H. C. 1979, Sol. Phys., 61, 363

Steiner, O., Grossmann-Doerth, U., Knoelker, M., & Schuessler, M. 1998, ApJ, 495, 468

Steiner, O., Hauschildt, P. H., & Bruls, J. 2001, A&A, 372, L13

Stix, M. 2002, The Sun. An Introduction (Springer-Verlag Berlin Heidelberg, 2nd edition)

Sütterlin, P., Schroter, E. H., & Muglach, K. 1996, Sol. Phys., 164, 311

Thomas, J. H. & Montesinos, B. 1993, ApJ, 407, 398

Thomas, J. H. & Weiss, N. O. 2004, ARA&A, 42, 517

Title, A. M., Frank, Z. A., Shine, R. A., et al. 1993, ApJ, 403, 780

Toenjes, K. & Woehl, H. 1982, Sol. Phys., 75, 63

Tomczyk, S., Elmore, D. F., Lites, B. W., et al. 1992, Bulletin of the American Astronomical Society, 24, 814

Tritschler, A., Schmidt, W., & Rimmele, T. 2002, in ESA SP-506: Solar Variability: From Core to Outer Frontiers, 477–478

Tritschler, A., Schmidt, W., & Rimmele, T. 2003, Astronomische Nachrichten Supplement, 324, 54

van Ballegooijen, A. A., Nisenson, P., Noyes, R. W., et al. 1998, ApJ, 509, 435

von der Lühe, O., Soltau, D., Berkefeld, T., & Schelenz, T. 2003, in Innovative Telescopes and Instrumentation for Solar Astrophysics. Edited by Stephen L. Keil, Sergey V. Avakyan . Proceedings of the SPIE, Volume 4853, 187

Weiss, N. O., Thomas, J. H., Brummell, N. H., & Tobias, S. M. 2004, ApJ, 600, 1073

Westendorp Plaza, C., del Toro Iniesta, J. C., Ruiz Cobo, B., et al. 1998, ApJ, 494, 453

Westendorp Plaza, C., del Toro Iniesta, J. C., Ruiz Cobo, B., & Pillet, V. M. 2001a, ApJ, 547, 1148

Westendorp Plaza, C., del Toro Iniesta, J. C., Ruiz Cobo, B., et al. 2001b, ApJ, 547, 1130

Wiehr, E., Bovelet, B., & Hirzberger, J. 2004, A&A, 422, L63

Wittmann, A. D. & Xu, Z. T. 1987, A&AS, 70, 83

Yang, G., Xu, Y., Wang, H., & Denker, C. 2003, ApJ, 597, 1190

Young, C. A. 1874, Astronomical register, 12, 211

Yurchyshyn, V. B., Wang, H., & Goode, P. R. 2001, ApJ, 550, 470

Zhang, J., Solanki, S. K., & Wang, J. 2003, A&A, 399, 755

Appendix A

Mueller matrices

The Mueller matrices describe the allowed transformations inside the Stokes "vector space". A formal derivation is given in, e.g., Shurcliff (1966) or Collett (1992).

A.1 The polarizer

A polarizer is an optical element with a transmission dependent on the oscillation direction of the electrical field and can be described by

$$E'_x = p_x \cdot E_x$$
$$E'_y = p_y \cdot E_y \,,$$

with $0 \leq p_x, p_y \leq 1$.
The Mueller matrix of a linear polarizer can be derived to be

$$\mathbf{M}_{\text{pol}} = \frac{1}{2} \begin{pmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2 \cdot p_x \cdot p_y & 0 \\ 0 & 0 & 0 & 2 \cdot p_x \cdot p_y \end{pmatrix} . \tag{A.1}$$

For an ideal polarizer p_y is equal to 0 and p_x equal to 1. High quality linear polarizers can reach $p_y = 10^{-4}$.

A.2 The retarder

A retarder has different propagation velocities for the two oscillation directions (x,y). It changes the relative phase between the field components. Here is

$$E'_x = e^{+i\Phi/2} \cdot E_x$$
$$E'_y = e^{-i\Phi/2} \cdot E_y \,,$$

and the Mueller matrix resulting is

$$\mathbf{M}_{\text{ret}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\Phi & \sin\Phi \\ 0 & 0 & -\sin\Phi & \cos\Phi \end{pmatrix} . \tag{A.2}$$

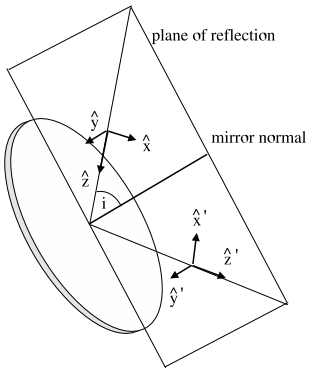


Figure A.1: Sketch of reflection on a flat mirror. The incoming ray is described in the right-handed orthogonal coordinate system given by $\{\hat{x},\hat{y},\hat{z}\}$, the reflected beam in the primed one. $\hat{x}$ and $\hat{x}'$ lie in the incidence-reflection plane, while $\hat{y}$ and $\hat{y}'$ are parallel to the mirror surface. i denotes the incidence angle.

A.3 Rotated optical elements

A rotator rotates (as its name says) the field components by a fixed angle θ. Here is

$$\begin{aligned} E'_x &= \;\; E_x \cdot \cos\theta + E_y \cdot \sin\theta \\ E'_y &= -E_x \cdot \sin\theta + E_y \cdot \cos\theta \,, \end{aligned}$$

and the matrix is

$$\mathbf{M}_{\text{rot}}(2\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\,\theta & \sin 2\,\theta & 0 \\ 0 & -\sin 2\,\theta & \cos 2\,\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{A.3}$$

The Mueller matrix of the rotator is especially important as it also allows one to determine the effects of rotated optical elements, which often occur in optical designs. The rotation can be included by two rotator matrices around the optical element itself by

$$\mathbf{M}_{\text{elem}}(\theta) = \mathbf{M}_{\text{rot}}(-2\theta) \cdot \mathbf{M}_{\text{elem}} \cdot \mathbf{M}_{\text{rot}}(2\theta) \,.$$

One example is the rotating retarder used by either the ASP or POLIS :

$$\mathbf{M}_{\text{ret}}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & c^2 + s^2 d & sc(1-d) & -se \\ 0 & sc(1-d) & s^2 + c^2 d & ce \\ 0 & se & -ce & d \end{pmatrix} , \tag{A.4}$$

with $c = \cos 2\theta, s = \sin 2\theta, d = \cos\delta$ and $e = \sin\delta$. δ is the retardance of the modulator and θ the angle between its fast axis and the x-direction, as always counter-clockwise when looking at the light source[1].

A.4 Reflection on a mirror

Figure A.1 sketches the reflection on a mirror: The Stokes vector of the incoming beam defined in the reference frame $\{\hat{x},\hat{y},\hat{z}\}$ is transformed into the Stokes vector of the outgoing beam in the reference system $\{\hat{x}',\hat{y}',\hat{z}'\}$. The two reference systems are defined such that $\hat{z}$ is pointing into the direction of propagation, $\hat{x}$ is in the incidence

[1] I really like that definition.

plane and is directed towards the mirror normal, and $\hat{y}$ is parallel to the mirror plane. The corresponding Mueller matrix which transforms the Stokes vector from basis $\{\hat{x},\hat{y},\hat{z}\}$ to $\{\hat{x}',\hat{y}',\hat{z}'\}$ is given by (cf. Skumanich et al. 1997):

$$\mathcal{M}(i_{\rm M},n,k) = \begin{pmatrix} \frac{(r_x+r_y)}{2} & \frac{(r_x-r_y)}{2} & 0 & 0 \\ \frac{(r_x-r_y)}{2} & \frac{(r_x+r_y)}{2} & 0 & 0 \\ 0 & 0 & \sqrt{r_x r_y}\cos\delta & \sqrt{r_x r_y}\sin\delta \\ 0 & 0 & -\sqrt{r_x r_y}\sin\delta & \sqrt{r_x r_y}\cos\delta \end{pmatrix}, \tag{A.5}$$

where r_x and r_y are the intensity reflection coefficients, and δ is the phase change, induced by the reflection, between the electric field components in the $\hat{x}'$ and $\hat{y}'$ direction. Under the assumption that the aluminum coatings are sufficiently thick such that the reflection is described as a one-layer reflection[2], the three parameters – r_x, r_y, and δ – are related to the incidence angle, $i_{\rm M}$, and two material properties: the index of refraction, n_λ, and the index of extinction, k_λ. Then, suppressing the wavelength subscript for n_λ and k_λ for simplicity, the relations are given by

$$r_\parallel = \frac{(n+{\rm i}k)^2\cos i_{\rm M} - u - {\rm i}v}{(n+{\rm i}k)^2\cos i_{\rm M} + u + {\rm i}v}, \tag{A.6}$$

$$r_\perp = \frac{\cos i_{\rm M} - u - {\rm i}v}{\cos i_{\rm M} + u + {\rm i}v}, \tag{A.7}$$

where u and v are calculated by

$$u = \sqrt{\tfrac{1}{2}\left[c + \sqrt{c^2+4n^2k^2}\right]}, \tag{A.8}$$

$$v = \sqrt{\tfrac{1}{2}\left[-c + \sqrt{c^2+4n^2k^2}\right]}. \tag{A.9}$$

with $c = n^2 - k^2 - \sin^2 i_{\rm M}$. With $r_\parallel = \sqrt{r_x}e^{{\rm i}\varepsilon_x}$, $r_\perp = \sqrt{r_y}e^{{\rm i}\varepsilon_y}$, $\sqrt{r_x r_y}\cos\delta = {\rm Re}(r_\parallel r_\perp^*)$, and $\sqrt{r_x r_y}\sin\delta = -{\rm Im}(r_\parallel r_\perp^*)$ the Mueller matrix parameters are defined.

[2]This corresponds to a value of $t = 0$ in Eq. (A1) in Skumanich et al. (1997).

Appendix B

Radiative transfer in the presence of magnetic fields

The radiation transfer equation for the Stokes vector **S** in the solar atmosphere is given by

$$\cos\theta \frac{d\mathbf{S}_\nu(\tau)}{d\tau} = \mathcal{M} \cdot \left(\mathbf{S}_\nu(\tau) - (B_\nu(\tau),0,0,0)^{\mathrm{T}}\right) \ . \tag{B.1}$$

The total absorption matrix $\mathcal{M}$ is given by

$$\kappa_c \cdot \mathcal{M} = \kappa_0 \cdot \eta + \kappa_c \cdot \mathbb{1} = \begin{pmatrix} \eta_I & \eta_Q & \eta_U & \eta_V \\ \eta_Q & \eta_I & \rho_V & -\rho_U \\ \eta_U & -\rho_V & \eta_I & \rho_Q \\ \eta_V & \rho_U & -\rho_Q & \eta_I \end{pmatrix} + \kappa_c \cdot \mathbb{1} \ , \tag{B.2}$$

The entries of η can be derived from the properties of the electric transition in the presence of magnetic fields (Landi degl'Innocenti & Landi degl'Innocenti 1972). They are given by

$$\eta_I = \frac{1}{2}\left[\eta_p \sin^2\gamma + \frac{1}{2}(\eta_r+\eta_b)(1+\cos^2\gamma)\right] \tag{B.3}$$

$$\eta_Q = \frac{1}{2}\left[\eta_p - \frac{1}{2}(\eta_r+\eta_b)\right]\sin^2\gamma\cos 2\phi \tag{B.4}$$

$$\eta_U = \frac{1}{2}\left[\eta_p - \frac{1}{2}(\eta_r+\eta_b)\right]\sin^2\gamma\sin 2\phi \tag{B.5}$$

$$\eta_V = \frac{1}{2}\left[\eta_p - \frac{1}{2}(\eta_r+\eta_b)\right]\sin^2\gamma\cos 2\phi \tag{B.6}$$

$$\rho_Q = \frac{1}{2}\left[\rho_p - \frac{1}{2}(\rho_r+\rho_b)\right]\sin^2\gamma\cos 2\phi \tag{B.7}$$

$$\rho_Q = \frac{1}{2}\left[\rho_p - \frac{1}{2}(\rho_r+\rho_b)\right]\sin^2\gamma\sin 2\phi \tag{B.8}$$

$$\rho_V = \frac{1}{2}\left[\rho_p - \rho_r\right]\cos\gamma \tag{B.9}$$

γ denotes the inclination between magnetic field and the line of sight (LOS), ϕ the field azimuth in the plane perpendicular to the LOS, $\eta_{p,b,r}$ and $\rho_{p,b,r}$ the general profile functions for absorption and anormal dispersion.

The profile functions are defined as the energy emission of a damped oscillator with a finite life time, Γ, that is additionally smeared out by the random distribution of the absorber velocities. The profile function is then given by the convolution of a Lorentz profile ($\propto \frac{\Gamma}{1+\Gamma^2}$) with a Gaussian ($\propto e^{-\nu^2}$):

$$\Phi(\nu) = \frac{\Gamma}{\sqrt{\pi}\Delta\nu_D}\int_{-\infty}^{+\infty} \frac{\exp\left[-(\nu-\nu')^2/\Delta\nu_D^2\right]}{(2\pi)^2(\nu'-\nu_0)^2+\Gamma^2/4} d\nu' \ , \tag{B.10}$$

with the thermally induced Doppler width, $\Delta\nu_D = \nu_0/c\sqrt{2kT/m}$, and the frequency of the transition, ν_0.
Using

$$y = (\nu - \nu')/\Delta\nu_D, \tag{B.11}$$
$$a = \Gamma 4\pi\Delta\nu_D, \text{ and} \tag{B.12}$$
$$v = (\nu - \nu_0)/\Delta\nu_D, \tag{B.13}$$

Eq. (B.10) can be written as

$$\Phi(\nu) = \frac{1}{\sqrt{\pi}\Delta\nu_D} H(v,a), \tag{B.14}$$

with the *Voigt function* $H(v,a)$ given by

$$H(v,a) = \frac{a}{\pi}\int_{-\infty}^{+\infty} \frac{e^{-y^2}dy}{(v-y)^2 + a^2}\,. \tag{B.15}$$

With the indices p for the π-component of a Zeeman triplet at the rest wavelength, b for the σ^--component, and r for the σ^+-component (cf. Sect. 3.1) the quantities $\eta_{p,b,r}$ and $\rho_{p,b,r}$ are given by

$$\eta_i = \sum w_i H(v - v_i, a) \tag{B.16}$$
$$\rho_i = \sum w_i F(v - v_i, a)\,, \tag{B.17}$$

with the *Faraday-Voigt function*

$$F(v,a) = \frac{a}{2\pi}\int_{-\infty}^{+\infty} \frac{(v-y)e^{y^2}dy}{(v-y)^2 + a^2}\,. \tag{B.18}$$

w_i are the statistical weights for the transition probability, and v_i is determined by the strength of the magnetic field by

$$v_i = v_0 \pm \frac{e}{4\pi m_e c^2}\frac{\lambda_0^2}{\Delta\lambda_D}\cdot B\,. \tag{B.19}$$

Doppler shifts due to the flow velocity in the atmosphere are added to v_i. The effect of the microturbulent velocity, v_{mic}, is added to the collision broadening coefficient, $\Gamma_{\rm tot} = \Gamma + \Gamma_{v_{mic}}$. The effect of the macroturbulent velocity, v_{mac}, is added to the thermal Doppler width.

Appendix C

Maps of the aligned VTT/DOT data

All maps of the aligned DOT and VTT data sets are displayed in Figs. C.1 and C.2. Shown are clockwise the DOT G-band image, the total integrated polarization, the difference intensity (DI) between G-band and blue continuum, and the TIP intensity in IR. For the first repetition, the mask of identified BPs in the VTT data is displayed instead of DI. Also for the first repetition the area used for the quiet Sun control sample is marked. A visual inspection shows that almost all features identified as G-band bright points in the difference image are co-spatial with polarization signal. A single polarization patch usually contains some BPs, and most patches show BPs at least in one of the repetitions. Most of the magnetic signals are present permanently during the 1 hour of observation time. This is in agreement with the interpretation that brightenings in the G-band are due to the presence of magnetic fields, where the high variability of the BPs is due to changes in the magnetic field configuration.

The canopy of the spot, which is located at the upper boundary of the FOV displayed, clearly extends beyond the white light boundary. It does not lead to a brightening in the G-band because the fields are rather horizontal. Rays passing through them do only encounter a short path length in an atmosphere with reduced density, and do not penetrate to deeper photospheric layers with higher temperatures at the end.

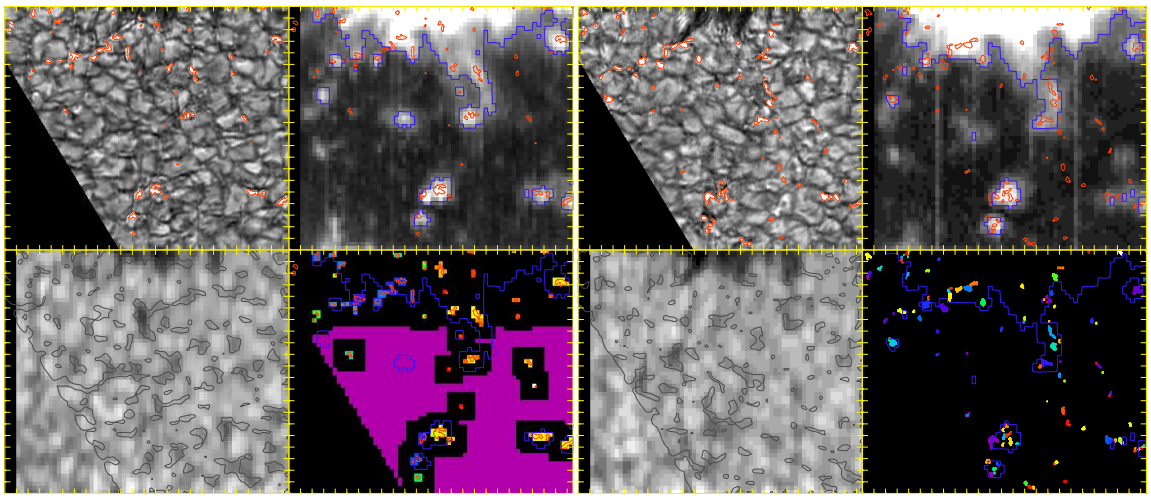


Figure C.1: **Clockwise, starting top left:** G-band intensity, total integrated polarization, VTT binary mask, IR intensity. Blue contours outline high polarization signal, red contours the selected BP areas. In the VTT binary mask the BPs are color coded according to their number. The uniform pink area marks the spatial points of the Quiet Sun control sample outside the canopy, and with 3 pixels ($\sim$ 1 arcsec) distance to any BP. For the other repetitions the DOT BP mask is displayed at this position.

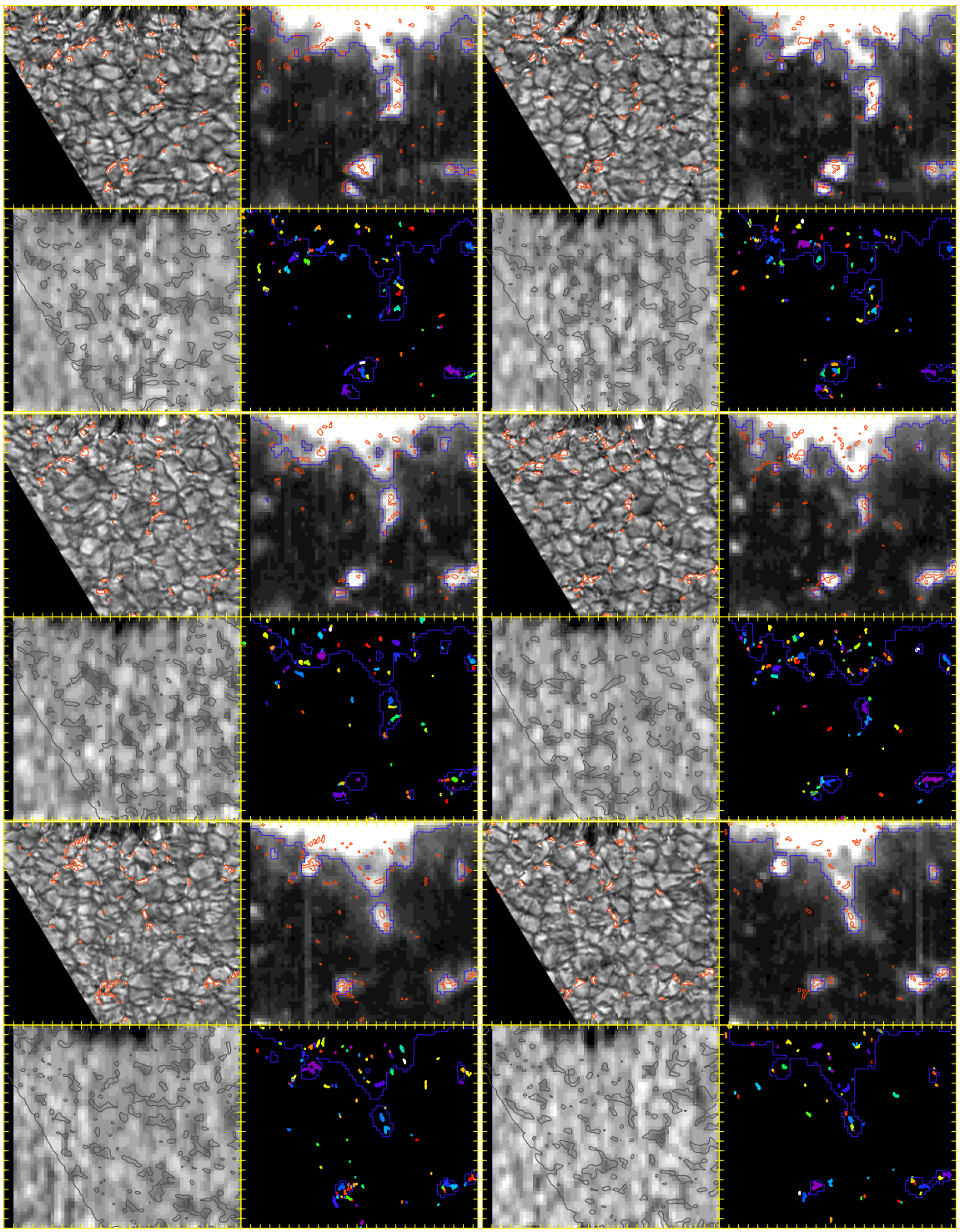


Figure C.2: Same as Fig. C.1 for the other repetitions of the scan.

Appendix D

Stokes profiles across an isolated Bright Point

The Stokes profiles along the cut through the Bright Point displayed in Fig. 4.23 are shown in Fig. D.1 for IR and Fig. D.2 for the visible lines, together with the best fit profile from the inversion. The inversion without gradients in physical parameters does not allow to match the asymmetries of the signal, but for the canopy field extending up to the 7th profile row the V signal is very regular. Only the two most magnetic sensitive lines, 1564.8 nm and 630.25 nm, show a considerable amount of linear polarization, mainly in Stokes Q, for the canopy field. The BP is characterized by an increase in field strength, almost no linear polarization signal, and definitely asymmetric Stokes V lobes, especially for the visible lines.

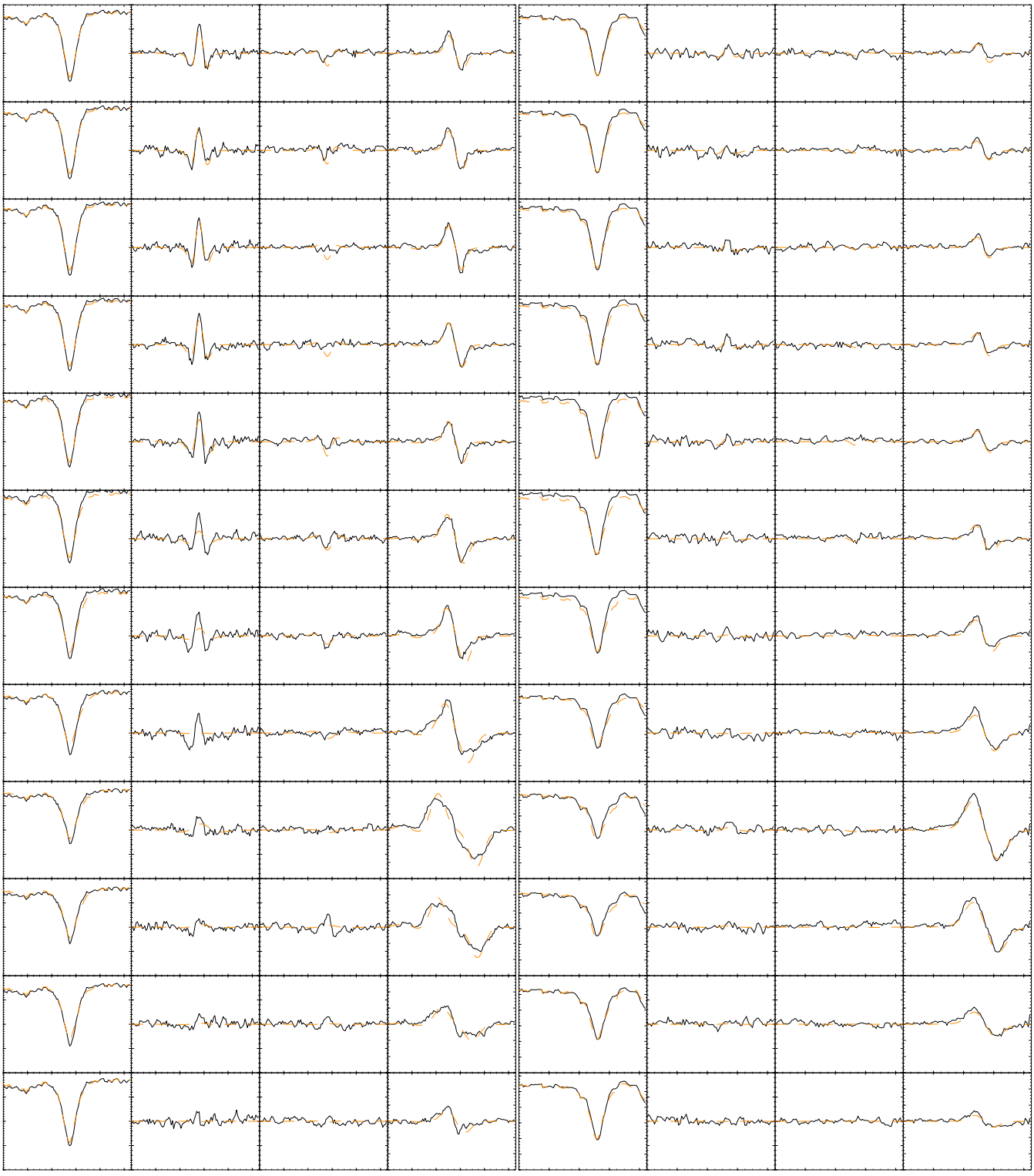


Figure D.1: Stokes profiles observed (**solid**) and best fit profile from the inversion (**dashed**) for the two infrared lines, from **left to right** IQU, and V for 1564.8 nm, and 1565.2 nm, respectively. The vertical dimension corresponds to the position of the pixel in the horizontal cut through a BP in Fig. 4.23. The range displayed is $\pm$ 1 % for the linear polarization and $\pm$1.5 % for V. The most magnetic sensitive line, 1564.8 nm, is the only one showing a very prominent linear polarization signal. It is most probably not connected to the BP at all, but the outermost edge of the sunspots' canopy. Note that the BP has a much larger field strength, and the linear polarization signal vanishes completely for the last four spatial positions. The co-spatial visible profiles are displayed in Fig. D.2.

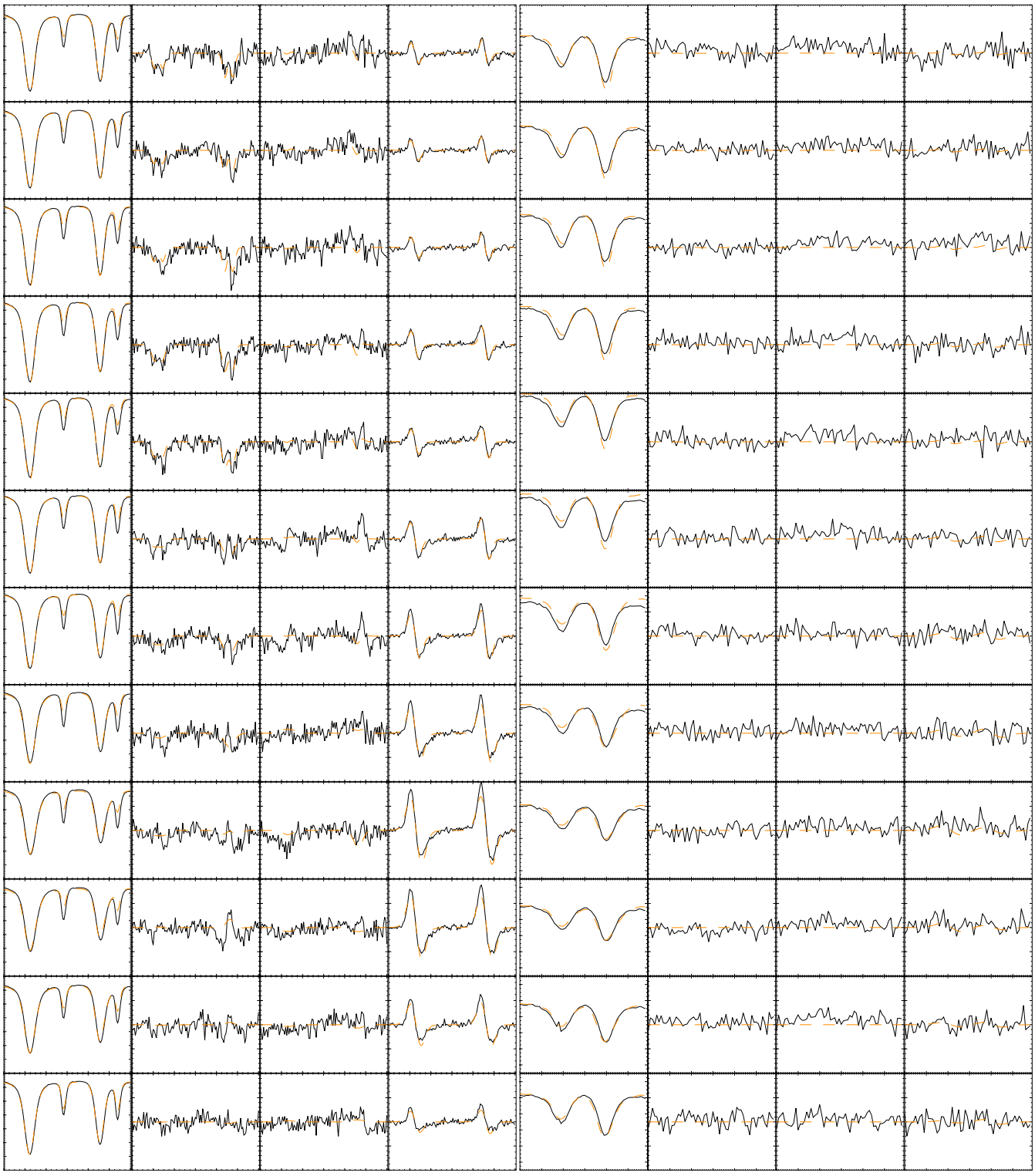


Figure D.2: Stokes profiles observed (**solid**) and best fit profile from the inversion (**dashed**) for the visible lines, from **left to right** IQU, and V for 630.15/25 nm, and 630.37 nm, respectively. The vertical dimension corresponds to the position of the pixel in the horizontal cut through a BP in Fig. 4.23. The range displayed is $\pm$ 1 % for the linear polarization and $\pm$4 % for V. The most magnetic sensitive line, 630.25 nm, is the only one showing a clear linear polarization signal. The inversion with constant atmospheric properties cannot reproduce the asymmetry of the Stokes V signal of the BP. The co-spatial IR profiles are displayed in Fig. D.1.

Acknowledgments

First of all, I want to thank my father and my mother for their neverending support. Without them, all this would not have been possible.

At the Kiepenheuer Institute for Solar Physics in Freiburg, I thank my supervisor, PD Dr. Wolfgang Schmidt, for the topic, support, advice, assistance, and helpful remarks in all scientific and technical and personal questions. I thank Rolf Schlichenmaier, Luis Ramon Bellot Rubio and Oskar Steiner for their advice in scientific topics and hours of discussion (and in some cases, drinking). In the technical and engineering section, I especially thank Thomas Kentischer for being allowed to call him on weekends, holidays, and/or sundays, when we got stuck again at the VTT ("Thomas, dein Motorenkäschtle geht nicht, der CAN-Bus hängt, die Kamera reagiert nicht mehr, der zweite Coelostatenspiegel ist am Endschalter, das Licht ist weg, das Spektrum kippt weg, die Spektrallinien sind schief und defokussiert, und meine Nerven sind am Ende"). Kentischer-Soft© is the backbone of most instrumentation at the VTT; I hope he never leaves the KIS. The same thank goes to Thomas Berkefeld (Adaptive Optics) and Joachim Staiger (Telescope Control) for keeping the VTT alive. A big "thank you" goes to my Lieblingsbeobachtungsassistent, the now finally retired Hanspeter Schilling. It was always a pleasure to be with you at the VTT. Oliver Wiloth kept me alive during my 5-weeks stay at the VTT by bringing his laptop, having a 2nd hard disk for the PC in the computer room, and, hence, allowing for extended gaming sessions (and thus the Allied won the Battle of Omaha beach, but unfortunately they lost at Stalingrad). I thank Thomas Keller for opening the dome usually a bit later than others. And to Thomas Sonner, I swear I will never touch the hairline mounting again ! Of course, a big thank also goes to Peter Caligari and Reiner Hammer for keeping the computers alive, even if Sokrates was dead at some time (I finally found him on a shelf, poor guy).

I thank my colleagues at the KIS in, close and up to two floors above the "Spielzimmer", in no special order: Daniel Müller, Le Monsieur Docteur Tayeb Aiouaz ("Dasch isch gesuuund !"), Alexandra Tritschler, Kai Langhans, Sven Bingert (Kopf hoch, vielleicht wird's beim nächsten Mal nur ein Kind, immer am Ball bleiben), Petri Whatever Käpylä, Friedrich Wöger, Kasia Mikurda, Christian Hupfer, Markus Roth, Mathieu Ossendrijver, Stefan Jendersie, for the nice events of KIS-Kino, KIS-Fussball, visits to the Kastaniengarten, and too many things more I cannot list.

For the people at the IAC in La Laguna, "Muchas gracias" go to Manolo Collados, whom I tortured with the telescope model, until he gave up and invited me to my first visit on Tenerife, which (unfortunately) was followed by about four visits per year with a total duration of around one month. I thank Valentin Martínez Pillet for teaching me how to hold a wine glass correctly (sorry, I'm german, and we Germans drink beer and eat Worscht and Sauerkraut, you should know that !). In the Max Planck Institut for Solar System Research in Katlenburg-Lindau, I thank Andreas Lagg, Godehard Monecke and Regina Aznar Cuadrado for the intensive "Tischtennis" session. Regina, it's a pity that you already were married when we first met. Good luck with everything ! At the IAA in Granada, I thank Daniel Cabrera for bringing his CDs with him to Tenerife. Otherwise the observations, where we had to switch from AO to CT fifteen times per hour, would have been even more terrible. In Göttingen, I thank Markus Sailer for a nice time, you were not the only one to enjoy your stay in Freiburg. I thank Bruce Lites, David Elmore and Kim Streander at the High Altitude Observatory in Boulder, Colorado (USA), for their support with POLIS.

And of course, a big thank goes to the weekly "Skat-Runde": Michael Engler, Stefan Henninger, and Carsten Baur.

To my friends in Alzenau, a very big thank goes to Hartmut Pachl and Oliver Fiebig, for always having a good time, when I was not in Freiburg. And last but not least, I thank my car for not falling to pieces in thirteen years. But guy, you are loosing oil. And your brakes are squeaking. Please keep on rolling !

For all those I forgot, you are included here: **Thank you.**